

# Learning and Memory in an Exactly Solvable Stochastic Spiking Network

Surya Ganguli

Sloan Swartz Center for Theoretical Neurobiology, UCSF

A central challenge in systems neuroscience involves understanding how memories are stored as stable patterns of synaptic efficacies in recurrent spiking networks, and how learning is mediated by global changes in these synaptic patterns. At the level of the single cortical synapse, it is thought that changes in efficacy are driven by local spike timing dependent plasticity (STDP) rules. Thus to understand learning and memory, one must understand the complex connection between these local rules, operating at individual synapses, and their effects at the macroscopic level of recurrent neural circuit behavior. In situations where there is a separation of time scales between plasticity rates and spiking dynamics, the statistics of spatiotemporal correlations between spikes drive synaptic changes. In such a restricted scenario, the crux of the difficulty in understanding global network level learning under local STDP lies in the intricate relation between synaptic states and spiking statistics; the spiking statistics are a complicated global function of all synaptic states in the network and in turn these emergent statistics drive further changes in all synaptic states.

We address this difficulty by proposing a recurrent spiking model that, remarkably, is exactly solvable in the sense that the spiking statistics can be computed in closed form for arbitrary recurrent and feed forward synaptic weights and arbitrary instantaneous correlations in external inputs. In this model each synapse has a weight  $p$  that represents the probability a presynaptic spike will cause a postsynaptic spike with a fixed latency  $\epsilon$ . If multiple input spikes arrive simultaneously on different synapses, the output spiking probability sums linearly. The weights are constrained so the output probability never exceeds one. This is simply the linear Poisson neuron (Kistler and van Hemmen 2002) with a delayed delta-function kernel. Furthermore all neurons are pairwise independent conditioned upon past inputs. Thus a network of  $N$  neurons is characterized by its synaptic weight matrix  $p_{ij}$ . There can also be synapses driven by external inputs with weights  $f_{ik}$ .

Given these definitions, one can calculate the cross correlation function  $\langle s_i(t)s_j(t+t') \rangle$ , where  $s_i(t)$  is the spike train of neuron  $i$ , by explicitly summing over all consistent spiking histories. Naively, histories are paths through the space of all  $2^N$  possible spike patterns at each time, but the sum over paths simplifies considerably in this particular model. The reduced sum (which is nonzero only when  $t' = n\epsilon$  for integer  $n$ ) can be visualized as a sum all over paths through the network of two given types. The first type of path starts at neuron  $i$  at time  $t$ , hops from neuron to neuron and visits neuron  $j$  in precisely  $n$  steps at time  $t'$ . The second type of path originates at a third neuron in the past, and sends off two nonintersecting branches through the network. One branch must terminate at neuron  $i$  at time  $t$  and the other at neuron  $j$  at time  $t+t'$ . Each path is associated with a product of probabilities  $p_{ij}$  corresponding to links in the path, and is weighted by the firing rate of the neuron at the path's point of origin. These paths are shown in figure 1.

Surprisingly, these paths can be summed and written in closed form. The resulting expression can be substituted into any STDP rule of the form  $\frac{d}{dt}p_{ij} = F(p_{ij}, \langle s_i s_j \rangle)$ . Fixed points of these equations represent the stable network

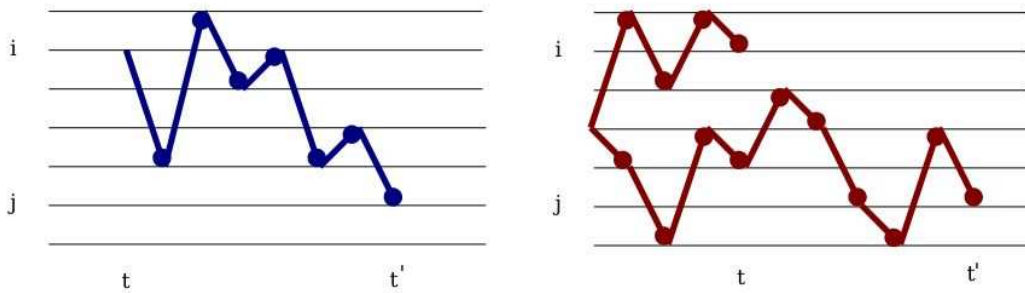


Fig 1: These two types of contributions (spacetime paths) to the cross correlation between neuron  $i$  and  $j$  capture the intuition that there are two ways neurons can be correlated: either there is a direct interaction between  $i$  and  $j$  through the network (blue path), or a third neuron can interact with both (red path).

memories. Changes in these fixed points in response to changes in external input correlations correspond to network learning.

An exact analysis of small 2 and 3 node networks yields the following intuitions. The existence of recurrent circuitry in addition to feed forward circuitry partitions the space of input correlations into three types: weak, strong positive, and strong negative. If correlations are weak, a large class of recurrent synaptic patterns are stable. Strong positive correlations destabilize these synaptic patterns and create a new synaptic fixed point, which reinforces the input spike correlations. If the network starts in this fixed point and is subsequently exposed to strong negative correlations, the fixed point will destabilize and the weights will move to a new fixed point that reinforces the negative input correlations. The stronger the positive correlations were, the harder it becomes to destabilize the resulting fixed point upon subsequent exposure of the network to negative correlations. Put simply, recurrent synapses, combined with local STDP rules, make it harder for a network to change its mind in the face of new observations that contradict old observations. These synapses make networks stubborn.

The analysis of small networks can be extended to larger networks, where questions about the learnability of various input correlations given initial synaptic weights arise. More generally, this work opens the door to an understanding of synaptic pattern formation in large networks with arbitrary feedbacks and sensorimotor connectivities, which is a prerequisite for a biophysical understanding of learning and memory at the spiking network level.