Recurrent Neural Networks as a Model to Probe Neuronal Timescales Specific to Working Memory

Robert Kim        Terrence J. Sejnowski

Computational Neurobiology Laboratory, Salk Institute
University of California, San Diego (UCSD)

NeurIPS Workshop

December 14, 2019
Motivation

· Neurons important for working memory (WM) have stable and long neuronal timescales
Motivation

- Neurons important for working memory (WM) have stable and long neuronal timescales
- Experimentally challenging to probe circuitry and connectivity mechanisms
Motivation

- Neurons important for working memory (WM) have stable and long neuronal timescales
- Experimentally challenging to probe circuitry and connectivity mechanisms
- Recurrent neural network (RNN) model can be useful
Motivation

- Neurons important for working memory (WM) have stable and long neuronal timescales
- Experimentally challenging to probe circuitry and connectivity mechanisms
- Recurrent neural network (RNN) model can be useful

Circuit/network mechanisms required for stable temporal receptive fields critical for WM maintenance
• **Continuous RNNs** converted to **Leaky Integrate-and-Fire (LIF) RNNs** [1]
· **Continuous RNNs** converted to **Leaky Integrate-and-Fire (LIF) RNNs** [1]

· 40 RNNs \((N = 200)\) trained to perform a **delayed match-to-sample (DMS)** task
- **Continuous RNNs** converted to **Leaky Integrate-and-Fire (LIF) RNNs** [1]

- 40 RNNs ($N = 200$) trained to perform a **delayed match-to-sample (DMS)** task

- 80% **excitatory** and 20% **inhibitory**
SPIKING RNN MODEL

[Graphs showing spiking neuron activity over time, with excitatory and inhibitory outputs depicted.]
Experimental Data

- Public dataset (crcns.org) – Constantinidis lab [2–4]
• Public dataset (crcns.org) – Constantinidis lab [2–4]

• **Four monkeys** trained to perform two delayed match-to-sample tasks: spatial and feature tasks
Public dataset (crcns.org) – Constantinidis lab [2–4]

Four monkeys trained to perform two delayed match-to-sample tasks: spatial and feature tasks

959 dorsolateral prefrontal cortex (dlPFC) units
Spiking RNN Model

Cue: 0.25 s
Delay: 0.75 s
Sample: 0.25 s

Experimental Data

Fixation: 1 s
Cue: 0.5 s
Delay: 1.5 s
Sample: 0.5 s

Saccade
• Spike-count autocorrelation during **fixation**
• Spike-count autocorrelation during **fixation**

• Spike counts in successive time bins \((w = 50 \text{ ms})\)
Neuronal Timescales

- Spike-count autocorrelation during **fixation**
- Spike counts in successive time bins ($w = 50$ ms)
· Spike-count autocorrelation during fixation
· Spike counts in successive time bins ($w = 50$ ms)
· Correlation between two time bins separated by a lag ($\Delta$)
Neuronal Timescales

The figure illustrates the relationship between time (ms) and the number of spikes across different trials. The top plot shows the time course of spikes over time, with each trial represented by a vertical bar. The bottom plot depicts the autocorrelation function (\(\bar{\rho}\)) as a function of time lag (\(\Delta\) ms), with the color scale indicating the number of spikes per trial.

- **Time (ms)**: 0, 200, 400, 600, 800, 1000
- **Trials**: 0, 1, 2
- **Number of Spikes**: 0, 1, 2
- **Autocorrelation (\(\bar{\rho}\))**: -0.2, -0.1, 0, 0.1, 0.2, 0.3
- **Time Lag, \(\Delta\) (ms)**: 0, 100, 300, 500
Neuronal Timescales

![Graph showing neuronal timescales with trials on the x-axis and time (ms) on the y-axis, along with autocorrelation values.](image)
Neuronal Timescales
\[ \bar{\rho}(\Delta) = A \left( \exp \left( -\frac{\Delta}{\sigma} \right) + B \right) \]
\[ \bar{\rho}(\Delta) = A \left( \exp \left( -\frac{\Delta}{\sigma} \right) + B \right) \]
Heterogeneous Neuronal Timescales

![Graphs showing autocorrelation and density distributions for Short and Long timescales for dIPFC and RNNs.](image)

- **Short Timescale**
  - dIPFC: Median 49 ms
  - RNN: Median 66 ms

- **Long Timescale**
  - dIPFC: Median 243 ms
  - RNN: Median 217 ms

**Untrained RNNs**
- Median 20 ms

**Trained RNNs**
- Median 120 ms

**dIPFC**
- Median 114 ms
LONG $\sigma$ UNITS INVOLVED WITH STABLE CODING

![Graph showing dIPFC Short $\sigma$ with Time (s) on the x-axis and Discriminability (au) on the y-axis.]
LONG $\sigma$ UNITS INVOLVED WITH STABLE CODING

Independent Split A

Independent Split B

dIPFC Short $\sigma$

dIPFC Long $\sigma$

Time (s)

Cue
Delay

Discriminability (au)

0 0.5 1.0

0 0.5 1.0

0 0.5 1.0

0 0.5 1.0
LONG $\sigma$ UNITS INVOLVED WITH STABLE CODING
WM-specific Neuronal Timescales

DMS

Median 120 ms

Density

Timescale (ms)
WM-specific Neuronal Timescales

DMS

Median 120 ms

Alternative Forced Choice (AFC)
WM-specific Neuronal Timescales

DMS

AFC

Median 46 ms

Median 120 ms
**WM-specific Neuronal Timescales**

![Graph showing normalized autocorrelation against time lag for different brain areas: AFC, DMS, PFC.](image-url)
Striking similarities b/w RNN model of WM and experimental data
• Striking similarities b/w RNN model of WM and experimental data

• Both utilize units with stable temporal receptive fields to perform WM
· Striking similarities b/w RNN model of WM and experimental data

· Both utilize units with stable temporal receptive fields to perform WM

· Need to characterize network/circuit dynamics that lead to long timescales

