Graph Reproduction and Synaptic Model Comparisons

For Professor Nicolas Brunel
By Riccardo Scibetta

25/06/2025

1 Graph Reproduction

This section reproduces the graph from the original Benna and Fusi (2016) paper. The synaptic model they propose consists of a synapse with m hidden variables u_k (typically $m = 4$–12), which interact bidirectionally. The first variable u_1 represents the synaptic efficacy, and the dynamics are defined by:

$$\Delta u_k = \alpha n^{-2k+2}(u_{k-1} - u_k) + \alpha n^{-2k+1}(u_{k+1} - u_k) \tag{1}$$

with boundary cases:

$$\Delta u_1 = I_{ij} + \alpha n^{-2k+1}(u_{k+1} - u_k) \tag{2}$$
$$\Delta u_m = \alpha n^{-2k+2}(u_{k-1} - u_k) \tag{3}$$

Each variable u_k takes a bounded number of discrete values (commonly $L_k = 30$–50, centered around 0, spaced at intervals of 1). These levels may decrease for deeper variables.

We consider an asynchronous, deterministic Hopfield network with $N = 10{,}000$ binary neurons $S_i \in \{+1, -1\}$ endowed with Bidirectional Cascade synapses. At each time step, one new memory pattern ξ^μ is stored using the Hebbian input rule:

$$I_{ij} = \xi_i^\mu \xi_j^\mu \tag{4}$$

The network is initialized at the stored pattern and evolved asynchronously for 10 iterations. The overlap is then computed as:

$$m(\tau) = \frac{1}{N} \sum_i S_i(\tau) \xi_i^\mu \tag{5}$$

This is repeated with initialization corrupted by noise $\epsilon = 0.25$. Blue lines indicate recall of memory after perfect initialization; red lines indicate recall from corrupted initial conditions.

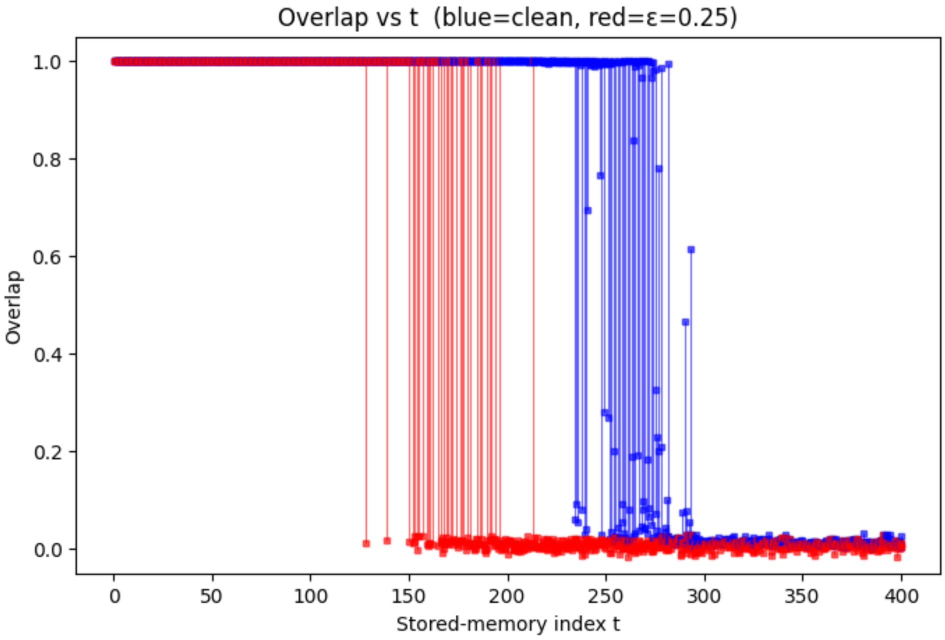


Figure 1: Reproduced graph with $N = 10{,}000$, $\alpha = 0.25$, $n = 2$, $m = 4$ hidden variables, and $L_i = 30$.

The result aligns with the original graph from Benna & Fusi, shown below for comparison ($N = 30{,}000$):

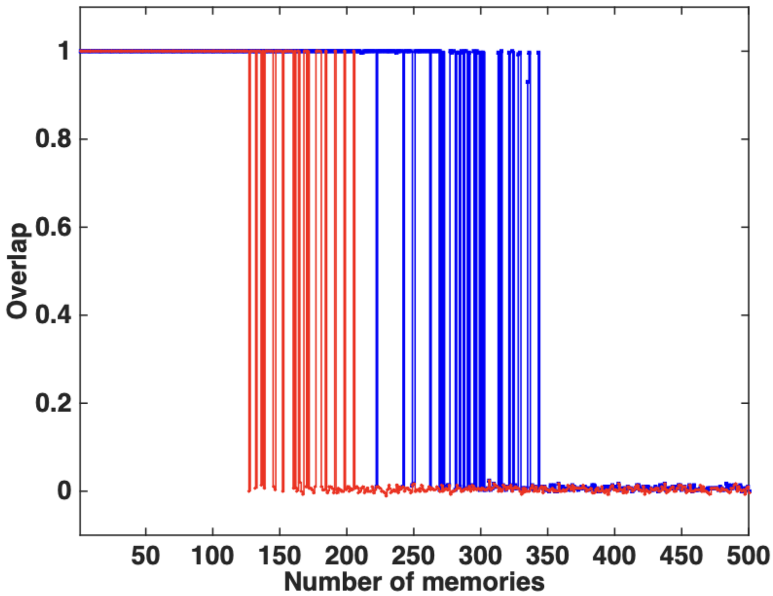


Figure 2: Original figure from Benna & Fusi (2016), ArXiv preprint.

2 Comparison of Synapse Models

We now compare the memory lifetime of four synapse models:

- **Unbounded Continuous** (Hopfield, 1982)
- **Bidirectional Cascade** (Benna & Fusi, 2016)

- **Double-Well** (Feng & Brunel, 2024)
- **Weight-Decaying** (Mézard et al., 1986)

a) Weight-Decaying Synapses

Update rule:

$$J_{ij}(t+1) = \alpha \frac{1}{N} \xi_i^\mu \xi_j^\mu + \lambda J_{ij}(t) \tag{6}$$

b) Double-Well Synapses

These evolve in a potential $U(J)$ with two minima at $\pm C$:

$$\Delta J_{ij} = -r_1 \frac{dU(J)}{dJ} + r_2 I(\xi_i^\mu, \xi_j^\mu) + r_3 \epsilon_{ij}^\mu \tag{7}$$

The interaction function I is:

$$I\left(\xi_i^\nu, \xi_j^\nu\right) = \begin{cases} (1-f)^2 & \text{if } \xi_i^\nu = \xi_j^\nu = 1 \\ -f(1-f) & \text{if } \xi_i^\nu \neq \xi_j^\nu \\ f^2 & \text{if } \xi_i^\nu = \xi_j^\nu = 0 \end{cases} \tag{8}$$

This reduces to the Hebbian rule when $f = 0.5$.

c) Bidirectional Cascade Synapses

Tested as described in Section 1.

d) Unbounded Continuous Synapses

Updated with standard Hebbian learning:

$$J_{ij} = \frac{1}{N} \sum_\mu \xi_i^\mu \xi_j^\mu \tag{9}$$

Important Note: The superior performance of Double-Well synapses may partly result from the use of **synchronous dynamics**, as per Feng & Brunel (2024). All other models used asynchronous dynamics, which introduce greater noise. This methodological asymmetry should be kept in mind when comparing results.

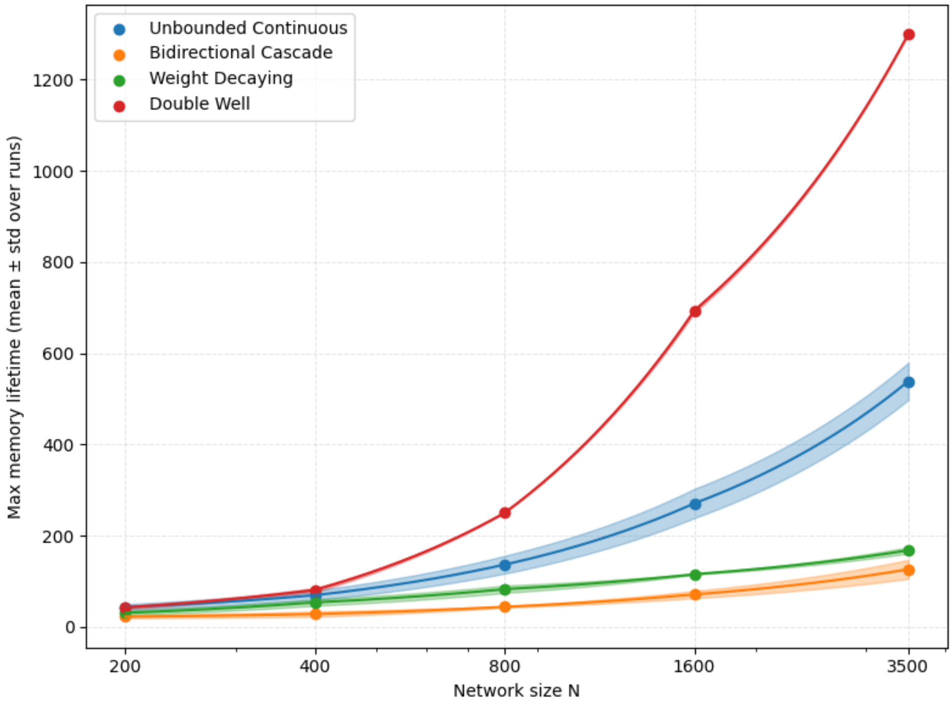


Figure 3: Maximum memory lifetime across synapse models. Lifetime is defined as the number of stored memories retrievable with $m > 0.97$ after 10 iterations. Averages taken over 5 runs. The parameters of synaptic models were to chosen to maximize memory lifetime but also to avoid catastrophic forgetting were possible. For Weight-decaying $\mu = 0.995$ and $\alpha = 4$. For Bidirectional Cascade $n = 2$, $\alpha = 0.25$, $m = 4$ and L_i monotonically decreasing from $L_1 = 35$ to $L_4 = 2$. For Double Well synapses $r_2 = 1$ and $r_3 = 0$, while r_1 and C were different for each level of N.

3 Sparse Coding Implementation

All models were also implemented in the sparse coding regime ($f \ll 1$) using the Tsodyks–Feigelman framework:

$$\tilde{\xi}_i^\mu = \xi_i^\mu - f \tag{10}$$

$$J_{ij} = \frac{1}{Nf(1-f)} \sum_\mu \tilde{\xi}_i^\mu \tilde{\xi}_j^\mu \quad (i \neq j), \quad J_{ii} = 0 \tag{11}$$

$$\tilde{S}_j(t) = S_j(t) - f \tag{12}$$

$$h_i(t) = \sum_j J_{ij} \tilde{S}_j(t) \tag{13}$$

$$S_i(t+1) = \text{sign}(h_i(t)) \tag{14}$$

$$m^\mu(t) = \frac{1}{Nf(1-f)} \sum_i (S_i(t) - f)\,(\xi_i^\mu - f) \tag{15}$$

However, I found testing memory lifetime in this regime is more difficult due to low steady-state overlap at limited load. I did not know how to procede with the analysis of memory lifetime in the sparse coding regime.

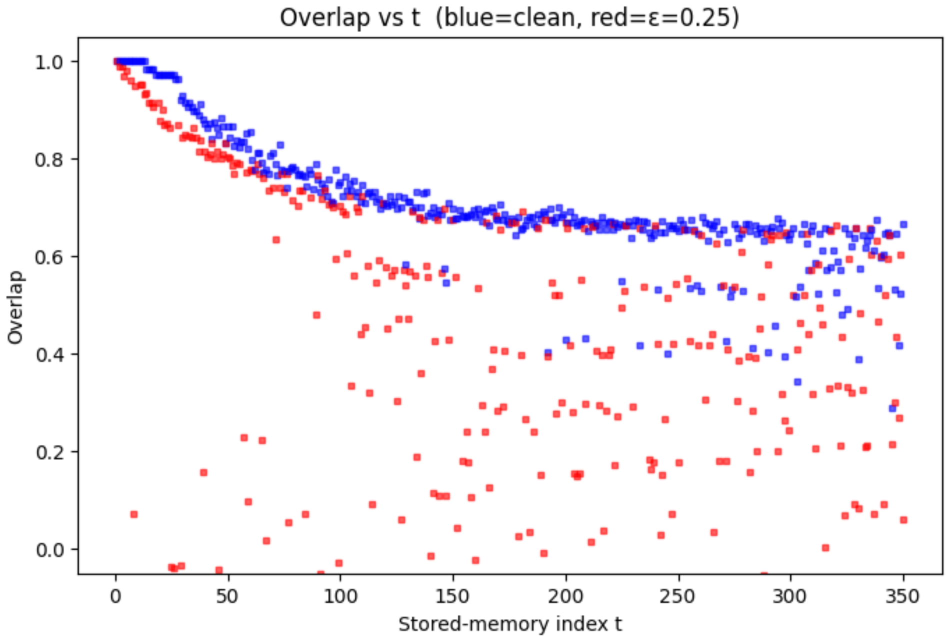
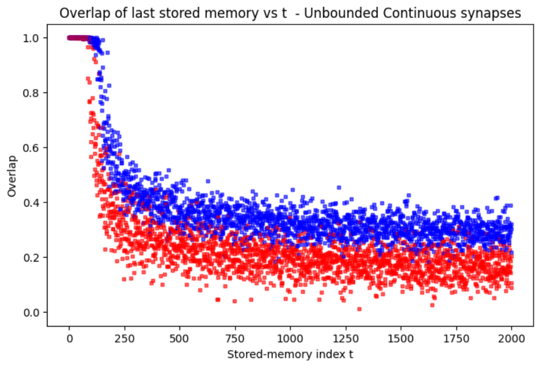


Figure 4: Overlap vs time for sparse coding ($f = 0.01$) with continuous unbounded synapses. $N = 800$, 10 asynchronous updates. One run shown.

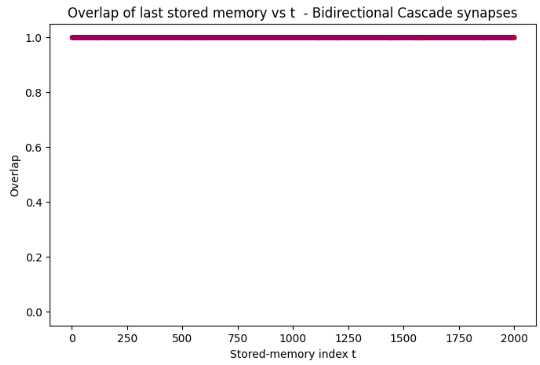
4 Catastrophic Forgetting

To test resilience against catastrophic forgetting, we initialize the network at the most recently stored pattern and track overlap over time.

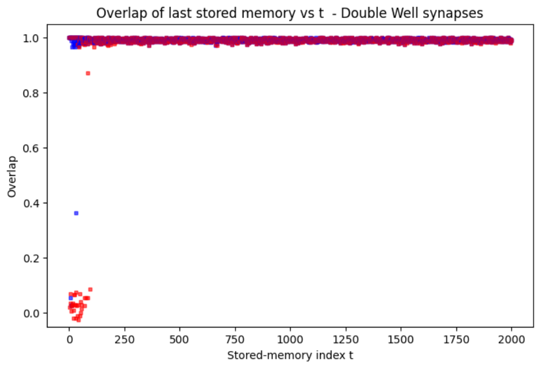
Below are the plots for $N = 800$ with the parameters of each model set to be equal to those of Figure 3.



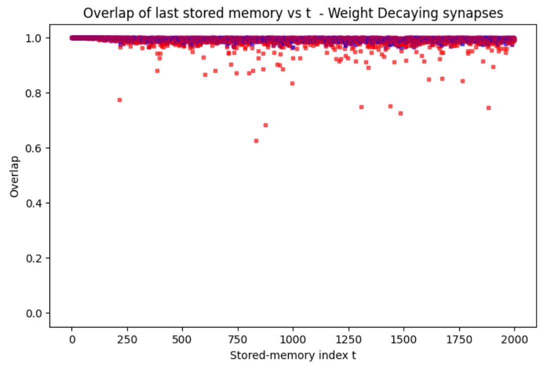
Unbounded Continuous



Bidirectional Cascade



Double Well



Weight Decaying

Hence, we can observe that all synapse models can be tuned to resist catastrophic forgetting when appropriate parameters are chosen.

5 Discussion

The primary objective of this study was to investigate the memory performance of the Bidirectional Cascade synaptic model within the framework of fully connected Hopfield networks. The results show that, in terms of memory lifetime and resistance to catastrophic forgetting, this model performs less favorably when compared to other synaptic models such as Double-Well and Weight-Decaying synapses.

However, it is important to highlight that one of the key theoretical advantages of the Bidirectional Cascade model lies in its ability to maintain a high signal-to-noise ratio (SNR), both immediately after memory encoding and over time. This feature, which is central to the original Benna and Fusi proposal, is not fully leveraged in the fully connected Hopfield network used in our simulations.

The reason lies in the structure of fully connected networks: each neuron connects to all others, resulting in N^2 synapses. Yet, due to symmetry and overlap in connectivity, only N of these synapses contribute independently to the input received by any given neuron. The rest are strongly correlated. This correlation limits the effective diversity of synaptic contributions and diminishes the advantage of having high SNR at the level of individual synapses, as the accumulated input becomes dominated by shared noise.

Therefore, while the Bidirectional Cascade model may not outperform others in the fully connected Hopfield setting, it could be more effective in architectures where synaptic independence is greater—such as sparsely connected or structured networks—where the advantages of high SNR might be more fully realized.