Random Projection
and
the Assembly Hypothesis

with Christos Papadimitriou, Mike Collins, Dan Mitropolsky

Thanks to Rosa Arriaga, Manuel Blum, Les Valiant,
Samantha Petti, Samira Samadi, John Wilmes,
Le Song, Wolfgang Maass, Robert Legenstein

Random Projection

- Replace original features by random linear combinations of them

- Feature vector $x = (x_1, x_2, \dots, x_n)$
- Random $k \times n$ matrix R with iid entries

- $y = Rx$

The Machine Learning Problem (to a theoretician)

Elephant

Not Elephant

?

Elephant

Not Elephant

The Brain Problem (to a theoretician, possibly the same one)

Daddy!

??

How does Cognition (learning, language,..)

arise

from Brain (neurons, synapses,…)?

How could the brain *possibly* accomplish all that it does?

- Visual invariants: how to define an “A”?
- Motor control: Swim/walk etc. Solving the equations of mechanics is not tractable on the fly; so what is going on?
- Language: How is language processed? How do we learn so well from so few examples?

High-level Challenge #1: Find a plausible (and algorithmic) explanation for how all this is even possible.

Random Projection for ML [Arriaga-V. 99,06]

- Random Projection preserves
  - lengths to within relative error, and
  - inner products to within additive err
- So robustly separated concept clas

An algorithmic theory of learning: Robust concepts and random projection

Rosa I. Arriaga · Santosh Vempala

- And the separator in the projected space is the projection of the separator in the original space
- E.g., if the original concepts are intersections of halfspaces, $W \cdot x \geq 0$, then for projected points $y = Rx$, the "projected" halfspaces WR^T is the approximate concept: $\mathrm{E}\left(\left(WR^T\right) \cdot (Rx)\right) = W \cdot x$ and $WR^T Rx \simeq Wx$
- Points not too close to the boundary will have their concepts preserved!
- Dimension $k = \tilde{O}\left(1/\gamma^2\right)$ suffices if the concept has margin γ.

Random Projection to featurize Kernels

- Often, similarity between x, y is $K(x, y) = \phi(x) \cdot \phi(y)$ where $\phi(.)$ is some function, typically a map to a much higher dimensional space.

- If x are from a distribution, then choose $x_1, x_2, \ldots, x_m$ iid and define
$$F(x) = \left(K(x, x_1), \ldots, K(x, x_m)\right)$$

- Then $F(x) \cdot F(y) \simeq K(x, y)$ and if the kernel has a classifier with margin γ, then with prob. $1 - \epsilon$ it suffices to have $m = \tilde{O}(1/\gamma^2)$ to preserve it. [BBV06]

Random Projection in deep learning

- Can replace all but the last layer with one large enough layer with random weights into it.

- Thm [V.-Wilmes 2019] Gradient descent on just the top-layer weights learns best fixed-degree polynomial approximation of arbitrary input functions for spherically symmetric input distributions, using poly time and samples.

- Thm. Matching Statistical Query lower bound!

Many applications since then, in theory and practice

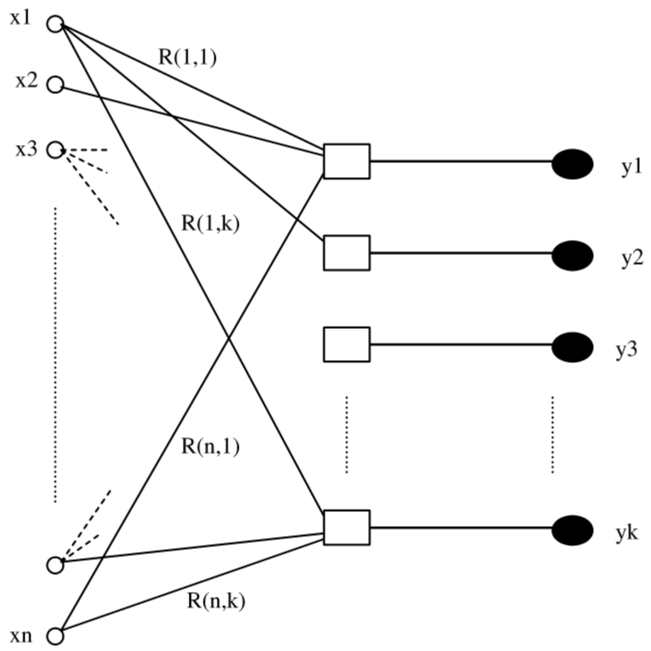
- Low dimensional representation --- efficiency of learning
- Also bounds number of samples needed for learning
- Useful for finding approximately nearest neighbors:
  - Locality sensitive hashing

  [Indyk-Motwani-Raghavan-V. 97; Indyk-Motwani 98, Andoni-Indyk 07]
- Rounding for approximation algorithms
- Matrix sketching
- Etc.

Random Projection for brain?

- Neuronal RP

Fig. 1 Neuronal Random Projection



The Brain, oversimplified

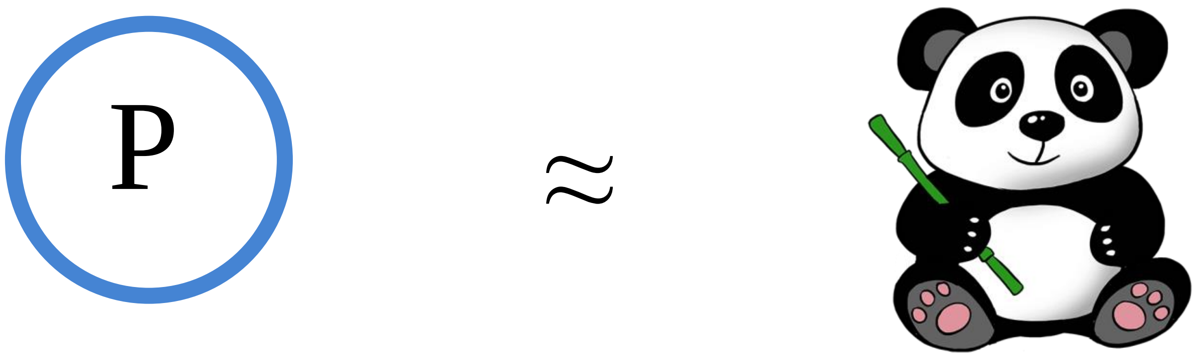
- A network of ~ 80 billion neurons
- $\sim 10^3 - 10^4$ connections per neuron
- Synapses (connections) have strengths, new synapses can form and existing ones might disappear during life
- Individual neurons "spike"/fire based on activation rules that are functions of signals on their input synapses.
  - Nonlinear activations; a common model is a threshold function
  - But there are >1000 types of neurons
  - Signals have a temporal aspect, with firing "rates" and "patterns".

Searching for Memory

- Concepts: “letter A”, “the internet”, ““Epiphany”, “lilac”, “chair”
- How are they represented?
- How are they created?
- How are they associated?
- How do they change over time?

Neural Assemblies

- An assembly = a subset of k neurons
- Representation of real-world idea, e.g. “Panda”



- Simultaneous activation of (most of) these neurons corresponds with recall, i.e., thinking of this concept

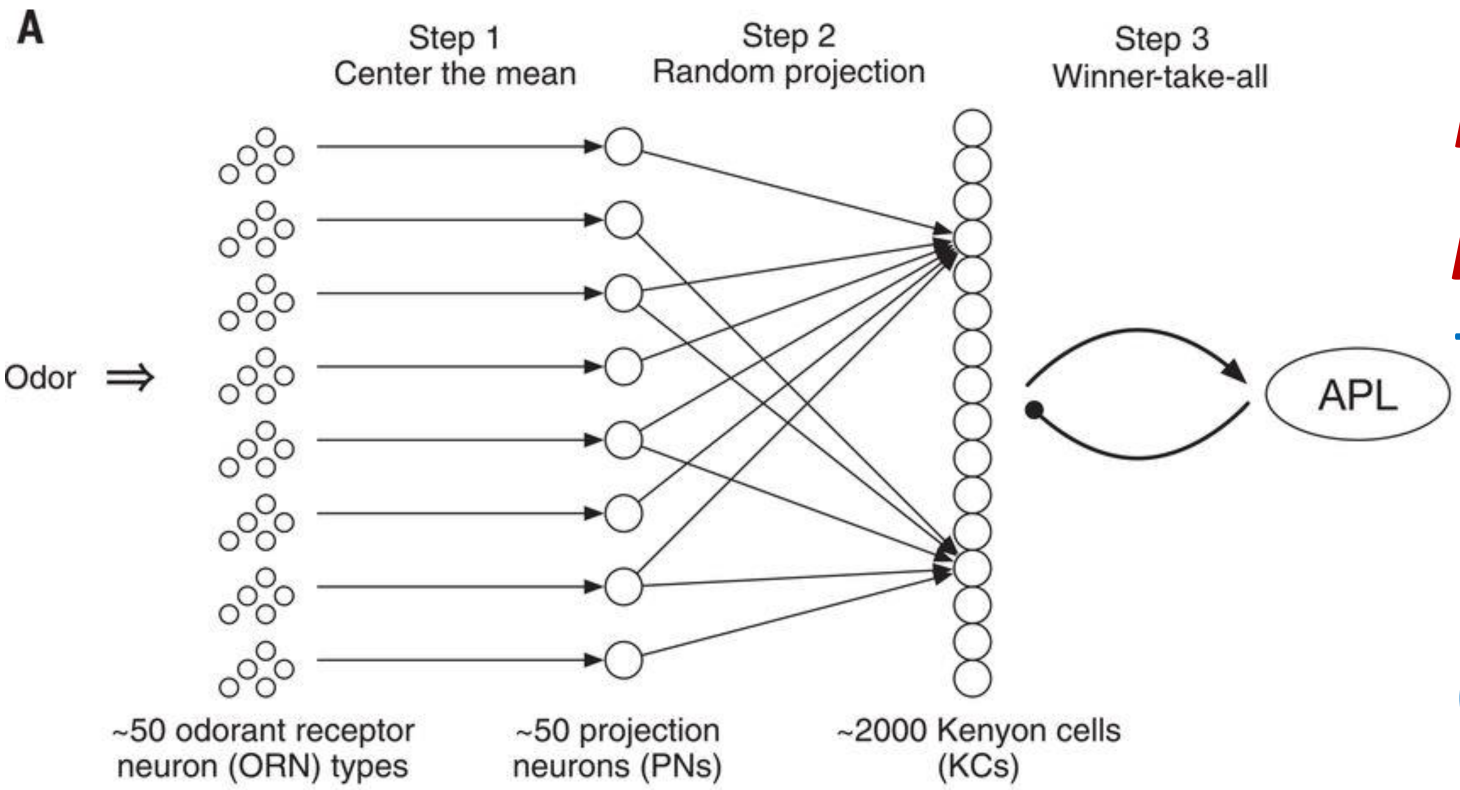
Representation: what is a memory?

1. Subset of neurons, such that if more than a certain fraction “fire”, then the concept is recognized.
2. Distribution over neurons
3. Activity pattern of neurons

Two experiments

1. Association of two assemblies observed in humans [2016]
2. Olfaction in fruit fly [2018] and mouse [2013]

How fruit flies remember smells

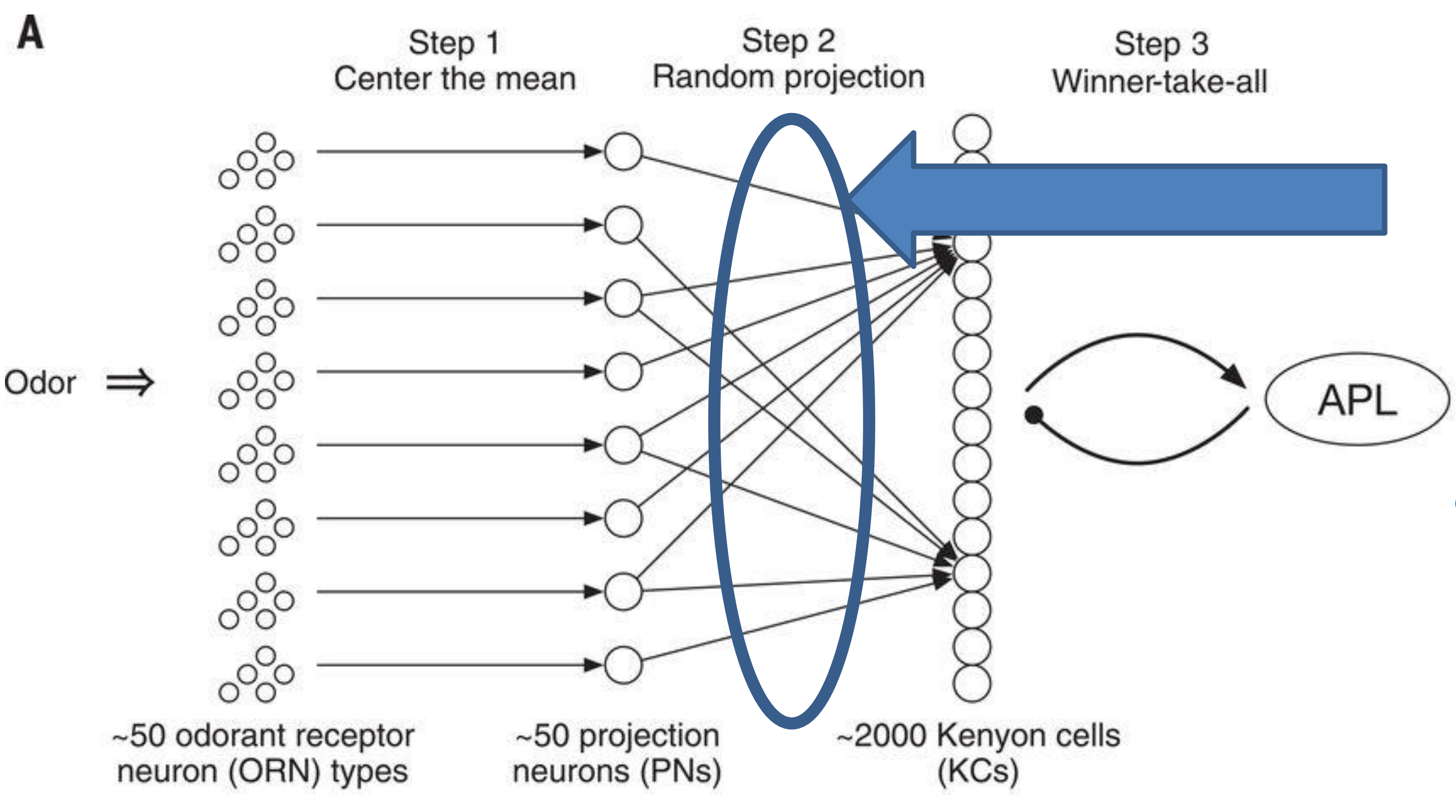


random projection followed by ***cap:*** 100 winners (out of 2000) take all

A neural algorithm for a fundamental computing problem

Sanjoy Dasgupta, Charles F. Stevens, and Saket Navlakha, Science AAAS *Nov 2017*

How fruit flies remember smells



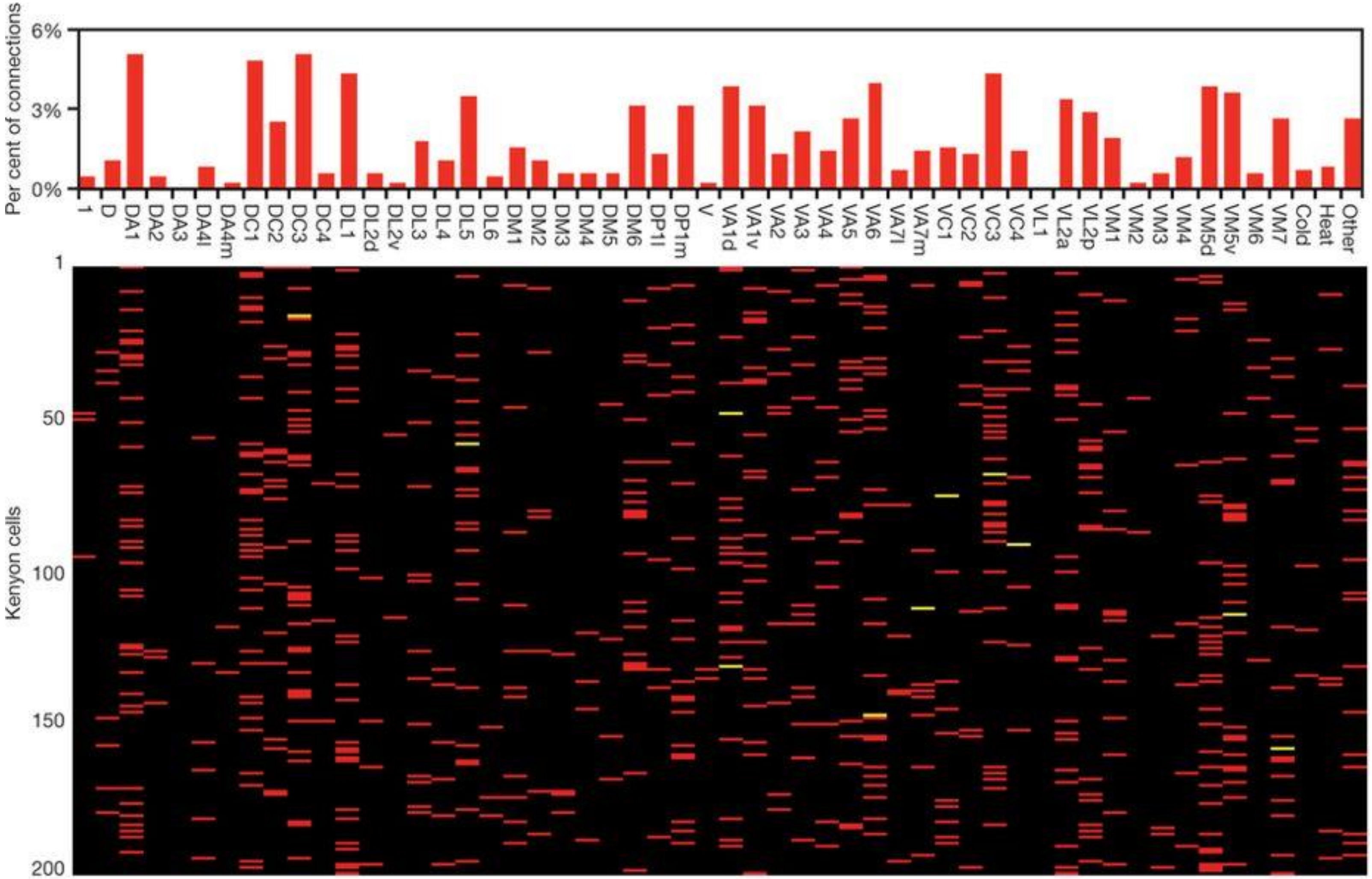
Q: *but* is this a random bipartite graph?

A neural algorithm for a fundamental computing problem
Sanjoy Dasgupta, Charles F. Stevens, and Saket Navlakha, Science AAAS *Nov 2017*

A: ***Random convergence of olfactory input in the Drosophila mushroom body***

by S. Caron, V. Ruta, L. Abbott, R. Axel, nature.com 2013

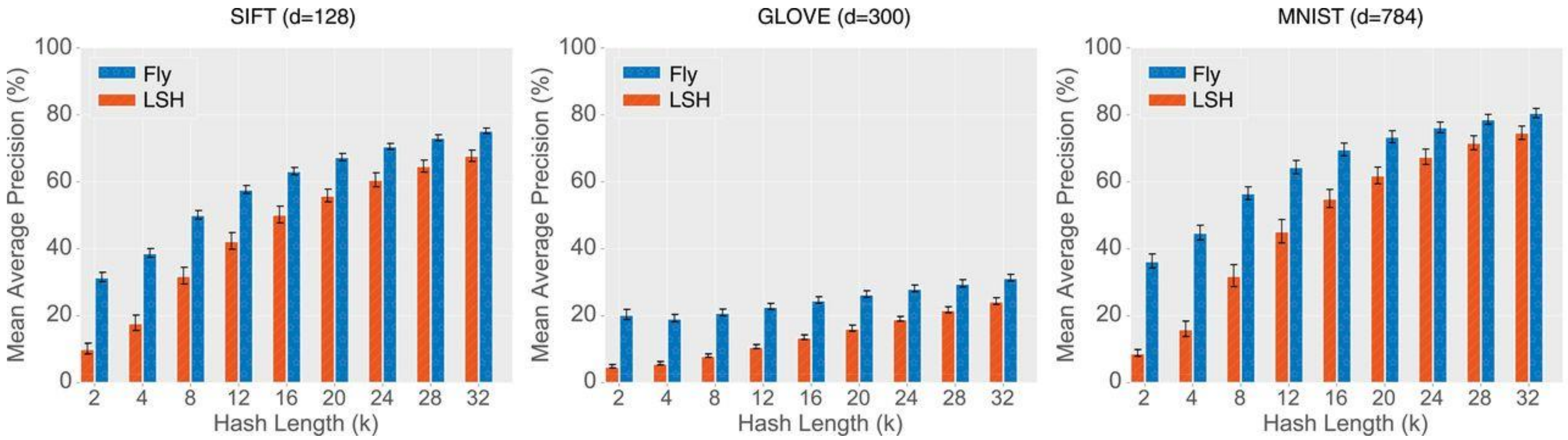
looks like a random bipartite graph, except that the degree distribution of the LHS is not uniform



A neural algorithm for a fundamental computing problem

Sanjoy Dasgupta, Charles F. Stevens, and Saket Navlakha, Science AAAS *Nov 2017*

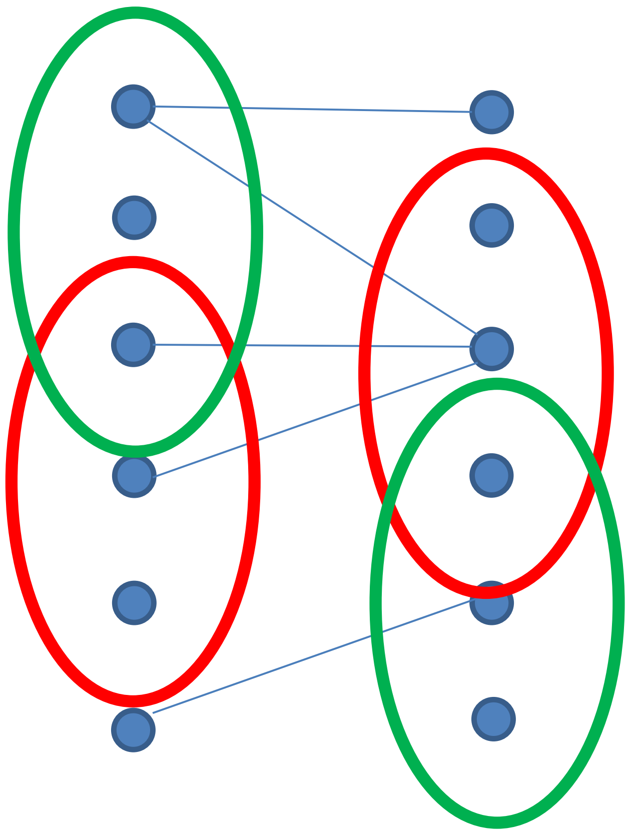
Surprise! the fly's algorithm (random projection followed by cap) ***outperforms**** ***locality sensitive hashing*** in standard datasets!



*: not the most recent versions!

The underlying mathematical reason?

- Random $n \times n$ bipartite graph
- A set **A** of k out of n nodes of the LHS fire
- Random projection, followed by a k-cap on the RHS, forming a new set **cap(A)**
- Repeat for **B, cap(B)**
- If **A** and **B** overlap in αk nodes, what is the overlap of **cap(A)** and **cap(B)**?



The underlying mathematical reason

Theorem [Papadimitriou-V. 2019]:
The intersection of **cap(A)** and **cap(B)** will be at least

$$\frac{\left(\frac{k}{n}\right)^{\frac{1-\alpha}{1+\alpha}}}{\left(\ln\frac{n}{k}\right)^{\frac{\alpha}{1+\alpha}}}$$

The underlying mathematical reason

Theorem [Papadimitriou-V.]

$$\frac{|A \cap B|}{k} = \alpha \quad \Rightarrow \quad \frac{|\mathrm{cap}(A) \cap \mathrm{cap}(B)|}{k} \geq \frac{\left(\frac{k}{n}\right)^{\frac{1-\alpha}{1+\alpha}}}{\left(\ln \frac{n}{k}\right)^{\frac{\alpha}{1+\alpha}}}$$

For fruit fly, $\frac{n}{k} \sim \frac{2000}{100}$ and the bound is substantial! ($\alpha = 0.5 \rightarrow 0.37,\ 0.75 \rightarrow 0.65$)

Proof idea. Probability of being in a cap is positively correlated with being in a previous cap.

The underlying mathematical reason

Proof idea. Probability of being in a cap is positively correlated with being in a previous cap, if stimuli overlap.

$$\Pr(y + z \geq t \mid x + y \geq t) \gg \frac{k}{n}$$

Inhibition and the k-cap

- K-cap is a simple, plausible toy model of inhibition
- In fact if we view the fraction of excitatory, inhibitory neurons as evolving by simple ODEs,
- It reaches stable, nontrivial solutions only if inhibitory neurons integrate synaptic input (a cons
- Which is act

"The membr ich faster

Neuron
Review

Rapid Neocortical Dynamics: Cellular and Network Mechanisms

Bilal Haider[1] and David A. McCormick[1,*]
[1]Department of Neurobiology, Kavli Institute for Neuroscience, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA
*Correspondence: david.mccormick@yale.edu
DOI 10.1016/j.neuron.2009.04.008

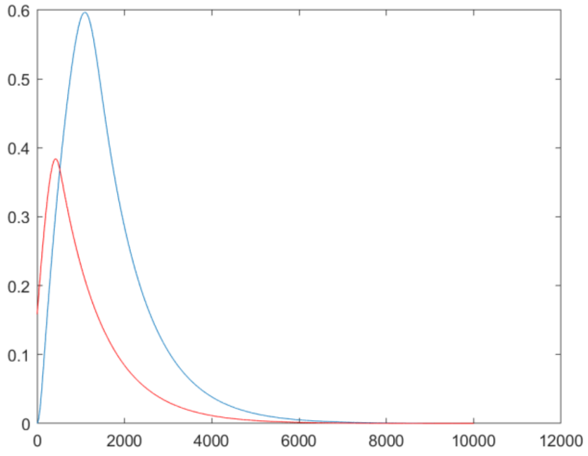
K-cap via inhibition

- Threshold T, step-size h
- n excitatory neurons, m inhibitory
- Inhibitory neurons integrate r times faster
- Repeat:

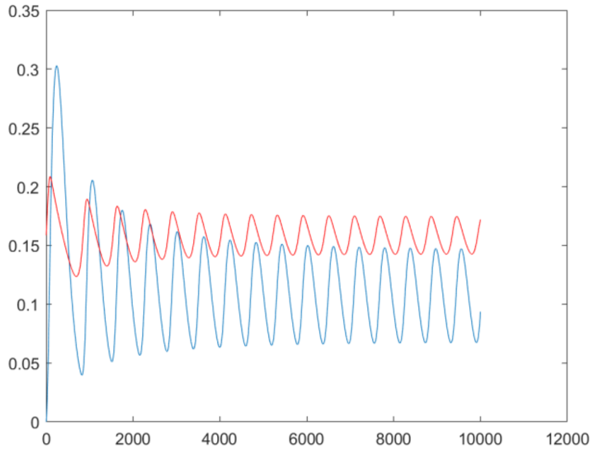
$$I(t+1) = I(t) + \boldsymbol{r} \cdot h\left(\phi\left(T, pnK(t), \sqrt{pnK(t)}\right) - I(t)\right);$$

$$K(t+1) = K(t) + h\left(\phi\left(T - pL, -pmI(t) + pnK(t), \sqrt{pnK(t) + pmI(t)}\right) \ - K(t)\right);$$

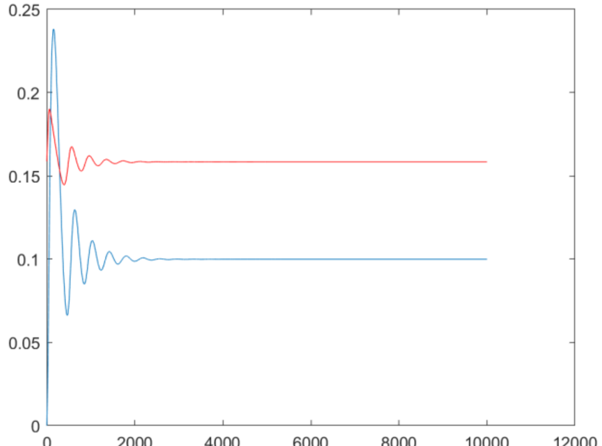
r=1



r=5



r=10



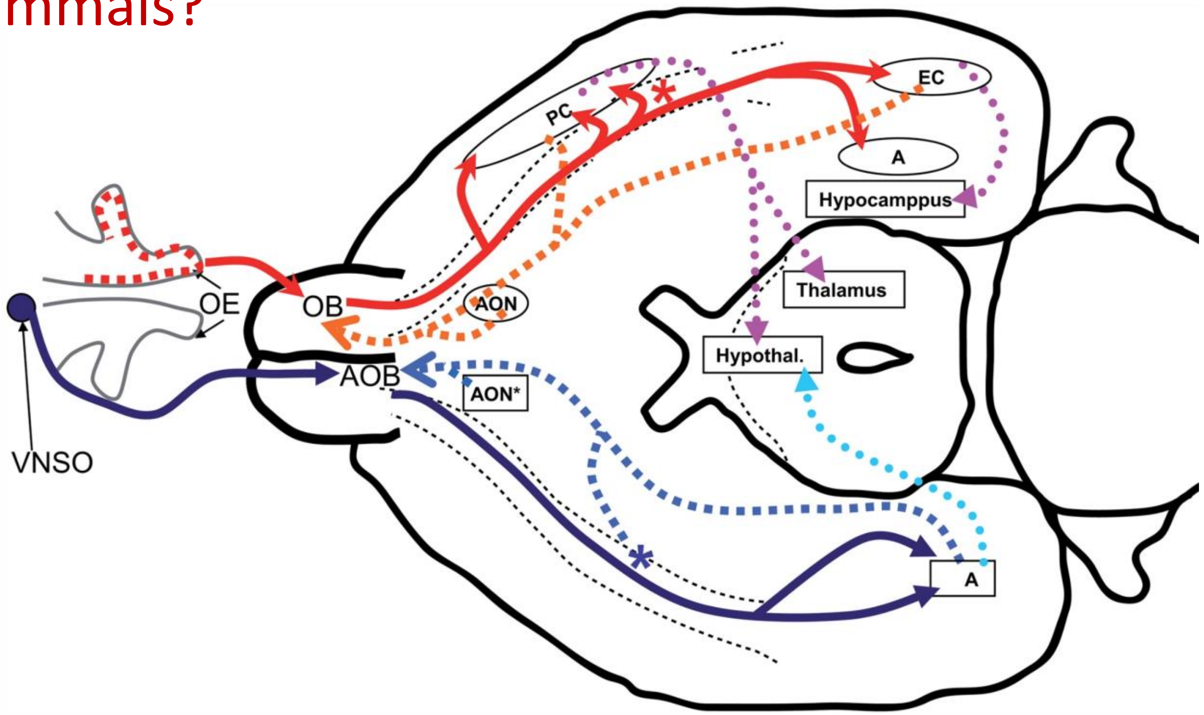
After the fruit fly…

Q: What about mammals?

Insects (fruit fly) vs Mammals (mice)

Q: Does something similar happen in mammals?

YES!



K. Franks, M. Russo, S. Sosulki, A. Mulligan, S. Siegelbaum, R. Axel
“Recurrent Circuitry Dynamically Shapes the Activation of Piriform Cortex,” Neuron, October 2011

From the *Discussion* section of Franks *et al.*

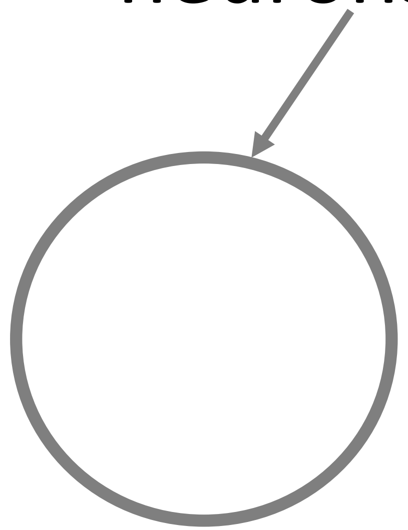
An odorant may [cause] a small subset of [PC] neurons [to fire].

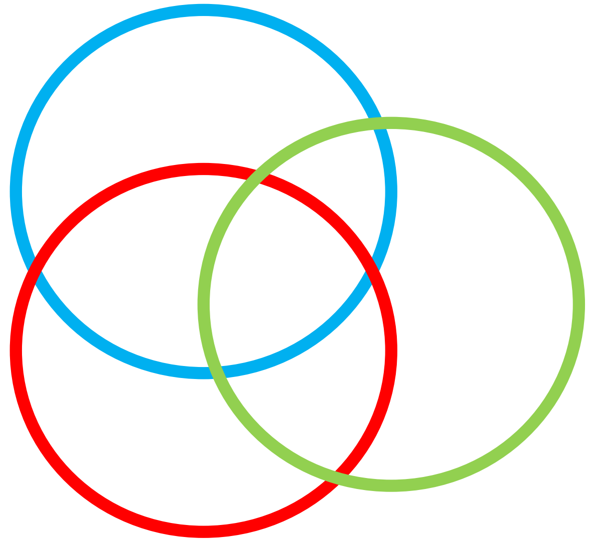
Inhibition triggered by this activity will prevent further firing

This small fraction of ... cells would then generate sufficient recurrent excitation to recruit a larger population of neurons.

In the extreme, some cells could receive enough recurrent input to fire ... without receiving [initial] input...

Recurrent Activation





Does this process converge? And does it preserve similarity?

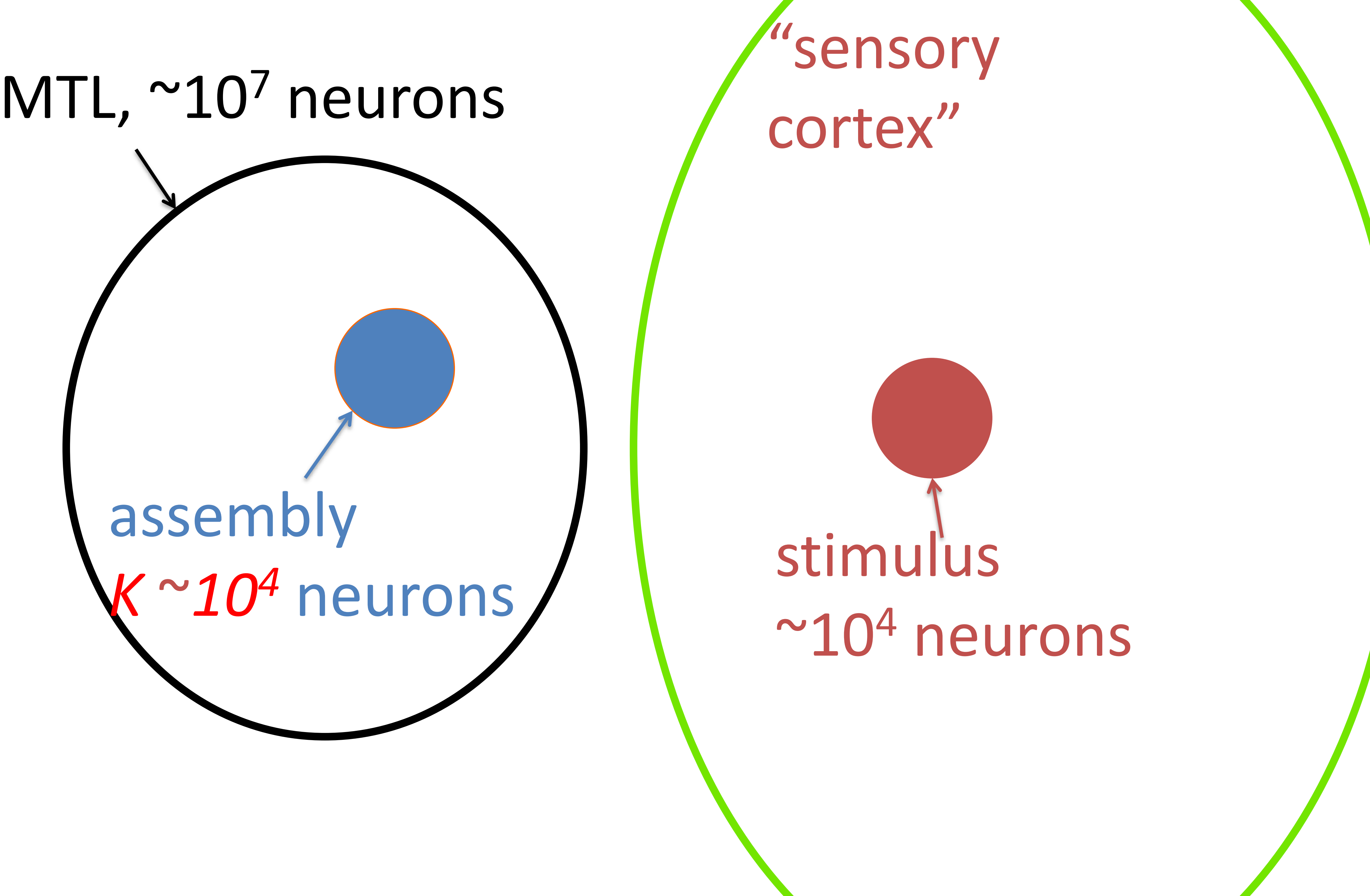
Repeat:
(1) Top k neurons fire,
(2) Synapse weights update

Does this process converge?

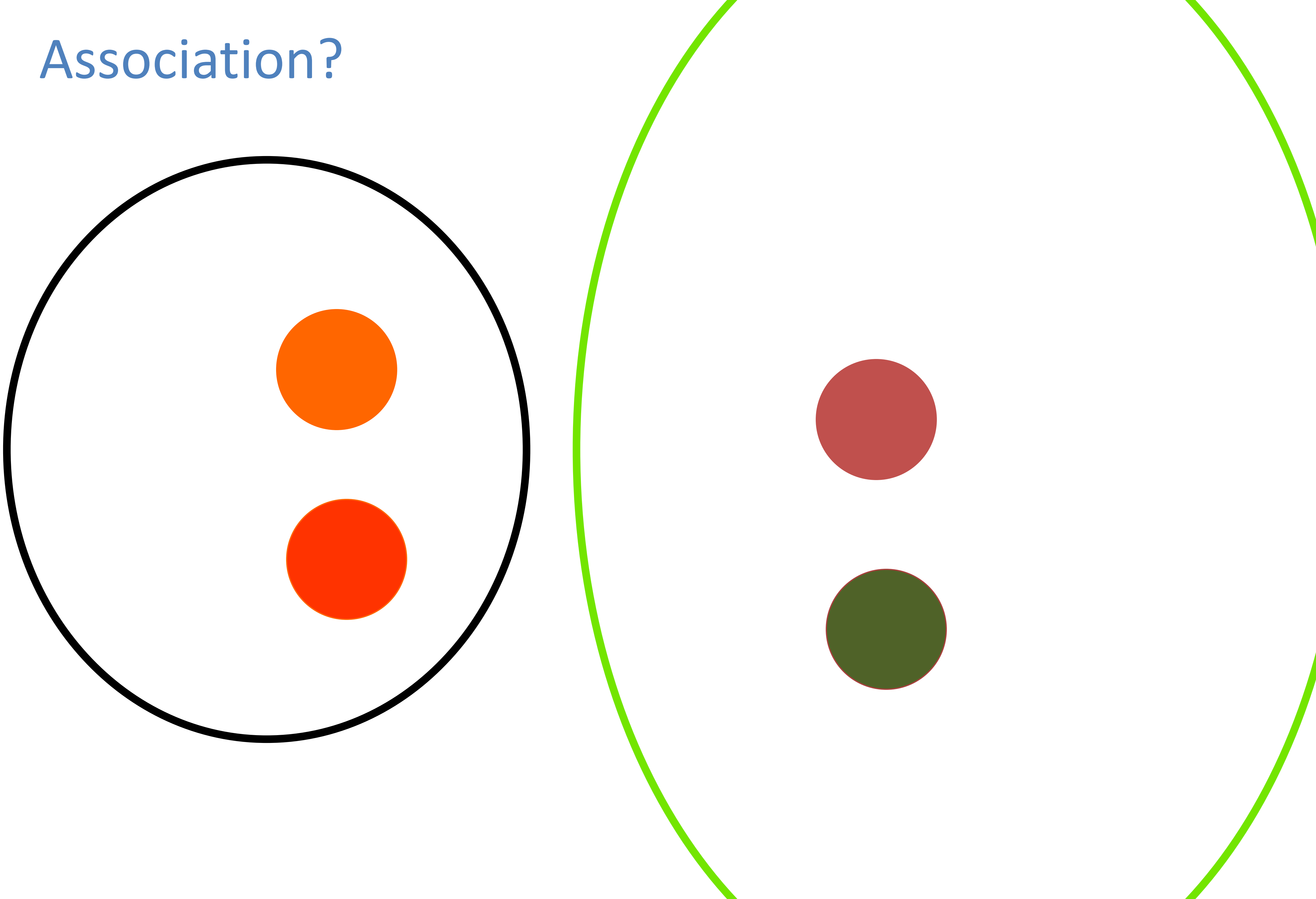
And does it preserve similarity?

Speculating on the Hardware

- Each memory is represented by an *assembly* of ***many*** (~10^4-10^5) neurons; *cf* [Hebb 1949], [Buzsaki 2003, 2010]
- *Highly connected*, therefore stable
- It is somehow formed by sensory stimuli
- Every time we think of this memory, most of these neurons fire
- Two memories can be associated by increasing their overlap


MTL, ~10^7 neurons
assembly
K ~10^4 neurons
"sensory cortex"
stimulus
~10^4 neurons

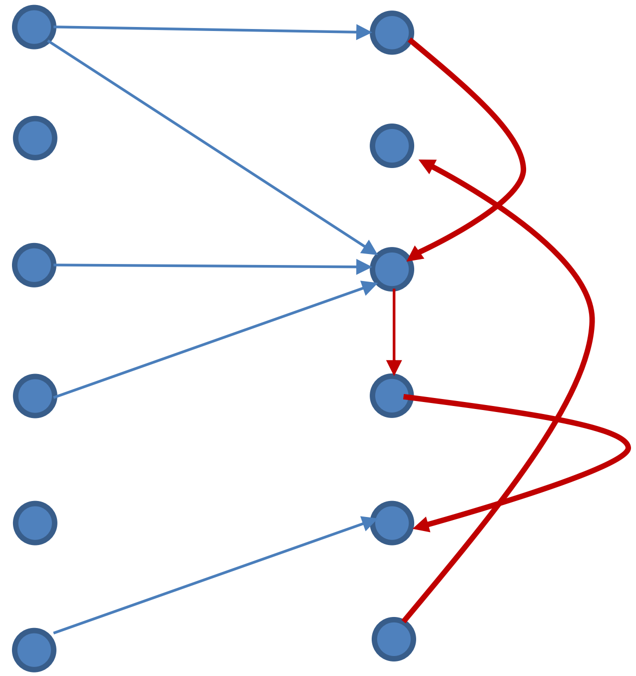
Association?



- demo

A Random brain model

Fruit fly ++:



- A brain area has n excitatory neurons and recurrent connectivity
- Random connections with prob. p
- k-cap
- Hebbian plasticity: synaptic weight $i \rightarrow j$ increased by β, or multiplied by $(1+\beta)$

W… arized system

"probability" of activation

Input from stimulus

$$x_i(t+1) = s_j + \Sigma_{i \to j} x_i(t)\, w_{ij}(t)$$

synaptic weights

(additive) plasticity

$$w_{ij}(t+1) = w_{ij}(t) + \beta\, x_i(t)\, x_j(t+1)$$

Linearized model: Result

Theorem [Papadimitriou-Legenstein-Maass-V.2018]: The linearized dynamics converges geometrically and with high probability to

$$x_j = s_j + \frac{\sum_{i \to j} x_i^2}{\sum_{i \to j} x_i}$$

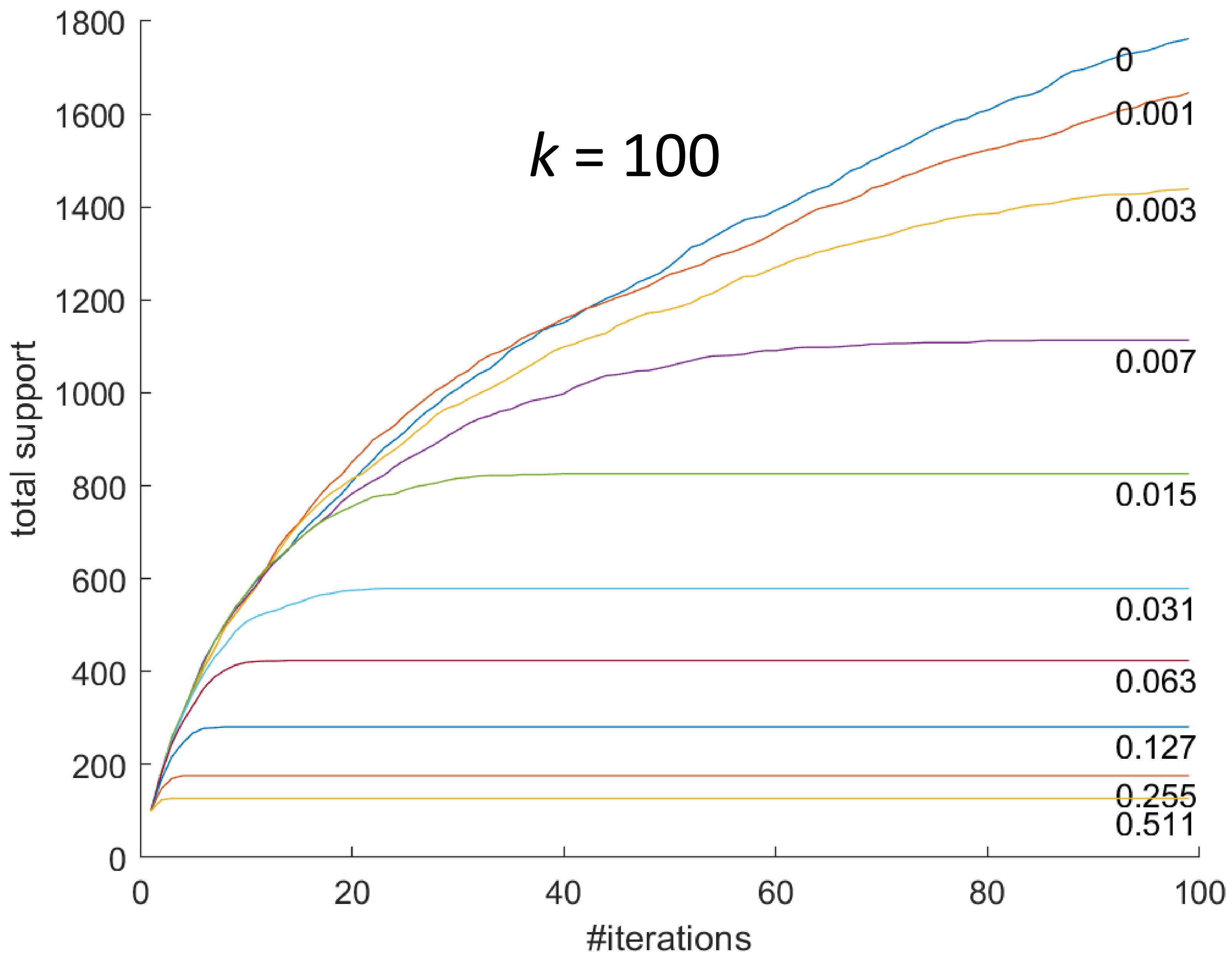
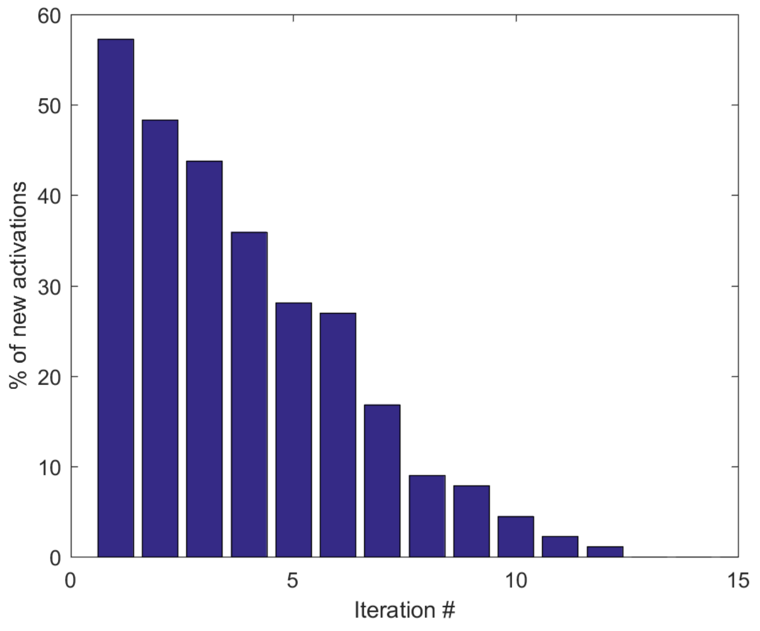
"To be successful, you either
have to be born rich,
or have many successful supporters,
or a little of both"

How about the realistic nonlinear system?

Theorem (Papadimitriou-V. 2019): The process converges exponentially fast, with high probability, and the *total number of cells involved* is at most:

- If $\beta \geq \beta^*$: $k + o(k)$
- If $0 < \beta < \beta^*$: $k \cdot \exp(0.17 \cdot \ln(n/k)/\beta)$

- $\beta^* = \frac{(\sqrt{2} - 1)}{1 + \frac{\sqrt{pk}}{\ln n}}$


k = 100
total support
#iterations
0
0.001
0.003
0.007
0.015
0.031
0.063
0.127
0.255
0.511
1800
1600
1400
1200
1000
800
600
400
200
0
0
20
40
60
80
100

% of new activations
Iteration #
60
50
40
30
20
10
0
0
5
10
15

Proof of Convergence

- First cap A_1: k neurons receiving highest input from stimulus, i.e., top k of n draws from $N(pk, pk)$

- Second cap: k neurons receiving highest input from stimulus+A_1, i.e., top k of n draws from $N(2pk, pn)$
  - Competition between k previous winners and much larger pool.
  - Let $\mu_1, \ldots, \mu_t$ be the fraction of new winners at each step

$$C_1 = pk + \sqrt{2pk \ln \frac{n}{k}}, \qquad C_t = 2pk + 2\sqrt{pk \ln \frac{n}{\mu_t k}}, \qquad t \geq 2.$$

Proof of Convergence

$$C_1 = pk + \sqrt{2pk \ln \frac{n}{k}}, \qquad C_t = 2pk + 2\sqrt{pk \ln \frac{n}{\mu_t k}}, \qquad t \geq 2.$$

- For step 2, for a neuron to stay in the cap, we need:

$$(1+\beta)C_1 + X \geq C_2 \quad \text{with} \quad X \sim N(pk, p(1-p)k)$$

where the $(1+\beta)$ is the plasticity boost for previous round winners

- This determines the threshold β^*
- Winners are more and more likely to survive: $(1+\beta)^i C_i + X \geq C_{i+1}$

Pattern Completion

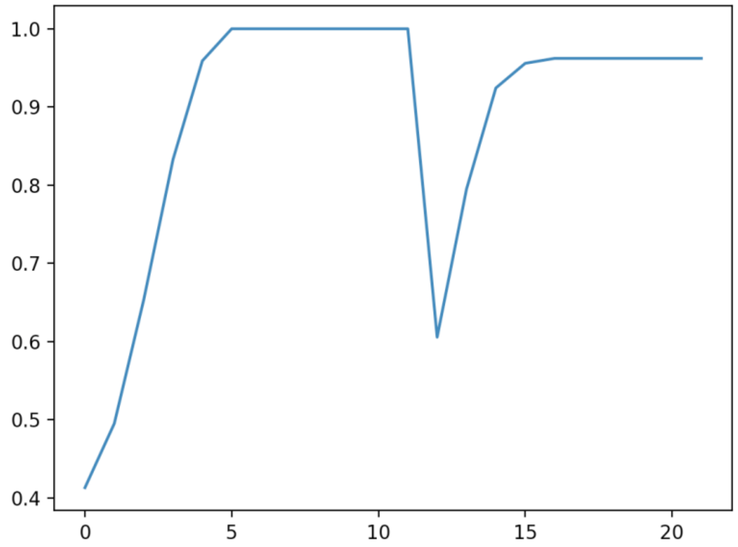
- Recall: Future presentation of the same or *similar* stimulus fires mostly the same assembly.
- What about firing a subset of the assembly itself?
- Yes! Suffices to fire fraction of created stimuli to complete to (almost) the rest.

- Thm [Collins-Mitropolsky-P.-V. 2019]. For any $\varepsilon \in (0,1)$, with $T > T(\varepsilon)$ presentations of a stimulus, the resulting assembly A has the property that firing ε fraction of the assembly results in completing to 90% of A.

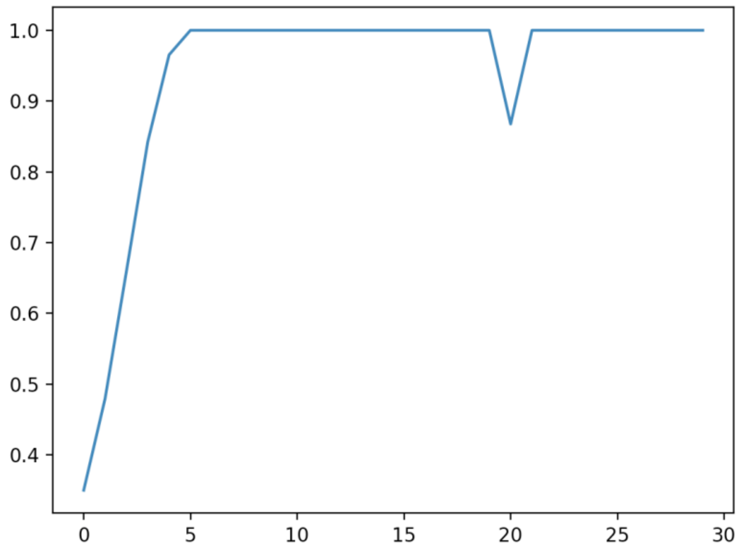
Pattern Completion

- Firing 20% completes to assembly!

#iter = 12



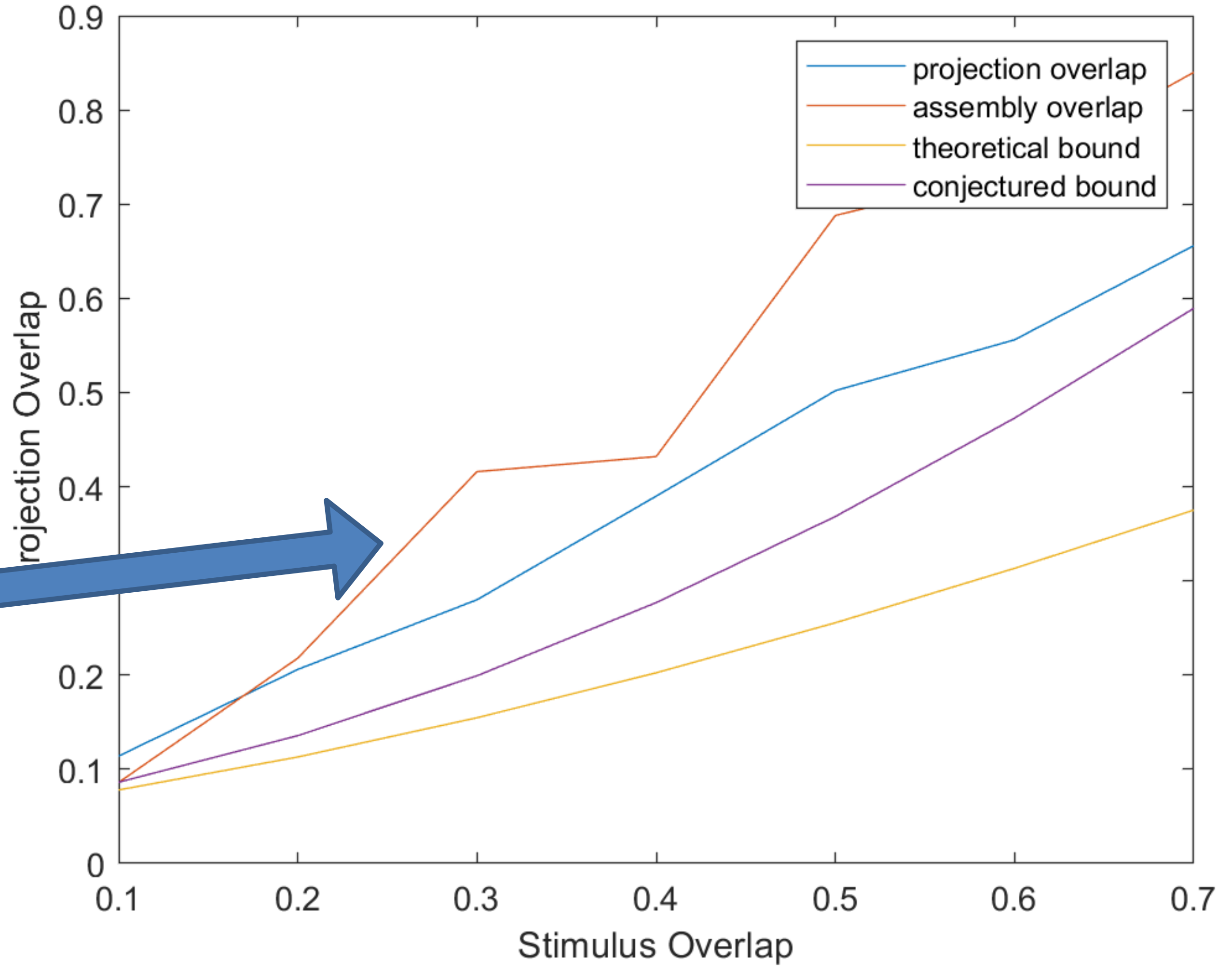
#iter = 20



Similarity preserved by assembly creation

Recall the fly and hashing

Similarity is well-preserved



Why assemblies instead of 1-step projections?

- Pattern completion!

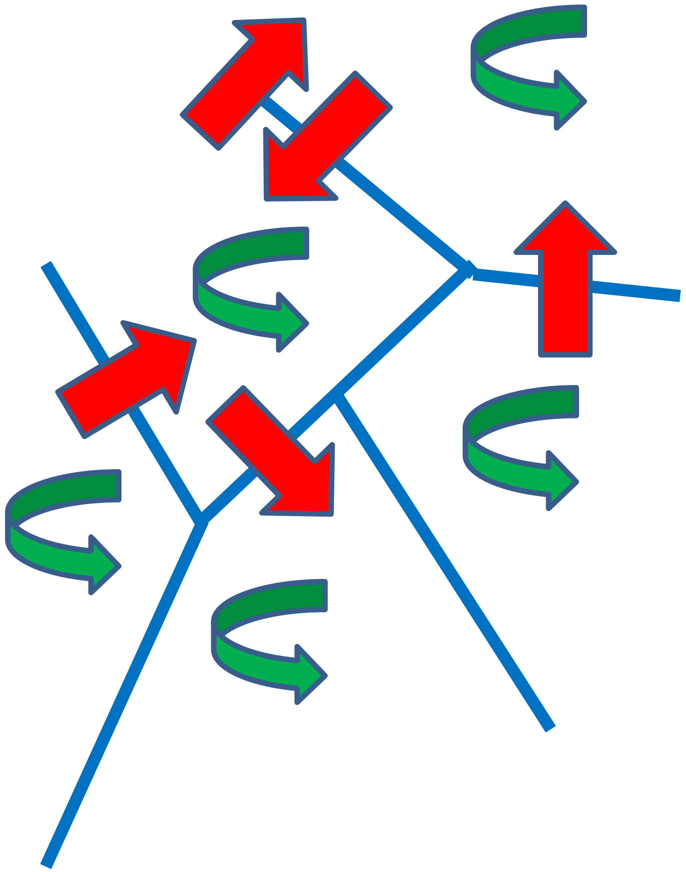
- and learning?
- Similarity is enhanced for sufficiently similar inputs by assembly creation

- Could this be the key to invariants?

The model

- Finite number of brain regions
- Each contains n neurons
- And has a forgetting (decay) constant for synapse weights
- Some pairs of areas are connected by directed bipartite $G_{n,p}$
- Each is connected within by directed $G_{n,p}$

- Neurons fire in discrete steps
- Inhibition: In each region, k with highest input fire
- Plasticity: If i fires and in the next step j fires, $w(i \rightarrow j)$ is multiplied by (1 + β)
- Connections between areas can be enabled/disabled
- (Roughly, $n = 10^7$, $k = 10^4$, $p = 10^{-3}$)



Open Questions

- Assembly support size: phase transition with plasticity parameter?
- Assembly convergence: limiting behaviors? (e.g., limit cycles?)
- Asynchronous models?
- A probabilistic composition lemma? (for reliable downstream assemblies)
- Richer base graph models? (e.g., with triangle completion, higher clustering coefficient etc.)
- Deterministic property replacing $G_{n,p}$?
- Computational power?
- Capacity: how many assemblies can be reliably supported?

- What is Brain Learning? How does the brain create invariants?
- How is language accomplished?

later today:

CP on “An Assembly Calculus for the Brain” !!!

Thank You!