CRISP: Challenging the Standard Framework of Hippocampal Memory Function

Laurenz Wiskott
Mehdi Bayati  Sen Cheng  Jan Melchior  Torsten Neher

supported by:
Deutsche Forschungsgemeinschaft - Sonderforschungsbereich 874
Theories of the Hippocampus
Entorhinal cortex (EC) serves as an interface between association areas of neocortex and the hippocampus.

The subareas of the hippocampus are connected in a loop: EC - DG - CA3 - CA1 - subiculum - EC.

Because of its recurrent connectivity, CA3 serves as the central autoassociative memory.

Dentate gyrus (DG) orthogonalizes similar patterns by sparsification.

CA1 helps expanding the highly compressed representation in CA3 on the way back to the association areas.

Subiculum has no specific function associated with it.

The entorhinal-hippocampal part has been implemented as a connectionist model.

(Treves and Rolls, 1994, Hippocampus 4(3):374–391)
- **Context Reset by dentate gyrus (DG)**
  - Dentage gyrus performs disambiguation of similar patterns.

- **Intrinsic Sequences in CA3**
  - Patterns are connected by association with pre-existing sequences.

- **Pattern completion in CA1**
  - Pattern storage and retrieval is done through feedforward hetero-association.

- **This is a conceptual model.**

(Cheng, 2013, Frontiers in Neural Circuits 7(88):1–14)
Memory Fidelity of Single Patterns
Cell numbers, connectivity and sparsity are derived from rat. Scaling factor for number of neurons is 100, for connections per neuron is 10.

Activation is $p_i(t + 1) = \sum w_{ij}p_j(t)$ with $k$-winners-take-all.

Autoassociative feedback loop in CA3 is run 15 times per pattern.

Learning rules exactly as in (Rolls, 1995).

Storage is done via DG, recall via EC→CA3 connections.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Following the Rolls (1995) Model

The Rolls model (top and lower left) used 1% activity in CA1 for 100 patterns and full connectivity from CA1 to EC. We changed that to 10% and sparse connectivity from CA1 to EC, and during storage CA1 was activated by EC → CA1.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Correlated Input

- Four modules of grid cells as mEC input.
- Population activity at random locations serves as input.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Performance on 252 Random and Correlated Patterns

Solid/dashed lines: with/without recurrent dynamics in CA3.
Learning in DG is disabled, because it drops performance. 252 patterns were used.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Performance on 252 Correlated Patterns

Solid/dashed lines: with/without recurrent dynamics in CA3.

Learning in DG is disabled, because it drops performance. 252 patterns were used.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Performance on Correlated Patterns

CA3

Recurrency No Recurrency

Top/bottom: without/with recurrent dynamics in CA3. Red: correlations with wrong patterns. Blue/cyan: correlations with correct pattern. Blue: cases where the recalled pattern is closer to a wrong than to a correct pattern. Black star: average correlation with correct pattern. Histograms taken at cue quality levels marked by red diamonds in previous graphs.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Performance with EC→CA1→EC Network

- **Storage**: Activity in CA1 triggered by EC→CA3→CA1, without plasticity. Connections EC→CA1→EC are plastic.
- **Retrieval**: EC→CA1→EC only is effective.

![Graphs showing Performance with EC→CA1→EC Network](image)

Solid/dashed lines: with/without recurrent dynamics in CA3.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Summary

- Qualitative behavior of a network can be very different for random and for more natural input patterns.

- Correlation between stored and retrieved patterns is only one measure of performance. Confusion rate might be more important.

- We found feed-forward hetero-association to be more powerful than recurrent auto-association.

- Recurrent dynamics in CA3 was even harmful.

- A simple EC→CA1→EC performed best on correlated input.

(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)
Instantaneous Sequential Storage and Retrieval of Pattern Sequences

$N = 200$.

Fixed connections were pre-trained with gradient descent, plastic connections were trained with Hebbian learning plus weight decay.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
A random sequence of 200 handwritten digits of size $28 \times 28 = 784$ from the MNIST database serves as raw input, shown here by rows from top left to bottom right.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Raw images are compressed with an auto-encoder network down to 220 dimensions to yield the EC representation.

This image shows the reconstructed images from the auto-encoder.
Full Recall from Cue 10 Without Noise

- Cue image is shown negative. Retrieved sequence is rotated for easier comparison.
- The recently stored patterns (lower right) are clearer than the earlier stored patterns (upper left). The quality loss is roughly linear.
- In this run 196/200 of the retrieved sequences are correct.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Full Recall from Cue 10 With 20% Input Noise

Same as before but with 20% input pixel noise.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Full Recall from Cue 10 Without Noise

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Full Recall from Cue 10 With 20% EC Noise

Same as before but with 20% noise in EC.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Reconstructed Input Patterns

| 5 0 4 1 9 2 1 3 1 4 3 5 3 6 1 7 2 8 6 9 |
| 4 0 9 1 1 2 4 3 2 7 3 8 6 9 0 5 6 0 7 6 |
| 1 8 7 9 3 9 8 5 9 3 3 0 7 4 9 8 0 9 4 1 |
| 4 6 0 4 5 6 0 0 1 7 1 6 3 0 2 1 1 7 |
| 9 0 2 6 7 8 3 9 0 4 6 7 4 6 8 0 7 8 3 1 |
| 5 7 1 7 1 6 3 0 2 9 3 1 1 0 4 9 2 0 0 |
| 2 0 2 7 1 8 6 4 1 6 3 4 1 9 \ 3 3 8 5 4 |
| 7 4 2 8 5 8 6 4 3 4 6 1 9 9 6 0 3 7 2 |
| 8 2 4 4 6 4 9 7 0 9 2 7 5 1 5 9 1 9 3 |
| 1 3 5 9 1 7 6 2 8 2 5 0 7 4 9 7 8 3 2 |

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Cue 121 does not trigger the correct sequence.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
But after about 60 time steps CA3 converges to the correct sequence shifted by 39.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Reconstructed Input Patterns

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Full Recall from Cue 14 Without Noise

- Cue 14 does not trigger the correct sequence and CA3 does not recover into the correct sequence at all.
- CA3 fluctuates around a spurious attractor state.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Summary

- It is possible to store a sequence of up to $1.5N$ random patterns in a recurrent CA3 network of $N$ units with gradient descent.

- It is possible to do instantaneous sequential hetero-association of a sequence of correlated patterns to the intrinsic sequence of patterns in a CA3 with some preprocessing (auto-encoder + DG).

- The system has no catastrophic interference/forgetting, quality of retrieved patterns degrades linearly.

- Sequential order is preserved reliably even for similar stimuli and overlapping sequences.

(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)
Thank you!