Understanding and Controlling Memory in RNN

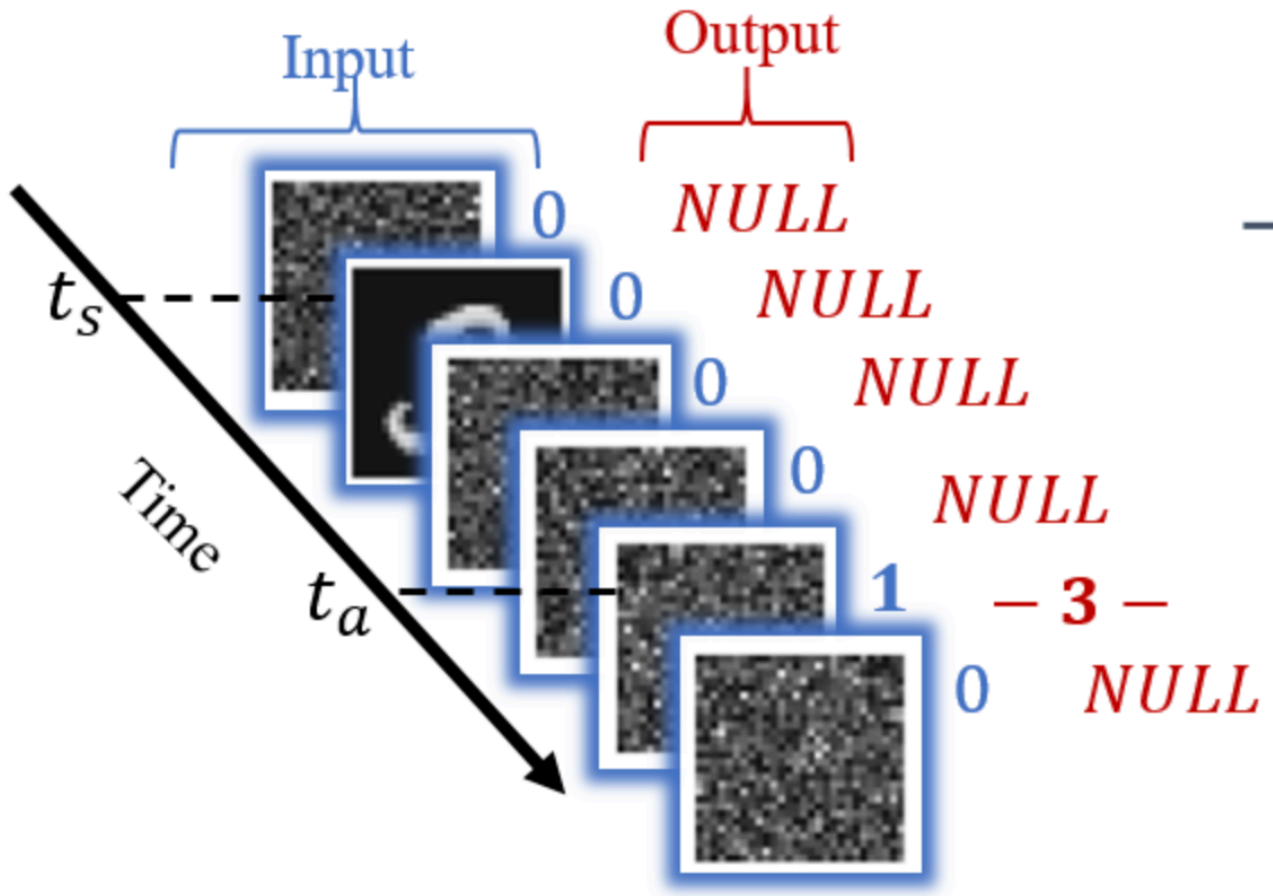
D. Haviv, A. Rivkind, O. Barak

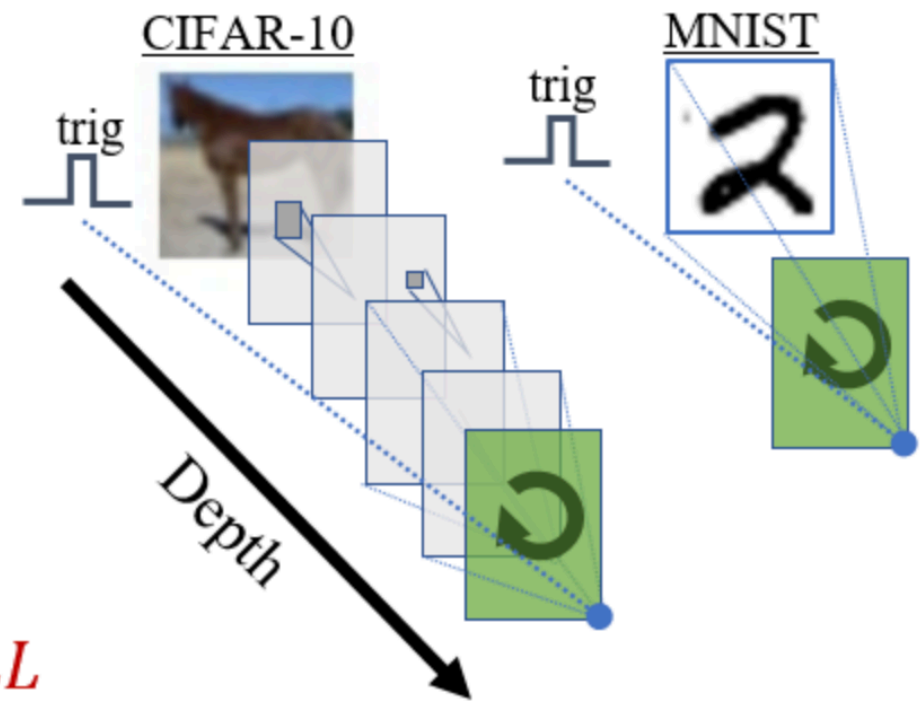
Network Biology Research Laboratories
Technion – Israel Institute of Technology

Objectives

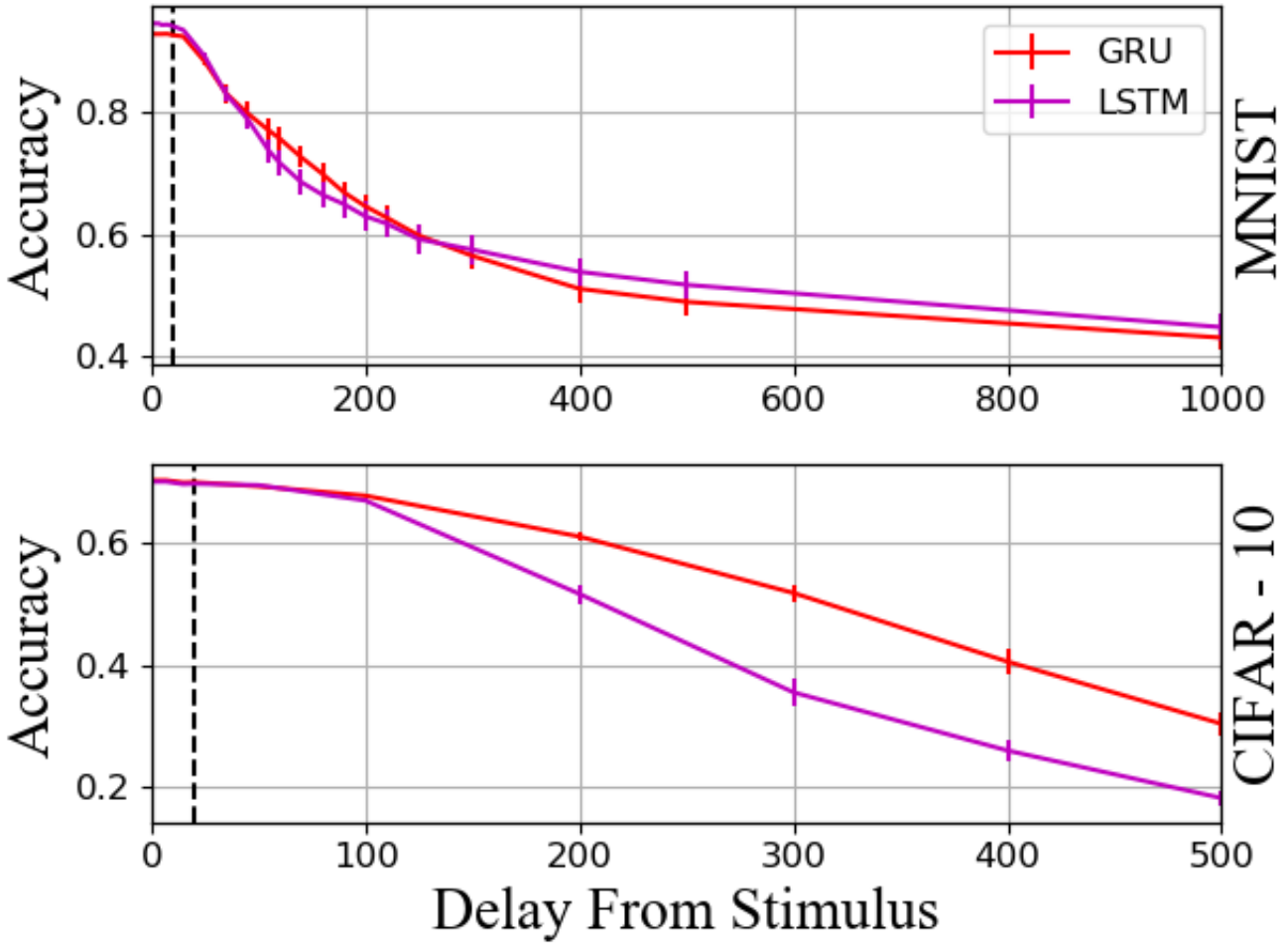
- RNNs are trained only for limited timesteps – Can they form long term memories?
- How are these memories (short or long-term) represented as dynamical objects?
- Can these dynamical objects be manipulated to explicitly demand long term memorization?

Task Definition



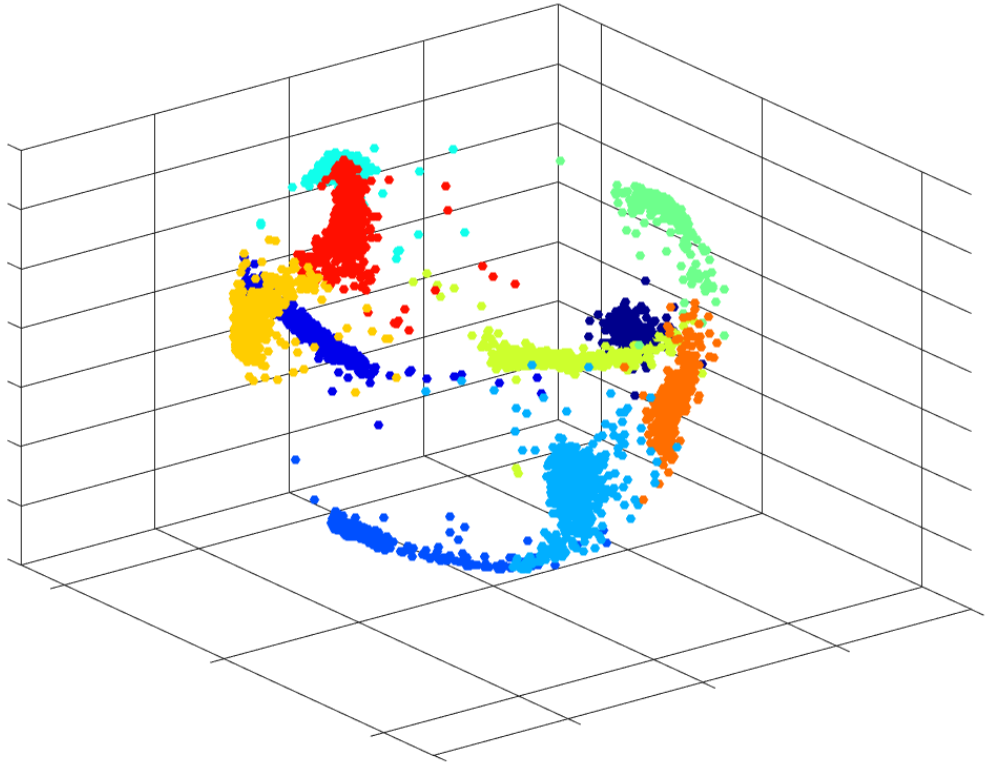


Can RNN Form Long-Term Memories?



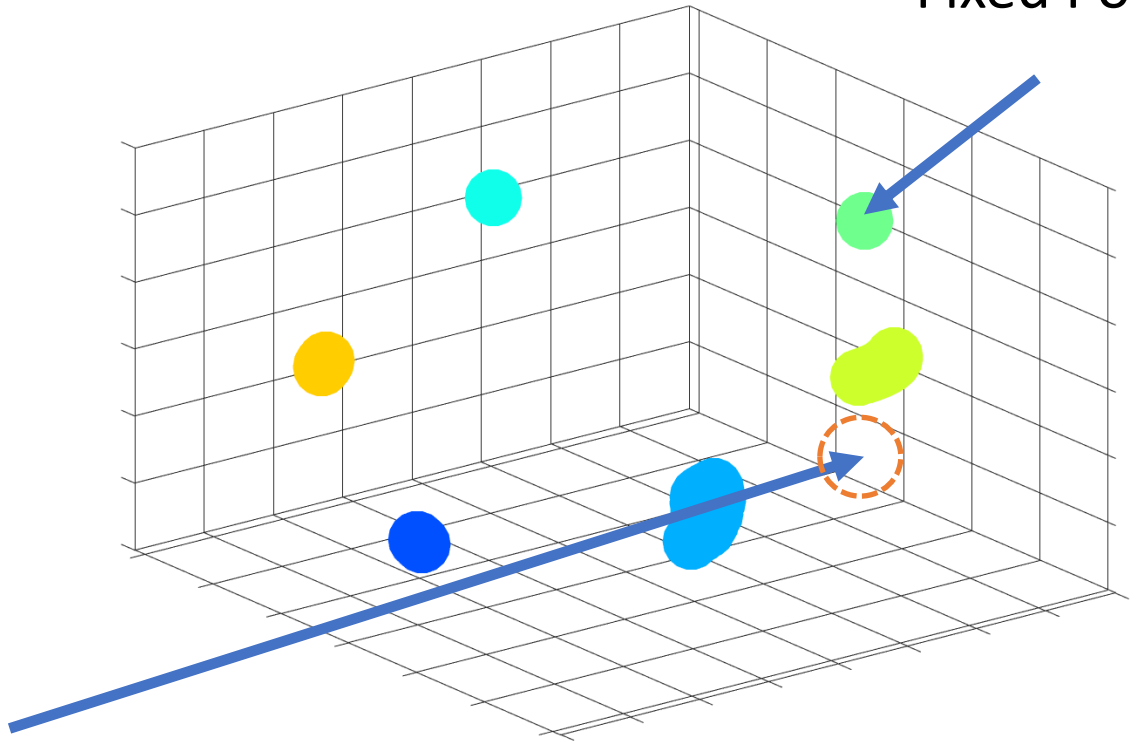
Can RNN Form Long-Term Memories?

20 Timesteps



1000 Timesteps

Fixed Point

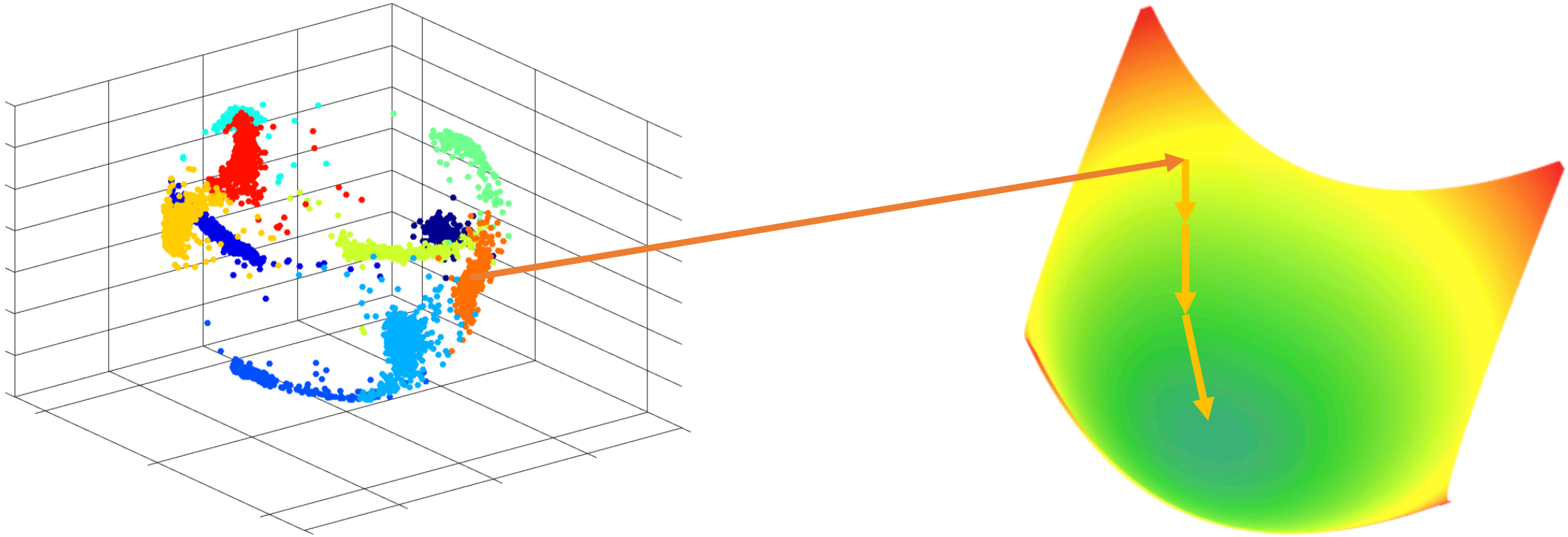


Slow Point?

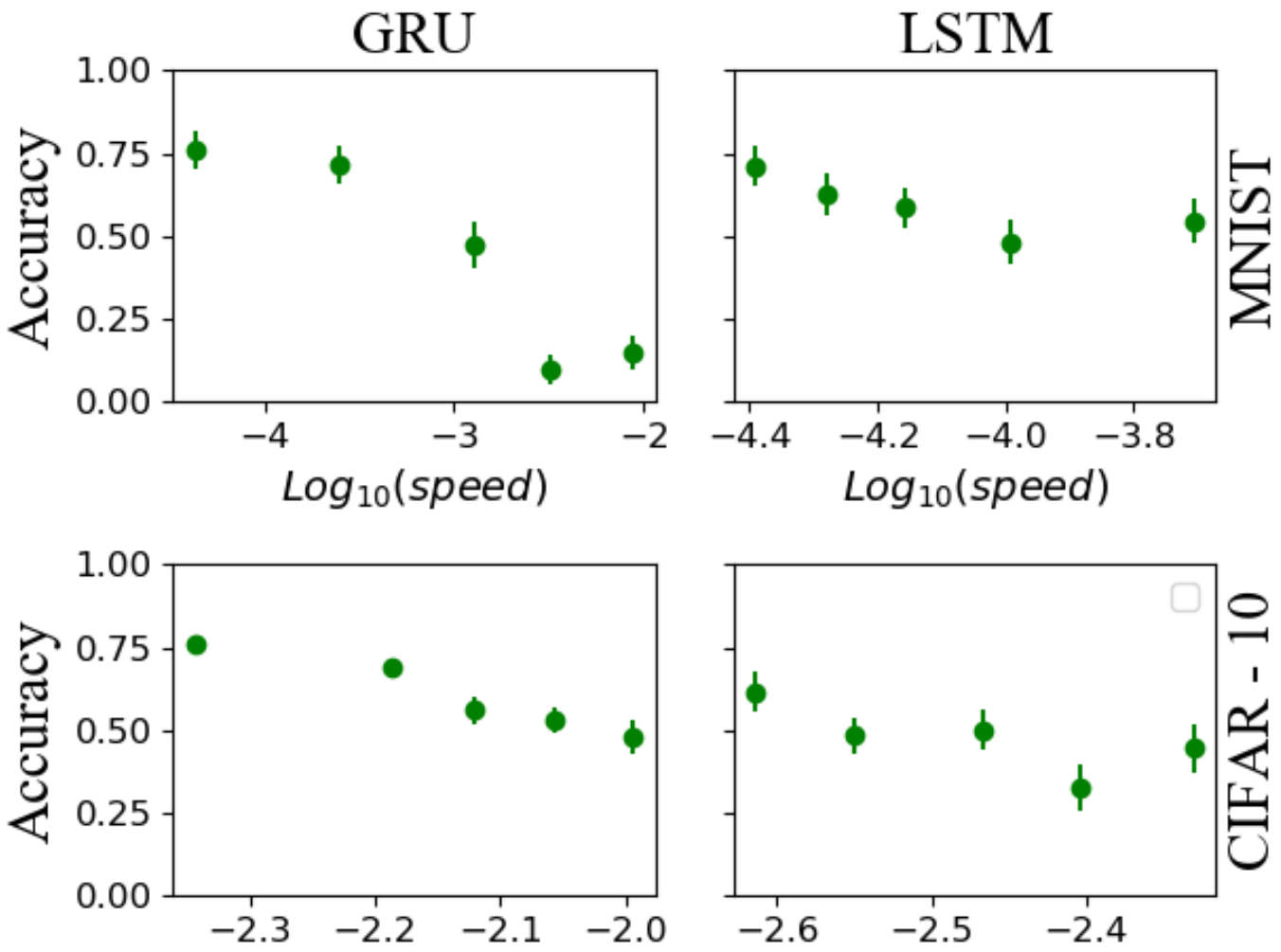
Slow-Points and How to Find Them

$$S(h_t, I) = \|h_{t+1} - h_t\|_2^2$$

$$\hat{h}_n = \hat{h}_{n-1} - \nabla S(h, I_\mu)\Big|_{\hat{h}_{n-1}}$$



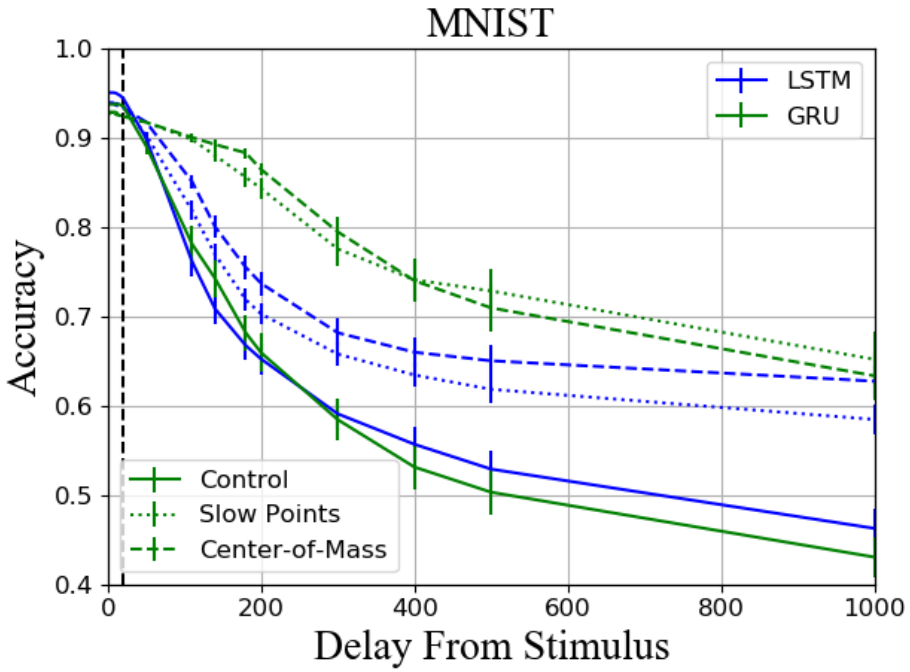
Slow-Point Speed Predicts Memory Robustness

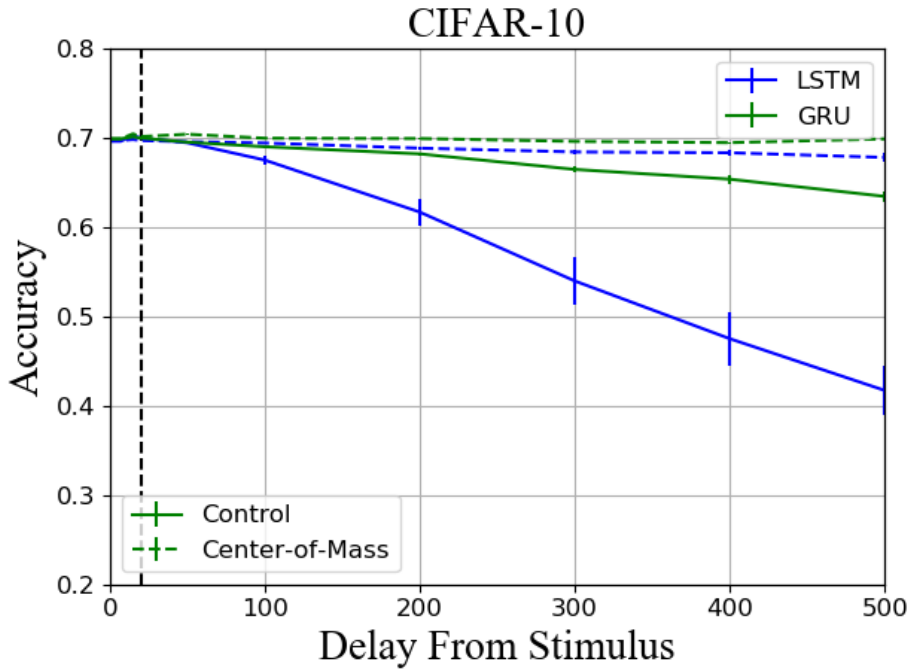


Regularize Speed for Long-Term Memories

Fine-tuning with modified loss:

$$\hat{L} = L_{CE} + \lambda \sum_{i \in V} S(h_i, I)$$





Key Findings

- RNNs can form long term memories, but not all memories are created equal
- Slow-Point speed is quantitatively correlated to memory robustness
- We can explicitly demand long-term memorization by regularizing the hidden-state speed

Thanks for Listening!

Poster **#258** at Pacific Ballroom

Code: https://github.com/DoronHaviv/MemoryRNN