(Samuel Eckmann, Yashar Ahmadian, Máté Lengyel, Cosyne 2024, Lisbon, Portugal)

Inhibition-stabilized supralinear memory ensembles

The hippocampus plays a central role in memory formation and retrieval. In order to avoid deleterious interference between stored and ongoing experience, theoretical considerations require a separation between phases of memory encoding and recall within the same neural circuit, putatively controlled by hippocampal theta oscillations.[1] However, the neural mechanisms subserving this separation remain unknown. Classical models either remain mute about these mechanisms,[2] or assume purpose-built neuromodulatory interactions that are in conflict with biologically realistic timescales and specificity of synaptic modulation.[3,4] In addition, computational models of memory recall typically do not consider inhibitory neurons at all,[2,5] or only for stabilizing the network globally.[6,7] In contrast, recent experiments suggest that structured inhibitory connections are crucial for memory retrieval.[8,9] Here, we develop an excitatory-inhibitory network model with structured connectivity between units conforming to a canonical circuit motif, the inhibition-stabilized supralinear network.[10,11] This network naturally gives rise to a separation between phases that are ideal for either the recall or the storage of memories, solely determined by the input strength of an external memory cue. For weak input, the cued memory is recalled, and neurons are strongly stabilized by inhibition. For strong input, the external cue is encoded, while inhibition stabilization is paradoxically weaker. Our model only requires a Hebbian and an anti-Hebbian form of biologically plausible plasticity that respectively store the positive and negative parts of the pattern covariance matrix. Specifically, patterns are stored in an input-dominated encoding regime via synapse-type-specific competitive Hebbian plasticity,[12] for the positive part, and—motivated by experimental results[13,14]—via anti-Hebbian plasticity at excitatory-to-inhibitory synapses, for the negative part. The resulting recurrent connectivity is highly structured and consistent with Dale's law. In summary, we present a model of hippocampal memory recall that meets key biological constraints and reveals a novel mechanism for alternating between storage and recall within the same circuit.

We consider E-I rate networks with dynamics

$$\dot{\mathbf{r}} \propto -\mathbf{r} + [\mathbf{W}\mathbf{r} + \mathbf{h}]_+^{\circ n}, \quad [x]_+ = \max(x, 0), \tag{1}$$

$$\mathbf{r} = \begin{pmatrix} \mathbf{r}_E \\ \mathbf{r}_I \end{pmatrix}, \ \mathbf{W} = \begin{pmatrix} \mathbf{W}_{EE} & -\mathbf{W}_{EI} \\ \mathbf{W}_{IE} & -\mathbf{W}_{II} \end{pmatrix}, \ \mathbf{h} = \begin{pmatrix} \mathbf{h}_E \\ \mathbf{h}_I \end{pmatrix}, \tag{2}$$

where bold symbols denote matrices and vectors, and $[\cdot]^{\circ n}$ denotes the element-wise raising to power n. We split synapses into two types $\mathbf{W} = \mathbf{W}^+ + \mathbf{W}^-$. For synapses w_{ij}^{+AB} connecting neurons i, j of type A, B, we assume a synapse-type-specific competitive Hebbian learning rule[12]

$$\dot{w}_{ij}^{+AB} \propto \left(r_i^A - \bar{r}_i^A\right) r_j^B - \gamma_i^{+A} w_{ij}^{+AB}, \quad A, B \in \{E, I\}, \tag{3}$$

where r are firing rates with means $\bar{r}$. The scalar γ maintains the total synaptic weight of all recurrent excitatory or inhibitory inputs such that

$$\sum_j w_{ij}^{+AB} = \mathcal{W}_{AB}^+, \quad \boldsymbol{\mathcal{W}}^+ \equiv \begin{pmatrix} \mathcal{W}_{EE}^+ & \mathcal{W}_{EI}^+ \\ \mathcal{W}_{IE}^+ & \mathcal{W}_{II}^+ \end{pmatrix}, \tag{4}$$

while we set negative weights to zero, adhering to Dale's law. We make the assumption that during memory storage, the network is dominated by the input pattern $\mathbf{h} = \mathbf{p}$ and we further ignore the neuronal non-linearity, such that $\mathbf{r} \propto \mathbf{p}$. The expected synaptic weight change becomes (cf. Eq. 3):

$$\langle \dot{\mathbf{W}}^+ \rangle \propto \mathbf{C} - \boldsymbol{\Gamma}^+ \mathbf{W}^+, \quad \mathbf{C} = \langle \mathbf{p}\mathbf{p}^\mathsf{T} \rangle - \langle \mathbf{p} \rangle \langle \mathbf{p}^\mathsf{T} \rangle, \tag{5}$$

where $\mathbf{C}$ is the pattern covariance matrix, and the diagonal matrix $\boldsymbol{\Gamma}$ holds the appropriate γ_i^{+A}. We make the simplifying assumption that for each excitatory neuron there is an inhibitory neuron that receives the same input, i.e., $\mathbf{h}_E = \mathbf{h}_I$ (cf. Fig. 1*A*). Then $\mathbf{C}$ becomes a 2×2 block matrix with four identical submatrices, so that the learning fixed point, for which $\langle \dot{\mathbf{W}}^{+*} \rangle = 0$, becomes a Kronecker product

$$\mathbf{W}^{+*} = \boldsymbol{\mathcal{W}}^+ \otimes \bar{\mathbf{C}}_+, \tag{6}$$

where $\bar{\mathbf{C}}_+$ is the positive part of the $\mathbf{C}$-submatrix, normalized such that each row sums to one and the total synaptic weights are set by the 2×2 matrix $\boldsymbol{\mathcal{W}}^+$. Synaptic weights that connect two neurons with negative covariance decay to zero.

Negative covariances are instead camptured by synapses $\mathbf{W}^-$, which are plastic according to an anti-Hebbian learning rule mediated by calcium-permeable AMPA receptors at excitatory-to-inhibitory synapses[13,14] (Fig. 1*B*):

$$\dot{w}_{ij}^{-IE} \propto -\left(r_i^I - \bar{r}_i^I\right) r_j^E - \gamma_i^{-I} w_{ij}^{-IE}, \tag{7}$$

with γ_i^{-A}, and $\boldsymbol{\mathcal{W}}^-$ below, defined analogously to competitive Hebbian plasticity (cf. Eqs. 3 & 4). From the entries of $\boldsymbol{\mathcal{W}}^-$, we chose only $\mathcal{W}_{IE}^-$ to be non-zero, since this form of plasticity has been exclusively reported at excitatory-to-inhibitory synapses.[13,14]

We can solve the weight dynamics analytically which results in the following fixed point

$$\mathbf{W}^* = \mathbf{W}^{+*} + \mathbf{W}^{-*} = \boldsymbol{\mathcal{W}}^+ \otimes \bar{\mathbf{C}}_+ + \boldsymbol{\mathcal{W}}^- \otimes \bar{\mathbf{C}}_-. \tag{8}$$

Our plasticity rule predicts, that a presynaptic excitatory neuron and a postsynaptic inhibitory neuron are either connected via a Hebbian or an anti-Hebbian synapse, reflecting either a positive or negative covariance, potentially mirrored by the presence or absence of calcium-permeable AMPA receptors, mediating a form of meta-plasticity.[15]

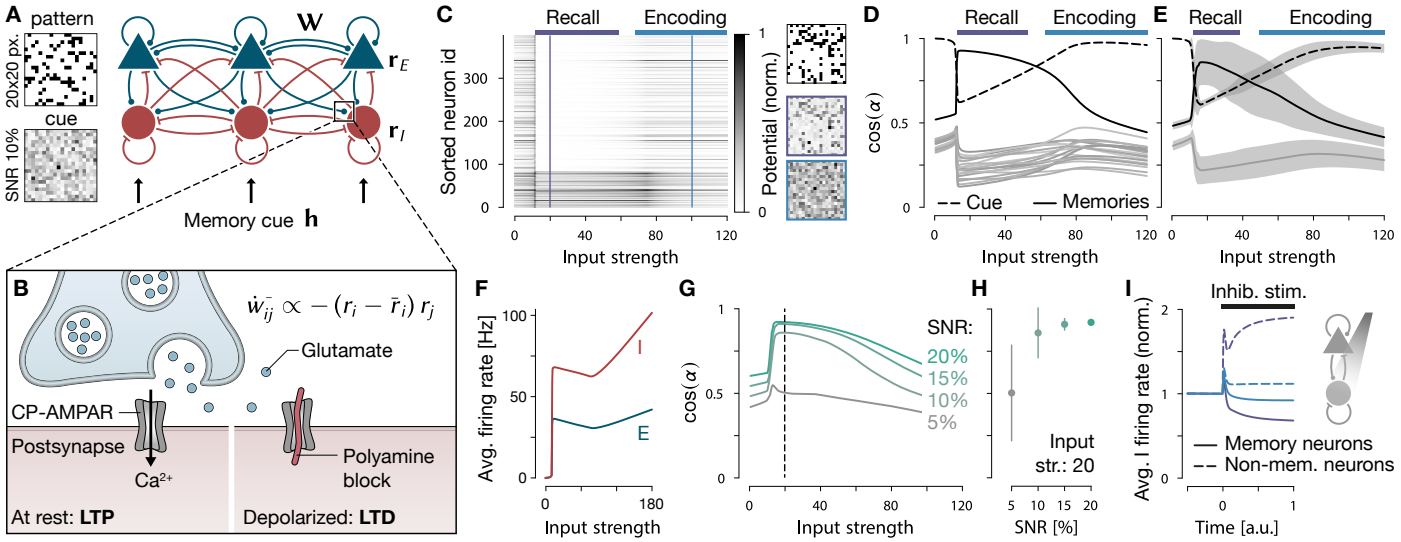

Figure 1. **(A)** 20×20 binary noise patterns are memorized by a recurrent network model of 400 excitatory, $\mathbf{r}_E$, and 400 inhibitory units, $\mathbf{r}_I$, and are recalled when memory cues of different signal-to-noise ratios (SNRs) are presented. Positive correlations are learned via competitive Hebbian plasticity at all recurrent synapse types.[12] **(B)** Negative correlations are learned via competitive anti-Hebbian plasticity at excitatory synapses onto inhibitory neurons: Calcium-permeable AMPA receptors allow for calcium influx during pre-synaptic stimulation while the postsynaptic neuron is at rest, triggering long-term potentiation (LTP).[13] When the postsynapse is depolarized, polyamine blockage prevents LTP and potentially allows for synaptic depression (LTD).[14] Top right: Resulting negative covariance plasticity rule. **(C)** Raster plot of normalized (norm.) steady state membrane potentials during cue presentation at different input strengths. Neurons that are part of the cued memory are sorted to the bottom. The network switches from a recall phase (purple) to an encoding phase (blue), depending on the strength of the input cue. Right: representation of the cued memory (top) and the network activity in the two regimes (bottom), at 20 and 100 input strength (vertical lines), respectively. **(D)** Cosine similarity between membrane potentials of the excitatory population and the cued memory (solid, black), non-cued memories (solid, grey), and the recall cue (dashed, black). Same data as in *C*. **(E)** Same as in *D*, averaged over 20 stored memories times 5 different noise cues. Shaded regions span two standard deviations. **(F)** Average excitatory (E) and inhibitory (I) population firing rates for neurons in *C*. **(G)** Same as *E*, for different SNRs (shades of green). Only the average overlap with the cued memories is shown. Means and standard deviations for an input strength of 20 (vertical dashed line) are presented in **(H)**. **(I)** Average population firing rates of inhibitory neurons that are either part of a cued memory pattern (solid line) or not (dashed line). The recall cue is presented at either 20 (purple) or 100 (blue) input strength. Inhibitory neurons received additional, unspecific external stimulation, starting at time zero.

We construct a network of 400 E-I modules and store 20 binary patterns with 80 active units per pattern (Fig.1*A*, right). To obtain a noisy memory cue, we linearly mix a memory pattern with a noise pattern. We present this cue to the network at different input strengths and record its steady state membrane potentials (Fig.1*A*, left). For low input strengths, the network remains silent, before activities discontinuously increase and represent the cued memory (cf. Fig.1*C* & *F*). For increasing input strengths, the network enters an encoding phase, where the noisy input cue is represented, suggesting a regime where input-dominated learning can take place, potentially via behavioral timescale synaptic plasticity at the troughs of theta cycles,[16,17] facilitated by slower cholinergic modulation.[3,4]

We quantify memory recall and encoding by measuring the similarity between membrane potentials, and the stored memories, or the input cue, respectively (Fig.1*D*). We find memory recall to be robust with respect to different memory patterns, and cues (Fig.1*E*), as well as different cue noise levels (Fig.1*G* & *H*).

The network is inhibition-stabilized and exhibits the paradoxical effect:[18] When inhibitory neurons are perturbed during stimulation with a memory cue, the activity of inhibitory memory neurons paradoxically decreases (Fig.1*I*). We quantified the extent of this decrease in the recall and encoding regimes and find a weaker decrease at higher input levels, contrasting results in less structured supralinear networks.[19,20]

References: ***1***. M. E. Hasselmo *et al.*, *Neural computation* (2002). ***2***. J. J. Hopfield, *Proceedings of the National Academy of Sciences* (1982). ***3***. M. E. Hasselmo, *Current opinion in neurobiology* (2006). ***4***. L. M. Giocomo, M. E. Hasselmo, *Molecular neurobiology* (2007). ***5***. M. Lengyel *et al.*, *Nature neuroscience* (2005). ***6***. Y. Roudi, P. E. Latham, *PLoS computational biology* (2007). ***7***. W. Gerstner *et al.* (Cambridge University Press, 2014). ***8***. X. He *et al.*, *Neuron* (2021). ***9***. A. Tzilivaki *et al.*, *Neuron* (2023). ***10***. Y. Ahmadian *et al.*, *Neural computation* (2013). ***11***. D. B. Rubin *et al.*, *Neuron* (2015). ***12***. S. Eckmann, J. Gjorgjieva, *bioRxiv* (2022). ***13***. K. P. Lamsa *et al.*, *Science* (2007). ***14***. F. Laezza *et al.*, *Science* (1999). ***15***. S.-Q. J. Liu, S. G. Cull-Candy, *Nature* (2000). ***16***. K. C. Bittner *et al.*, *Nature neuroscience* (2015). ***17***. K. C. Bittner *et al.*, *Science* (2017). ***18***. M. V. Tsodyks *et al.*, *Journal of neuroscience* (1997). ***19***. Y. Ahmadian, K. D. Miller, *Neuron* (2021). ***20***. S. Sadeh, C. Clopath, *Elife* (2020).