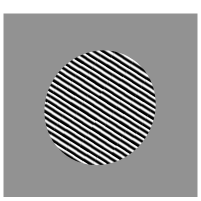
Dimensionality and dynamics in neural networks

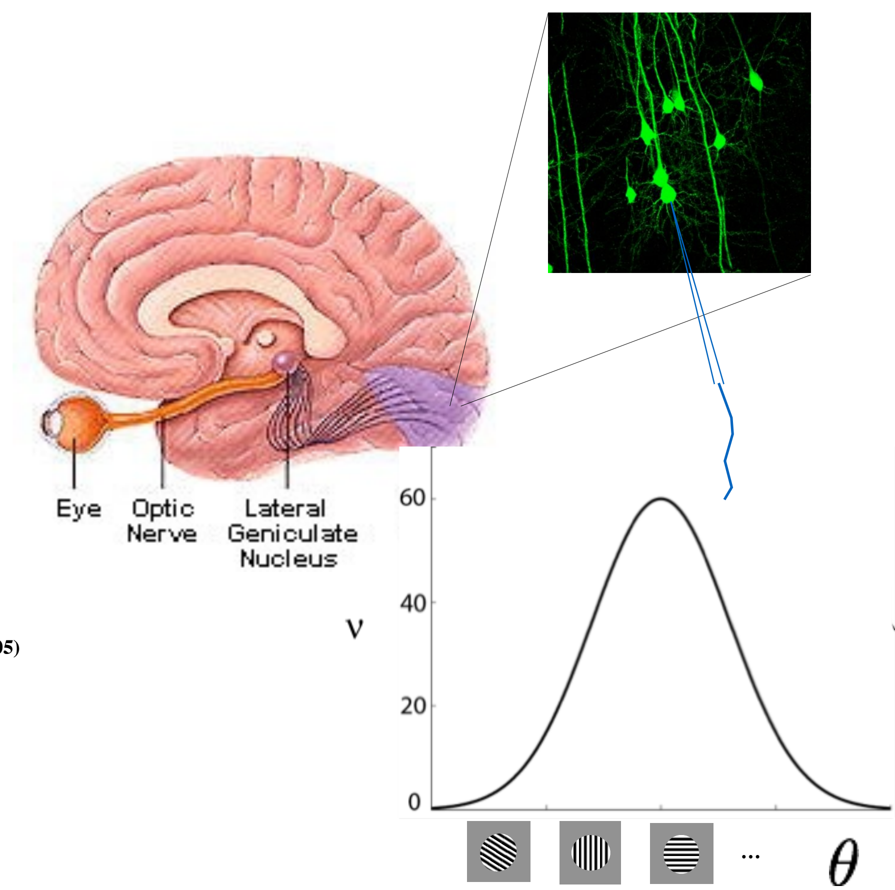
Eric Shea-Brown	(University of Washington)
Stefano Recanatesi	(University of Washington)
Matt Farrell	(University of Washington)
Merav Stern	(University of Washington)
Michael Buice	(Allen Institute)
Gabe Ocker	(Allen Institute)
Guillaume Lajoie	(U. Montreal)
Mattia Rigotti	(IBM)
Madhu Advani	(Harvard)
Sophie Deneve	(ENS)
David Dahmen	(Julich)
Moritz Helias	(Julich)

UNIVERSITY of WASHINGTON
COMPUTATIONAL NEUROSCIENCE CENTER

UNIVERSITY of WASHINGTON
APPLIED MATHEMATICS

The neural code

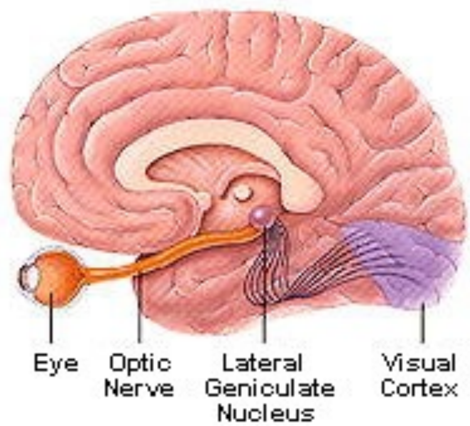


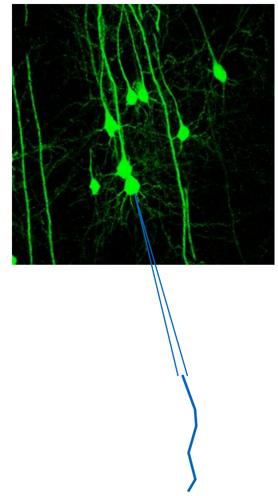


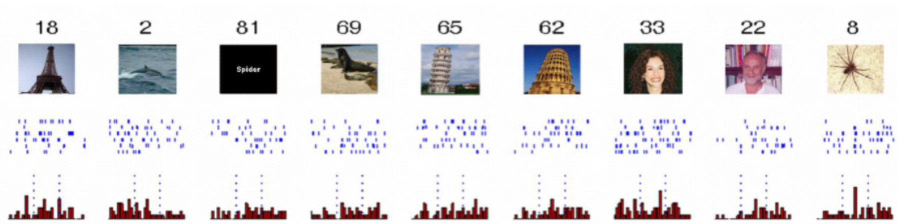
Hubel and Weisel. J. Physiol.,1962

Reddy, Kreiman, Koch, and Fried (2005)

The neural code

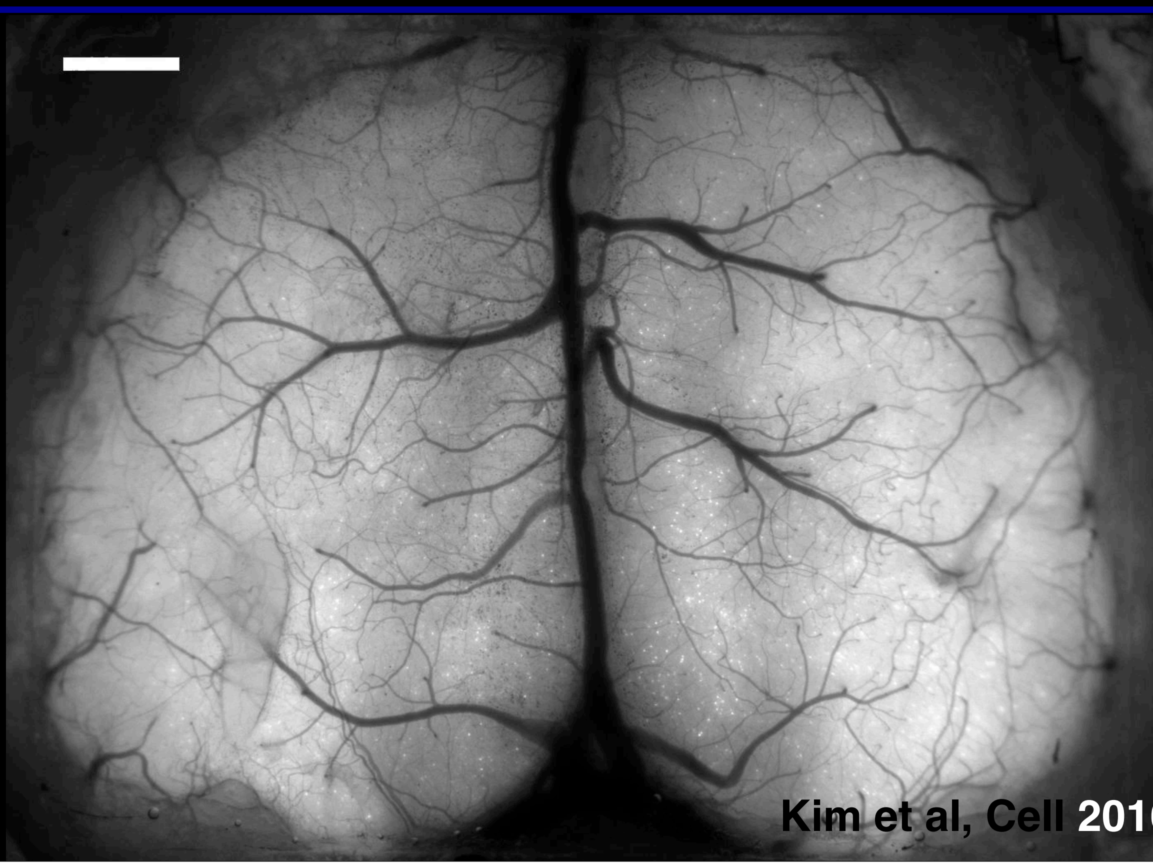






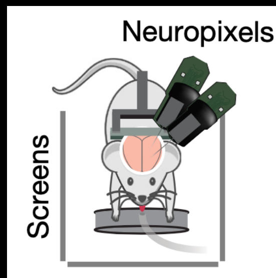
Quiroga, Koch et al, Nature 2005

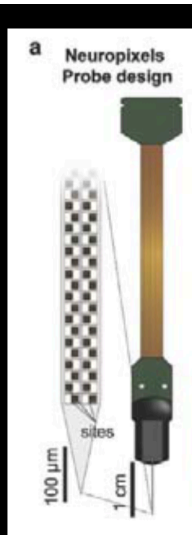
The neural code today

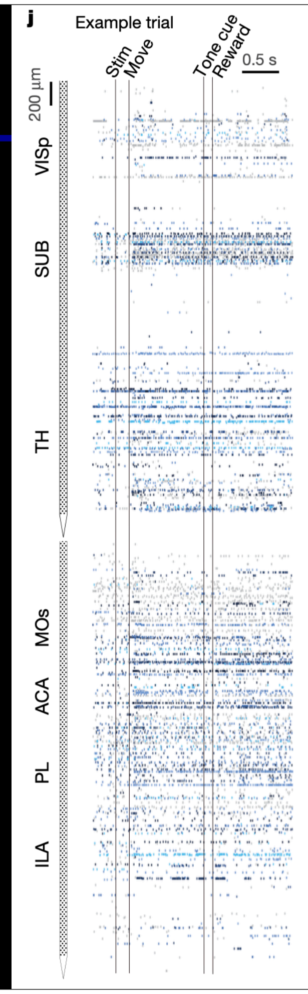


Kim et al, Cell 2016

The neural code today







Jun et al, 17

Steinmetz et al, Nature '19

Question 1:

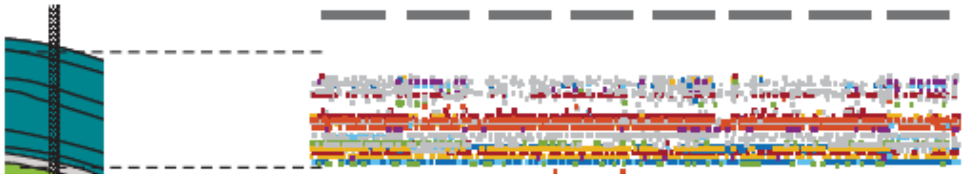
How different is this neural code?

A big question! Simple(st?) version:

How many dimensions does neural activity really fill up?

How many dimensions does neural activity really fill up?

Probe 1



Jun el al, 2017

$$\begin{bmatrix}4\\10\\\vdots\\5\end{bmatrix}\begin{bmatrix}11\\3\\\vdots\\12\end{bmatrix}\cdots$$

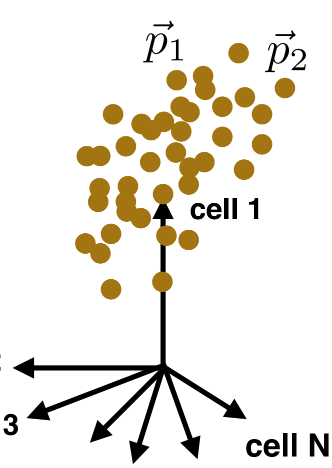
$$\vec{p}_1 \quad \vec{p}_2 \quad \cdots$$

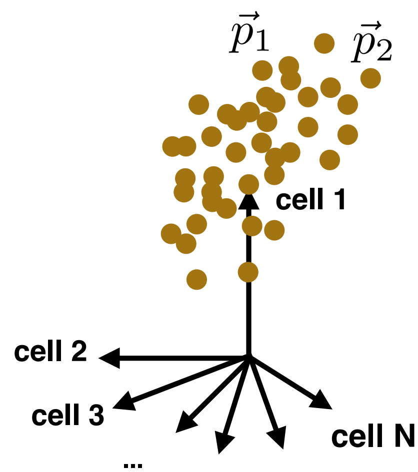
How many dimensions does neural activity really fill up?

Probe 1

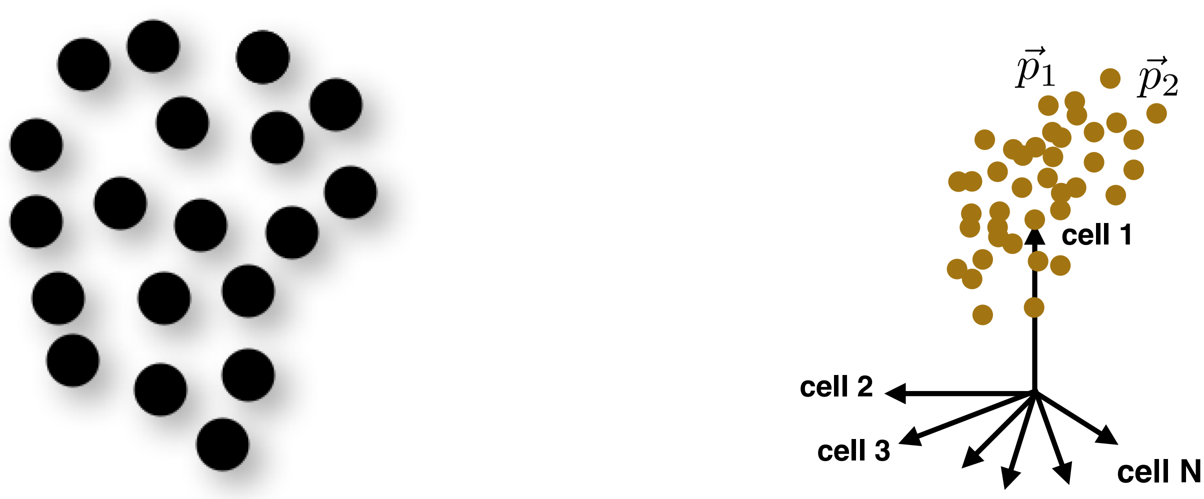
Jun el al, 2017

$$\begin{bmatrix}4\\10\\\vdots\\5\end{bmatrix}\begin{bmatrix}11\\3\\\vdots\\12\end{bmatrix}\cdots$$

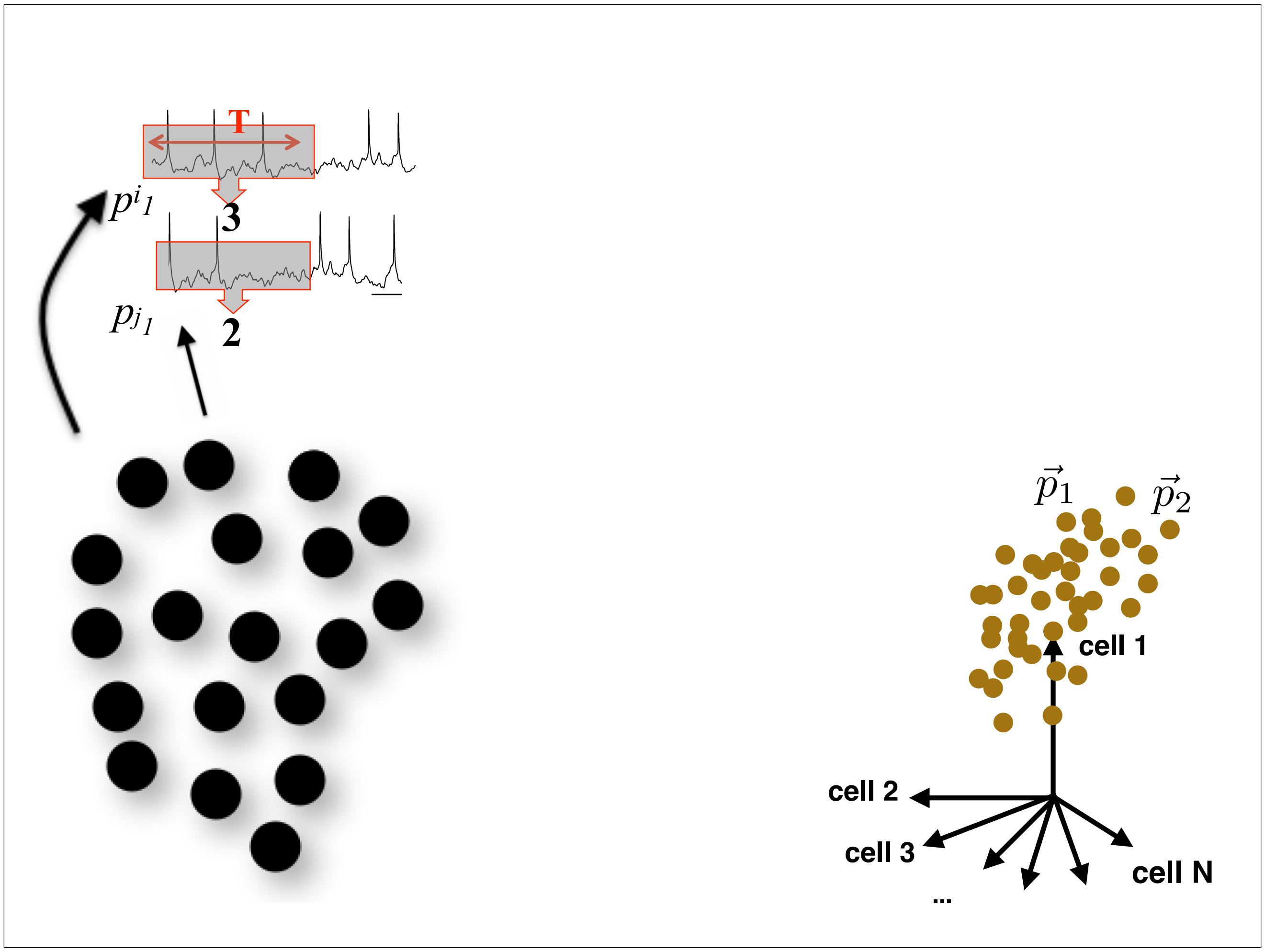




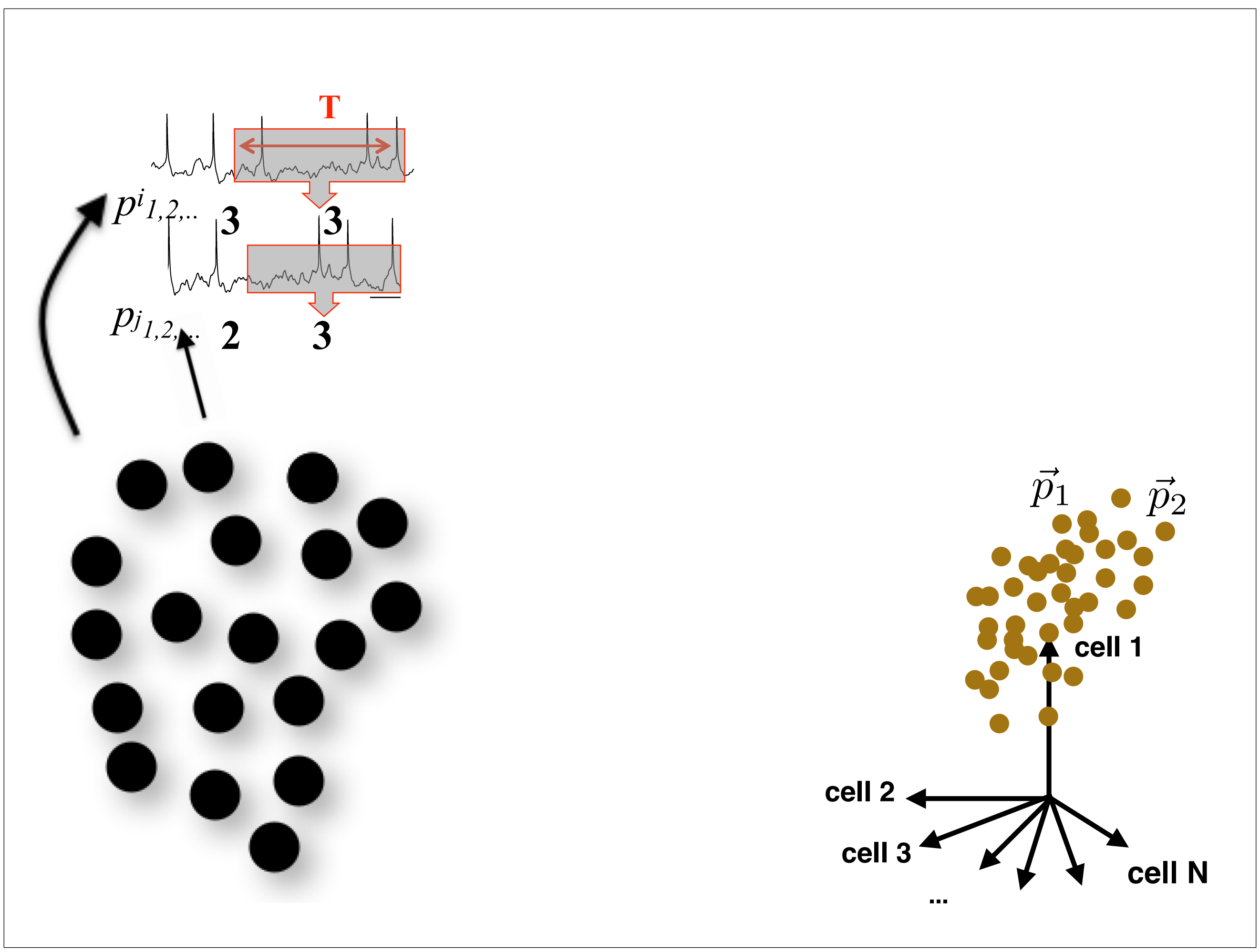
$\vec{p}_1$
$\vec{p}_2$
cell 1
cell 2
cell 3
...
cell N



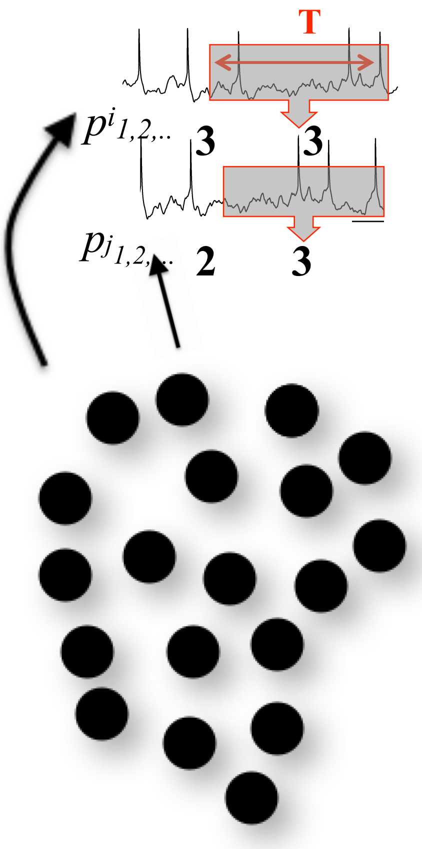
$\vec{p}_1$
$\vec{p}_2$
cell 1
cell 2
cell 3
...
cell N



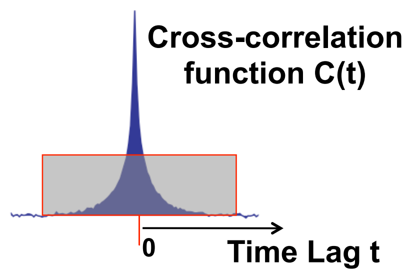
T
$p^i{}_1$
3
p_{j_1}
2
$\vec{p}_1$
$\vec{p}_2$
cell 1
cell 2
cell 3
...
cell N



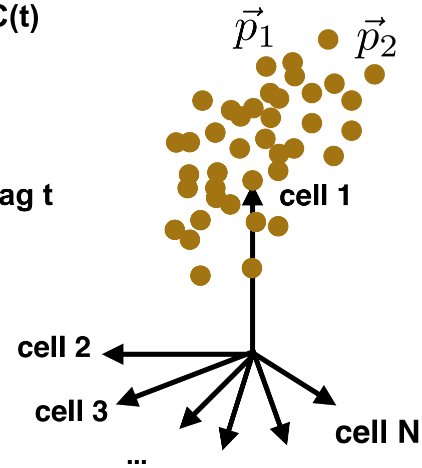
T
$p^i{}_{1,2,..}$
3
3
$p_{j_{1,2,..}}$
2
3
$\vec{p}_1$
$\vec{p}_2$
cell 1
cell 2
cell 3
...
cell N

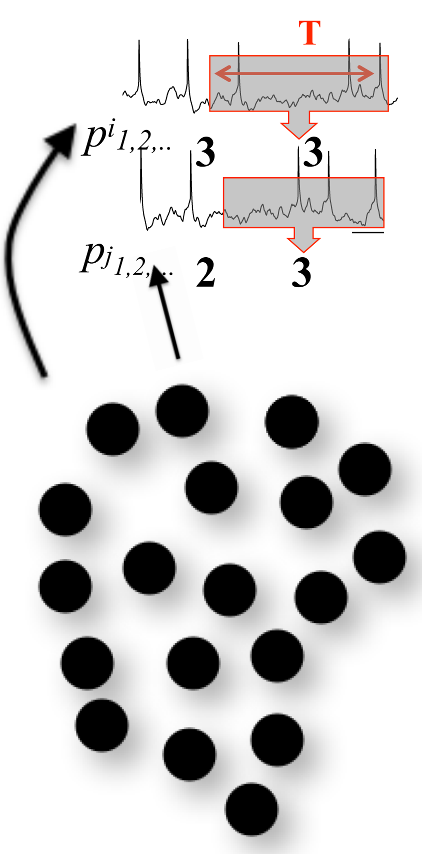


$$Cov(p^i, p^j) = \mathbb{E}\left[(p^i - \mathbb{E}(p^i))(p^j - \mathbb{E}(p^j))\right]$$
$$= C_{ij}$$

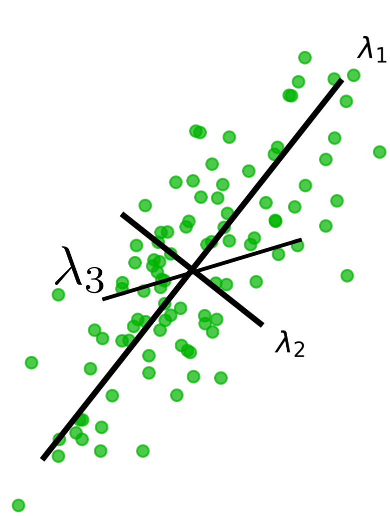


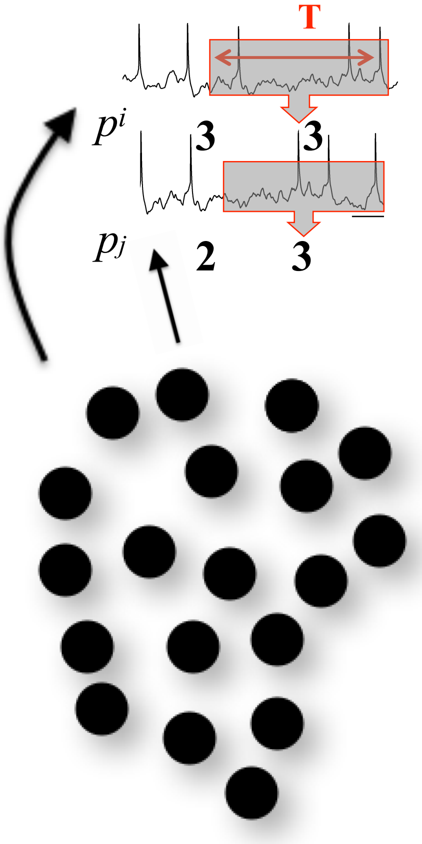
Bair et al
J. Neurosci ‘01





$$Cov(p^i, p^j) = \mathbb{E}\left[(p^i - \mathbb{E}(p^i))(p^j - \mathbb{E}(p^j))\right]$$
$$= C_{ij}$$

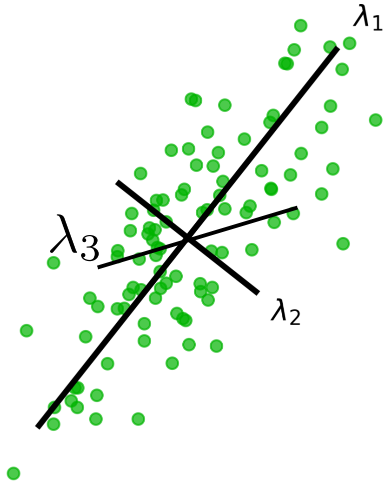


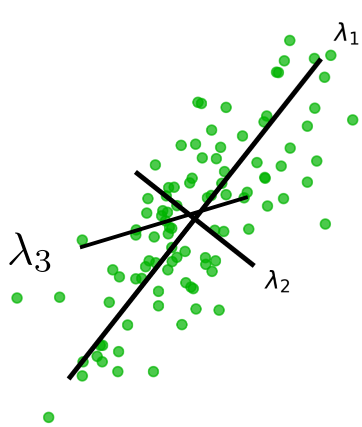


$$Cov(p^i, p^j) = \mathbb{E}\left[(p^i - \mathbb{E}(p^i))(p^j - \mathbb{E}(p^j))\right]$$
$$= C_{ij}$$

eigenvectors $\{v_i\}$
eigenvalues $\{\lambda_i\}$

$\langle C \rangle = \langle C_{ij} \rangle$
average cov.

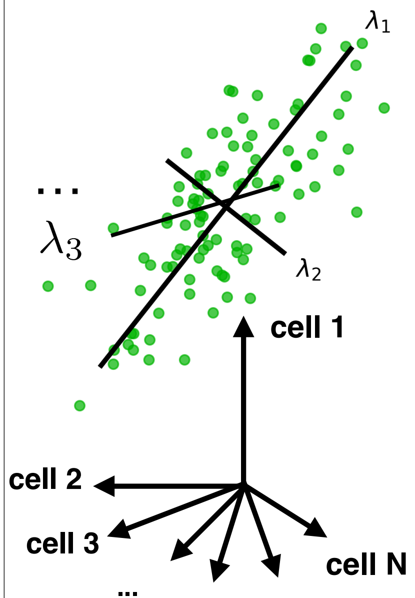


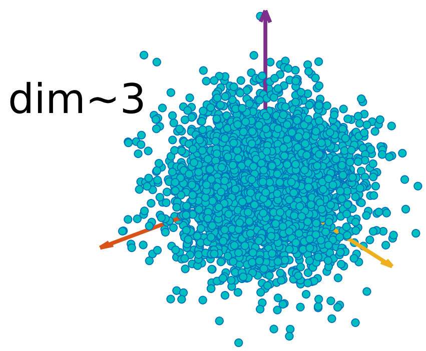


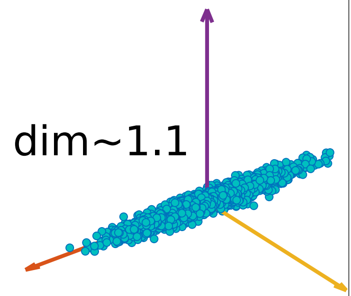
How many dimensions does neural activity really fill up?

Abbott et al, ’11
Mazzucato et al. ’15
Gao and Ganguli ’17
Cayco-Gajic et al ’17
Litwin-Kumar et al ‘17

$$\tilde{\lambda}_i = \frac{\lambda_i}{\sum_j \lambda_j} \qquad dim = \frac{1}{\sum_i \tilde{\lambda}_i^2}$$





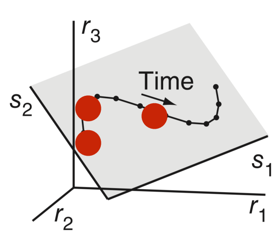


$\tilde{\lambda}_1 = 1/3$	$\tilde{\lambda}_1 = .95$
$\tilde{\lambda}_2 = 1/3$	$\tilde{\lambda}_2 = .025$
$\tilde{\lambda}_3 = 1/3$	$\tilde{\lambda}_3 = .025$
$\tilde{\lambda}_4 = 0$	$\tilde{\lambda}_4 = 0$
$\cdots$	$\cdots$
$\tilde{\lambda}_N = 0$	$\tilde{\lambda}_N = 0$

How many dimensions does neural activity really fill up?
Many Implications

Low dim:

- **denoising**
- **decoding from fewer cells**
- **generalization from fewer examples**



Review:
Cunningham and Yu ’14

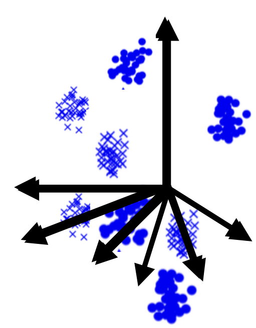
Goldilocks dim:

Good classification
+ good generalization

Reviews:
Fusi, Miller, Rigotti ’16
Bengio ’09 “Learning Deep Architectures”
DiCarlo and Rust ’12
Anselmi and Poggio ‘14

High dim:

easy classification



Reviews:
Hertz, Krough, Palmer ’91;
Vapnik ‘00

How many dimensions does neural activity really fill up?

Many Approaches

<u>Measure empirically, and interpret!</u>

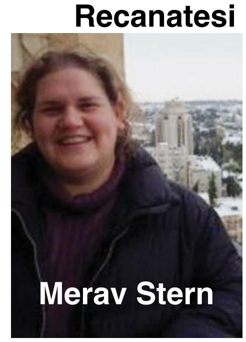
e.g. Stinger, Pachitariu et al, Science, Nature '19
Yu, Cunningham Nature Rev. Neurosci. '14
Gao and Ganguli, ArXiv 2017

<u>Model computationally, and propose origins in network dynamics</u>

Our focus in these 2 lectures

How many dimensions does neural activity really fill up?

Stefano Recanatesi



<u>PART 1 — Bottom up:</u>

Connectivity → dimension?

<u>PART 2 — Top down:</u>

Task → dimension?

What about connectivity matters for network-wide dimension ?

A: Something strong, even when avg. correlations <C> are low.

Low **<C> found throughout brain:**

DATA: LGN: ***Alonso et al 1996 …*** **V1:** ***Kohn and Smith 2005, Ecker et al 2010 …*** **Motor cortex:** ***Vaadia et al 1995***

THEORY: ***Van Vreeswijk/Sompolnisky '96, Renart, de la Rocha et al 2010; Doiron, Rosenbaum 2010's: THE BALANCED STATE***

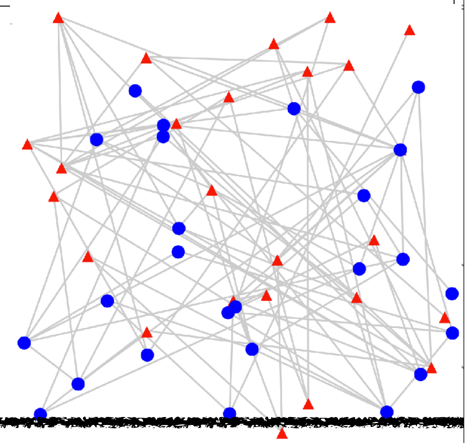
$$\langle C_{ij} \rangle \approx \frac{c}{N} \qquad Var\left(C_{ij}\right) \approx \frac{k}{N}$$

Renart, de la Rocha et al 2010 (balanced networks)

$$dim = f\left(\langle C_{ij} \rangle, Var\left(C_{ij}\right)\right)$$

Mazzucato et al. 2015. Litwin-Kumar et al. 2017

$$\Longrightarrow \lim_{N\to\infty} \frac{dim}{N} = \frac{1}{1+k}$$



Weak pairwise correlations imply strongly correlated network states in a neural population

Elad Schneidman[1,2,3], Michael J. Berry II[2], Ronen Segev[2] & William Bialek[1,3]

What about connectivity matters for network-wide dimension ?

Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks

Francesca Mastrogiuseppe • Srdjan Ostojic [3] • Show footnotes

low-rank connectivity

RESEARCH ARTICLE

Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models

Ryan C. Williamson[1,2,3], Benjamin R. Cowley[1,3], Ashok Litwin-Kumar[4], Brent Doiron[1,5], Adam Kohn[6,7,8], Matthew A. Smith[1,9,10,11], Byron M. Yu[1,12,13]*

Circuit Models of Low-Dimensional Shared Variability in Cortical Networks

Chengcheng Huang[1,2], Douglas A. Ruff[2,3], Ryan Pyle[4], Robert Rosenbaum[4,5], Marlene R. Cohen[2,3], Brent Doiron[1,2,6]

spatial / temporal connectivity

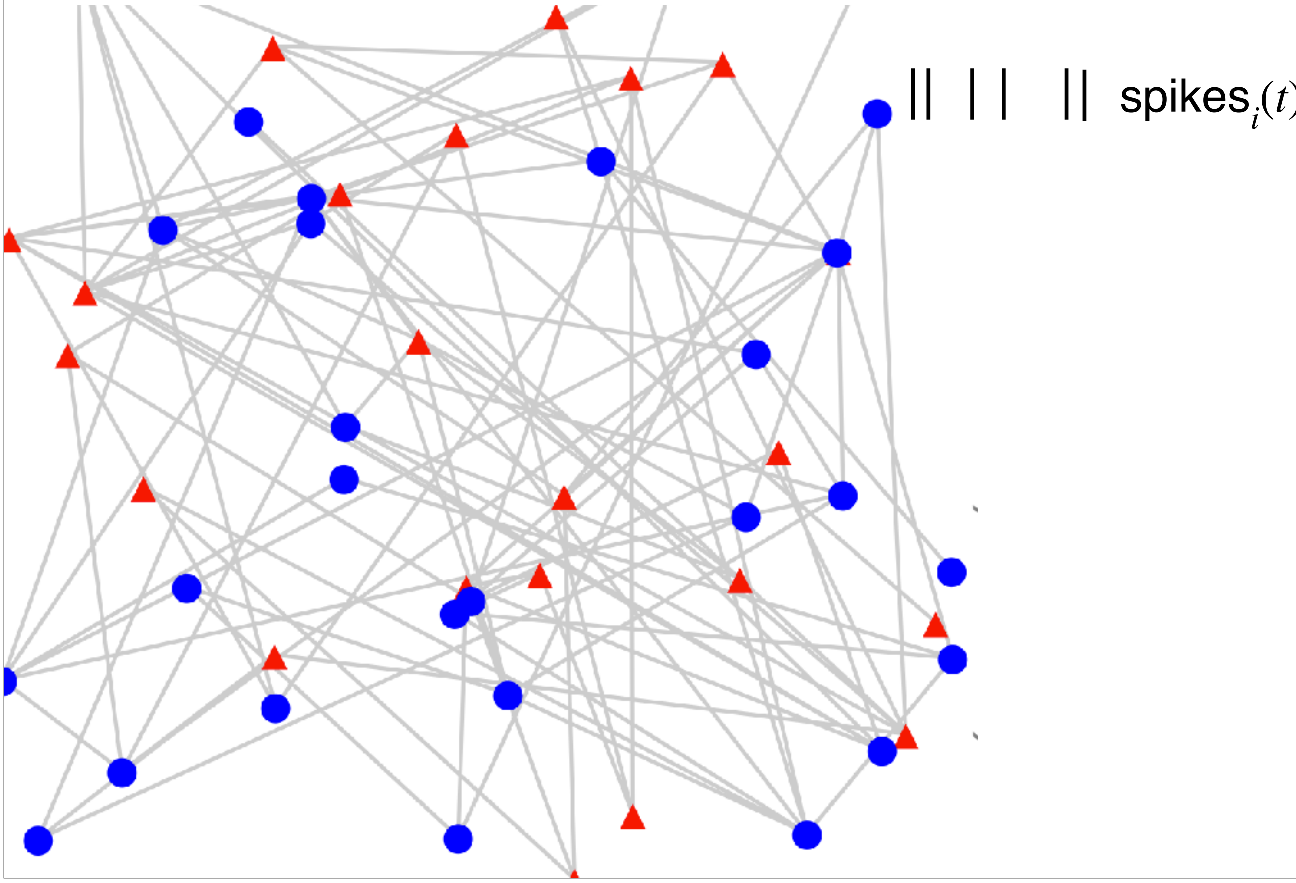
[Yu Hu et al, 13, 14, 17
Ocker, Buice et al 18
Recanatesi et al 19
Dahmen, Recanatesi et al '21]

network motifs local connectivity

… limited regime, but general answer

What about connectivity matters for network-wide dimension ?

Limit to linear response regime for spiking networks



What about connectivity matters for network-wide dimension ?

Limit to linear response regime for spiking networks

spikes$_i(t)$ is a point process with intensity $rate_i(t)$

$$\text{spikes}_i(t) = \sum_{t_i} \delta(t - t_i)$$

$$\Pr\left[t_i \in [t, t + dt] = \text{rate}_i(t)\, dt\right]$$

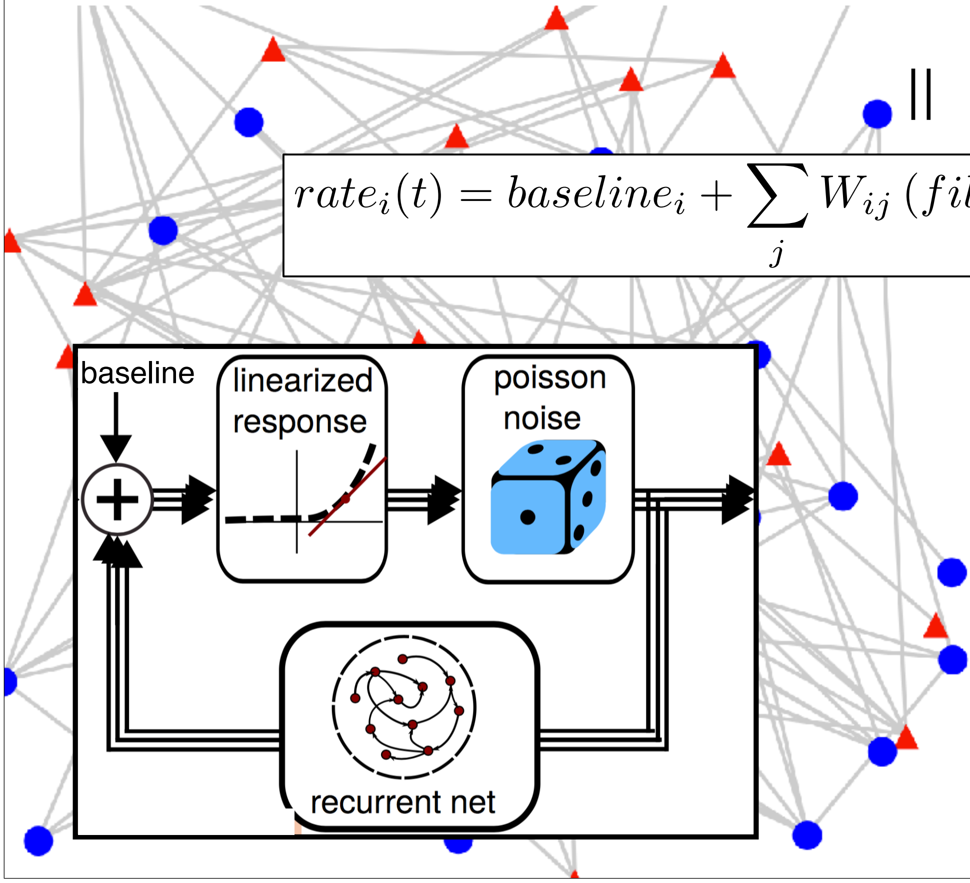
Point Processes, Cox and Isham '80

What about connectivity matters for network-wide dimension ?

Limit to linear response regime for spiking networks

$||\ |\ |\ \ \ ||\ \ spikes_j(t)$

$$rate_i(t) = baseline_i + \sum_j W_{ij}\left(filter * spikes_j(t)\right)$$



Coupled
point process

Hawkes process
(Hawkes, '72)

Brown, Kass et al J. Neurophys '04

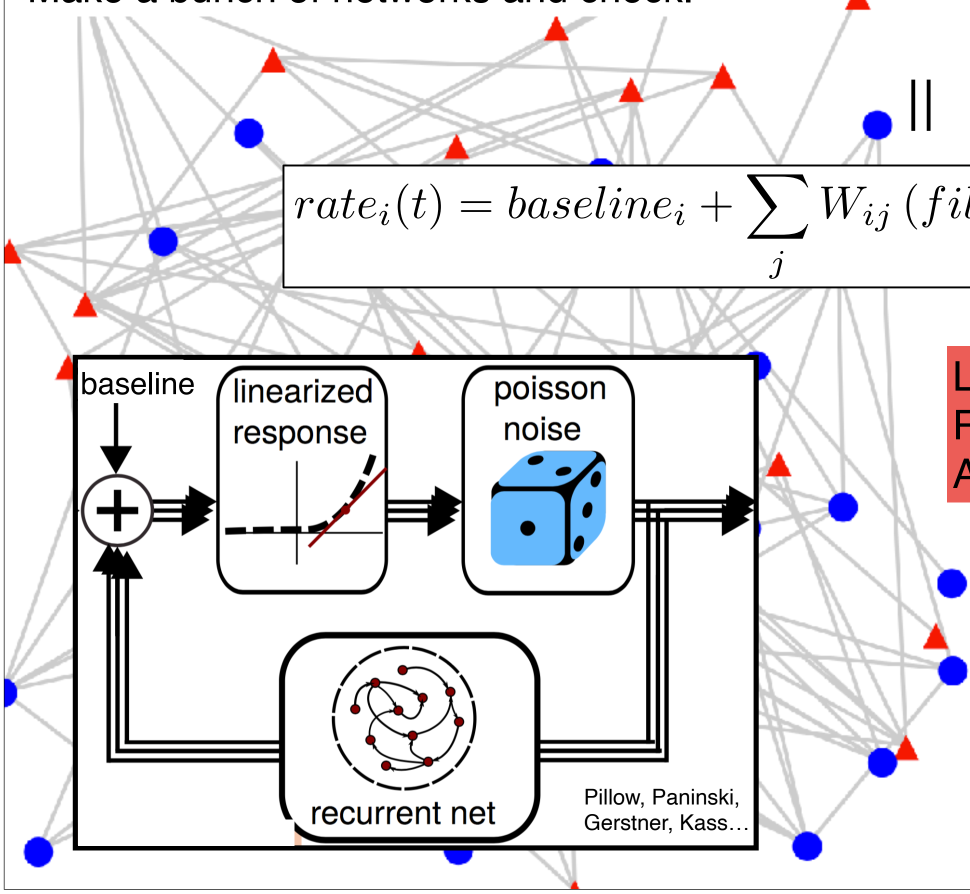
Pillow, Paninski et al, Nature '08

What about connectivity matters for network-wide dimension ?

Make a bunch of networks and check.

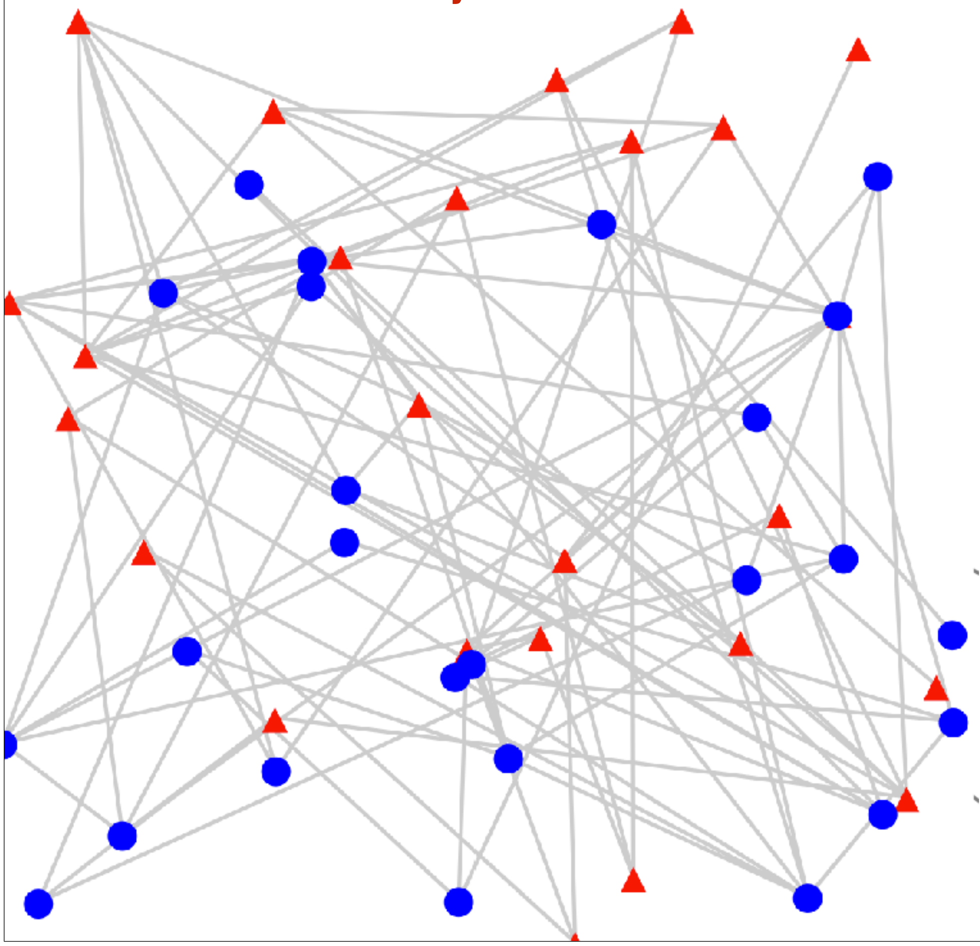
$||\ |\ |\ \ \ ||\ \ spikes_j(t)$

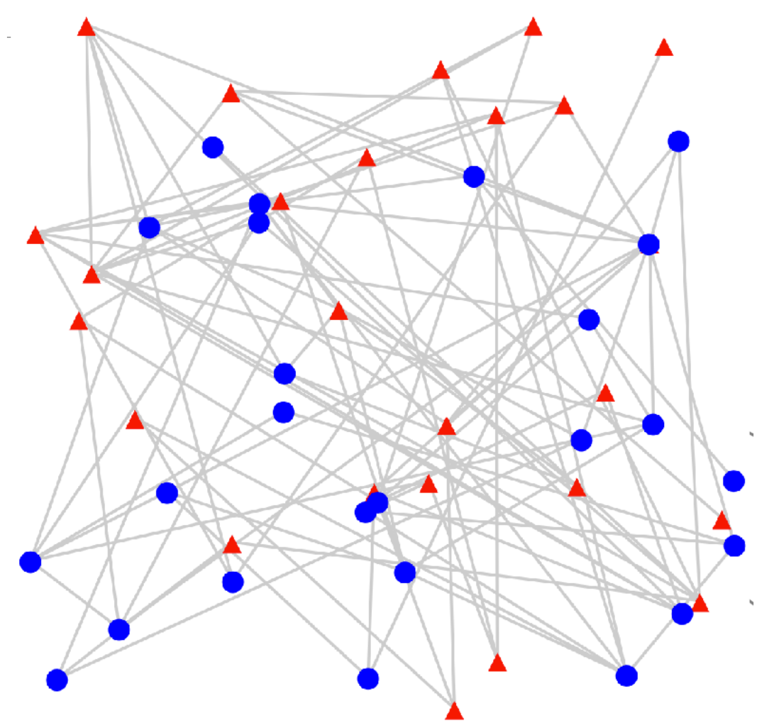
$$rate_i(t) = baseline_i + \sum_j W_{ij}\left(filter * spikes_j(t)\right)$$

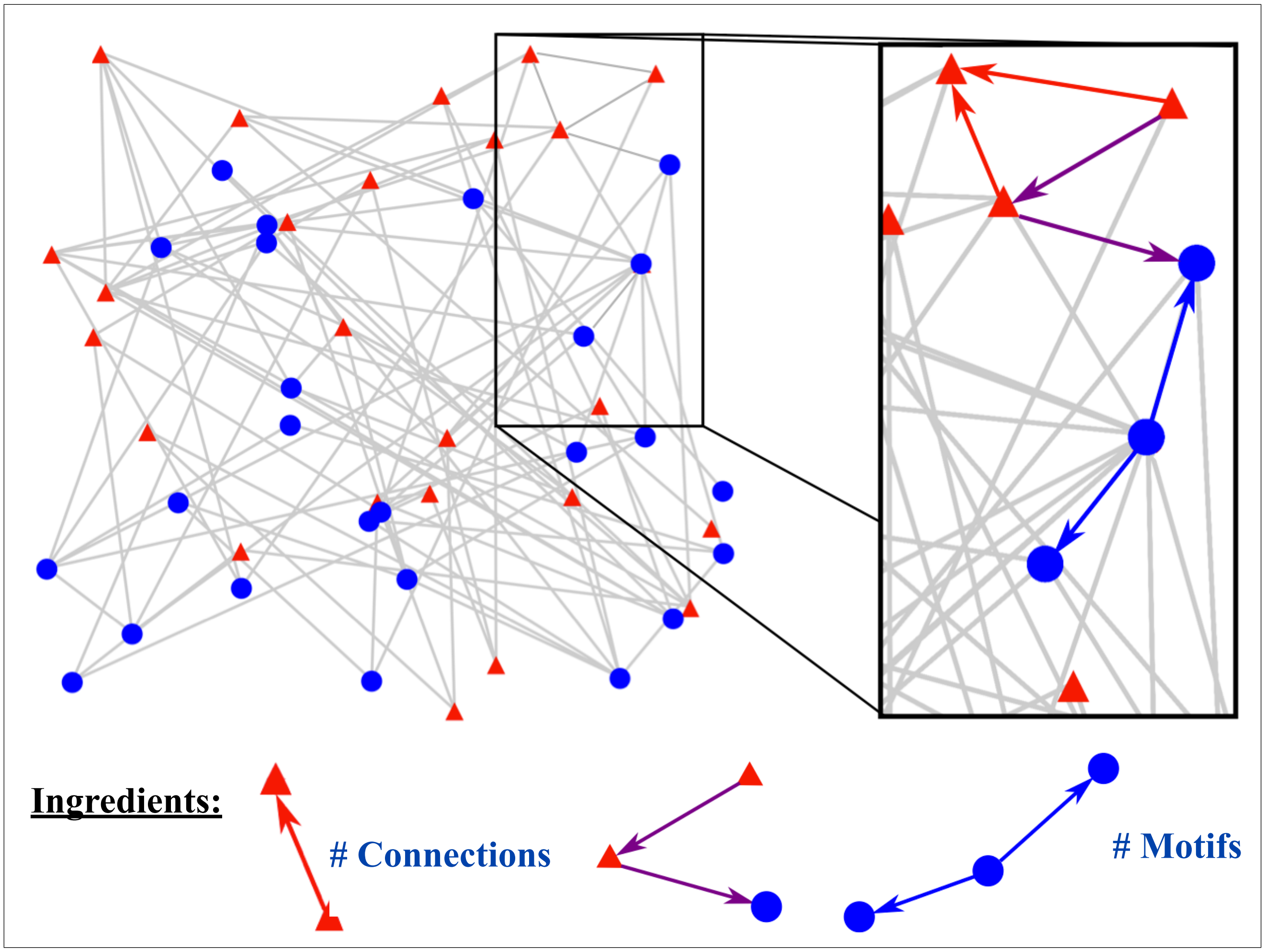


LINEAR RESPONSE APPROXIMATION

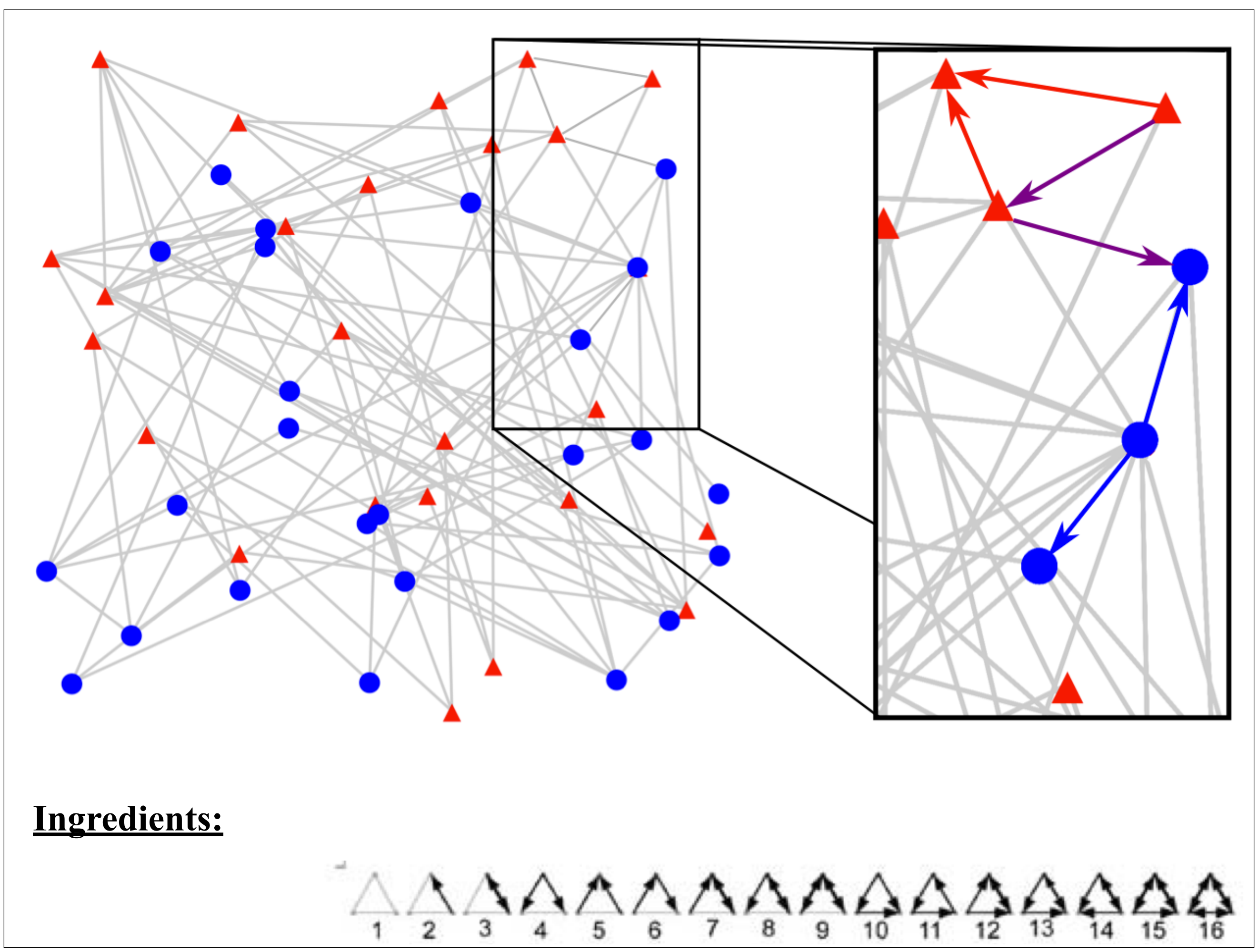
What about connectivity matters for network-wide dimension ?







Ingredients:
Connections
Motifs

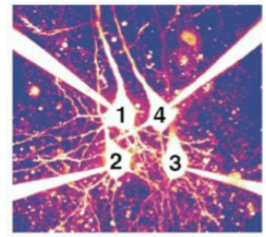


Ingredients:
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Are small motifs ENOUGH to predict network-wide dimension?

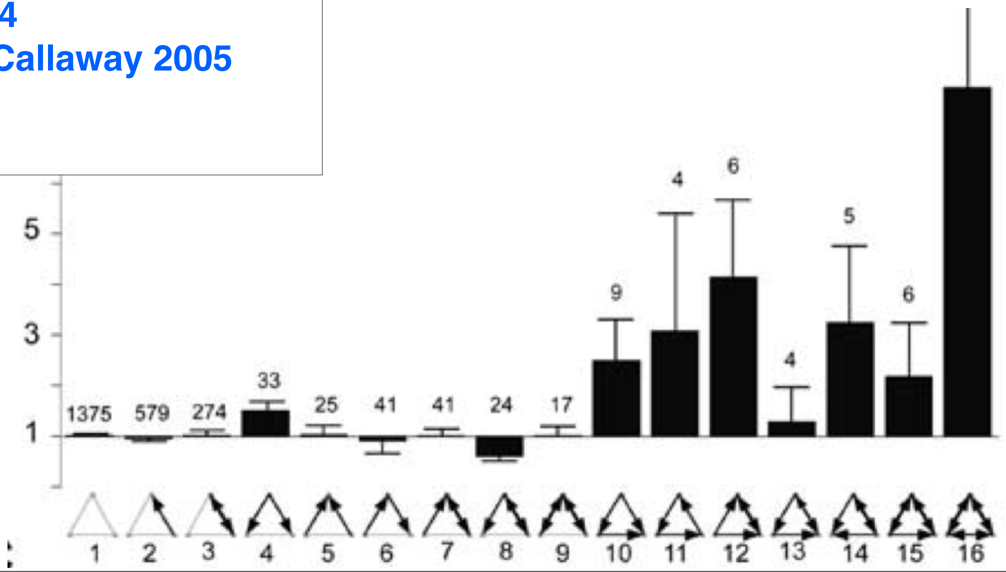
Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits

Sen Song[1], Per Jesper Sjöström[2,3], Markus Reigl[1], Sacha Nelson[2], Dmitri B. Chklovskii[1*]



Sporns et al 2004
Yoshimura and Callaway 2005
Perrin et al 2011
Bock et al 2011

Ingredients:



Are small motifs ENOUGH to predict network-wide dimension?

Yes!

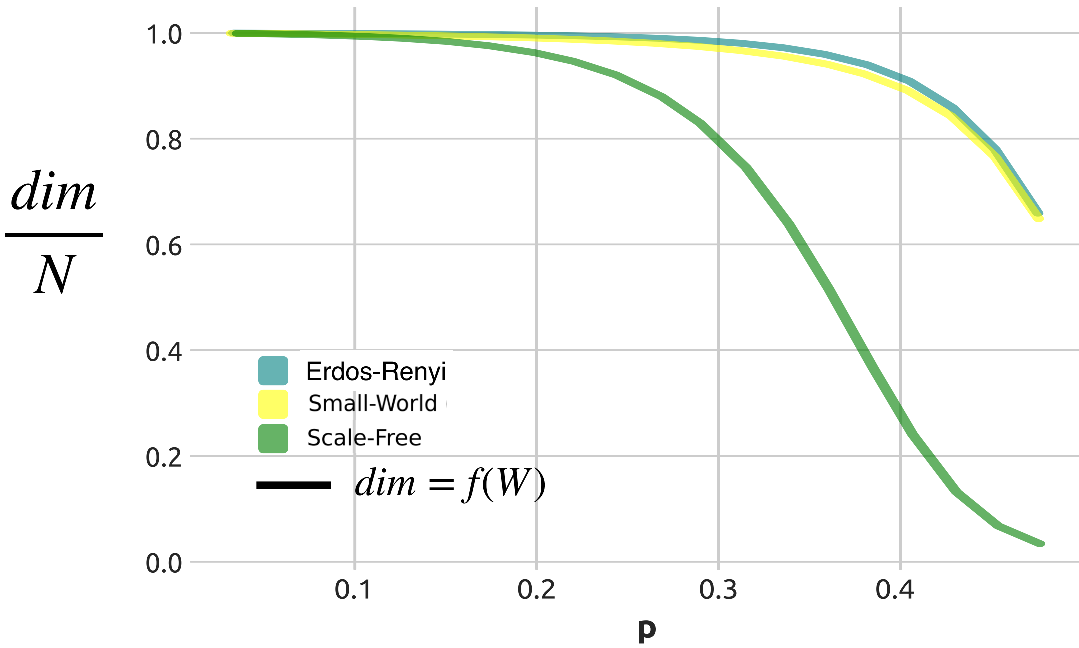
[cf. Yu Hu et al, 13, 14, 17
Ocker, Buice et al 18
Recanatesi et al 19
Dahmen, Recanatesi et al '21]

* at least when linearize

* most efficient for INTRINSIC, not stimulus-driven dynamics

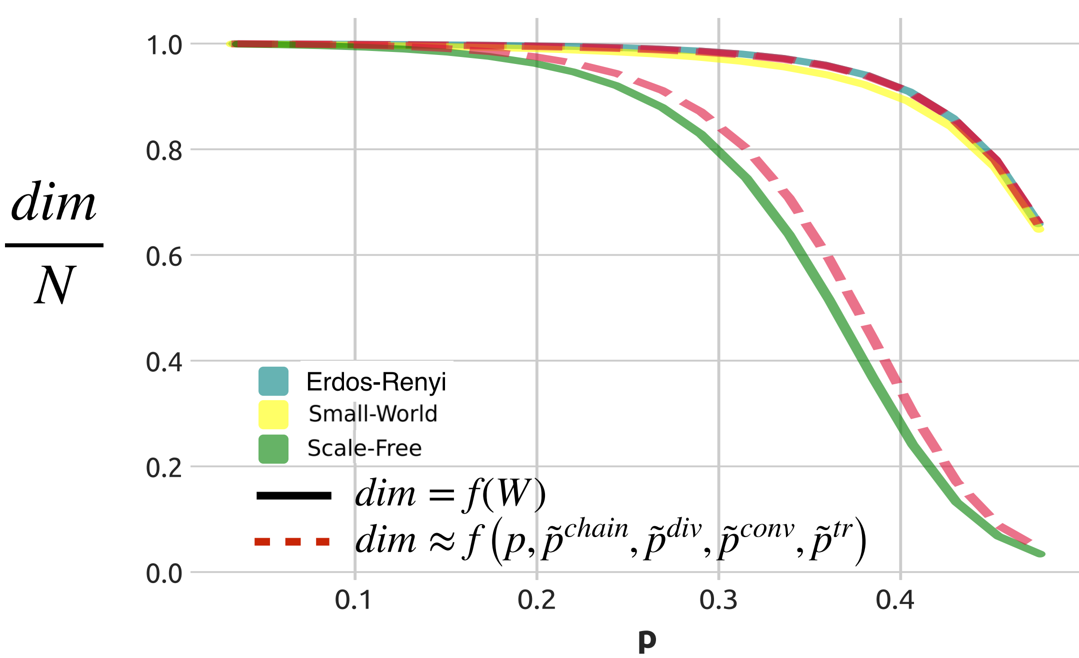
Are small motifs ENOUGH to predict network-wide dimension?

PURELY EXCITATORY



Are small motifs ENOUGH to predict network-wide dimension?

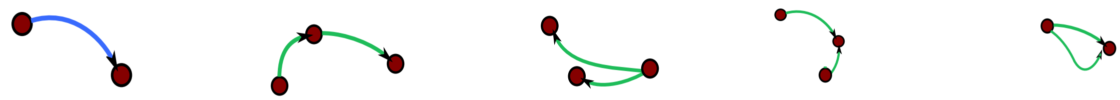
PURELY EXCITATORY



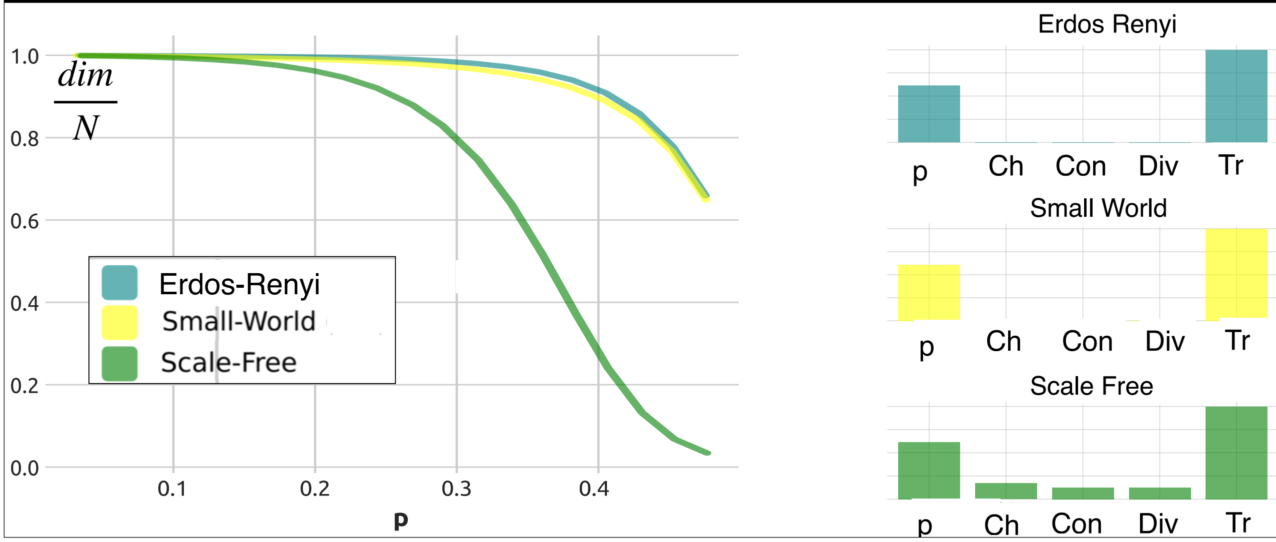
A recipe for dimension: Motif importance factors

$$dim \approx f\left(p, \tilde{p}^{chain}, \tilde{p}^{div}, \tilde{p}^{conv}, \tilde{p}^{tr}\right)$$

$$\approx \alpha_p p + \alpha_{chain}\, \tilde{p}^{chain} + \alpha_{div}\, \tilde{p}^{div} + \alpha_{con}\, \tilde{p}^{conv} + \alpha_{tr}\, \tilde{p}^{tr}$$



(prove!) < 0 < 0 < 0



Bottom up:

Connectivity → **dimension?**

A: When linearize: network motifs via their importance factors

Recanatesi et al, PLOS CB, '19; Recanatesi, Dahmen et al ArXiv '21

*** most efficient for INTRINSIC, not stimulus-driven dynamics**

A general approach, but others work in nonlinear regimes:

Mastroguiseppe and Ostojic '18; Williamson et al '16

low-rank connectivity

Circuit Models of Low-Dimensional Shared Variability in Cortical Networks

Chengcheng Huang [1, 2], Douglas A. Ruff [2, 3], Ryan Pyle [4], Robert Rosenbaum [4, 5], Marlene R. Cohen [2, 3], Brent Doiron [1, 2, 6]

spatial / temporal connectivity

Dimensionality and entropy of spontaneous and evoked neural rate dynamics

Rainer Engelken[1,2,3,4*], Fred Wolf[1,2,3,4]

PART 2

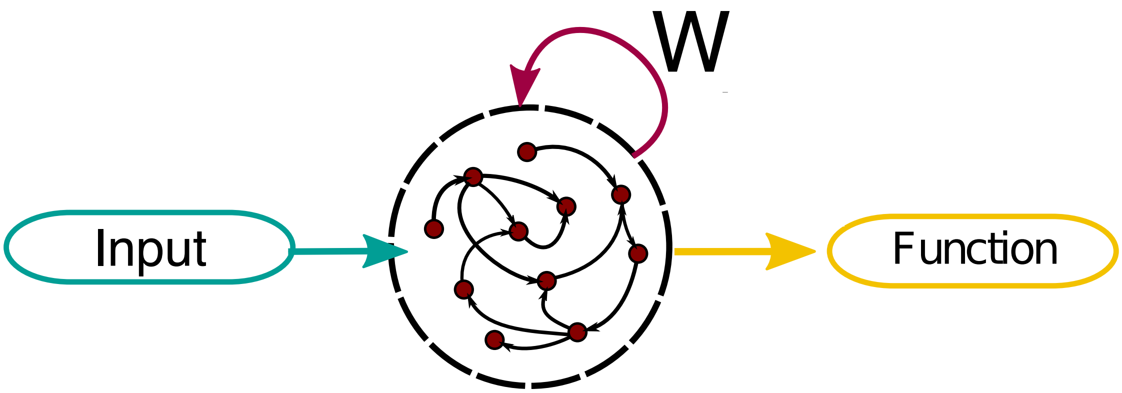
Bottom up:

Connectivity → dimension?

Top down:

Task → dimension?

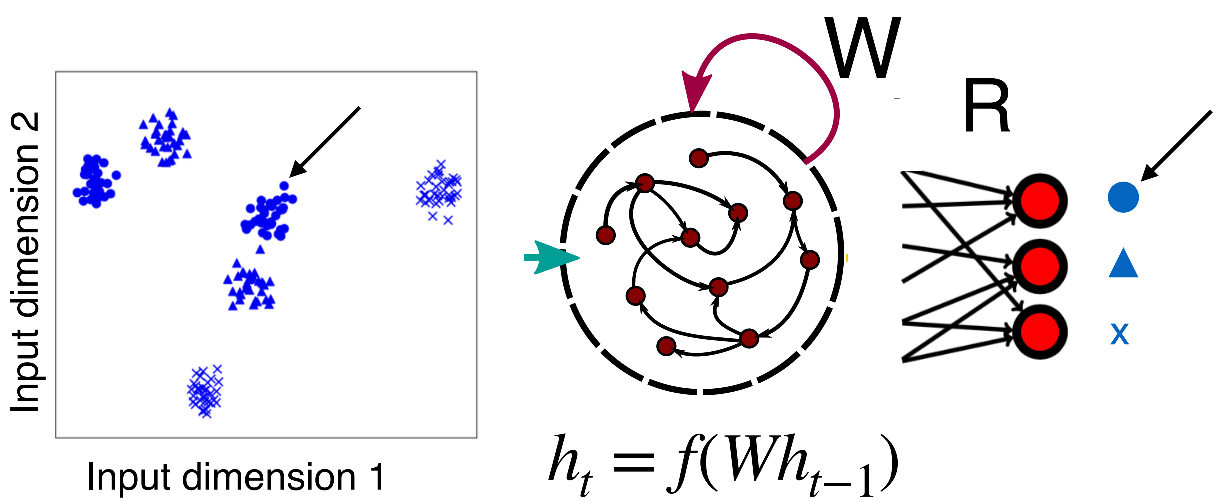
Train "artificial" neural networks to solve tasks



$$h_{i,t} = f(\sum_j W_{ij}\, h_{j,t-1})$$

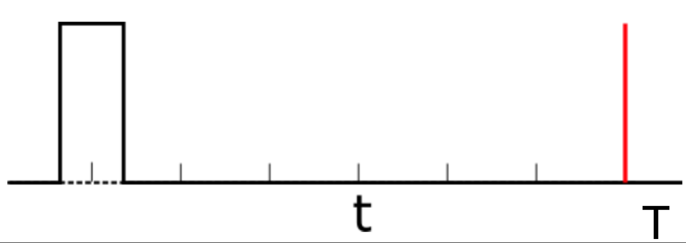
$$h_t = f(Wh_{t-1})$$

Train "artificial" neural networks to solve tasks

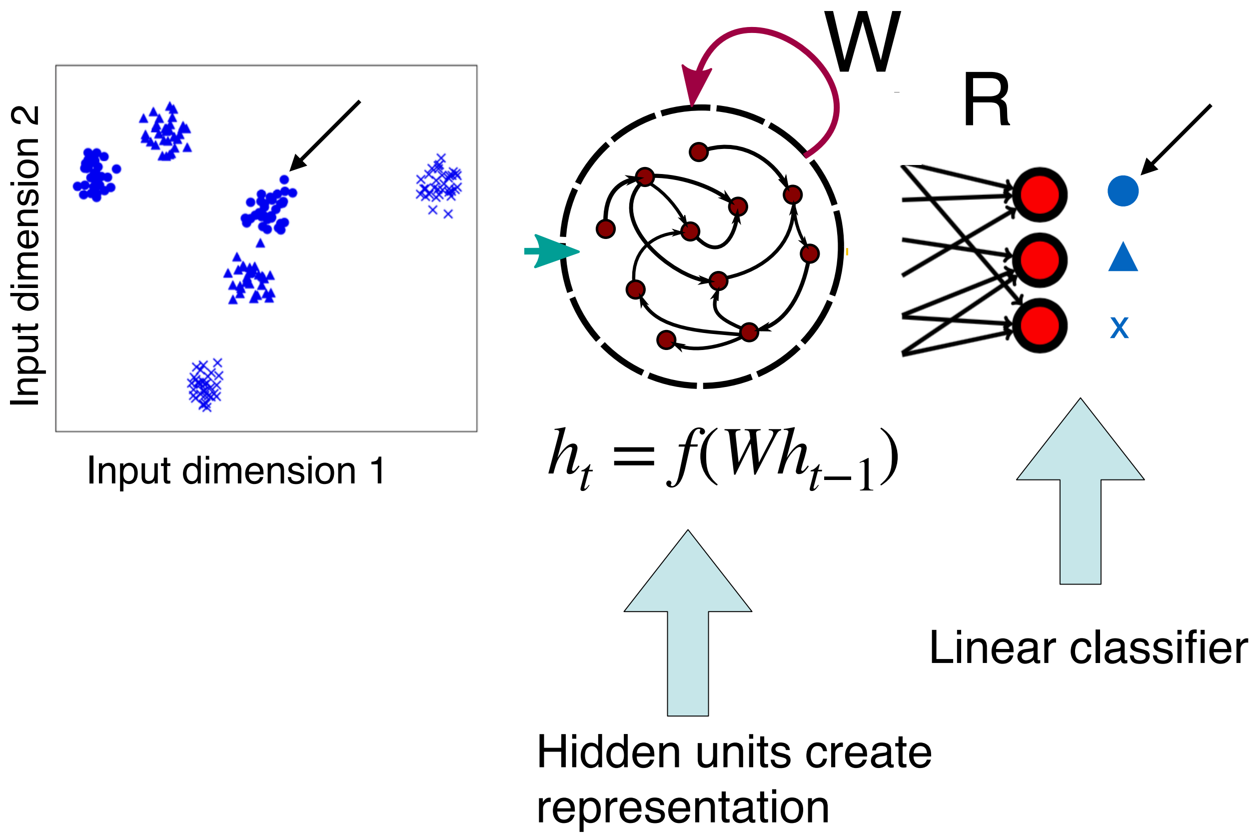


$$h_t = f(Wh_{t-1})$$

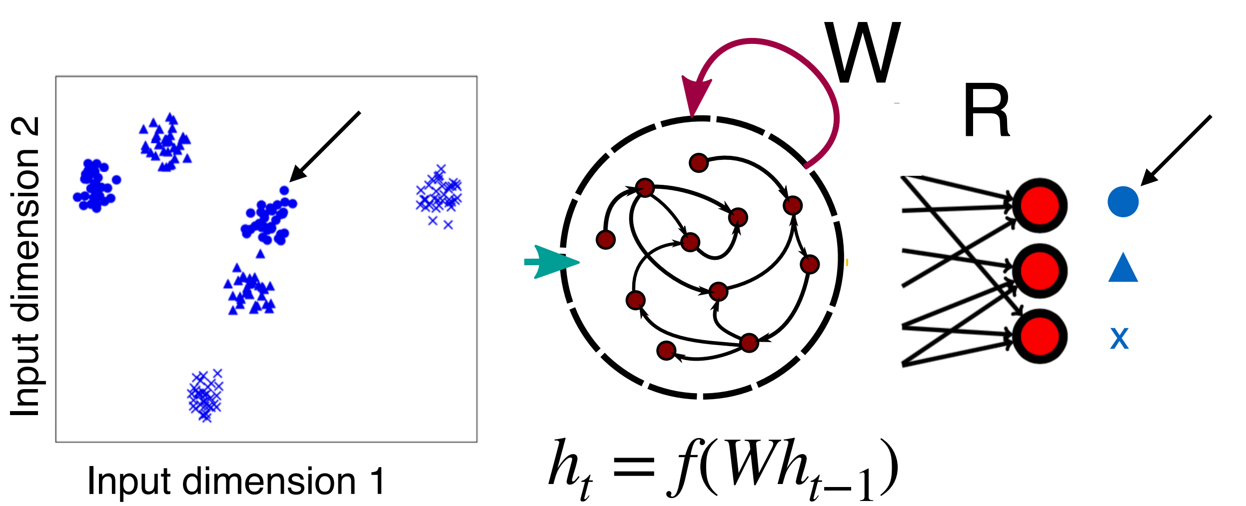
Delayed stimulus classification Task



Train “artificial” neural networks to solve tasks



Train “artificial” neural networks to solve tasks



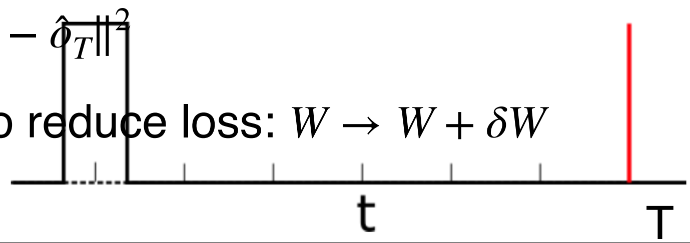
target: o_T

network: $\hat{o}_T = Rf(h_T)$

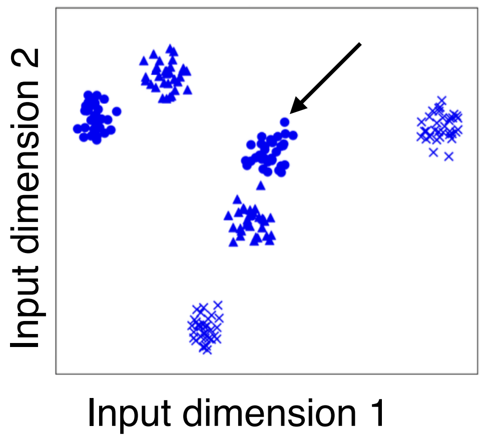
Loss: $L = \frac{1}{2}\|o_T - \hat{o}_T\|^2$

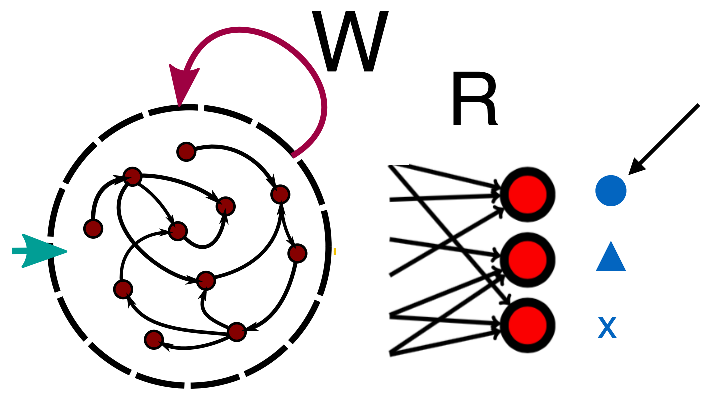
Train weights to reduce loss: $W \rightarrow W + \delta W$

$\delta W = -\eta \nabla_W L$



Train “artificial” neural networks to solve tasks





$$h_t = f(Wh_{t-1})$$

target: o_T

network: $\hat{o}_T = Rf(h_T)$

Loss: $L = \frac{1}{2}\|o_T - \hat{o}_T\|^2$

$\delta W = -\eta \nabla_W L$

$\delta R = -\eta \nabla_R L$

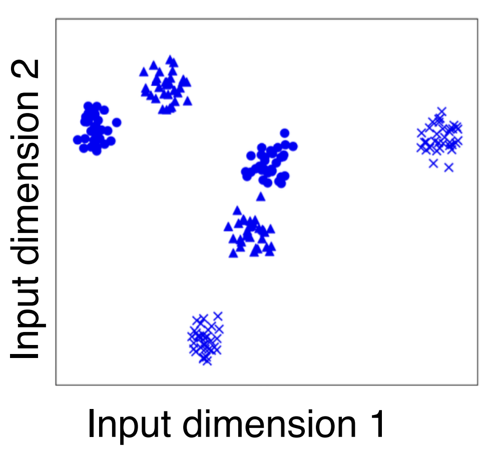
$$\nabla_W L = -\sum_k (o_{T,k} - \hat{o}_{T,k}) \nabla_W \hat{o}_{T,k}$$

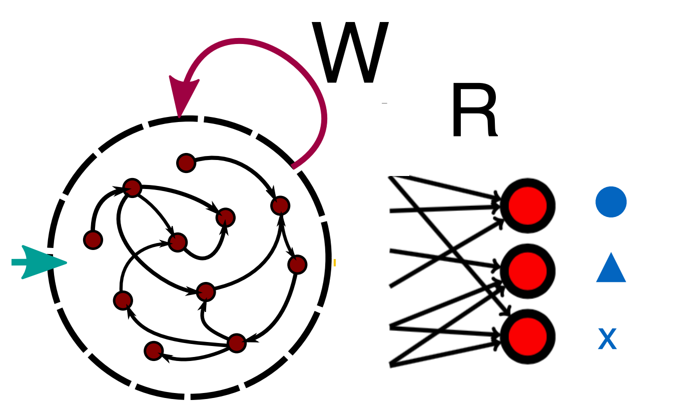
$$\nabla_W \hat{o}_{T,k} = \nabla_W \sum_l R_{kl} f(h_{T,l})$$

$$= \sum_l R_{kl} \nabla_W f\left(f(Wh_{T-1,l}\right)$$

$$\ldots$$

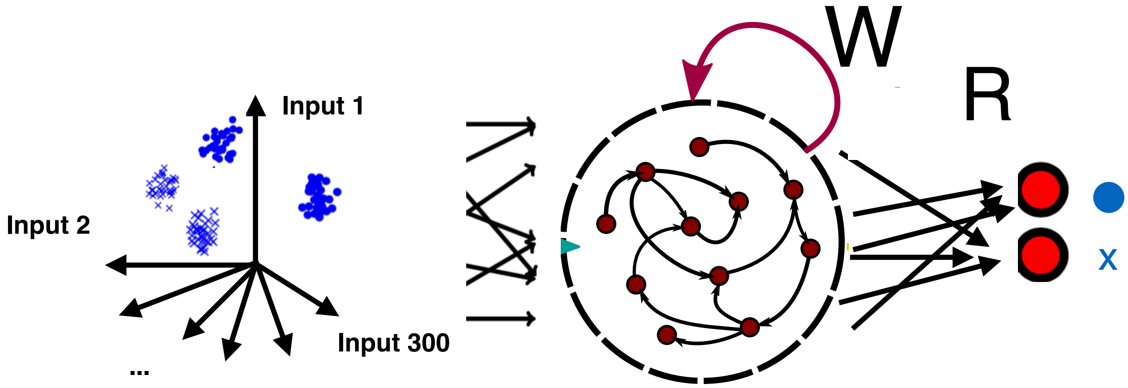
Train “artificial” neural networks to solve tasks





What’s dimension that network learns?

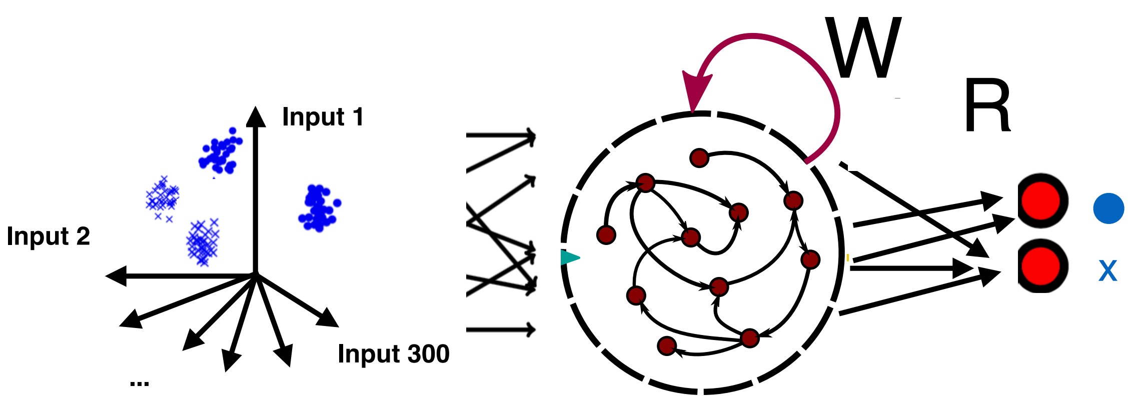
Train “artificial” neural networks to solve tasks



Easy classification:
Input has high dim.

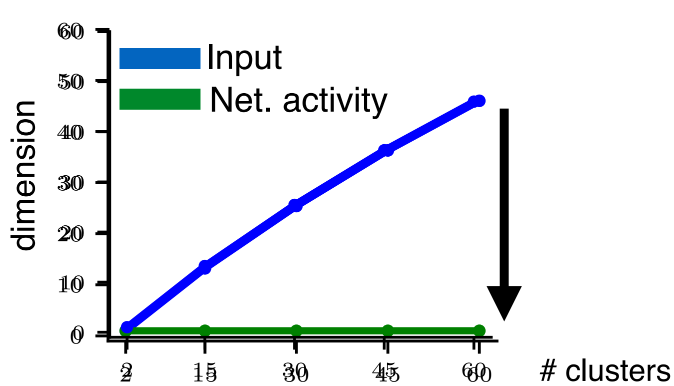
What’s dimension that
network learns?

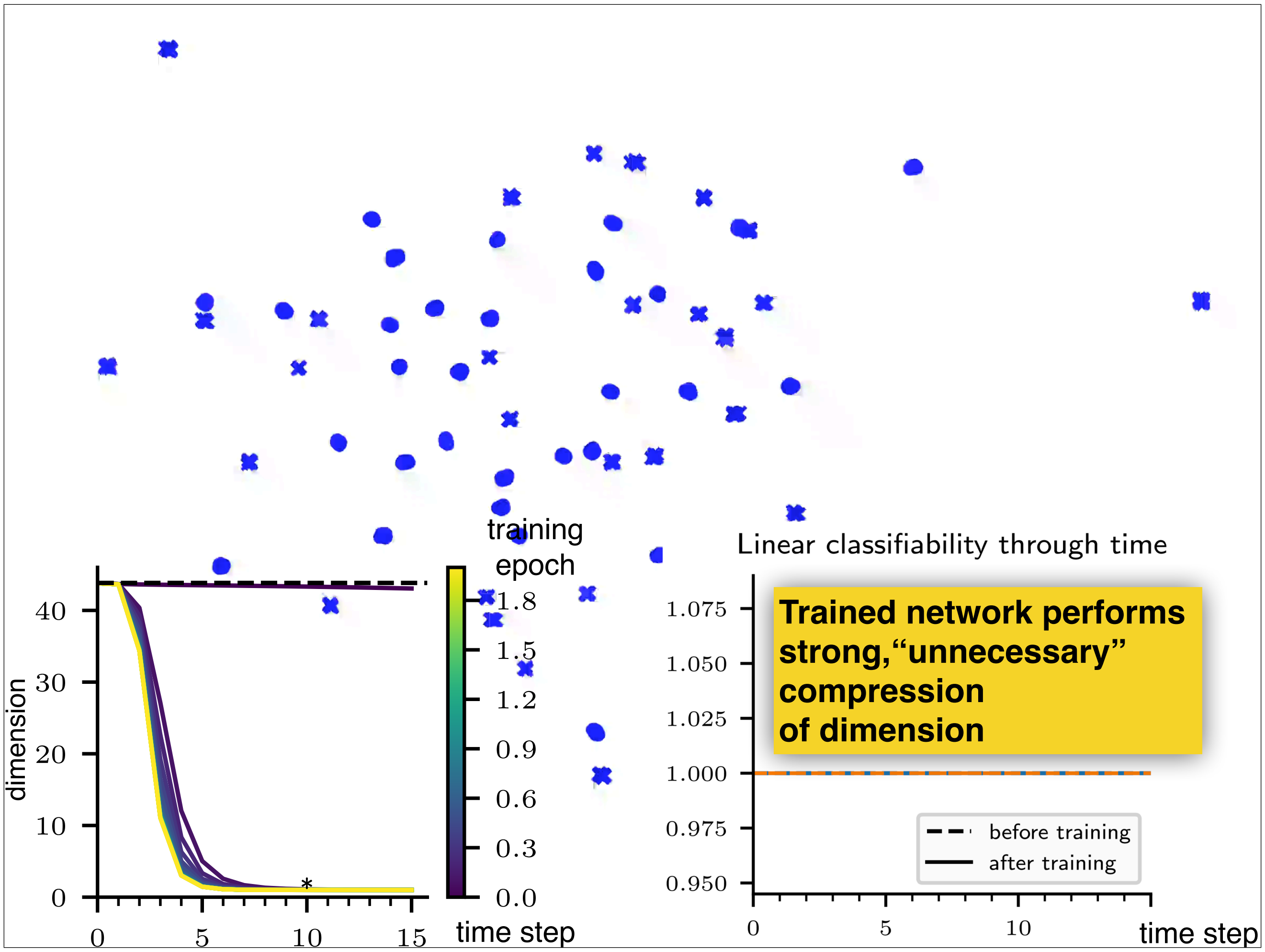
Train “artificial” neural networks to solve tasks

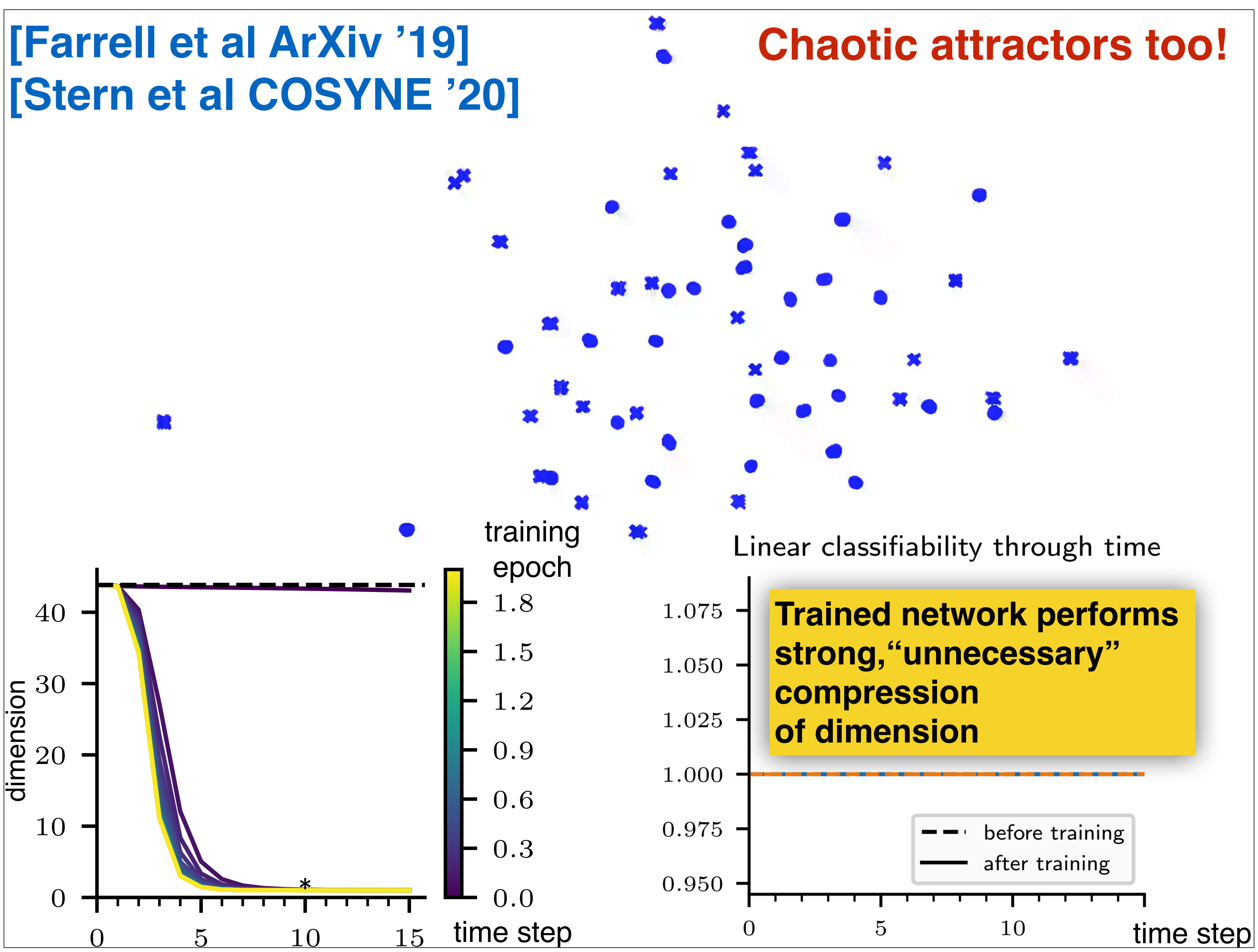


Easy classification:
Input has high dim.

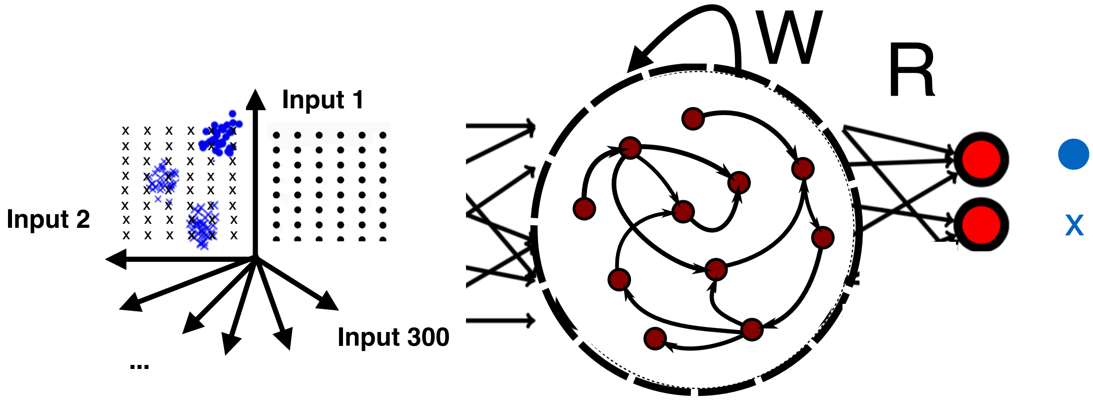
What’s dimension that
network learns?




training
epoch
dimension
time step
Linear classifiability through time
Trained network performs strong,"unnecessary" compression of dimension
before training
after training
time step


[Farrell et al ArXiv '19]
[Stern et al COSYNE '20]
Chaotic attractors too!
training
epoch
dimension
time step
Linear classifiability through time
Trained network performs strong,"unnecessary" compression of dimension
before training
after training
time step

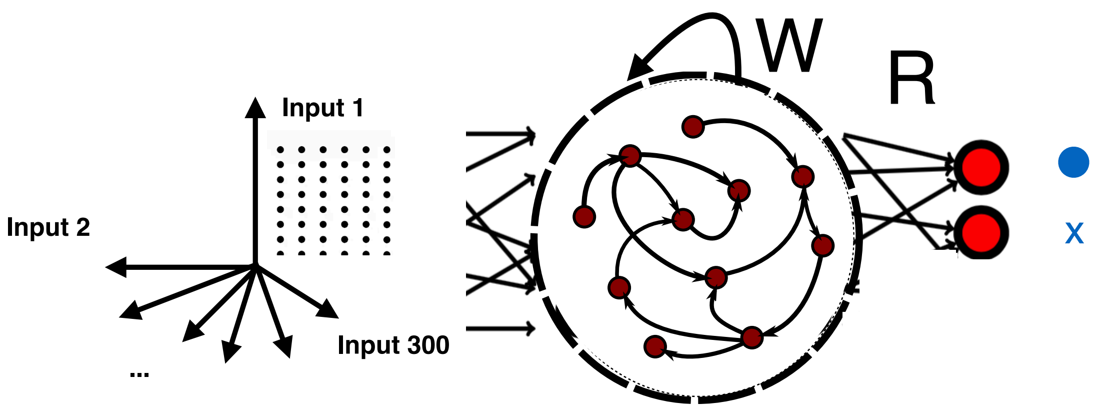
Collapse of dimension: mechanism?



$$h_t = f(Wh_{t-1})$$

Toy model:

Collapse of dimension: mechanism?

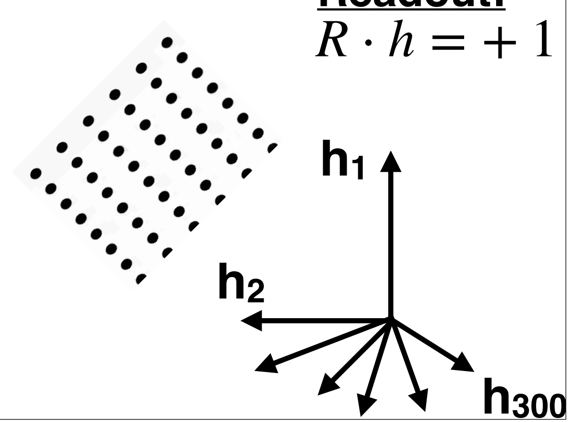


$$h_t = f(Wh_{t-1})$$

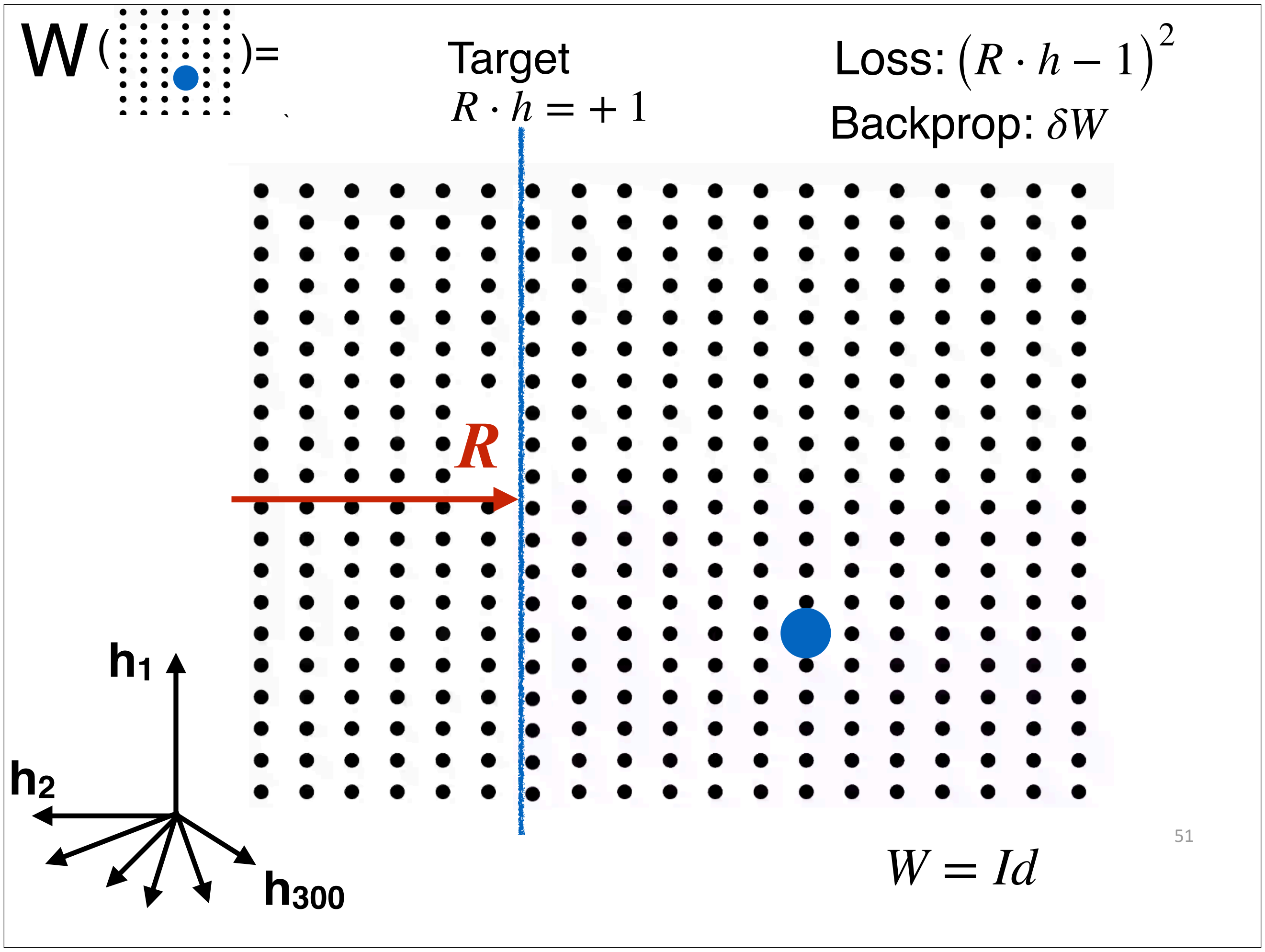
Toy model:

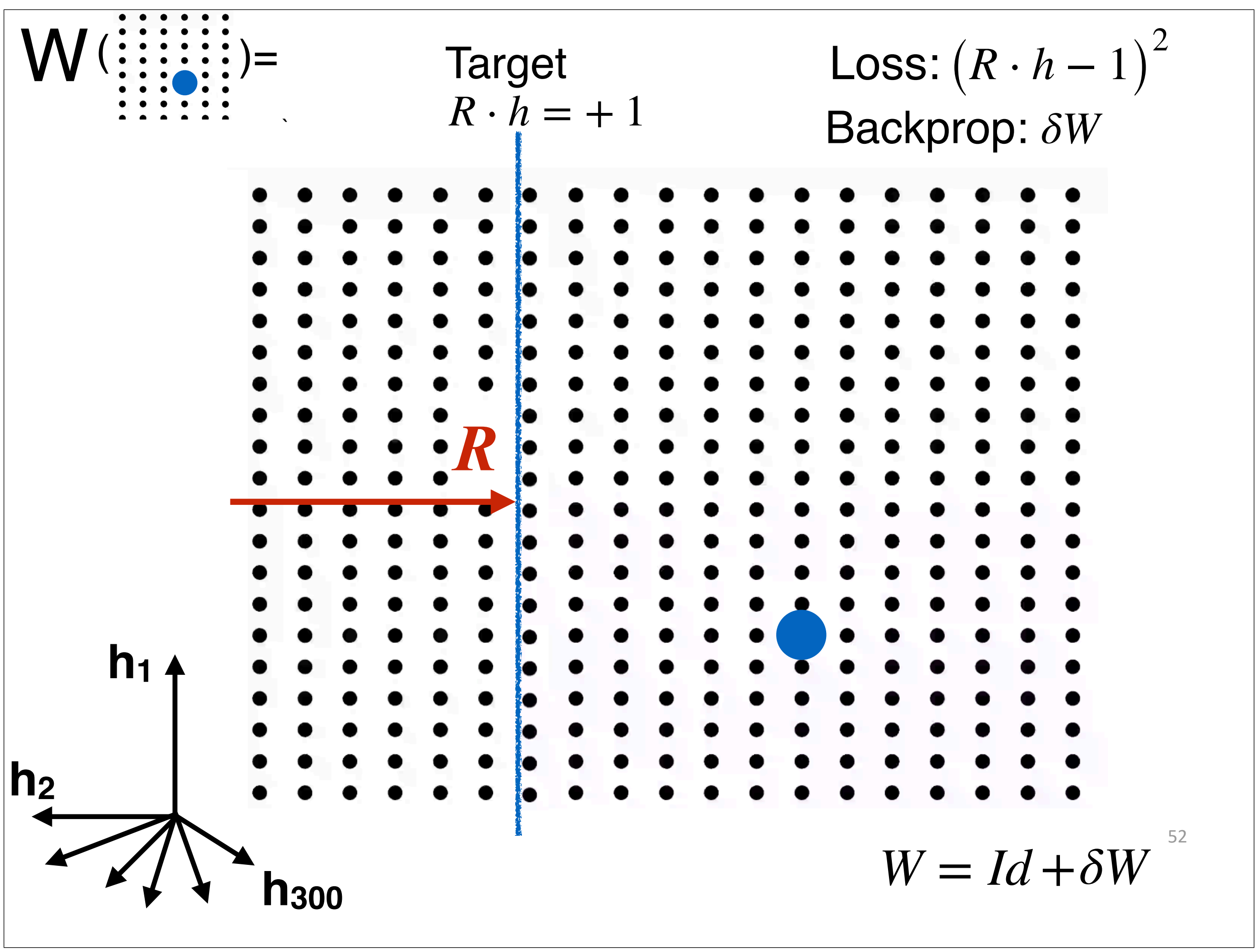
Readout:
$R \cdot h = +1$

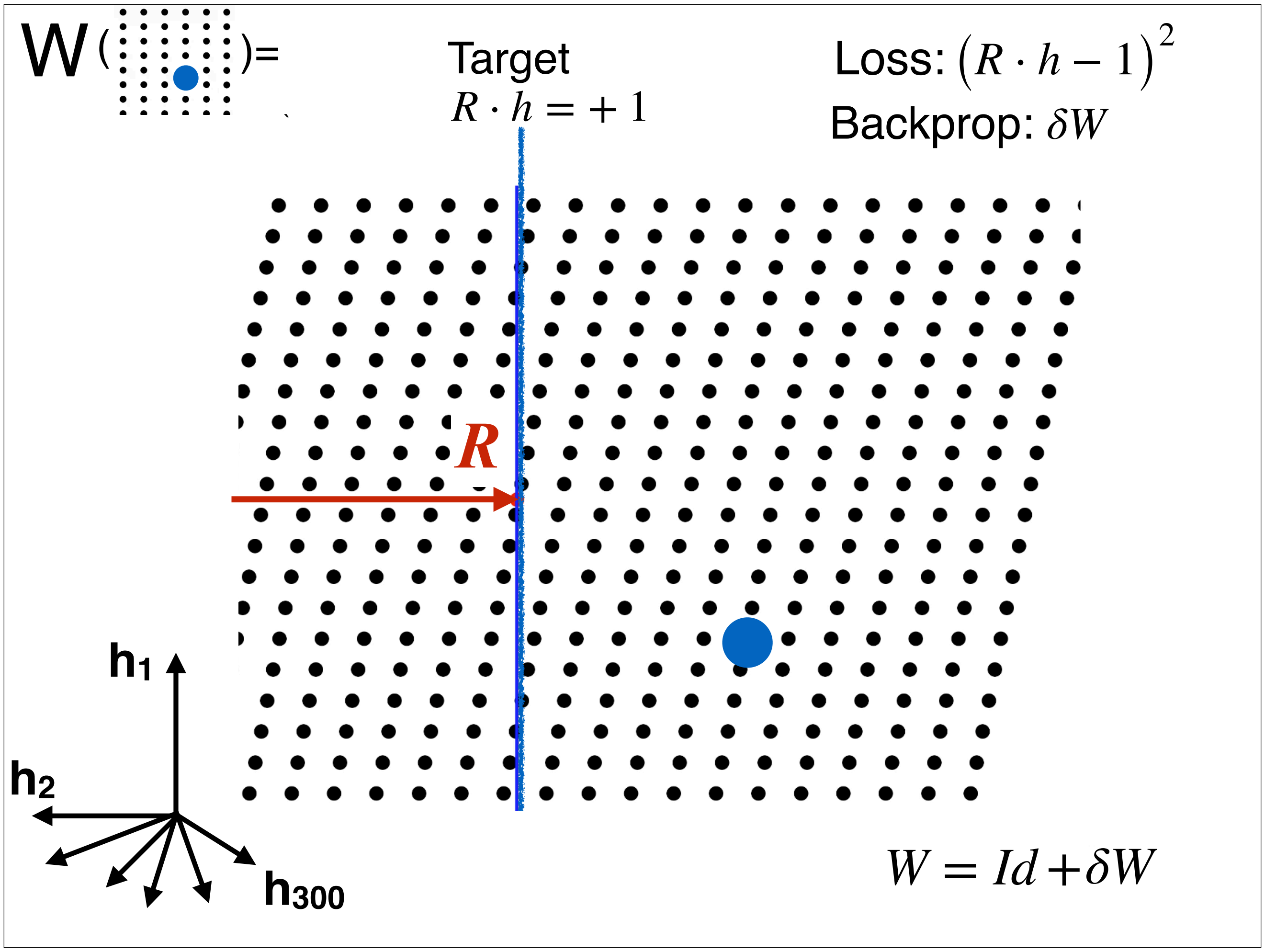
W(   )=

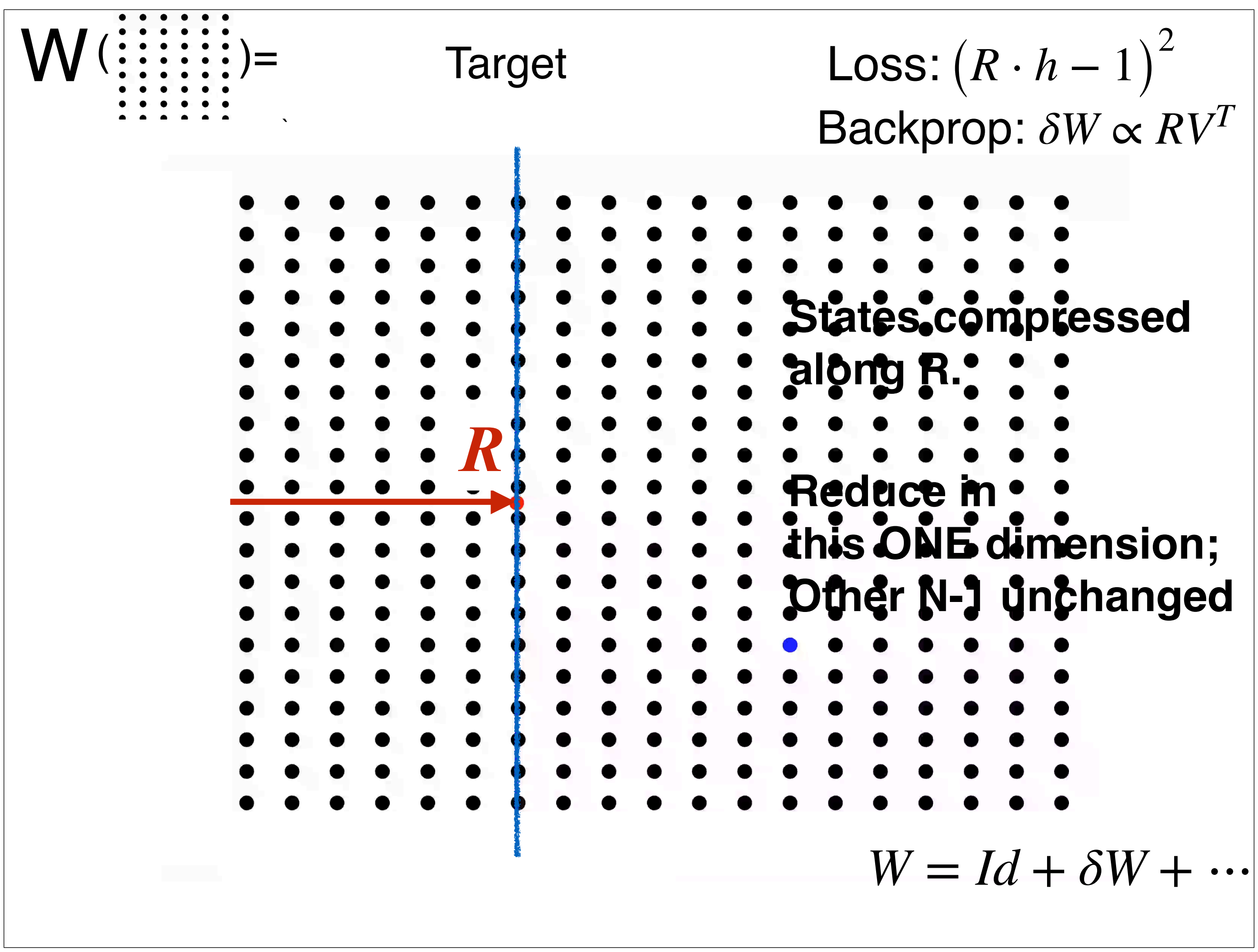


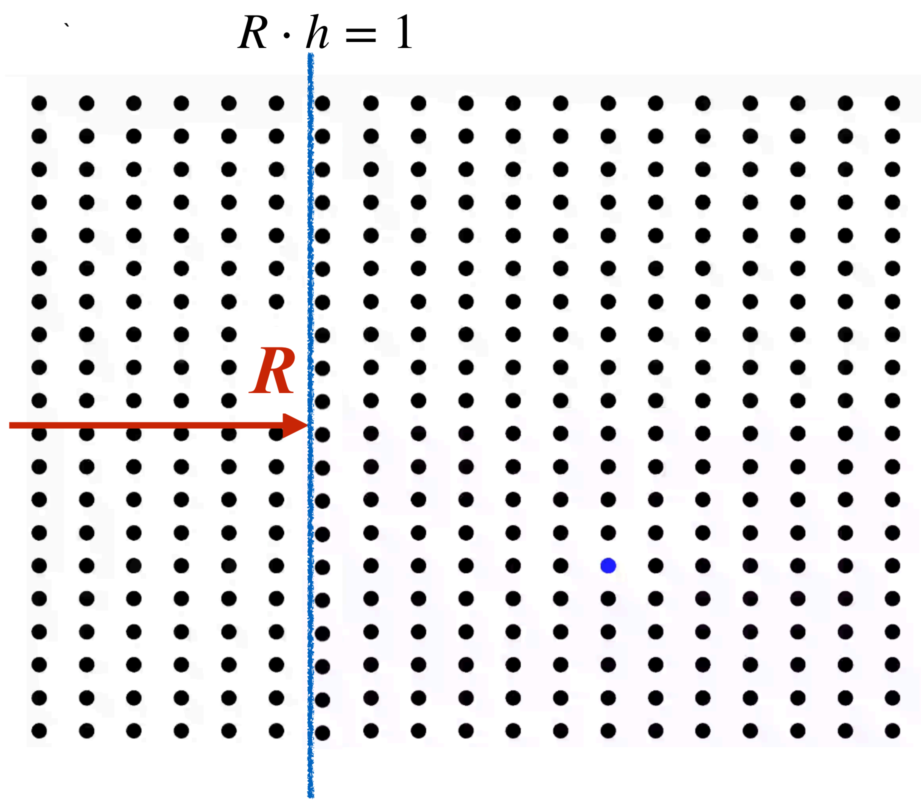
Linearized
dynamics

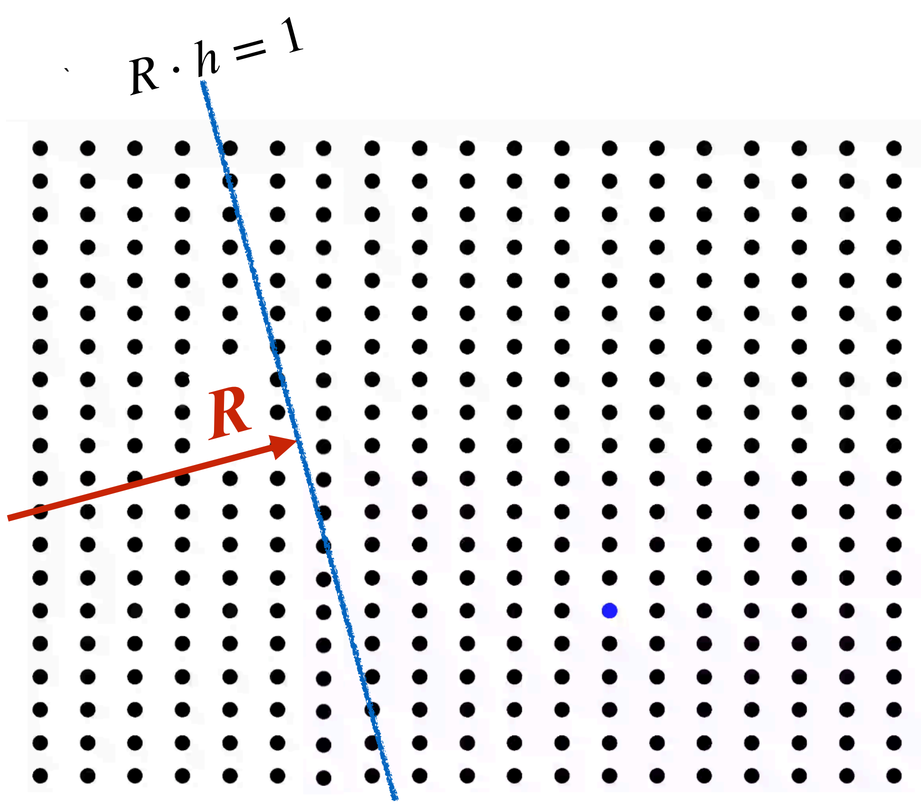

W( )=
Target
$R \cdot h = +1$
Loss: $(R \cdot h - 1)^2$
Backprop: δW
R
h1
h2
h300
$W = Id$
51

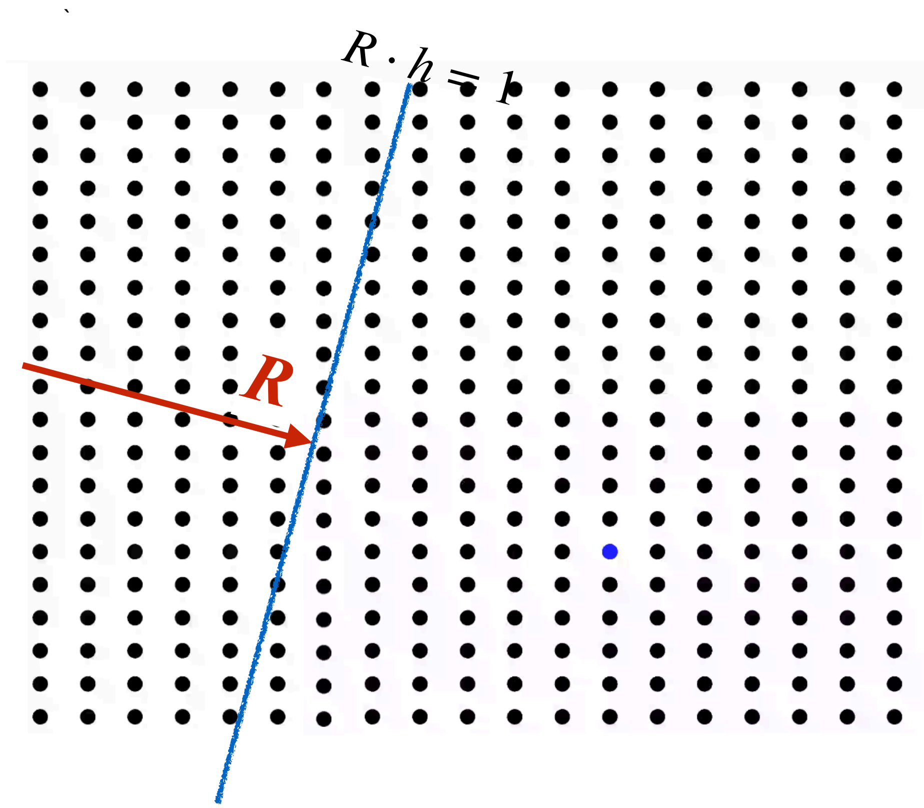

W( )=
Target
$R \cdot h = +1$
Loss: $(R \cdot h - 1)^2$
Backprop: δW
R
h1
h2
h300
$W = Id + \delta W$
52

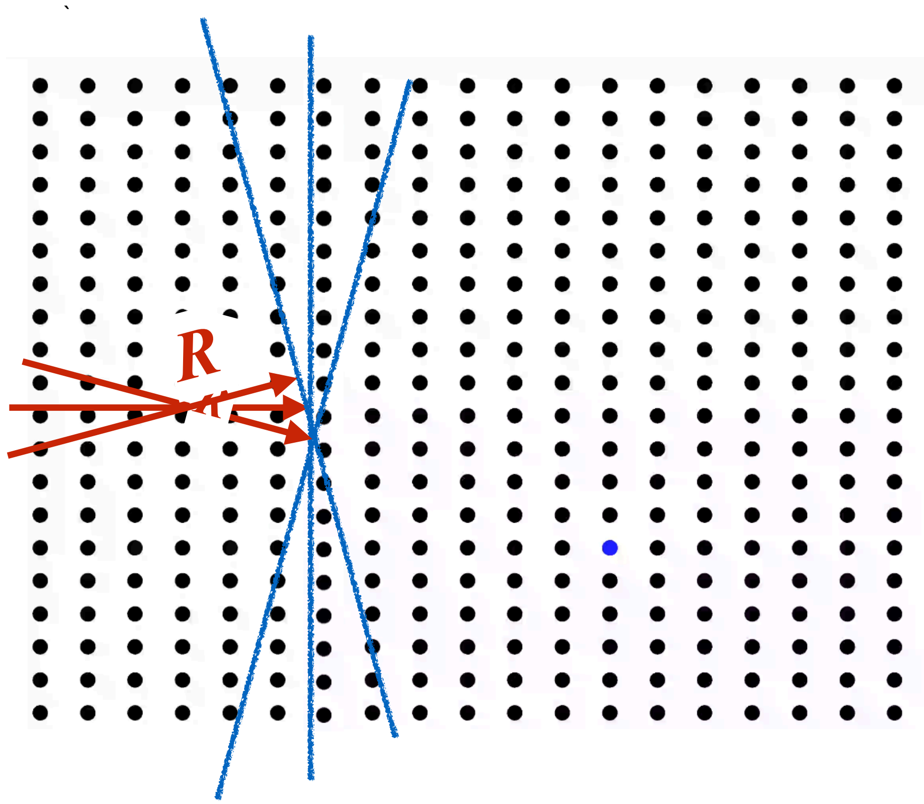

W( )=
Target
$R \cdot h = +1$
Loss: $(R \cdot h - 1)^2$
Backprop: δW
R
h1
h2
h300
$W = Id + \delta W$

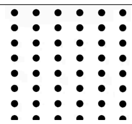

W( )=
Target
Loss: $(R \cdot h - 1)^2$
Backprop: $\delta W \propto R V^T$
States compressed along R.
Reduce in this ONE dimension; Other N-1 unchanged
R
$W = Id + \delta W + \cdots$


$R \cdot h = 1$
R


$R \cdot h = 1$
R

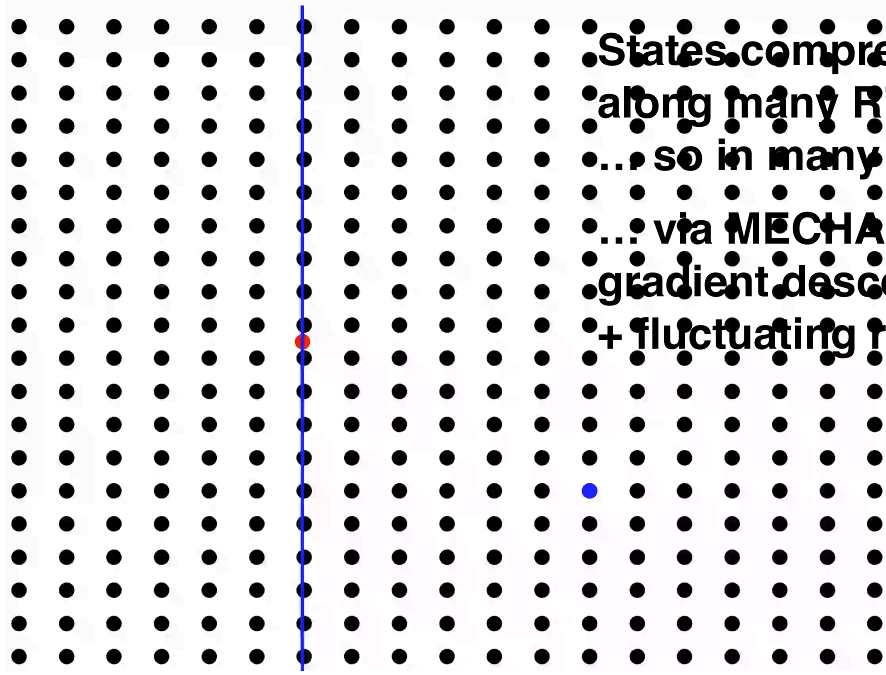

$R \cdot h = 1$
R


R

W()=

Loss: $(R \cdot h - 1)^2$

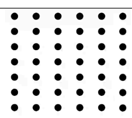
Backprop: $\delta W \propto RV^T$



States compressed along many R's,
… so in many dim's.

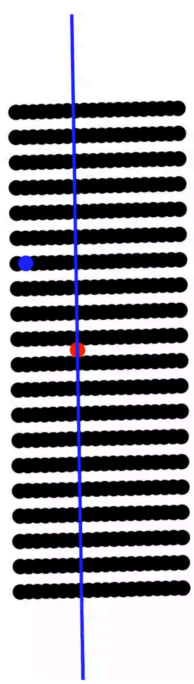
… via MECHANISM:
gradient descent on W
+ fluctuating readout R

$$W = Id + \delta W + \cdots$$

W()=

Loss: $(R \cdot h - 1)^2$

Backprop: $\delta W \propto RV^T$



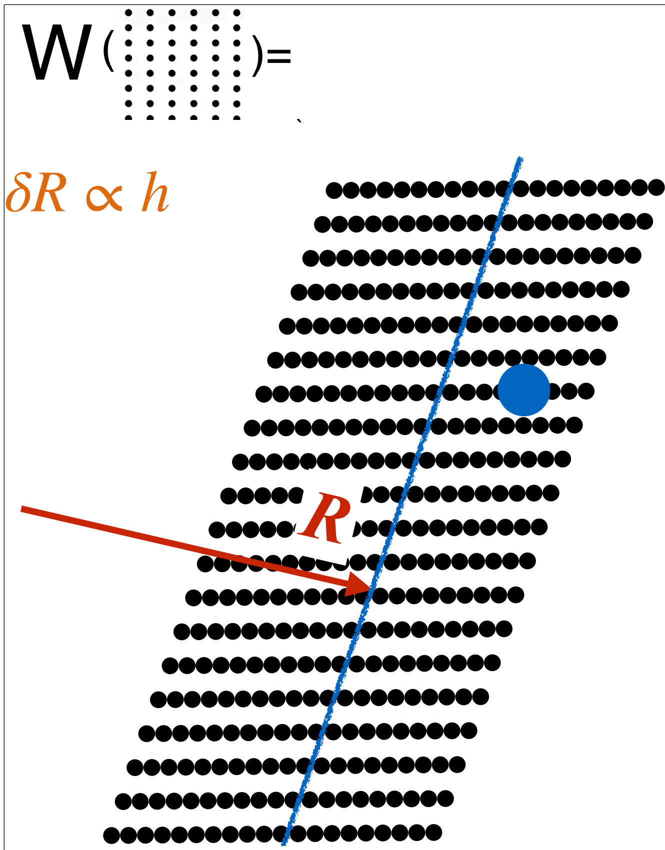
States compressed along many R's,
… so in many dim's.

… via MECHANISM:
gradient descent on W
+ fluctuating readout R

… not on every step,
but on <u>average</u>

<u>Prove:</u>
w/ $R \sim \bar{R}$ + noise :
Recanatesi, Advani, Farrell, et al
+ Farrell et al, ArXiv 2019

w/ $R \sim \bar{R} + \delta R$ **from SGD: ?**



$\delta R \propto h$

Loss: $(R \cdot h - 1)^2$

Backprop: $\delta W \propto RV^T$

States compressed along many R's, … so in many dim's.

… via MECHANISM gradient descent on W + fluctuating readout R

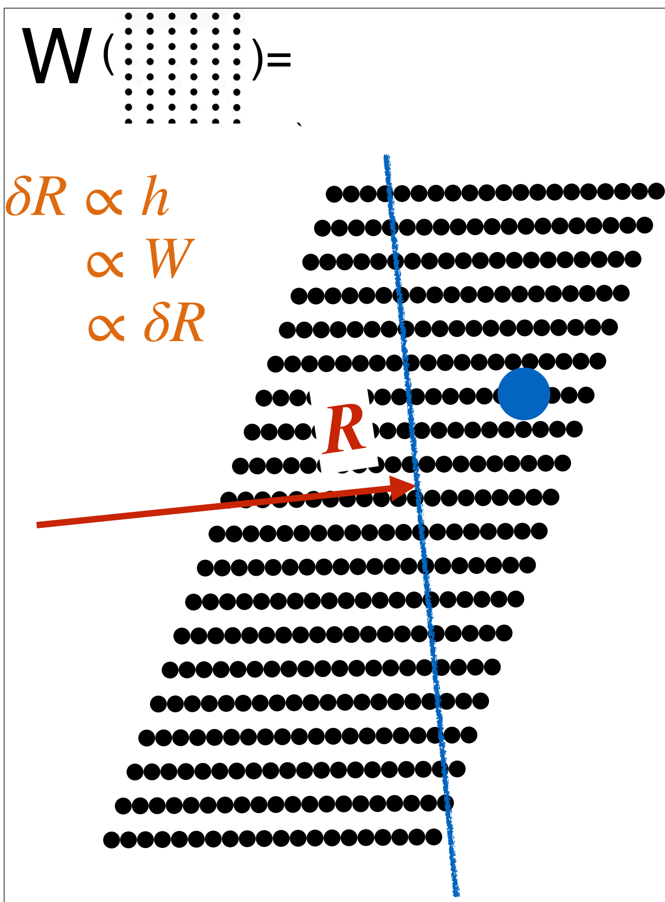
… not on every step, but on <u>average</u>

<u>Prove:</u>

w/ $R \sim \bar{R}$ + noise :

Recanatesi, Advani, Farrell, et al
+ Farrell et al, ArXiv 2019

w/ $R \sim \bar{R} + \delta R$ **from SGD: ?**



$\delta R \propto h$
$\propto W$
$\propto \delta R$

Loss: $(R \cdot h - 1)^2$

Backprop: $\delta W \propto RV^T$

States compressed along many R's, … so in many dim's.

… via MECHANISM gradient descent on W + fluctuating readout R

… not on every step, but on <u>average</u>

<u>Prove:</u>

w/ $R \sim \bar{R}$ + noise :

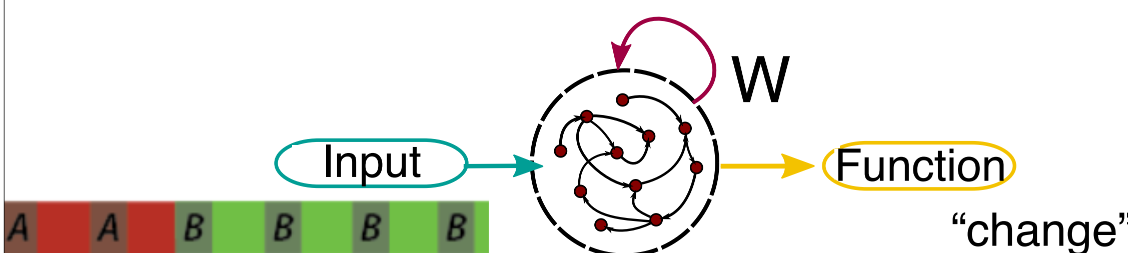
Recanatesi, Advani, Farrell, et al
+ Farrell et al, ArXiv 2019

w/ $R \sim \bar{R} + \delta R$ **from SGD: ?**

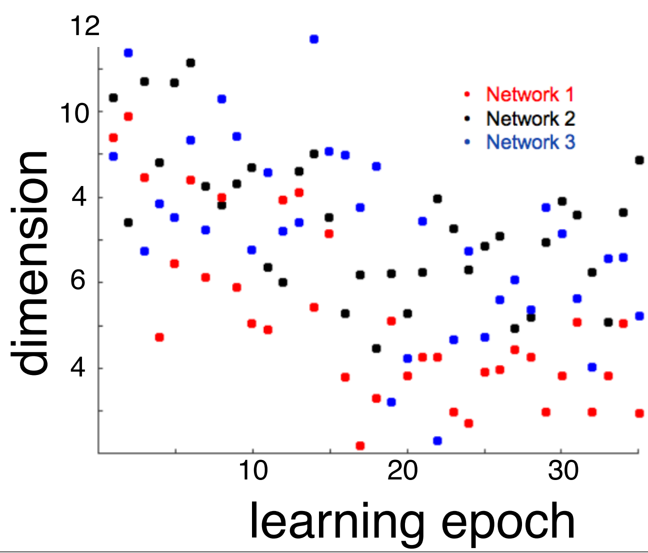
Change detection task

Merav Stern, Matt Valley, Shawn Olsen, Sahar Manavi, Doug Ollerenshaw, Kathleen Champion, …

Change detection task

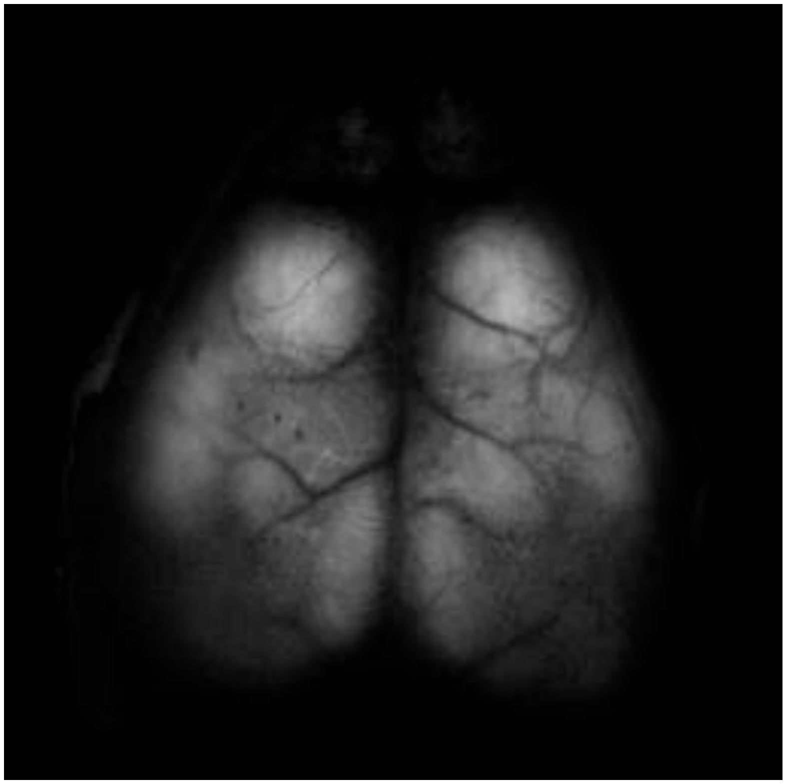


Dimension of trained net activity

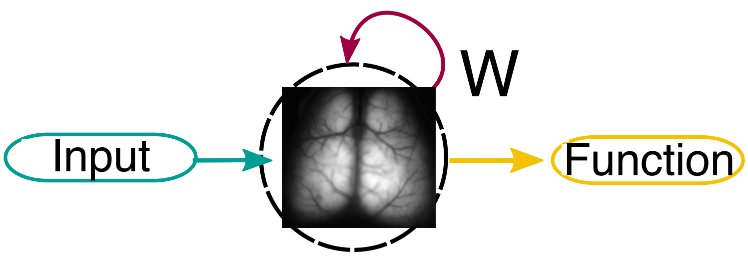


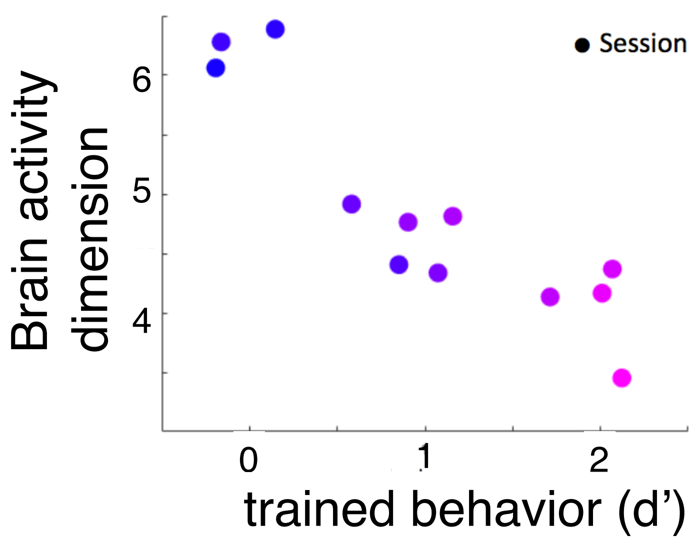
Stern et al
COSYNE
2020

Change detection task



Change detection task





Stern et al COSYNE 2020

To do: is dim. collapsing in "task-irrelevant directions?"

Top Down: Task → dimension?

"Easy" classification task: Dimension compressed

... in "task-irrelevant" directions

... creating "useful" compact representations?

ex. of Tishby's Info Bottleneck (2018)?

... signature of (some) learning rules in neural data?

Different than, e.g. Todorov '05, Druckmann + Chklovskii '12

cf. Cueva, Fusi et al, BioRXiv '18

Top Down: Task → dimension?

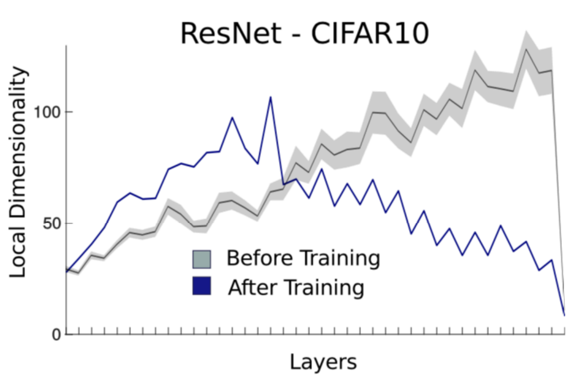
"Easy" classification task: Dimension compressed

... in "task-irrelevant" directions

... creating "useful" compact representations?

... signature of (some) learning rules in neural data?

... same trend in deep FFW nets



Papyan et al
PNAS '20

Cohen et al,
Separability and geometry …
PRX 2019/20

Recanatesi, Farrell et al,
Dimension compression and expansion deep nets
ArXiv 2019

Top Down: Task → dimension?

"Easy" classification task: Dimension compressed

… in "task-irrelevant" directions

… creating "useful" compact representations?

… signature of (some) learning rules in neural data?

… same trend in deep FFW nets

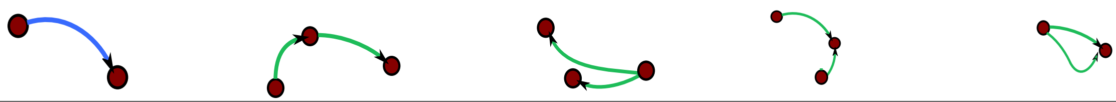
… and for predictive learning [Recanatesi et al '19]

Bottom-up: Connectivity → dimension?

Via network motifs

Bridge? Goldilocks dimension via local plasticity

$$dim \approx \alpha_p p + \alpha_{chain}\, \tilde{p}^{chain} + \alpha_{div}\, \tilde{p}^{div} + \alpha_{con}\, \tilde{p}^{conv} + \alpha_{tr}\, \tilde{p}^{tr}$$



Thank you

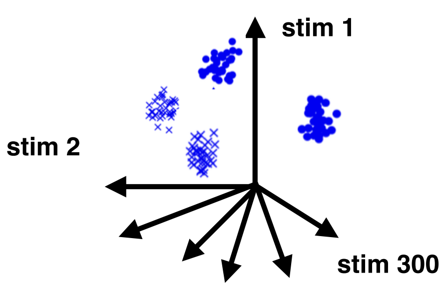
Stefano Recanatesi (University of Washington)
Matt Farrell (University of Washington)
Merav Stern (University of Washington)
Michael Buice (Allen Institute)
Gabe Ocker (Allen Institute)
Yu Hu (HKUST)
Madhu Advani (Harvard U.)
Guillaume Lajoie (U. Montreal)
Sophie Deneve (ENS)
Mattia Rigotti (IBM)

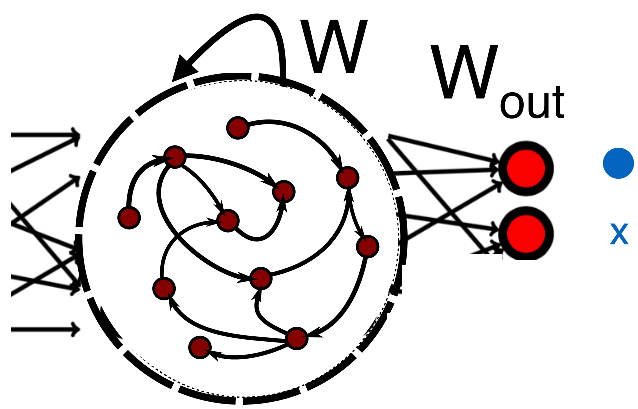
UWIN, NSF, NIH Comp Neuro TG,
Swartz Foundation

We thank the Allen Institute for Brain Science founder Paul G. Allen for his vision, encouragement, and support.

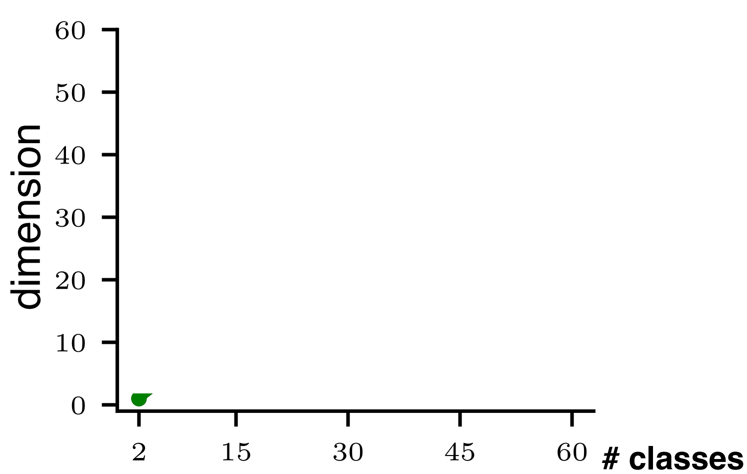
extras

Delayed Stimulus Classification Task

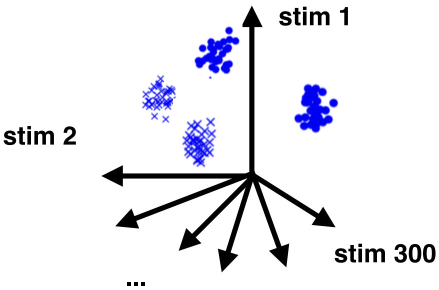


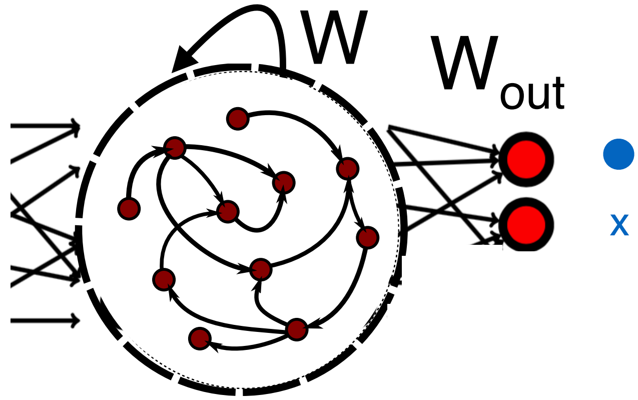


Easy classification.
Stim lives in high dim.

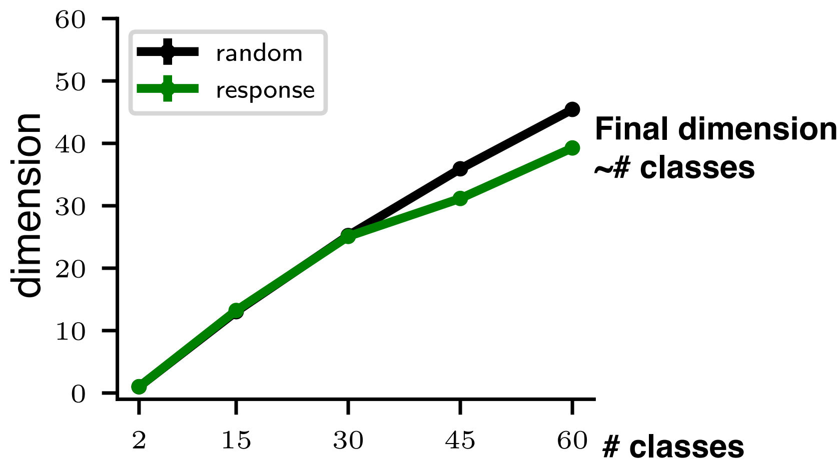


Delayed Stimulus Classification Task

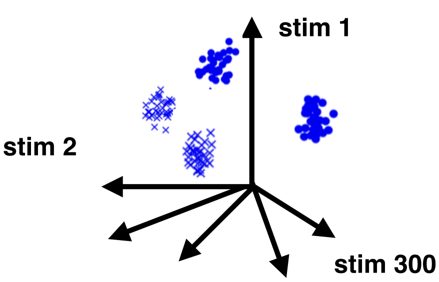




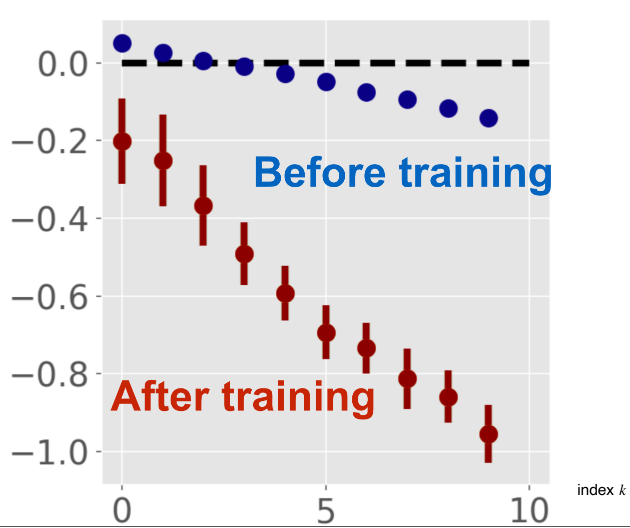
Easy classification.
Stim lives in high dim.



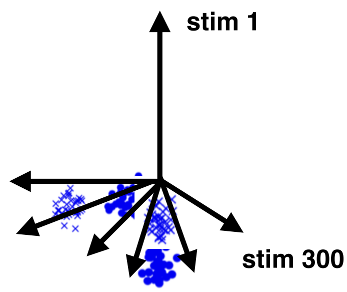
Collapse of dimension: mechanism



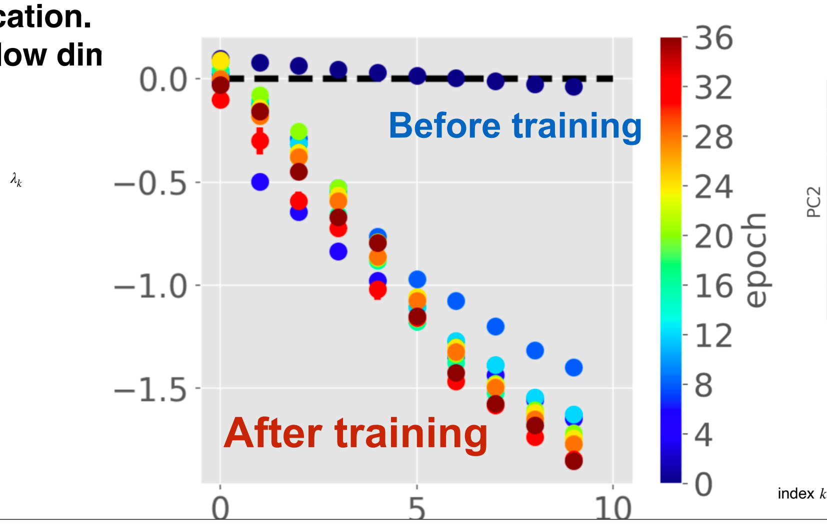
Easy classification.
Stim lives in high dim.

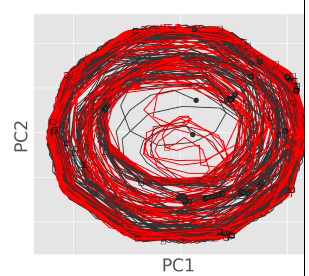


Collapse of dimension: mechanism



Hard classification.
Stim lives in low din





Showing that our dim collapse is due to correlated activity, not just very heterogeneous levels of variance / activity from one cell to next

- red squares are theory values (as in the last plot of the paper)
- red Xs are covariance matrix where diagonal values have been replaced the mean of the diagonal: C -> C-diag(diag(C)) + eye(N)*mean(diag(C))
- blue dots are covariance matrices where the diagonal is kept and outer diagonal entries are randomly reassigned C -> randomize(triu(C)) + randomize(tril(C)) + diag(diag(C))

I think this is good because the last C structure keeps average correlations but destroys the eigenvalue structure. I hope it is clear, but let me know if not.