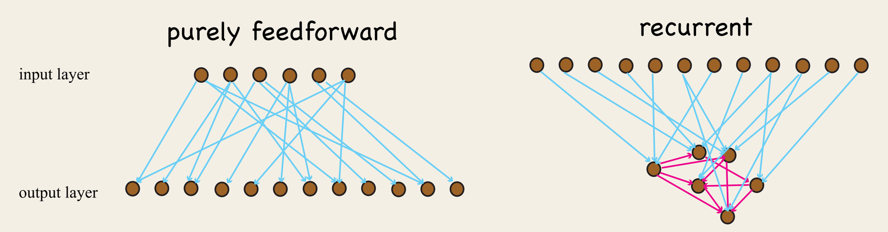
Recurrent vs. feedforward networks: differences in neural code topology

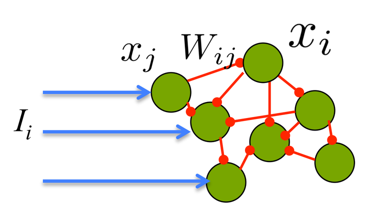
Vladimir Itskov[1], Anda Degeratu[2], Carina Curto[1]
[1]Department of Mathematics, University of Nebraska-Lincoln; [2]Albert-Ludwigs-Universität Freiburg, Germany.

Question:

Is there anything a recurrent network can do that feedforward network can not?



Stimulus representation by a neural network with attractor dynamics



$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \varphi(\sum_{j=1}^{n} W_{ij}x_j + I_i),$$

$x_i \geq 0$

the network receives a slowly-changing input I and evolves to an attractor x*

stimulus → steady state x*

Neural code of a network (combinatorial code)

stimulus → steady state x* → stable set

stable set = a set of neurons that can fire together at a steady state in response to a particular input

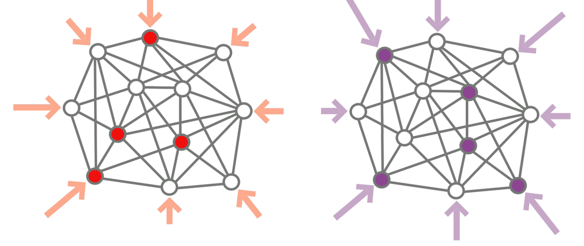
$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \varphi(\sum_{j=1}^{n} W_{ij}x_j + I_i),$$

$$x(t) \longrightarrow x^* \text{ as } t \to \infty$$

$$\sigma = \text{supp}(x^*) = \{i \in \{1,\ldots,n\} \mid x_i^* > 0\}$$

Neural code =collection of all stable sets,

i.e. all the combinatorial patterns of activity observed at all the attractors



Note that here we disregard all information about the input-output relationship.

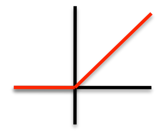
Question: Can we directly relate the set of encodable memory patterns to properties of the synaptic matrix?
Answer: Yes.

a standard firing rate model:

$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \varphi(\sum_{j=1}^{n} W_{ij}x_j + I_i),$$

$x_i \geq 0$

$$\varphi(x) = [x]_+ = \max(0, x)$$



$$\dot{x} = -Dx + [Wx + I]_+$$

$$D = \text{diag}\left(\frac{1}{\tau_1}, \frac{1}{\tau_2}, \ldots, \frac{1}{\tau_n}\right)$$

Observation: If all inputs I to the network are possible, then

σ is a stable set if and only if $(-D+W)_\sigma$ is a stable matrix.

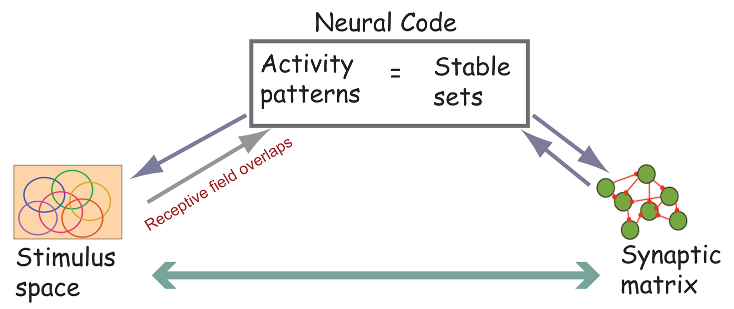
Neural code = all stable principal submatrices of (-D+W)

A principal submatrix is obtained by restricting to a subset of indices:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \quad A_{\{124\}} = \begin{pmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{pmatrix}$$

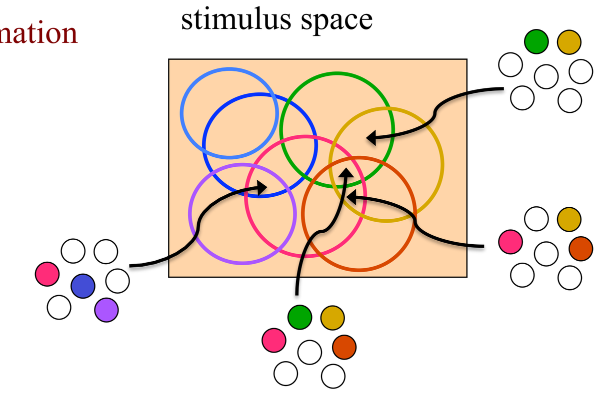
A matrix is stable iff all its eigenvalues have negative real part.

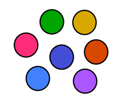
Relating the network structure to the represented stimuli



what cells can fire together ⟷ receptive field intersection information

If neurons have convex receptive fields, then allowed activity patterns are determined by receptive field intersections.

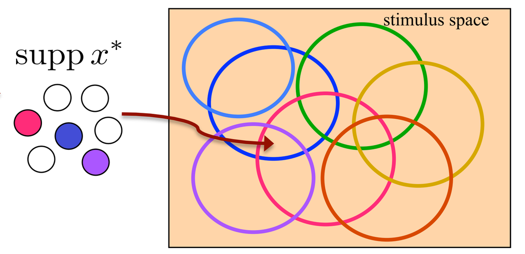



7 example neurons, each with a RF

In a network with attractor dynamics, receptive field overlaps correspond to the stable sets:

$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \varphi(\sum_{j=1}^{n} W_{ij}x_j + I_i),$$

slow-changing inputs I_i → steady state x* →

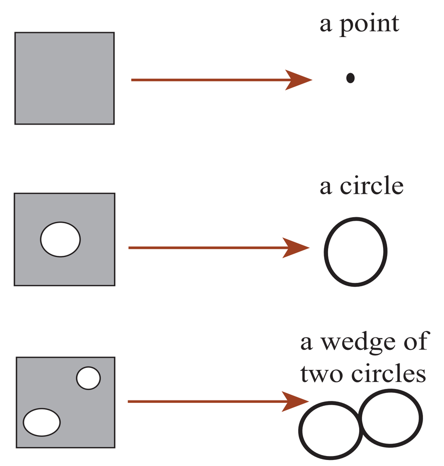


Question: What features of the represented stimuli are captured by the combinatorics of the neural code itself without the information about RFs?

Answer: Topological features of the space of represented stimuli.

If neurons have *convex receptive fields*, then the *homotopy type* of the space of represented stimuli is completely determined by the combinatorics of the neural code.

Q: What is homotopy type?
A: Two spaces have the same homotopy type of one can be deformed into another by a continuous transformation. Note that this can change the space's dimension.



Why does the combinatorics of the neural code yield topological information about the space?

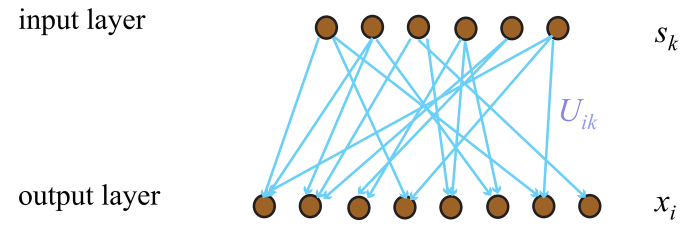
[1 2], [2 3], [3 4], [4 5], [5 6], [6 7], [7 1]

[1 2], [2 3], [3 5], [2 5] , [2 4], [1 4], [4 5], [4 6], [1 2 4], [2 3 5], [2 4 5]

Question: Can a network encode any homotopy type?

Two extreme architectures:

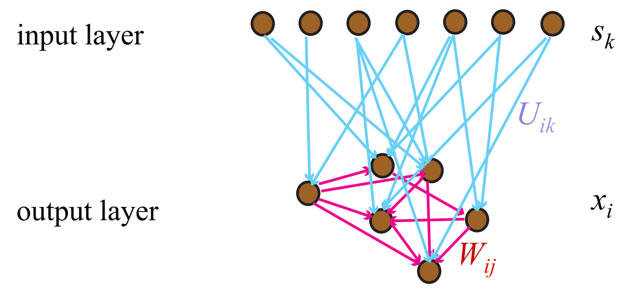
purely feedforward



$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \phi\left(\sum_{k=1}^{m} U_{ik}s_k + \theta_i\right)$$

The neural code is determined by the structure of the two-layer perceptron

recurrent



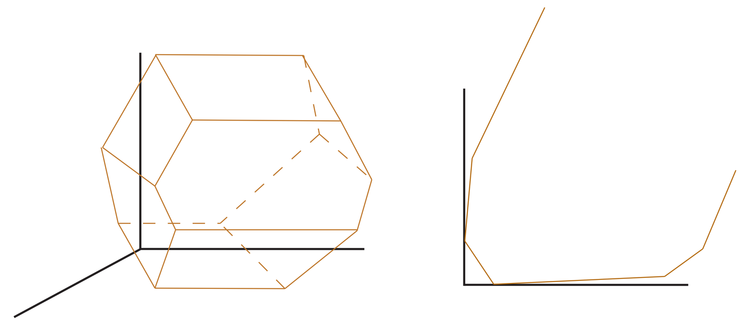
$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \phi\left(\sum_{j=1}^{N} W_{ij}x_j + \sum_{k=1}^{m} U_{ik}s_k + \theta_i\right)$$

If the input layer is larger than the output layer, then the neural code is determined only by the stable submatrices of (-D+W)

The Main Result (a theorem):

Recurrent networks can represent spaces with any prescribed homotopy type, but for purely feedforward networks, the homotopy type is highly restricted.

- Excitatory purely feedforward networks: Can only represent contractible spaces, like balls (i.e., with homotopy type of a point).
- Purely feedforward networks: (both E and I synapses) Can only represent spaces with homotopy type of an orthant with a convex region cut out.



In contrast,

- Recurrent networks: Can encode spaces with any prescribed homotopy type!!

Conclusions:

- It is possible to directly relate topological features of the space of represented stimuli to the synaptic matrix.
- Recurrent networks can represent stimuli with any prescribed homotopy type, while purely feedforward networks can not.

References:

[1] C. Curto, V. Itskov. Cell groups reveal structure of stimulus space. PLoS Computational Biology 4(10), 2008.
[2] C. Curto, A. Degeratu, V. Itskov. Flexible memory networks. Bulletin of Mathematical Biology, 74:590-614, 2012.
[3] C. Curto, A. Degeratu, V. Itskov. Network architecture imposes topological constraints on represented stimuli. 2012 (*to be submitted*)