Geometric Priors for Generalizable World Models via Vector Symbolic Architecture

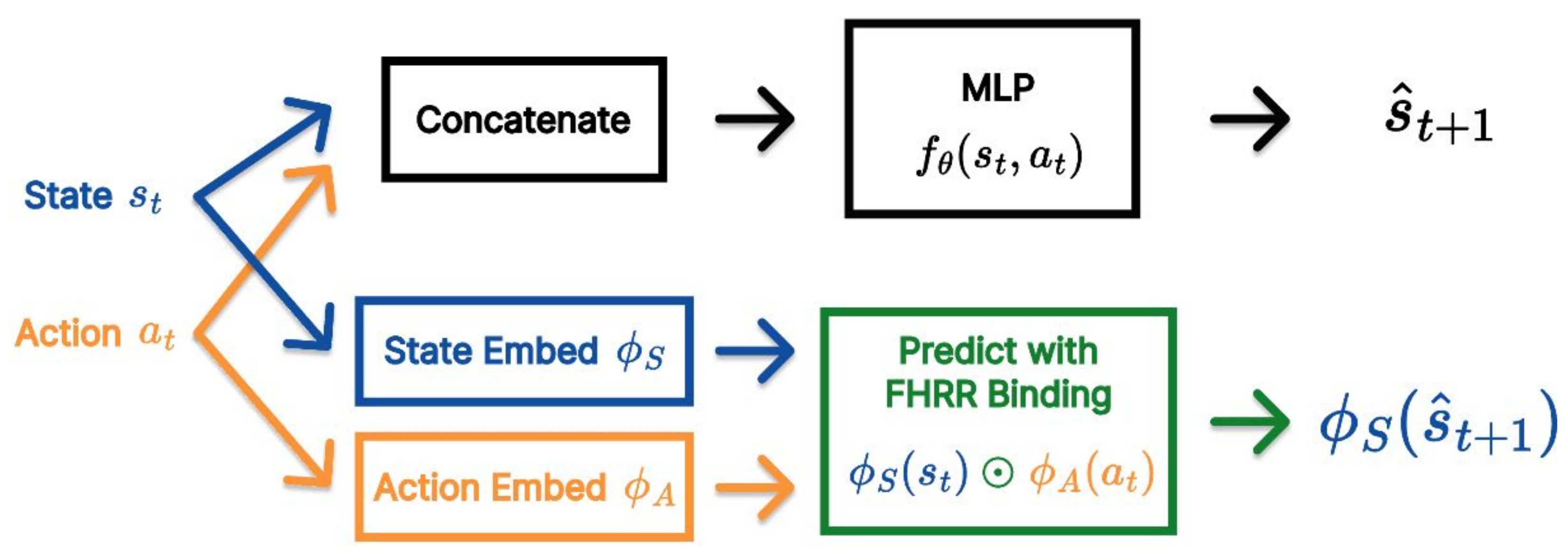
William Youngwoo Chung[1], Calvin Yeung[1], Hansen Jin Lillemark[2]
Zhuowen Zou[1], Xiangjian Liu[1], Mohsen Imani[1]
University of California, Irvine[1], University of California, San Diego[2]

NEURAL INFORMATION PROCESSING SYSTEMS

Unstructured Latent Spaces for World Modeling leads to Poor Generalization and Sample Efficiency

Most world models approximate a transition functions with large unstructured neural networks. They fail to capture underlying symmetries leading to poor sample efficiency and generalization to unseen states.

- Incorporating principles from ***Vector Symbolic Architectures*** (VSA) [1] can help enforce structure in the latent space to help with generalization.
- Specifically, we base our model off ***Fourier Holographic Reduced Representations*** (FHRR) [2] embeddings
- We define transition function as binding between state and action embeddings.



Latent Geometric Structure through FHRR

FHRR Representation:

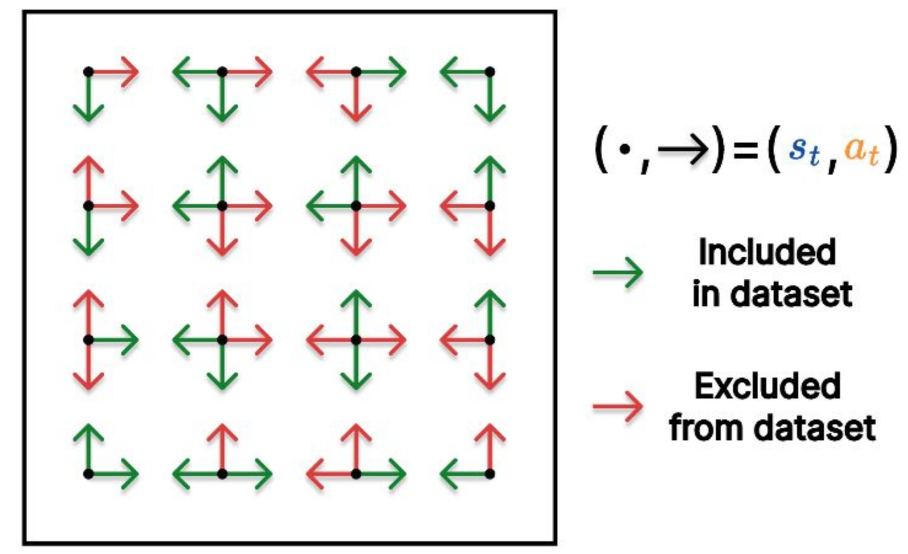
- Encoders: We map states and actions to high-dimensional unitary complex vectors via learnable complex encoders:
  - $\phi_S(s) = [e^{i\theta_{j,s}^\top s}]_{j=1}^D, \quad \phi_A(a) = [e^{i\theta_{j,a}^\top a}]_{j=1}^D.$
- Dynamics as Binding: Latent transitions are models as *binding* (element-wise complex multiplication), which act as rotations in the complex plane
  - Transition: $\phi_S(s_{t+1}) = \phi_S(s_t) \odot \phi_A(a_t).$

Training Objectives: We train the encoders using three loss function to enforce VSA structure:

- **Binding**: $\mathcal{L}_{\text{bind}} = \|\phi_S(s_{t+1}) - \phi_S(s_t) \odot \phi_A(a_t)\|^2$
- **Invertibility**: $\mathcal{L}_{\text{inv}} = \sum_{(a,a^{-1})} \|\phi_A(a) \odot \phi_A(a^{-1}) - \mathbf{1}\|^2,$
- **Orthogonality**: $\mathcal{L}_{\text{ortho}} = \sum_{i \neq j} (\langle \phi_S(s_i), \phi_S(s_j) \rangle)^2.$

Dynamics Modeling Experiments on GridWorld

Setup: A 10x10 discrete Grid World (100 states, 4 actions)
Challenge: Train on 80% of transitions, test on the held-out 20% (Zero-Shot)

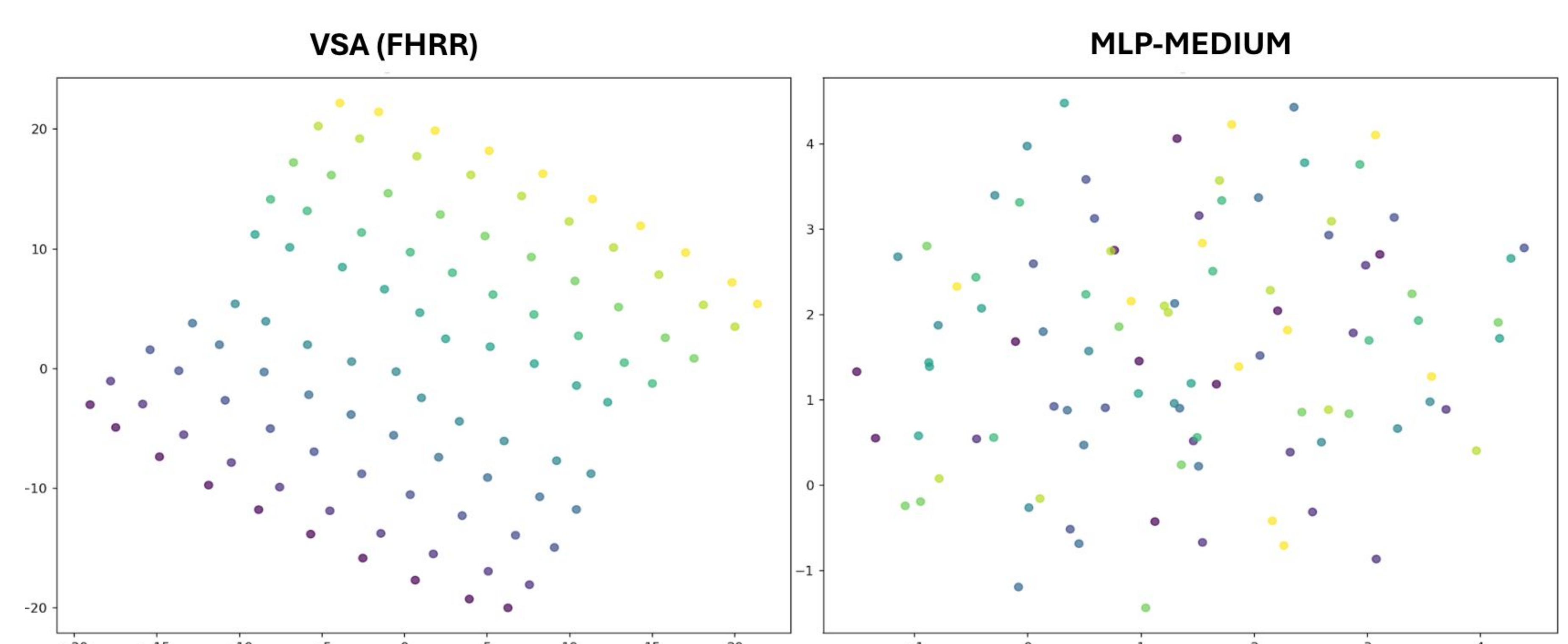


FHRR-based World Model outperforms MLP-based models on Grid World.

Task	FHRR (Ours)	MLP-Small	MLP-Medium	MLP-Large
1-step Accuracy	**96.3%**	80.0%	80.0%	80.25%
1-step Accuracy (Zero-Shot)	**87.5%**	0.0%	0.0%	1.25%
Cosine Similarity	**83.0**	79.5	79.9	80.6
Cosine Similarity (Zero-Shot)	**80.5**	0.9	0.15	3.1
Rollout (5 steps)	**74.6%**	39.8%	38.0%	40.8%
Rollout (20 steps)	**34.6%**	2.0%	4.0%	6.2%
Rollout (20 steps + Clean)	**61.4%**	5.4%	7.8%	8.4%
Rollout (100 steps)	1.8%	0.8%	1.8%	**2.0%**
Rollout (100 steps + Clean)	**38.6%**	2.8%	4.0%	3.2%

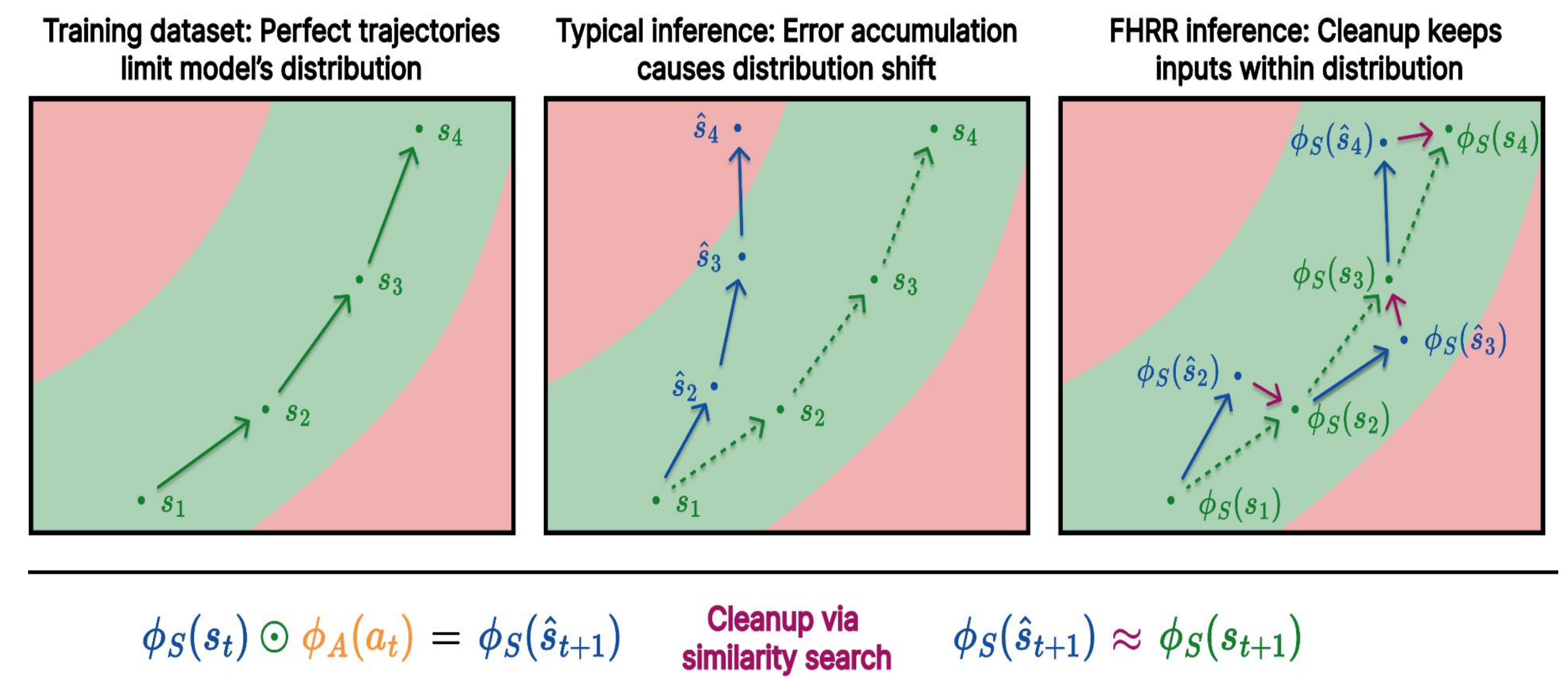
- FHRR-based model outperforms all MLP variants regardless on scale
- MLP models struggle on Zero-Shot tests
- Cleanup shows strong performance increase on Rollout tests for FHRR but not for MLP

t-SNE Visualization of Learned State Embeddings

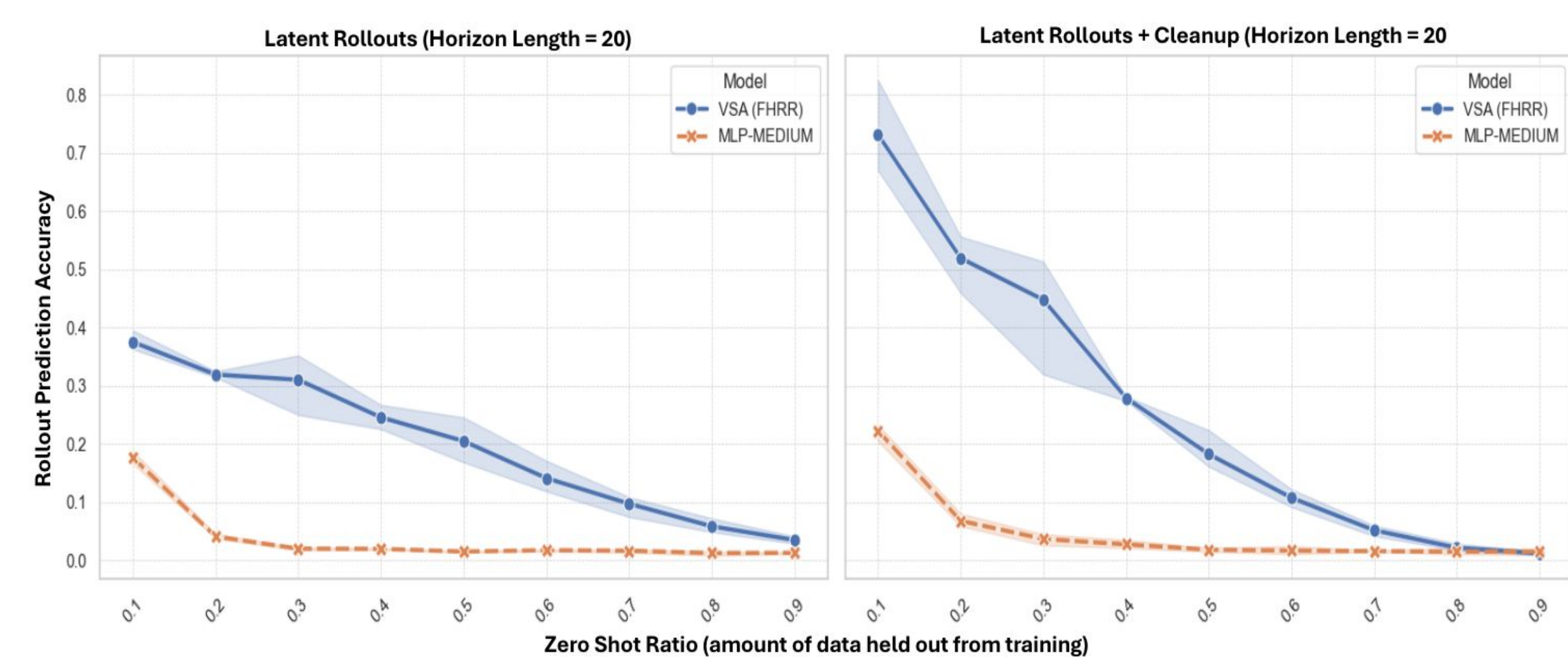


Cleanup: Correcting Drift in Latent Space

- In standard models, small prediction errors accumulate over time causing predictions to drift over long horizons
- In VSA, vectors are high-dimensional and quasi-orthogonal and a noisy prediction remains closer to the true state than other states [3].
- We "clean" the prediction by taking the state most similar to the prediction via:
  - $\phi_S(\hat{s}_{t+1}) = \arg\max_{s \in \mathcal{S}} \text{Re} \langle \phi_S(\hat{s}_{t+1}), \phi_S(s) \rangle$



Latent Rollouts with Cleanup Results (L = 20)



- Performing the cleanup helps with latent rollout performance especially at smaller zero-shot ratios

References

[1] Pentti Kanerva (2009). "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors." In: Cognitive computation, 1:139–159.

[2] Tony A Plate (2003). "Holographic Reduced Representation: Distributed representation for cognitive structures." In: CSLI Publications Stanford volume 150.

[3] Tony Plate et al. (1991). "Holographic reduced representations: Convolution algebra for compositional distributed representations." In: IJCAI, pages 30–35.