Sufficient conditions for offline reactivation in recurrent neural networks

Nanda H Krishna[1,2], Colin Bredenberg[1,2], Daniel Levenstein[1,3], Blake Aaron Richards[1,3], Guillaume Lajoie[1,2]

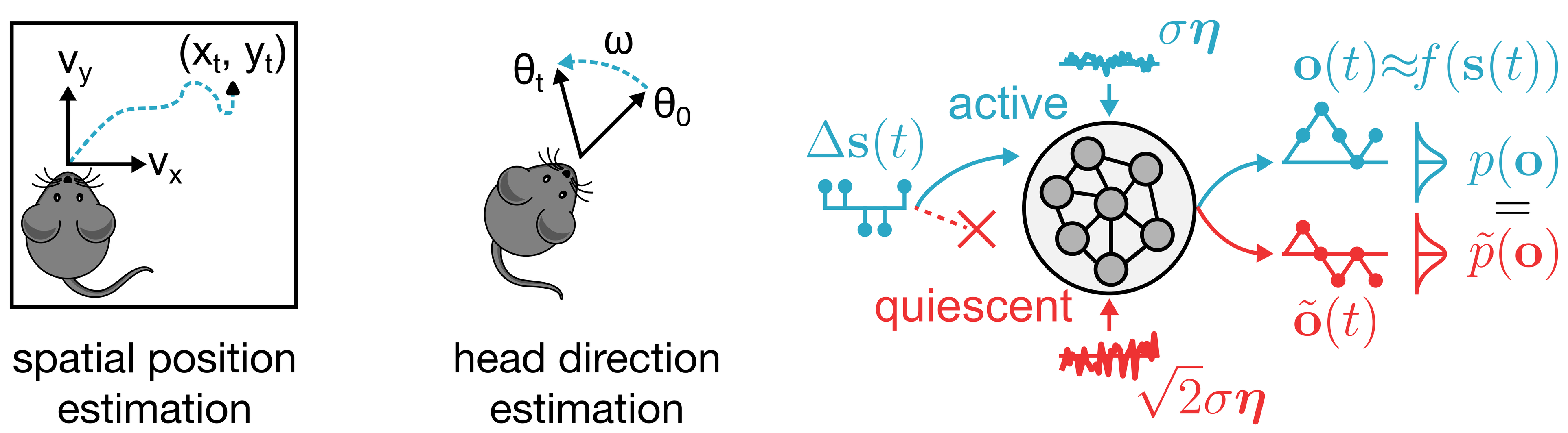
[1]Mila – Quebec AI Institute, [2]Université de Montréal, [3]McGill University, Correspondence: {nanda.harishankar-krishna,colin.bredenberg}@mila.quebec

Introduction

Neural circuits recapitulate task-like activity[1] during quiescence.

This reactivation has many functional benefits in neural networks[2,3].

Key question: How does offline reactivation emerge?



We show that diffusive reactivation can emerge as a consequence of optimally encoding environmental variables in the presence of noise!

Greedily Optimal Dynamics

Conditions – the recurrent neural network must:

1. Implement a **continuous-time dynamical system**.
2. Perform **noisy path integration** using change-based inputs.
3. Estimate state variables **near-optimally**.

$$\mathcal{L}(t+\Delta t) = \mathbb{E}_{\eta} \left\| f(\mathbf{s}(t+\Delta t)) - \mathbf{D}\big(\mathbf{r}(t) + \overbrace{\Delta\mathbf{r}_1(t) + \Delta\mathbf{r}_2(t) + \sigma\boldsymbol{\eta}(t)}^{\Delta\mathbf{r}(t)}\big) \right\|_2$$

$$\approx \mathbb{E}_{\eta} \left\| \frac{\mathrm{d}f(\mathbf{s}(t))}{\mathrm{d}\mathbf{s}(t)} \Delta\mathbf{s}(t) - \mathbf{D}\Delta\mathbf{r}_2(t) + f(\mathbf{s}(t)) - \mathbf{D}\big(\mathbf{r}(t) + \Delta\mathbf{r}_1(t) + \sigma\boldsymbol{\eta}(t)\big) \right\|_2$$

$$\mathcal{L}_{\text{upper}} = \left\| \frac{\mathrm{d}f(\mathbf{s}(t))}{\mathrm{d}\mathbf{s}(t)} \Delta\mathbf{s}(t) - \mathbf{D}\Delta\mathbf{r}_2(t) \right\|_2 + \|\mathbf{D}\|_2 \sqrt{\mathbb{E}_{\eta} \|\Delta\mathbf{r}_1(t) + \sigma\boldsymbol{\eta}(t-\Delta t)\|_2^2}$$

The greedily optimal neural dynamics during the active phase are:

$$\boxed{\Delta\mathbf{r}^*(t)} = \Big[\underbrace{\mathbf{D}^{\dagger} \frac{\mathrm{d}f(\mathbf{s}(t))}{\mathrm{d}\mathbf{s}(t)} \frac{\mathrm{d}\mathbf{s}(t)}{\mathrm{d}t}}_{\text{state estimation}} + \underbrace{\sigma^2 \frac{\mathrm{d}}{\mathrm{d}\mathbf{r}(t)} \log p(\mathbf{r}(t))}_{\text{denoising}} \Big] \Delta t + \sigma\boldsymbol{\eta}(t)$$

Emergent Offline Reactivation

The dynamics in the absence of inputs, i.e. during quiescence, are:

$$\boxed{\Delta\tilde{\mathbf{r}}(t)} = \left[\sigma^2 \frac{\mathrm{d}}{\mathrm{d}\mathbf{r}(t)} \log p(\mathbf{r}(t)) \right] \Delta t + \sqrt{2}\sigma\boldsymbol{\eta}(t)$$

This is exactly **Langevin sampling** of $p(\mathbf{r})$!

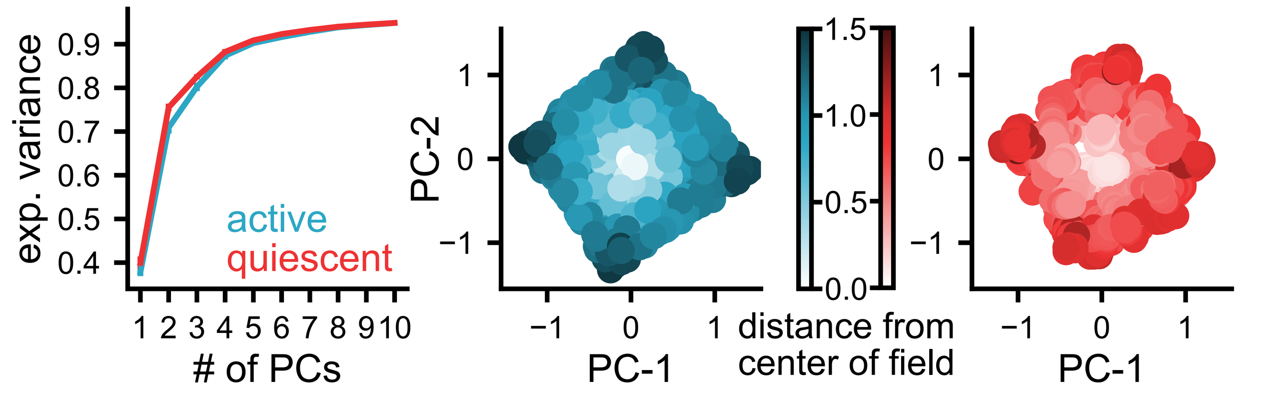
Doubled noise variance is not necessary for reactivation to emerge.

Diffusion along attractor manifold[4] → statistically faithful reactivation

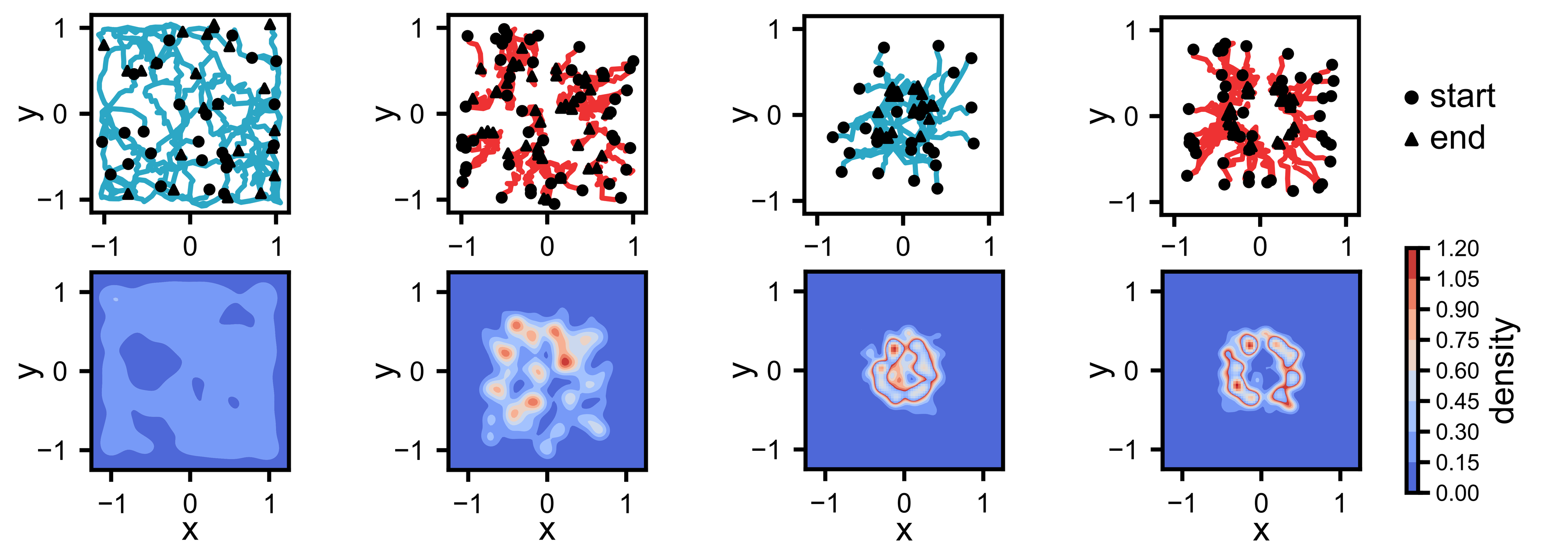
$$p(\mathbf{r}) = \tilde{p}_{t\to\infty}(\mathbf{r}) \;\Rightarrow\; p(\mathbf{o}) = \tilde{p}_{t\to\infty}(\mathbf{o})$$

Reactivation in Spatial Position Estimation

The neural activity manifold is low-D for both the active and quiescent phases.

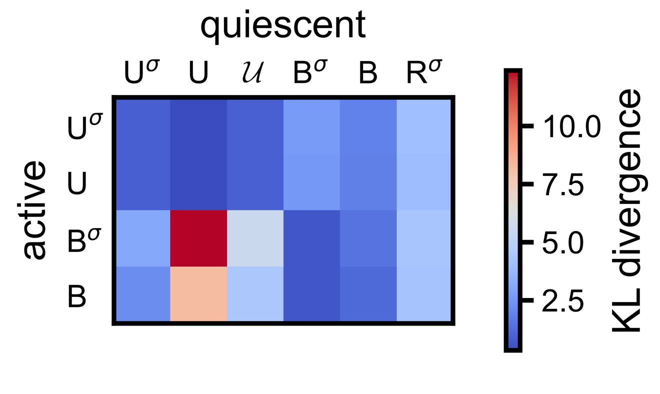


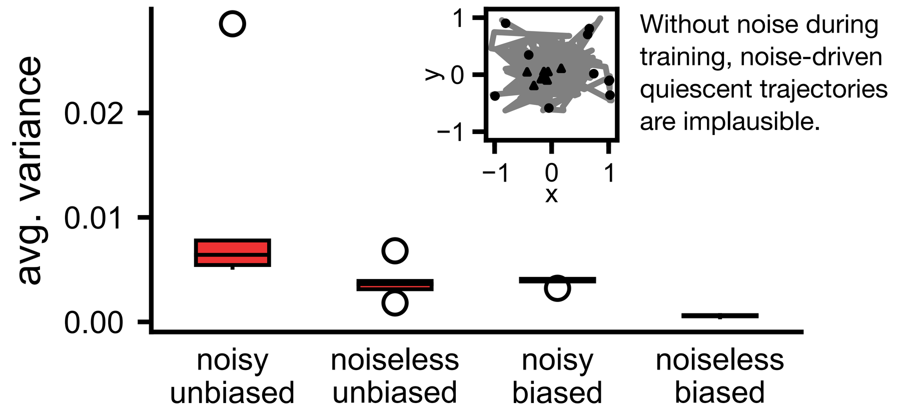
Decoded output trajectories from both phases tile space similarly.



Quiescent distributions resemble corresponding active distributions.

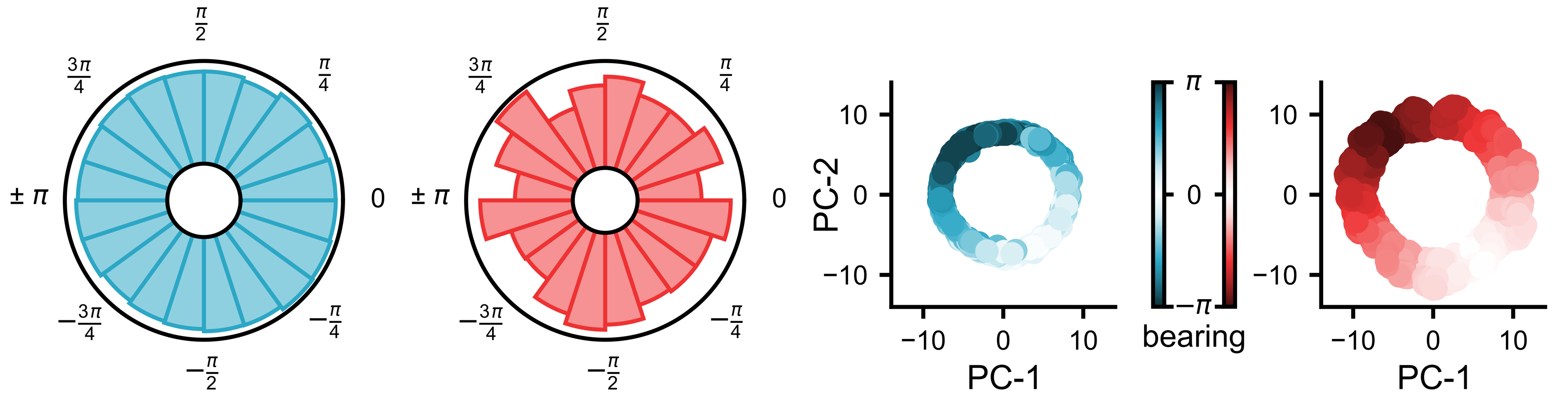
Quiescent trajectories fail to explore in the absence of noise.



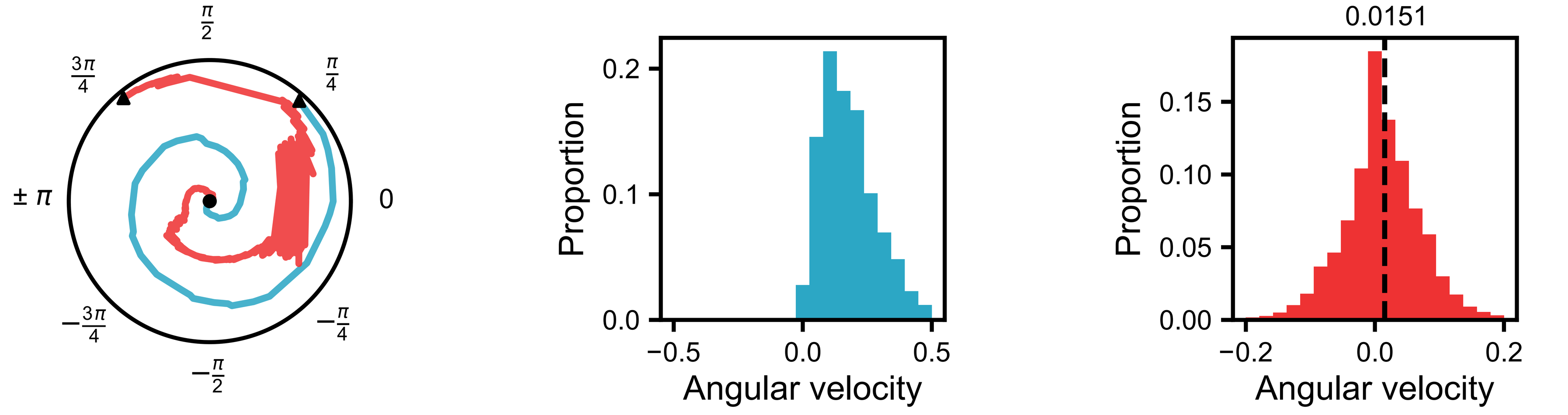


Reactivation in Head Direction Estimation

Distributions are similar across both phases in this task as well.



By biasing motion to be counter-clockwise, we observe that the transition structure is partially recapitulated during quiescence.



Conclusions and Future Work

Reactivaton could be a natural consequence of learning in the presence of noise.

Our results could be extended to explain reactivation in brain regions dedicated to higher-order cognition & decision making[5].

Future work could study the transition dynamics[6] of reactivations & their functional utility, e.g. in memory consolidation[7].

References

[1] Tingley & Peyrache. *Phil. Trans. R. Soc. B* **375** (2020).
[2] Buzsáki. *Neuroscience* **31** (1989).
[3] Mnih et al. *Nature* **518** (2015).
[4] Burak & Fiete. *PLOS Comput. Biol.* **5** (2009).
[5] Peyrache et al. *Nat. Neurosci.* **12** (2009).
[6] Peyrache et al. *Nat. Neurosci.* **18** (2015).
[7] McClelland et al. *Psychol. Rev.* **102** (1995).

UNIQUE CENTRE

IVADO

CIFAR