


Discrete attractor dynamics underlies persistent activity in the frontal cortex¹

Reproducing Neuron Dynamics with Highly Structured and Trained
Chaotic Random RNN Models

Elena Zoppellari

Final Project for Physical Models of Living Systems course

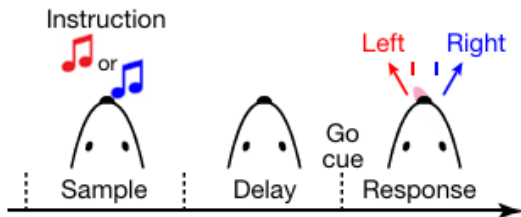
¹Inagaki, H.K., Fontolan, L., Romani, S. *et al.* *Nature* **566**, 212–217 (2019). 

Motivation

- **Persistent activity**: neural firing that continues after the triggering stimulus goes away
- Persistent activity supports short-term memory and motor planning.
- Inagaki et al. (2019): **network dynamics**, rather than intrinsic cell properties, sustain short-term memory.
→ **Structured discrete attractor models** match experimental data.
- What happens if we train a **random** recurrent network (RNN) to reproduce short-term activity observed during a **delayed paired-association task**?

Experimental Task (Inagaki et al., 2019)

- Delayed-response licking task (left/right)
- Task epochs: Sample → Delay → Go cue
- Neural recordings from the ALM (anterior lateral motor cortex) using silicon probes
- Persistent activity observed during the delay epoch



Excluding Cell-Autonomous Mechanisms

- 1 **Membrane time constants** similar in selective and non-selective neurons, not large enough to sustain activity ($\tau_{memb} \sim 20$ ms vs $\tau_{delay} \sim 2$ s)
- 2 **Spike bursts** could activate voltage-dependent channels, but they are rare and did not increase during delay
- 3 **Conductances activated by depolarization:** hyperpolarizing the cells don't remove cell's selectivity
→ single cell mechanisms are not enough to explain persistent activity.

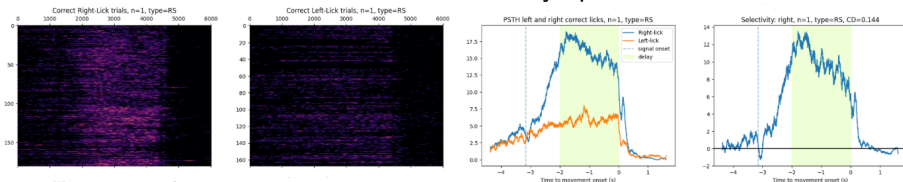
Project Roadmap

- 1 Organize data from Inagaki et al.'s extracellular recordings and replicate the Coding Direction (CD) analysis
- 2 Reproduce their results using the structured attractor model
- 3 Implement the FORCE learning algorithm, following Rajan et al. (2016)²
- 4 Train a random RNN with FORCE using the Inagaki's extracellular recording data
- 5 Analyze and compare the results

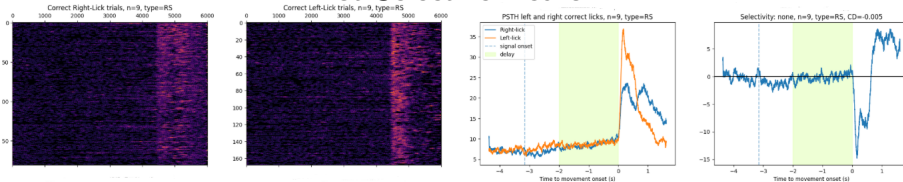
²Rajan, K., Harvey, C.D., Tank, D.W. Recurrent network models of sequence generation and memory. *Neuron*, **90**, 128–142 (2016).

Neuron Selectivity

Selective Neuron: spike rate significantly different between correct R and correct L trials in delay epoch



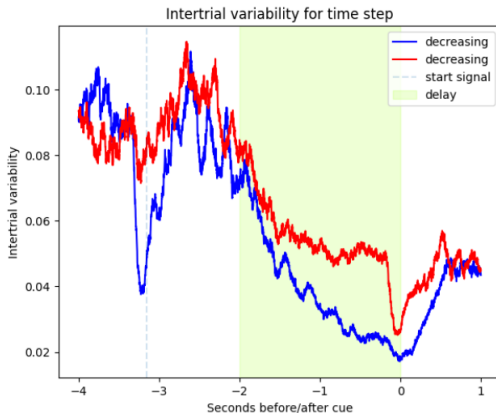
Not Selective Neuron



(Spikes averaged in 100ms time window)

Intertrial Variability During Delay

- Analyzed across-trial variability for correct L/R trials using variance in selective neurons
- Found that variability **decreases** during delay → signature of discrete attractor



Coding Direction (CD) Method

- Map from \mathbb{R}^N to \mathbb{R}
- **CD** is calculated using 50% of **correct** trials (*training*)
- **CD** is the weighted max difference in delay epoch between left and right trials

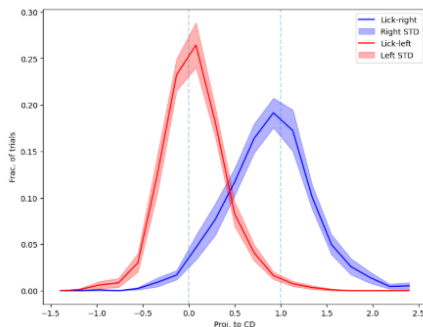
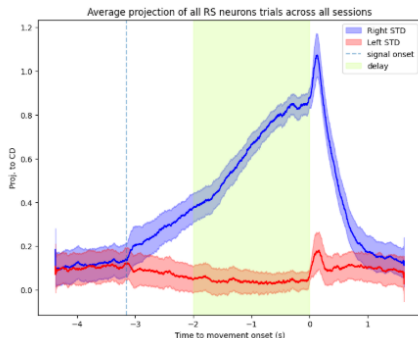
$$\mathbf{CD} = \frac{\bar{\mathbf{r}}_{\text{right}} - \bar{\mathbf{r}}_{\text{left}}}{\|\bar{\mathbf{r}}_{\text{right}} - \bar{\mathbf{r}}_{\text{left}}\|} \in \mathbb{R}^N \quad (1)$$

- For each trial, population activity projected as:

$$\text{proj}_{\mathbf{CD},t} = \langle \mathbf{x}_t^{\text{test}}, \mathbf{CD} \rangle \in \mathbb{R} \quad (2)$$

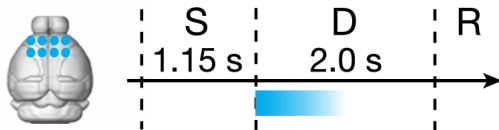
CD Projections in Data

- Left and right trials are maximally differentiated in CD space
- Endpoint distributions are peaked
- Ramping activity can be eliminated performing random delay sessions



Photoinhibition Experiment

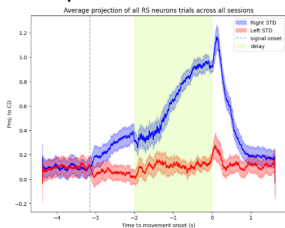
- **Perturbation of persistent activity:** how does neurons activity change during correct/incorrect trials?
- Blue light used to activate inhibitory PV neurons in ALM
- Silenced excitatory activity during the first 0.6s of delay epoch



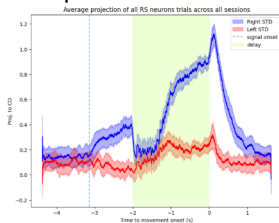
- Mice **recovered correct response** or showed **attractor switches**
→ **2 discrete attractors**

Photoinhibited Correct Trials

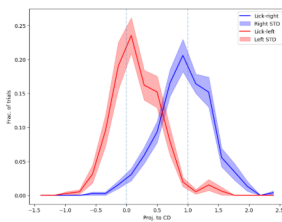
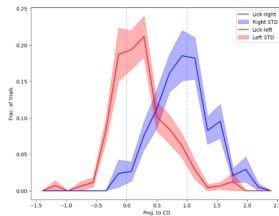
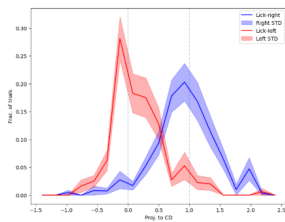
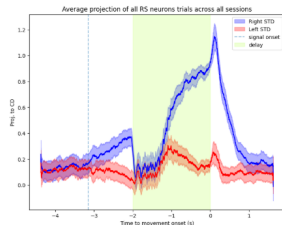
ph = 0.1mW



ph = 0.2mW

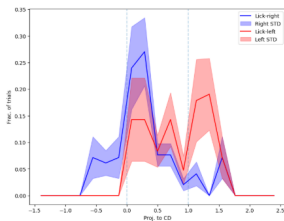
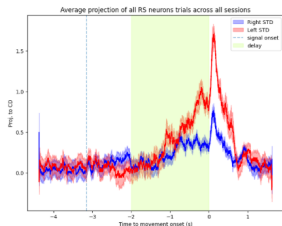


ph = 0.3mW

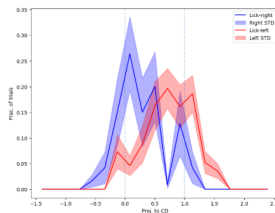
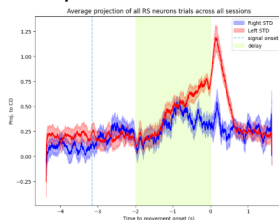


Photoinhibited Incorrect Trials

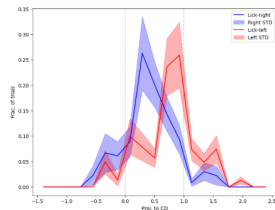
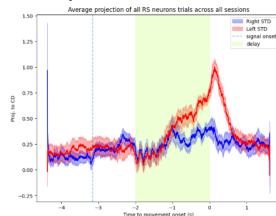
ph = 0.1mW



ph = 0.2mW

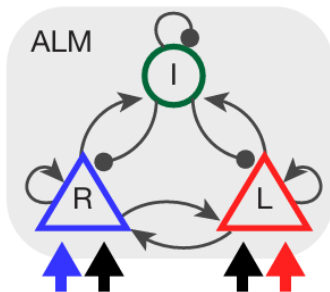


ph = 0.3mW



Three-Populations Network (Inagaki et al.) (1)

- Excitatory Left (L), Excitatory Right (R), Inhibitory (I)
- Recurrent and inhibitory populations dynamics create **stable attractors** with **fixed parameters**



$$\tau_i \frac{dh_i(t)}{dt} = -h_i(t) + \sum_{j=L,R} \tilde{W}_{ij} G_E(h_j(t)) + \tilde{I}_i^{\text{nonsel}}(t) + I_i^{\text{sel}}(t) + \eta_i(t) \quad (3)$$

Three-Populations Network (2)

- They have assumed instantaneous inhibitory integration:

$$\tilde{W}_{ij} = W_{ij} - \frac{W_{il} W_{lj}}{1 + W_{ll}} \quad (4)$$

$$\tilde{I}_i^{\text{nonsel}}(t) = I_E^{\text{nonsel}}(t) - \frac{W_{il} I_l^{\text{nonsel}}(t)}{1 + W_{ll}} \quad (5)$$

- For excitatory populations, transduction function is:

$$G_E(h_j) = u_j^* x_j^* g(h_j), \quad g(h_j) = k \log(1 + e^{h_j/k}) \quad (6)$$

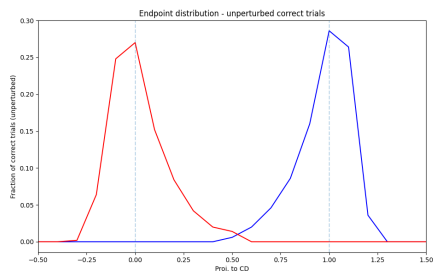
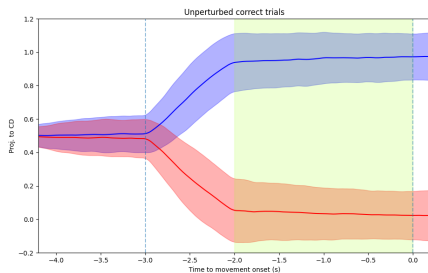
- Where short-term facilitation u_j and depression x_j are stationary:

$$u_j^* = \frac{g(h_j) U \tau_f + U}{1 + U g(h_j) \tau_f}, \quad x_j^* = \frac{1}{1 + \tau_D u_j^* g(h_j)} \quad (7)$$

(U = synaptic release probability)

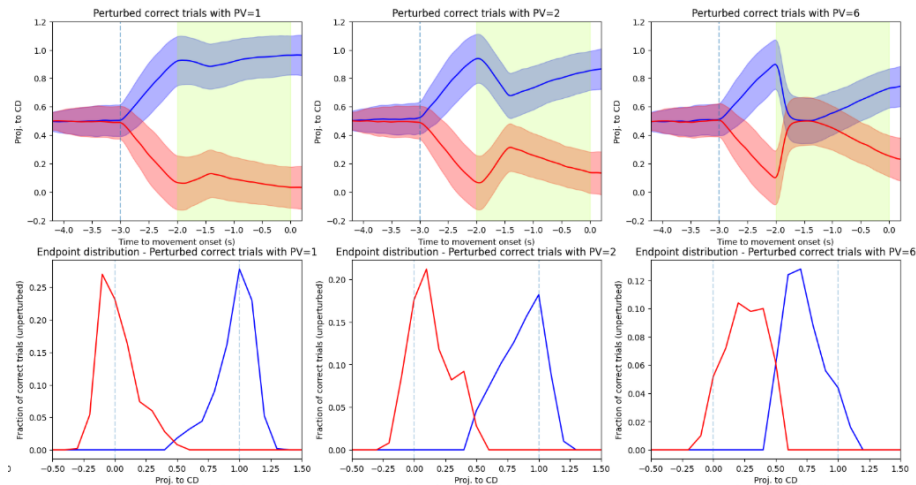
Structured Model Results

- Network shows bifurcation into L/R attractors
- CD projections match recorded data
- Correct trials = during a L/R trial, the activity at the end of delay on L/R population is higher than R/L one

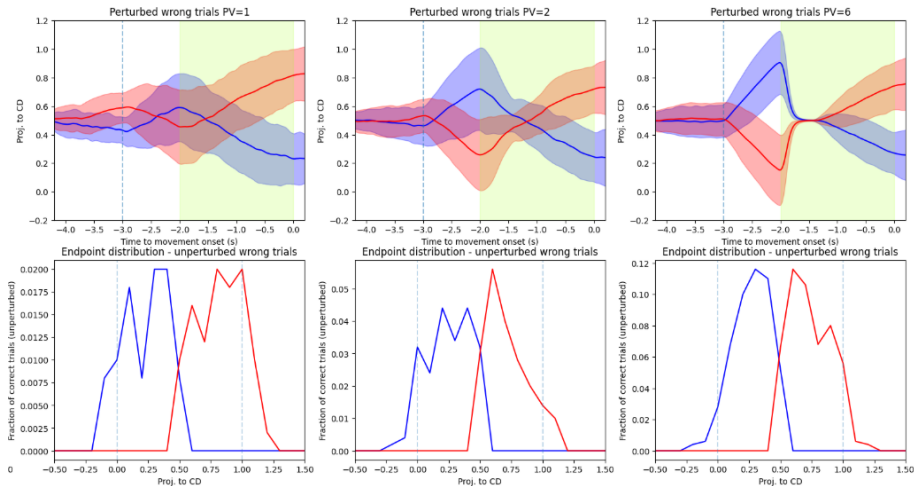


Simulated Correct Photoinhibited Trials

Photoinhibition for inhibitory population: $I_i^{stim}(t) \propto PV \cdot I_i^{nonse}(t)$

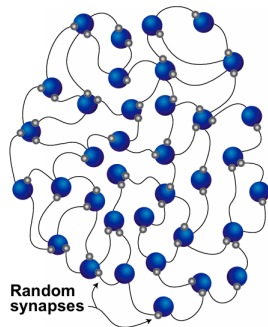


Simulated Incorrect Photoinhibited Trials



Random RNN Dynamics

- Random matrix $J \in \mathbb{R}^{N \times N}$, values from gaussian with $\mu = 0$ and $\sigma^2 = \frac{g^2}{N}$, $g = 1.2$ (chaotic)



- Dynamics for each neuron i :

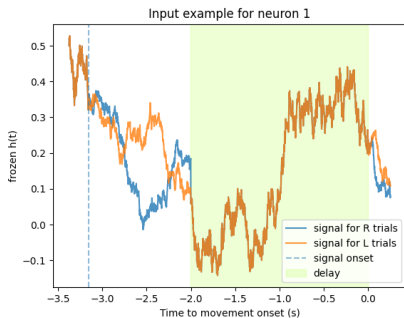
$$\tau \frac{dx_i}{dt} = -x_i + \sum_j J_{ij} \phi(x_j) + h_i(t), \quad r_i = \phi(x_i) = \frac{1}{1 + e^{-x_i}} \quad (8)$$

Input Design for L/R Trials

- External input $h_i(t)$:

$$\tau_{WN} \frac{dh_i}{dt} = -h_i + h_0 \eta_i(t) \quad (9)$$

- Frozen across learning trials for each neuron i
- Equal for R/L trials to mimick the absence of a specific stimulus



FORCE Update Rule (1)

- Target currents: $f_i(t) = \log\left(\frac{R_i(t)}{1-R_i(t)}\right)$
- $R_i(t)$ is the true trial-averaged firing rate from data
- $r_i(t)$ model's prediction
- Predicted internal currents: $z_i(t) = \sum_j J_{ij} r_j(t)$
- Error: $e_i(t) = z_i(t) - f_i(t)$
- Update Rule:

$$J_{ij}(t) = J_{ij}(t-1) - \Delta J_{ij} \quad (10)$$

FORCE Update Rule (2)

$$\Delta J_{ij}(t) = ce_i(t) \sum_k P_{jk}(t) r_k(t), \quad c = \frac{1}{1 + r^T(t)P(t)r(t)} \quad (11)$$

where j and k are restricted to p trainable synapses³

$P_{ij} = \langle r_i r_j \rangle^{-1}$ inverse cross-correlation matrix

$$P(t) = P(t-1) - \frac{P(t-1)r(t)r^T(t)P(t-1)}{1 + r^T(t)P(t-1)r(t)}, \quad P(0) = \alpha \mathbb{I} \quad (12)$$

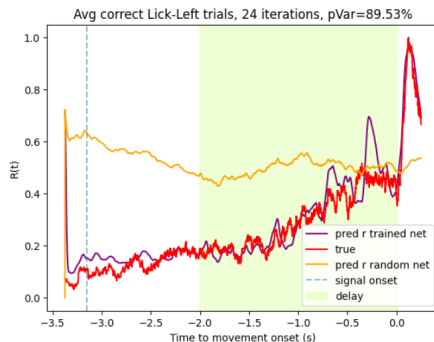
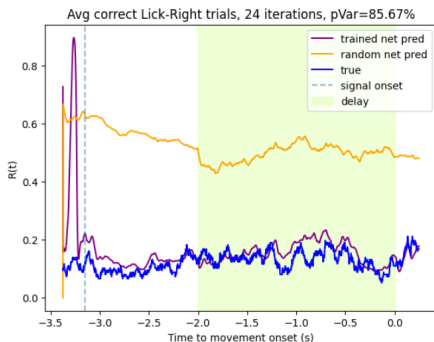
Convergence criteria:

$$pVar = 1 - \frac{\langle R_i(t) - r_i(t) \rangle^2}{\langle R_i(t) - \bar{R}_i(t) \rangle^2} > 0.85 \quad (13)$$

³as in Rajan et al. (2016)

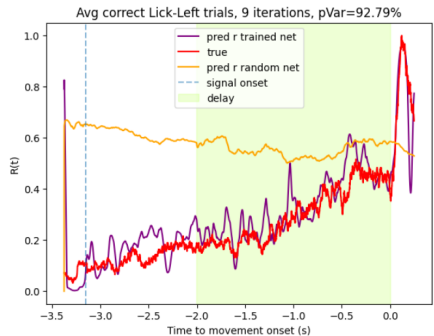
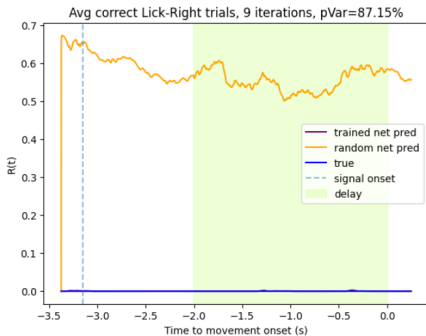
Trial-Averaged Training, N=87

- Convergence only if the network is large enough
- For N=87, convergence for almost full training: $p > 0.98$
- Selectivity recovered: $\langle sel \rangle_{delay} = 0.91$



Variant: explicitly suppress activity in selective neurons

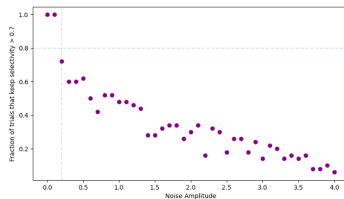
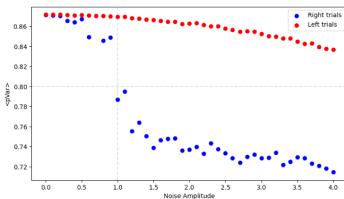
- Inspiration from the dataset used by Rajan et al. (2016): given a selective L/R neuron, its activity during R/L trials is suppressed
- Selectivity recovered: $\langle sel \rangle_{delay} = 0.94$



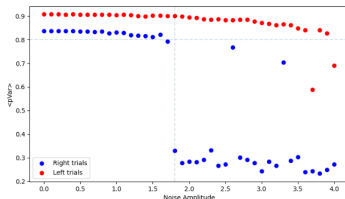
Noise Tolerance Limits

- Tried different noise amplitudes A , for each of them 50 trials
- $\langle pVar \rangle_{right}$ drops after $A^* = \{1, 1.8\}$ values
- frac. of trials with $\langle sel \rangle_{delay} > 0.7$ drops after $A_{sel} = \{0.2, 1.2\}$

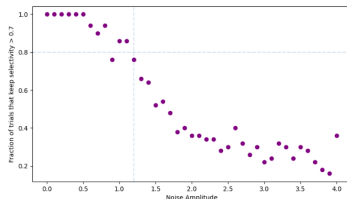
Noise Amplitude effect for trained network, 50 generation per amplitude



Noise Amplitude effect for trained network with activity suppression in



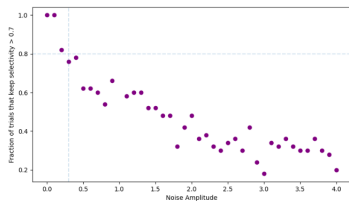
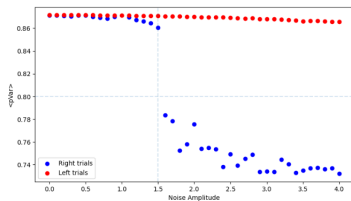
selective neuron, 50 generation per amplitude



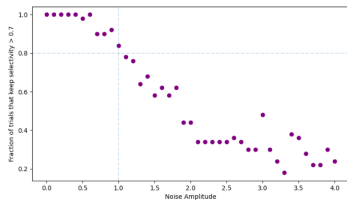
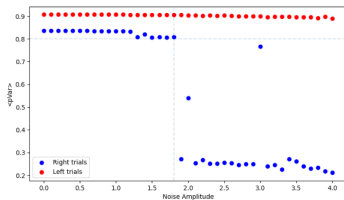
Noise tolerance during delay

$$A^* = \{1.5, 1.8\}, A_{sel} = \{0.4, 1.2\}$$

Noise Amplitude effect during first 0.6s of delay epoch for trained network, 50 generation per amplitude

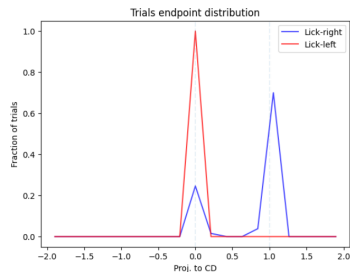
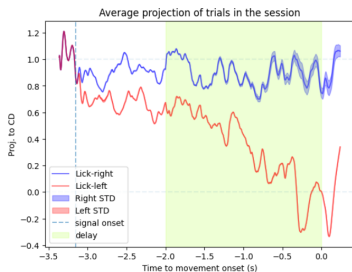


Noise Amplitude effect during first 0.6s of delay epoch for trained network with activity suppression in -selective neurons, 50 generation per amplitude

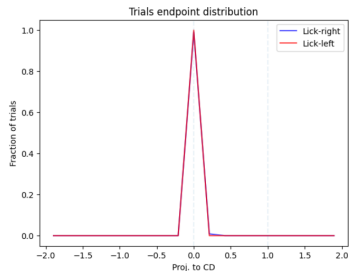
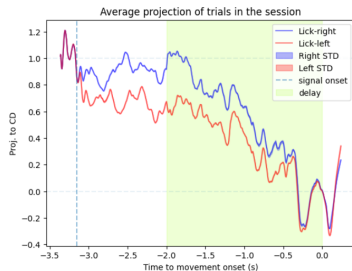


CD projection, noise perturbation $A^* = 1$

Correct trials:
 $\langle sel \rangle_d > 0.7$

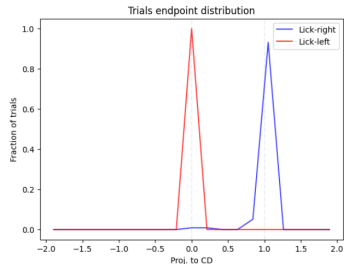
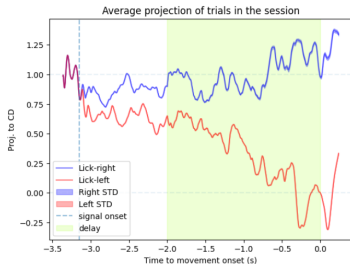


Incorrect trials

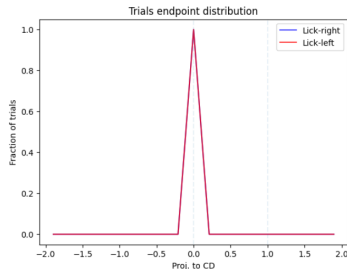
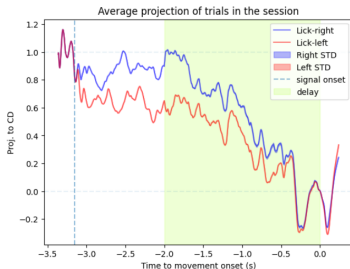


CD projection, noise delay $A_{sel} = 1.5$

Correct trials



Incorrect trials

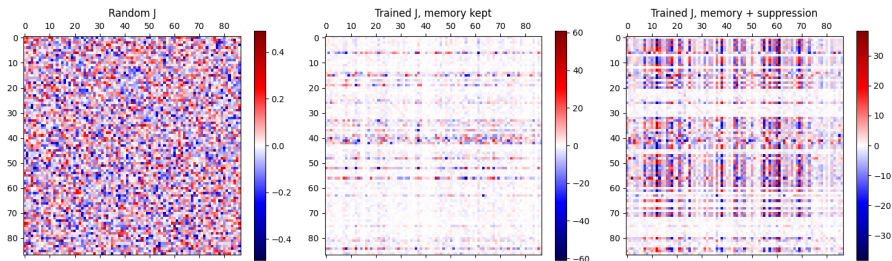


Observations

- During delay epoch, the correct trials recover the same learned trajectory with small variability
- During incorrect trials, the network fails to recover right dynamics and converge it towards left dynamics, erasing differences between trajectory types
- The convergence appears few seconds before go cue \rightarrow shorter delay epochs might succeed to keep the memory for stronger A
- Similar results are found using the trajectories from $J_{suppress}$

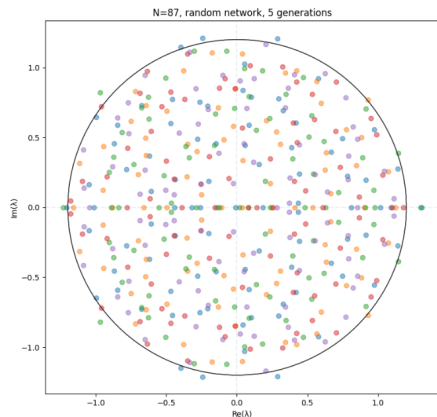
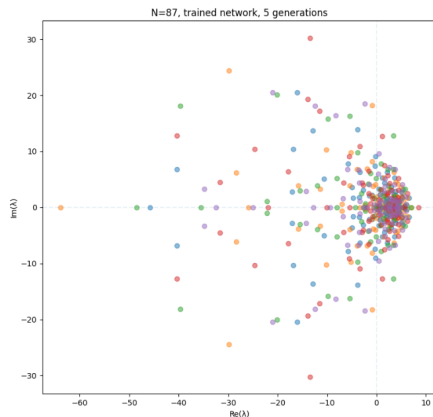
Synaptic Structure Changes

- Some synaptic connections are dominant $J_{ij} > |20|$
- In $J_{suppress}$ the intensity of strong connections is in the same range \rightarrow the pattern is evident



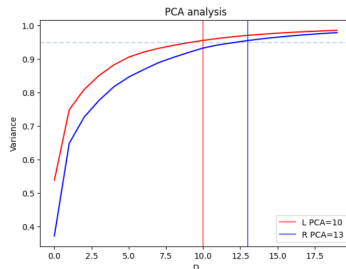
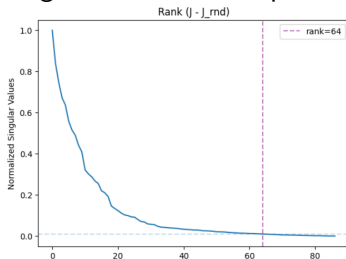
Eigenvalue Spectra of J

- Pre-training: Circular Law holds
- Post-training: ellipse + strong $\text{Re}(\lambda) < 0$ outliers, all complex eigenvalues are symmetric, $\max(\text{Re}(\lambda)) \sim 7$



Rank and PCA analysis (1)

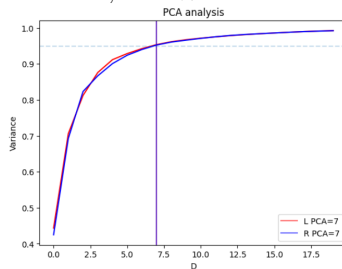
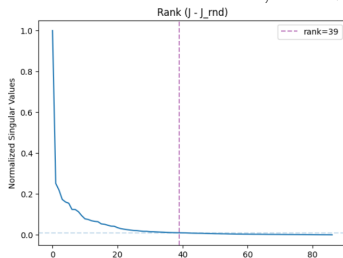
- PCA Random: $D_{L,cvar95\%}^{random} = 9$, $D_{R,cvar95\%}^{random} = 9$, $\alpha \sim 10^{-13}$
- PCA Target: $D_{L,cvar95\%}^{target} = 16$, $D_{R,cvar95\%}^{target} = 18$, $\alpha \sim 10^{-13}$
- Singular Value Decomposition: $\Delta J = USV^T$, $s_i > 0$



- $D_{L,cvar95\%}^{network} = 10$, $D_{R,cvar95\%}^{network} = 13$, $\alpha \sim 10^{-13}$
- $rank = 64 \rightarrow$ a large number of directions are modified with learning

Rank and PCA analysis (2)

- Suppressed target: $D_{L,cvar95\%} = 14$, $D_{R,cvar95\%} = 15$, $\alpha \sim 10^{-13}$



- PCA Target: $D_{L,cvar95\%}^{network} = 7$, $D_{R,cvar95\%}^{network} = 7$, $\alpha \sim 10^{-13}$
- L and R trials share the same subspace
- $rank = 39$

Conclusions

- **Highly structured models** with finely tuned parameters that produce discrete attractors can **replicate** neural dynamics from delayed-pair association tasks
- **Simulated photoinhibition** reveals **attractor recovery or switch**, consistent with experimental findings
- **Random RNNs** trained on trial-averaged activity **can memorize left/right inputs before a delay period**
- **Adding stochastic noise** to the input can cause right-trial trajectories to **collapse** into their corresponding left-trial dynamics during delay but not vice versa, losing the memory of correct input before go cue