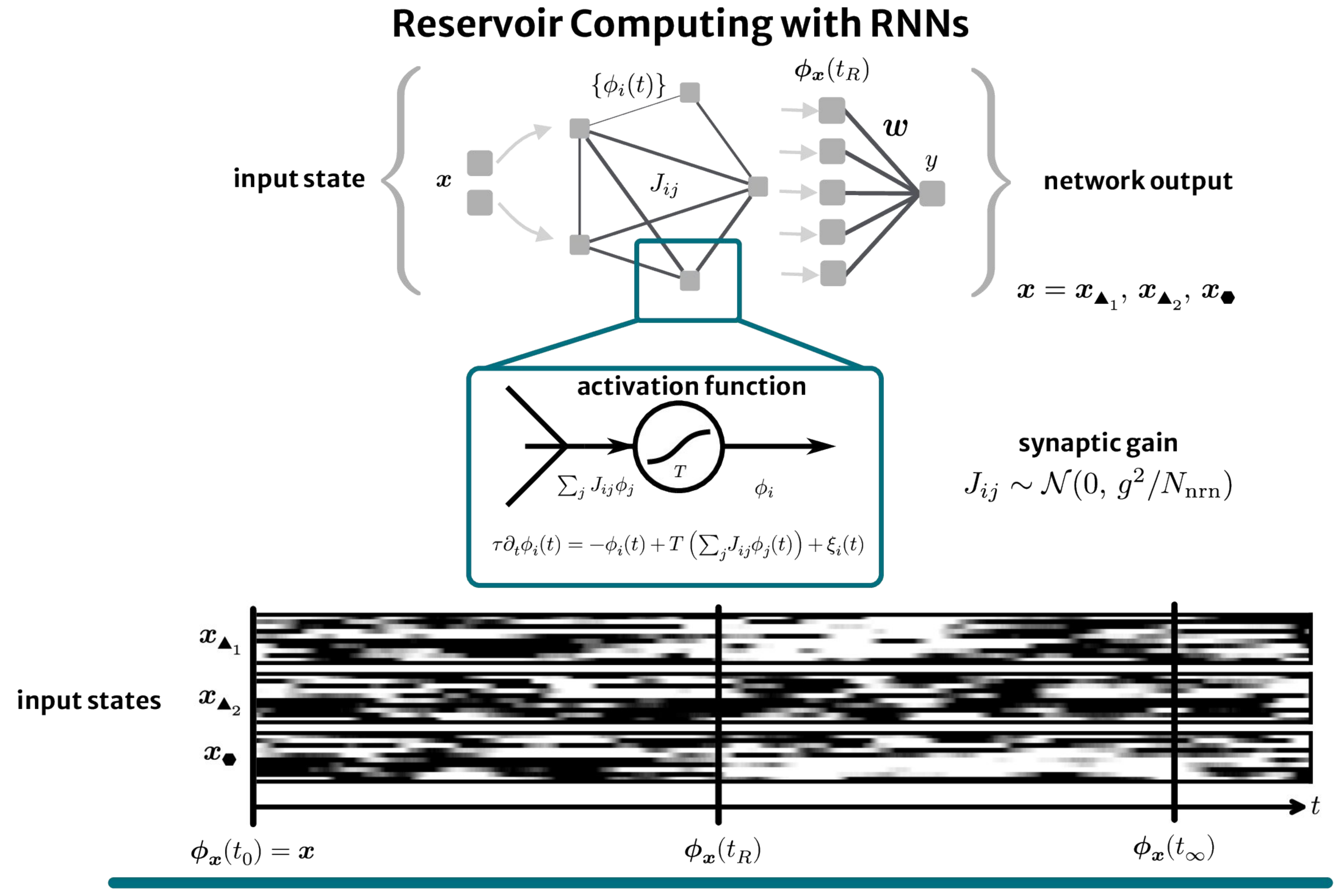
Discrete communication mediates effective regularization in chaotic recurrent networks

Jan Philipp Bauer[1,2] · Jonathan Kadmon[1] · Moritz Helias[2]

[1] The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Israel
[2] Institute of Neuroscience and Medicine (INM-6), Jülich Research Centre, Germany

Reservoir Computing with RNNs



state space picture

supervised learning with ridge regression loss

$$w^* = \mathrm{argmin}_w \mathcal{L}(w) = \sum_{p \in \mathcal{D}_{\mathrm{train}}} \left(Y_{X^p} - w \cdot \phi^J_{X^p}\right)^2$$

find the optimal w^*

$$y_{x^*} = w^* \cdot \phi^J_{x^*} = Y_X^T \Phi_X^J \left(\Phi_X^J \Phi_X^{J\,T}\right)^{-1} \cdot \phi^J_{x^*} = K^J_{x^* X} \left(K^J_{XX}\right)^{-1} Y_X$$

rephrase using kernel matrix

mean-field theory $N_{\mathrm{nrn}} \to \infty$

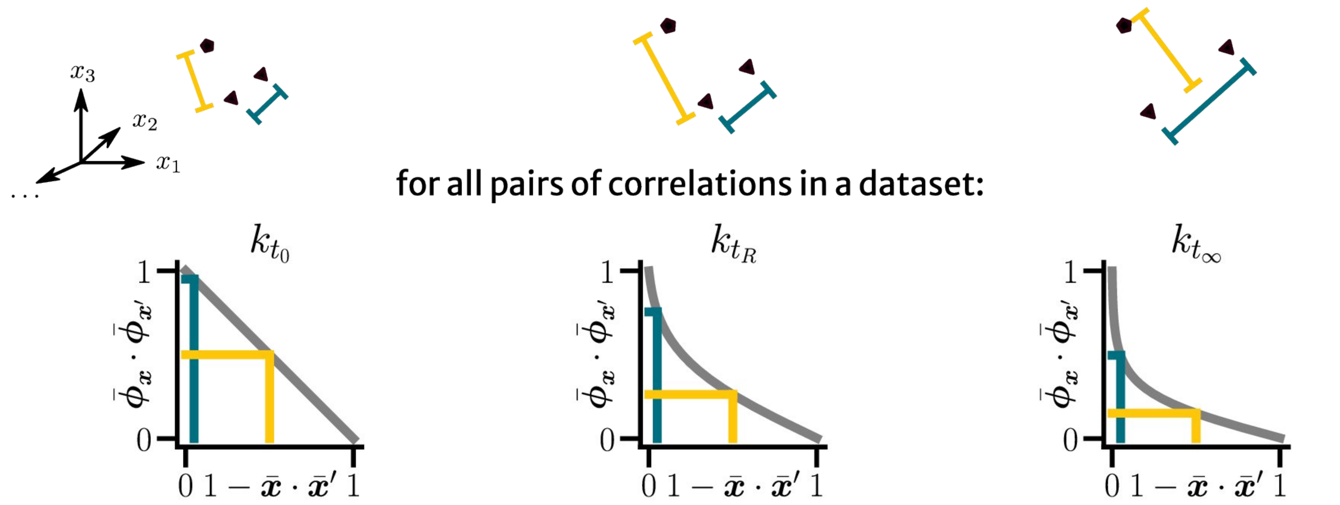
Gaussian process (GP) picture

$$(K^J)_{x,x'} = \langle \phi^J_x \cdot \phi^J_{x'} \rangle_J \overset{N_{\mathrm{nrn}} \to \infty}{=} \langle \phi\, \phi' \rangle =: k_t(x \cdot x')$$

→ independence of microscopic details J by **disorder average**

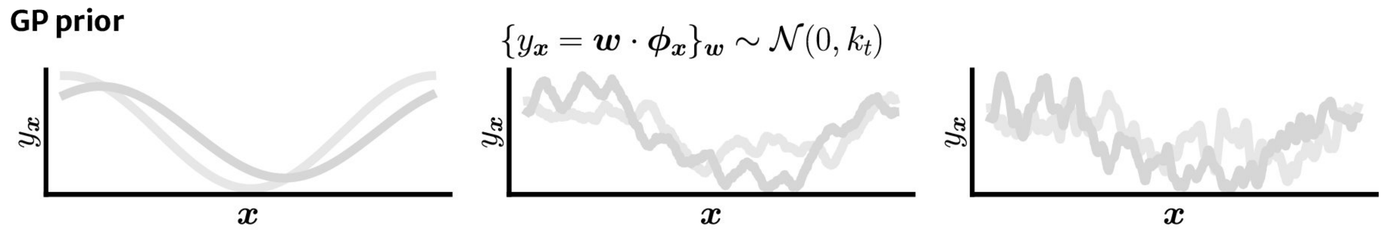
This **kernel function** k captures how **input correlations** $x \cdot x'$ map to **output correlations** $\phi_x \cdot \phi_{x'}$

for all pairs of correlations in a dataset:



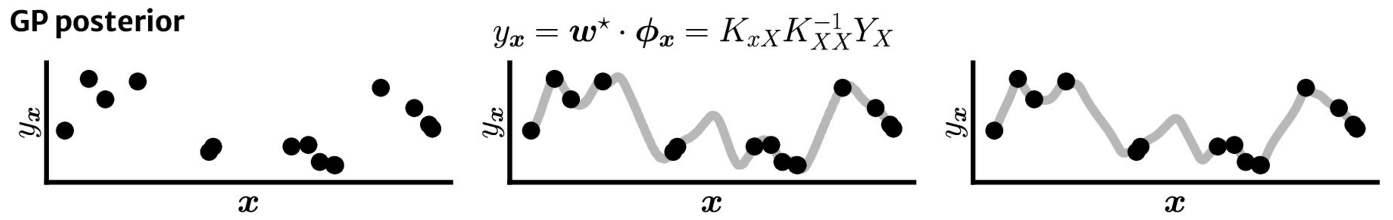
With it, we can quantify the roughness of the network output before

GP prior $\{y_x = w \cdot \phi_x\}_w \sim \mathcal{N}(0, k_t)$



and after training of w^* via **Bayesian inference**:

GP posterior $y_x = w^* \cdot \phi_x = K_{xX} K_{XX}^{-1} Y_X$



Summary

Disordered networks with discrete signaling are considered a poor substrate for computation, yet they are ubiquitous in the brain.

We show that such **large chaotic** networks can support **reliable computation**, with a surprisingly **long working memory**. To this end, we reformulate the recurrent network's activity in terms of an effective kernel.

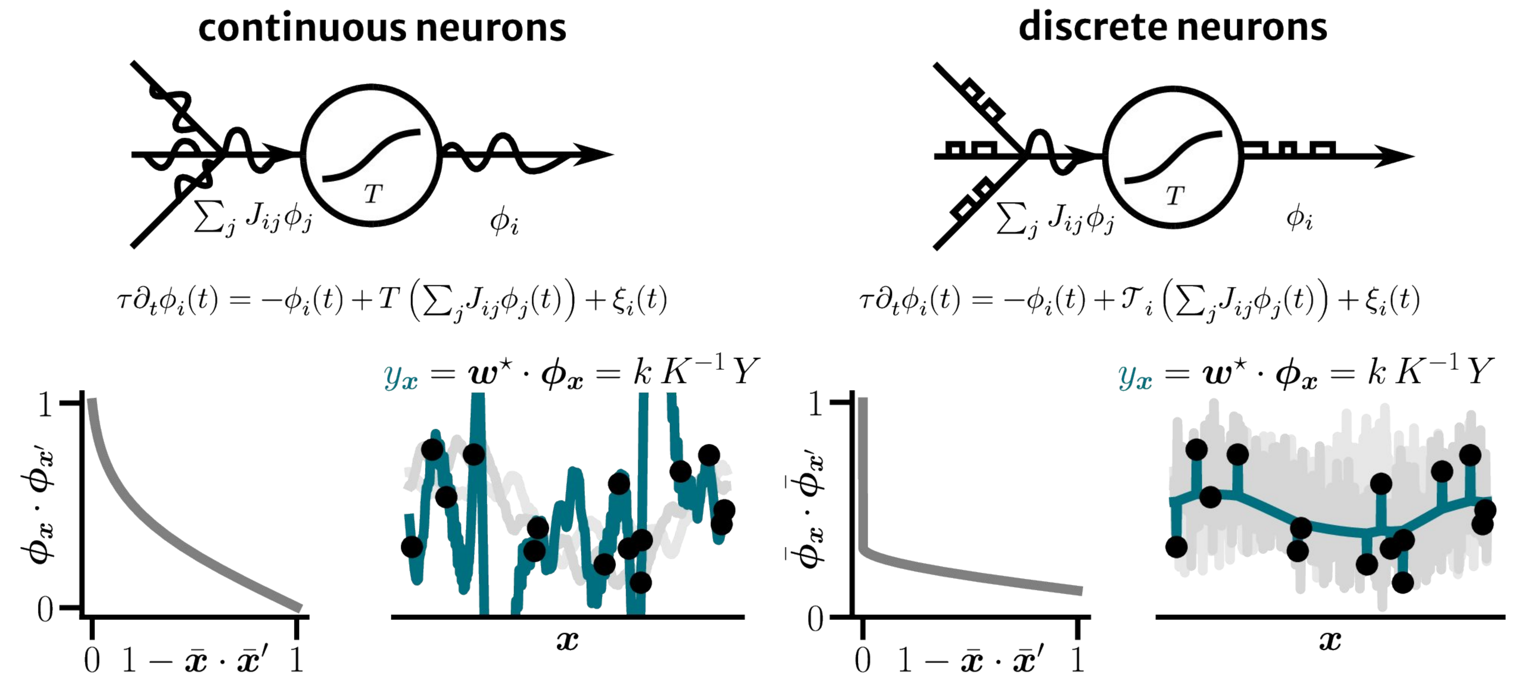
In particular, we find that

- chaos in a discrete signaling network acts as an effective regularizer,
- rich and robust computation is possible in the chaotic regime and
- at the edge of chaos, reliable computation persists even longer.

References

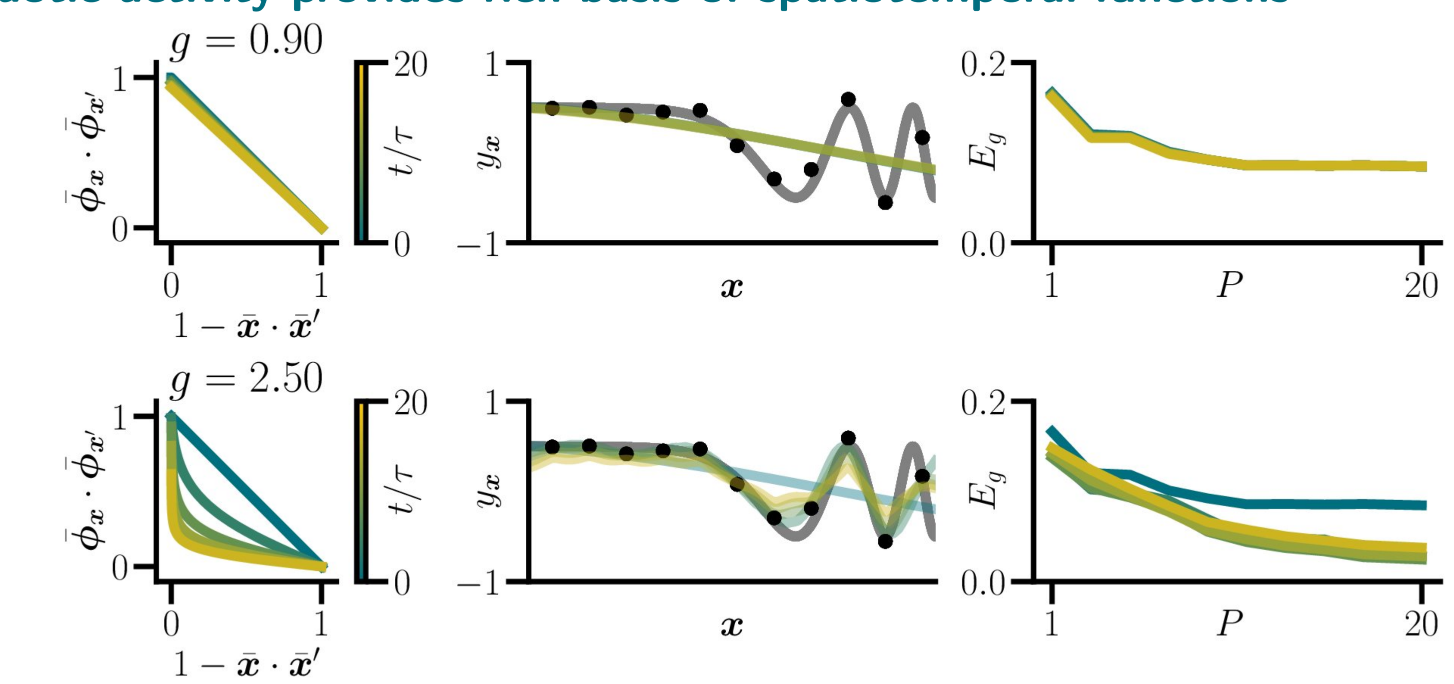
[1] C Keup, M Helias et al. (2021). "Transient Chaotic Dimensionality Expansion by Recurrent Networks". Physical Review X.
[2] J Kadmon and H Sompolinsky (2015). "Transition to chaos in random neuronal networks". Physical Review X.
[3] C Stringer, K Harris et al. (2019). "High-dimensional geometry of population responses in visual cortex". Nature.
[4] B Bordelon and C Pehlevan (2022). "Population codes enable learning from few examples by shaping inductive bias". eLife.

Discrete signaling is an effective regularizer



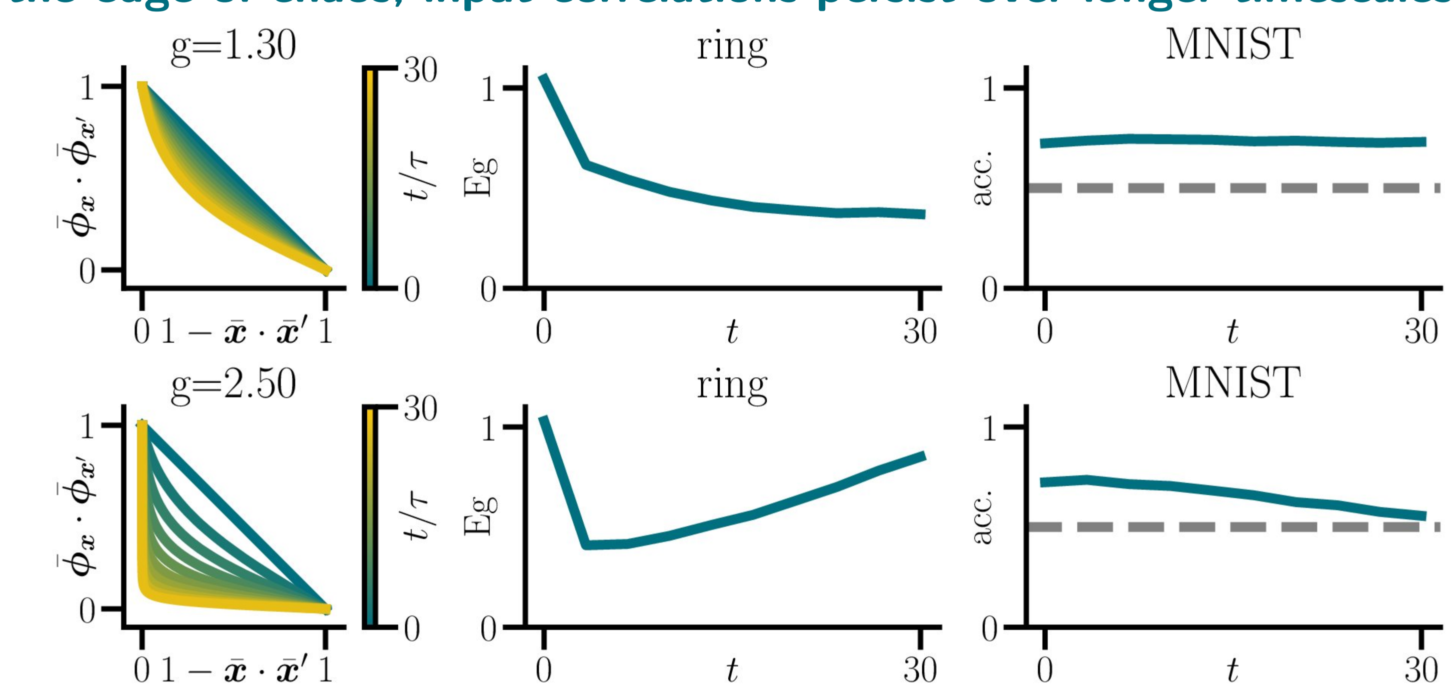
- Discrete signaling leads to a **sharp drop** in the kernel function.
- This strong chaos makes trajectories diverge **rapidly, but deterministically.**
- As a result, noisy observations are **regularized via benign overfitting.**

Chaotic activity provides rich basis of spatiotemporal functions



- Sub-critical networks have mostly **linear** kernels.
- Extrapolation in **regression tasks** is hence limited to **slowly-changing** functions.
- In contrast, chaotic networks have a **richer inductive bias.**

At the edge of chaos, input correlations persist over longer timescales



- At the edge of chaos, neural networks show **critical slowing down.**
- Similarly, **input correlations also persist** on longer timescales.
- This allows the network to **maintain accuracy** on generalization tasks.

Contact: j.bauer@mail.huji.ac.il