A BIOLOGICALLY PLAUSIBLE ASSOCIATIVE MEMORY NETWORK

Mohadeseh Shafiei Kafraj
Gatsby Computational Neuroscience Unit
University College London
London W1T 4JG, UK
mohadeseh.kafraj.22@ucl.ac.uk

Dmitry Krotov
MIT-IBM Watson AI Lab
IBM Research
Cambridge, MA, USA
krotov@ibm.com

Brendan A. Bicknell
Gatsby Computational Neuroscience Unit
University College London
London W1T 4JG, UK
brendan.bicknell@gmail.com

Peter E. Latham
Gatsby Computational Neuroscience Unit
University College London
London W1T 4JG, UK
pel@gatsby.ucl.ac.uk

ABSTRACT

The Hopfield network has been the leading model for associative memory for over four decades, culminating in the recent 2024 Nobel Prize. However, the vanilla version of the Hopfield network has a capacity that scales with the number of connections per neuron. In the mammalian brain, that's about 1,000, leading to a capacity of about 50 memories in a spiking network—regardless of its size. Therefore, it cannot possibly account for the capacity of human memory. To address this limitation, various modifications to the Hopfield network have been proposed. One promising variant is Dense Associative Memory, which significantly increases capacity and could be implemented in a two-layer architecture consisting of memory and feature neurons. However, from the point of view of biological plausibility, this comes with a downside: during recall of a specific memory, all neurons but one cued neuron (a neuron that is associated with the recalled memory) in the memory layer are silent, whereas in the brain, neurons are rarely silent for extended periods. This is not easy to fix: the memory layer contains a large number of neurons, and allowing non-cued neurons (neurons that are not associated with the recalled memory) to exhibit even low firing rates can introduce an unacceptable level of noise, preventing the perfect recall of the cued memory. To address this challenge, we propose a novel architecture that introduces nonlinear dendrites in the Dense Associative Memory network. This model supports a capacity that is polynomial in the number of memory neurons while enabling non-cued memory neurons to be unsilenced. The proposed architecture adheres to other key biological constraints, including the presence of both excitatory and inhibitory populations that obey Dale's law and maintain non-saturated firing rates and sparsity in the connections. These properties enhance the model's biological plausibility while achieving polynomial capacity, bridging the gap between theoretical and biological constraints on associative memory.

1 INTRODUCTION

Associative memory models as a variant of an attractor network could be powerful analogs for understanding human memory (Wills et al. (2005)). However, current models face significant limitations regarding capacity and biological plausibility. In the case of the standard Hopfield Network, memory capacity is constrained by the number of synaptic connections. Since the brain has sparse connections, the number of synaptic connections is far fewer than the number of neurons, resulting in an impractically low memory capacity (Roudi & Latham (2007)).

To address the limitation on capacity, Krotov and Hopfield (Krotov & Hopfield (2016)) introduced the Dense Associative Memory model, also known as the Modern Hopfield Network, which dramatically increases capacity. However, this network still falls short in its ability to describe biological networks at a true microscopic level. Specifically, it incorporates many-body synaptic terms (synapses that are shared

between more than two neurons) in the equations governing its dynamics, which limits its biological plausibility.

Later, Krotov and Hopfield proposed a more biologically plausible implementation of Dense Associative Memory, which uses only two-body synapses (Krotov & Hopfield (2021)). This model divides neurons into two populations: memory neurons and feature neurons. While the network's capacity grows exponentially with the number of feature neurons, it is still limited by the number of memory neurons (Krotov (2021)). Additionally, a major biological concern arises during memory recall, where most memory neurons, which are not associated with the recalled memory, remain silent, contradicting the well-established observation that a large group of neurons in the brain rarely stay completely inactive. Other models that attempt to enhance associative memory capacity also fail to meet critical biological constraints, including, but not limited to, making the majority of neurons silent, not adhering to Dale's law, firing at saturation level, or losing capacity under sparsity of connections (Chandra et al. (2025), Kozachkov et al. (2023)). As such, they cannot serve as potential models for human memory.

In this paper, we propose a novel architecture for an associative memory network that enhances Dense Associative Memory with nonlinear dendritic computation (Poirazi et al. (2003)). This expanded model supports a polynomial capacity by distributed encoding of memories and key biological plausibility constraints, including: it does not require all non-cued neurons to be silent; neurons are allowed to have a low but nonzero firing rate; there are both excitatory and inhibitory populations in the network; all neurons obey Dale's law and fire at a non-saturated rate; and it supports sparsity in synaptic connections.

2 MODEL

To address challenges on capacity and biological plausibility, we introduce a new architecture for associative memory (Figure 1a). The network consists of two excitatory populations, E_1 (memory) and E_2 (feature), inspired by Dense Associative Memory. One key biological plausibility issue we address is avoiding the requirement that all non-cued neurons be silent when recalling a memory. This is achieved by augmenting each feature neuron with N_D nonlinear dendritic branches.

When recalling a memory without dendrites, a feature neuron would receive noisy input from many non-cued memory neurons and a signal from only a few neurons coding for the recalled memory, all arriving at the soma. This would make it difficult to distinguish signal from noise unless the associated neurons fire at an exceptionally high rate. With dendrites, each branch processes input from only a small subset of memory neurons reducing noise variance and enhancing signal clarity. Therefore, this architecture improves the robustness of memory retrieval in the presence of noise from non-cued neurons, without requiring excessively high firing rates for the cued neurons.

To ensure that branches of E_2 neurons receiving input from non-cued E_1 neurons have minimal effect on the soma, the input to these branches must be very small, while E_1 neurons adhere to Dale's law. This is achieved by introducing the inhibitory population I_1 , which ensures that the input to non-active branches remains very small if they do not receive input from a cued neuron (see Appendix).

For the same reason, since all memory neurons—even when not cued—receive input from feature neurons, we need to ensure that non-cued E_1 neurons maintain low activity and receive a small total input. This is achieved by introducing the inhibitory population I_2 into the model. The presence of these inhibitory groups stabilizes the network, preventing runaway activity and pathologically high firing rates as well. The firing rate dynamics of neurons in each population are described as follows:

$$\tau_{E_1} \frac{d\nu_i^{E_1}}{dt} = F\left(h_i^{E_1}\right) - \nu_i^{E_1}, \tag{1a}$$

$$\tau_{E_2} \frac{d\nu_i^{E_2}}{dt} = \frac{1}{\sqrt{N_D}} \sum_{d_i}^{N_D} J_{id_i}^{E_2 D} G\left(h_{d_i}^{E_2}\right) - \nu_i^{E_2}, \tag{1b}$$

$$\tau_{I_1} \frac{d\nu_i^{I_1}}{dt} = \left[h_i^{I_1}\right]^+ - \nu_i^{I_1}, \tag{1c}$$

$$\tau_{I_2} \frac{d\nu_i^{I_2}}{dt} = \left[h_i^{I_2}\right]^+ - \nu_i^{I_2}. \tag{1d}$$

Here, τ_Q is the time constant for each population Q, and h^Q is the synaptic drive to neurons in Q population for $Q \in \{E_1, E_2, I_1, I_2\}$, described as.

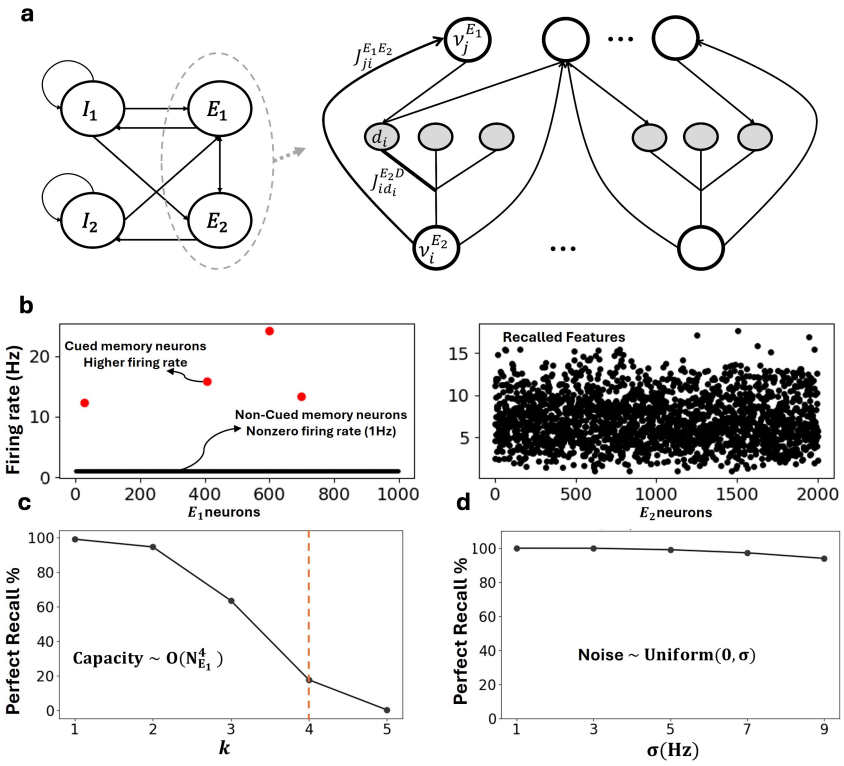

Figure 1: (a) The network architecture on the left, and its excitatory subnetwork on the right. (b) A sample activation of E_1 (memory) and E_2 (feature) neurons during memory recall . Red dots represent cued E_1 neurons that are actively engaged in recall, while black dots denote non-cued neurons, which remain inactive but are not artificially silenced. (c) The percentage of perfectly recalled memories as a function of the number of neurons encoding each memory, k. The network maintains high recall accuracy for $k \leq 4$, resulting in a capacity of $O(N_{E_1}^4)$. (d) The percentage of perfectly recalled memories as a function of the induced noise level in feature neurons.

$$h_i^{E_1} = \frac{1}{\sqrt{cN_{E_2}}} \sum_j^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2} \nu_j^{E_2} - \frac{1}{\sqrt{cN_{I_2}}} \sum_j^{N_{I_2}} c_{ij}^{E_1I_2} \tilde{J}_{ij}^{E_1I_2} \nu_j^{I_2} - \frac{1}{\sqrt{cN_{I_1}}} \sum_j^{N_{I_1}} c_{ij}^{E_1I_1} \tilde{J}_{ij}^{E_1I_1} \nu_j^{I_1} + h_{i,ext}^{E_1}, \tag{2a}$$

$$h_{d_i}^{E_2} = \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}} \sum_j^{N_{E_1}} c_{d_ij}^{DE_1} J_{d_ij}^{DE_1} \nu_j^{E_1} - \frac{1}{\sqrt{\frac{cN_{I_1}}{N_D}}} \sum_j^{N_{I_1}} c_{d_ij}^{DI_1} \tilde{J}_{d_ij}^{DI_1} \nu_j^{I_1} + h_{d_i,ext}^{E_2}, \tag{2b}$$

$$h_i^{I_1} = \frac{1}{cN_{E_1}} \sum_j^{N_{E_1}} c_{ij}^{I_1E_1} J_{ij}^{I_1E_1} \nu_j^{E_1} - \frac{1}{cN_{I_1}} \sum_j^{N_{I_1}} c_{ij}^{I_1I_1} \tilde{J}_{ij}^{I_1I_1} \nu_j^{I_1}, \tag{2c}$$

$$h_i^{I_2} = \frac{1}{cN_{E_2}} \sum_j^{N_{E_2}} c_{ij}^{I_2E_2} J_{ij}^{I_2E_2} \nu_j^{E_2} - \frac{1}{cN_{I_2}} \sum_j^{N_{I_2}} c_{ij}^{I_2I_2} \tilde{J}_{ij}^{I_2I_2} \nu_j^{I_2}. \tag{2d}$$

And, F, and G are nonlinearities, defined as follows:

$$G(z) = \begin{cases} g_{min} & \text{if } z \leq g_t \\ \alpha(z - g_t) + g_{min} & \text{if } 0 < \alpha(z - g_t) < g_{max} - g_{min} \\ g_{max} & \text{if } \alpha(z - g_t) \geq g_{max} - g_{min} \end{cases} \tag{3a}$$

$$F(z) = \begin{cases} r_{min} & \text{if } z \leq f_t \\ (z - f_t) + r_{min} & \text{if } 0 < (z - f_t) < r_{max} - r_{min} \\ r_{max} & \text{if } (z - f_t) \geq r_{max} - r_{min} \end{cases} \tag{3b}$$

Here, α is a positive constant. g_{min} and g_{max} represent the minimum and maximum dendritic currents, respectively, while r_{min} and r_{max} denote the minimum and maximum firing rates, which are 1 Hz and 100 Hz, respectively. The parameters g_t and f_t must be determined based on stability criteria to ensure stable fixed points in the dynamics (see Appendix). The threshold-linear operator $[z]^+$ is defined as $[z]^+ = z$ if $z > 0$ and 0 otherwise.

The selected nonlinearities play a crucial role in satisfying key biological constraints. Specifically, they ensure that non-cued neurons maintain a nonzero firing rate. The maximum firing rate is set high ($r_{\text{max}} = 100$ Hz), and the fixed points of the dynamics are stabilized within the linear regime. As a result, neurons remain in an unsaturated operating state even when they participate in the recall of a memory.

J_{ij}^{QR} represents the synaptic weight from the jth neuron in population R to the ith neuron in population Q. Similarly, $J_{d_{ij}}^{DR}$ denotes the synaptic weight between neuron j in population R and the dendritic branch d_i of neuron i in population E_2. The term $J_{id_i}^{E_2D}$ represents the dendritic weight of the d_ith branch of neuron i in the E_2 population.

All J_{ij}^{QR} are random and positive. Additionally, they are non-symmetric, except for the weights between E_1 and E_2 neurons, which are dependent, that is, if neuron i in population E_1 is connected to neuron j in population E_2 via branch ν_j, then

$$J_{ij}^{E_1E_2} = J_{j\nu_j}^{E_2D}. \tag{4}$$

Furthermore, all neurons in the network obey Dale's law.

Here, c_{ij}^{QR} determines whether neuron j in populations R is connected to neuron i in population Q defined as:

$$c_{ij}^{QR} = \begin{cases} 1 & \text{with probability } c \\ 0 & \text{with probability } 1 - c \end{cases} \tag{5}$$

Similarly, $c_{d_ij}^{DR}$ determines whether neuron j in populations R is connected to the dendritic branch d_i of neuron i in population E_2, for $R \in \{E_1, I_1\}$, and is defined as:

$$c_{ij}^{DR} = \begin{cases} 1 & \text{with probability } \frac{c}{N_D} \\ 0 & \text{with probability } 1 - \frac{c}{N_D} \end{cases} \tag{6}$$

The appropriate distribution and scaling of the weights ensure that when a memory is recalled, the average synaptic drive to the cued E_1 neurons is positive, while the drive to the non-cued E_1 neurons is negative. Similarly, a dendritic branch receiving input from a cued E_1 neuron experiences a net positive current, whereas a non-activated branch receives a negative. Additionally, the weights are scaled such that the variance of the synaptic drive to each neuron remains independent of the number of postsynaptic neurons (see Appendix).

3 RESULT

Each neuron in the E_2 population receives excitatory inputs from both cued and non-cued neurons in the E_1 population, which are allowed to have a low, nonzero firing rate. Additionally, E_2 neurons receive inhibitory inputs from I_1 neurons. If a dendritic branch does not receive a large input from a cued E_1 neuron, its total input remains very small and, therefore, it does not significantly affect the soma. However, branches connected to cued E_1 neurons receive strong positive input. Consequently, the activity of E_2 neurons is primarily determined by their connectivity with cued E_1 neurons, even in the presence of substantial noise from non-cued neurons.

Because the dependency in recurrent weights from E_2 to E_1 (Equation 4), the input to the cued E_1 neurons becomes more aligned with these recurrent weights, resulting in a stronger recurrent input. The input to the non-cued memory neurons, however, remains small, as the weights are not aligned with the input from E_2. This mechanism enables perfect recall of the cued memory: only the neurons encoding the cued memory have a high firing rate (red dots in Figure 1a), while the rest maintain a low firing rate (black dots in Figure 1b).

In the proposed model, each memory can be encoded in 1, 2, ..., or k neurons in the E_1 population, which k depends on network parameters. Therefore, the capacity—the number of stable fixed points—corresponds to the number of possible ways memories can be encoded in the memory layer, is

$$\text{capacity} \sim \beta_1 \binom{N_{E_1}}{1} + \cdots + \beta_k \binom{N_{E_1}}{k} \sim O(N_{E_1}^k),$$

where β_k is the percentage of memories that can be perfectly recalled when the memory is encoded in k neurons in the E_1 layer. Figure 1c shows that, for an example fully connected network with $N_{E_1} = 1000$ (memory neurons) and $N_{E_2} = 2000$ (feature neurons), the network is able to perfectly recall a large number of memories for k up to 4.

Beyond capacity, robustness—defined as the size of the basin of attraction for each fixed point (memory)—is also critical. Figure 1d shows that the fixed points of the dynamics have a large basin of attraction. Here, E_2 neurons are initialized with a noisy memory, i.e., features associated with a memory plus a noise $\sim$ Uniform$(0, \sigma)$. Given the large number of E_2 neurons, this noise level is substantial, further demonstrating the robustness of memory recall.

4 Discussion

In this paper, we propose a novel architecture for associative memory by augmenting Dense Associative Memory with nonlinear dendrites. Our model achieves two crucial properties: high-capacity storage with large basins of attraction, and biological plausibility. Unlike previous implementations of Dense Associative Memory models—which rely on a single active memory neuron per memory during recall—our approach allows for the simultaneous activation of multiple ($1 \leq k \ll N_D$) memory neurons. This distributed representation enables each memory neuron to participate in encoding multiple patterns, resulting in an efficient storage capacity of $\sim O(N_{E_1}^k)$. This marks a significant improvement over earlier implementations, which lacked such an efficient, distributed memory framework.

To address a key biological constraint—the unsilencing of non-cued neurons during memory recall—we leverage dendritic computations, a fundamental property of real neurons. In our model, dendritic nonlinearities filter out noise from non-cued memory neurons before it reaches the soma, preventing interference while allowing these neurons to remain active. This mechanism aligns with experimental observations that large groups of neurons rarely remain completely silent for extended periods. Additionally, our model incorporates other crucial biological features. Specifically, it ensures the presence of both excitatory and inhibitory neurons, adherence to Dale's law, operation at non-saturated firing rates, and robustness of capacity to sparse connectivity.

Our proposed model establishes a robust foundation for associative memory networks, offering a biologically plausible framework with polynomial capacity. However, several open questions remain. Are synaptic weights in the memory network relatively stable, or do they change dynamically over time as new memories are added? If the weights are fixed, how does the brain associate different memories with these fixed points? Conversely, if the weights are dynamic, what learning rule governs their updates? Addressing these questions requires further investigation into the neural circuits that support attractor networks.

REFERENCES

Sarthak Chandra, Sugandha Sharma, Rishidev Chaudhuri, and Ila Fiete. Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature*, 2025.

Leo Kozachkov, Jean-Jacques Slotine, and Dmitry Krotov. Neuron-astrocyte associative memory. *arXiv preprint arXiv:2311.08135*, 2023.

Dmitry Krotov. Hierarchical associative memory. *arXiv preprint arXiv:2107.06446*, 2021.

Dmitry Krotov and John J Hopfield. Dense associative memory for pattern recognition. *Advances in neural information processing systems*, 29, 2016.

Dmitry Krotov and John J Hopfield. Large associative memory problem in neurobiology and machine learning. In *International Conference on Learning Representations*, 2021.

Panayiota Poirazi, Terrence Brannon, and Bartlett W Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37:989–999, 2003.

Yasser Roudi and Peter E Latham. A balanced memory network. *PLoS computational biology*, 3:e141, 2007.

Tom J Wills, Colin Lever, Francesca Cacucci, Neil Burgess, and John O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308:873–876, 2005.

A APPENDIX

In this section, we provide details on the stability analysis of the memories.

The firing rate dynamics of neurons in each population are described by Equation 1. Without loss of generality, we set $h^{E_1}_{i,\text{ext}}$ and $h^{E_2}_{d_i,\text{ext}}$ to zero for the remainder of the analysis, and we assume memories are encoded in a single memory neuron. The same reasoning applies when memories are encoded in $1 < k \ll N_D$ neurons in the E_1 population.

The weights are scaled such that when a memory is recalled, the average synaptic drive to the cued E_1 neurons is positive, while the drive to the non-cued E_1 neurons is negative and nearly balanced. Similarly, a dendritic branch receiving input from a cued E_1 neuron experiences a net positive current, whereas a non-activated branch receives a negative and nearly balanced input drive. Furthermore, the weights are scaled so that the variance of the synaptic drive to each neuron remains independent of the number of postsynaptic neurons. We will further demonstrate this point later. The required scaling are therefore as follows,

$$\tilde{J}^{E_1 I_2}_{ij} \sim \sqrt{\frac{N_{E_2}}{N_{I_2}}} J^{E_1 I_2}_{ij} \tag{7a}$$

$$\tilde{J}^{E_1 I_1}_{ij} \sim \sqrt{\frac{N_{E_2}}{N_{I_1}}} J^{E_1 I_1}_{ij} \tag{7b}$$

$$\tilde{J}^{D I_1}_{d_i j} \sim \sqrt{\frac{N_{E_1}}{N_{I_1}}} J^{D I_1}_{d_i j} \tag{7c}$$

$$\tilde{J}^{I_1 I_1}_{ij} \sim \frac{1}{cN_{I_1}} J^{I_1 I_1}_{ij} \tag{7d}$$

$$\tilde{J}^{I_2 I_2}_{ij} \sim \frac{1}{cN_{I_2}} J^{I_2 I_2}_{ij} \tag{7e}$$

All J^{QR}_{ij} are assumed to be random and positive, drawn from a uniform distribution with mean J. Additionally, they are non-symmetric, except for $J^{E_1 E_2}_{ij}$, which are dependent on dendritic weights, as described in 4 and are drawn from a gamma distribution.

krBy substituting the scaled weights, the synaptic drive to the inhibitory neurons can be approximated as:

$$\begin{aligned} h^{I_1}_i &= \frac{1}{cN_{E_1}} \sum_j^{N_{E_1}} c^{I_1 E_1}_{ij} J^{I_1 E_1}_{ij} \nu^{E_1}_j - \frac{1}{cN_{I_1}} \sum_j^{I_1} c^{I_1 I_1}_{ij} \tilde{J}^{I_1 I_1}_{ij} \nu^{I_1}_j \\ &\approx \overline{c^{I_1 E_1}_{ij} J^{I_1 E_1}_{ij} \nu^{E_1}_j} + O\left(\frac{1}{\sqrt{cN_{E_1}}}\right) - \overline{c^{I_1 I_1}_{ij} \frac{1}{cN_{I_1}} J^{I_1 I_1}_{ij} \nu^{I_1}_j} - O\left(\frac{1}{cN_{I_1}\sqrt{cN_{I_1}}}\right) \\ &\approx cJ\overline{\nu^{E_1}_j}, \end{aligned} \tag{8a}$$

$$\begin{aligned} h^{I_2}_i &= \frac{1}{cN_{E_2}} \sum_j^{N_{E_2}} c^{I_2 E_2}_{ij} J^{I_2 E_2}_{ij} \nu^{E_2}_j - \frac{1}{cN_{I_2}} \sum_j^{I_2} c^{I_2 I_2}_{ij} \tilde{J}^{I_2 I_2}_{ij} \nu^{I_2}_j \\ &\approx \overline{c^{I_2 E_2}_{ij} J^{I_2 E_2}_{ij} \nu^{E_2}_j} + O\left(\frac{1}{\sqrt{cN_{E_2}}}\right) - \overline{c^{I_2 I_2}_{ij} \frac{1}{cN_{I_2}} J^{I_2 I_2}_{ij} \nu^{I_2}_j} - O\left(\frac{1}{cN_{I_2}\sqrt{cN_{I_2}}}\right) \\ &\approx cJ\overline{\nu^{E_2}_j}. \end{aligned} \tag{8b}$$

And by assuming that the inhibitory neurons have a faster dynamics compare to the excitatory neurons, we can reduce the four dimensional equation 25 to a two dimensional equation:

$$\tau_{E_1}\frac{d\nu_i^{E_1}}{dt} = F\left(h_i^{E_1}\right) - \nu_i^{E_1}, \tag{9a}$$

$$\tau_{E_2}\frac{d\nu_i^{E_2}}{dt} = \frac{1}{\sqrt{N_D}}\sum_{d_i}^{N_D} J_{id_i}^{E_2 D} G\left(h_{d_i}^{E_2}\right) - \nu_i^{E_2}, \tag{9b}$$

$$\nu_i^{I_1} = \left[h_i^{I_1}\right]^+ \approx cJ\overline{\nu_j^{E_1}}, \tag{9c}$$

$$\nu_i^{I_2} = \left[h_i^{I_2}\right]^+ \approx cJ\overline{\nu_j^{E_2}}. \tag{9d}$$

And the synaptic drives could be approximated as:

$$h_i^{E_1} = \frac{1}{\sqrt{cN_{E_2}}}\sum_j^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2}\nu_j^{E_2} - \frac{1}{\sqrt{cN_{I_2}}}\sum_j^{N_{I_2}} c_{ij}^{E_1I_2}\tilde{J}_{ij}^{E_1I_2}(cJ\overline{\nu_j^{E_2}}) - \frac{1}{\sqrt{cN_{I_1}}}\sum_j^{N_{I_1}} c_{ij}^{E_1I_1}\tilde{J}_{ij}^{E_1I_1}(cJ\overline{\nu_j^{E_1}}), \tag{10a}$$

$$h_{d_i}^{E_2} = \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}}\sum_j^{N_{E_1}} c_{d_ij}^{DE_1} J_{d_ij}^{DE_1}\nu_j^{E_1} - \frac{1}{\sqrt{\frac{cN_{I_1}}{N_D}}}\sum_j^{N_{I_1}} c_{d_ij}^{DI_1}\tilde{J}_{d_ij}^{DI_1}(cJ\overline{\nu_j^{E_1}}). \tag{10b}$$

by substituting the scaled weights we will get:

$$h_i^{E_1} = \frac{1}{\sqrt{cN_{E_2}}}\sum_j^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2}\nu_j^{E_2} - \frac{1}{\sqrt{cN_{I_2}}}\sum_j^{N_{I_2}} c_{ij}^{E_1I_2}\left(\sqrt{\frac{cN_{E_2}}{cN_{I_2}}}\left\langle J_{ij}^{E_1E_2}\right\rangle_{i,j} J_{ij}^{E_1I_2}\right)(cJ\overline{\nu_j^{E_2}}) - \frac{1}{\sqrt{cN_{I_1}}}\sum_j^{N_{I_1}} c_{ij}^{E_1I_1}\left(\sqrt{\frac{cN_{E_2}}{cN_{I_1}}}J_{ij}^{E_1I_1}\right)(cJ\overline{\nu_j^{E_1}}), \tag{11a}$$

$$h_{d_i}^{E_2} = \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}}\sum_j^{N_{E_1}} c_{d_ij}^{DE_1} J_{d_ij}^{DE_1}\nu_j^{E_1} - \frac{1}{\sqrt{\frac{cN_{I_1}}{N_D}}}\sum_j^{N_{I_1}} c_{d_ij}^{DI_1}\left(\sqrt{\frac{cN_{E_1}}{cN_{I_1}}}J_{d_ij}^{DI_1}\right)(cJ\overline{\nu_j^{E_1}}). \tag{11b}$$

And , the synaptic drives will be simplified as:

$$\begin{aligned} h_i^{E_1} &= \frac{1}{\sqrt{cN_{E_2}}}\sum_j^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2}\nu_j^{E_2} - cJ^2\sqrt{cN_{E_2}}\left\langle J_{ij}^{E_1E_2}\right\rangle_{i,j}\overline{\nu_j^{E_2}} - O\left(\sqrt{\frac{N_{E_2}}{N_{I_2}}}\right) - \\ &\quad cJ^2\sqrt{cN_{E_2}}\overline{\nu_j^{E_1}} - O\left(\sqrt{\frac{N_{E_2}}{N_{I_1}}}\right) \\ &\sim \frac{1}{\sqrt{cN_{E_2}}}\sum_j^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2}\nu_j^{E_2} - cJ^2\sqrt{cN_{E_2}}\left(\left\langle J_{ij}^{E_1E_2}\right\rangle_{i,j}\overline{\nu_j^{E_2}} + \overline{\nu_j^{E_1}}\right), \end{aligned} \tag{12a}$$

$$h_{d_i}^{E_2} = \frac{1}{\sqrt{\frac{cN_{E_1}}{D}}}\sum_j^{N_{E_1}} c_{d_ij}^{DE_1} J_{d_ij}^{DE_1}\nu_j^{E_1} - cJ^2\sqrt{\frac{cN_{N_{E_1}}}{N_D}}\overline{\nu_j^{E_1}} - O\left(\sqrt{\frac{N_{E_1}}{N_{I_1}}}\right). \tag{12b}$$

Now we want to ask whether and under what condition there are Equibliria in the network associated with stored memories such that when a memory is perfectly recalled, one neuron in E_1 population is highly active $\nu_\mu^{E_1} = \nu_\mu^{*E_1}$, where $r_{min} \ll \nu_\mu^{*E_1} \ll r_{max}$, all other E_1 neurons remain are at a low activity equal to $\nu_{j\neq\mu}^{E_1} = r_{min}$, for $j \neq \mu \in (1, 2, \ldots, N_{E_1})$. In this case, $\overline{\nu_j^{E_1}} \approx r_{min}$. For simplicity, for the rest of analysis we set $r_{min} = 1$, and $J = 1$. if neuron μ in E_1 is connected to neuron i in E_2 via branch ν_i,

then the synaptic drive to the activated branch ν_i, and other will be

$$h_{\nu_i}^{E_2} = \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}} c_{\nu_i\mu}^{N_D E_1} J_{\nu_i\mu}^{DE_1} \nu_\mu^{*E_1} + \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}} \sum_{j\neq\mu}^{N_{E_1}} c_{\nu_i j}^{DE_1} J_{\nu_i j}^{DE_1} - c\sqrt{\frac{cN_{E_1}}{N_D}} - O(\sqrt{\frac{N_{E_1}}{N_{I_1}}}) \approx$$

$$\frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}} c_{\nu_i\mu}^{DE_1} J_{\nu_i\mu}^{DE_1} \nu_\mu^{*E_1} + \sqrt{\frac{cN_{E_1}}{N_D}} + O(1) - c\sqrt{\frac{cN_{E_1}}{N_D}} - O(\sqrt{\frac{N_{E_1}}{N_{I_1}}}). \tag{13a}$$

$$h_{\nu_i}^{E_2} \approx \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}} c_{\nu_i\mu}^{DE_1} J_{\nu_i\mu}^{DE_1} \nu_\mu^{*E_1} - O(\sqrt{\frac{N_{E_1}}{N_{I_1}}}) + (1-c)\sqrt{\frac{cN_{E_1}}{N_D}} \tag{13b}$$

$$h_{d_i}^{E_2} = \frac{1}{\sqrt{\frac{cN_{E_1}}{N_D}}} \sum_{j\neq\mu}^{N_{E_1}} c_{d_i j}^{DE_1} J_{d_i j}^{DE_1} - c\sqrt{\frac{cN_{E_1}}{N_D}} - O(\sqrt{\frac{N_{E_1}}{N_{I_1}}}) \approx \sqrt{\frac{cN_{E_1}}{N_D}} + O(1) - c\sqrt{\frac{cN_{E_1}}{N_D}} - O(\sqrt{\frac{N_{E_1}}{N_{I_1}}}). \tag{13c}$$

$$h_{d_i}^{E_2} \approx -O(\sqrt{\frac{N_{E_1}}{N_{I_1}}}) + (1-c)\sqrt{\frac{cN_{E_1}}{N_D}} \tag{13d}$$

With an appropriate choice of g_t in Equation 3, we can ensure that in the steady state, only the active branch receives a large input, while the remaining branches receive small inputs and do not significantly affect the soma,

$$\alpha(h_{\nu_i}^{E_2} - g_t) > 0 \Rightarrow G(h_{\nu_i}^{E_2}) = c_{\nu_i\mu}^{DE_1} g_{max} + (1 - c_{\nu_i\mu}^{DE_1}) g_{min} \tag{14a}$$

$$h_{d_i}^{E_2} - g_t < 0 \Rightarrow G(h_{d_i}^{E_2}) = g_{min}. \tag{14b}$$

Therefore in the steady state,

$$\nu_i^{*E_2} = \frac{1}{\sqrt{N_D}} J_{i\nu_i}^{E_2 D} \left(c_{\nu_i\mu}^{DE_1} g_{max} + (1 - c_{\nu_i\mu}^{DE_1}) g_{min} \right) + \frac{1}{\sqrt{N_D}} \sum_{d_i}^{D} J_{id_i}^{E_2 D} g_{min}, \tag{15}$$

Now, let's consider the excitatory input to the cued neuron μ, the first term in equation12a, remembering that, $J_{\mu i}^{E_1 E_2} = J_{i\nu_i}^{E_2 D}$, Equation 4,

$$\frac{1}{\sqrt{cN_{E_2}}} \sum_{j}^{N_{E_2}} c_{\mu j}^{E_1 E_2} J_{\mu j}^{E_1 E_2} \nu_j^{E_2} \approx \frac{1}{\sqrt{cN_D N_{E_2}}} \sum_{j}^{N_{E_2}} c_{\mu j}^{E_1 E_2} J_{\mu j}^{E_1 E_2} \left(J_{j\nu_j}^{E_2 D} c_{\nu_j\mu}^{DE_1} g_{max} \right)$$
$$+ \frac{1}{\sqrt{cN_D N_{E_2}}} \sum_{j}^{N_{E_2}} c_{\mu j}^{E_1 E_2} J_{\mu j}^{E_1 E_2} \left(\sum_{d_j}^{N_D} J_{j d_j}^{E_2 D} g_{min} \right), \tag{16}$$

$$= \frac{1}{\sqrt{cN_D N_{E_2}}} \sum_{j}^{N_{E_2}} J_{\mu j}^{E_1 E_2} \left(J_{\mu j}^{E_1 E_2} c_{\nu_j\mu}^{DE_1} g_{max} \right)$$
$$+ \frac{1}{\sqrt{cN_D N_{E_2}}} \sum_{j}^{N_{E_2}} c_{\mu j}^{E_1 E_2} J_{\mu j}^{E_1 E_2} \left(\sum_{d_j}^{N_D} J_{j d_j}^{E_2 D} g_{min} \right), \tag{17}$$

$$\approx \left\langle \left(J_{ij}^{E_1 E_2} \right)^2 \right\rangle_{i,j} \frac{\sqrt{cN_{E_2}}}{N_D \sqrt{N_D}} g_{max} + O\left(\frac{1}{N_D} \right) +$$
$$\left\langle J_{ij}^{E_1 E_2} \right\rangle_{i,j} \sqrt{cN_D N_{E_2}} g_{min} + O(1). \tag{18}$$

$$\approx \left\langle \left(J_{ij}^{E_1 E_2} \right)^2 \right\rangle_{i,j} \frac{\sqrt{cN_{E_2}}}{N_D \sqrt{N_D}} g_{max} + \left\langle J_{ij}^{E_1 E_2} \right\rangle_{i,j} \sqrt{cN_D N_{E_2}} g_{min} \tag{19}$$

Following the same approximation, the excitatory input to the non-cued neuron $i \neq \mu$ is:

$$\frac{1}{\sqrt{cN_{E_2}}} \sum_{j}^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2} \nu_j^{E_2} \approx \left(\left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \right)^2 \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + O\left(\frac{1}{N_D}\right) + \left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \sqrt{cN_D N_{E_2}} g_{min} + O(1). \tag{20}$$

$$\frac{1}{\sqrt{cN_{E_2}}} \sum_{j}^{N_{E_2}} c_{ij}^{E_1E_2} J_{ij}^{E_1E_2} \nu_j^{E_2} \approx \left(\left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \right)^2 \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + \left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \sqrt{cN_D N_{E_2}} g_{min} \tag{21}$$

By defining, (See Equation 12a and Equation 16):

$$\theta = \left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \sqrt{cN_D N_{E_2}} g_{min} - cJ^2 \sqrt{cN_{E_2}} \left(\left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \overline{\nu_j^{E_2}} + \overline{\nu_j^{E_1}} \right) \tag{22}$$

$$= \sqrt{cN_{E_2}} \left(-cJ^2 \left(\left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \overline{\nu_j^{E_2}} + \overline{\nu_j^{E_1}} \right) + \left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \sqrt{N_D} g_{min} \right) \tag{23}$$

we can finally express the synaptic drive in equation 12a to the cued and non-cued neurons as:

$$h_\mu^{E_1} \approx \left\langle \left(J_{ij}^{E_1E_2} \right)^2 \right\rangle_{i,j} \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + \theta, \tag{24a}$$

$$h_{i\neq\mu}^{E_1} \approx \left(\left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \right)^2 \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + \theta. \tag{24b}$$

Finally chosing the right distribution over $J_{ij}^{E_1E_2}$ and setting the appropriate f_t for the F nonlinearity in equation 3b, we can make sure that there is an equiblirium such that:

$$\nu_\mu^{*E_1} = F\left(\left\langle \left(J_{ij}^{E_1E_2} \right)^2 \right\rangle_{i,j} \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + \theta \right) = \left\langle \left(J_{ij}^{E_1E_2} \right)^2 \right\rangle_{i,j} \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + \theta - r_t + r_{min}, \tag{25a}$$

$$\nu_{i\neq\mu}^{*E_1} = F\left(\left(\left\langle J_{ij}^{E_1E_2} \right\rangle_{i,j} \right)^2 \frac{\sqrt{cN_{E_2}}}{N_D\sqrt{N_D}} g_{max} + \theta \right) = r_{min}. \tag{25b}$$

And since around the equilibrium,

$$\frac{\partial G(h_{d_i}^{E_2})}{\partial h_{d_i}^{E_2}} = 0, \quad \frac{\partial F(h_{i\neq\mu}^{E_1})}{\partial h_i^{E_1}} = 0, \quad \frac{\partial F(h_\mu^{E_1})}{\partial h_\mu^{E_1}} = 1, \quad \text{and} \quad \frac{\partial h_\mu^{E_1}}{\partial \nu_\mu^{E_1}} < 0, \tag{26}$$

the equilibria are stable.