CRISP: Challenging the Standard Framework of Hippocampal Memory Function

Laurenz Wiskott

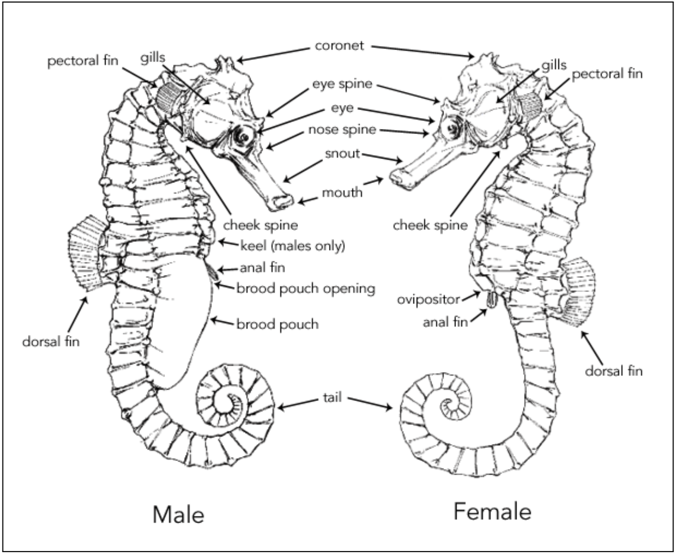
Mehdi Bayati Sen Cheng Jan Melchior Torsten Neher

INSTITUT FÜR NEUROINFORMATIK

RUHR UNIVERSITÄT BOCHUM

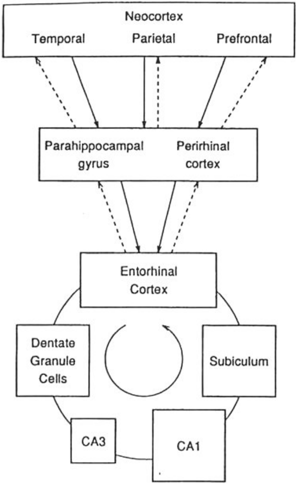
supported by:
Deutsche Forschungsgemeinschaft - Sonderforschungsbereich 874

Theories of the Hippocampus



(2016-02-15 http://www.fusedjaw.com/wp-content/uploads/2011/12/seahorse-anatomy-male-female.png)

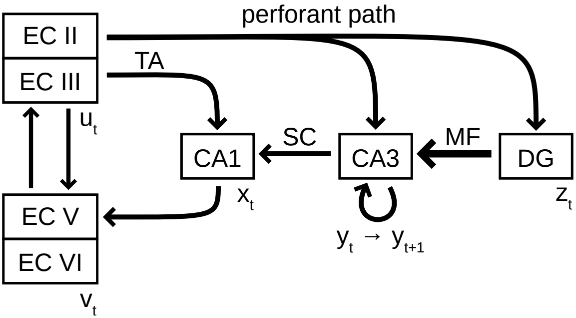
Standard Framework of the Hippocampus



- 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)

CRISP Theory



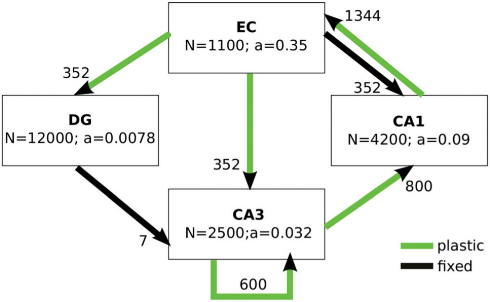
- Context Reset by dentate gyrus (DG)
  - Dentage gyrus performs disambiguation of similar patterns.
- Intrinsic Sequences in CA3
  - Patterns are connected by association with pre-existing sequences.
- Pattern completion in CA1
  - Pattern storage and retrieval is done through feedforward hetero-association.
- This is a conceptual model.

(Cheng, 2013, Frontiers in Neural Circuits 7(88):1–14)

Memory Fidelity of Single Patterns

(2018-01-26 https://pixabay.com/en/loving-memory-memorial-grief-1207568/)

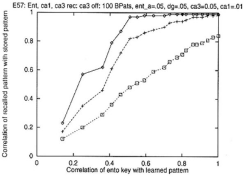
Network

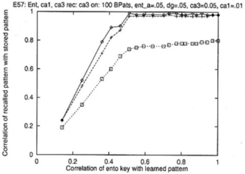


- Cell numbers, connectivity and sparsity are derived from rat.
  Scaling factor for number of neurons is 100, for connections per neuron is 10.
- Activation is $p_i(t+1) = \sum w_{ij} p_j(t)$ with k-winners-take-all.
- Autoassociative feedback loop in CA3 is run 15 times per pattern.
- Learning rules exactly as in (Rolls, 1995).
- Storage is done via DG, recall via EC$\rightarrow$CA3 connections.

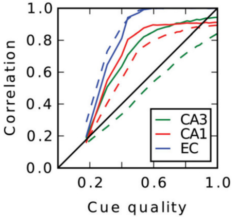
(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)

Following the Rolls (1995) Model

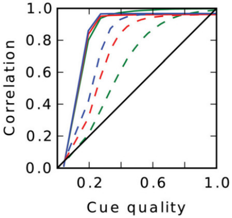




(Rolls, 1995, Intl. J. of Neural Systems 6:51-70)



'Rolls model'



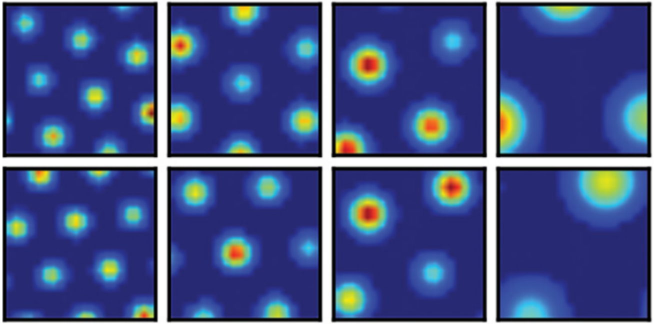
'modified Rolls model'

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

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$\rightarrow$CA1.

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

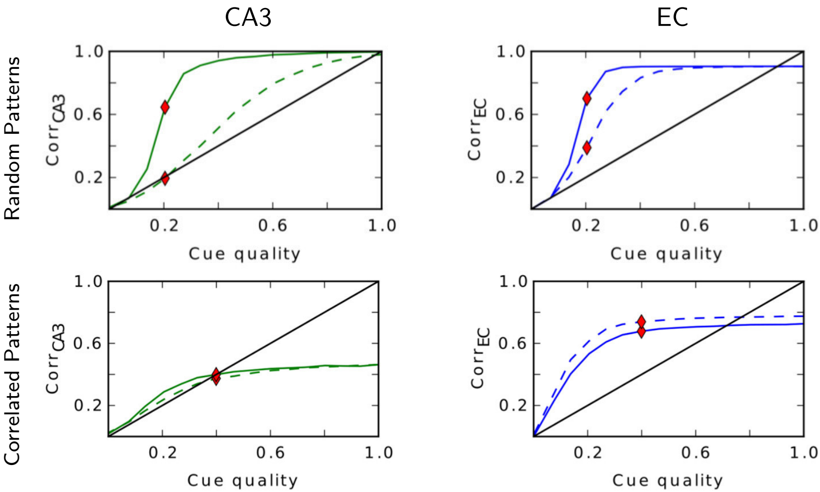
Correlated Input



- 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

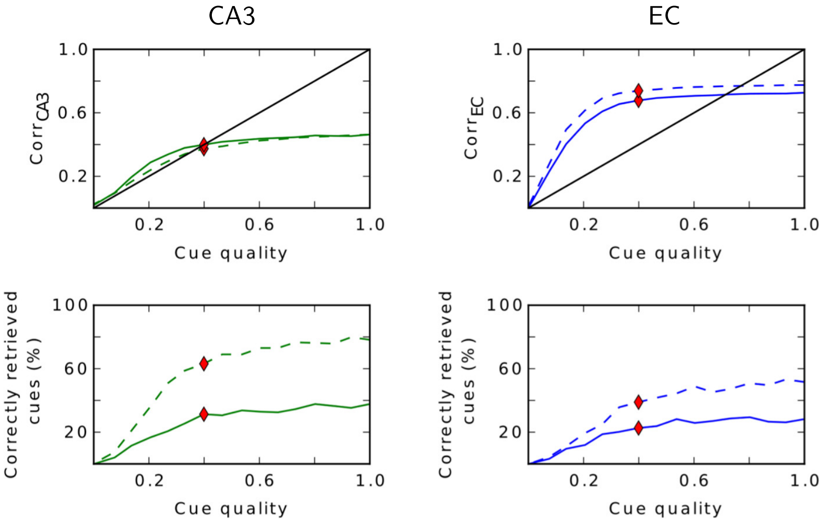


Solid/dashed lines: with/without recurrent dynamics in CA3.

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

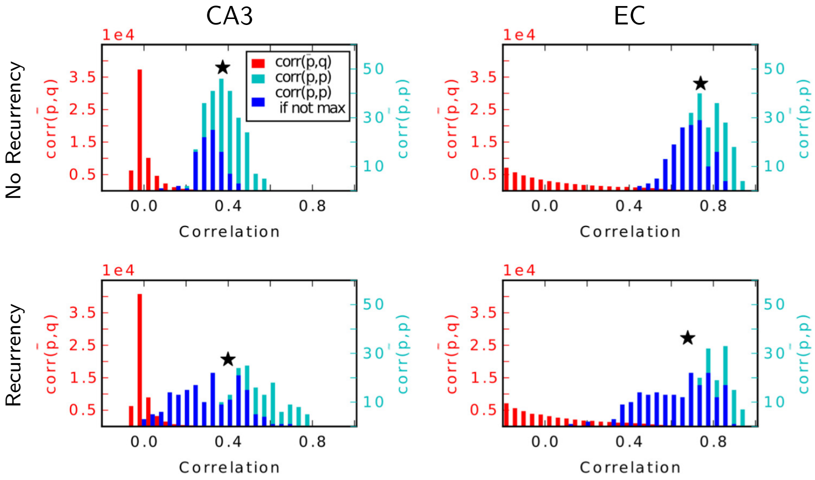


Solid/dashed lines: with/without recurrent dynamics in CA3.

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

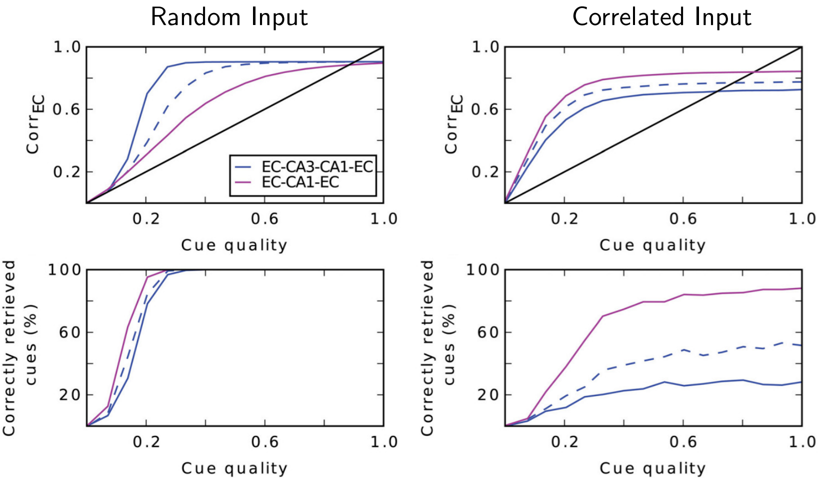


Top/bottom: without/with recurrent dynamics in CA3. Red: correlations with wrong patterns. Blue/cyan: correlations with correct pattern. Blue: cases where the recalled pattern is closer to a wrong than to a correct pattern. Black star: average correlation with correct pattern. Histograms taken at cue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



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

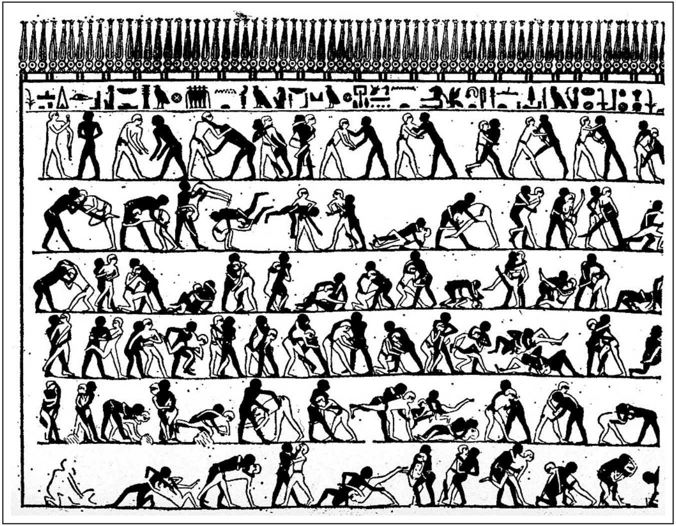
(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)

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$\rightarrow$CA1$\rightarrow$EC performed best on correlated input.

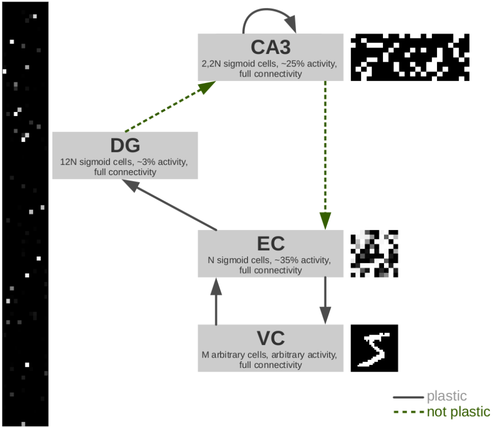
(Neher, Cheng, & Wiskott, 2015, PLoS Comp. Biol. 11:e1004250)

Instantaneous Sequential Storage and Retrieval of Pattern Sequences



(2018-02-07 https://commons.wikimedia.org/wiki/File:Egyptmotionseries.jpg)

Network



- $N = 200$.
- 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)

Raw Input Patterns

- 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.

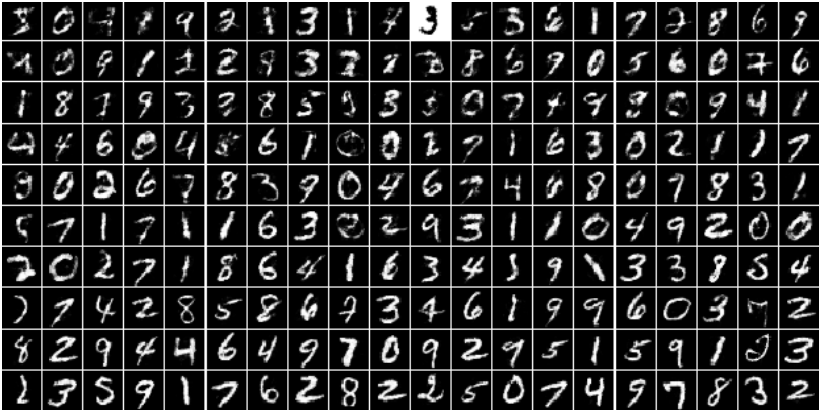
(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)

Reconstructed Input Patterns

- 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.

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

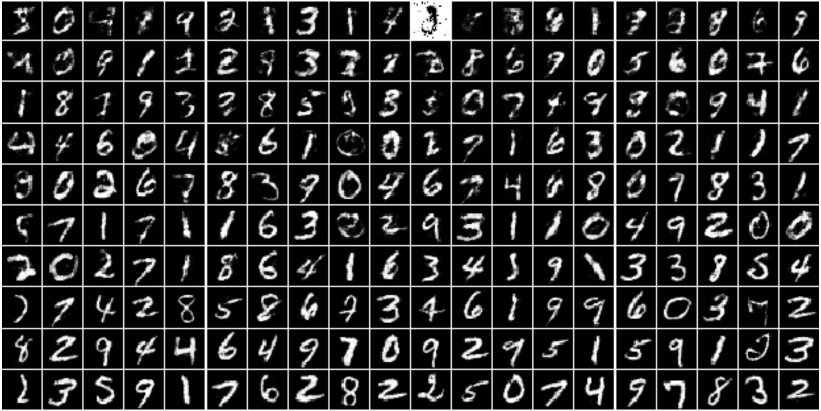
Full Recall from Cue 10 Without Noise



- 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)

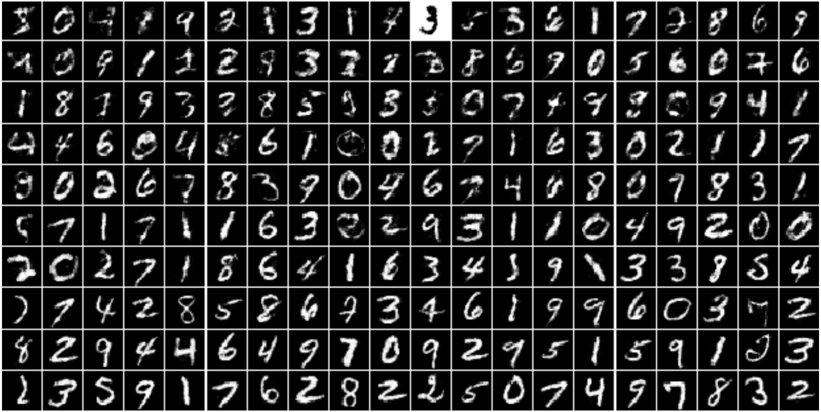
Full Recall from Cue 10 With 20% Input Noise



- Same as before but with 20% input pixel noise.

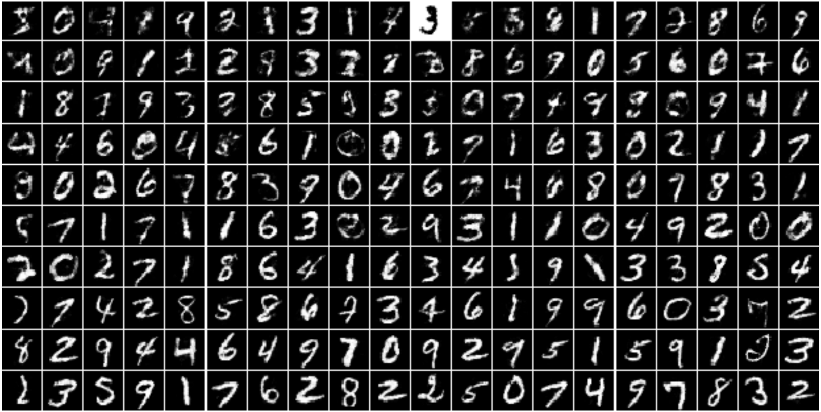
(Melchior, Bayati, Cheng, & Wiskott, 2018, in preparation)

Full Recall from Cue 10 Without Noise



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

Full Recall from Cue 10 With 20% EC Noise



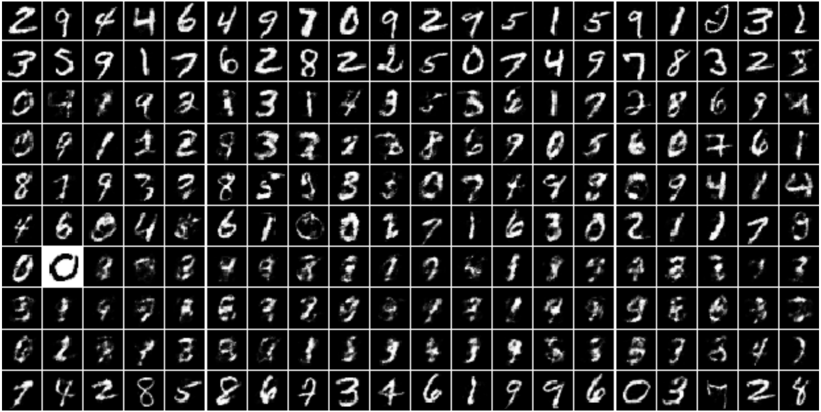
- Same as before but with 20% noise in EC.

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

Reconstructed Input Patterns

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

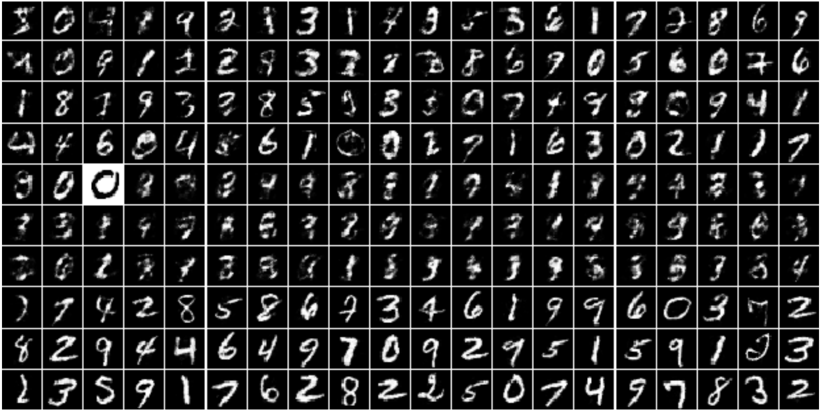
Full Recall from Cue 121 Without Noise



- Cue 121 does not trigger the correct sequence.

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

Full Recall from Cue 121 Without Noise Shifted by 39



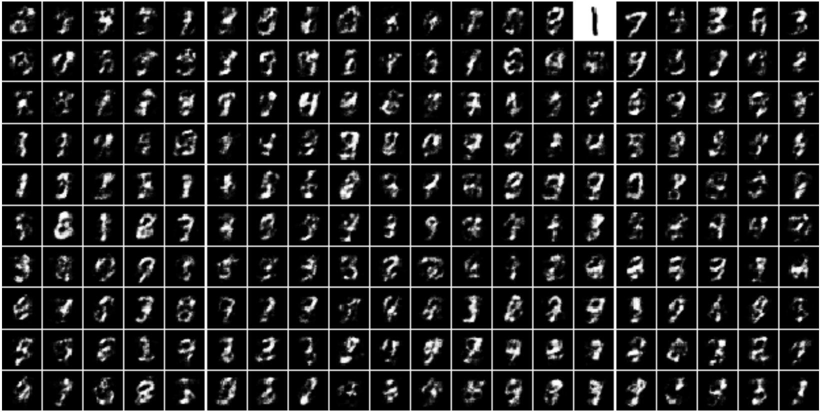
- But after about 60 time steps CA3 converges to the correct sequence shifted by 39.

(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!