Learning Associations With A Neurally-Computed Global Novelty Signal

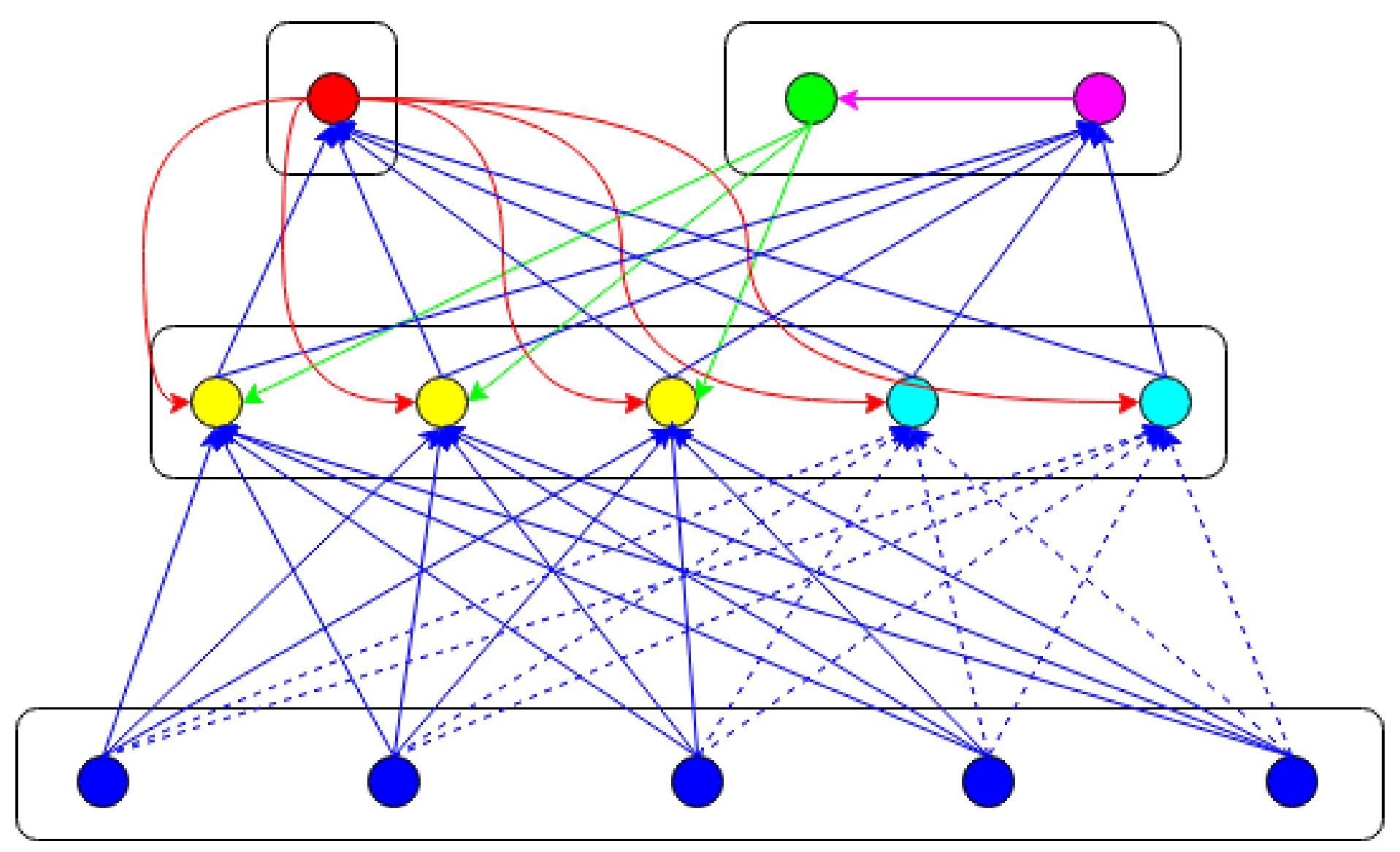
Mohammad Javad Faraji[1], Kerstin Preuschoff[2], and Wulfram Gerstner[1]

1. School of Life Sciences, Brain Mind Institute & School of Computer and Communication Sciences, EPFL, Switzerland.
2. Geneva Finance Research Institute (GFRI), University of Geneva, Switzerland.

Abstract

Novelty is a crucial factor in both learning and memory formation. In order to efficiently learn new memories without altering past useful memories, it is essential to detect novel stimuli. It is also necessary to incorporate novelty information (related to the structure of the environment) in addition to reward information in models of reinforcement-based learning. We address (1) how novelty is computed in a neurally plausible way and (2) how it affects (synaptic) learning rules.

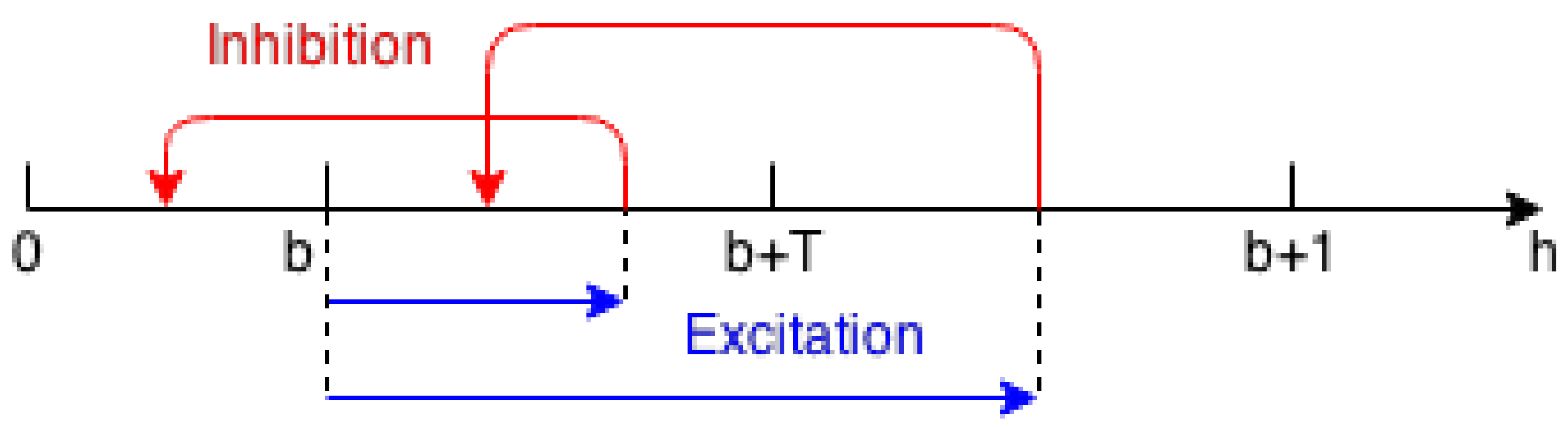
Computing novelty using NN



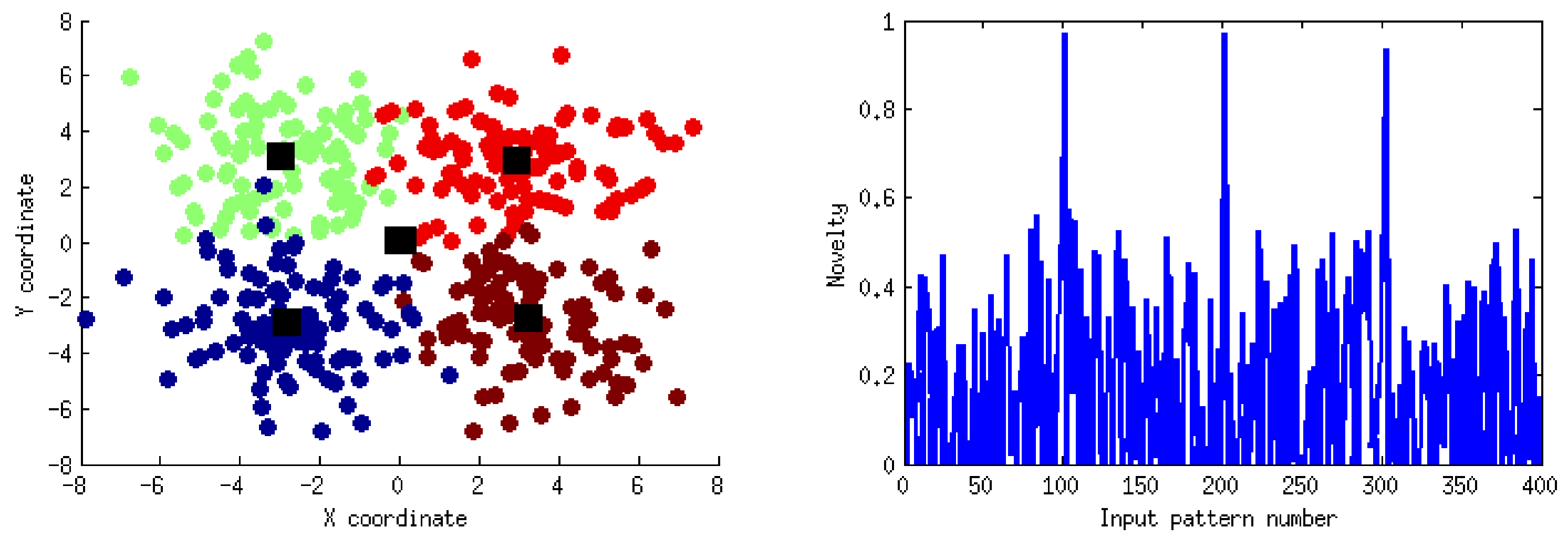
$$
\begin{aligned}
x_i &= g(h_i)\\
h_i &= b + E_i - I_i\Theta(\bar{x}_i)\\
E_i &= \sum_j w_{ij}x_j\\
I_i &= c\mathcal{N}(X)\\
\mathcal{N}(X) &= 1 - \max_i\,[E_i]\\
&\propto 1 - \max_i\,[x_i]\\
c &= T(1-T)^{-1}
\end{aligned}
$$

Input units (blue circles in bottom layer) provide excitatory input for excitatory decision units in middle layer consisting of both always-loser units (cyan circles) and active units (yellow circles) through weak (dashed arrows) and strong (solid arrows) connections, respectively. WTA between decision units is implemented via mutual inhibition through a common inhibitory pool (red circle). Additionally, the decision units project to an inhibitory unit (magenta circle) via predefined non-plastic connections to determine their maximum activity. A novelty detector (green circle) strongly inhibits all active units if the input pattern may belong to non of the existing clusters.

$$\max_i\,[E_i] < T \Rightarrow \mathcal{N}(X) > 1 - T \Rightarrow I_i > T > E_i \Rightarrow h_i < b$$

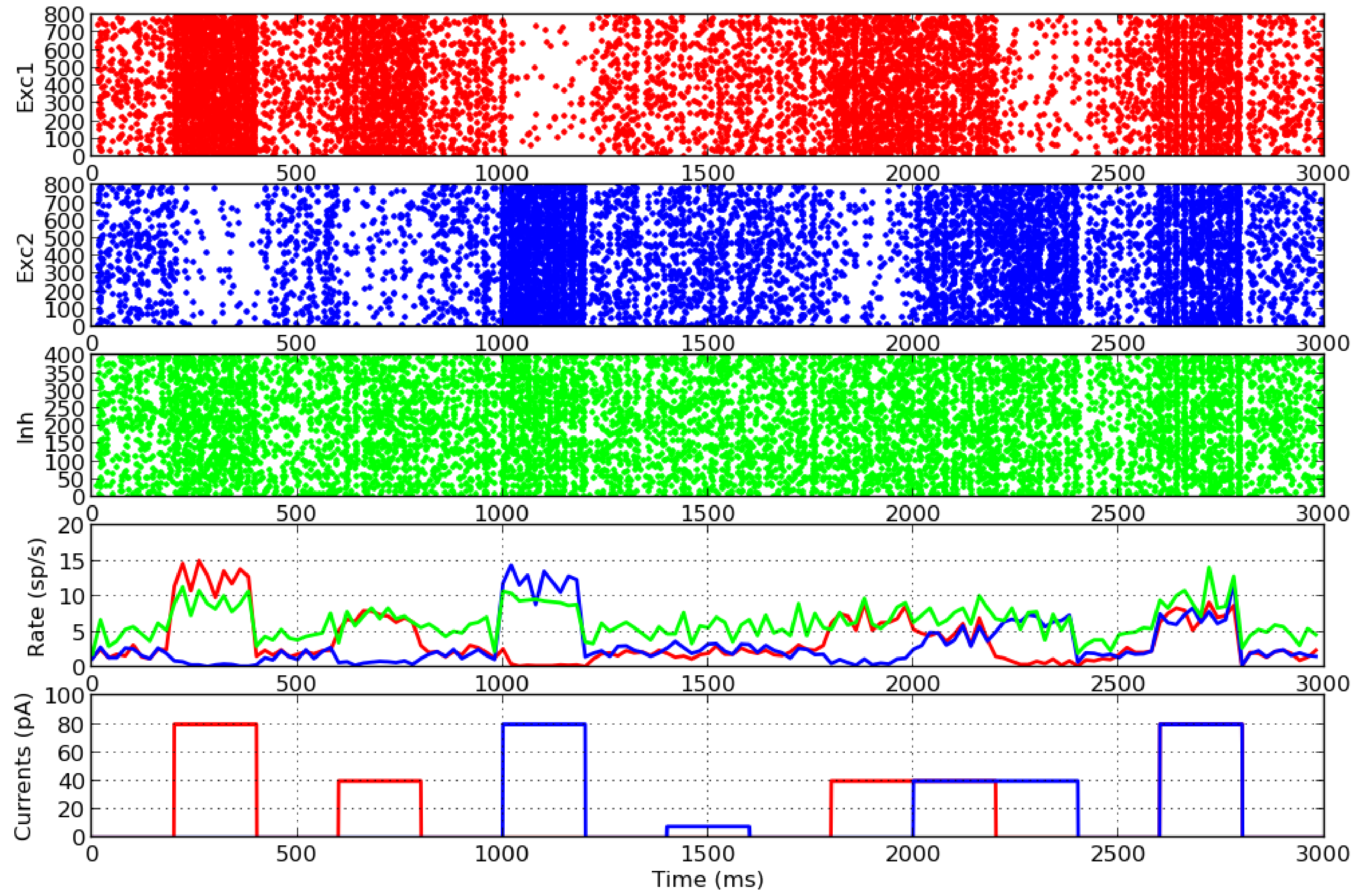


Novelty triggers new clusters



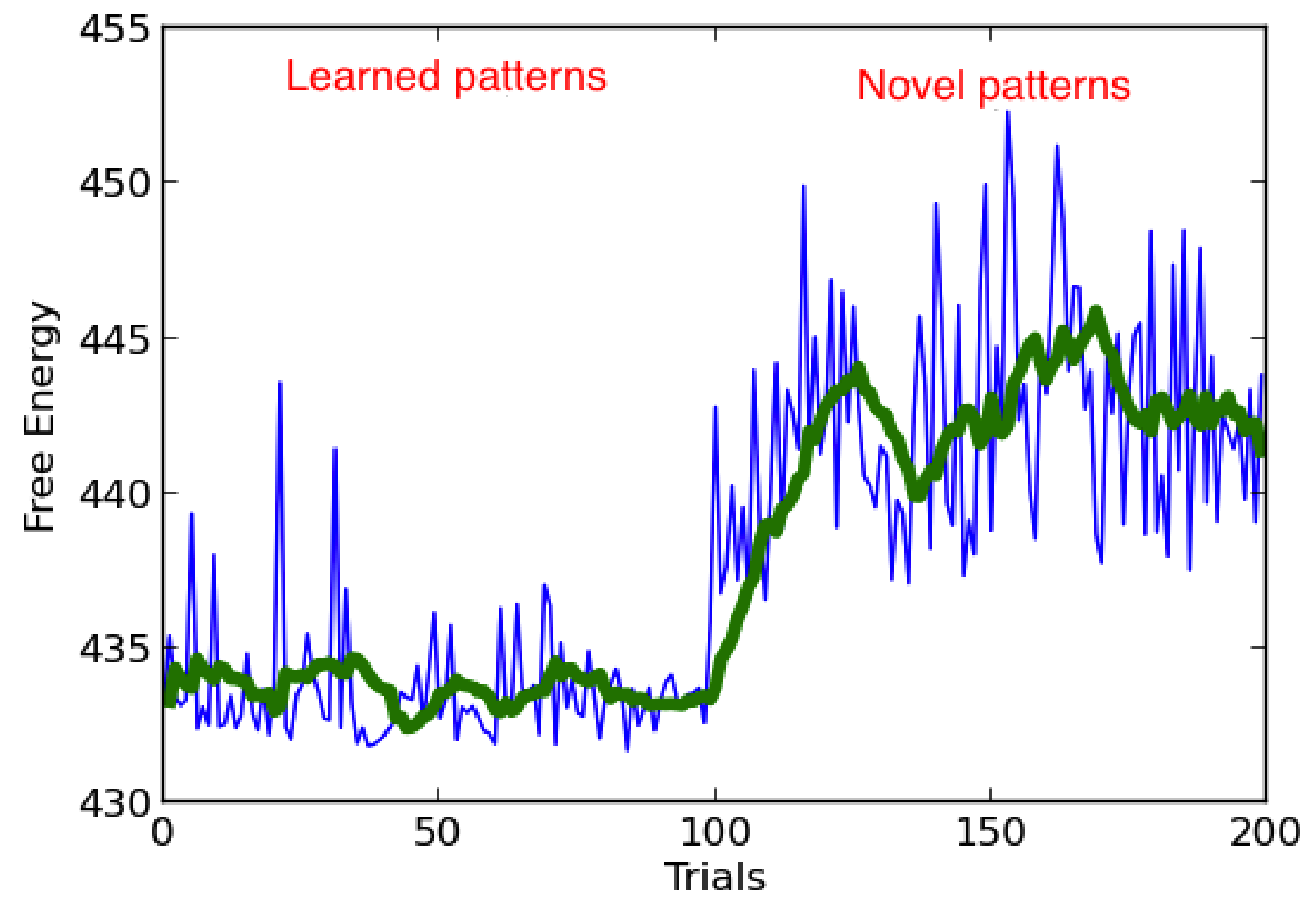
Input patterns (colored dots) are presented in four $\mathbf{100}$-sample blocks (samples from multivariate normal distribution). Each color indicates the cluster into which samples are classified. The black squares among colored dots represent the center of each learned cluster. Black squares at $\mathbf{(0,0)}$ correspond to weak synapses afferent to always-loser units which have not yet been used for learning a new cluster. Right figure depicts the measured novelty for each input pattern. Whenever the distribution from which samples are drawn is changed (i.e., trials $\mathbf{101, 201,}$ and $\mathbf{301}$), the novelty signal increases leading to the creation of a new cluster.

Implementation in spiking networks



The novelty signal in our framework can be interpreted as a (global) modulatory signal, corresponding to the diffusion of a non-specific neuromodulator (e.g., NE released from locus coeruleus (LC) neurons) and can modulate the local Hebbian factors.

Free energy in neural networks

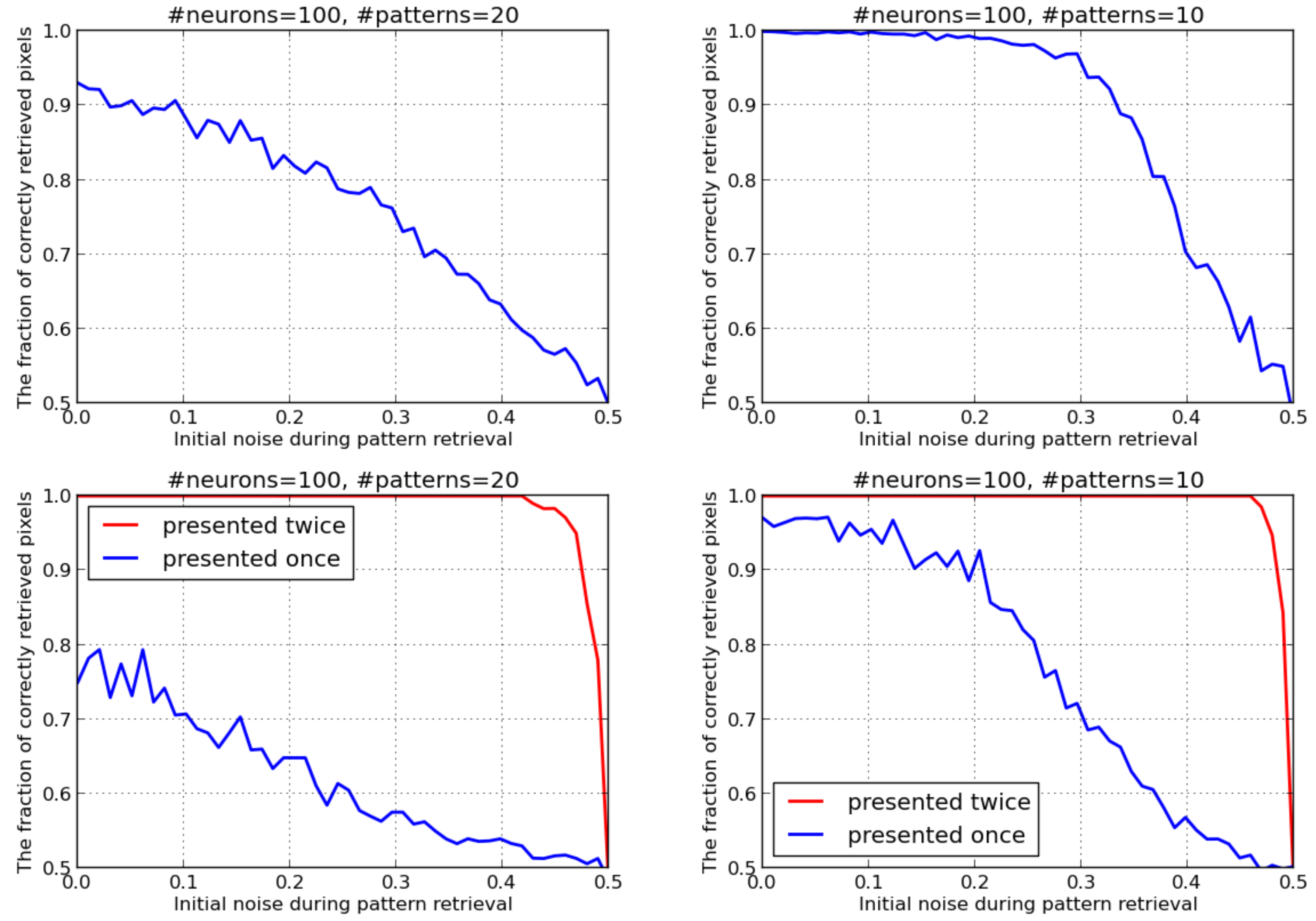


The variational free energy (blue line), used for estimating the likelihood of the input patterns (digits from the MNIST dataset) in a Boltzmann machine can also be used as a novelty measure.

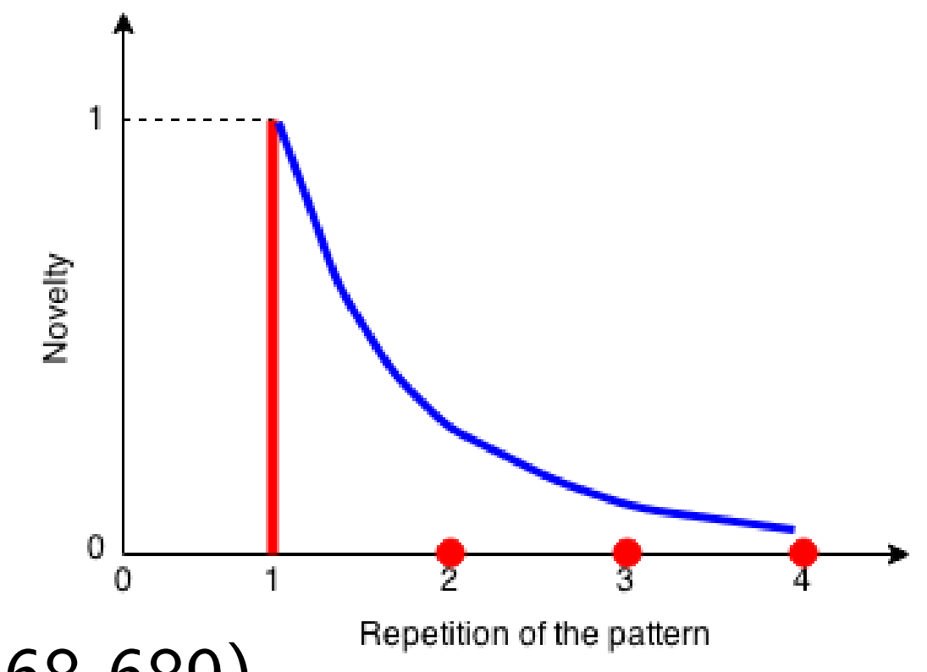
Hopfield model modulated by novelty

Classic Hopfield model of associative memory (in which $\mathbf{\Delta w_{ij}} \propto \xi_i^\mu \xi_j^\mu$) ignores the effect of pattern repetition during learning. A pattern, which has a higher contribution in strengthening a synapse, occupies a wider basin of attraction.

$$\mathbf{w_{ij}} = \mathbf{c}\sum_\mu \lambda_\mu \xi_i^\mu \xi_j^\mu \Rightarrow \mathbf{E}[\vec{\mathbf{S}}] = -\mathbf{cn}^2\sum_\mu \lambda_\mu \mathbf{m}_\mu^2$$



A Hebbian learning rule modulated by the novelty signal ($\mathbf{\Delta w_{ij}} \propto \xi_i^\mu \xi_j^\mu \mathcal{N}(\mu)$) can resolve that issue. Novelty as a function of repetition could be the following.



Research was supported by the ERC (grant no. 268 689).