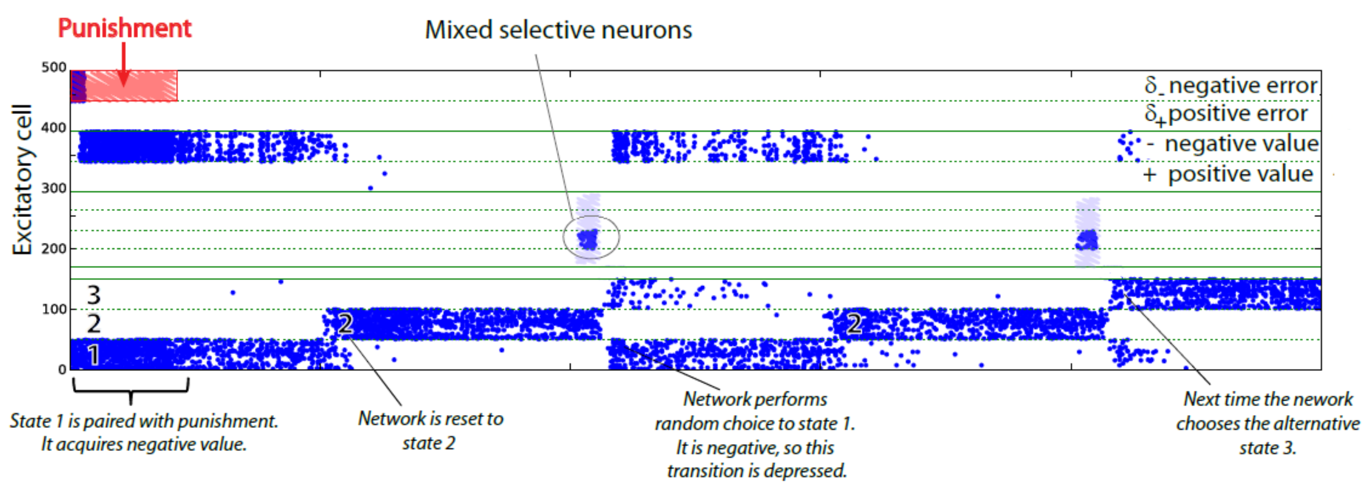
An adaptive spiking neural network for decision making in partially observable environments

Mattia Rigotti, Daniel Ben Dayan Rubin, Nathaniel Daw, Stefano Fusi

Predicting future events is essential for deciding between alternative courses of action, some of which may lead to rewarding outcomes, others which may result in loss of resources. Reinforcement Learning (RL) offers a theoretical framework for formalizing the problem of predicting the result of interactions with the environment and adapting behavior so as to choose an optimal course of action. RL's popularity in neuroscience is at least partly due to its success in modeling neural data as exemplified by the well-known interpretation of dopamine neurons activity as a temporal-difference reward-prediction error. Considerable effort has recently been devoted to extending the RL formalism to the case of only partially observable environments, a situation which is commonly abstractly formalized as a Partially Observable Markov Decision Process (POMDP). In an attempt to bridge the abstract algorithmic RL formalism with a biophysically plausible neural network implementation, we study a spiking network model endowed with representations of internal states and their values, in addition to observable features of the external environment. Our model defines a probabilistic policy which corresponds to a set of transition probabilities between internal states conditional on an observed external stimulus. We show how this architecture is formally equivalent to a finite-state controller (FSC), a finite policy parametrization which has been shown to efficiently approximate an optimal POMDP policy under many circumstances. This structure has several advantages. FSCs induce finite-state Markov chains in a POMDP, and thus simplify the problem of computing policy values. FSCs deal with a finite set of discrete states, as opposed to the high-dimensional belief-space of exact POMDP methods. From an implementational point of view, the discrete nature of FSC states allow us to represent them as attractors of the neural dynamics of a spiking network.
Fundings: DARPA SyNAPSE; Gatsby, Kavli, Sloan-Swartz and Swiss National Science Foundations

We show how to implement a FSC in a recurrent attractor network of spiking integrate-and-fire neurons endowed with slow NMDA synaptic channels and random synaptic projections to an additional network of neurons selective to external observables. This latter network of mixed selectivity neurons provides a high-dimensional representation of the conjunction of external and internal state, and, thanks to reward-modulated plastic synapses, it can adaptively mediate the stimulus-driven transitions between the attractor states of the recurrent network.

We implement a simple actor/critic architecture where the short-lived phasic reward-prediction error signal is utilized to update the time-extended value of a given set of internal states and to modify a probabilistic policy at every state. This gives a natural way to reactively cope with a partially observable environment and actuate an initial (suboptimal) policy. Motivated by previous work we show how to improve on this initial policy by iteratively adding new internal attractor states.



Learning state transitions through reinforcement learning. The figure shows the spiking activity of a network with 3 internal attractor states as a function of time. State 1 is paired with a punishing outcome. External stimuli trigger transitions between the states. The network learns to avoid the negative state 1.