A THEORY OF MEMORY STABILITY IN HIPPOCAMPUS AREA CA3

Uri Cohen[1], Roland Mason Rodriguez[2], Ole Paulsen[2], Máté Lengyel[1,3]

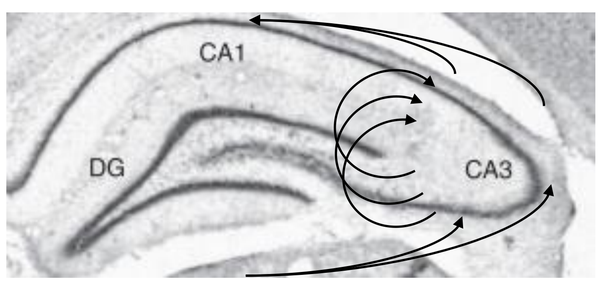
[1] Computational and Biological Learning Lab, Department of Engineering, University of Cambridge; [2] Department of Physiology, Development and Neuroscience, University of Cambridge; [3] Department of Cognitive Science, Central European University

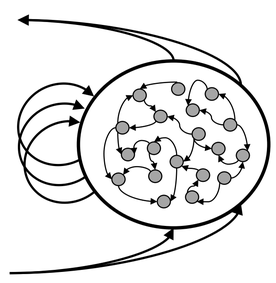
@uricohen42

uc231@cam.ac.uk

Get poster!

Hippocampal area CA3 has been suggested to implement auto-associative memory [1-6].





We provide a new analysis of the conditions for high-capacity storage of stable memories using biologically plausible recall dynamics.

network dynamics

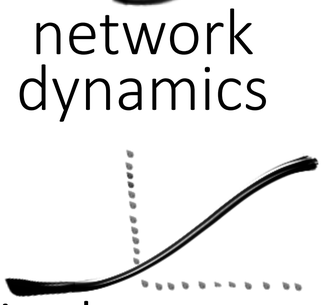
A network with nonlinear voltage dynamics:

$$\dot{\boldsymbol{v}} = -\boldsymbol{v} + W g(\boldsymbol{v}) - \theta$$

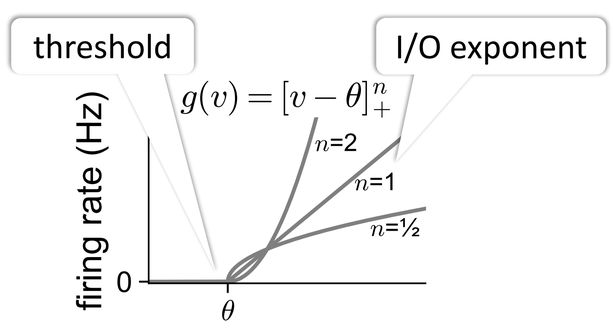
Memories are stable fixed points of the dynamics:

$$\boldsymbol{r}^{\mu} = g(W\boldsymbol{r}^{\mu} - \theta)$$

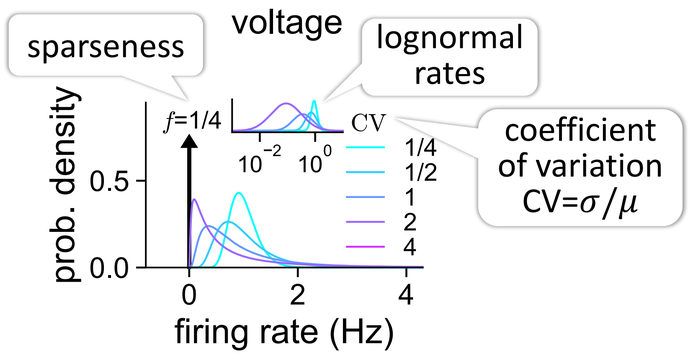
How do single-neuron properties support stable recall of memories?

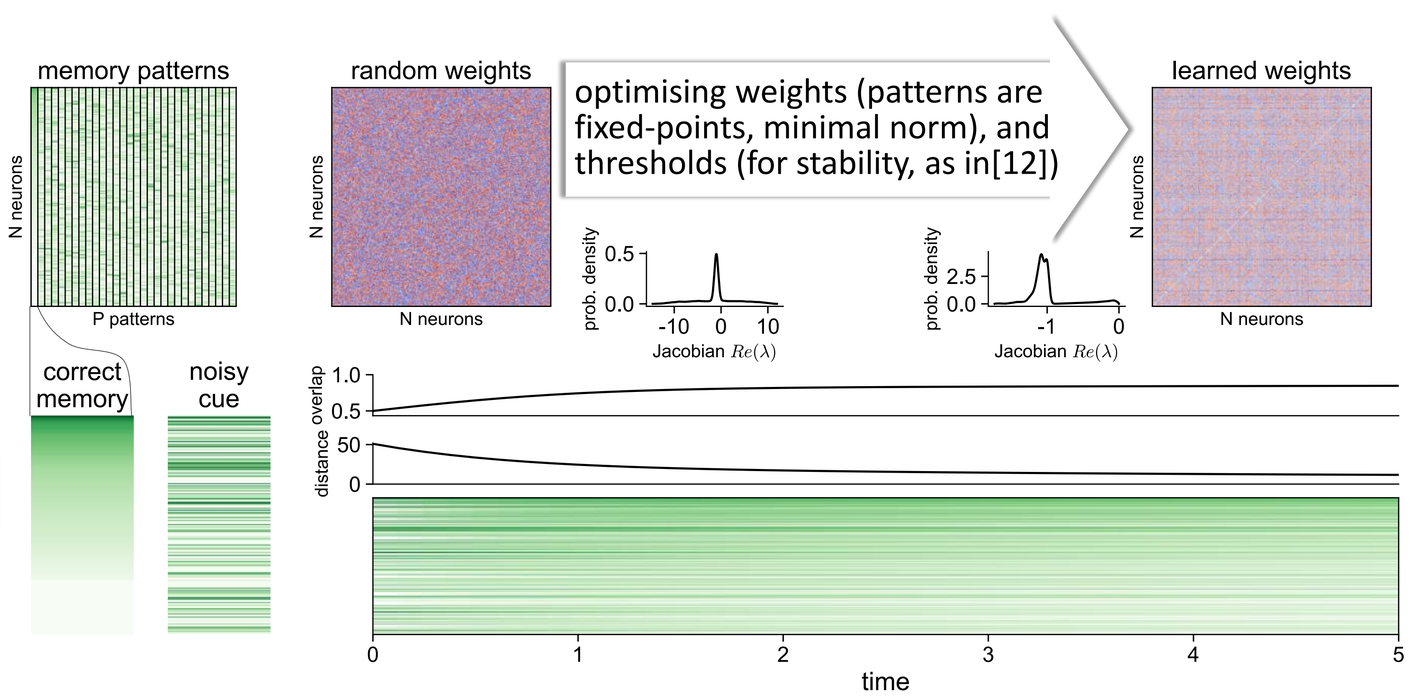


single-neuron nonlinearity



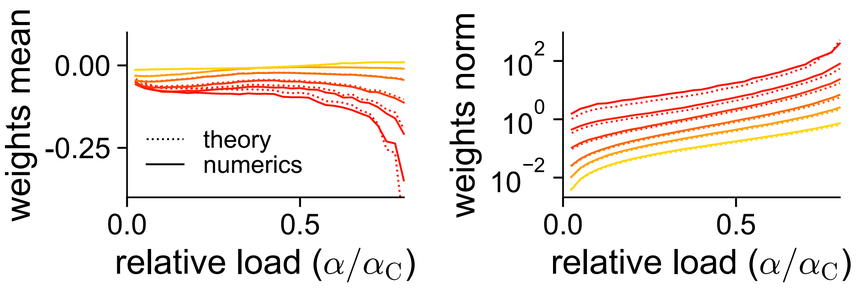
memory patterns





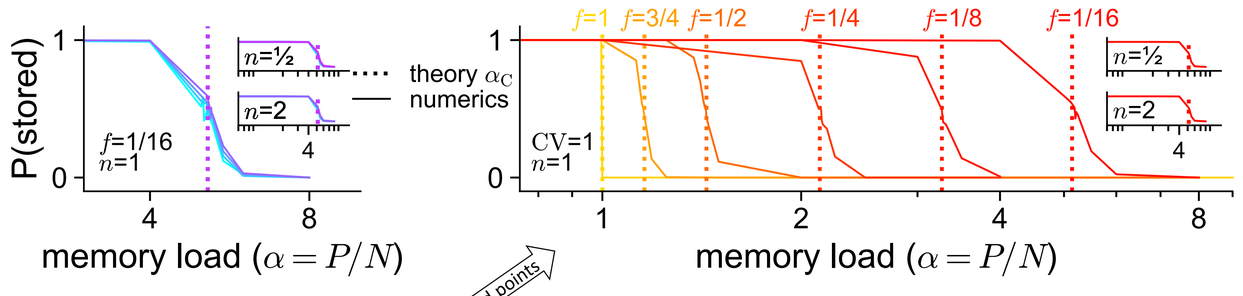
mean-field theory

Mean-field theory (à la [13-15]) fully describes the ability to store patterns as fixed-points.

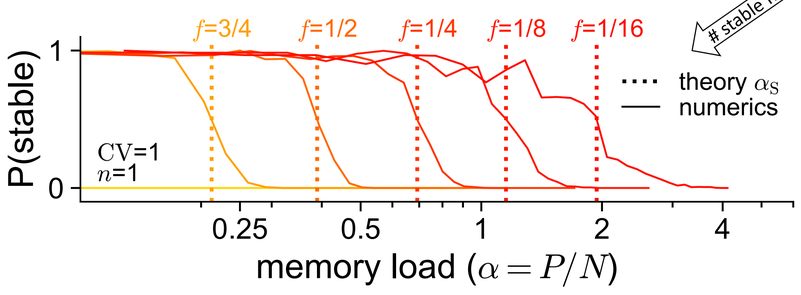


Number of fixed-points is invariant to CV and single-neuron non-linearity, depends only on sparseness (extending [15]).

Non-trivial stable memories regime; e.g., no stable fixed points for $n \geq 1$ and $\theta \geq 0$.



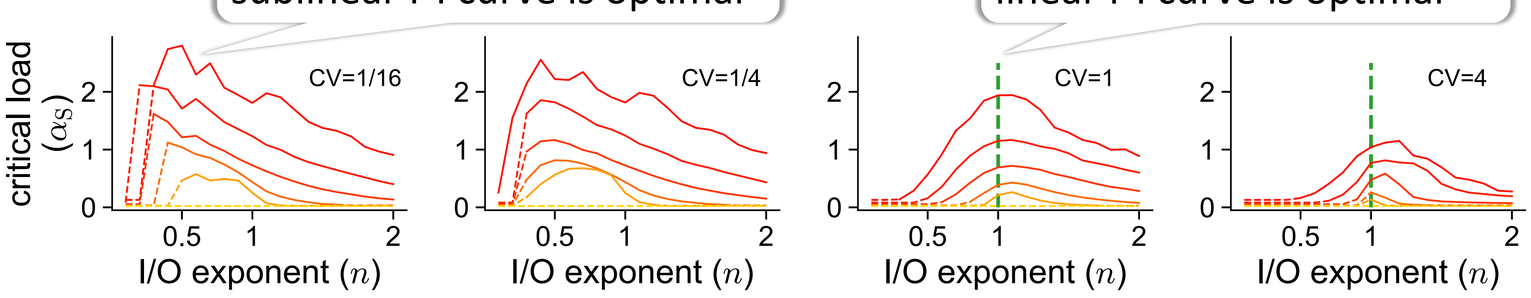
In the stable regime, memories are all stable up to a critical load.



stable fixed points ≠ # fixed points

At low firing rate CV, a sublinear f-I curve is optimal

At high firing rate CV, a linear f-I curve is optimal



technical novelty

The Hopfield model [7] and various extensions [8-11] all use *gradient-dynamics* to store stable memories in attractor networks: $\dot{r} \propto -\nabla E$

- ✓ stable low energy memories
- ✗ non-biological dynamics

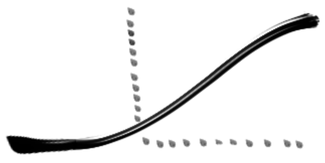
Our attractor networks has:

- ✓ biological plausible dynamics
- ✗ possibly unstable memories
- ✓ the regime where memories tend to be all stable provides new experimentally testable predictions

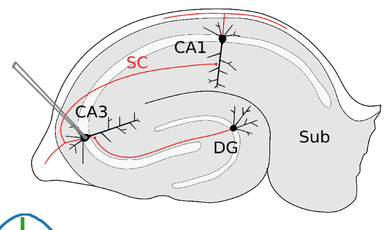
experimental data analysis

Due to its high-CV memory patterns, CA3 pyramidal neurons should have near-linear f-I curves, and negative threshold (i.e., should spontaneously fire if recurrent connections are suppressed)

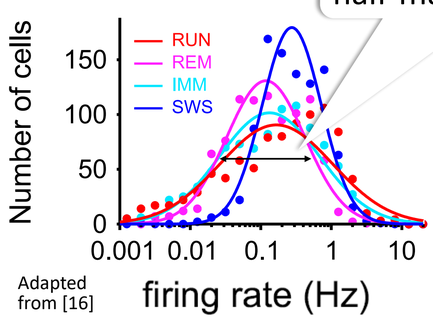
analysis of place cell activity in CA3 of freely moving rates [16-17]

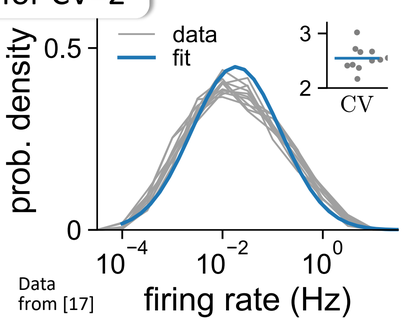


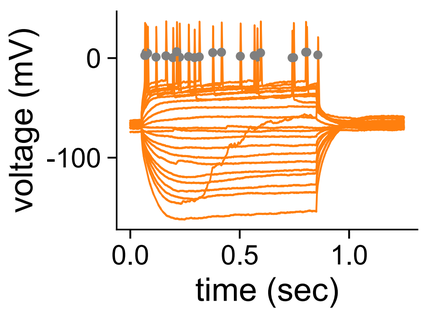
Recording activity from mice CA3 pyramidal neurons in in-vitro whole-cell experiments, we find near-linear f-I curves, and see similar results in existing in-vivo data [18].

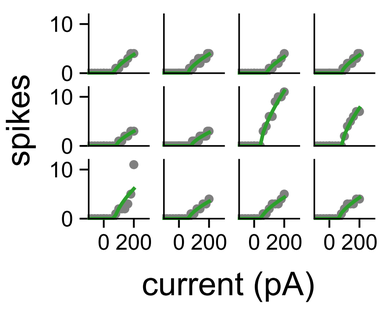


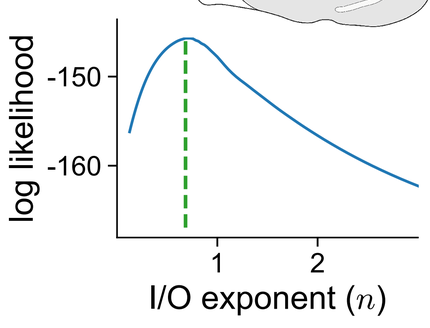
Expected full-width half-maximum for CV=2











References: **[1]** Marr D (1971) Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. 262:23-81. **[2]** McNaughton BL, Morris RG (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends Neurosci. 10:408-15. **[3]** Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. Hippocampus 4:374-91. **[4]** Nakazawa K, Quirk MC, Chitwood RA, Watanabe M, Yeckel MF, Sun LD., … Tonegawa S (2002) Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science, 297:211-8. **[5]** Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. Science 308:873-6. **[6]** Khona M, Fiete IR (2022) Attractor and integrator networks in the brain. Nat. Rev. Neurosci. 23:744-66. **[7]** Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79:2554-8. **[8]** Tsodyks MV, Feigel'man MV (1988) The enhanced storage capacity in neural networks with low activity level. EPL 6:101. **[9]** Treves A (1990) Graded-response neurons and information encodings in autoassociative memories. Phys. Rev. A 42:2418. **[10]** Krotov D, Hopfield JJ (2016) Dense associative memory for pattern recognition. Adv. Neural Inf. Process. 29. **[11]** Lengyel M, Kwag J, Paulsen O, Dayan P (2005) Matching storage and recall: hippocampal spike timing-dependent plasticity and phase response curves. Nat. Neurosci., 8:1677-83. **[12]** Festa D, Hennequin G, Lengyel M (2014) Analog memories in a balanced rate-based network of EI neurons. Adv. Neural Inf. Process. 27. **[13]** Gardner E (1988) The space of interactions in neural network models. J. Phys. A Math. Theor. 21:257. **[14]** Amit DJ, Gutfreund H, Sompolinsky H (1985) Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys. Rev. Lett. 55:1530. **[15]** Schönsberg F, Roudi Y, Treves, A (2021) Efficiency of local learning rules in threshold-linear associative networks. Phys. Rev. Lett. 126:018301. **[16]** Mizuseki K, Buzsáki G (2013). Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. Cell reports, 4:1010-21. **[17]** Alme CB, Miao C, Jezek K, Treves A, Moser EI, Moser MB (2014). Place cells in the hippocampus: eleven maps for eleven rooms. Proc. Natl. Acad. Sci. USA, 111:18428-35. **[18]** Malezieux M, Kees AL, Mulle C (2020). Theta oscillations coincide with sustained hyperpolarization in CA3 pyramidal cells, underlying decreased firing. Cell Reports 32.

wellcome

BLAVATNIK CAMBRIDGE FELLOWSHIPS

BRITISH COUNCIL