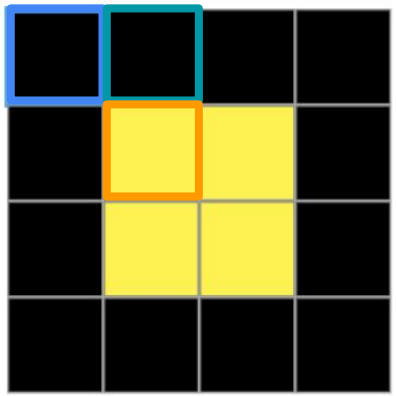
Long- and short-term memory using attractor network with triple-well connection decay

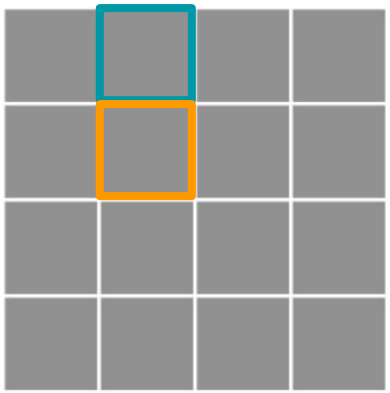
Considerations on LTM decay and on LTM-STM interactions

Aisling Mellon aismellonw@gmail.com
John Hughes johnniallhughes@gmail.com
Francisco Passos francisco.passos@gmail.com
Maxwell Buckley maxwbuckley@gmail.com

Intuition about the model

present

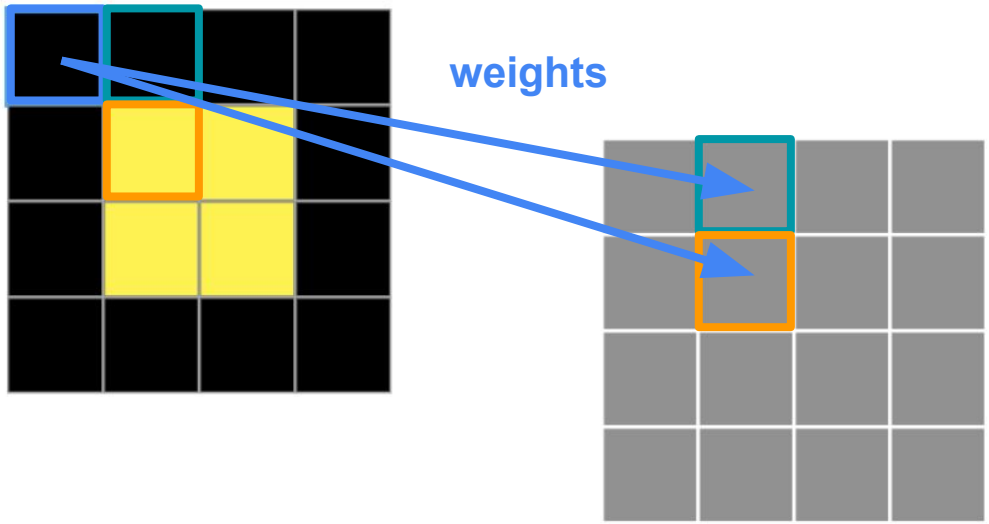




Learning:

- Look at pixel pair (a, b) values

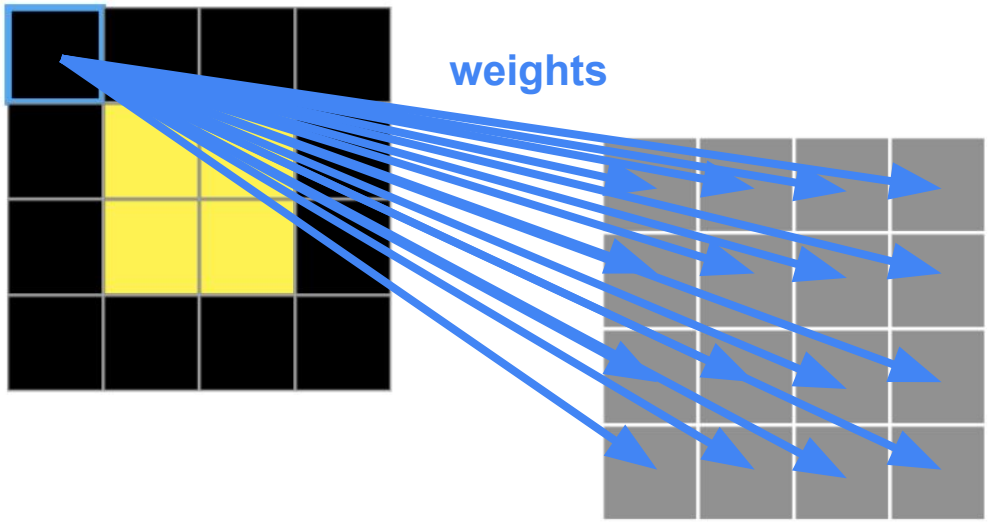
Intuition about the model



Learning:

- Look at pixel pair (a, b) values
- Reflect their relationship into w_{ab}

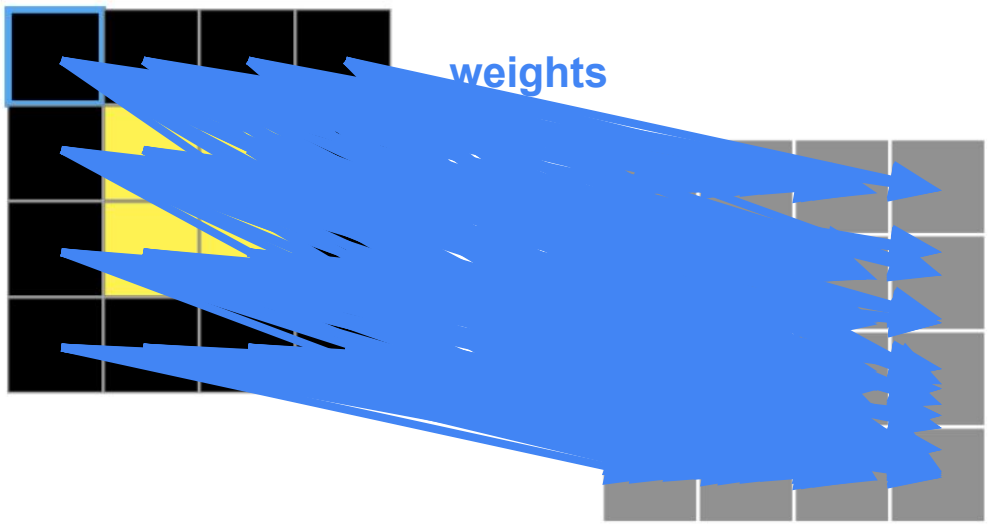
Intuition about the model



Learning:

- Look at pixel pair (a, b) values
- Reflect their relationship into w_{ab}
- Do that for all destination pixels from that pixel

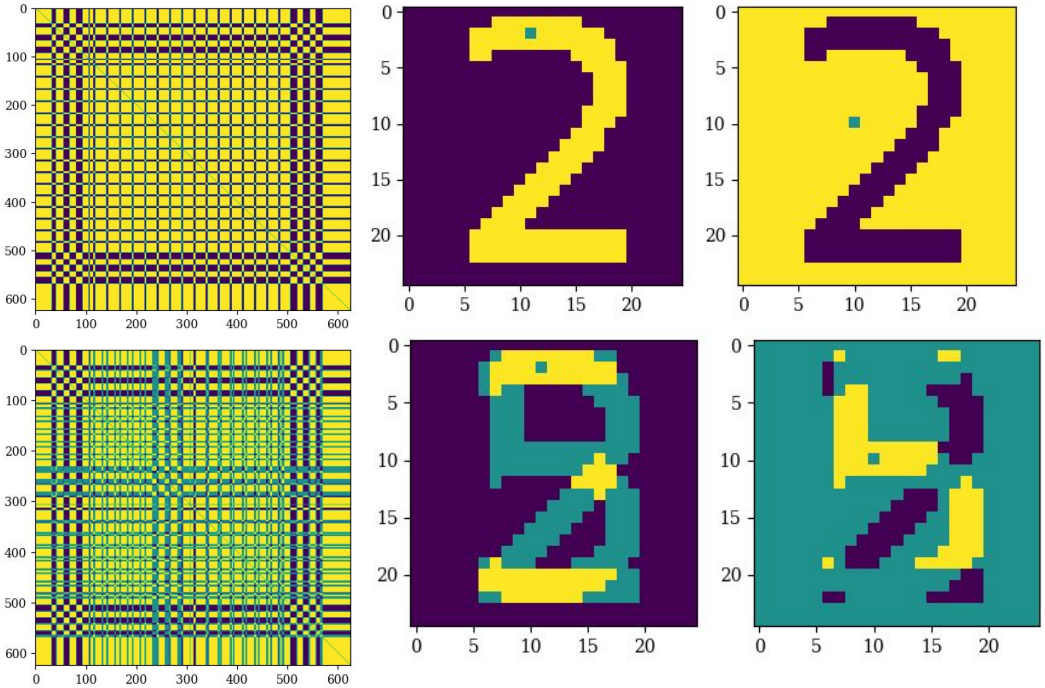
Intuition about the model



Learning:

- Look at pixel pair (a, b) values
- Reflect their relationship into w_{ab}
- Do that for all destination pixels from that pixel
- Repeat that from every pixel source

Weights



Images are presented in sequence without delay

A negative weight represents an inverse correlation to a certain pixel

A weight of zero is shown as turquoise colour.

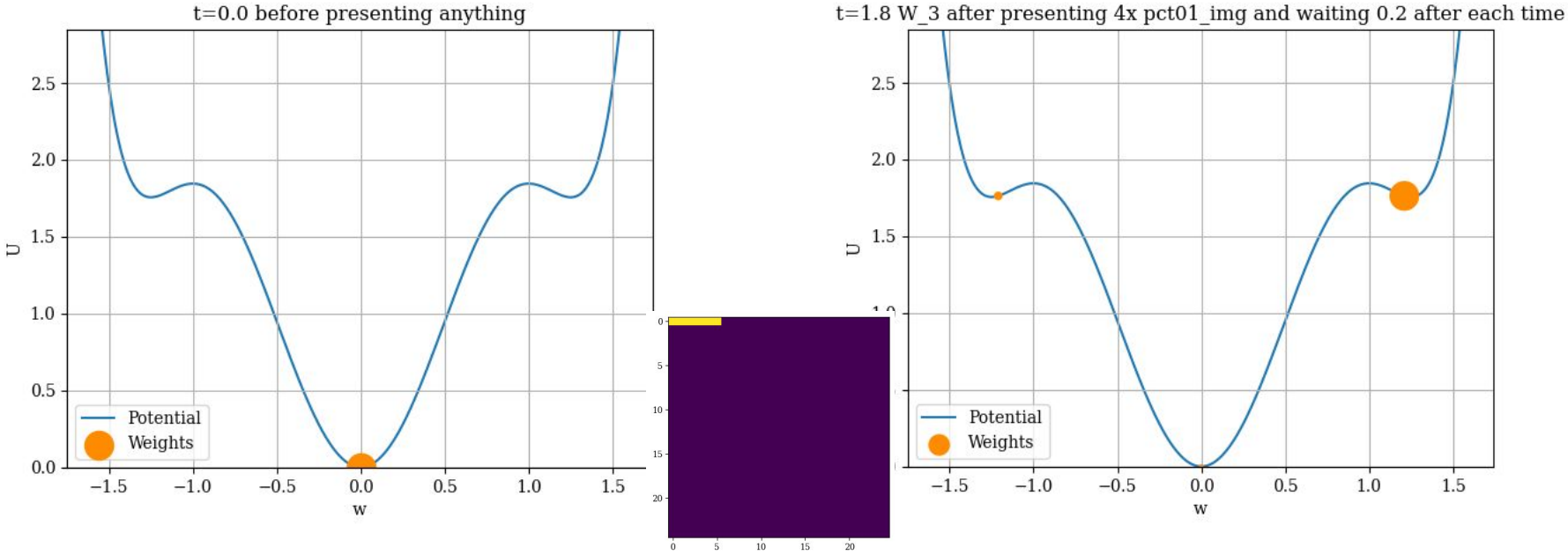
This is seen by the isolated pixel having a turquoise connection and the model requires no self connections.

$$W_{ii} = 0$$

LTM decay caused by subthreshold learning (1)

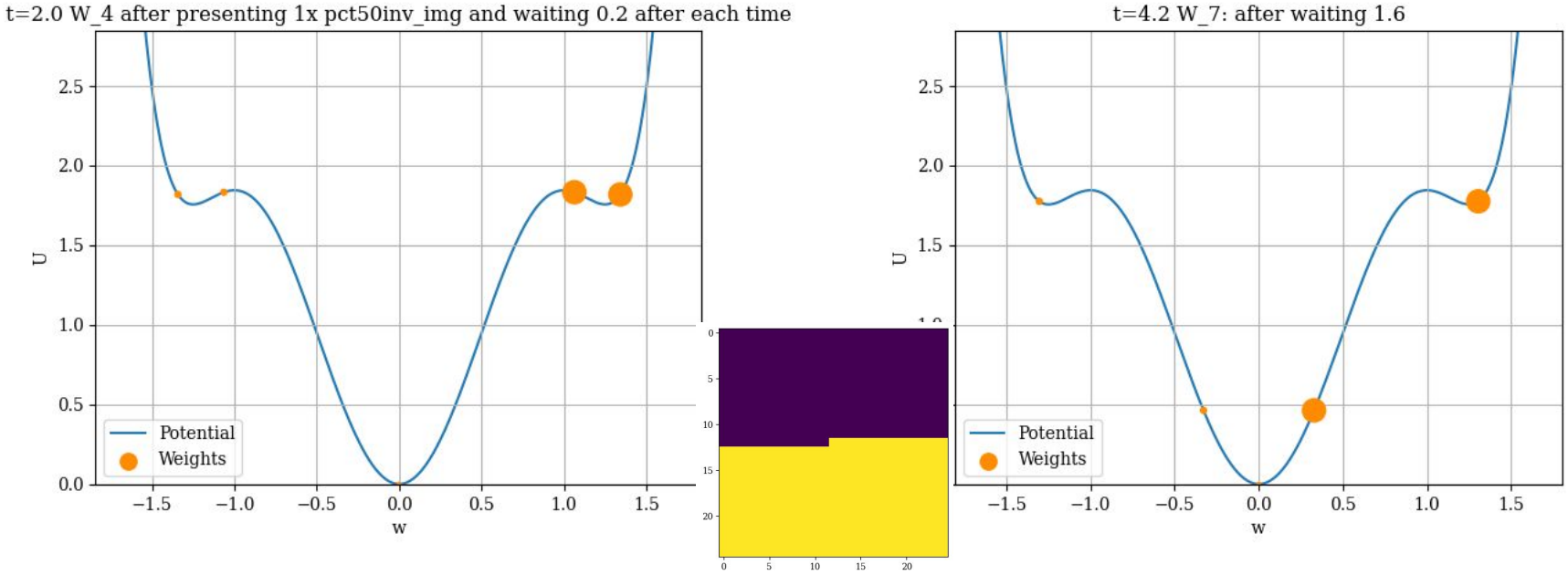
A long-term memory forms when some - or all - of its weights enter an LTM well.

Temporal decay does not move weights outside the well.



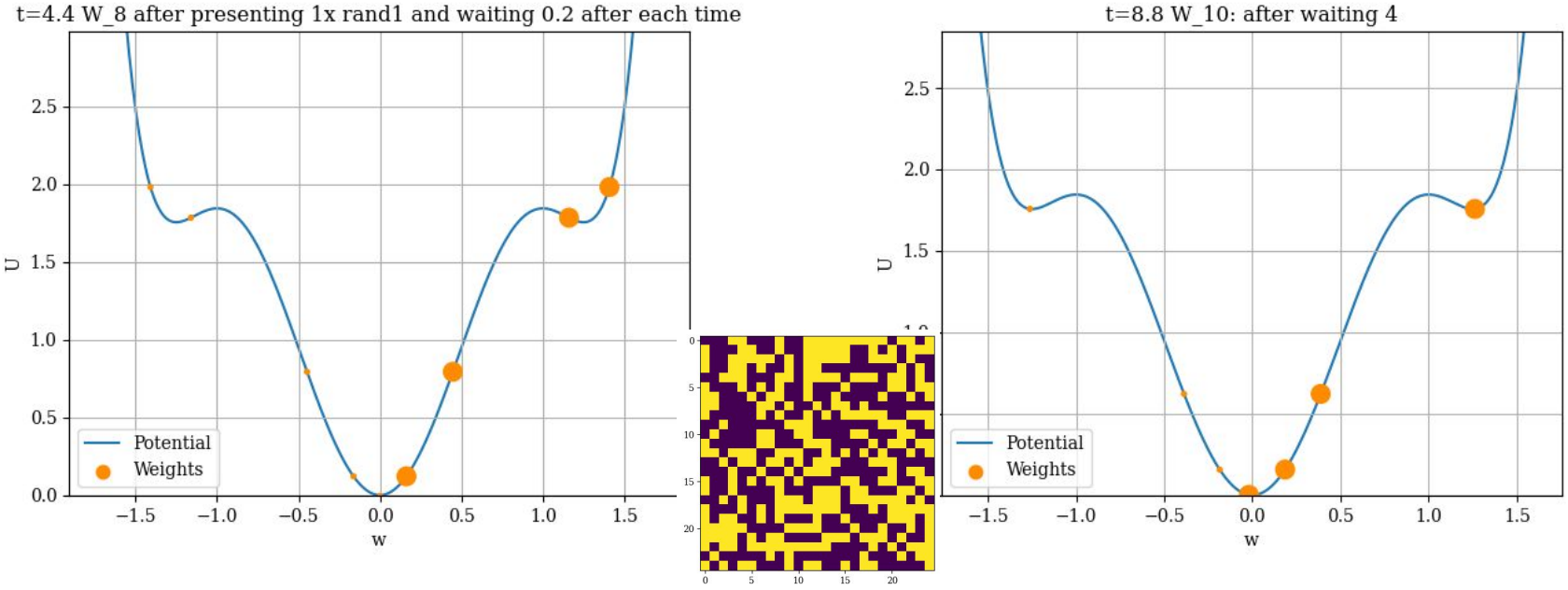
LTM decay caused by subthreshold learning (2)

Repeated exposure to patterns different to the ones in LTM memories pulls some weights towards lower absolute values (and reinforces others).



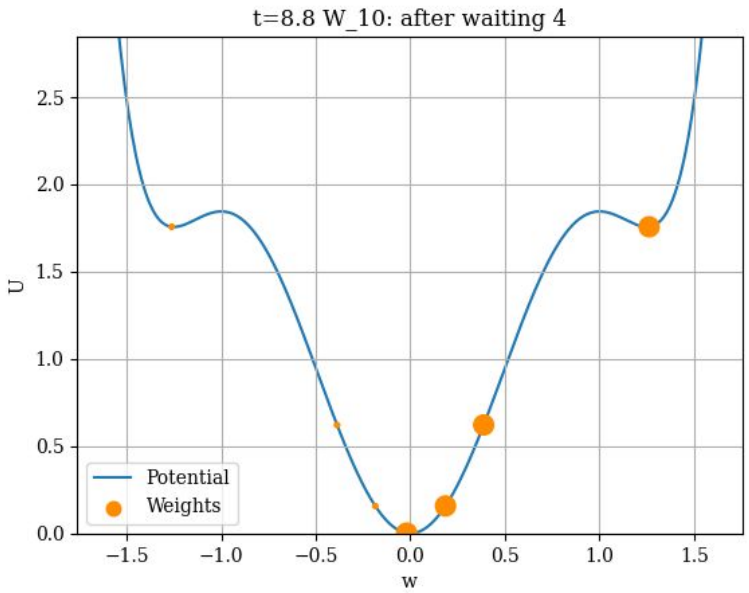
LTM decay caused by subthreshold learning (3)

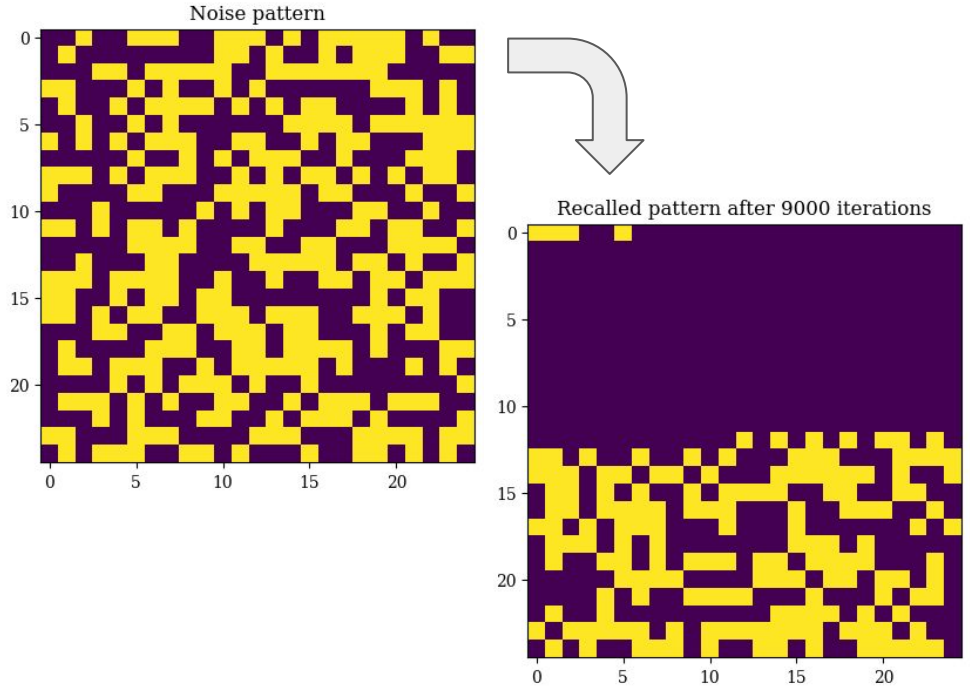
Weights outside of the LTM well become vulnerable to temporal decay



LTM decay caused by subthreshold learning (4)

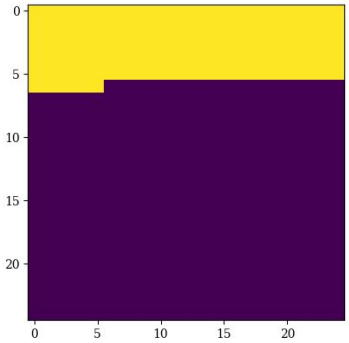
At some point enough weights have been affected and the attractor may have deteriorated

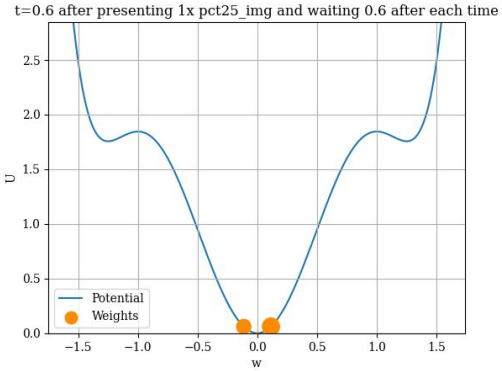




Memory Interference (1)

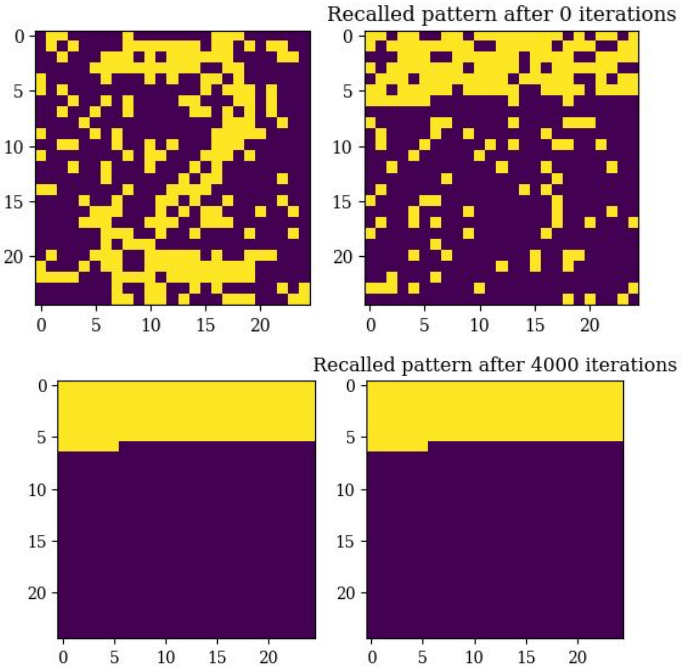
Training image





Model was trained with one image, then memory was allowed to decay for a 0.6 timestep.

The model showed perfect recall with highly correlated and uncorrelated input images



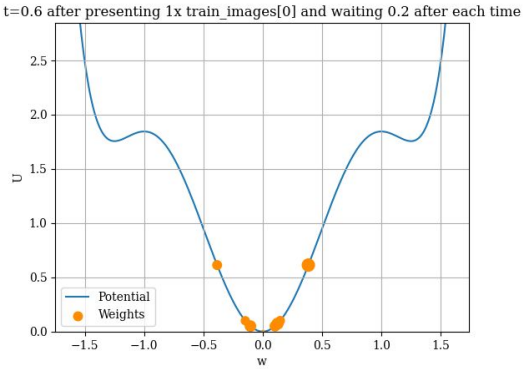
Memory Interference (2)

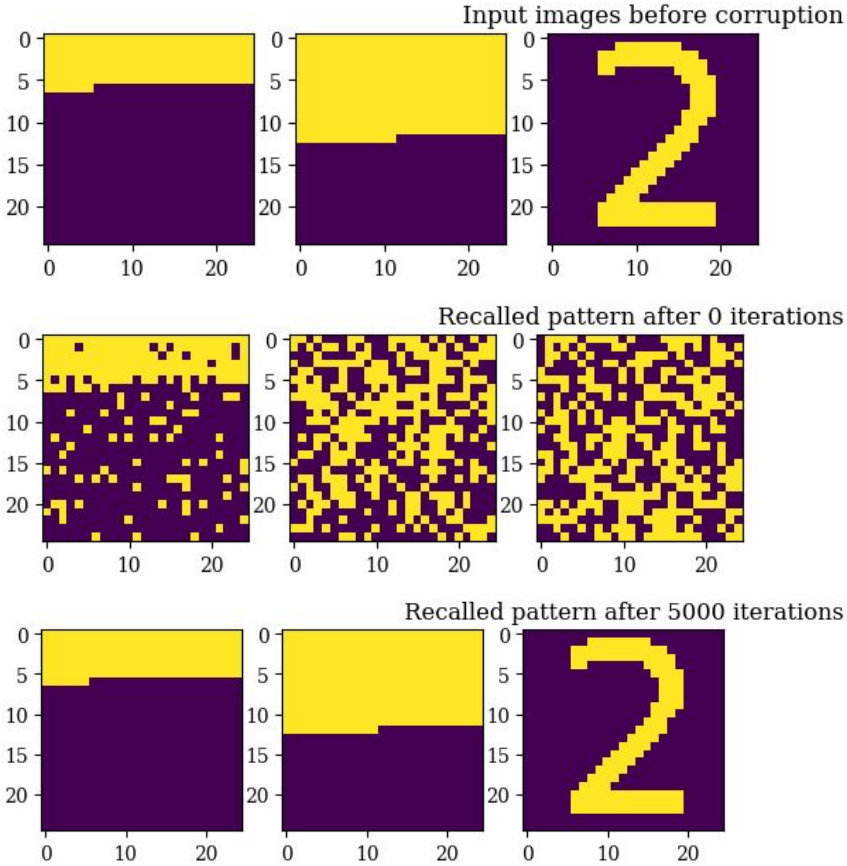
Same total time delay, but much less robust memory caused by relative weight difference

Figures show maximum corruption of input image which will still result in correct recall

Oldest image recalled with corruption factor of 0.162433
Intermediate image recalled with corruption factor 0.42
Most recently trained image recalled with corruption factor of 0.45

Heavily biased towards most recent image





Possible future work

- Quantify recall speed vs patterns memorized
- Analytical demonstration of the boundaries of retrievability
- Quantification of attractor interference
- Mechanisms of implanting intentional false attractors

Thank you for your attention

Questions?

Aisling Mellon aismellonw@gmail.com
John Hughes johnniallhughes@gmail.com
Francisco Passos francisco.passos@gmail.com
Maxwell Buckley maxwbuckley@gmail.com