

Learning in a Simplified Neural Compartment Model

Fabian Schubert

August 31, 2020

1 Dendritic Computation for Sequence Prediction

Shai et al. proposed a phenomenological model describing neural activity as a function of basal and distal synaptic input in pyramidal neurons [1], see Fig. 1. We further simplified this model to the following form:

$$y(I_p, I_d) = \sigma(I_p - \theta_1^p) \sigma(I_d - \theta^d) + \alpha \sigma(I_p - \theta_0^p) [1 - \sigma(I_d - \theta^d)] \quad (1)$$

$$\sigma(x) = \frac{1}{1 + \exp(-4g \cdot x)} \quad (2)$$

See Fig. 2 for a visualization of the relevant parameters.

We expected the nonlinearity of the neuronal output to be selective for correlated proximal and distal input, such that a Hebbian learning rule allows the neuron to be selective to synaptic input generating the most coherence between distal and proximal input. We tested this hypothesis with a single neuron using the setup illustrated in Fig. 3. The neuron receives $n = 10$ proximal input signals ($y_1^p(t), \dots, y_n^p(t)$) and a single distal input signal $y^d(t)$. Proximal inputs were generated by simulating a chaotic random network and recording the activity of a randomly chosen unit. Each simulation run corresponds to a single proximal input signal to prevent possible correlations. The distal input was generated as a linear combination of the proximal input streams. In particular, we chose it to be an exact copy of $y_1^p(t)$ as the most simple case for initial testing. Proximal weights were subject to Hebbian plasticity and weight normalization. The exact mathematical description of the setup is given in (3) – (8) and Table 1.

$$I_p(t) = \sum_{k=1}^n w_k^p(t) y_k^p(t) \quad (3)$$

$$y^d(t) = \sum_{k=1}^n a_k x_k^p(t), \quad \sum_{k=1}^n a_k = 1 \quad (4)$$

$$I_d(t) = w^d x^d(t) \quad (5)$$

$$y(t) = \sigma(I_p(t) - \theta_1^p) \sigma(I_d(t) - \theta^d) + \alpha \sigma(I_p(t) - \theta_0^p) [1 - \sigma(I_d(t) - \theta^d)] \quad (6)$$

$$\Delta w_i^p(t) = \epsilon_w (y(t) - \langle y \rangle) (y_i^p(t) - \langle y_i^p \rangle) \quad (7)$$

$$w_i^p(t+1) = w_{\text{total}}^p \frac{w_i^p(t) + \Delta w_i^p(t)}{\sum_{k=1}^n w_k^p(t) + \Delta w_k^p(t)} \quad (8)$$

A First result is shown in Fig. 4. The correct proximal weight is chosen such that the proximal total input follows the distal input signal. However, when testing the

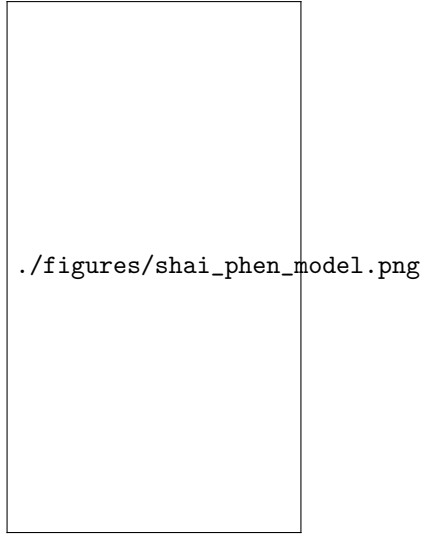


Figure 1: Firing rate as a function of distal and proximal input of the rate model proposed in [1].

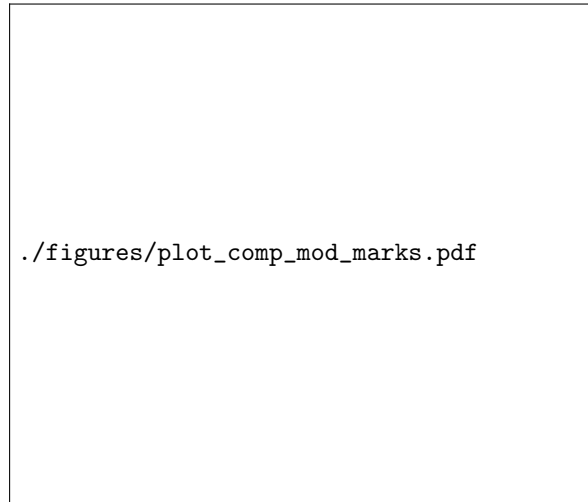
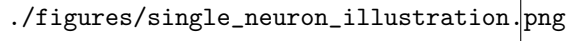


Figure 2: Output firing rate as a function of proximal and distal input as given by (1)

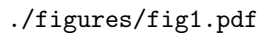
Table 1: Parameter settings for the setup described in (3) – (8).

w^d	1
θ_0^p	0.5
θ_1^p	0.2
θ^d	0.5
α	0.3
g	5
ϵ_w	10^{-3}
w_{total}^p	1




./figures/single_neuron_illustration.png

Figure 3: A single neuron receiving multiple proximal inputs and a single distal signal.



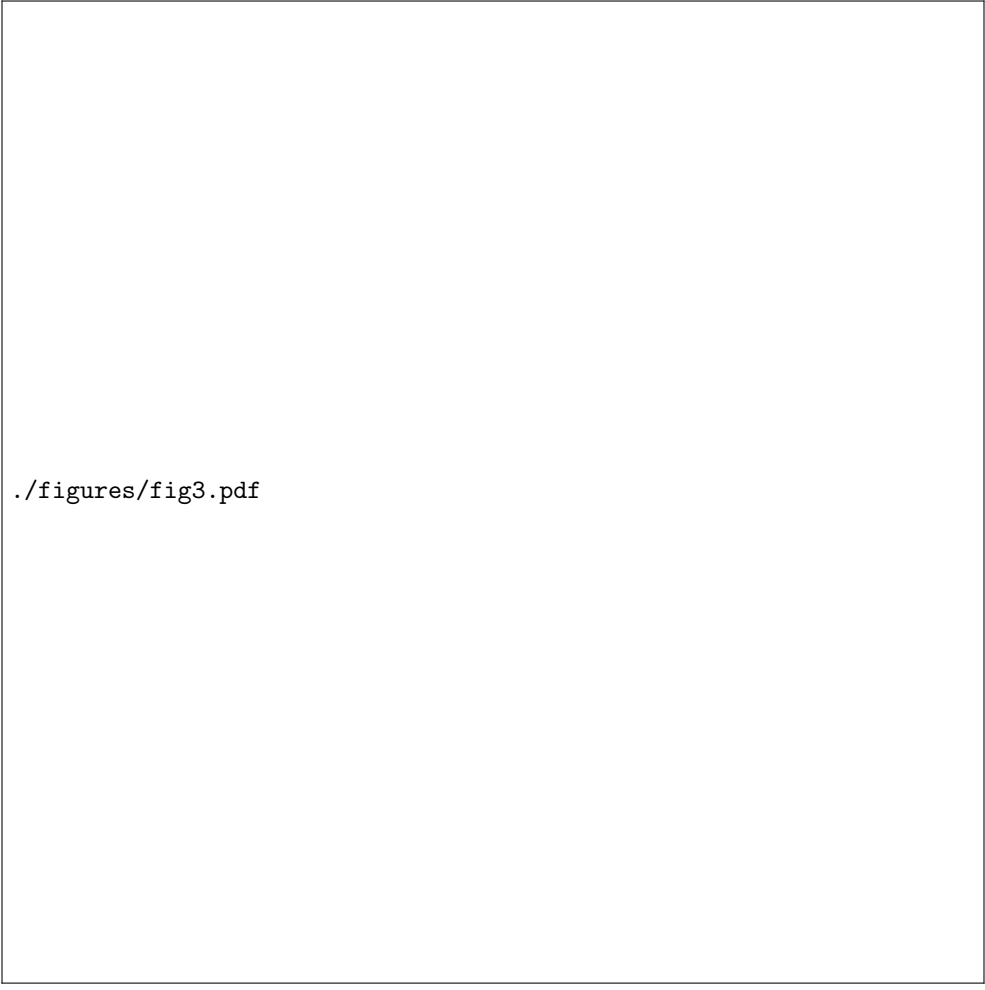
./figures/fig1.pdf

Figure 4: Weights and input signals before and after Hebbian learning of proximal weights, using nonlinear proximal-distal interaction.



./figures/fig2.pdf

Figure 5: Weights and input signals before and after Hebbian learning of proximal weights, using linear proximal-distal summation.



./figures/fig3.pdf

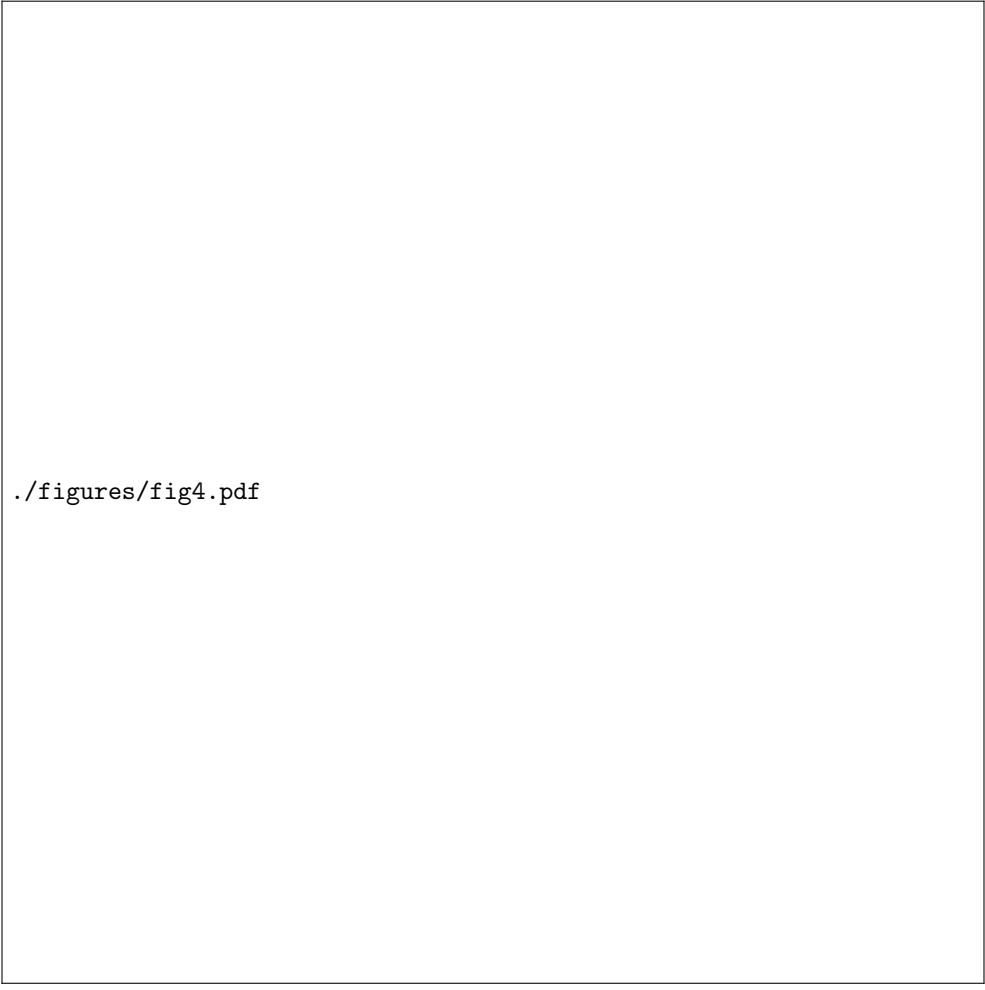
Figure 6: Same setup as in Fig. 5, but with the signal of the second proximal input channel scaled by a factor of two.

same setup except for a point-neuron summation of both proximal and distal inputs, the same result could be achieved, as shown in Fig. 5.

As a further test, we doubled the standard deviation of the second proximal input, thereby making it preferential for the classic scheme of Hebbian learning combined with linear superposition of inputs. This revealed a difference between the proximal-distal activation function and the point neuron: While the proximal-distal scheme was able to select the proximal input that would maximize correlation between proximal and distal signals, the point neuron selected the principal component of the proximal input, in this case being the second input signal. These results are shown in Fig. 6 and 7.

1.1 Output Dynamics After Learning

In the given case that after learning, the proximal input is aligned to the distal input, one might consider the case where this alignment is temporarily broken. Which input conveys more information about the output? Of course, different scenarios might be



`./figures/fig4.pdf`

Figure 7: Same setup as in Fig. 4, but with the signal of the second proximal input channel scaled by a factor of two.

./figures/output_modulation_sweep.pdf

Figure 8: Output activity for a set of constant distal inputs (red to blue) applied after learning. Green line corresponds to the distal input also used for the learning phase.

considered here. First, we could consider the case where no input, or a constant level of input arrives at the distal compartment. This case is shown in Fig. 8. Changes in the distal input has a modulatory effect on the maximal output, but also shifts the bias. Speaking in terms of information transmission, this relates to two different effects: While changing the maximal output alters the overall impact of this neuron’s efferent connections, changing the bias changes the actual gating for proximal input. In the next scenario, we presented uncorrelated, fluctuating input to the distal compartment. In this case, we gradually increased the maximum value for the distal input from zero to one, see Fig. 9. Obviously, increased fluctuations in I_d made it harder to actually extract information about I_p from the output. Therefore, one might wonder which input stream conveys more information to the output activity, given that both are uncorrelated and have the same overall strength. This is illustrated in Fig. 10. Although the projection onto the distal input space resulted in a wider spread of activity for strong driving, a numerical estimate of the mutual information resulted in $I(I_d; y) \approx 2.75$ bits and $I(I_p; y) \approx 1.87$ bits, respectively. This indicates that in the described scenario, the distal input actually conveys more information about the output as compared to the proximal input.

1.2 Analytic Approximation of Weight Dynamics

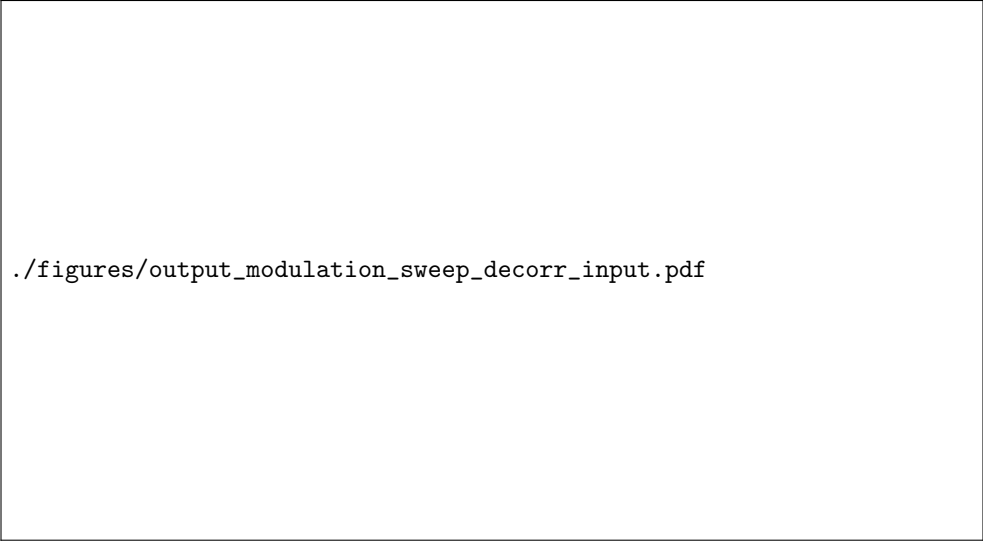
Assuming that inputs to (1) practically never reach a region where θ_{p1} becomes relevant, we remove this threshold, resulting in:

$$x(I_p, I_d) = \sigma(I_d - \theta_d) + \alpha \sigma(I_p - \theta_{p0}) \sigma(-(I_d - \theta_d)) . \quad (9)$$

We can set both thresholds to zero without loss of generality by assuming that $\langle I_p \rangle = 0$ and $\langle I_d \rangle = 0$.

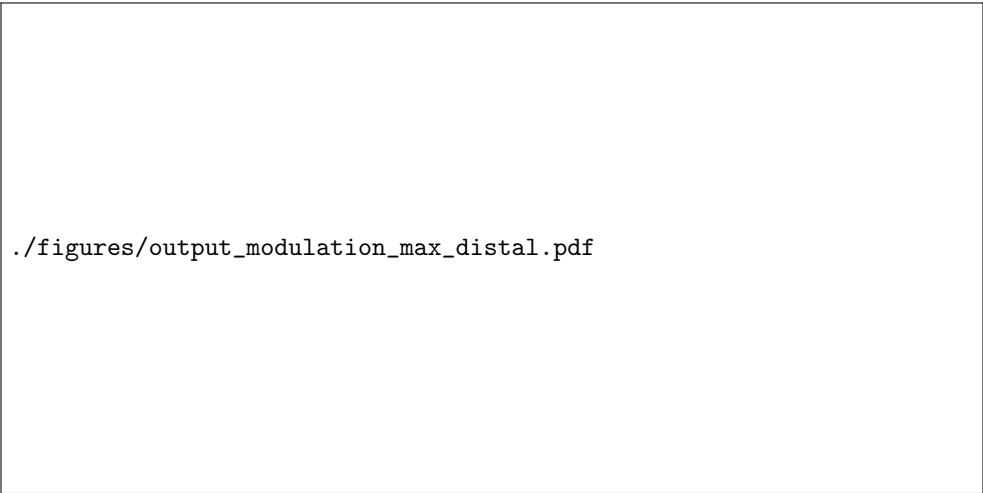
To approximate the weight dynamics, we expand the activation function around the mean input (being zero by assumption) to first order. This gives

$$y(I_p, I_d) = y_0 + p \cdot g\alpha/2 + d \cdot g(1 - \alpha/2) + \mathcal{O}(I_p^2, I_d^2) . \quad (10)$$



`./figures/output_modulation_sweep_decorr_input.pdf`

Figure 9: Output activity for a set of fluctuating distal inputs (red to blue) that are uncorrelated with the proximal input (applied after learning). Green line corresponds to the distal input also used for the learning phase.



`./figures/output_modulation_max_distal.pdf`

Figure 10: Activity for independently fluctuating values of I_d and I_p both ranging from 0 to 1, projected onto I_d (left) and I_p (right).

Plugging this approximation into (7), we get

$$\Delta w_i^p \approx \epsilon_w g (I_p \alpha/2 + I_d(1 - \alpha/2) - \langle I_p \alpha/2 + I_p(1 - \alpha/2) \rangle) (x_i^p(t) - \langle x_i^p \rangle) \quad (11)$$

Further simplification of this equation leads to

$$\Delta w_i^p \approx \epsilon'_w \left[\alpha \sum_j C_{ij}^{yy} w_j^p + (2 - \alpha) C_i^{dy} \right] \quad (12)$$

$$\epsilon'_w \equiv \epsilon_w g/2 \quad (13)$$

$$C_{ij}^{yy} \equiv \langle (y_i^p - \langle y_i^p \rangle) (y_j^p - \langle y_j^p \rangle) \rangle \quad (14)$$

$$C_i^{dy} \equiv \langle (d - \langle d \rangle) (y_i^p - \langle y_i^p \rangle) \rangle \quad (15)$$

We can compare (12) with the expression that we get if we calculate the negative derivative of the mean squared error between proximal and distal input, scaled by some arbitrary proportionality factor γ :

$$\Delta w_i^p \propto -\partial_{w_i^p} \frac{1}{2} \langle (I_p - \gamma I_d)^2 \rangle = - \sum_j C_{ij}^{yy} w_j^p + \gamma C_i^{dy} . \quad (16)$$

Note that if we set α to negative values, this equation would have the same mathematical form as (12).

What has not been included in this analysis yet is synaptic normalization. In this respect, it is important to note that the fixed point that one would get from $\Delta w_i^p = 0$, $\forall i$ is not the fixed point that is actually attained during the learning process since it is generally not compatible with the normalization condition $\|\mathbf{w}\|_1 = 1$. Rather, stable weights are achieved by a positive feedback on the weights which is constantly canceled by the synaptic normalization. By defining

$$\hat{A} \equiv \alpha \hat{C}^{yy} \quad (17)$$

$$\mathbf{b} \equiv (2 - \alpha) \mathbf{C}^{dy} \quad (18)$$

and omitting the p-superscript in the proximal weights, the actual stationarity condition is given by

$$\frac{(1 + \epsilon'_w \hat{A}) \mathbf{w} + \epsilon'_w \mathbf{b}}{\|(1 + \epsilon'_w \hat{A}) \mathbf{w} + \epsilon'_w \mathbf{b}\|_1} = \mathbf{w} \quad (19)$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm.

Expanding this expression to first order gives

$$\mathbf{w} + \epsilon'_w \left[\frac{\hat{A} \mathbf{w} + \mathbf{b}}{\|\mathbf{w}\|_1} - \frac{\mathbf{w}}{\|\mathbf{w}\|_1^2} \|\hat{A} \mathbf{w} + \mathbf{b}\|_1 \right] = \mathbf{w} \quad (20)$$

$$\hat{A} \mathbf{w} + \mathbf{b} - \mathbf{w} \|\hat{A} \mathbf{w} + \mathbf{b}\|_1 = 0 , \quad (21)$$

using the fact that $\|\mathbf{w}\|_1 = 1$ by construction. In the special case of \hat{A} being simply a rescaled identity matrix, the solution would be $\mathbf{w} = \mathbf{b}/\|\mathbf{b}\|_1$. Moreover, the fixpoint of (16) would also be aligned with $\mathbf{w} = \mathbf{b}$. An alternative case where $\mathbf{b}/\|\mathbf{b}\|_1$ would be an approximately correct solution of (19) is when \hat{C}^{yy} is weak compared to \mathbf{C}^{dy} . However, decreasing the overall scale \hat{C}^{yy} does not affect the direction of \mathbf{w} corresponding to the fixpoint of (16). Therefore, differences in the alignment between the solutions of (16) and (19) are to be expected.

It should be noted that $\mathbf{w} = \mathbf{b}/\|\mathbf{b}\|_1$ is also the stationary solution of a learning rule trying to maximize the covariance between I_p and I_d under a multiplicative normalization constraint, since (omitting means)

$$\partial_{\mathbf{w}^p} \langle I_p I_d \rangle = \langle \mathbf{y} I_d \rangle = \mathbf{C}^{dy} . \quad (22)$$

However, maximizing covariance is not necessarily the same as maximizing correlation:

$$\partial_{\mathbf{w}^p} \rho(I_p, I_p) \propto \mathbf{C}^{dy} - \left[\mathbf{w}^{pT} \hat{C}^{yy} \mathbf{w}^p \right] \left[\mathbf{w}^{pT} \mathbf{C}^{dy} \right] \hat{C}^{yy} \mathbf{w}^p \quad (23)$$

1.3 Comparison to a Two-Compartment Model with Error Minimization

In a publication by Urbanczik and Senn, somatic and dendritic interaction takes place in a two-compartment approach, where both compartments are modeled as passive capacitors [2]. Somatic spiking is modeled as a Poisson process with a rate given by the sigmoid function of the somatic voltage. In contrast to our model, dendritic plasticity is then driven by an error term between somatic spiking activity and the activity that would be present in the soma if only dendritic input was present. This allows single neurons to precisely learn a teaching signal. Furthermore, it was shown that such a learning algorithm allows the formation of associative memory in a recurrent network, as well as the emergence of feed-forward topographic maps (onto the dendritic compartment). The authors claim that using the two-compartment model justifies the use of an error driven plasticity rule. While it is reasonable that compartments could temporally store two different values—voltages in this case—which are to be compared, it is not obvious how such a comparison might be implemented. In particular, the rule proposed in the paper contains some hypothetical somatic activity, derived from the dendritic voltage, whose biological origin remains unclear.

1.4 Physiological Basis of the Architecture

- Feedback connections project from L5 to L1/2 [3]. However, there is no evidence that feedback connections terminate *exclusively* at the apical parts of L5 neurons [4]. Therefore, the functioning or mode of operation of these feedback connections might change depending on their termination site.
- In the predictive coding scheme, feedback connections are actually assumed to be mostly inhibitory [5]. The common picture is that predictive feedback tries to eliminate or cancel out input from feed-forward or later connections. A possible explanatory mechanism could be via inhibitory interneurons [6].
- Feed-forward connections project from L4/5 to L5 [3]

1.4.1 What is the Right Plasticity Rule for Distal Synapses?

While a simple Hebbian rule is a plausible candidate for proximal plasticity, experimental studies suggest that distal synapses can actually undergo LTP or LTD, depending on the mode of neural firing [7, 8]: So-called backpropagation-activated calcium spikes (BACs), which appear as bursts of action potentials, can be elicited when both basal and apical excitatory input is present. In this case, LTP (i.e. regular Hebbian-like) plasticity can be observed in the apical synapses. However, in the

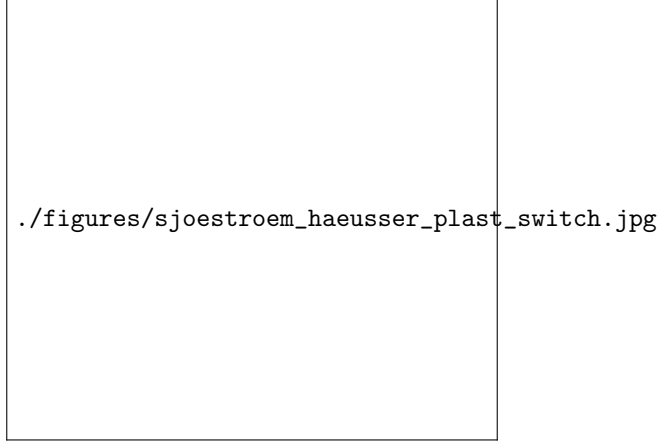


Figure 11: From Sjöström and Häusser [7]: Somatic depolarization determines whether distal synapses undergo LTD or LTP upon stimulation.

absence of this firing mode, apical synapses tend to undergo LTD, which could be described as a form of anti-Hebbian plasticity. Fig. 11 illustrates that the amount of somatic depolarization can determine the sign of synaptic weight changes in distal compartments.

On the other hand, under some conditions, LTP can occur even in the absence of somatic spiking [9], and this effect appears to be linked with long-lasting NMDA depolarizations.

1.5 Apical and Basal Compartments from an Information Theoretic Perspective

1.5.1 Partial Information Decomposition

Recently, an information theoretic approach to characterizing multiple input channels called partial information decomposition (PID, see [10]) has been proposed as a framework for the understanding of the interplay between basal and apical inputs and its effects on information processing [11]. In particular, Kay and Philips [12] analyzed the model proposed by Shai et al. [1] using PID. Roughly speaking, PID decomposes the total mutual information between neural input and output into three types of information: redundant, unique and synergistic information. In the case of a separation into apical and basal inputs, this would allow splitting the total information into a part that is provided by both input channels, unique information that is only provided by either apical or basal input, and information that is retrieved by means of the interaction between both input channels. Kay and Philips found that in the rate model by Shai et al., apical inputs provide very little unique information, but a lot of synergistic information is present. Taken together, this is interpreted as apical input acting mostly “modulatory” rather than driving onto the neural output.

1.5.2 Interaction Information

Interaction information is an extension of mutual information to more than two variables [13]. For three random variables, the interaction information $I(X; Y; Z)$ is given

by

$$I(X; Y; Z) = I(X; Y) - I(X; Y|Z) \quad (24)$$

$$= I(X; Z) - I(X; Z|Y) \quad (25)$$

$$= I(Y; Z) - I(Y; Z|X) \quad (26)$$

where e.g. $I(X; Y)$ is the mutual information between X and Y and $I(X; Y|Z)$ is the conditional mutual information. The latter is simply the expected value over Z of the mutual information between the conditional probability distribution $P(X, Y|Z)$. Interpreting the given definition, interaction information quantifies if the knowledge of a third variable increases or decreases (on average) the mutual information between two other variables. Note that this quantity is not strictly positive. A positive sign of $I(X; Y; Z)$ indicates that Z accounts for some of the correlation between X and Y . This is a rather common case, e.g. if X can be partly explained by Z . An example for negative interaction information would be a logical XOR gate $Z = \text{XOR}(X, Y)$, given that X and Y are independent binary variables. In this case, knowledge about the input Y increases the mutual information between the input X and the output Z .

Interaction information can also be used to write the mutual information $I(Z; X, Y)$ between some hypothetical output function $Z = Z(X, Y)$ and a pair of input channels X, Y as

$$I(Z; X, Y) = I(X; Z) + I(Y; Z) - I(X; Y; Z) . \quad (27)$$

That means, if interaction information is negative, more information about Z is conveyed via X, Y taken together than the sum of their individual information content about Z . In this scenario, X, Y jointly provide some cooperative information that can not be retrieved by inspecting X and Y individually, in accordance with the interpretation of negative interaction information given above. In the other case, where interaction information is positive, the actual information provided by X, Y is less than their individual information content. This could, for example indicate some form of redundancy, which, again is in accordance to the interpretation of positive interaction information previously given.

2 Possible Future Work

- As explained, plasticity in distal synapses does not follow a simple Hebbian learning scheme. Rather, experimental evidence suggests that the amount of somatic depolarization can change the sign of plasticity in distal synapses. Therefore, it might be worthwhile investigating the plasticity dynamics of a simple learning rule that incorporates these findings. Actually, if we combine a simple Hebbian learning rule $y_{\text{pre}} \cdot y_{\text{post}}$ with a multiplicative factor $(y_{\text{post}} - \theta)$ that changes sign depending on postsynaptic activity, we essentially get a BCM-like rule:

$$\Delta w_{ij} \propto y_j y_i (y_i - \theta) . \quad (28)$$

Such a type of learning rule has been studied extensively, however, has not been combined with a basal-distal compartment model.

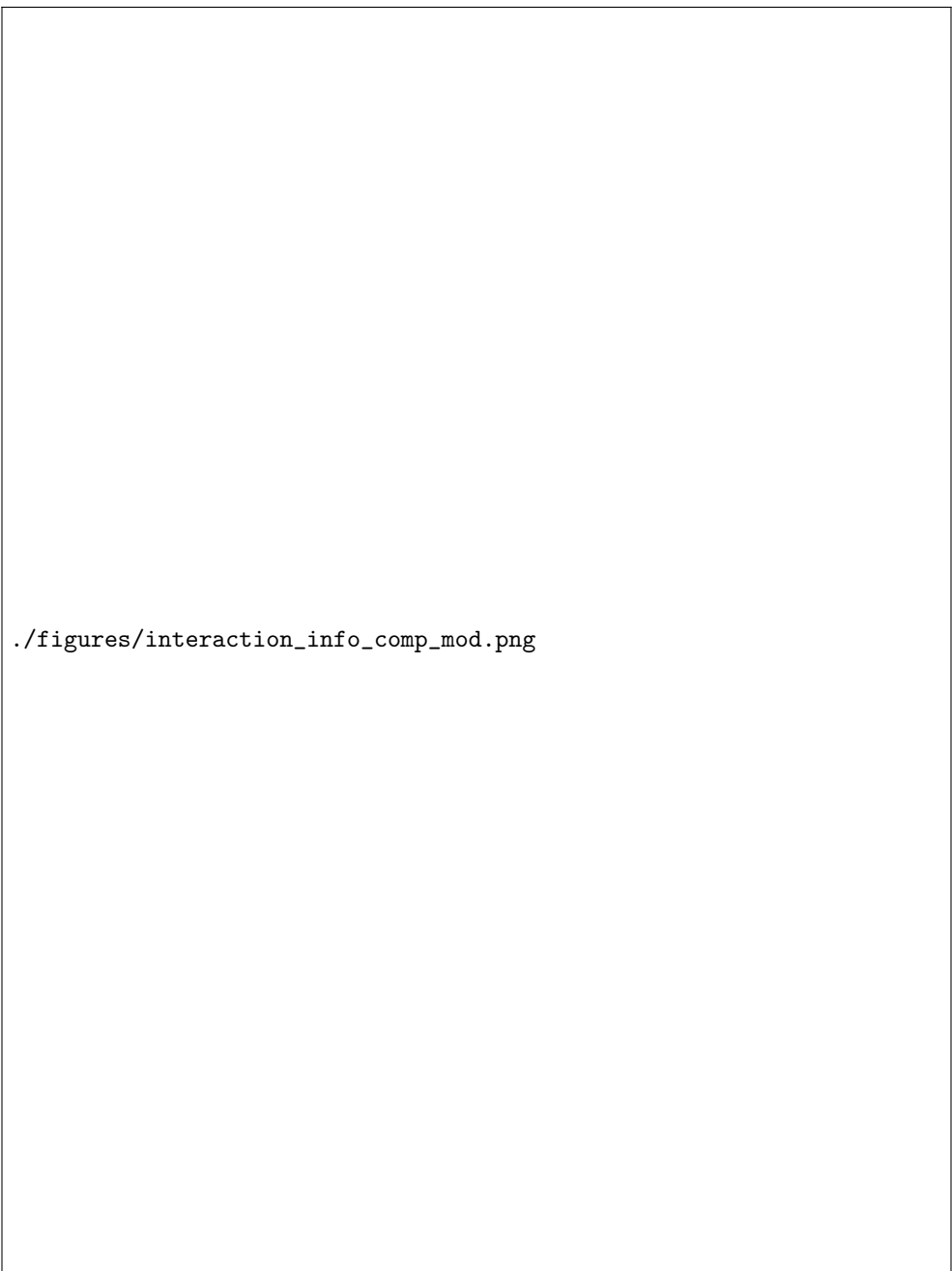
An alternative view on plasticity in distal dendrites would be to consider it a distinct synaptic integration zone. From this perspective, one could propose a Hebbian learning rule in which the postsynaptic term is just a function of the local sum of postsynaptic potentials. Similar to what we already found in

Sect. ??, one would expect such a learning rule to align weights to the PC of all distal inputs.

- Interaction information is a relatively simple quantity that provides some insight into how two (or more) variables are linked with respect to their effect on a third variable. For example, in Fig. 12, this measure was numerically calculated for different input statistics of the two-compartment firing rate model given in (1). Interestingly, increasing the correlation between proximal and distal inputs reverses the direction of the gradient in the space of proximal and distal input variances. However, this should just serve as an exemplary calculation, and it remains to be discussed whether this measure could provide insight e.g. into how such non-linear compartment model differs in its information processing to a simple point neuron model. Furthermore, one could investigate the effects of plasticity onto this quantity.
- Moldwin and Segev [14]: Perceptron learning in a biophysical compartment model. → maybe test these capabilities in our model?

References

- [1] A. S. Shai, C. A. Anastassiou, M. E. Larkum, and C. Koch. Physiology of Layer 5 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. *PLOS Computational Biology*, 11(3), 2015.
- [2] R. Urbanczik and W. Senn. Learning by the Dendritic Prediction of Somatic Spiking. *Neuron*, 81(3):521–528, feb 2014.
- [3] H. Barbas. General Cortical and Special Prefrontal Connections: Principles from Structure to Function. *Annual Review of Neuroscience*, 38(1):269–289, jul 2015.
- [4] Matthew E. Larkum, Lucy S. Petro, Robert N. S. Sachdev, and Lars Muckli. A perspective on cortical layering and layer-spanning neuronal elements. *Frontiers in Neuroanatomy*, 12:56, 2018.
- [5] Andre M. Bastos, W. Martin Usrey, Rick A. Adams, George R. Mangun, Pascal Fries, and Karl J. Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, Nov 2012.
- [6] Hanno S. Meyer, Daniel Schwarz, Verena C. Wimmer, Arno C. Schmitt, Jason N. D. Kerr, Bert Sakmann, and Moritz Helmstaedter. Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5a. *Proceedings of the National Academy of Sciences*, 108(40):16807–16812, 2011.
- [7] P. J. Sjöström and M. Häusser. A Cooperative Switch Determines the Sign of Synaptic Plasticity in Distal Dendrites of Neocortical Pyramidal Neurons. *Neuron*, 51(2):227–238, jul 2006.
- [8] 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.
- [9] Frédéric Gambino, Stéphane Pagès, Vassilis Kehayas, Daniela Baptista, Roberta Tatti, Alan Carleton, and Anthony Holtmaat. Sensory-evoked ltp driven by dendritic plateau potentials in vivo. *Nature*, 515(7525):116–119, Nov 2014.



`./figures/interaction_info_comp_mod.png`

Figure 12: Interaction information between and proximal input, distal input and the resulting firing rate of the compartment model given by (1). Inputs are Gaussian random variables with mean values of 0.5 and standard deviations σ_P and σ_D respectively, and a correlation given by ρ .

- [10] Paul L. Williams and Randall D. Beer. Nonnegative decomposition of multivariate information. *CoRR*, abs/1004.2515, 2010.
- [11] Michael Wibral, Viola Priesemann, Jim W. Kay, Joseph T. Lizier, and William A. Phillips. Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain and Cognition*, 112:25 – 38, 2017. Perspectives on Human Probabilistic Inferences and the 'Bayesian Brain'.
- [12] Jim W. Kay, W. A. Phillips, Jaan Aru, Bruce P. Graham, and Matthew E. Larkum. Bayesian modeling of bac firing as a mechanism for apical amplification in neocortical pyramidal neurons. *bioRxiv*, 2019.
- [13] William J. McGill. Multivariate information transmission. *Psychometrika*, 19(2):97–116, Jun 1954.
- [14] Toviah Moldwin and Idan Segev. Perceptron learning and classification in a modeled cortical pyramidal cell. *Frontiers in Computational Neuroscience*, 14:33, 2020.