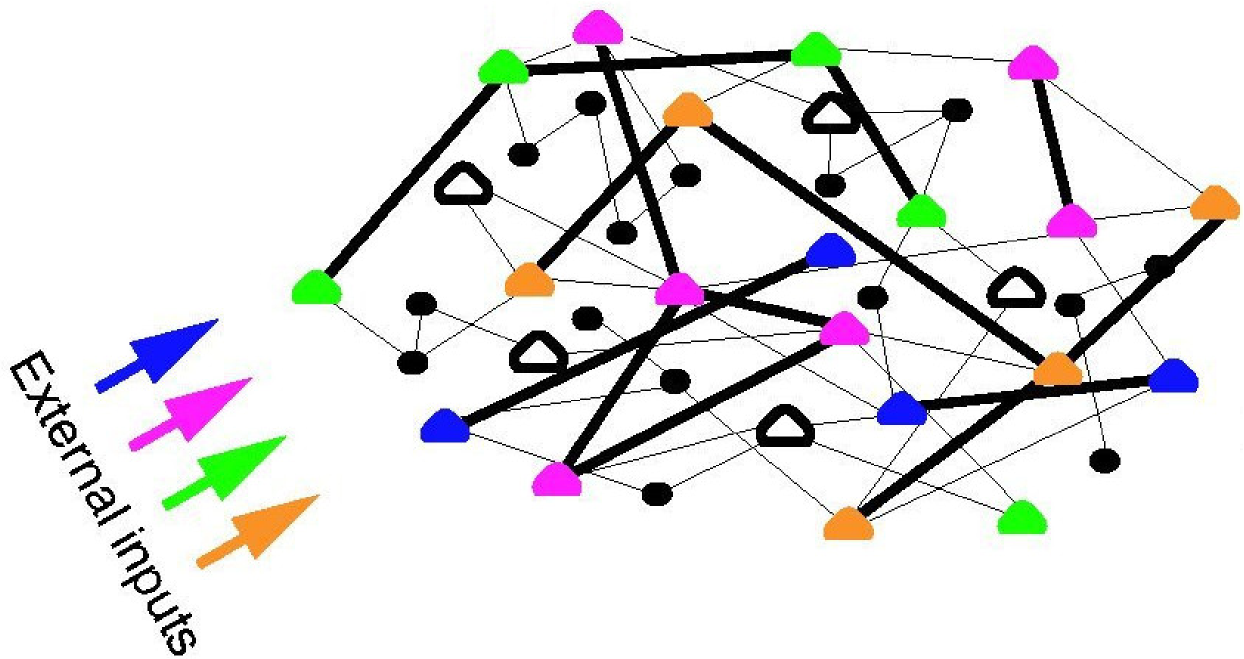
Dynamics of neural networks with learning rules inferred from data

Nicolas Brunel

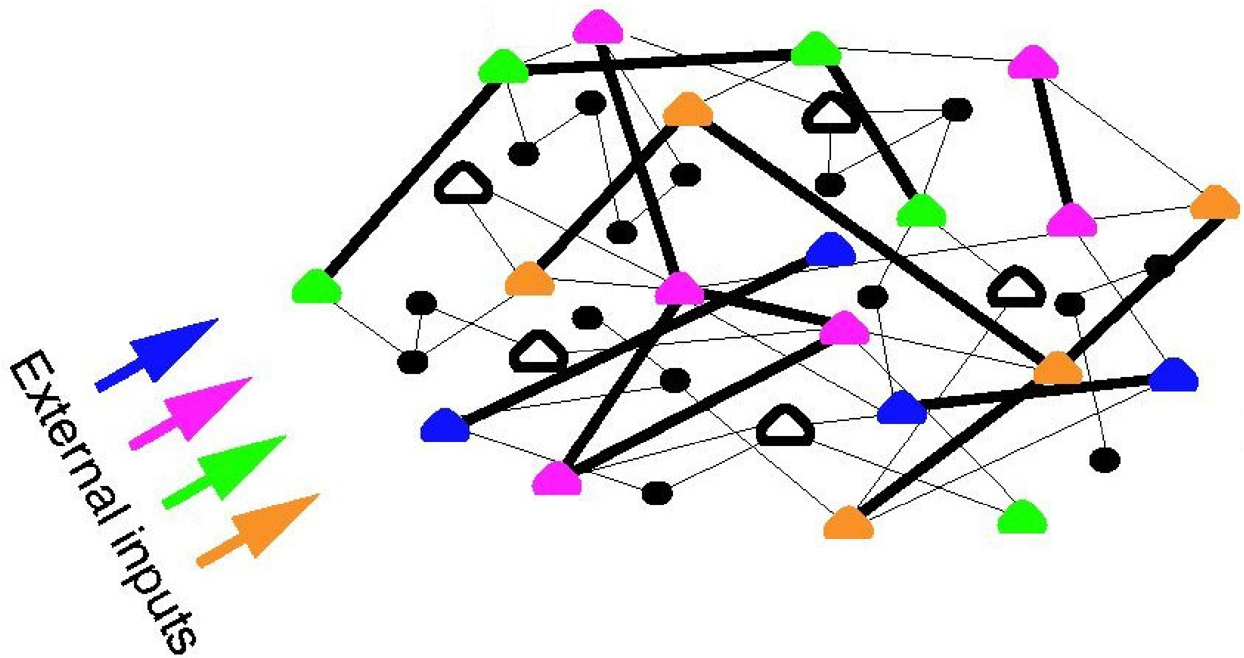
Departments of Neurobiology and Physics, Duke University

How do brain networks store information?

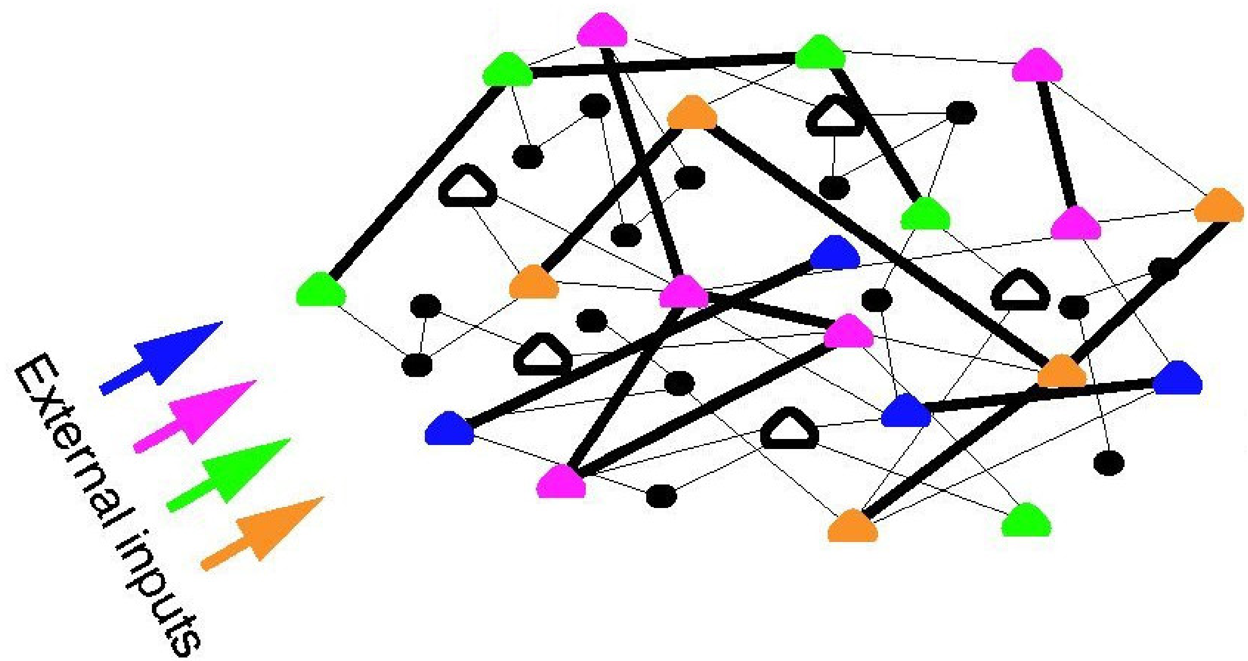


Questions

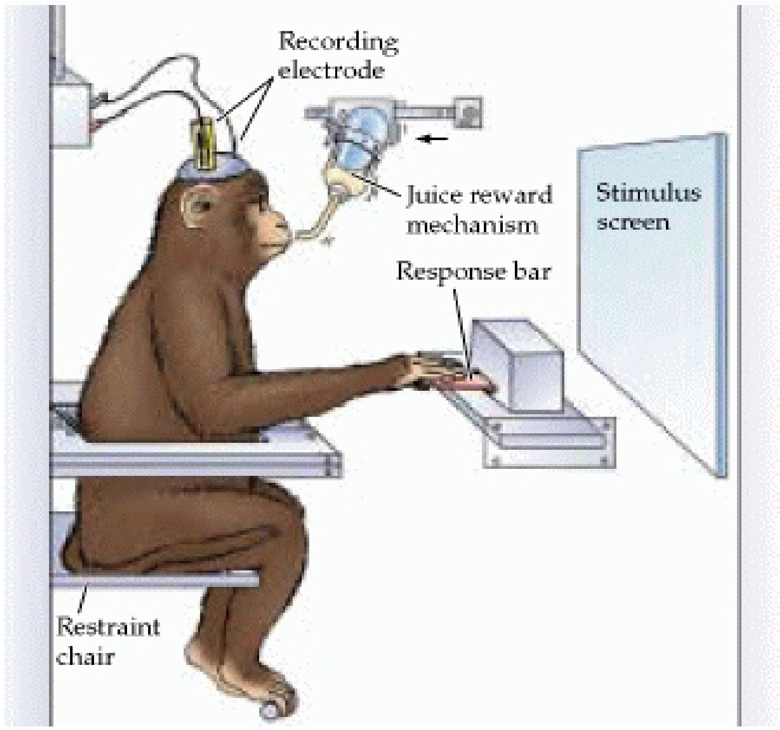
1. What are the rules governing synaptic plasticity ('learning rules')?
2. How does synaptic plasticity affect network dynamics?
3. What is the storage capacity of neural networks?

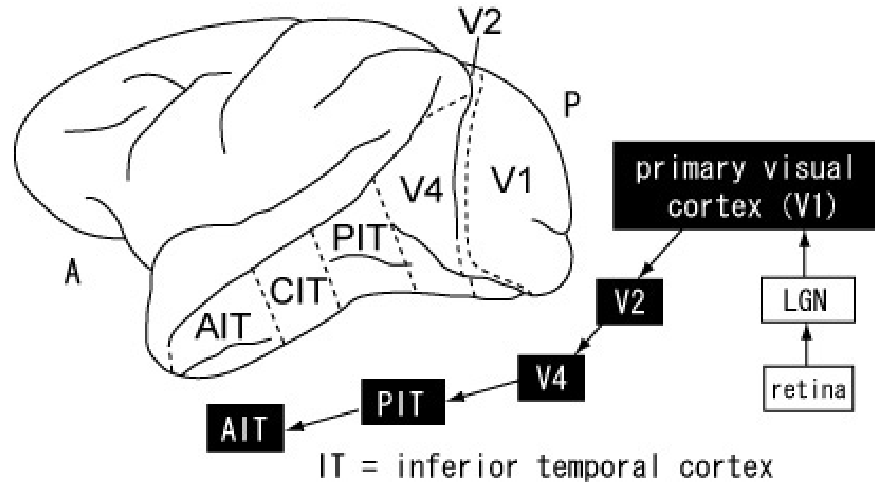


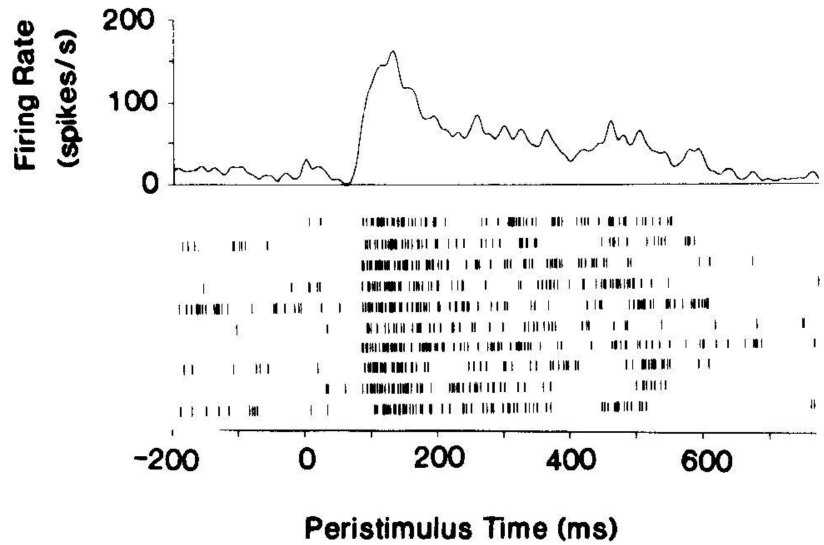
Inferring learning rules from in vivo data



Electrophysiological recordings in ITC of awake monkeys

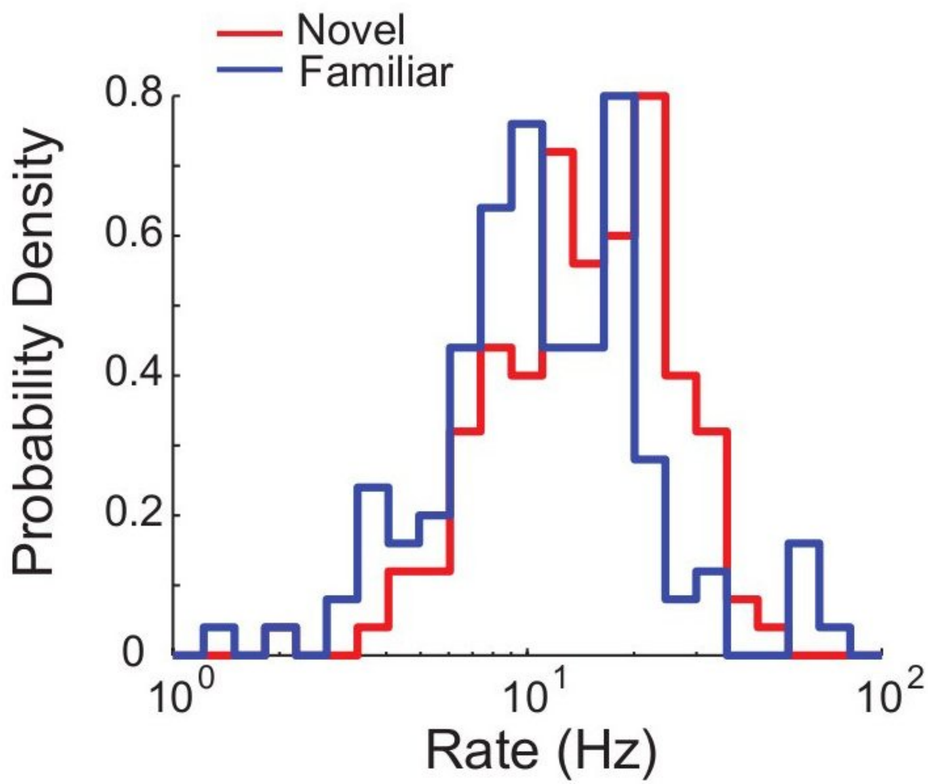






- Data from Woloszyn and Sheinberg (2012): 125 novel and 125 familiar images per session
- Average visual response in the 75-200ms interval for each neuron and each stimulus

Distribution of average visual responses for novel/familiar stimuli



For most neurons:

- Mean(Familiar) $<$ Mean(Novel)
- Best(Familiar) $>$ Best(Novel)

How do distributions of firing rates evolve with learning?

- Firing rate model, with $N \gg 1$ neurons described by a firing rate r_i;
- Total inputs to neuron i

$$h_i = I_{iX} + \frac{1}{N} \sum_j J_{ij} r_j$$

- Firing rate $\tau dr_i/dt = -r_i + \Phi(h_i)$
- When a novel stimulus is shown, $r_i = v_i$ where v_i is drawn from $P_{nov}(v)$
- Induces changes in synaptic connectivity according to an unsupervised rule

$$J_{ij} \rightarrow J_{ij} + \Delta J(v_i, v_j)$$

- We assume $\Delta J(v_i, v_j) = f(v_i) g(v_j)$
- What is the new distribution of rates for the (now familiar) stimulus?

How do distributions of rates evolve with learning?

- Changes in total input due to learning are

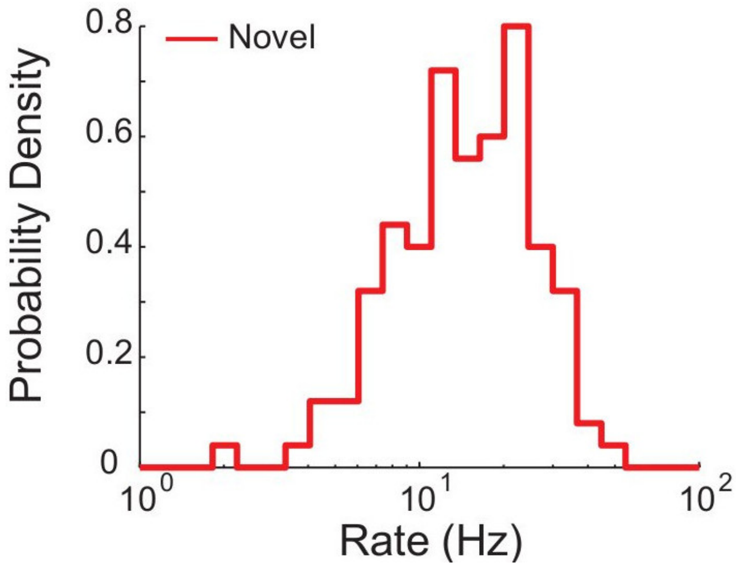
$$\Delta h_i \approx \frac{1}{N}\sum_j \Delta J_{ij} v_j$$
$$\approx f(v_i)\overline{g(v)v}$$

- Change in inputs Δh_i depend on visual response v_i, through f
- Sign of f determines whether response increases or decreases with learning

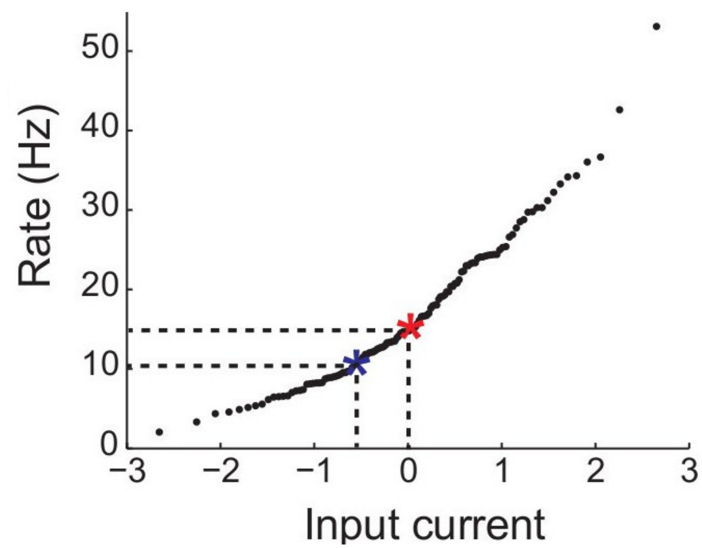
Inferring transfer function

- Infer transfer function Φ, from
  - Empirical distribution of rates for novel stimuli;
  - Assumed Gaussian distribution of inputs for novel stimuli

$P_{nov}(v_i)$



$\Phi(h_i)$



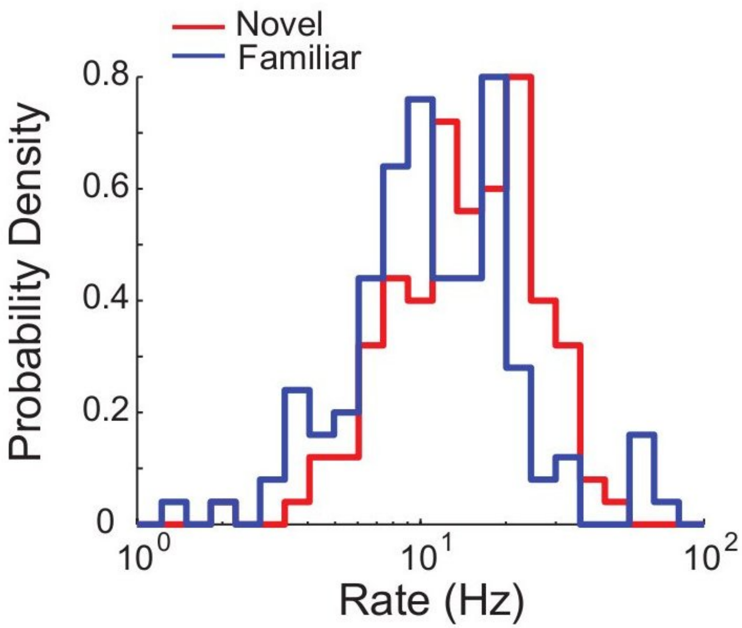
- Supra-linear transfer functions, consistent neurons operating in fluctuation-driven regime

Inferring learning rule

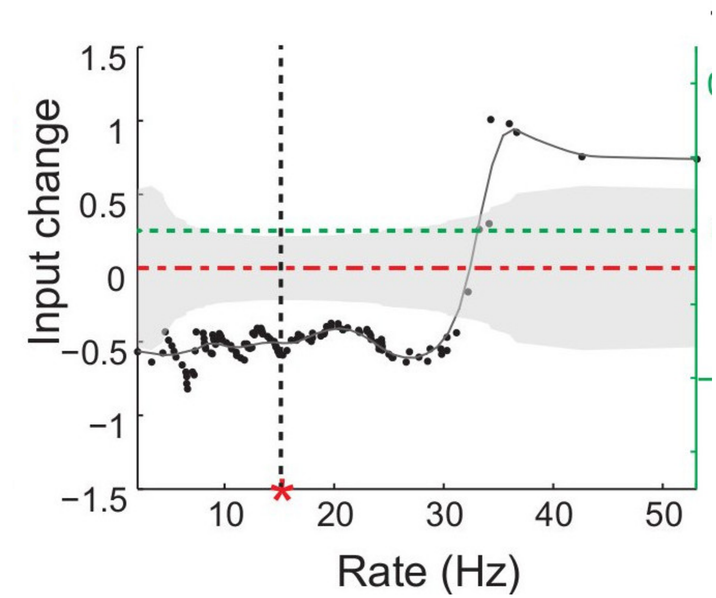
- Goal: Infer plasticity rule $\Delta J(v_i, v_j) = f(v_i)g(v_j)$ from $P_{nov}(v_i)$ and $P_{fam}(v_i)$?
- Assumptions:
  - Stationarity (currently familiar stimuli had, when they were novel, the same distribution as currently novel stimuli)
  - Learning rule preserves rank
- With these assumptions, it is possible to infer $f(v_i)$ - the dependence of the rule on the post-synaptic firing rate - from $P_{nov}(v_i)$ and $P_{fam}(v_i)$
- $g(v)$ undetermined, but it has to satisfy

$$\begin{aligned} \int g(v)P_{nov}(v)dv &= 0 \\ \int g(v)vP_{nov}(v)dv &> 0 \end{aligned}$$

$P_{nov}(v_i)$, $P_{fam}(v_i)$



$f(v_i)$



- Consistent with a Hebbian rule whose dependence on post-synaptic firing rate is non-linear, and biased towards depression

Conclusions I

- Inferred post-synaptic dependence of learning rule from in vivo data
- Data consistent with unsupervised Hebbian plasticity
- Firing rate dependence is consistent with a BCM rule
- Sparsening of representations in ITC
- Simple readout for stimulus familiarity (average network activity)

Lim, McKee, Woloszyn, Amit, Freedman, Sheinberg and Brunel (Nat. Neurosci. 2015)

Dynamics of networks with learning rules inferred from data

- Data consistent with a non-linear Hebbian rule whose post-synaptic dependence is dominated by depression
- Does such a rule lead to attractor dynamics?
- What is the storage capacity of such a rule?

The model

- N neurons, whose firing rate obey

$$\tau \frac{dr_i}{dt} = -r_i + \Phi \left(I_i + \sum_{i \neq j}^{N} J_{ij} r_j \right)$$

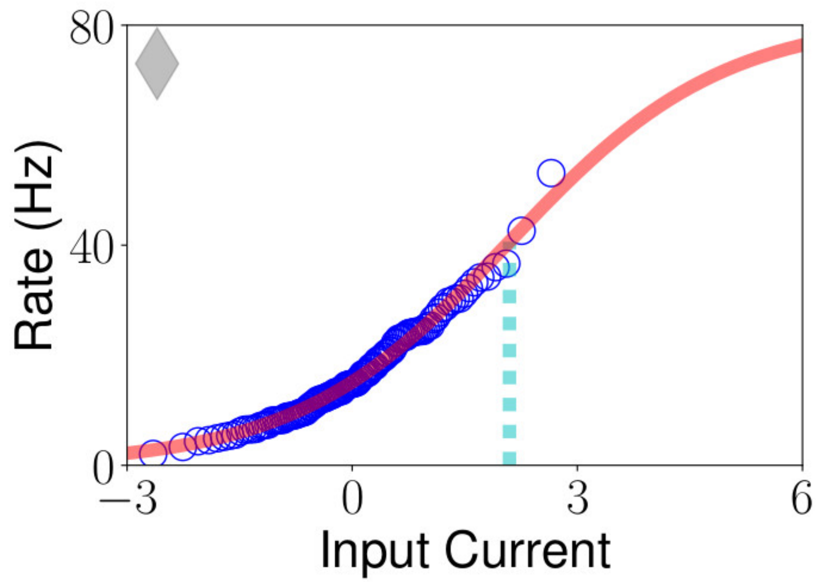
- p random uncorrelated Gaussian input patterns $\xi_i^\mu \sim \mathcal{N}(0,1)$
- Connectivity matrix

$$J_{ij} = \frac{c_{ij}}{cN} \sum_{\mu=1}^{p} \tilde{f}\left(\xi_i^\mu\right) \tilde{g}\left(\xi_j^\mu\right)$$

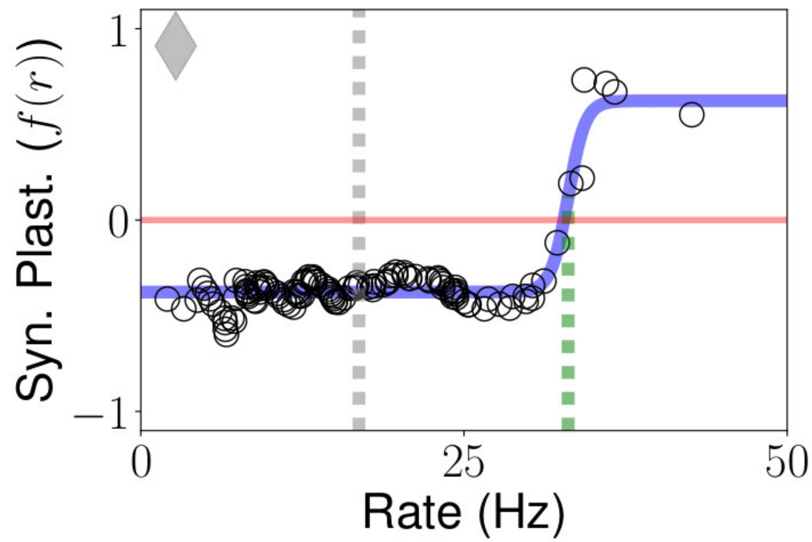
where $\tilde{f}(x) = f(\Phi(x)), \tilde{g}(x) = g(\Phi(x))$,
$c_{ij} =$ ER 'structural' connectivity matrix ($c_{ij} = 1$ with prob. $c \ll 1$),
g such that $\int Dx\tilde{g}(x) = 0$, $\int Dx\tilde{g}(x)\Phi(x)dx > 0$.

Transfer functions and learning rules inferred from data

Transfer function Φ



Post dependence of learning rule f



- Fit both Φ and f by sigmoidal functions, for all neurons with significant 'Hebbian' plasticity rules;
- Take g as a sigmoidal function, with threshold and gain identical to f, and offset given by the condition $\int Dx\tilde{g}(x) = 0$
- Simulate and analyze the dynamics of a network with median parameters

Mean-field theory

- Can the network retrieve a stored pattern (i.e. converge to an attractor that is correlated with the pattern)?
- Define order parameters

$$m = \langle \frac{1}{N} \sum_i \tilde{g}(\xi_i^1) r_i \rangle \qquad \text{(Overlap with retrieved pattern)}$$

$$\sigma^2 = \langle \frac{1}{N^2} \sum_{\mu>1,j} \tilde{f}^2(\xi_i^\mu) \tilde{g}^2(\xi_j^\mu) r_j^2 \rangle \qquad \text{(Quenched noise due to other patterns)}$$

- In the limits $N \to \infty$, $N \gg cN \gg 1$, $p \sim cN$, order parameters given by MF equations

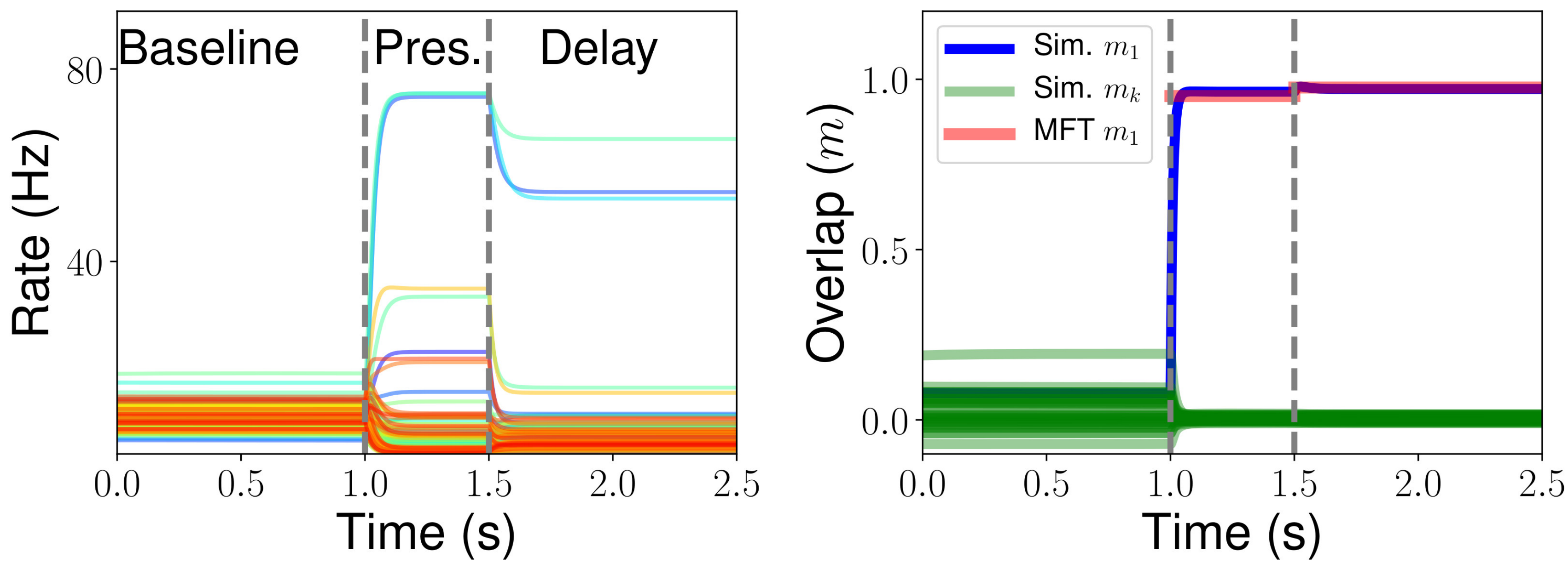
$$m = \int D\xi Dz \tilde{g}(\xi) \Phi(\tilde{f}(\xi) m + \sigma z)$$

$$\sigma^2 = \alpha \int D\xi \tilde{f}^2(\xi) \int D\xi \tilde{g}^2(\xi) \int D\xi Dz \Phi^2(\tilde{f}(\xi) m + \sigma z)$$

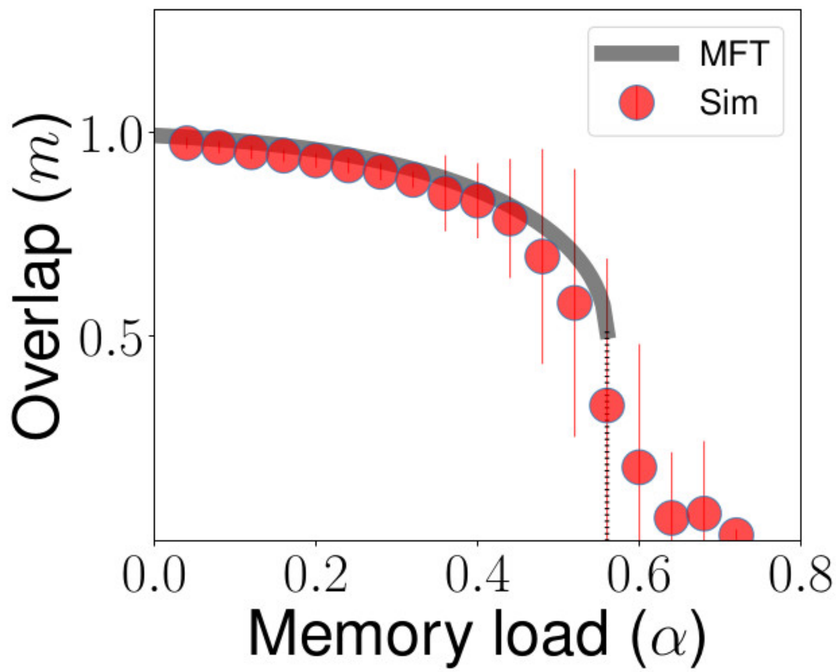
where $\alpha = p/(cN)$.

- Retrieval states: Solutions such that $m > 0$;
- Storage capacity: largest α for which retrieval states exist.

Learning rules inferred from data lead to attractor dynamics and delay period activity

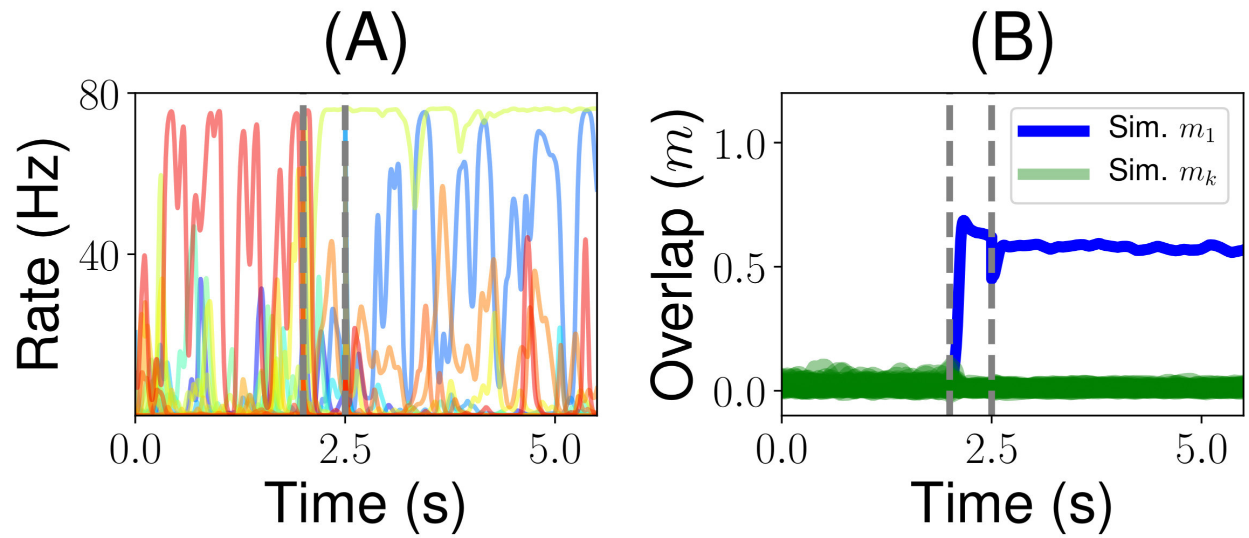


Storage capacity



- Storage capacity for median parameters close to 0.6;
- Close to optimal capacity ($\alpha_{max} \sim 0.8$), in the space of sigmoidal functions f and g.
- Optimal learning rule in such a space: Both f and g are step functions with high thresholds

Transition to chaos at strong coupling



- Increasing coupling strength leads to chaotic retrieval states
- Similar to chaotic states in simpler asymmetric rate models (Sompolinsky et al 1988, Tirozzi and Tsodyks 1991)
- Reproduces strong irregularity and diversity of temporal profiles of activity seen in delay periods in PFC

Conclusions

- Network model with distribution of patterns and learning rule inferred from data exhibits attractor dynamics
- Learning rule inferred from data close to optimal in terms of storage capacity (in the space of Hebbian learning rules with sigmoidal dependence on pre and post rates)
- Transition to chaos at sufficiently strong coupling - leads to strong irregularity and diversity of temporal profiles of activity in the delay period, similar to observations in PFC

Pereira and Brunel (Neuron 2018)

Acknowledgements

- **Sukbin Lim** (Chicago $\rightarrow$ NYU Shanghai)
- **Ulises Pereira** (Chicago $\rightarrow$ NYU)
- Yali Amit (Chicago)
- Luke Woloszyn (Brown)
- David Sheinberg (Brown)
- Jill McKee (Chicago)
- David Freedman (Chicago)