Error-correcting columnar networks: high-capacity memory under sparse connectivity

Haozhe Shan^{1,2} **Ludovica Bachschmid Romano**^{3} **Haim Sompolinsky**^{1,4}
^{1}Center for Brain Science, Harvard University
^{2}Program in Neuroscience, Harvard Medical School
^{3}Department of Neurobiology, Duke University
^{4}Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem

Abstract

Neurons with recurrent connectivity can store memory patterns as attractor states in their dynamics [1], forming a plausible basis for associative memory in the brain. Classic theoretical results on fully connected recurrent neural networks (RNNs) with binary neurons and Hebbian learning rules state that they can store at most $O(N)$ memories [2, 3], where N is the number of neurons. However, under the physiological constraint that neurons are sparsely connected, this capacity is dramatically reduced to $O(K)$ [4, 5, 6, 7], where K is the average degree of connectivity (estimated to be $O(10^3 \sim 10^4)$ in the mammalian neocortex [8]). This reduced capacity is orders-of-magnitude smaller than experimental estimates of human memory capacity. In this work, we propose the error-correcting columnar network (ECCN) as a plausible model of how the brain realizes high-capacity memory storage despite sparse connectivity. In the ECCN, neurons are organized into "columns": in each memory, neurons from the same column encode the same feature(s), similar to columns in primary sensory areas [9]. A column-synchronizing mechanism utilizes the redundancy of columnar codes to perform error correction. We analytically computed the memory capacity of the ECCN via a dynamical mean-field theory. The results show that for a fixed column size M, the capacity grows linearly with network size N until it saturates at $\propto MK$. For optimal choice of M for each N, the capacity is $\propto \sqrt{NK}$.

1 Introduction

RNNs with binary neurons provide concrete and well-studied models of how the brain encodes long-term memories in synaptic connectivity [1]. During memory retrieval, the network starts near a memory state and converges to a fixed point corresponding to the (approximately) true memory. An important quantity is the maximal number of memories that can be stored in the plastic synaptic weights such that they are (approximate) fixed points of the dynamics. We will call this number the network's memory capacity, denoted P_c. In fully connected networks, this number depends on the learning rule but generally scales linearly with the network size N. However, in sparsely connected networks where each neuron is connected to only (random) K neighbors, previous studies have found that P_c scales linearly with K instead [4, 5, 6, 7]. In human cortex, K is at most $O\left(10^4\right)$ (on average, each presynaptic neuron makes 30,000 synapses [8] and many of them likely connect to the same postsynaptic target as well as to subcortical targets). Thus, these models would predict a human memory capacity of $O\left(10^3 \sim 10^4\right)$, substantially lower than experimental estimates. For example, the English vocabulary of a college graduate can be of size $O\left(10^5\right)$ [10] (The vocabulary size in modern large language models is $O(10^4 - 10^5)$ [11]). In vision, human recognition memory does not show signs of saturation after learning 10^4 images [12, 13]. Altogether, it is reasonable to assume

Associative Memory & Hopfield Networks in 2023. NeurIPS 2023 workshop.

that the number of linguistic and non-linguistic tokens in human long-term memory (which we call here the "cortical dictionary") is at least $O(10^5)$, which is apparently inconsistent with the sparseness of connectivity.

This discrepancy could be resolved by splitting the dictionary into many parts, each stored in a subnetwork. However, such schemes would require complex learning rules and retrieval mechanisms. Here, we propose and analyze the error-correcting columnar network (ECCN), which contains a simple and biologically plausible mechanism that recovers high memory capacity in sparsely connected networks. Neurons in the ECCN are organized into groups ("columns"); each column of neurons collectively acts as one computational unit representing one bit of memory. This organization, inspired by columnar structures in cortex [9], is in itself insufficient for overcoming sparse connectivity (further details below). To leverage the within-column redundancy, the ECCN performs within-column error correction by a fast synchronization mechanism. We compute the memory capacity of the resultant model by building a dynamical mean-field theory using the path-integral formulation [14] and show that the ECCN has memory capacity that is much larger than the degree of connectivity and is consistent with the connectivity and size of circuits in human cortex.

Related work RNNs where neurons are organized into columns have been studied before [15, 5]. Most relevantly, [5] studied an RNN of threshold-linear neurons with a columnar organization for storing memories with a hierarchical structure (each global memory is a concatenation of sub-memories, each of which is stored in a different column). Their model does not include any error-correcting mechanism and achieves a memory capacity at most linear in the degree of connectivity. [15] simulated an RNN of Hodgkin-Huxley neurons organized into columns and showed that it exhibits biologically realistic physiological properties and did not study its capacity. Therefore, to our best knowledge, our results provide the first computational basis for storing a large number of memories (more than linear in K) under sparse connectivity.

2 ECCNs: architecture, learning and dynamics

2.1 Columnar organization of neurons

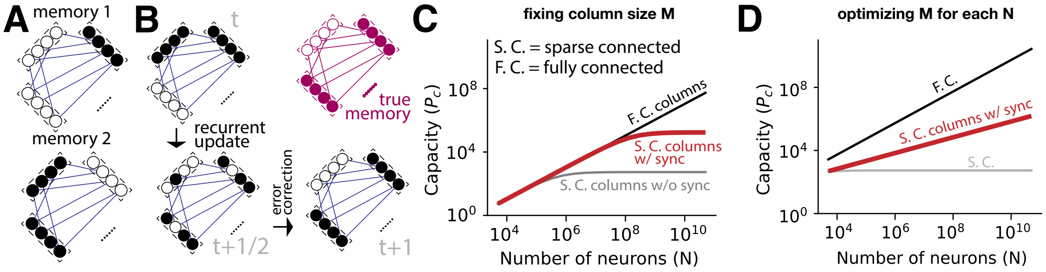

Figure 1: **A** Schematics of the ECCN. N sparsely connected neurons are divided into G columns of size M. Neurons from each column are all on or all off in each memory. **B** Schematics of network dynamics. In each iteration, the network first performs a recurrent update before doing an error-correction step. **C** Memory capacity of the fully connected model (black), the sparsely connected model without synchronization (gray), and the sparsely connected model with synchronization (ECCN; red). In all three, G is being varied while $M = 500, f = 0.1$ stay fixed. $K = 1000$ for sparsely connected models and $K = N$ in the fully connected model. **D** Same as C, but for every value of N, a different column size M that maximizes the capacity is used (for models without sync., optimal M is 1; for the model with sync. (ECCN) it is $O(\sqrt{N/K})$).

We consider an RNN with N binary neurons divided into G columns of size M ($GM = N$) and denote the activity of the i-th neuron in the g-th column at time t as $s_i^g(t) \in \{0, 1\}$ [3], $i = 1, ..., M$ and $g = 1, ..., G$ (Fig. 1A) [1]. The network stores P memories with a corresponding columnar structure: Each memory pattern, indexed by $\mu = 1, .., P$, is a $G-$dim binary vector,$\vec{x^\mu}$, where x_g^μ denotes the activity of neurons in the g-th column in the μ-th memory (Fig. 1A). Importantly, all neurons in a given memory and column share the same activity. The memories have a low level of activity [3]: $\{x_g^\mu\}_{\mu=1,...,P;g=1,...,G}$ are i.i.d. Bernoulli variables with $p(x_g^\mu = 1) = f$ and $f \ll 1$.

---

[1] While not discussed further here, our analysis and qualitative conclusions also apply in the case of ±1 neurons [1].

The synaptic connection between neuron i from column g and neuron j from column g', $J_{ij}^{g,g'}$, is given by the Hebb rule with random dilution [6]

$$J_{ij}^{g,g'} = \frac{1}{f(1-f)K} C_{ij}^{g,g'} \sum_{\mu}^{P} \delta x_g^\mu \delta x_{g'}^\mu ,\ g \neq g' \tag{1}$$

where $\delta x_g^\mu \equiv x_g^\mu - f$ and $\{C_{ij}^{g,g'}\}_{i,j=1,\ldots,M;g,g'=1,\ldots,G}$ are i.i.d. Bernoulli variables with $p(C_{ij}^{g,g'} = 1) = K/N$ such that, on average, every neuron is connected to K others.

The columnar organization by itself is insufficient for raising the capacity beyond $O(K)$. To see this, consider now a network state where all neurons are in memory pattern 1. The local field for the i-th neuron in column g is given by

$$h_g^i = \sum_{j=1}^{M} \sum_{g' \neq g}^{G} J_{ij}^{g,g'} x_{g'}^1 - \theta \tag{2}$$

$$= \delta x_g^1 + \frac{1}{f\left(1-f\right)G} \sum_{g' \neq g, \mu > 1} \delta x_g^\mu \delta x_{g'}^\mu x_{g'}^1 + \frac{1}{f\left(1-f\right)K} \sum_{j, g' \neq g} \delta C_{ij}^{g,g'} \sum_{\mu} \delta x_g^\mu \delta x_{g'}^\mu x_{g'}^1 - \theta, \tag{3}$$

where $\delta C_{ij}^{g,g'} \equiv C_{ij}^{g,g'} - K/N$. The first term constitutes the $O(1)$ signal; the second term is the noise coming from interference from the other $\mu - 1$ patterns; the last term is the noise coming from connectivity dilution. At the sparse connectivity limit ($K \ll N$), interference noise has variance Pf/G and dilution noise has variance Pf/K. These expressions show that, regardless of M, the dilution noise overpowers the signal when $P \gg K/f$, destabilizing the memory state and causing retrieval failure. Thus, the capacity still scales with K and additional dynamics are needed to take advantage of the columnar redundancy. In particular, we note that while the signal and the interference noise are identical across neurons in each column, the dilution noise depends on index i and is uncorrelated among neurons (as $\delta C_{ij}^{g,g'}$ is i.i.d.). Thus, in sparsely connected networks where the dilution noise dominates, it is possible to improve the signal-to-noise via within-column averaging, as done in the ECCN.

2.2 Error-correcting dynamics

In the ECCN, each iteration in the discrete-time dynamics consists of two steps,

$$\text{recurrent update:} \quad s_i^g(t+1/2) = \Theta\left(\sum_{j=1}^{M} \sum_{g' \neq g}^{G} J_{ij}^{g,g'} s_j^{g'}(t) - \theta \right) \tag{4}$$

$$\text{error-correcting synchronization:} \quad s_i^g(t+1) = \Theta\left(M^{-1} \sum_{j=1}^{M} s_j^g(t+1/2) - T \right) \tag{5}$$

where Θ is the Heaviside step function, and θ and T are global thresholds (see below regarding their choice). The first step is the standard recurrent update (which can be asynchronous); the second step is an error-correcting synchronization mechanism that updates all neurons within each column (Fig. 1B) by a "majority" vote. Biologically, this could be implemented by thalamo-cortical loops that realize fast synchrony among neurons in cortex [16]. Note that after the synchronization step, $\forall i : s_i^g(t+1) = s^g(t+1)$.

2.3 Dynamical mean-field theory

We next compute the memory capacity of the model. Memory μ is considered successfully recalled if one initializes the network near this memory state and the network converges to a state highly similar to the memory state, with similarity measured by the overlap

$$m^\mu(t) = N^{-1} \sum_{i,g} \frac{1}{f(1-f)} s_i^g(t) \delta x_g^\mu, \tag{6}$$

which is 1 if s is identical to the μ-th memory and zero if it is uncorrelated with it. Unlike classic RNN models of associative memory, our model does not have an obvious energy function due to the error-correcting dynamics. We thus analyzed the capacity with a dynamical mean-field theory, exact in the limits of large N, P, Kf. The mean-field equations read

$$s_i^g(t+1/2) = \Theta\left(m^\mu(t)\delta x_g^\mu + \eta^g(t) + z_i^g(t) - \theta\right) \tag{7}$$

$$m^\mu(t+1) = \frac{1}{f(1-f)}\langle \delta x_g^\mu s^g(t+1)\rangle_{\eta^g(t),\{z_i^g(t)\},x_g^\mu}. \tag{8}$$

where $\eta^g(t)$, $z_i^g(t)$ are zero-mean Gaussian noise with variance Pf/G and Pf/K, respectively. $\eta^g(t)$ represents the interference noise; $z_i^g(t)$ is uncorrelated among neurons, representing the dilution noise. Substituting in Eq. 5 and computing $m^\mu(t+1)$ involve averaging $s^g(t+1)$ over both $\eta^g(t)$ as well as the residual noise induced by $z_i^g(t)$ after within-column averaging. The second averaging can be performed analytically if we assume that the network is sparse enough ($K \ll G$) such that $z_i^g(t)$ is the dominating fluctuation, resulting in

$$\forall i : s_i^g(t+1) = \Theta(m^\mu(t)\delta x_g^\mu + \epsilon^g(t) - T_{\text{eff}}) \tag{9}$$

where T_{eff} depends on θ, T (for convenience, we choose $T = 0.5$ in which case $T_{\text{eff}} = \theta$). $\epsilon^g(t)$ is zero-mean Gaussian noise with variance

$$\langle(\epsilon^g(t))^2\rangle = \frac{Pf}{G}\left(1 + \frac{\pi G}{2KM}\right). \tag{10}$$

The first term is the contribution from $\eta^g(t)$. The second is the residual noise from $z_i^g(t)$; compared to the variance of $z_i^g(t)$, it has been reduced by a factor M because of the synchronization.

3 Results

Computing the capacity from Eqs 9 10 is analogous to the analysis for the fully connected model [3, 17]. The ECCN capacity is given by the condition $(\epsilon^g(t))^2 = \theta^2/(2|\log f|)$, resulting in

$$P_c = \frac{\theta^2 G}{2f\,|\log f|\left(1 + \frac{\pi G}{2KM}\right)}. \tag{11}$$

This result implies that (1) for a fixed column size M, capacity first grows linearly with the number of neurons N but saturates at a value $\propto MK/f$ (Fig. 1C, red line); (2) for a given N larger than K , the optimal size of a column is $M \propto \sqrt{N/K}$, which results in a capacity growing as $P_c \propto \sqrt{NK}/f$ (Fig. 1D, red line; simulations in Fig. 2). The capacity can be several orders of magnitude greater than a standard sparse model (gray lines in Fig. 1C, D). As a concrete example, to achieve memory capacity of 10^5 with connectivity sparsity $K = 10^3$, activity level $f = 10^{-1}$, and threshold $\theta = 0.5$, N should be approximately 2×10^8 or larger (Fig. 1D), not unreasonable for the human brain (with more than 10^{10} cortical neurons [18]). Each of these patterns consists of $\sim 10^4$ bits. Under these parameters, the standard model with sparse connectivity would have a capacity of at most ≈ 543 regardless of N.

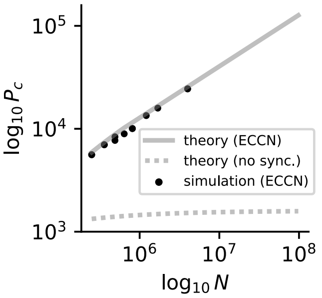

Figure 2: Theoretical capacity is validated by simulations. Here, N is increased as M is adjusted to be at the optimal size ($\propto \sqrt{N/K}$) for each N. The theory predicts that P_c increases as $\sqrt{N}$ in this case, which we observe. To make simulations feasible, we take $f = 0.02$ in this figure.

4 Discussion

We show that high-capacity memory can be realized in sparsely connected networks by taking advantage of redundancy in neural codes using a local error-correcting mechanism. Our model provides a possible explanation of the high memory capacity of the human brain, particularly in language, despite the sparseness of connectivity. The theory also suggests that the grouping of neurons into different feature encoders in a given neural memory system could be quantitatively linked to this system's memory demand and connectivity sparseness.

Acknowledgements This research is supported by the Gatsby Charitable Foundation, the Swartz Foundation, and ONR grant No.N0014-23-1-2051.

References

[1] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[2] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55(14):1530, 1985.

[3] Mikhail V Tsodyks and Mikhail V Feigel'man. The enhanced storage capacity in neural networks with low activity level. *EPL (Europhysics Letters)*, 6(2):101, 1988.

[4] Haim Sompolinsky. The theory of neural networks: The hebb rule and beyond. In *Heidelberg colloquium on glassy dynamics*, pages 485–527. Springer, 1987.

[5] Dominic O'Kane and Alessandro Treves. Short-and long-range connections in autoassociative memory. *Journal of Physics A: Mathematical and General*, 25(19):5055, 1992.

[6] MV Tsodyks. Associative memory in asymmetric diluted network with low level of activity. *EPL (Europhysics Letters)*, 7(3):203, 1988.

[7] Bernard Derrida, Elizabeth Gardner, and Anne Zippelius. An exactly solvable asymmetric neural network model. *EPL (Europhysics Letters)*, 4(2):167, 1987.

[8] Javier DeFelipe. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Frontiers in neuroanatomy*, page 29, 2011.

[9] David H Hubel, Torsten N Wiesel, and Michael P Stryker. Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology*, 177(3):361–379, 1978.

[10] James Milton and Jeanine Treffers-Daller. Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1):151–172, 2013.

[11] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[12] Lionel Standing. Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*, 25(2):207–222, 1973.

[13] Timothy F Brady, Talia Konkle, George A Alvarez, and Aude Oliva. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38):14325–14329, 2008.

[14] A Crisanti and H Sompolinsky. Path integral approach to random neural networks. *Physical Review E*, 98(6):062120, 2018.

[15] Erik Fransén and Anders Lansner. A model of cortical associative memory based on a horizontal network of connected columns. *Network: Computation in Neural Systems*, 9(2):235–264, 1998.

[16] Raul Vicente, Leonardo L Gollo, Claudio R Mirasso, Ingo Fischer, and Gordon Pipa. Dynamical relaying can yield zero time lag neuronal synchrony despite long conduction delays. *Proceedings of the National Academy of Sciences*, 105(44):17157–17162, 2008.

[17] H Sompolinsky and OL White. Theory of large recurrent networks: from spikes to behaviors. In *Methods and Models in Neurophysics: Lecture Notes of the Les Houches Summer School 2003*. Elsevier, 2004.

[18] Bente Pakkenberg and Hans Jørgen G Gundersen. Neocortical neuron number in humans: effect of sex and age. *Journal of comparative neurology*, 384(2):312–320, 1997.