Binding by Oscillatory Dynamics in Neural Architectures for Relational Reasoning

Elder G. Veliz, Awni Altabaa, John D. Lafferty

Yale University: Department of Statistics & Data Science

Introduction

Relational reasoning tasks (e.g., same/different judgments or analogies) challenge neural networks due to the lack of explicit binding mechanisms. Inspired by neuroscience, we address the *binding problem* using **oscillatory dynamics**, leveraging neural synchrony to dynamically group features. We introduce a **relational bottleneck** to enforce abstract, generalizable relational representations.

Model Architecture

Our model binds object features through Kuramoto oscillator synchronization:

- **Feature Extraction:** Object images z_i encoded by CNN: $E_i = f_\theta(z_i) \in \mathbb{R}^D$.
- **Oscillatory Binding (Kuramoto Dynamics):** Features drive oscillators $\mathbf{x}_{i,d}(t) \in S^{N-1}$ on a unit sphere, evolving as:

$$\dot{\mathbf{x}}_{i,d}(t) = \mathbf{\Omega}_d \mathbf{x}_{i,d}(t) + \mathrm{Proj}_{\mathbf{x}_{i,d}(t)}\left(\mathbf{c}_{i,d} + \sum_{d'} \mathbf{J}^{\mathrm{IN}}_{d,d'} \mathbf{x}_{i,d'}(t) + \sum_{d'} \mathbf{J}^{\mathrm{OUT}}_{d,d'} \mathbf{x}_{3-i,d'}(t)\right)$$

  Dynamics include:
  - *Natural frequency* ($\mathbf{\Omega}_d$): Skew-symmetric matrices generating intrinsic rotations, ensuring diverse oscillator frequencies and preventing trivial synchronization.
  - *Conditional input* ($\mathbf{c}_{i,d} = W_d \cdot E_i[d] + \mathbf{b}_d$): Encodes object-specific features into oscillator dynamics, guiding task-relevant synchronization patterns. W_d, $\mathbf{b}_d$ are learnable parameters.
  - *Within-object coupling* ($\mathbf{J}^{\mathrm{IN}}_{d,d'}$): Synchronizes features within the same object.
  - *Between-object coupling* ($\mathbf{J}^{\mathrm{OUT}}_{d,d'}$): Synchronizes features across different objects.
- **Relational Bottleneck:** Coherence measures synchronization strength:

$$\rho_d = \|\mathbf{x}_{1,d}(T) + \mathbf{x}_{2,d}(T)\|_2$$

- **Classification:** Classifier predicts relationship based on coherence signals.

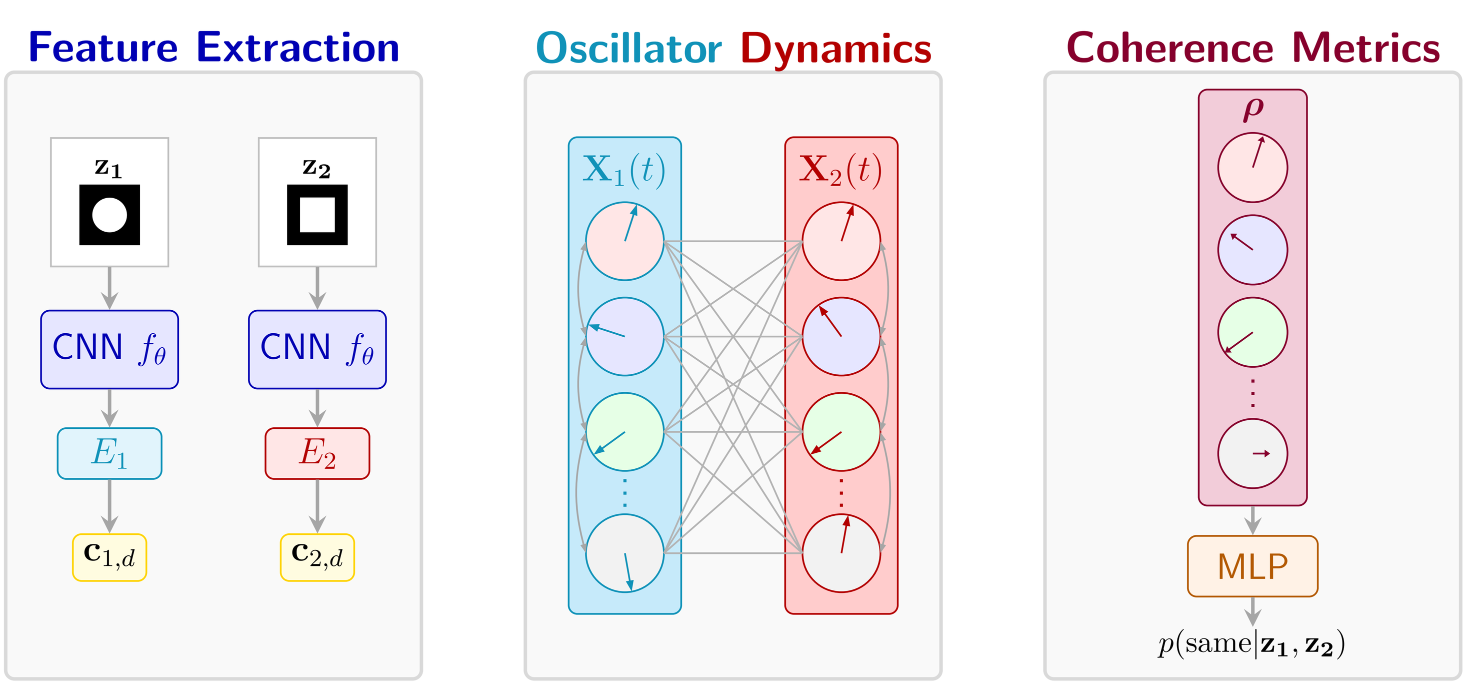


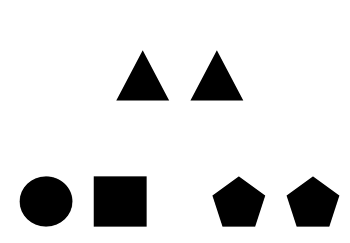
Figure 1. Model architecture overview.

Tasks

We evaluate on two visual reasoning tasks, using 100 Unicode characters with random scaling/positioning:

- **Same/Different:** Decide if two objects belong to the *same* category.
- **Relational Match-to-Sample (RMTS):** Given a source pair (defining "same" or "different"), choose which of two target pairs matches that relation. This is reasoning about *relations of relations*.
- **Generalization Regimes:** Train on $m \in \{95, 50, 15\}$ icons, test on the remaining $\{5, 50, 85\}$.

(a) Same/Different



(b) RMTS

Figure 2. Task Illustrations.

Transfer Learning Approach

- **Stage 1:** Train the full CNN + Oscillator network on the Same/Different task.
- **Stage 2:** Freeze CNN and oscillator parameters; extract coherence vectors for each pair and train a new MLP on the RMTS task.

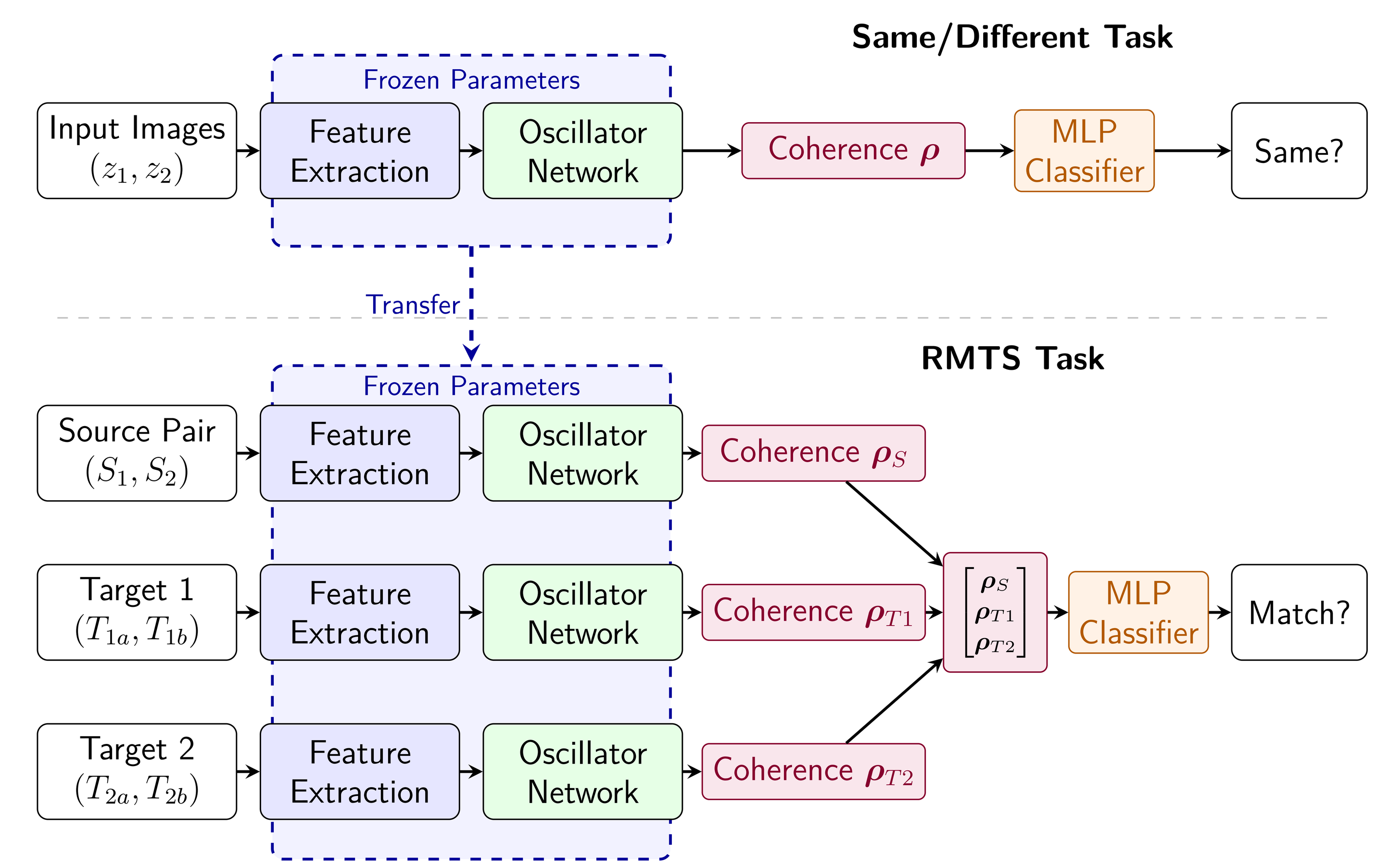
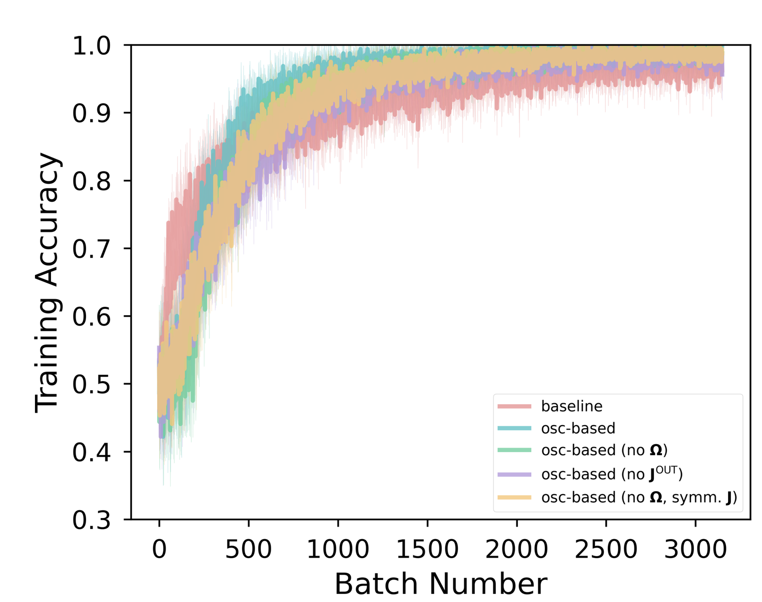
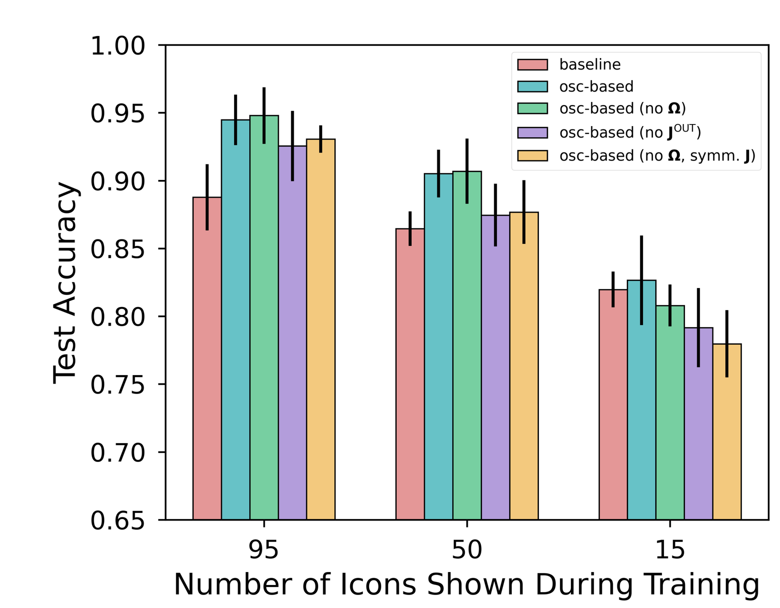


Figure 3. Two-stage transfer learning paradigm.

Same/Different Performance

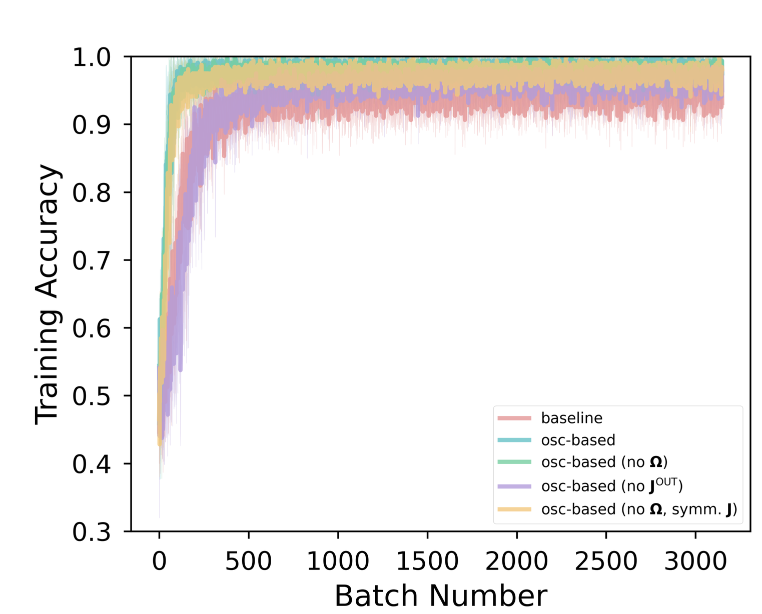


(a) Training accuracy over batches ($m = 95$)

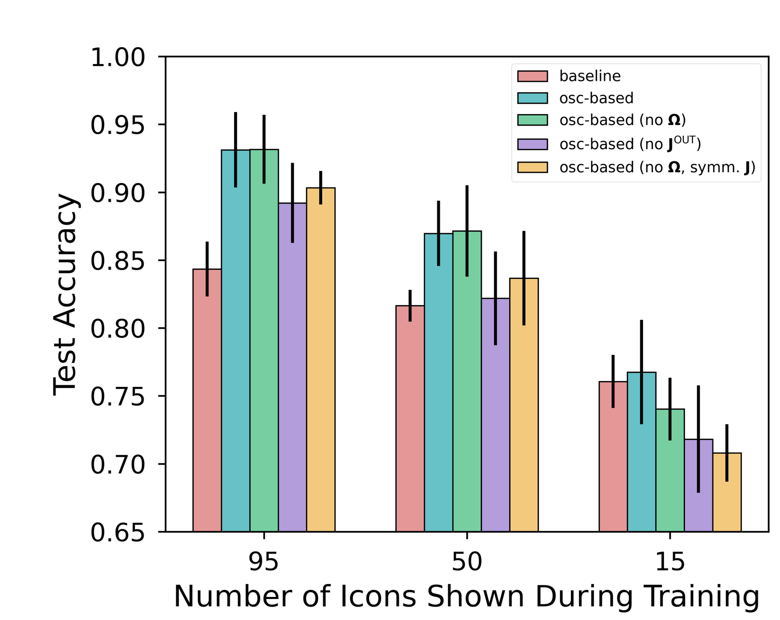


(b) Test accuracy across generalization regimes

RMTS Performance



(a) RMTS training accuracy ($m = 95$)



(b) RMTS test accuracy across regimes

Energy Minimization

Under symmetric coupling ($\mathbf{J} = \mathbf{J}^\top$) and zero natural frequencies ($\mathbf{\Omega} = 0$), the Kuramoto network minimizes the Lyapunov energy

$$\begin{aligned} E &= -\tfrac{1}{2}\sum_{i,j=1}^{2} \mathbf{x}_i^\top \mathbf{J}_{ij} \mathbf{x}_j - \sum_{i=1}^{2} \mathbf{c}_i^\top \mathbf{x}_i \\ &= -\tfrac{1}{2}\Big(\mathbf{x}_1^\top \mathbf{J}^{\mathrm{IN}} \mathbf{x}_1 + \mathbf{x}_2^\top \mathbf{J}^{\mathrm{IN}} \mathbf{x}_2 + \mathbf{x}_1^\top \mathbf{J}^{\mathrm{OUT}} \mathbf{x}_2 + \mathbf{x}_2^\top \mathbf{J}^{\mathrm{OUT}} \mathbf{x}_1\Big) - \left(\mathbf{c}_1^\top \mathbf{x}_1 + \mathbf{c}_2^\top \mathbf{x}_2\right). \end{aligned}$$

During each integration step, E decreases monotonically. Empirically, "same" input pairs converge to *lower* energy minima than "different" pairs.

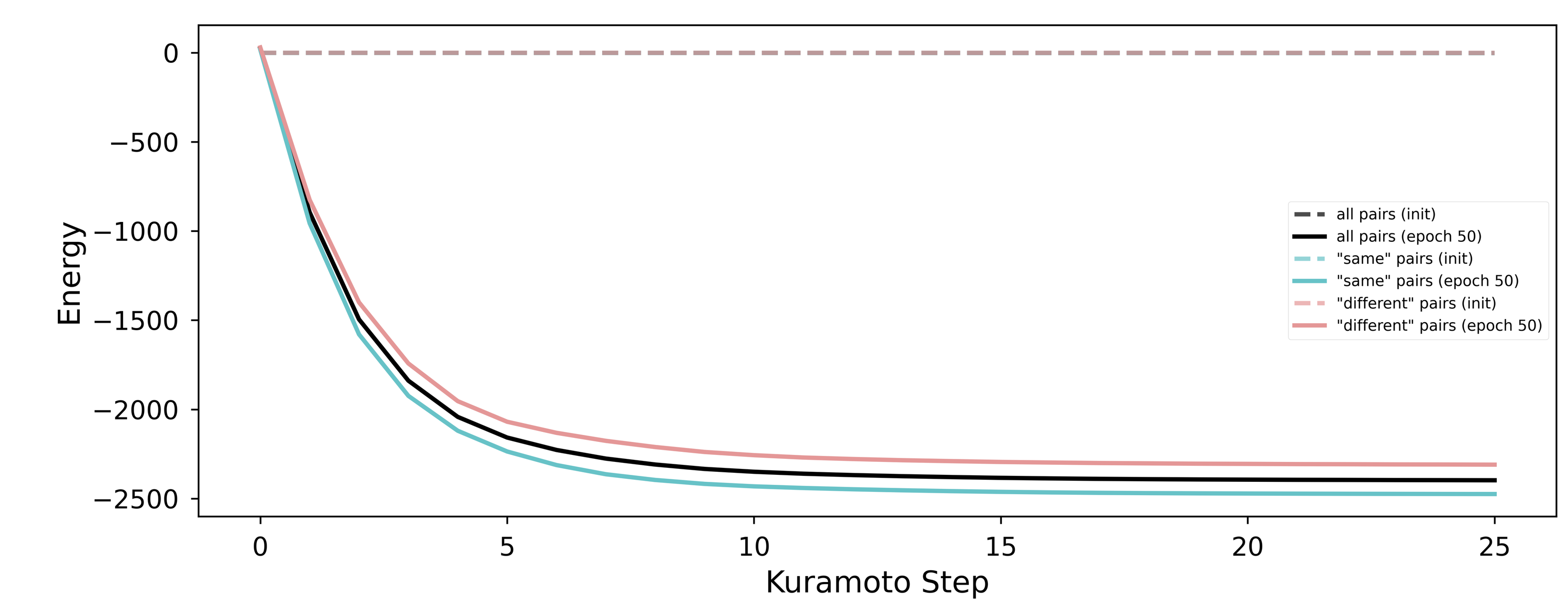
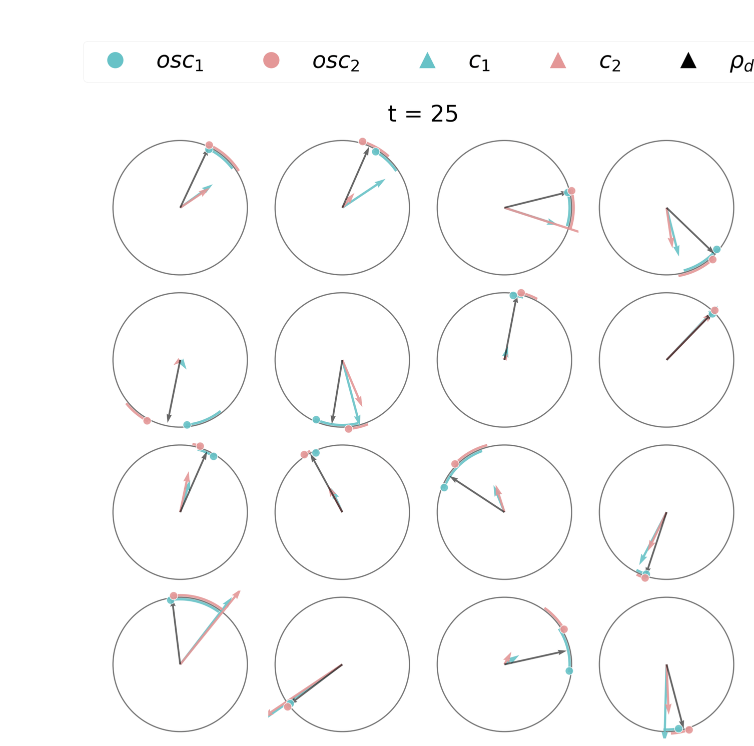


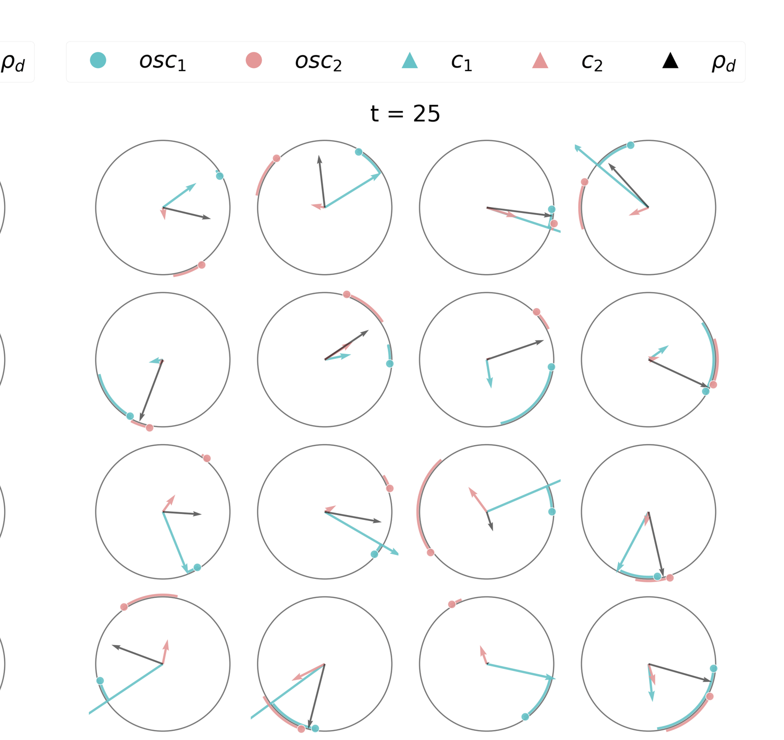
Figure 6. Average energy trajectory for "same" vs "different" pairs over $T = 25$ steps.

Oscillator State Dynamics

- **Initialization:** All D oscillators begin with random phases on the circle.
- **Trained "Same" Pair:** For every dimension d, the two corresponding oscillators lock in phase—i.e. they overlap as tight pairs on the circle.
- **Trained "Different" Pair:** Only shared–feature dimensions phase–lock; others remain asynchronous, producing a mix of tight and dispersed pairs.



(a) "Same" pair oscillators (epoch 50)



(b) "Different" pair oscillators (epoch 50)

Figure 7. All D oscillator phases at $t = T$ for sample "same" vs. "different" pairs.

References

[1] Awni Altabaa, Taylor Webb, Jonathan Cohen, and John Lafferty. Abstractors and relational cross-attention: An inductive bias for explicit relational reasoning in transformers, 2024.

[2] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In Huzihiro Araki, editor, *International Symposium on Mathematical Problems in Theoretical Physics*, pages 420–422, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[3] Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial kuramoto oscillatory neurons, 2025.

[4] Taylor W. Webb, Steven M. Frankland, Awni Altabaa, Simon Segert, Kamesh Krishnamurthy, Declan Campbell, Jacob Russin, Tyler Giallanza, Zack Dulberg, Randall O'Reilly, John Lafferty, and Jonathan D. Cohen. The relational bottleneck as an inductive bias for efficient abstraction, 2024.

[5] Taylor W. Webb, Ishan Sinha, and Jonathan D. Cohen. Emergent symbols through binding in external memory, 2021.

Github: Elder453

Poster Session

Email: elder.veliz@yale.edu