Deep Learning of dynamical systems for mechanistic insight and prediction in psychiatry

Daniel Durstewitz

Dept. of Theoretical Neuroscience

Central Institute for Mental Health/ Heidelberg University

(I) Brief primer on DST
→ role in neuroscience & psychiatry

(II) Deep learning of dynamical systems from time series observations

(III) Applications in psychiatry
[fMRI, mobile data]

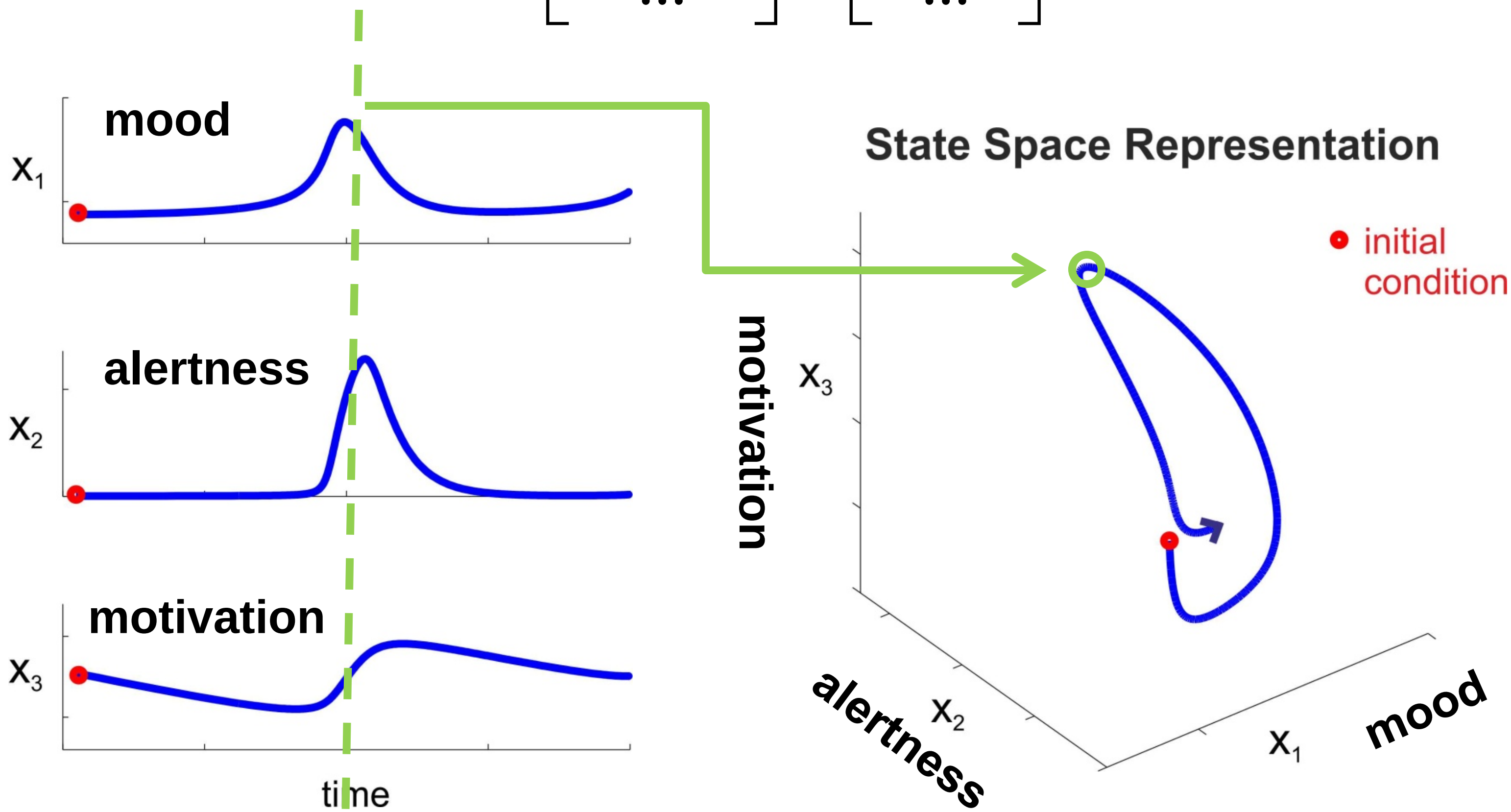
(I) Brief primer on DST
→ role in neuroscience & psychiatry

(II) Deep learning of dynamical systems from time series observations

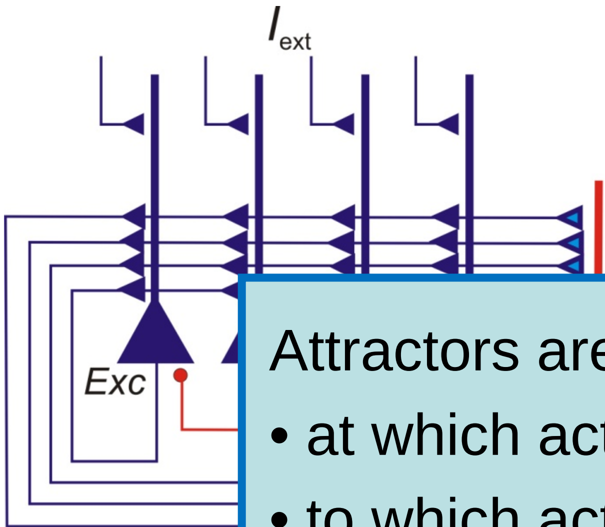
(III) Applications in psychiatry [fMRI, mobile data]

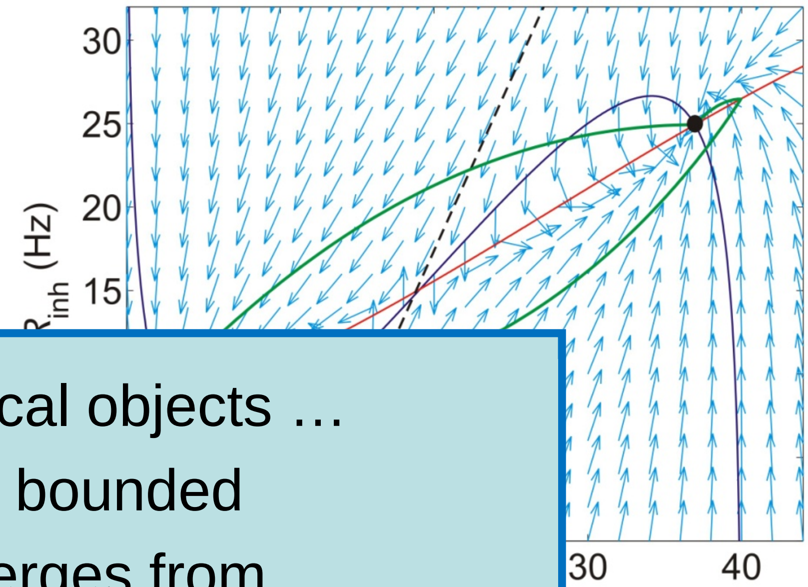
What is a dynamical system?

$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} dx_1/dt \\ dx_2/dt \\ dx_3/dt \\ \ldots \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \\ \ldots \end{bmatrix}$$



Attractor states in state space



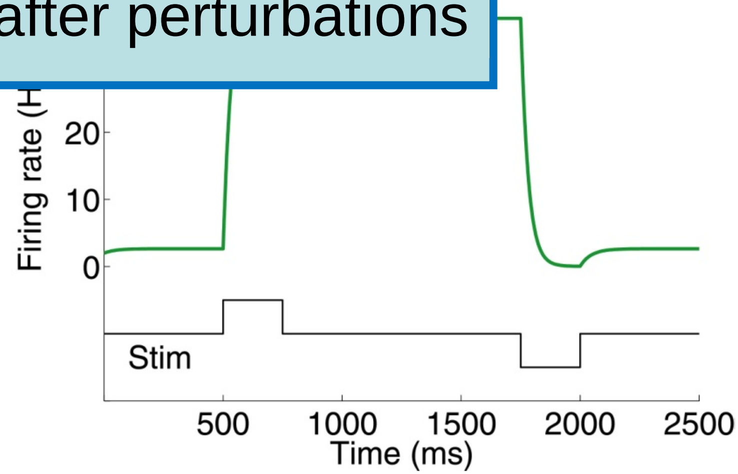


Attractors are geometrical objects …

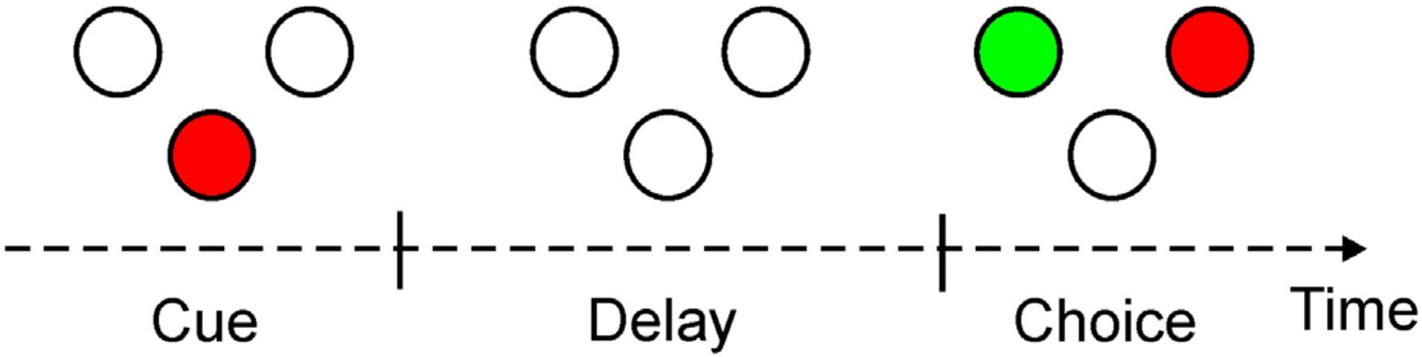
- at which activity stays bounded
- to which activity converges from neighborhood (basin of attraction)
- to which activity returns after perturbations

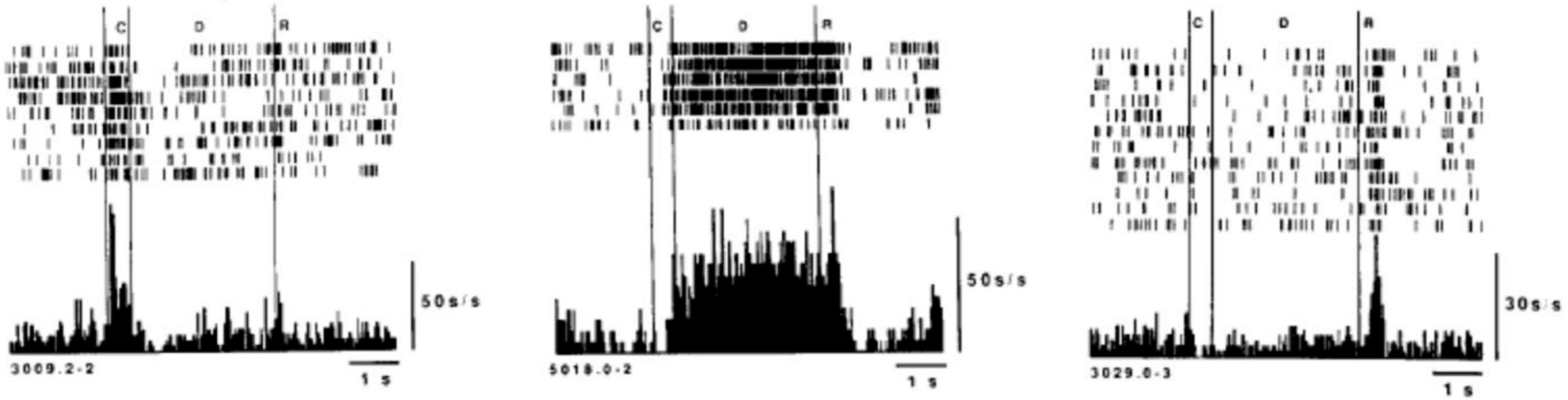
$$\tau_{exc}\frac{dR_{exc}}{dt} = -R \ldots$$

$$\tau_{inh}\frac{dR_{inh}}{dt} = -R_{inh} + \sigma\left(N_e w_{ie} R_{exc} + I_{ext,i}\right) = 0$$



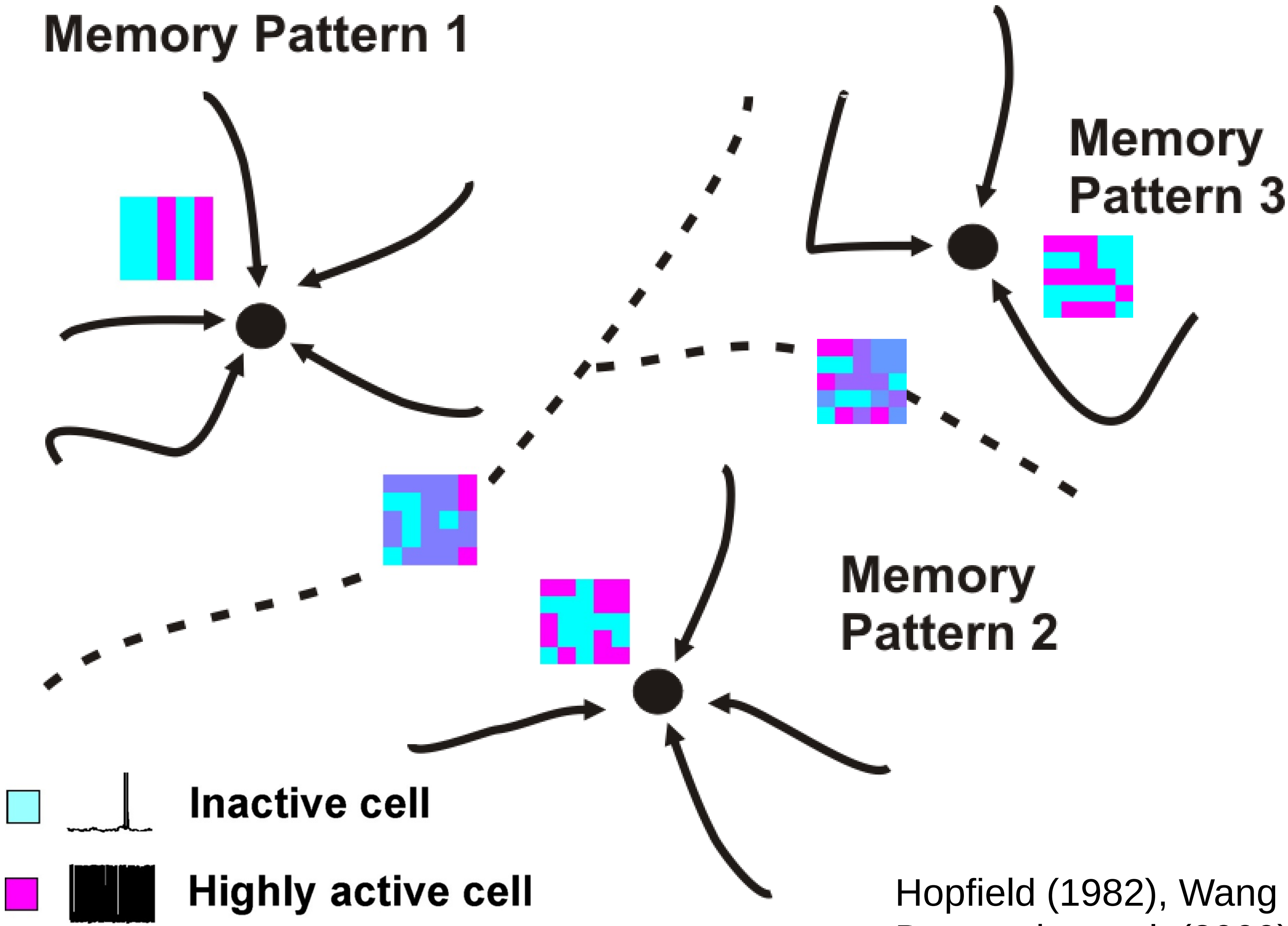
Working memory tasks & persistent activity





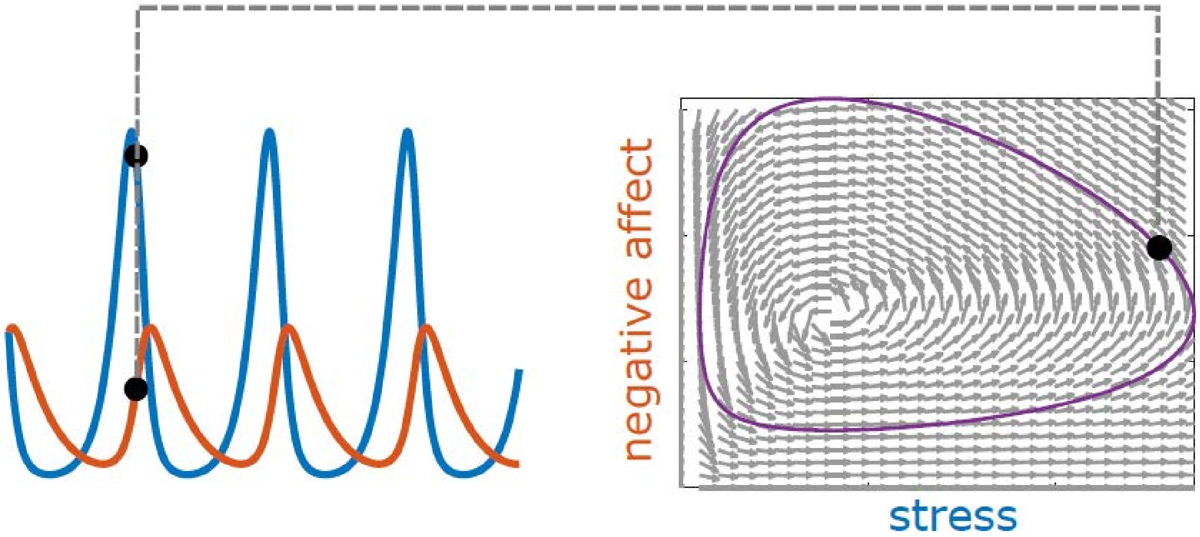
Goldman-Rakic (1990)

Memory patterns as attractor states

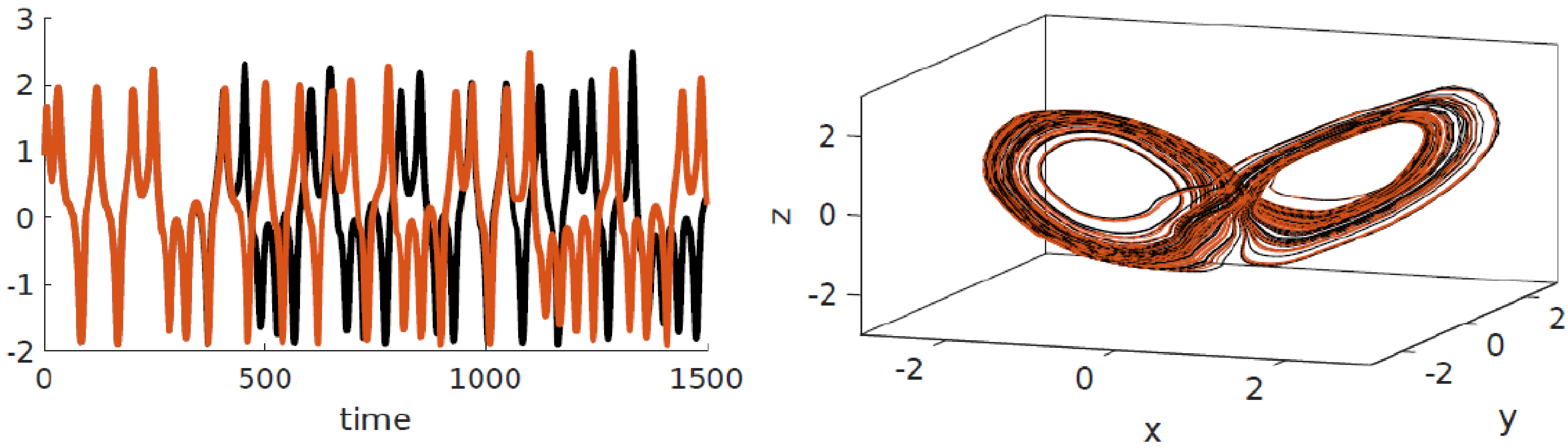


Hopfield (1982), Wang (1999), Durstewitz et al. (2000)

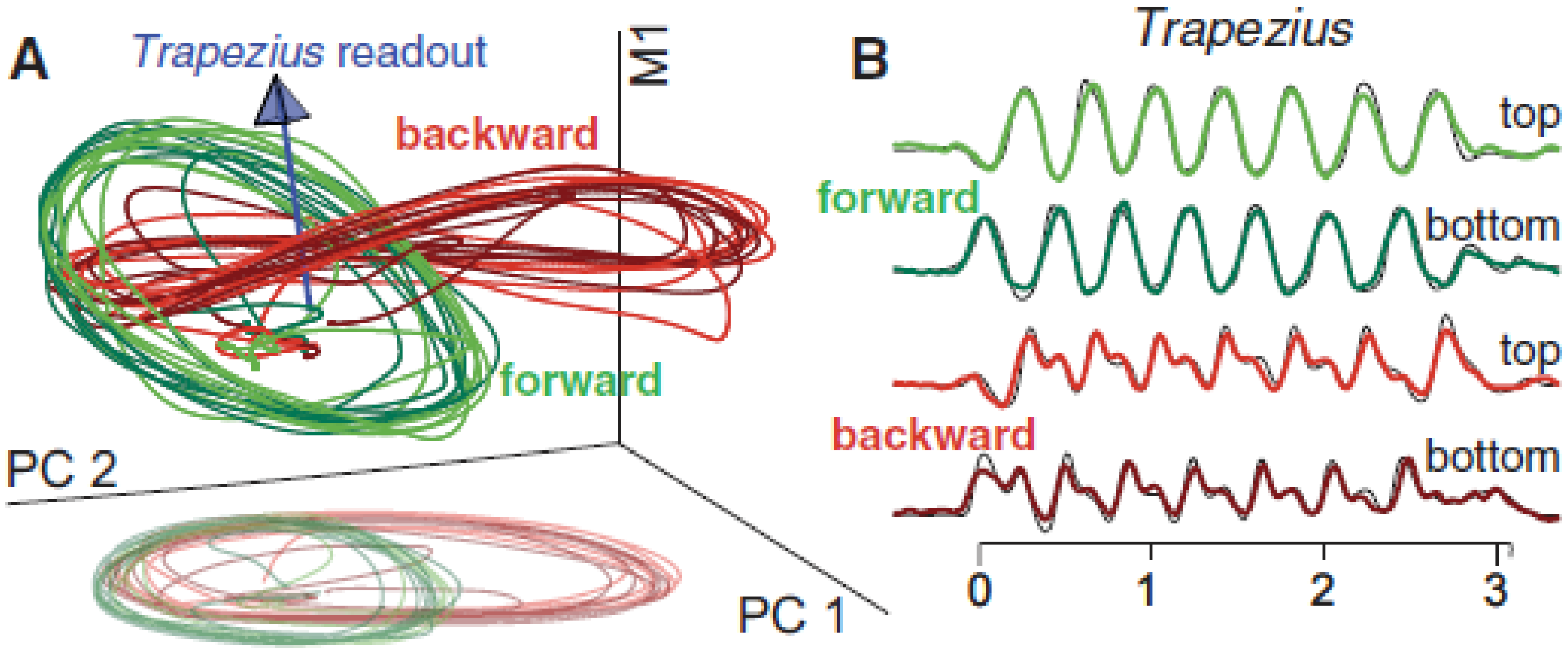
Limit cycles ...



... and chaotic attractors

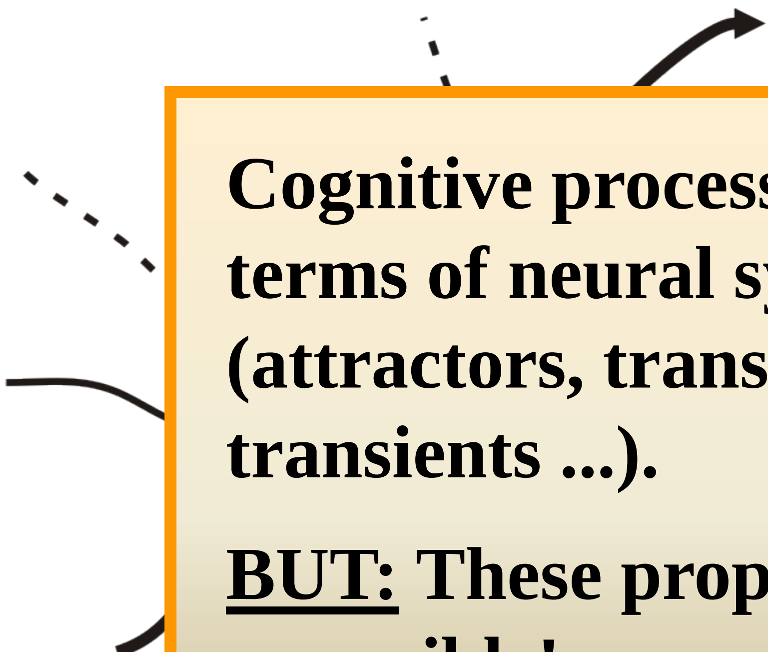


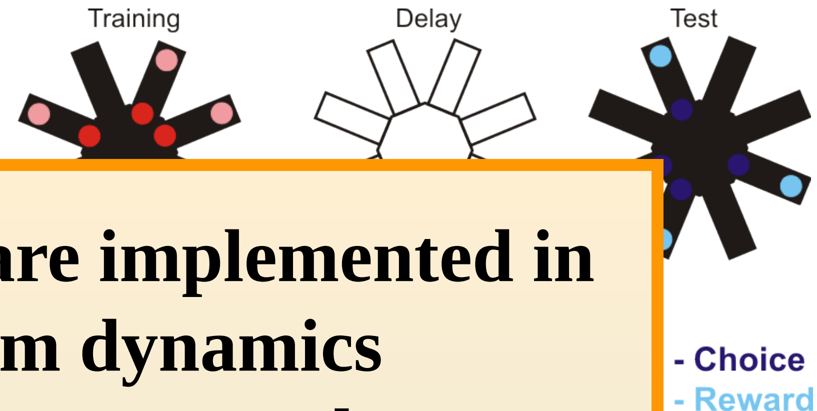
Limit cycles in motor behavior



Russo, Churchland et al (2018), Neuron

Action/ thought sequences as “heteroclinic channels”

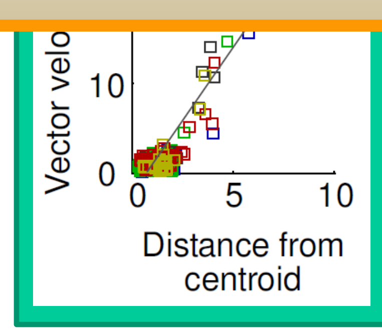


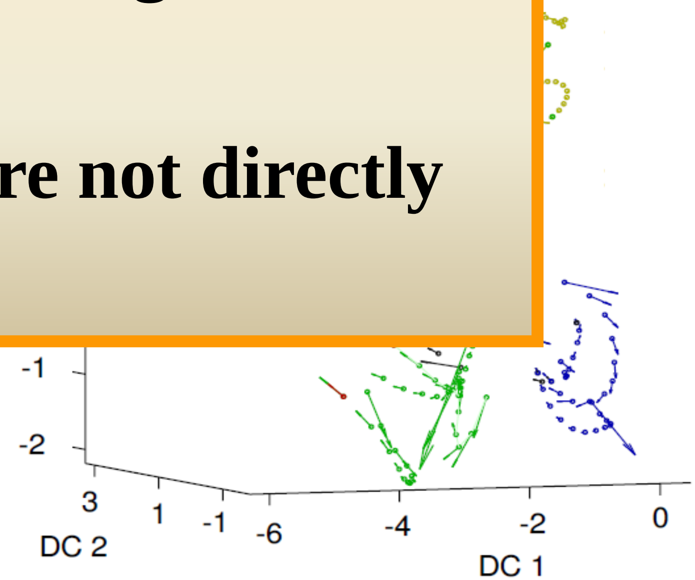


Cognitive processes are implemented in terms of neural system dynamics (attractors, transitions among these, transients ...).

<u>BUT:</u> These properties are not directly accessible!

e.g. Rabinovich et al. (2008)

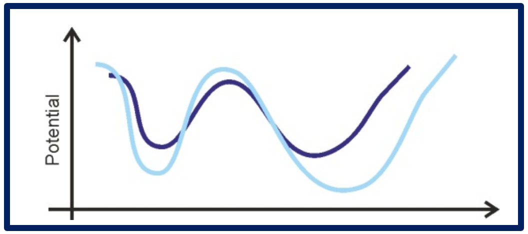




Balaguer-Ballester, Lapish et al. (2011) *PLoS Comp Biol*
Lapish, Balaguer-Ballester et al. (2015) *J Neurosci*

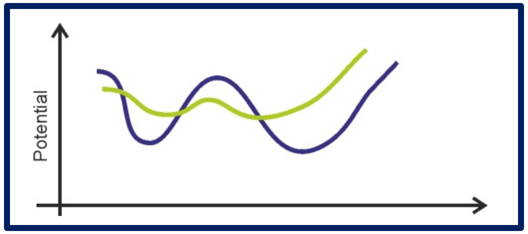
Altered dynamics in psychiatric states

Durstewitz, Huys, Koppe (2020) *Biological Psychiatry: CNNI*



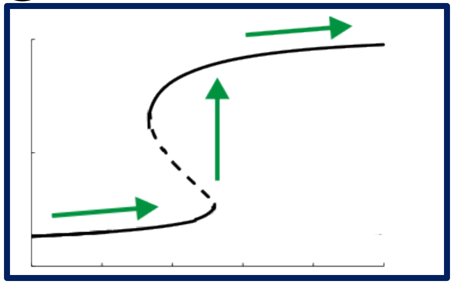
Overly strong attractors:

- Tinnitus
- Rumination
- Compulsivity
- Motor stereotypies
- Changes in cortical oscillations



Too weak attractors:

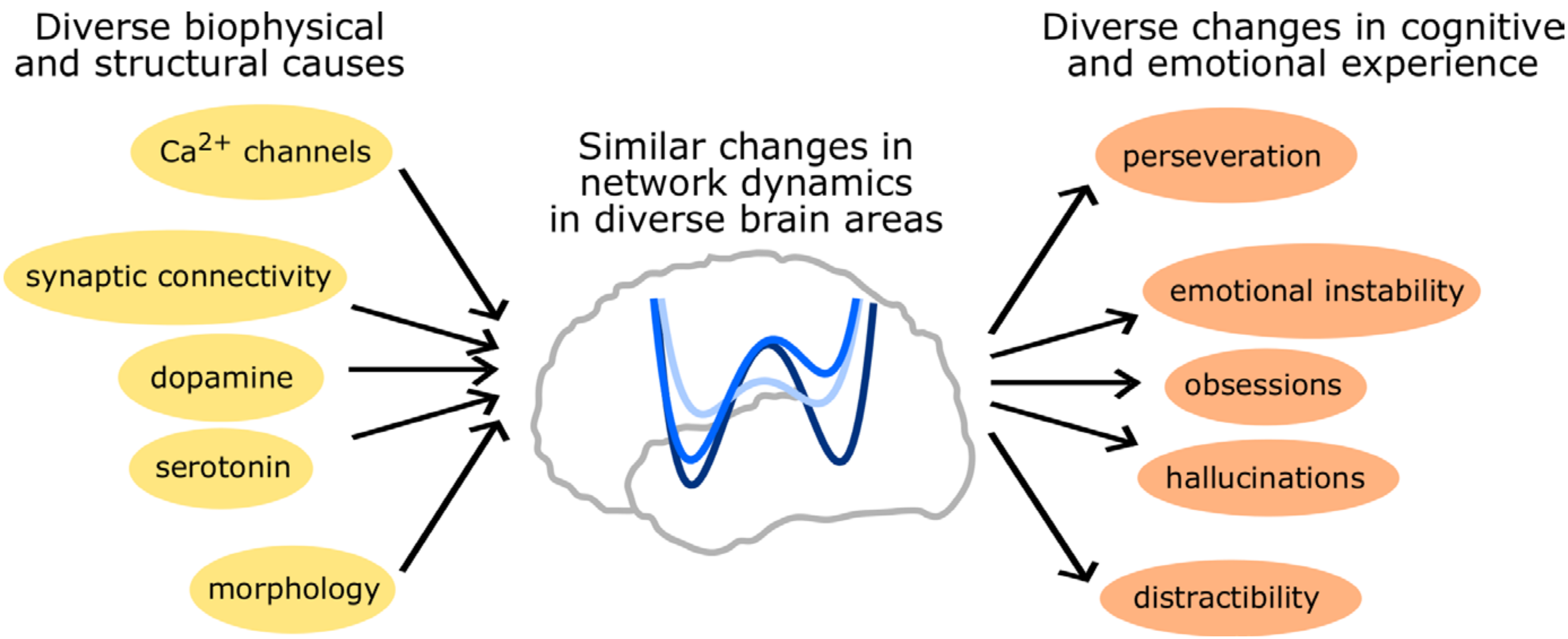
- Incoherent thought
- Hallucinations
- Attention deficits



Bifurcations:

- Sudden onset/ offset of symptoms

Dynamical systems as a central layer of convergence



➔ Implications for treatment

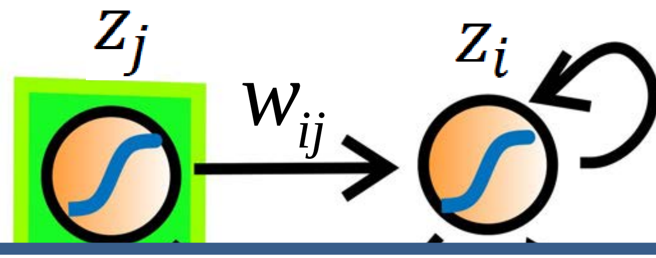
Durstewitz, Huys, Koppe (2020) Biological Psychiatry: CNNI

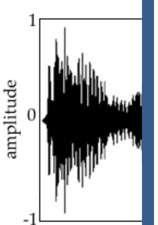
(I) Brief primer on DST
→ role in neuroscience & psychiatry

(II) Deep learning of dynamical systems from time series observations

(III) Applications in psychiatry
[fMRI, mobile data]

Recurrent Neural Network (→ time series)





Dynamically universal: Can approximate arbitrarily closely flow field of any other DS

- 2-layer NN approximates any real-valued F on compact set (Cybenko 1989)

$$y = F(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \sigma(\boldsymbol{\omega}^T \mathbf{x} + h_i) \,, \mathbf{x} \in \mathbb{R}^K$$

- approx. flow-field by 2L-NN, reformulate as RNN (Funahashi & Nakamura 1993; Kimura & Nakano 1998)

external input

Making RNN deep in time: "Vanishing/ exploding gradient problem"

"Inputs": $\boldsymbol{S} = \{\boldsymbol{s}_t\}$ "Outputs": $\widetilde{\boldsymbol{Z}} = \{\tilde{\boldsymbol{z}}_t\}, t = 1 \ldots T$

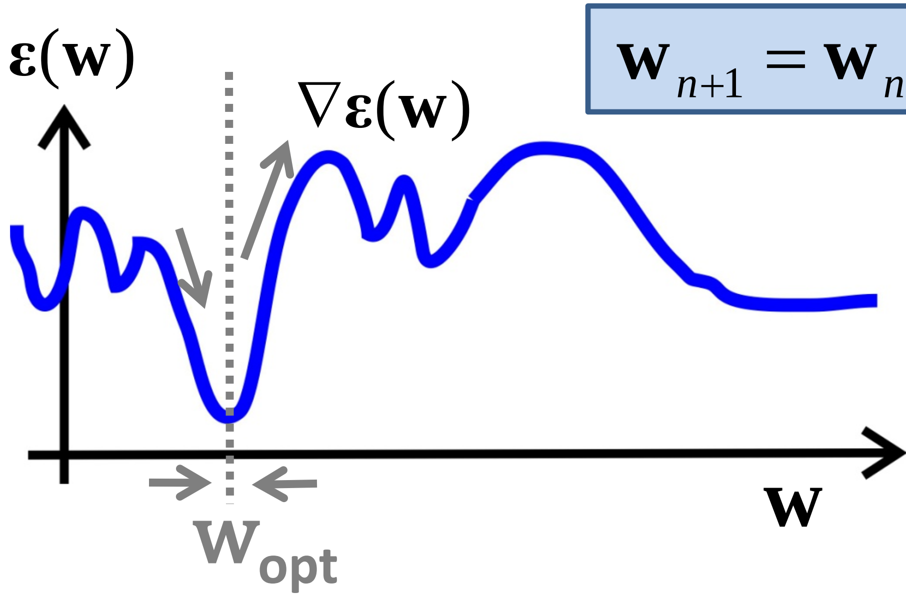
RNN $\boldsymbol{z}_t \in \mathbb{R}^M$

$$\boldsymbol{z}_t = \phi(\boldsymbol{W}\boldsymbol{z}_{t-1} + \boldsymbol{h} + \boldsymbol{C}\boldsymbol{s}_t)$$

Loss function

$$\boldsymbol{\varepsilon} = \sum_{r=1}^{R} \sum_{t=1}^{T} \|\tilde{\boldsymbol{z}}_t - \boldsymbol{z}_t\|_2^2$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \gamma \nabla \boldsymbol{\varepsilon}(\mathbf{w}_n)$$



Making RNN deep in time: "Vanishing/ exploding gradient problem"

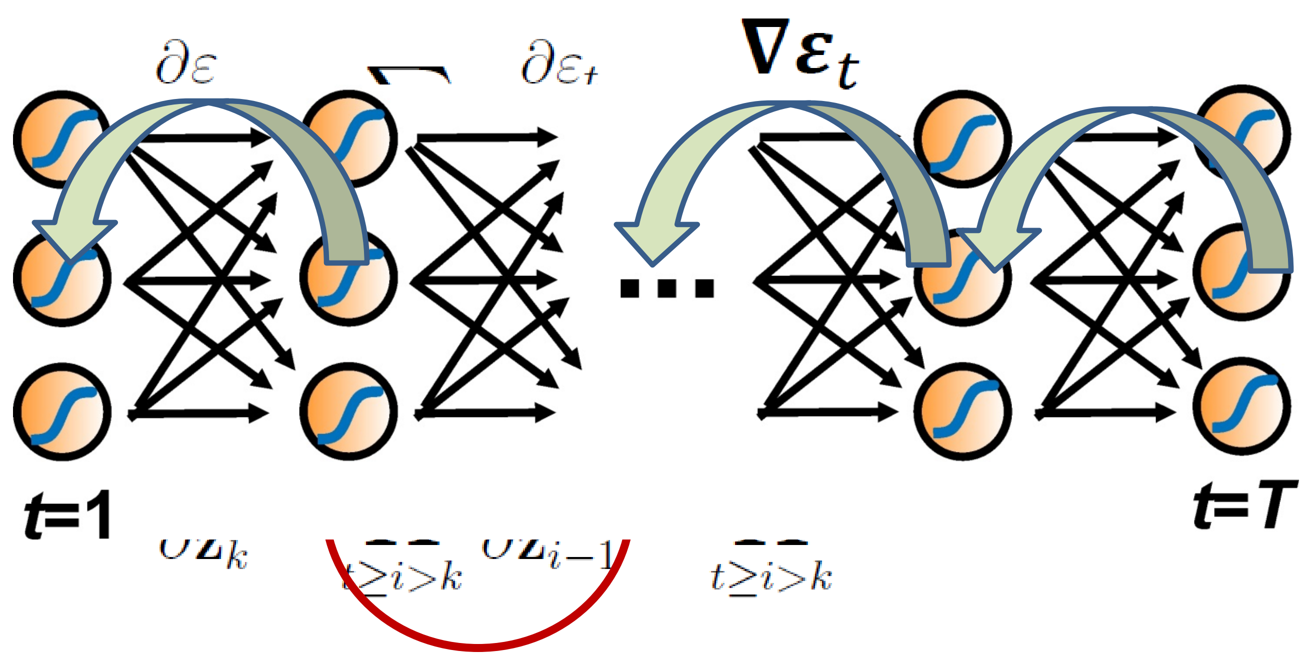
"Inputs": $\boldsymbol{S} = \{\boldsymbol{s}_t\}$ "Outputs": $\widetilde{\boldsymbol{Z}} = \{\tilde{\boldsymbol{z}}_t\}, t = 1 \dots T$

RNN $\boldsymbol{z}_t \in \mathbb{R}^M$

$$\boldsymbol{z}_t = \phi(\boldsymbol{W}\boldsymbol{z}_{t-1} + \boldsymbol{h} + \boldsymbol{C}\boldsymbol{s}_t)$$

Loss function

$$\boldsymbol{\varepsilon} = \sum_{r=1}^{R} \sum_{t=1}^{T} \|\tilde{\boldsymbol{z}}_t - \boldsymbol{z}_t\|_2^2$$



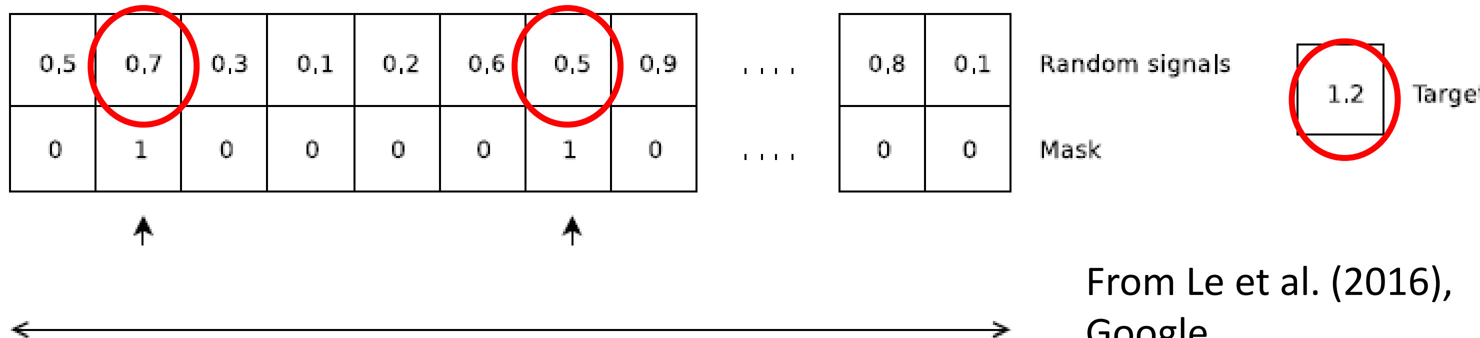
Making RNN deep in time …

Example: Addition problem

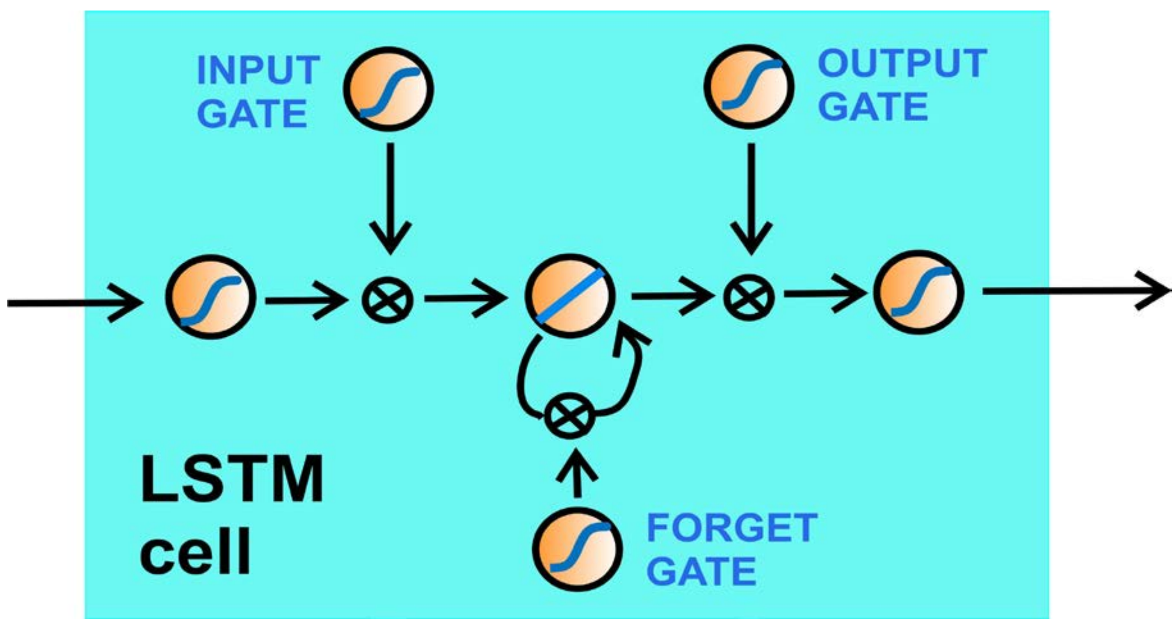
$$\tilde{z}_T = s_{1,t_1} + s_{1,t_2}$$

$s_{1,:} \in [0,1]$

$s_{2,:} \in \{0,1\}$



From Le et al. (2016), Google

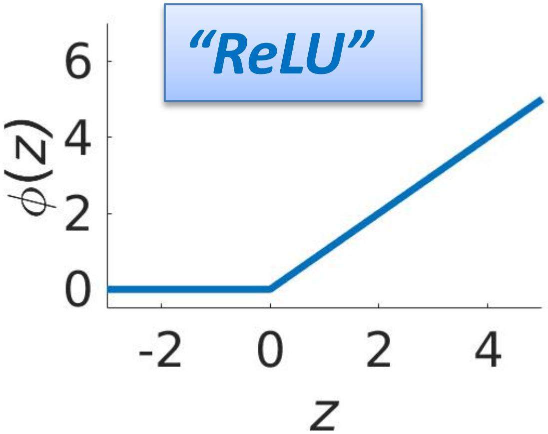


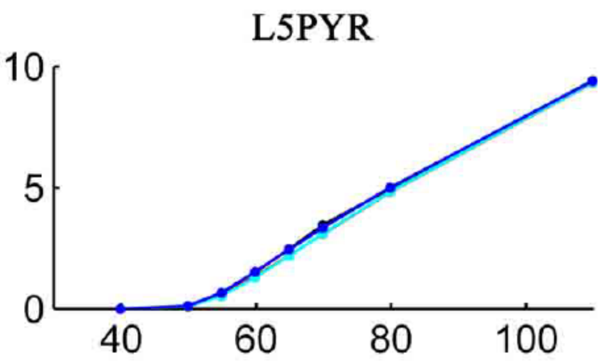
Hochreiter & Schmidhuber (1997)

Piecewise-Linear (PL) RNN

$$z_t = Az_{t-1} + W\phi(z_{t-1}) + Cs_t + h$$
$$\phi(z)_i = \max(0, z_i), i \in \{1, \dots, M\}$$

- Physiological motivation





- Many properties analytically accessible: fixed points, limit cycles, eigenvalue spectra (stability) ...
- Can be transformed into piecewise ODE system
- Efficient inference!

Line-attractor regularization

$$\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{z}_{t-1} + \boldsymbol{W}\phi(\boldsymbol{z}_{t-1}) + \boldsymbol{C}\boldsymbol{s}_t + \boldsymbol{h}$$

$$\boldsymbol{A} = \mathrm{diag}([a_{11} \ldots a_{MM}]) \in \mathbb{R}^{M \times M}$$

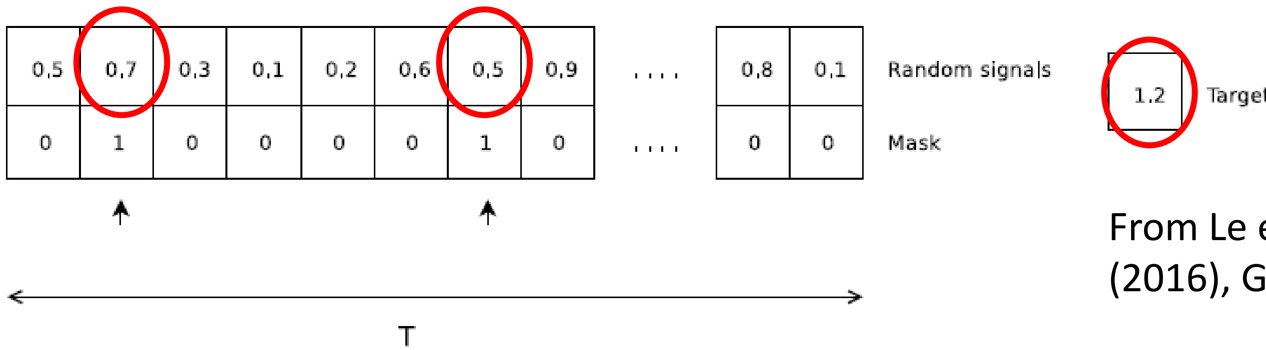
$$\mathrm{L}_{\mathrm{reg}} = \tau_A \sum_{i=1}^{M_{\mathrm{reg}}} (A_{i,i} - 1)^2 + \tau_W \sum_{i=1}^{M_{\mathrm{reg}}} \sum_{\substack{j=1 \\ j \neq i}}^{M} W_{i,j}^2 + \tau_h \sum_{i=1}^{M_{\mathrm{reg}}} h_i^2$$

penalty term $M_{\mathrm{reg}} \leq M$

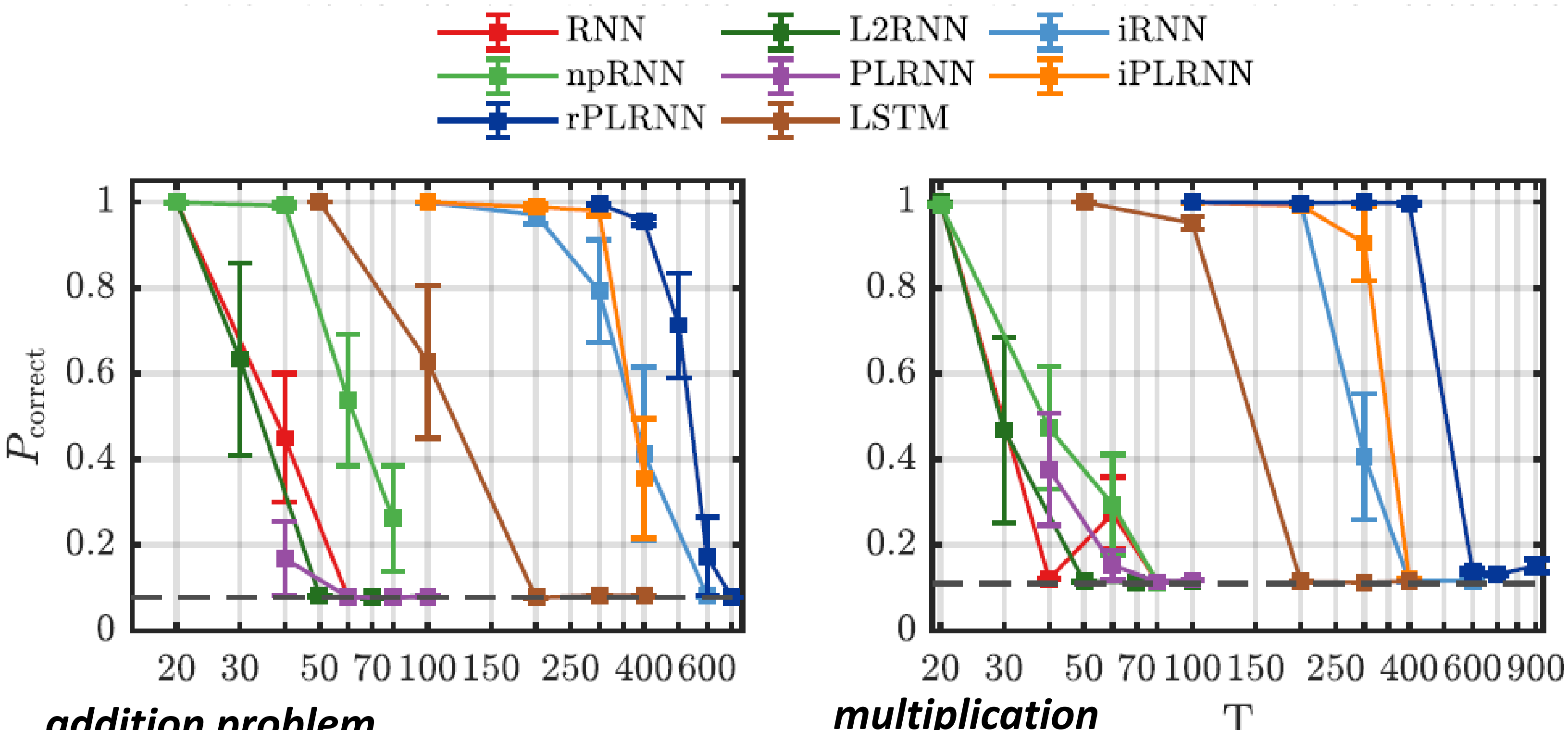
$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & 0 & \times & 0 \\ 0 & 0 & 0 & 0 & 0 & \times \end{pmatrix} \quad \boldsymbol{W} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \times & \times & \times & 0 & \times & \times \\ \times & \times & \times & \times & 0 & \times \\ \times & \times & \times & \times & \times & 0 \end{pmatrix} \quad \boldsymbol{h} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \times \\ \times \\ \times \end{pmatrix}$$

Schmidt et al. (2019) arXiv

Performance on ML benchmarks



From Le et al. (2016), Google

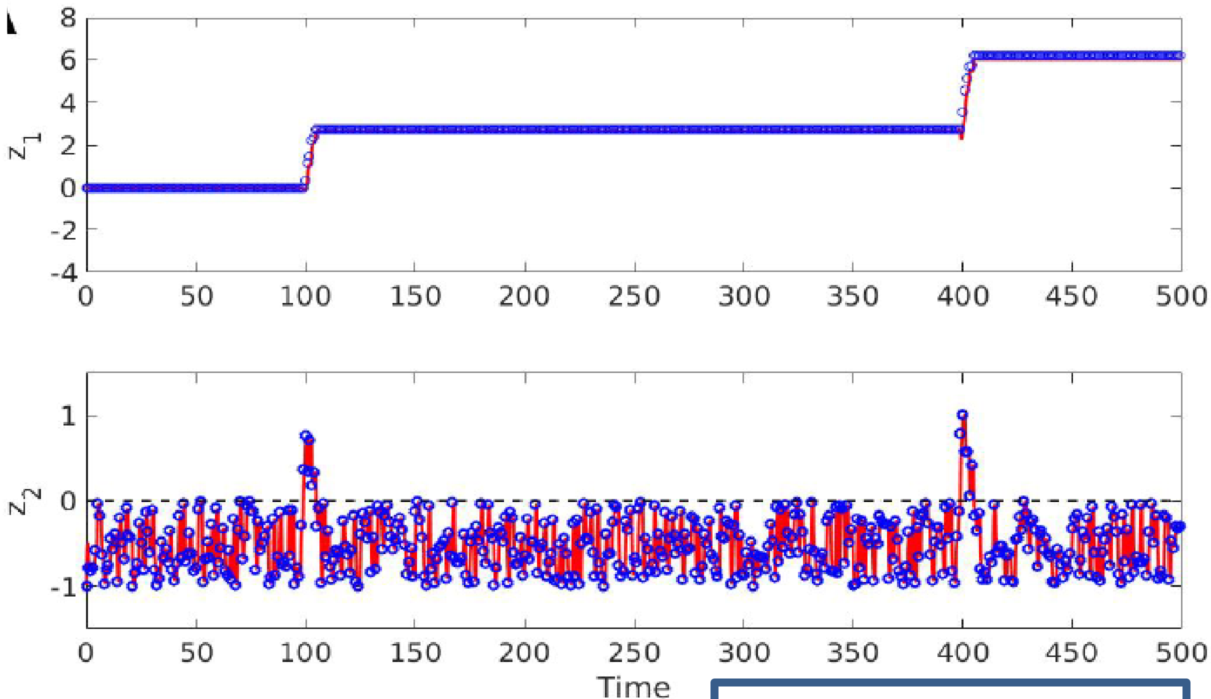


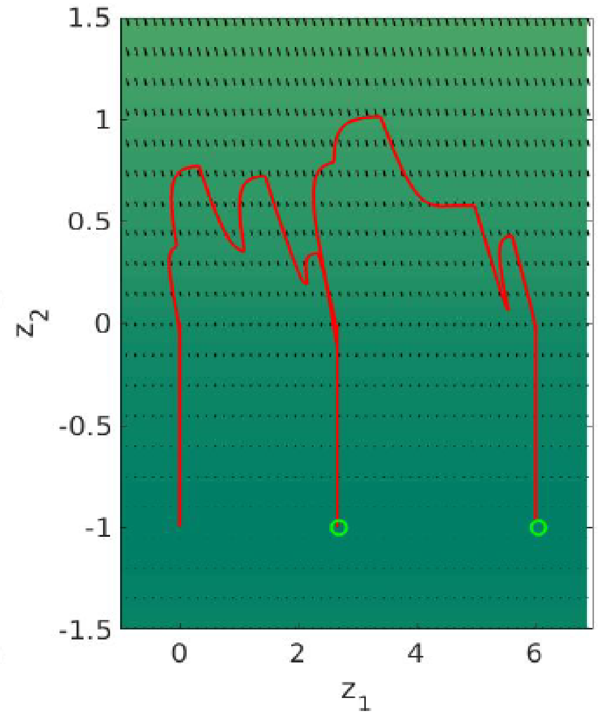
addition problem

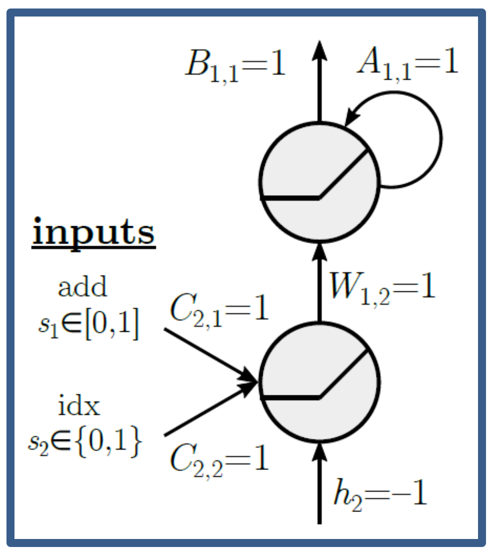
multiplication problem

Schmidt et al. (2019) arXiv

Line-attractors and solving long-range tasks

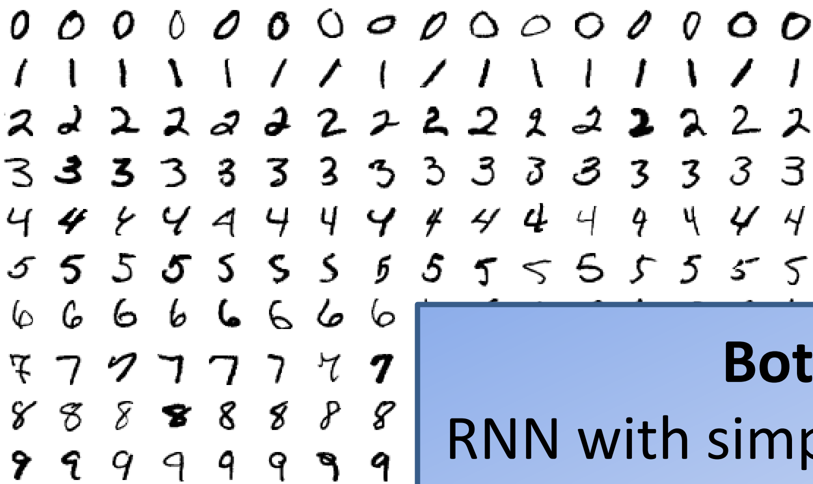


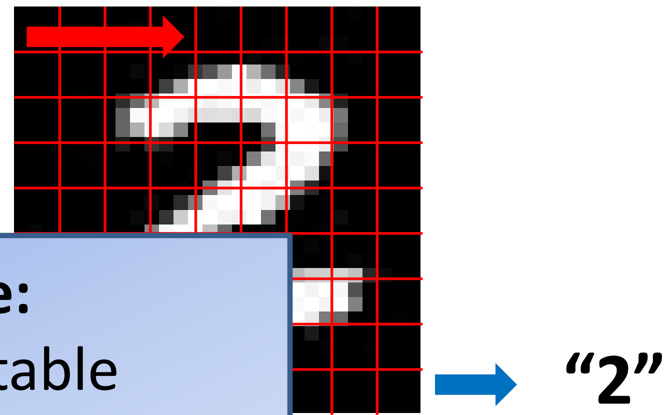




Schmidt et al. (2019) arXiv,
Monfared & Durstewitz (submitted)

Sequential MNIST benchmark

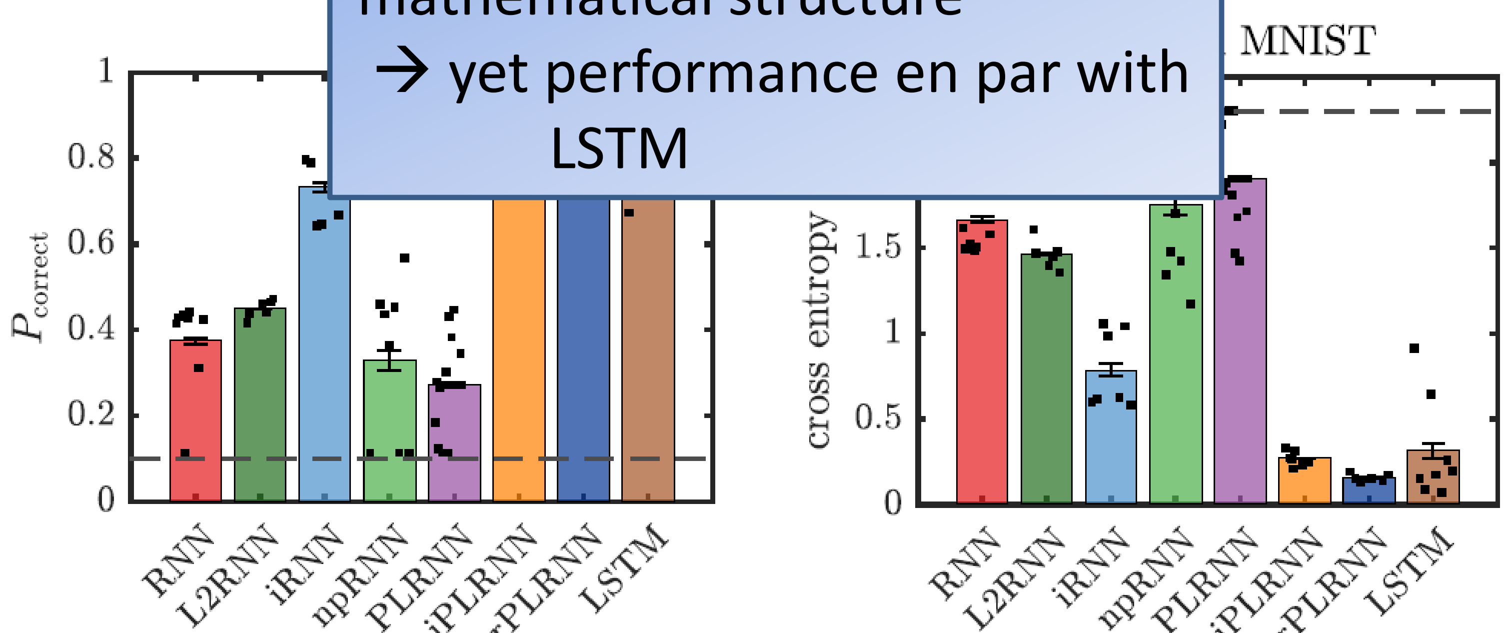




Bottom line:

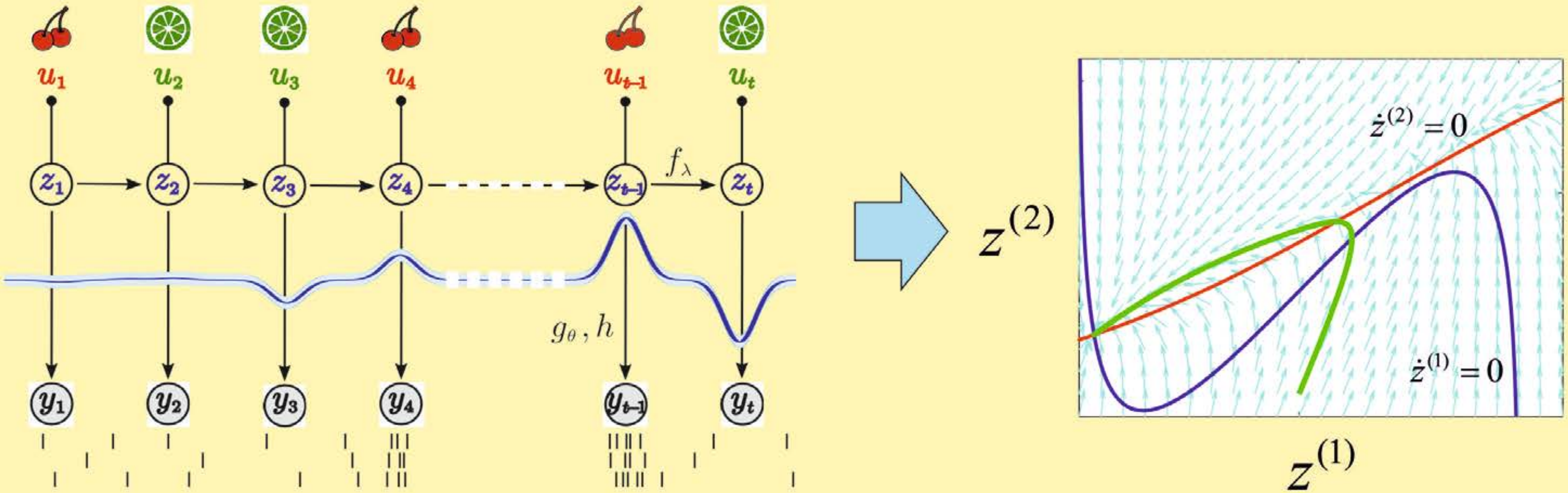
RNN with simple, tractable mathematical structure

→ yet performance en par with LSTM



Schmidt et al. (2019) arXiv

Generative PLRNN for dynamical systems

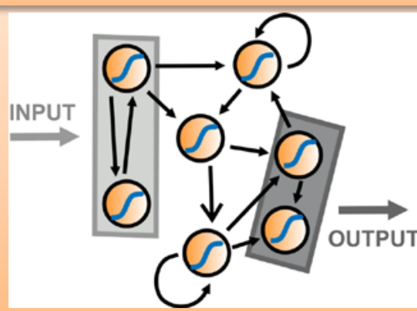


$$p_{\boldsymbol{\theta}}(\boldsymbol{Z},\boldsymbol{Y}) = p_{\boldsymbol{\theta}_{obs}}(\boldsymbol{Y}|\boldsymbol{Z})\,p_{\boldsymbol{\theta}_{lat}}(\boldsymbol{Z})$$

Based on:
Durstewitz, Koppe, Toutounji (2016) Current Opinion in Behavioral Sciences;
Durstewitz (2009) Neural Networks

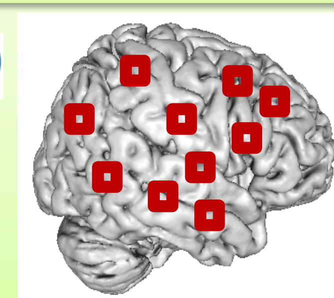
"Generic" dynamical systems model

$p_{\boldsymbol{\theta}_{lat}}(\boldsymbol{Z})$

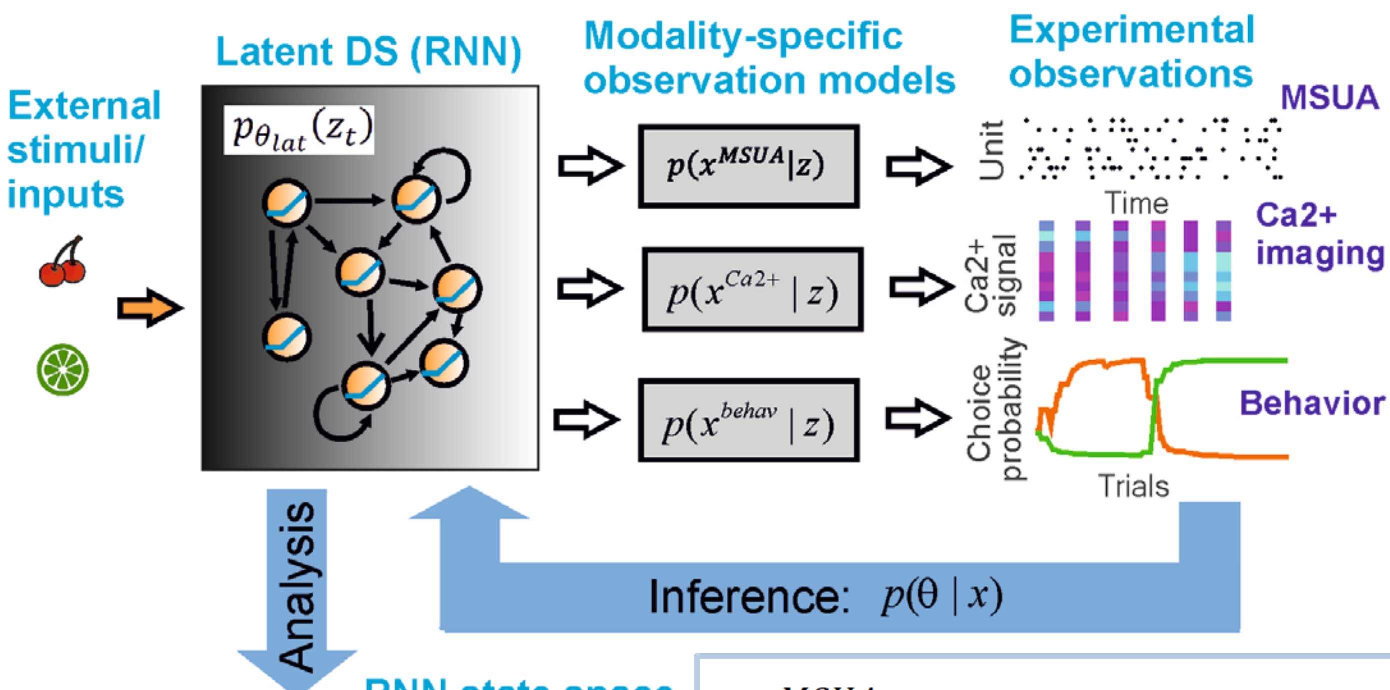


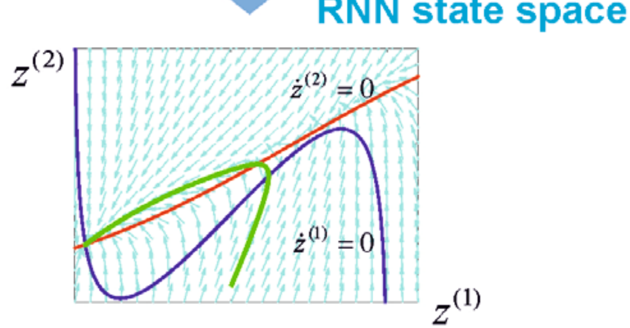
observation model

$p_{\boldsymbol{\theta}_{obs}}(\boldsymbol{Y}|\boldsymbol{Z})$



Generative PLRNN for dynamical systems





$$\mathbf{z}_t|\mathbf{z}_{t-1} \sim N(\boldsymbol{\mu}_t^{(z)}, \boldsymbol{\Sigma}_t^{(z)})$$

$$\boldsymbol{\mu}_t^{(z)} = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{W}\max[0, \mathbf{z}_{t-1}] + \mathbf{h} + \mathbf{C}\mathbf{s}_t$$

$$\mathbf{x}_t^{MSUA}|\mathbf{z}_t \sim \text{Poisson}(\boldsymbol{\lambda}_t = g_{\boldsymbol{\theta}}(\mathbf{z}_t))$$

$$\mathbf{x}_t^{Ca2+}|\mathbf{z}_t \dots \mathbf{z}_{t-\tau} \sim N(\boldsymbol{\mu}_t^{(x)}, \boldsymbol{\Sigma}_t^{(x)})$$

$$\boldsymbol{\mu}_t^{(x)} = f_{\boldsymbol{\theta}}^{(\mu)}(\mathbf{z}_t \dots \mathbf{z}_{t-\tau})$$

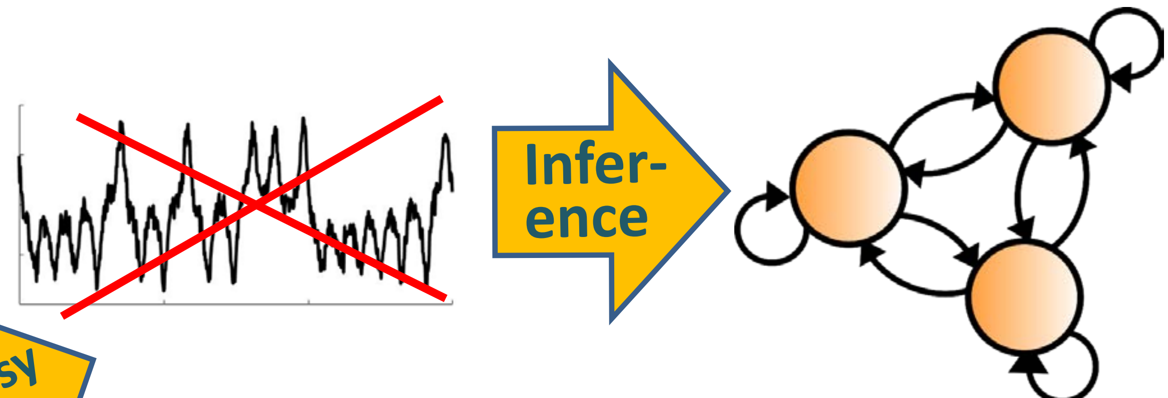
$$\boldsymbol{\Sigma}_t^{(x)} = f_{\boldsymbol{\theta}}^{(\Sigma)}(\mathbf{z}_t \dots \mathbf{z}_{t-\tau})$$

$$\mathbf{x}_t^{behavior}|\mathbf{z}_t \sim \text{Catgorical}(\boldsymbol{\pi}_t = f_{\boldsymbol{\theta}}^{(\pi)}(\mathbf{z}_t))$$

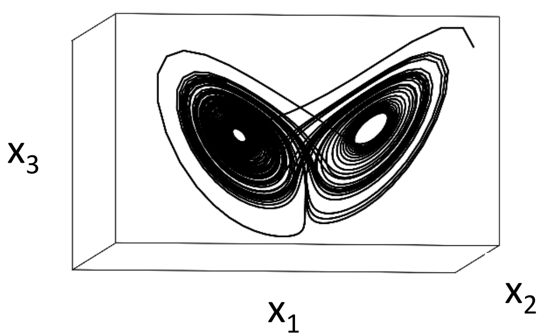
Reconstructing dynamical systems

Lorenz system

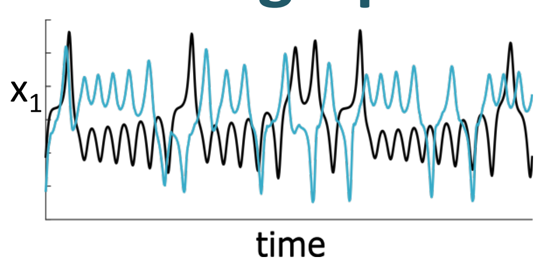
$$dx_1 / dt = s(x_2 - x_1)$$
$$dx_2 / dt = rx_1 - x_2 - x_1x_3$$
$$dx_3 / dt = x_1x_2 - bx_3$$



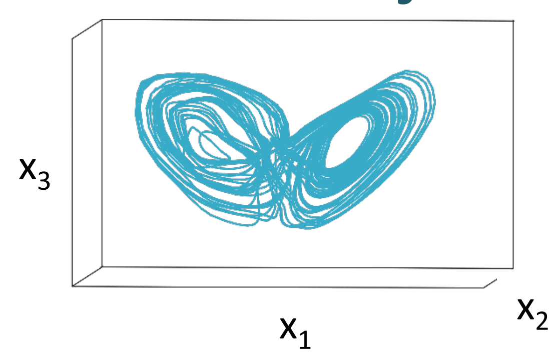
True trajectory



Time graph



Simulated trajectory



Statistical inference for small data: Expectation-Maximization

Estimation by Maximum Likelihood

observed

latent

Maximize $\log p(\mathbf{X} \mid \boldsymbol{\theta}) = \log \int_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) d\mathbf{Z}$

w.r.t. $\boldsymbol{\theta} = \{\boldsymbol{\mu}_0, \mathbf{A}, \mathbf{W}, \mathbf{h}, \boldsymbol{\Sigma}, \mathbf{B}, \boldsymbol{\Gamma}\}$

$$\log \int_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) d\mathbf{Z} \geq \int_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})}{q(\mathbf{Z})} d\mathbf{Z}$$

$$= \mathrm{E}_q[\log p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})] + \mathrm{H}[q(\mathbf{Z})]$$

$$= \log p(\mathbf{X} \mid \boldsymbol{\theta}) - \mathrm{KL}[q(\mathbf{Z}) \,\|\, p(\mathbf{Z} \mid \mathbf{X})]$$

M step: Maximize w.r.t. θ

E step: Determine $p(\mathbf{Z}|\mathbf{X})$

Expectation-Maximization Algorithm

$$\log p(\mathbf{X}|\boldsymbol{\theta}) \geq \mathcal{L}(\boldsymbol{\theta}, q)$$ **Evidence Lower Bound (ELBO)**

$$:= \mathrm{E}_q[\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})] + H\big(q(\mathbf{Z}|\mathbf{X})\big)$$

$$= \log p(\mathbf{X}|\boldsymbol{\theta}) - KL\big(q(\mathbf{Z}|\mathbf{X}), p_{\boldsymbol{\theta}}(\mathbf{Z}|\mathbf{X})\big)$$

E-step: $q^* := \arg\max_q \mathcal{L}(\boldsymbol{\theta}^*, q) \quad given\ \boldsymbol{\theta}^*$

M-step: $\boldsymbol{\theta}^* := \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, q^*) \quad given\ q^*$

→ **Laplace/ fixed-point-iteration algorithm efficiently exploiting piecewise-linear structure**

→ **Analytical derivation of moments of** $p_\theta(\mathrm{Z} \mid \mathrm{X})$

Durstewitz (2017), *PLoS Comp Biol* ; Koppe et al. (2019), *PLoS Comp Biol*

Statistical inference for big data: Sequential VAE & SGVB

$$\log p(\mathbf{X}|\boldsymbol{\theta}) \geq \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Psi}) := \mathrm{E}_q[\log p_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{Z})] + H\big(q_{\boldsymbol{\Psi}}(\mathbf{Z}|\mathbf{X})\big)$$

$$\approx \frac{1}{L}\sum_{l=1}^{L}\Big(\log p_{\boldsymbol{\theta}}\big(\mathbf{X}, \mathbf{Z}^{(l)}\big) - \log q_{\boldsymbol{\Psi}}\big(\mathbf{Z}^{(l)}|\mathbf{X}\big)\Big), \mathbf{Z}^{(l)} \sim q_{\boldsymbol{\Psi}}(\mathbf{Z}|\mathbf{X})$$

$$= \frac{1}{L}\sum_{l=1}^{L}\Big(\log p_{\boldsymbol{\theta}}\big(\mathbf{X}, g_{\boldsymbol{\Psi}}(\boldsymbol{\varepsilon}^{(l)}, \mathbf{X})\big) - \log q_{\boldsymbol{\Psi}}\big(g_{\boldsymbol{\Psi}}(\boldsymbol{\varepsilon}^{(l)}, \mathbf{X})|\mathbf{X}\big)\Big)$$

$$\boldsymbol{\varepsilon}^{(l)} \sim p(\boldsymbol{\varepsilon}), \text{ e.g. } p(\boldsymbol{\varepsilon}) = N(\mathbf{0}, \mathbf{I})$$

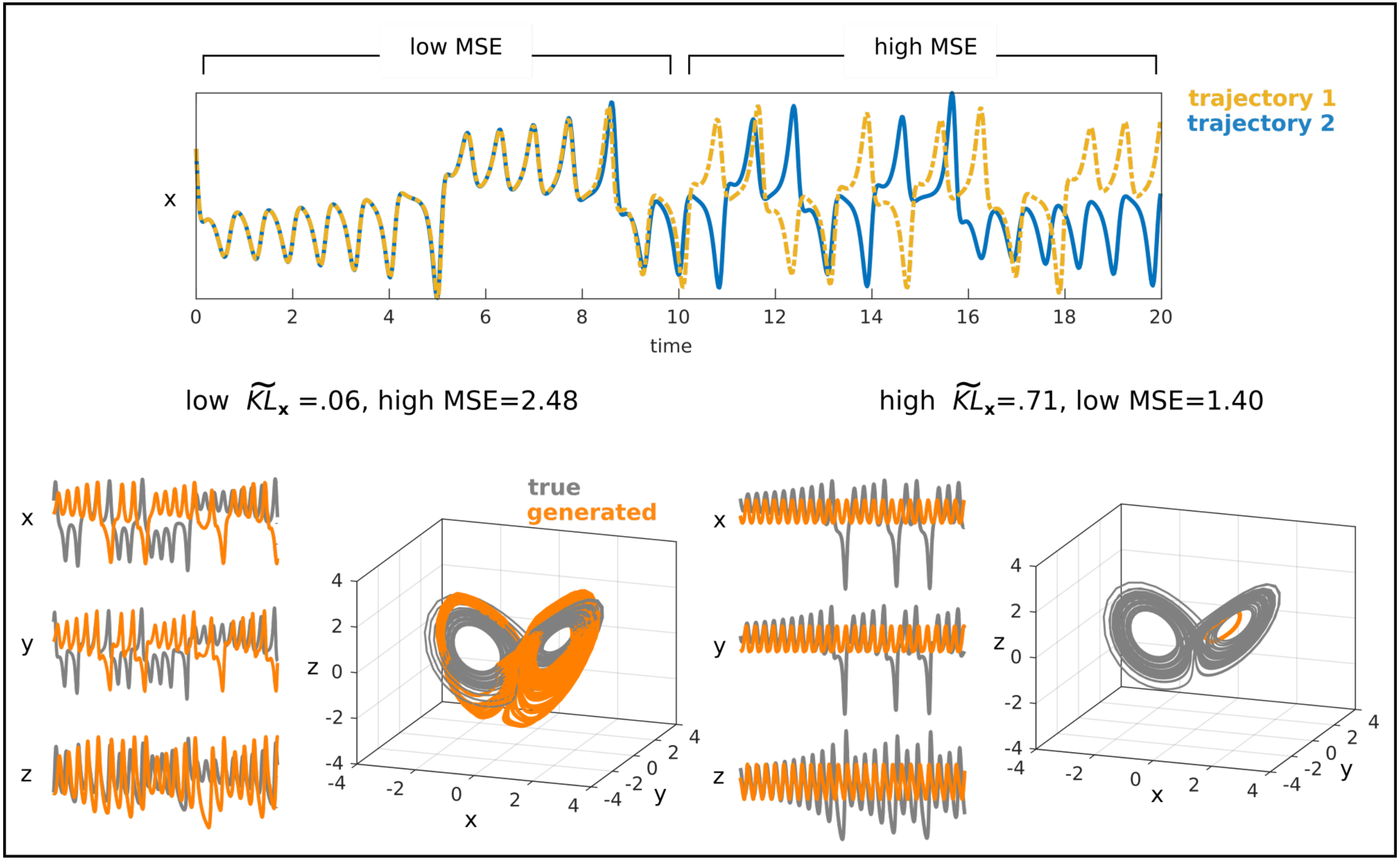
$q_{\boldsymbol{\Psi}}(\mathbf{Z}|\mathbf{X})$ **: Variational density ('encoder model'), e.g. MLP, NF**

$p_{\boldsymbol{\theta}_{obs}}(\mathbf{X}|\mathbf{Z})$ **: Obs. model ('decoder model'), e.g. MLP**

$p_{\boldsymbol{\theta}_{lat}}(\mathbf{Z})$ **: Prior (latent space) model, e.g. PLRNN**

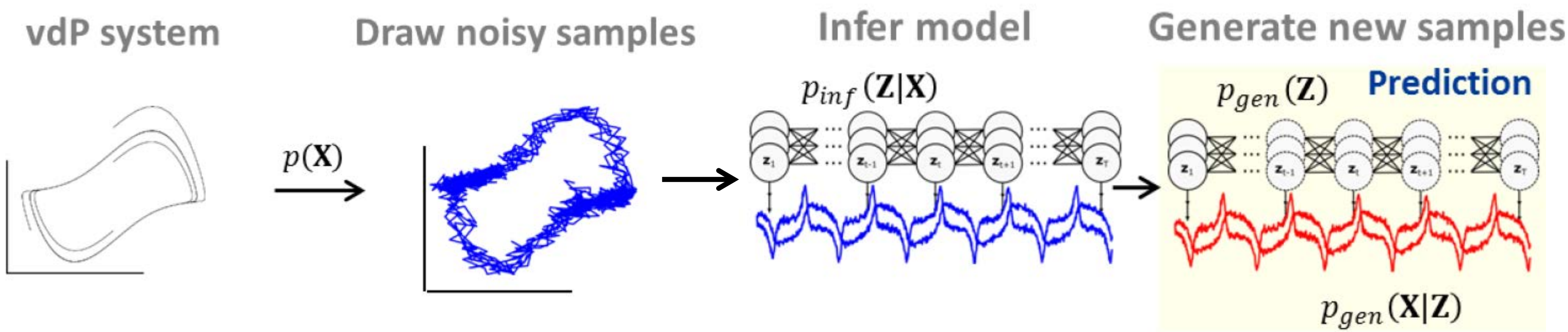
→ Inference through SGD using sampling & 're-parameterization trick' (Kingma & Welling 2014, Rezende et al. 2014)

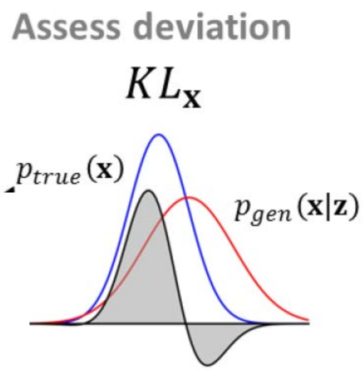
Simple ahead prediction errors may be meaningless



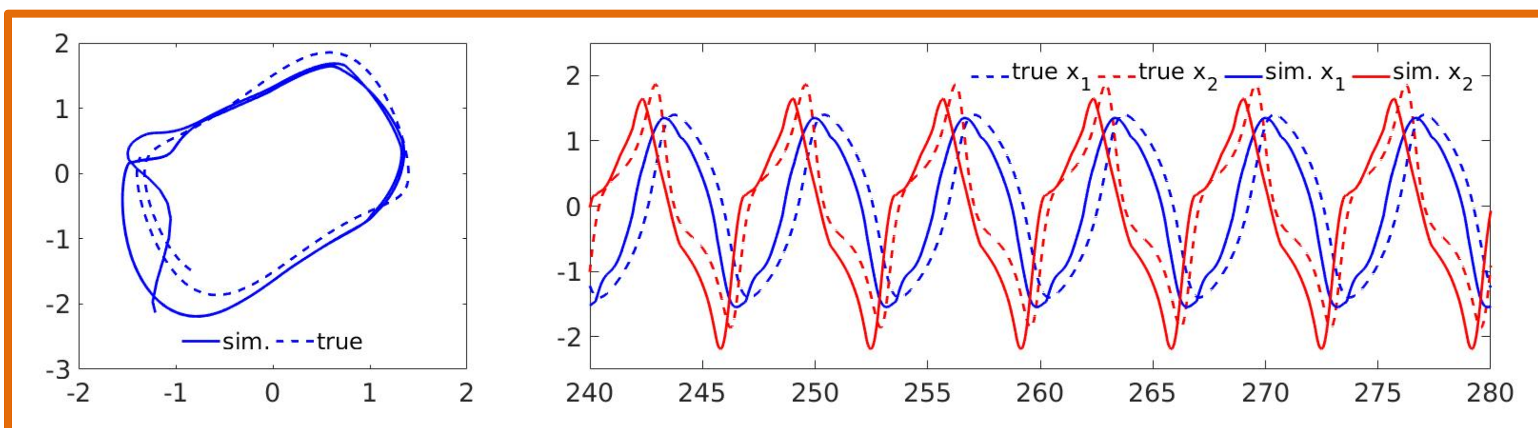
Koppe et al. (2019), *PLoS Comp Biol*

Reconstructing DS benchmarks

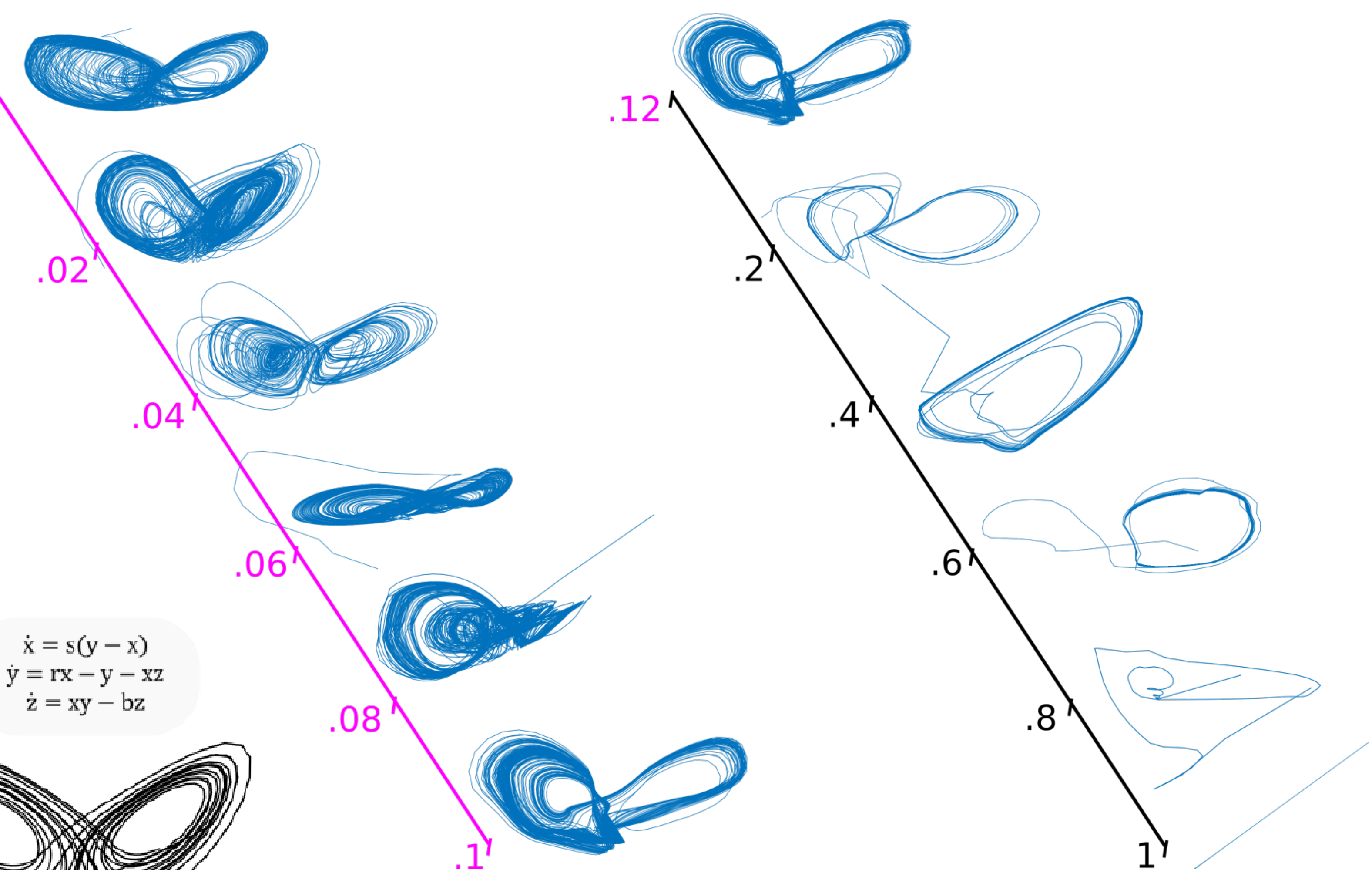




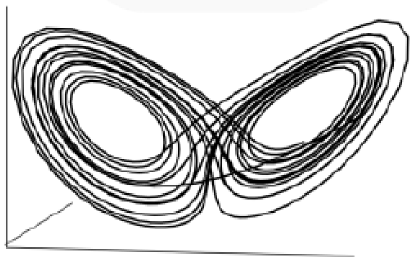
$$KL_{\mathbf{x}}\left(p_{true}(\mathbf{x}), p_{gen}(\mathbf{x}|\mathbf{z})\right) := \int_{\mathbf{x}\in\mathbb{R}^N} p_{true}(\mathbf{x}) \log \frac{p_{true}(\mathbf{x})}{p_{gen}(\mathbf{x}|\mathbf{z})} \mathrm{d}\mathbf{x}$$



Reconstructing DS: Lorenz system

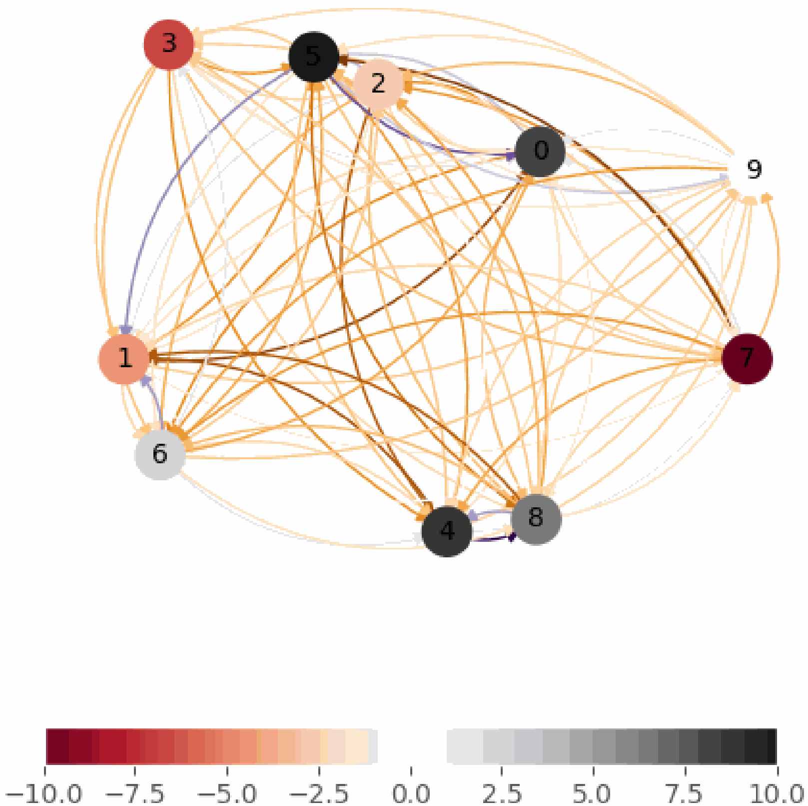


$$\dot{x} = s(y - x)$$
$$\dot{y} = rx - y - xz$$
$$\dot{z} = xy - bz$$

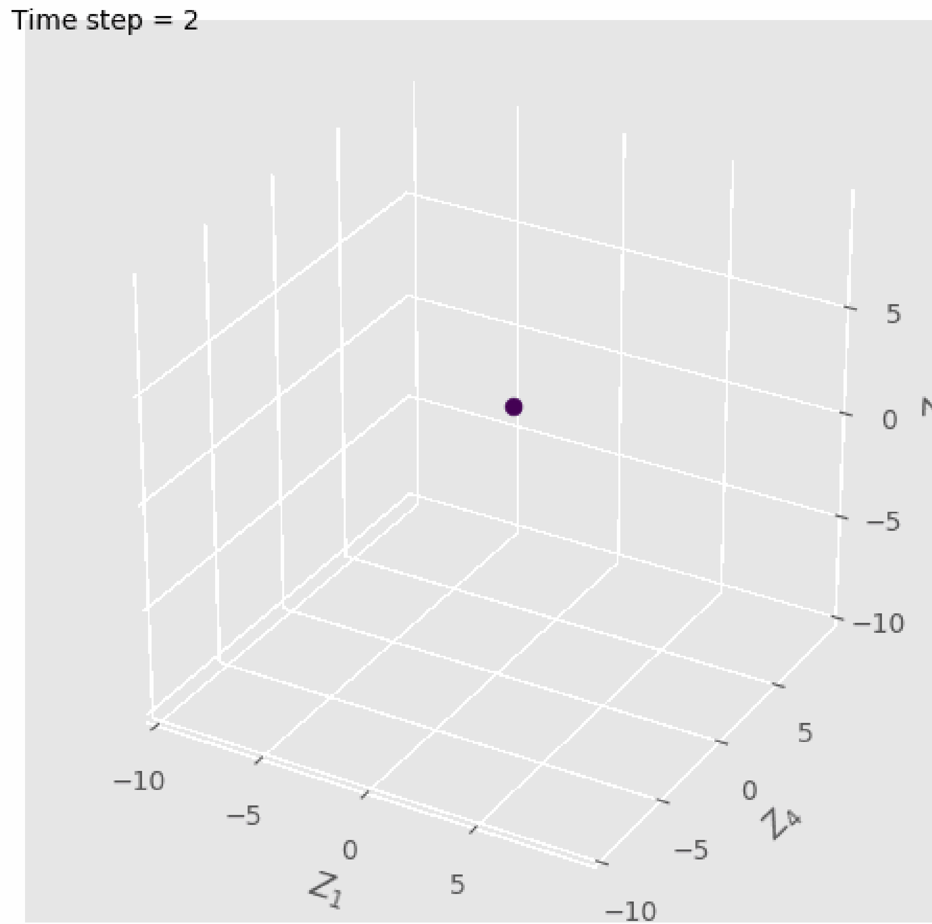


Koppe et al. (2019), *PLoS Comp Biol*

state space

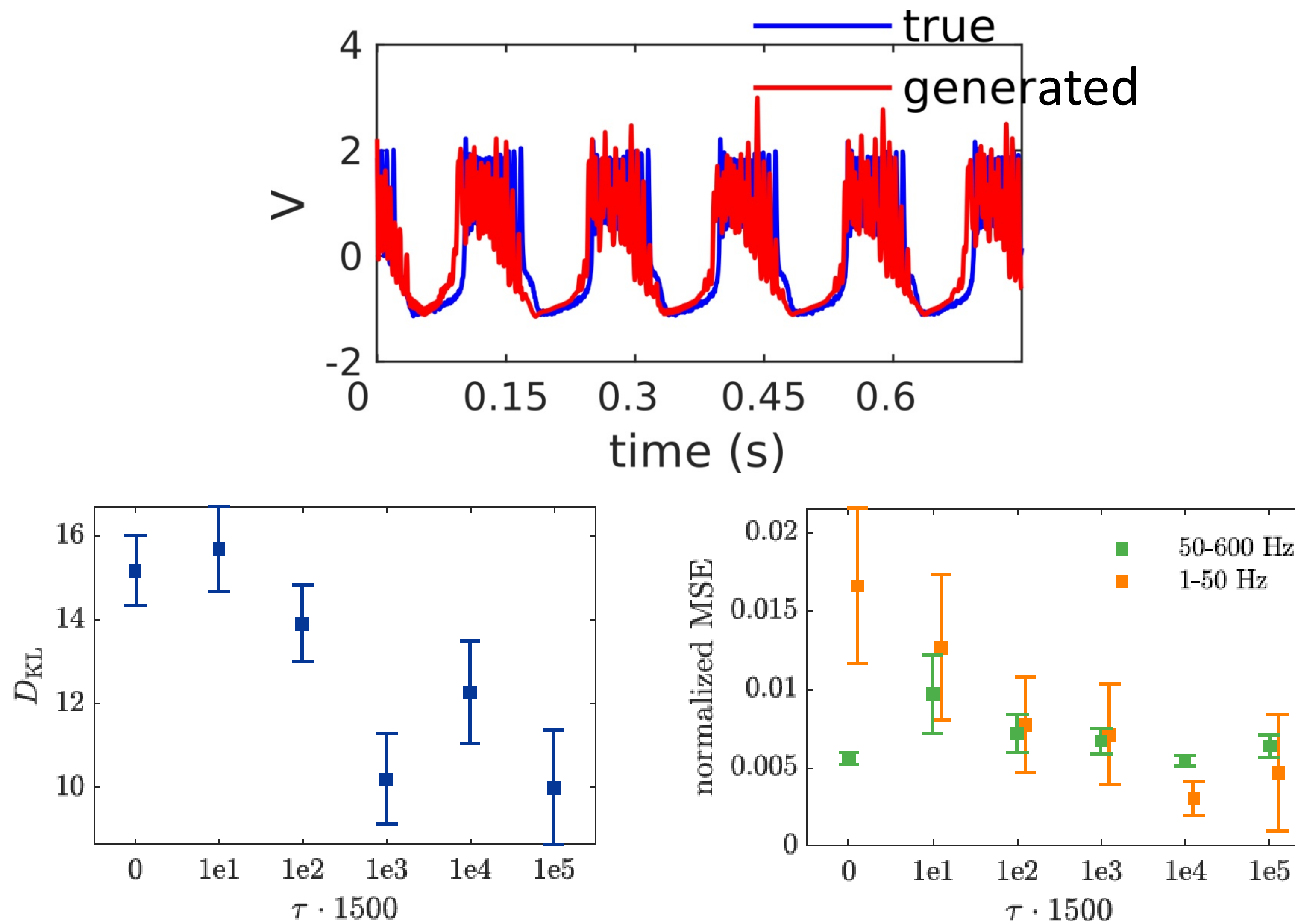


unit activity



Video: Saran Nambusubramaniyan

Enforcing line attractor directions helps to capture multiple time scales



Schmidt et al. (2019) arXiv

(I) Brief primer on DST
→ role in neuroscience & psychiatry

(II) Deep learning of dynamical systems from time series observations

(III) Applications in psychiatry [fMRI, mobile data]

Inferring PLRNN from fMRI data

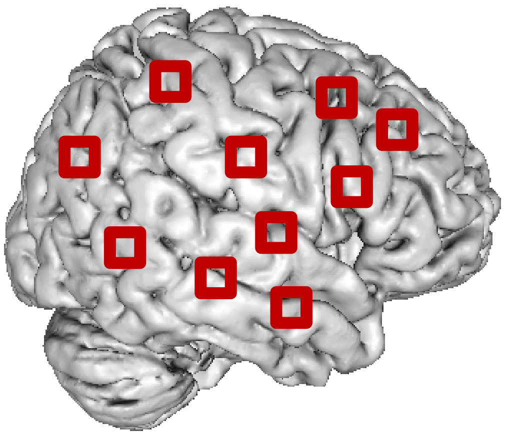
process model

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{W}\phi(\mathbf{z}_{t-1}) + \mathbf{h} + \mathbf{C}\mathbf{u}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

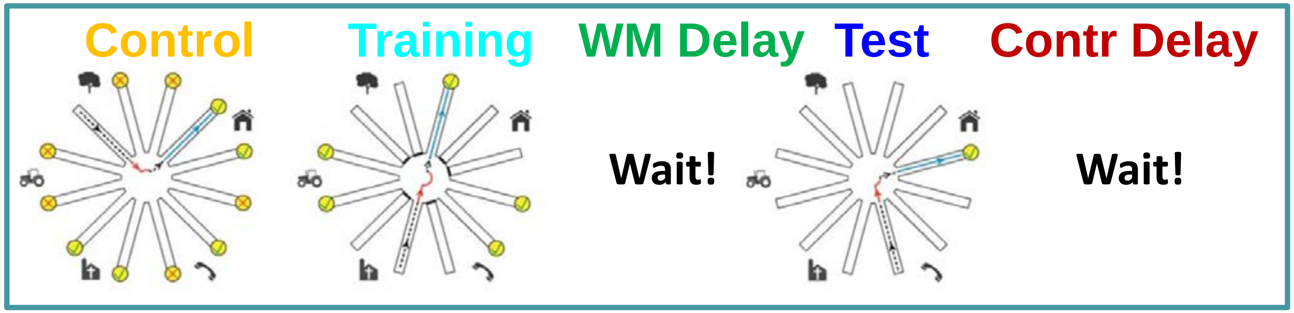
fMRI observation model

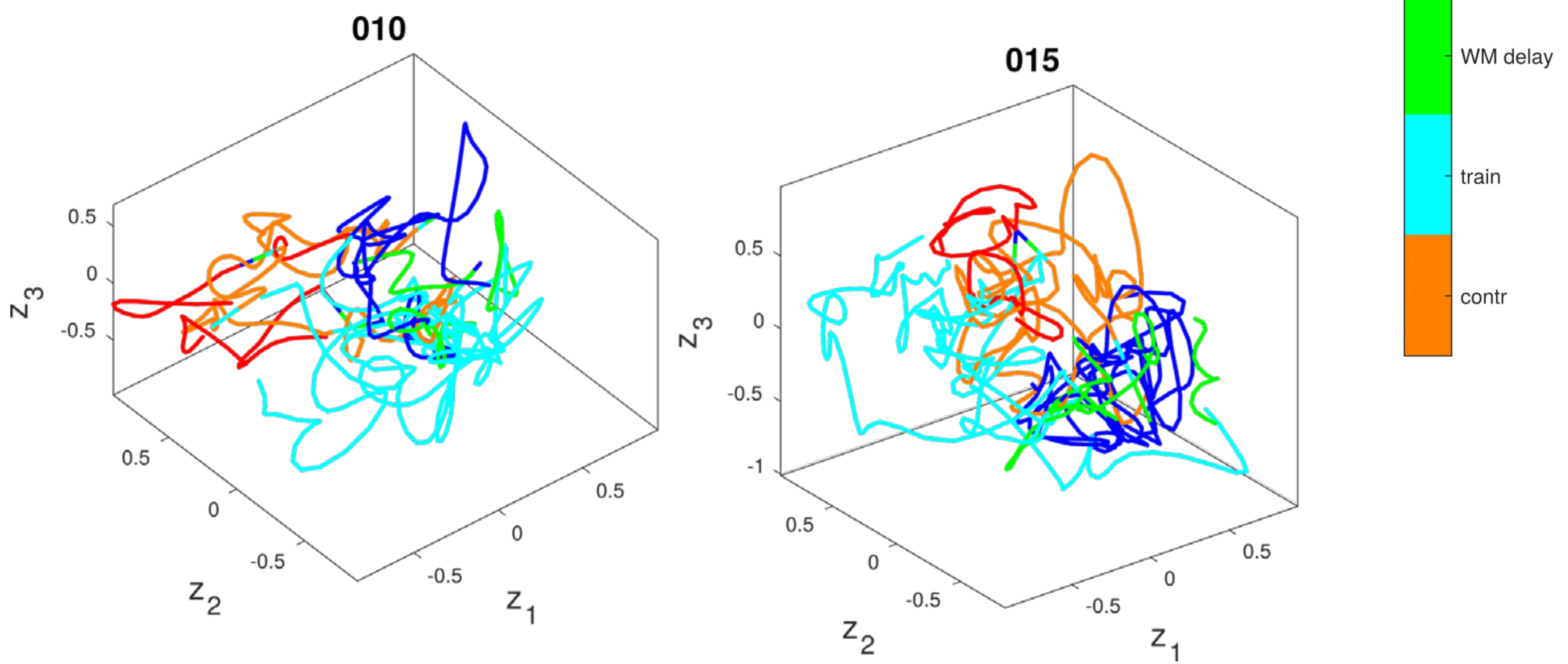
$$\mathbf{x}_t \mid \mathbf{z}_t \sim \mathrm{N}(\mathbf{B}[hrf * \mathbf{z}_{t-\tau:t}] + \mathbf{M}\mathbf{r}_t, \boldsymbol{\Gamma})$$



Koppe et al. (2019), *PLoS Comp Biol*

Reconstructed state space trajectories from fMRI

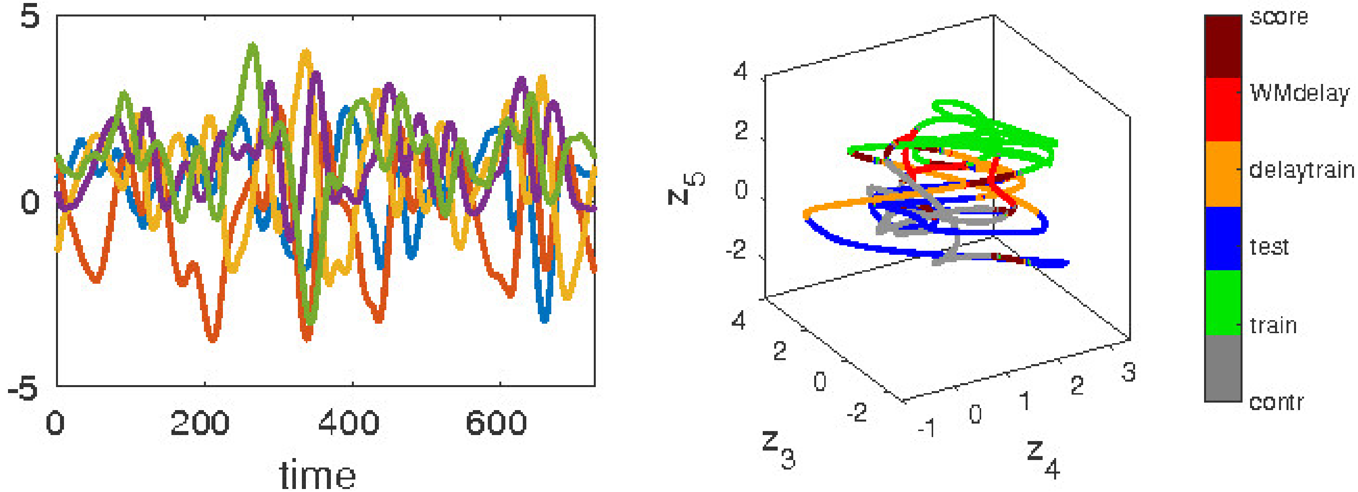




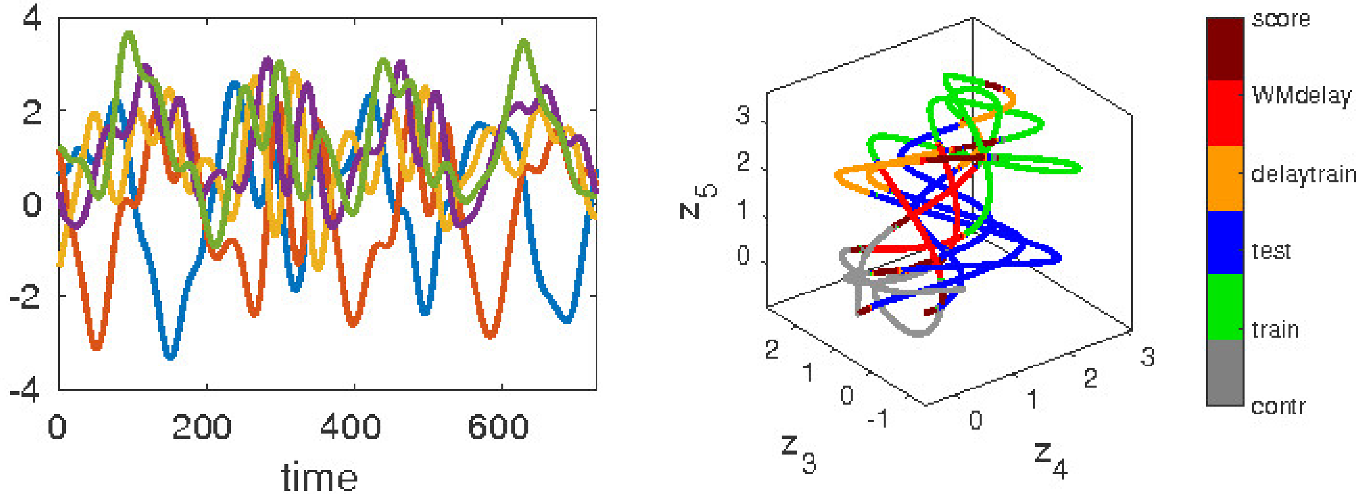
fMRI data: Bähner et al. (2015) ; Koppe et al. (2019), *PLoS Comp Biol*

Does PLRNN really capture measured dynamics?

Inferred (posterior) state estimates $p_\theta(\mathbf{z}_t \mid \mathbf{x}_{1:T})$



Generated (prior) state trajectories $p_\theta(\mathbf{z}_t)$



Does PLRNN really capture measured dynamics?

$$KL_{\mathbf{z}}\left(p_{inf}(\mathbf{z}|\mathbf{x}), p_{gen}(\mathbf{z})\right) = \int_{\mathbf{z}\in\mathbb{R}^M} p_{inf}(\mathbf{z}|\mathbf{x}) \log \frac{p_{inf}(\mathbf{z}|\mathbf{x})}{p_{gen}(\mathbf{z})} \mathrm{d}\mathbf{z}$$

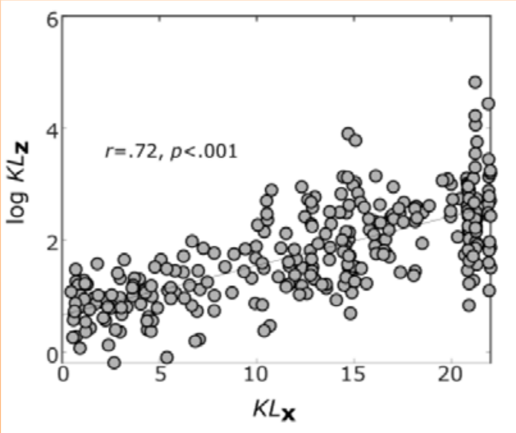
$$p_{inf}(\mathbf{z}|\mathbf{x}) \approx \frac{1}{T}\sum_{t=1}^{T} p(\mathbf{z}_t|\mathbf{x}_{1:T})$$

$$p_{gen}(\mathbf{z}) \approx \frac{1}{T}\sum_{t=1}^{T} p(\mathbf{z}_t|\mathbf{z}_{t-1})$$

Gaussian Mixture Models

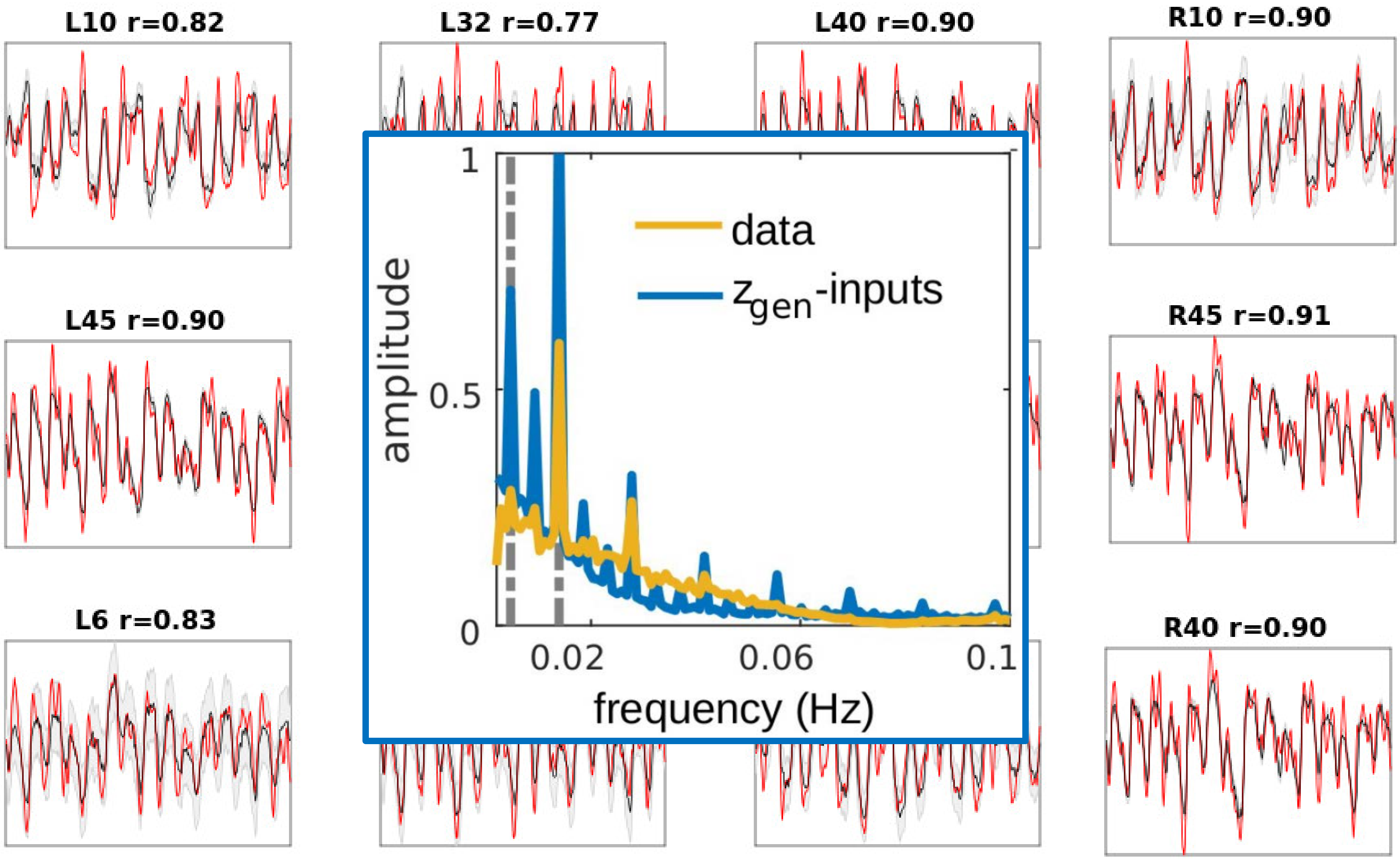
$$KL_{\mathbf{z}}^{(variational)}\left(p_{inf}(\mathbf{z}|\mathbf{x}), p_{gen}(\mathbf{z})\right) \approx \frac{1}{T}\sum_{t=1}^{T} \log \frac{\sum_{j=1}^{T} e^{-KL\left(p_{inf}(\mathbf{z}_t|\mathbf{x}_{1:T}), p_{inf}(\mathbf{z}_j|\mathbf{x}_{1:T})\right)}}{\sum_{k=1}^{T} e^{-KL\left(p_{inf}(\mathbf{z}_t|\mathbf{x}_{1:T}), p_{gen}(\mathbf{z}_k|\mathbf{z}_{k-1})\right)}}$$

Hershey & Olsen (2007), *IEEE*



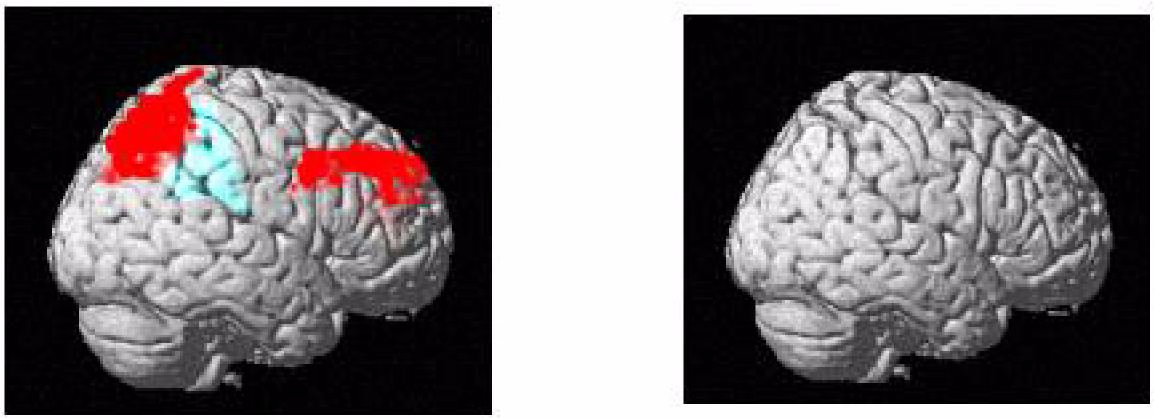
Koppe et al. (2019), *PLoS Comp Biol*

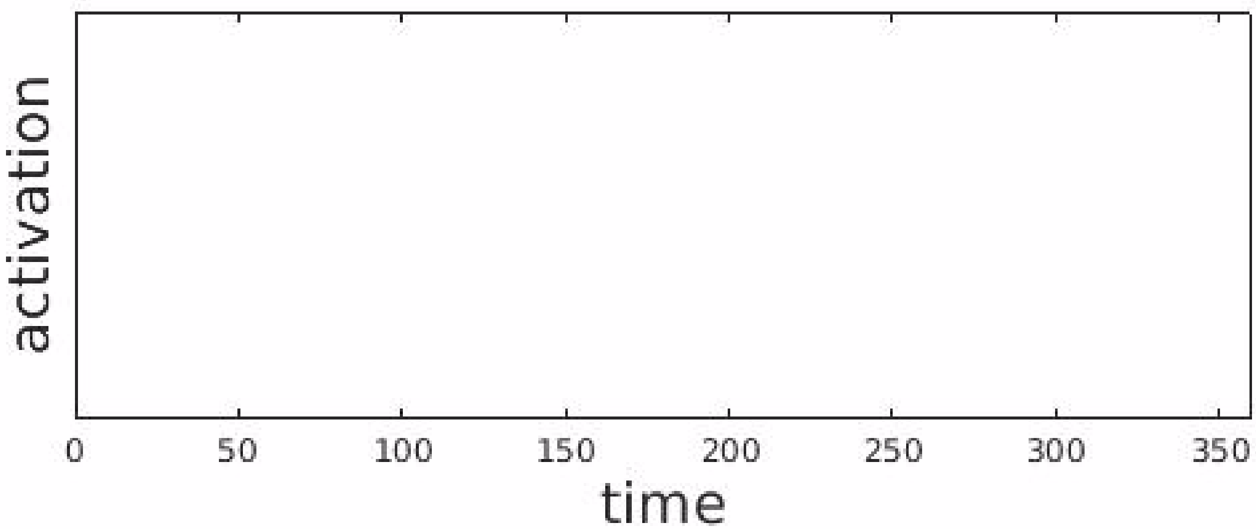
Observed and predicted BOLD traces



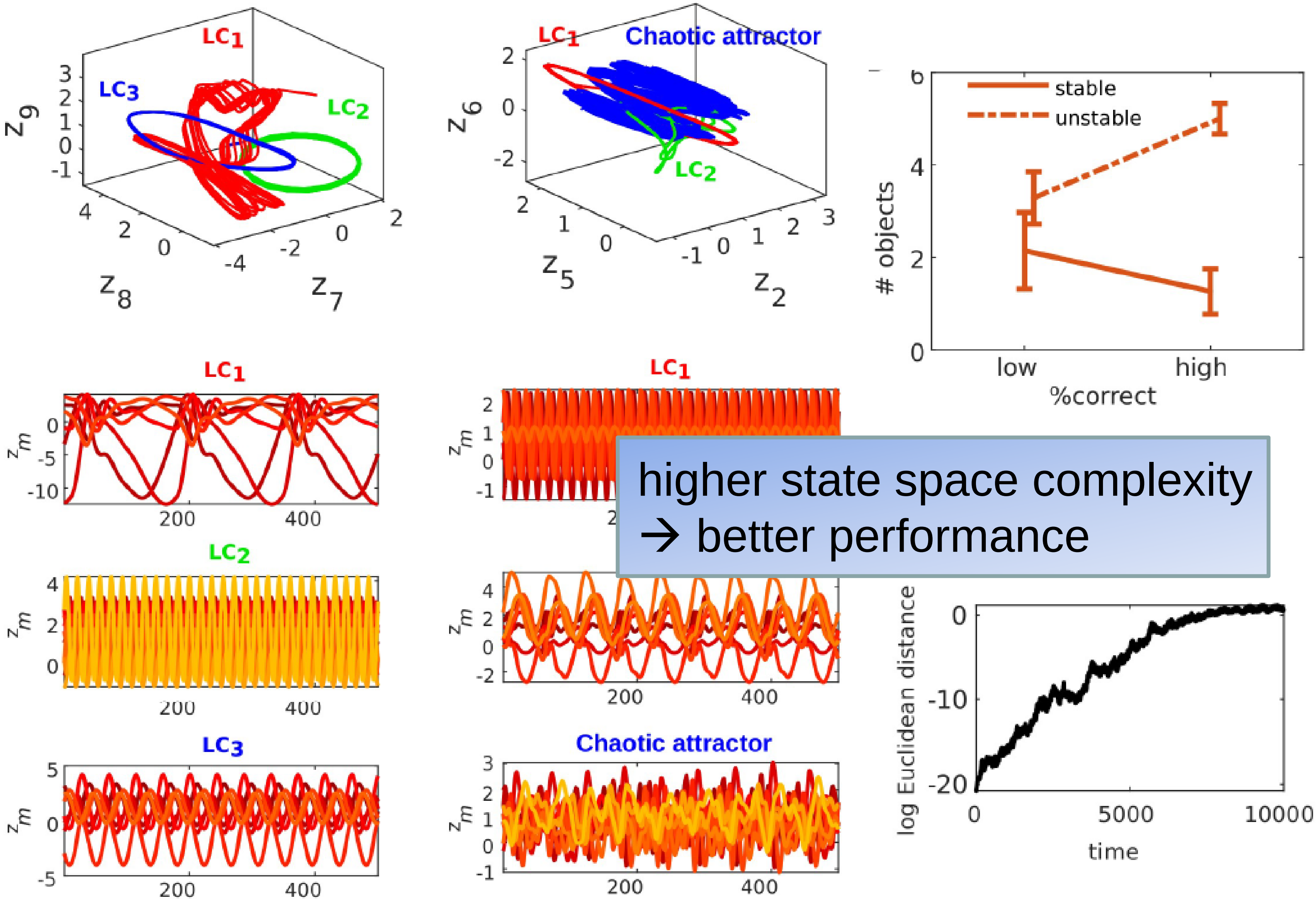
Koppe et al. (2019), *PLoS Comp Biol*

The fMRI Turing-Test





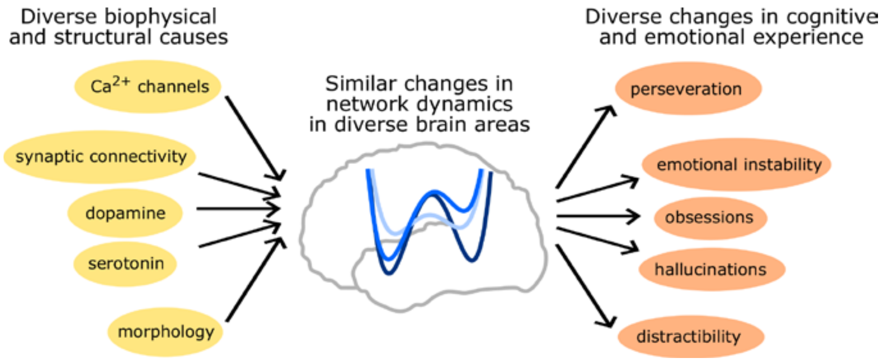
Examples: Nonlinear phenomena from fMRI data

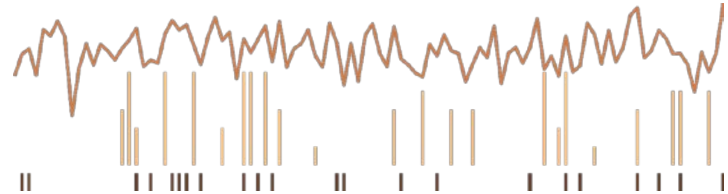


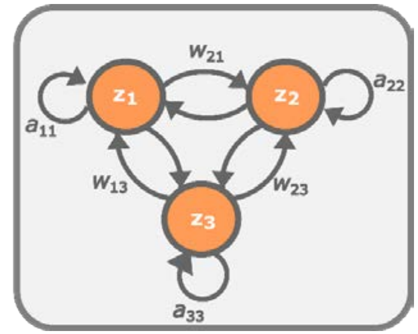
higher state space complexity
→ better performance

Koppe et al. (2019), *PLoS Comp Biol*

Take home's







- Dynamical systems theory = central layer for connecting neural and behavioral phenomena → new perspective on psychiatric symptoms and their treatment
- Deep generative RNN → enable reconstruction of DS from observed time series
- ... enables ‚diagnosis' of systems-dynamical states & prediction of individual disease trajectories
- ... enables simulation & prognosis of effects of interventions in personalized fashion

Many thanks for your attention!

Collaborations:
Ulrich Reininghaus
Heike Tost
Florian Bähner
Andreas Meyer-Lindenberg
Ulrich Ebner-Priemer
TRR-265
Chris Lapish
Jeremy Seamans

Bundesministerium für Bildung und Forschung
bccn heidelberg-mannheim
e:Med SYSTEMS MEDICINE
STRUCTURES CLUSTER OF EXCELLENCE
Deutsche Forschungsgemeinschaft DFG