Applying a Reduced Model of Apical and Basal Dendritic Compartments to Sequence Learning in Recurrent Neural Networks

Fabian Schubert, Claudius Gros

Institute for Theoretical Physics, Goethe University Frankfurt a.M.

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Introduction

- Experiments suggest that, depending on the amount of apical (distal) and basal dendritic synaptic drive, layer 5 pyramidal neurons can exhibit quietness, low frequency spiking and high frequency bursting [1, 2].
- High frequencies occur when distal and basal inputs coincide in time. A simplified, spiking compartment model of this effect has been used to gate plasticity of basal connections by means of distal synaptic inputs [3].
- In our framework, coincidence detection of distal and basal input modulates plasticity.
- This error-driven learning could be used to extract a predictive signal from proximal connections.

Model Description

We used a discrete time rate encoding neuron model based on a phenomenological Layer 5 Pyramidal cell model by Shai et al. [2], whose output frequency is given by

$$x\left(I_p, I_d\right) = 2\left[\sigma\left(I_p-\theta_{p1}\right)\sigma\left(I_d-\theta_d\right) + \alpha\sigma\left(I_p-\theta_{p0}\right)\sigma\left(-\left(I_d-\theta_d\right)\right)\right]-1$$

$$\sigma\left(x\right) = \frac{1}{1+\exp(-g\cdot x)}$$

$$I_p(t) = \sum_{k=1}^{n_p} w_{p,k}(t)x_{p,k}(t)$$

$$I_d(t) = \sum_{k=1}^{n_d} w_{d,k}(t)x_{d,k}(t)$$

Proximal and distal weights were subject to the same form of Hebbian plasticity and weight normalization.

$$\Delta w_i(t) = \epsilon_w\left(x(t)-\langle x\rangle\right)\left(x_i(t)-\langle x_i\rangle\right)$$

$$w_i(t+1) = w_{\text{total}}\frac{w_i(t)+\Delta w_i(t)}{\sqrt{\sum_{k=1}^{n}\left[w_{pk}(t)+\Delta w_{pk}(t)\right]^2}}$$

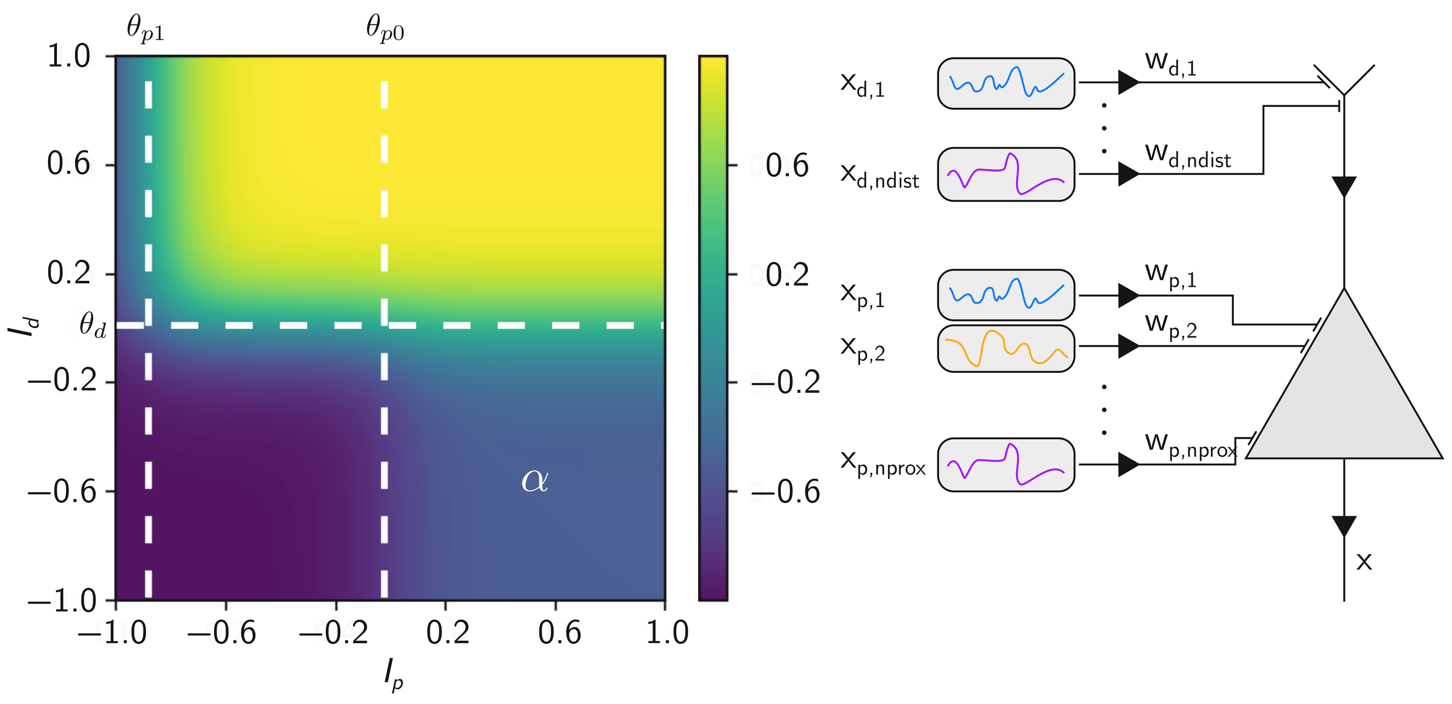

Figure 1: Left: Output rate as a function of proximal (basal) and distal (apical) dendritic input. Right: Illustration of the accumulation of inputs in the neuron model. α denotes the low-frequency rate.

Results for a Single Neuron

- We fed a single neuron with 10 randomly fluctuating proximal inputs and one distal input that was exactly correlated with one of the proximal signals.
- As a "distraction", the standard deviation of another proximal input was increased, giving the proximal inputs had a dominant principal component.
- As shown in Fig. 3, the distal input acts as a guiding input, such that plasticity potentiates the proximal input that maximizes correlation with the distal input, in spite of the distracting signal.
- If the distal input is uncorrelated with any of the proximal input, proximal weights select the principal component.
- Learning of multiple distal inputs is dominated by the principal component.

Analytic Approximation

An analytic approximation of the proximal weight dynamics consists of two dominant terms, containing the proximal covariance matrix and proximal-distal covariances.

$$\Delta w_{p,i} \approx \epsilon_w' \sum_j \alpha C_{ij}^{xx} + (2-\alpha)C_i^{dx}$$

$$C_{ij}^{xx} \equiv \langle\left(x_i-\langle x_i\rangle\right)\left(x_j-\langle x_j\rangle\right)\rangle$$

$$C_i^{dx} \equiv \langle\left(d-\langle d\rangle\right)\left(x_i-\langle x_i\rangle\right)\rangle$$

$$\epsilon_w' \equiv \epsilon_w\frac{g}{8}$$

Coupling with an Echo State Network for Sequence Prediction

- We hypothesized that a single unit sending its output to a random Echo State network while also receiving proximal input from the same reservoir, the neuron could learn to predict the guiding signal fed in as a distal input, as illustrated in Fig. 2.
- In contrast to the usual Echo State architecture, input and output are merged into a single unit.
- Prediction of the guiding signal succeeded if the learning rule was modified such that it included a local error signal:

$$\Delta w_{d,i}(t) = \epsilon_w\left(I_p(t)-I_d(t)\right)x_{d,i}$$

- After learning, the proximal input acted as a predictor of the distal guiding signal, see Fig. 2.
- Compared to separate input and output units, the convergence of the learning process was prolonged, since the input signal was disturbed by the initially random proximal input.

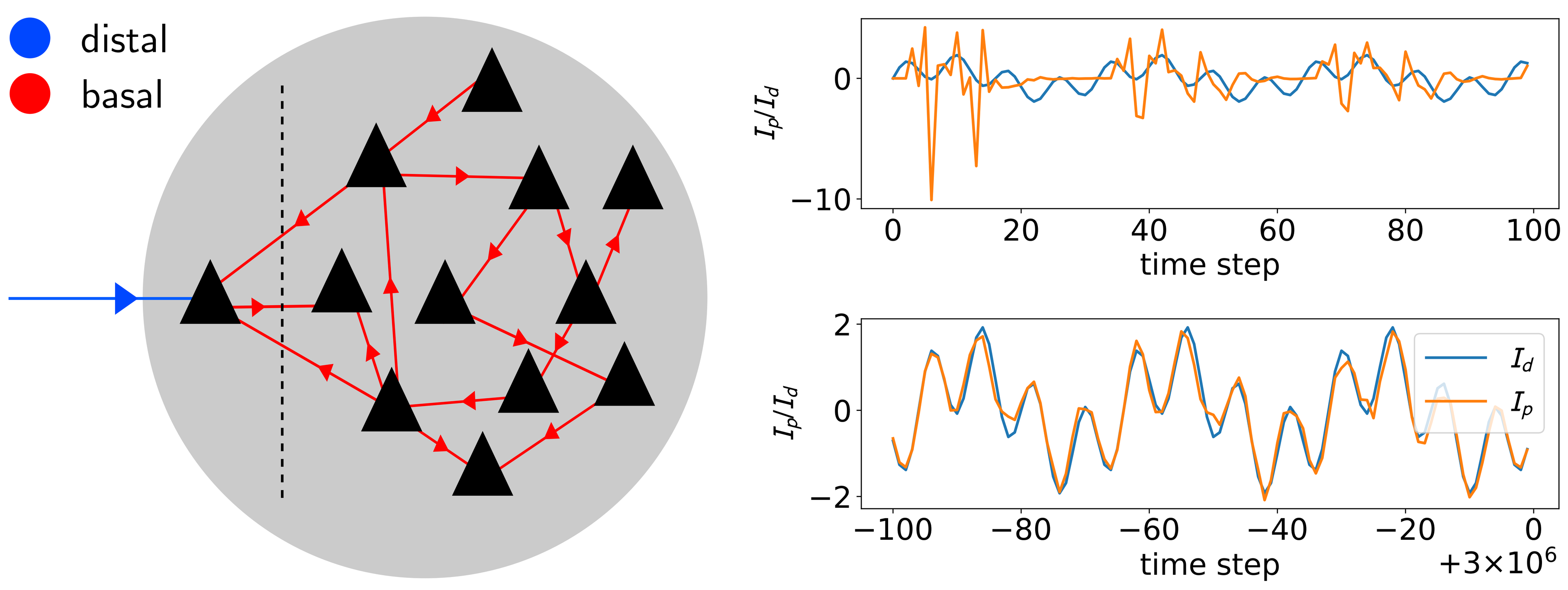

Figure 2: Combining the neuron model with an Echo State network for sequence prediction.

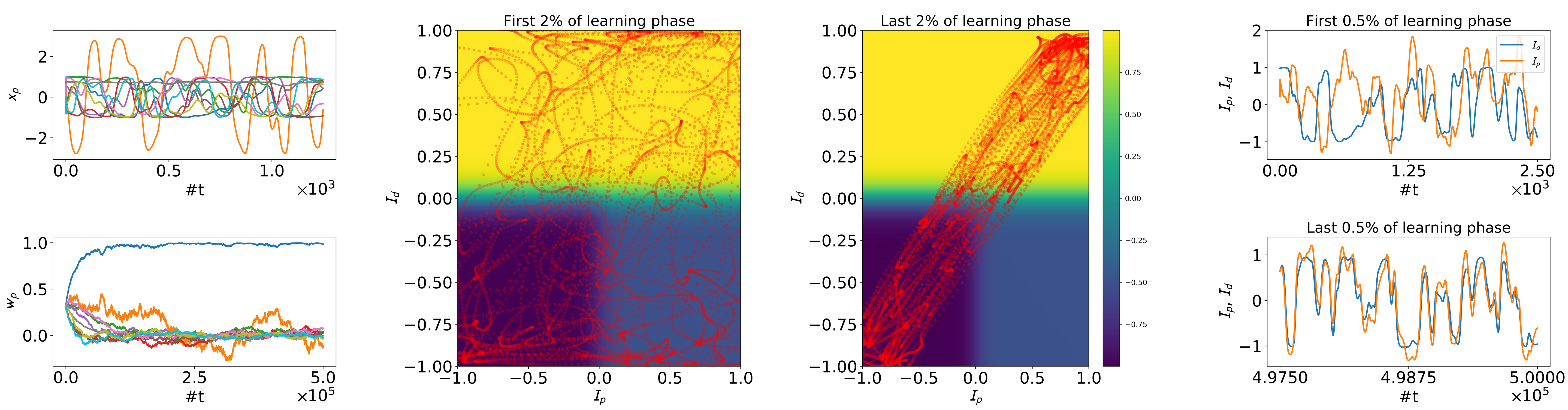

Figure 3: Results for a single distal input, correlated with one proximal input

Conclusions

- The specific activation function of our model neuron acts as a coincidence detector between proximal and distal inputs.
- This property can gate plasticity, where the distal input acts as a guiding signal.
- We hypothesize that this property can be used in a predictive framework, e.g. by means of a recurrent, proximally connected dynamic reservoir.
- Future work: Can external proximal input be included? If so, can we extend this model to a hierarchical predictive structure?

[1] Johannes J. Letzkus, Björn M. Kampa, and Greg J. Stuart. Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse Location. *Journal of Neuroscience*, 26(41):10420–10429, 2006.

[2] Adam S. Shai, Costas A. Anastassiou, Matthew E. Larkum, and Christof Koch. Physiology of Layer 5 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. *PLOS Computational Biology*, 11(3):1–18, 03 2015.

[3] J. Bono and C. Clopath. Modeling somatic and dendritic spike mediated plasticity at the single neuron and network level. *nature communications*, 2017.