

# Nonlinear Dendritic Coincidence Detection for Supervised Learning

Fabian Schubert<sup>1,\*</sup> and Claudius Gros<sup>1</sup>

<sup>1</sup>*Institute for Theoretical Physics, Goethe University Frankfurt am Main, Germany*

Correspondence\*:

Institute for Theoretical Physics  
Goethe University Frankfurt am Main  
Max-von-Laue-Str. 1  
60438 Frankfurt am Main, Germany  
fschubert@itp.uni-frankfurt.de

## 2 ABSTRACT

Cortical pyramidal neurons have a complex dendritic anatomy, whose function is an active research field. In particular, the segregation between its soma and the apical dendritic tree is believed to play an active role in processing feed-forward sensory information and top-down or feedback signals. In this work, we use a simple two-compartment model accounting for the nonlinear interactions between basal and apical input streams and show that standard unsupervised Hebbian learning rules in the basal compartment allow the neuron to align the feed-forward basal input with the top-down target signal received by the apical compartment. We show that this learning process, termed coincidence detection, is robust against strong distractions in the basal input space and demonstrate its effectiveness in a linear classification task.

**Keywords:** Dendrites, Pyramidal Neuron, Plasticity, Coincidence Detection, Supervised Learning

## 1 INTRODUCTION

In recent years, a growing body of research has addressed the functional implications of the distinct physiology and anatomy of cortical pyramidal neurons (Spruston, 2008; Hay et al., 2011; Ramaswamy and Markram, 2015). In particular, on the theoretical side, we saw a paradigm shift from treating neurons as point-like electrical structures towards embracing the entire dendritic structure (Larkum et al., 2009; Poirazi, 2009; Shai et al., 2015a). This was mostly due to the fact that experimental work uncovered dynamical properties of pyramidal neuronal cells that simply could not be accounted for by point models (Spruston et al., 1995; Häusser et al., 2000).

An important finding is that the apical dendritic tree of cortical pyramidal neurons can act as a separate nonlinear synaptic integration zone (Spruston, 2008; Branco and Häusser, 2011). Under certain conditions, a dendritic  $\text{Ca}^{2+}$  spike can be elicited that propagates towards the soma, causing rapid, bursting spiking activity. One of the cases in which dendritic spiking can occur was termed ‘backpropagation-activated  $\text{Ca}^{2+}$  spike firing’ (‘BAC firing’): A single somatic spike can backpropagate towards the apical spike initiation zone, in turn significantly facilitating the initiation of a dendritic spike (Stuart and Häusser, 2001; Spruston, 2008; Larkum, 2013). This reciprocal coupling is believed to act as a form of coincidence detection: If apical and basal synaptic input co-occurs, the neuron can respond with a rapid burst of spiking activity. The firing rate of these temporal bursts exceeds the firing rate that is maximally achievable under basal

synaptic input alone, therefore representing a form of temporal coincidence detection between apical and basal input.

Naturally, these mechanisms also affect plasticity and thus learning within the cortex (Sjöström and Häusser, 2006; Ebner et al., 2019). While the interplay between basal and apical stimulation and its effect on synaptic efficacies is subject to ongoing research, there is evidence that BAC-firing tends to shift plasticity towards long-term potentiation (LTP) (Letzkus et al., 2006). Thus, coincidence between basal and apical input appears to also gate synaptic plasticity.

In a supervised learning scheme, where the top down input arriving at the apical compartment acts as the teaching signal, the most straight-forward learning rule for the basal synaptic weights would be derived from an appropriate loss function, such as a mean square error, based on the difference between basal and apical input, i.e.  $I_p - I_d$ , where indices  $p$  and  $d$  denote ‘proximal’ and ‘distal’, in equivalence to basal and apical. Theoretical studies have investigated possible learning mechanisms that could utilize an intracellular error signal (Urbanczik and Senn, 2014; Schiess et al., 2016; Guerguiev et al., 2017). However, a clear experimental evidence for a physical quantity encoding such an error is—to our knowledge—yet to be found. On the other hand, Hebbian-type plasticity is extensively documented in experiments (Gustafsson et al., 1987; Debanne et al., 1994; Markram et al., 1997; Bi and Poo, 1998). Therefore, our work is based on the question whether the non-linear interactions between basal and apical synaptic input could, when combined with a Hebbian plasticity rule, allow a neuron to learn to reproduce an apical teaching signal in its proximal input.

We investigate coincidence learning by combining a phenomenological model that generates the output firing rate as a function of two streams of synaptic input (subsuming basal and apical inputs) with classical Hebbian, as well as BCM-like plasticity rules on basal synapses. In particular we hypothesized that this combination of neural activation and plasticity rules would lead to an increased correlation between basal and apical inputs. Furthermore, the temporal alignment observed in our study could potentially facilitate apical inputs to act as top-down teaching signals, without the need for an explicit error-driven learning rule. Thus, we also test our model in a simple linear supervised classification task and compare it with the performance of a simple point neuron equipped with similar plasticity rules.

## 2 MODEL

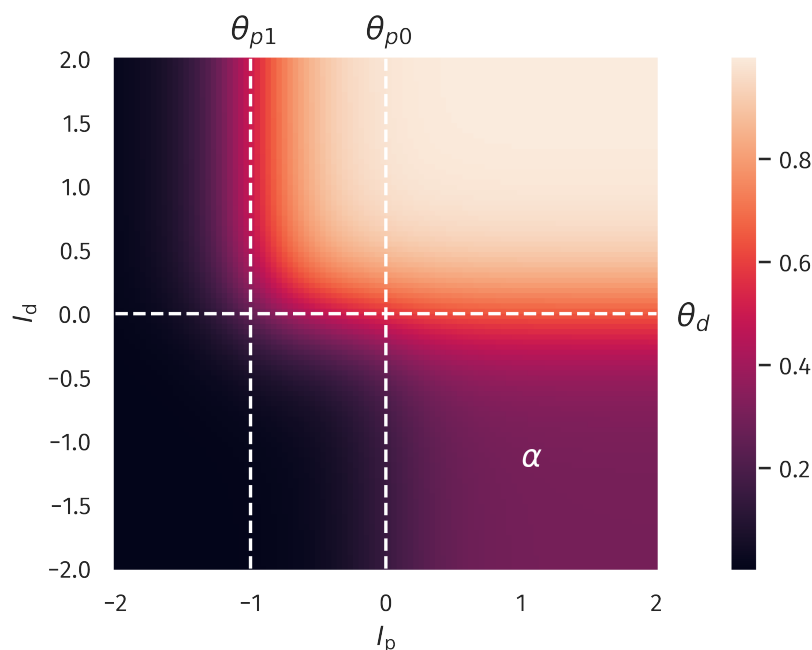
### 2.1 Compartamental Neuron

The neuron model used throughout this study is a discrete-time rate encoding model that contains two separate input variables, subsuming the total synaptic input current injected arriving at the basal (proximal) and apical (distal) dendritic structure of a pyramidal neuron, respectively. The model is a slightly simplified version of a phenomenological model proposed by Shai et al. (2015b). Denoting the input currents  $I_p$  (proximal) and  $I_d$  (distal), the model is written as

$$y(t) = \alpha \sigma(I_p(t) - \theta_{p0}) [1 - \sigma(I_d(t) - \theta_d)] + \sigma(I_d(t) - \theta_d) \sigma(I_p(t) - \theta_{p1}) \quad (1)$$

$$\sigma(x) \equiv \frac{1}{1 + \exp(-4x)} . \quad (2)$$

Here,  $\theta_{p0} > \theta_{p1}$  and  $\theta_d$  are threshold variables with respect to proximal and distal inputs. Overall, equation (1) describes two distinct regions of neural activation in the  $(I_p, I_d)$ -space which differ in their maximal firing rates, which are set to 1 and  $\alpha$ , where  $0 < \alpha < 1$ . A plot of (1) is shown in Fig. 1.



**Figure 1. Two-compartment rate model.** The firing rate as a function of proximal and distal inputs  $I_p$  and  $I_d$ , see (1). The thresholds  $\theta_{p0}$ ,  $\theta_{p1}$  and  $\theta_d$  define two regions of neural activity, with a maximal firing rate of unity and a plateau at  $\alpha = 0.3$ .

61 When both input currents  $I_d$  and  $I_p$  are large, viz larger than the thresholds  $\theta_d$  and  $\theta_{p1}$ , the second term in  
 62 (1) dominates, which leads to  $y \approx 1$ . An intermediate activity plateau, of strength  $\alpha$  emerges in addition  
 63 when  $I_p > \theta_{p0}$  and  $I_d < \theta_d$ . As such, the compartment model (1) is able to distinguish neurons with a  
 64 normal activity level, here encoded by  $\alpha = 0.3$ , and strongly bursting neurons, where the maximal firing  
 65 rate is unity. The intermediate plateau allows neurons to process the proximal inputs  $I_p$  even in the absence  
 66 of distal stimulation. The distal current  $I_d$  acts therefore as an additional modulator.

67 In our numerical experiments we compare the compartment model with a classical point neuron, as given  
 68 by

$$y(t) = \sigma(I_p(t) + I_d(t) - \theta) . \quad (3)$$

69 The apical input  $I_d$  is generated ‘as is’, meaning, it is not dynamically calculated as a superposition of  
 70 multiple presynaptic inputs. For concreteness, we used

$$I_d(t) = n_d(t)x_d(t) - b_d(t) , \quad (4)$$

71 where  $n_d(t)$  is a scaling factor,  $x_d(t)$  a pre-generated discrete time sequence and  $b_d(t)$  a bias. Note that  $n_d$   
 72 and  $b_d$  are time dependent since they are subject to adaptation processes, which will be described in the  
 73 next section. Similarly, the proximal input  $I_p(t)$  is given by

$$I_p(t) = n_p(t) \sum_{i=1}^N x_{p,i}(t)w_i(t) - b_p(t) , \quad (5)$$

74 where  $N$  is the number of presynaptic afferents,  $x_{p,i}(t)$  the corresponding sequences,  $w_i(t)$  the synaptic  
 75 efficacies and  $n_p(t)$  and  $b_p(t)$  the (time dependent) scaling and bias. Typical values for the parameters used  
 76 throughout this study are presented in Table 1.

## 2.2 Homeostatic Parameter Regulation

The bias variables entering the definitions (4) and (5) of the distal proximal current,  $I_d$  and  $I_p$ , are assumed to adapt according to

$$b_p(t+1) = b_p(t) + \mu_b [I_p(t) - I_p^t] \quad (6)$$

$$b_d(t+1) = b_d(t) + \mu_b [I_d(t) - I_d^t] , \quad (7)$$

where  $I_p^t = 0$  and,  $I_d^t = 0$  are preset targets and  $1/\mu_b = 10^3$  the timescale for the adaption process. Over time, both the distal and the proximal currents,  $I_d$  and  $I_p$ , average out.

Adaptation rules for the bias entering a transfer function, such as (7) and (6), have the task to regulate overall activity levels. The overall magnitude of the synaptic weights, which are determined by synaptic rescaling factors, here  $n_d$  and  $n_p$ , as defined in (4) and (5), will regulate in contrast the variance of the neural activity, and not the average level (Schubert and Gros, 2021). In this spirit we consider

$$n_d(t+1) = n_d(t) + \mu_n \left[ V_d^t - \left( I_d(t) - \tilde{I}_d(t) \right)^2 \right] \quad (8)$$

$$n_p(t+1) = n_p(t) + \mu_n \left[ V_p^t - \left( I_p(t) - \tilde{I}_p(t) \right)^2 \right] \quad (9)$$

$$\tilde{I}_d(t+1) = (1 - \mu_{av})\tilde{I}_d(t) + \mu_{av}I_d(t) \quad (10)$$

$$\tilde{I}_p(t+1) = (1 - \mu_{av})\tilde{I}_p(t) + \mu_{av}I_p(t) \quad (11)$$

Here,  $V_p^t$  and  $V_d^t$  define targets for the temporal averaged variances of  $I_p$  and  $I_d$ . The dynamic variables  $\tilde{I}_p$  and  $\tilde{I}_d$  are simply low-pass filtered running averages of  $I_p$  and  $I_d$ . Overall, the framework specified here allows the neuron to be fully flexible, as long as the activity level and its variance fluctuate around preset target values (Schubert and Gros, 2021). A list of the parameter values used throughout this investigation is also given in Table 1. Our choices of target means and variances are based on the assumption that neural input should be tuned towards a certain working regime of the neural transfer function. In the case of the presented model, this means that both proximal and distal input cover an area where the nonlinearities of the transfer function are reflected without oversaturation.

**Table 1.** Model parameters, as defined in sections 2.1 and 2.3.

$\theta_{p0}$	0	$V_d^t$	0.25
$\theta_{p1}$	-1	$\mu_b$	$10^{-3}$
$\theta_d$	0	$\mu_n$	$10^{-4}$
$\alpha$	0.3	$\mu_{av}$	$5 \cdot 10^{-3}$
$\mu_w$	$5 \cdot 10^{-5}$	$I_p^t$	0
$\epsilon$	0.1	$I_d^t$	0
$V_p^t$	0.25		

## 2.3 Synaptic Plasticity

The standard Hebbian plasticity rule for the proximal synaptic weights is given by

$$w_i(t+1) = w_i(t) + \mu_w [(x_{p,i}(t) - \tilde{x}_{p,i}(t)) (y(t) - \tilde{y}) - \epsilon w_i(t)] \quad (12)$$

$$\tilde{x}_{p,i}(t+1) = (1 - \mu_{av})\tilde{x}_{p,i}(t) + \mu_{av}x_{p,i}(t) \quad (13)$$

$$\tilde{y}(t+1) = (1 - \mu_{av})\tilde{y}(t) + \mu_{av}y(t) \quad (14)$$

The trailing time averages  $\tilde{x}_{p,i}$  and  $\tilde{y}$ , respectively of the presynaptic basal activities,  $x_{p,i}$ , and of the neural firing rate  $y$ , enter the Hebbian learning rule (12) as reference levels. Pre- and post-synaptic neurons are considered to be active/inactive when being above/below the respective trailing averages. The timescale of the averaging,  $1/\mu_{av}$ , is typically over 200 time steps, see Table 1. Since classical Hebbian learning does not keep weights bounded, we use an additional proportional decay term  $\epsilon w_i$  which prevents runaway growth using  $\epsilon = 0.1$ . With  $1/\mu_w = 2 \cdot 10^4$ , learning is assumed to be considerably slower, as usual for statistical update rules. For comparative reasons, the point neuron model (3) is equipped with the same plasticity rule for the proximal weights as (12).

Apart from classical Hebbian learning, we also considered a BCM-like learning rule for the basal weights (Bienenstock et al., 1982; Intrator and Cooper, 1992). The form of the BCM-rule used here reads

$$w_i(t+1) = w_i(t) + \mu_w [y(y - \theta_M)x_i - \epsilon w_i], \quad (15)$$

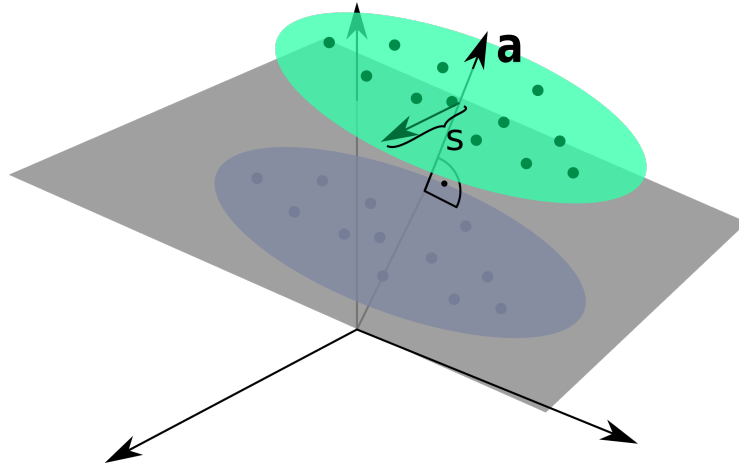
where  $\theta_M$  is a threshold defining a transition from long-term potentiation (LTP) to long-term depression (LTD) and, again,  $\epsilon$  is a decay term on the weights preventing unbounded growth. In the variant introduced by Law and Cooper (1994), the sliding threshold is simply the temporal average of the squared neural activity,  $\theta_M = \langle y^2 \rangle$ . In practice, this would be calculated as a running average, thereby preventing the weights from growing indefinitely.

However, for our compartment model, we chose to explicitly set the threshold to be the mean value between the high- and low-activity regime in our compartment model, i.e.  $\theta_M = (1 + \alpha)/2$ . By doing so, LTP is preferably induced if both basal and apical input are present at the same time. Obviously, for the point model, the reasoning behind our choice of  $\theta_M$  did not apply. Still, to provide some level of comparability, we also ran simulations with a point model where the sliding threshold was calculated as a running average of  $y^2$ .

## 3 RESULTS

### 3.1 Unsupervised Alignment between Basal and Apical Inputs

As a first test, we quantify the neuron's ability to align its basal input to the apical teaching signal. This can be done using the pearson correlation coefficient  $\rho[I_p, I_d]$  between the basal and apical input currents. We determined  $\rho[I_p, I_d]$  after the simulation, which involves all plasticity mechanisms, both for the synaptic weights and for the intrinsic parameters. The input sequences  $x_{p,i}(t)$  is randomly drawn from a uniform distribution, in  $[0, 1]$ , which is done independently for each  $i \in [1, N]$ .



**Figure 2. Input Space for the Linear Classification Task.** Two clusters of presynaptic basal activities were generated from multivariate Gaussian distributions. Here,  $s$  denotes the standard deviation orthogonal to the normal vector  $\mathbf{a}$  of the classification hyperplane, as defined by (16).

For the distal current  $I_d(t)$  to be fully ‘reconstructable’ by the basal input,  $x_d(t)$  has to be a linear combination

$$x_d(t) = \sum_{i=1}^N a_i x_{p,i}(t) \quad (16)$$

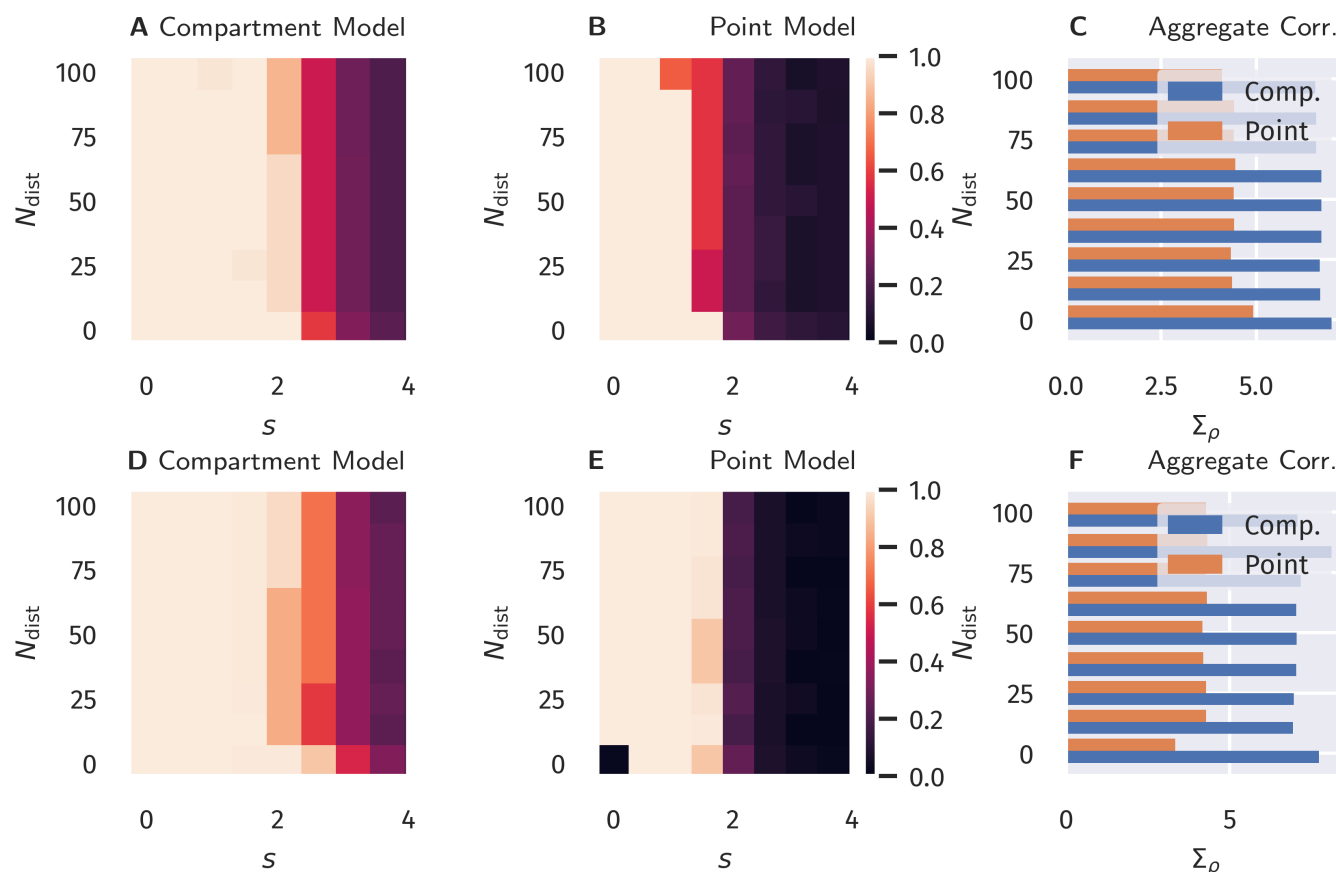
of the  $x_{p,i}(t)$ , where the  $a_i$  are the components of a random vector  $\mathbf{a}$  of unit length.

Given that we use with (12) a Hebbian learning scheme, one can expect that the direction and the magnitude of the principal components of the basal input may affect the outcome of the simulation significantly: A large variance in the basal input orthogonal to the ‘reconstruction vector’  $\mathbf{a}$  is a distraction for the plasticity. The observed temporal alignment between  $I_p$  and  $I_d$  should hence suffer when such a distraction is present.

In order to test the effects of distracting directions, we applied a transformation to the input sequences  $x_{p,i}(t)$ . For the transformation, two parameters are used, a scaling factor  $s$  and the dimension  $N_{\text{dist}}$  of the distracting subspace within the basal input space. The  $N_{\text{dist}}$  randomly generated basis vectors are orthogonal to the superposition vector  $\mathbf{a}$ , as defined by (16), and to each others. Within this  $N_{\text{dist}}$ -dimensional subspace, the input sequences  $x_{p,i}(t)$  are rescaled subsequently by the factor  $s$ . After the learning phase, a second set of input sequences  $x_{p,i}(t)$  and  $x_d(t)$  is generated for testing purposes, using the identical protocol, and the cross correlation  $\rho[I_p, I_d]$  evaluated. During the testing phase plasticity is turned off.

The overall aim of our portocal is to evaluate the degree  $\rho[I_p, I_d]$  to which the proximal current  $I_p$  aligns in the temporal domain to the distal input  $I_d$ . We recall that this is a highly non-trivial question, given that the proximal synpatic weights are adapted via Hebbian plasticity, see (12). The error  $(I_p - I_d)^2$  does not enter the adaption rules employed. Results are presented in Fig. 3 as a function of the distraction parameters  $s$  and  $N_{\text{dist}} \in [0, N - 1]$ . The total number of basal inputs is  $N = 100$ .

For a comparison, in Fig. 3 data for both the compartment model and for a point neuron are presented (as defined respectively by (1) and (3)), as well as results for both classical Hebbian and BCM learning rules. A decorrelation transition as a function of the distraction scaling paramerer  $s$  is observed for both models and plasticity rules. In terms of the learning rules, only marginal differences are present. However, the



**Figure 3. Unsupervised Alignment between Basal and Apical Input.** Color encoded is the Pearson correlation  $\rho[I_p, I_d]$  between the proximal and distal input currents,  $I_p$  and  $I_d$ . A–C: Classical Hebbian plasticity, as defined by (12). D–F: BCM rule, see (15). Data for a range  $N_{\text{dist}} \in [0, N - 1]$  of the orthogonal distraction directions, and scaling factors  $s$ , as defined in Fig. 2. The overall number of basal inputs is  $N = 100$ . In the bar plot on the right the sum  $\Sigma_{\text{acc}}$  over  $s = 0, 0.5, 1.0 \dots$  of the results is shown as a function of  $N_{\text{dist}}$ . Blue bars represents the compartment model, orange the point model.

138 compartment model is able to handle a significantly stronger distraction as compared to the point model.  
 139 These findings support the hypothesis examined here, namely that nonlinear interactions between basal and  
 140 apical input improve learning guided by top-down signals.

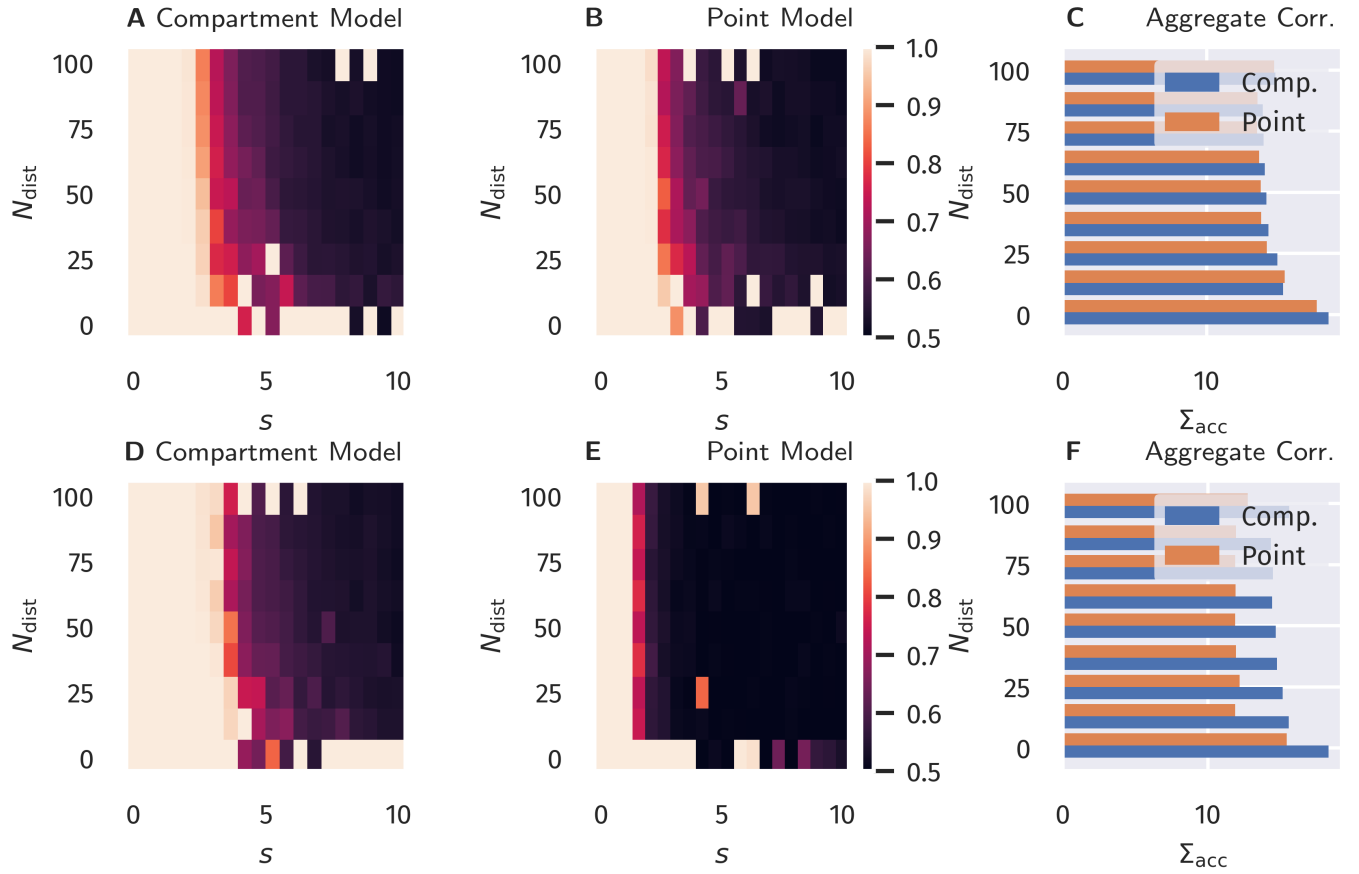
### 141 3.2 Supervised Learning in a Linear Classification Task

142 Next, we investigated if the observed differences would also improve the performance in an actual  
 143 supervised learning task. For this purpose, we constructed presynaptic basal input  $x_p(t)$  as illustrated in  
 144 Fig. 2. Written in vector form, each sample from the basal input is generated from,

$$\mathbf{x}_p(t) = \mathbf{b} + \mathbf{a}[c(t) + \sigma_a \zeta_a(t)] + s \cdot \sum_{i=1}^{N_{\text{dist}}} \zeta_{\text{dist},i}(t) \mathbf{v}_{\text{dist},i}, \quad (17)$$

145 where  $\mathbf{b}$  is a random vector drawn uniformly from  $(0, 1)^N$ ,  $\mathbf{a}$  is random unit vector as introduced in  
 146 Section 3.1,  $c(t)$  is a binary variable drawn from  $\{-0.5, 0.5\}$  with equal probability and  $\zeta_a(t)$  and the  
 147  $\zeta_{\text{dist},i}(t)$  are independent random Gaussian variables with zero mean and unit variance. Hence,  $\sigma_a$  simply  
 148 denotes the standard deviation of each Gaussian cluster along the direction of the normal vector  $\mathbf{a}$  and  
 149 was set to  $\sigma_a = 0.25$ . Finally, the set of  $\mathbf{v}_{\text{dist},i}$  forms a randomly generated orthogonal basis of  $N_{\text{dist}}$





**Figure 4. Binary Classification Accuracy.** Fraction of correctly classified patterns as illustrated in Fig. 2, see Section 3.2. A–C: Classical Hebbian plasticity. D–F: BCM rule. In the bar plot on the right the sum  $\Sigma_{\text{acc}}$  over  $s = 0, 0.5, 1.0 \dots$  of the results is given as a function of  $N_{\text{dist}}$ . Blue bars represents the compartment model, orange the point model.

150 unit vectors which are—as in Section 3.1—also orthogonal to  $\mathbf{a}$ . The free parameter  $s$  parameterizes the  
 151 standard deviation along this subspace orthogonal to  $\mathbf{a}$ . As indicated by the time dependence, the Gaussian  
 152 and binary random variables are drawn for each time step. The vectors  $\mathbf{b}$ ,  $\mathbf{a}$ , and  $\mathbf{v}_{\text{dist},i}$  are generated once  
 153 before the beginning of a simulation run.

For the classification task, we use two output neurons, indexed 0 and 1, receiving the same basal presynaptic input, with the respective top-down inputs  $x_{d,0}$  and  $x_{d,1}$  encoding the desired linear classification in a one-hot scheme,

$$x_{d,0}(t) = 1 - \Theta \left( (\mathbf{x}_p(t) - \mathbf{b})^T \mathbf{a} \right) \quad (18)$$

$$x_{d,1}(t) = \Theta \left( (\mathbf{x}_p(t) - \mathbf{b})^T \mathbf{a} \right), \quad (19)$$

154 where  $\Theta(x)$  is the Heaviside step function.

155 As in the previous experiment, we ran a full simulation until all dynamic variables reached a stationary  
 156 state. After this, a test run without plasticity and with the apical input turned off was used to evaluate the  
 157 classification performance. For each sample, the index of the neuron with the highest activity was used as  
 158 the predicted class. Accuracy was then calculated as the fraction of correctly classified samples.



159 The resulting accuracy as a function of  $N_{\text{dist}}$  and  $s$  is shown in Fig. 4, again for all four combinations of  
160 neuron models and learning rules.

161 For classical Hebbian plasticity, the differences between compartmental and point neuron are small.  
162 Interestingly, the compartment model performs measurably better in the case of the BCM rule (15), in  
163 particular when the overall accuracies for the tested parameter range are compared, see Fig. 4D. This  
164 indicates that the compartmental neuron makes better use, during learning, of the three distinct activity  
165 plateaus at 0,  $\alpha$  and 1, when the BCM rule is at work. Compare Fig. 1. We point out in this respect that the  
166 sliding threshold  $\theta_M$  in (15) has been set to the half-way point between the two non-trivial activity levels,  
167  $\alpha$  and 1.

168 It should be noted that the advantage of the compartment model is also reflected in the actual correlation  
169 between proximal and distal input as a measure of successful learning (as done in the previous section), see  
170 Fig. 5 in the appendix. Interestingly, the discrepancies are more pronounced when measuring the correlation  
171 as compared to the accuracy. Moreover, it appears that above-chance accuracy is still present for parameter  
172 values where alignment is almost zero. We attribute this effect to the fact that the classification procedure  
173 predicts the class by choosing the node that has the higher activity, independent of the actual “confidence”  
174 of this prediction, i.e. how strong activities differ relative to their actual activity levels. Therefore, marginal  
175 differences can still yield the correct classification in this isolated setup, but it would be easily disrupted by  
176 finite levels of noise or additional external input.

## 4 DISCUSSION

177 The workhorse of the brain, pyramidal neurons, possess distinct apical/basal (distant/proximal) dendritic  
178 trees. It is hence likely that models with at least two compartments are necessary for describing the  
179 functionality of pyramidal neurons. For a proposed two-compartment transfer function (Shai et al., 2015b),  
180 we have introduced both unsupervised and supervised learning schemes, showing that the two-compartment  
181 neuron is significantly more robust against distracting components in the proximal input space than a  
182 corresponding (one-compartment) point neuron.

183 The apical and basal dendritic compartments of pyramidal neurons are located in different cortical layers  
184 Park et al. (2019), receiving top-down and feed-forward signals, respectively. The combined action of these  
185 two compartments is hence the prime candidate for the realization of backpropagation in multi-layered  
186 networks (Bengio, 2014; Lee et al., 2015; Guerguiev et al., 2017).

187 In the past, backpropagation algorithms for pyramidal neurons concentrated on learning rules that are  
188 explicitly dependent on an error term, typically the difference between top-down and bottom up signals.  
189 In this work, we considered an alternative approach. We postulate that the correlation between proximal  
190 and distal input constitutes a viable objective function, which is to be maximized in combination with  
191 homeostatic adaptation rules that keeps proximal and distal inputs within desired working regimes. Learning  
192 correlations between distinct synaptic or compartmental inputs is as standard task for Hebbian-type learning,  
193 which implies that the here proposed framework is based not on supervised, but on biologically viable  
194 unsupervised learning schemes.

195 The proximal input current  $I_p$  is a linear projection of the proximal input space. Maximizing the  
196 correlation between  $I_p$  and  $I_d$  (the distal current), can therefore be regarded as a form of canonical  
197 correlation analysis (CCA) (Härdle and Simar, 2007). The idea of using CCA as a possible mode of  
198 synaptic learning has previously been investigated by Haga and Fukai (2018). Interestingly, according to  
199 the authors, a BCM-learning term in the plasticity dynamics accounts for a principal component analysis

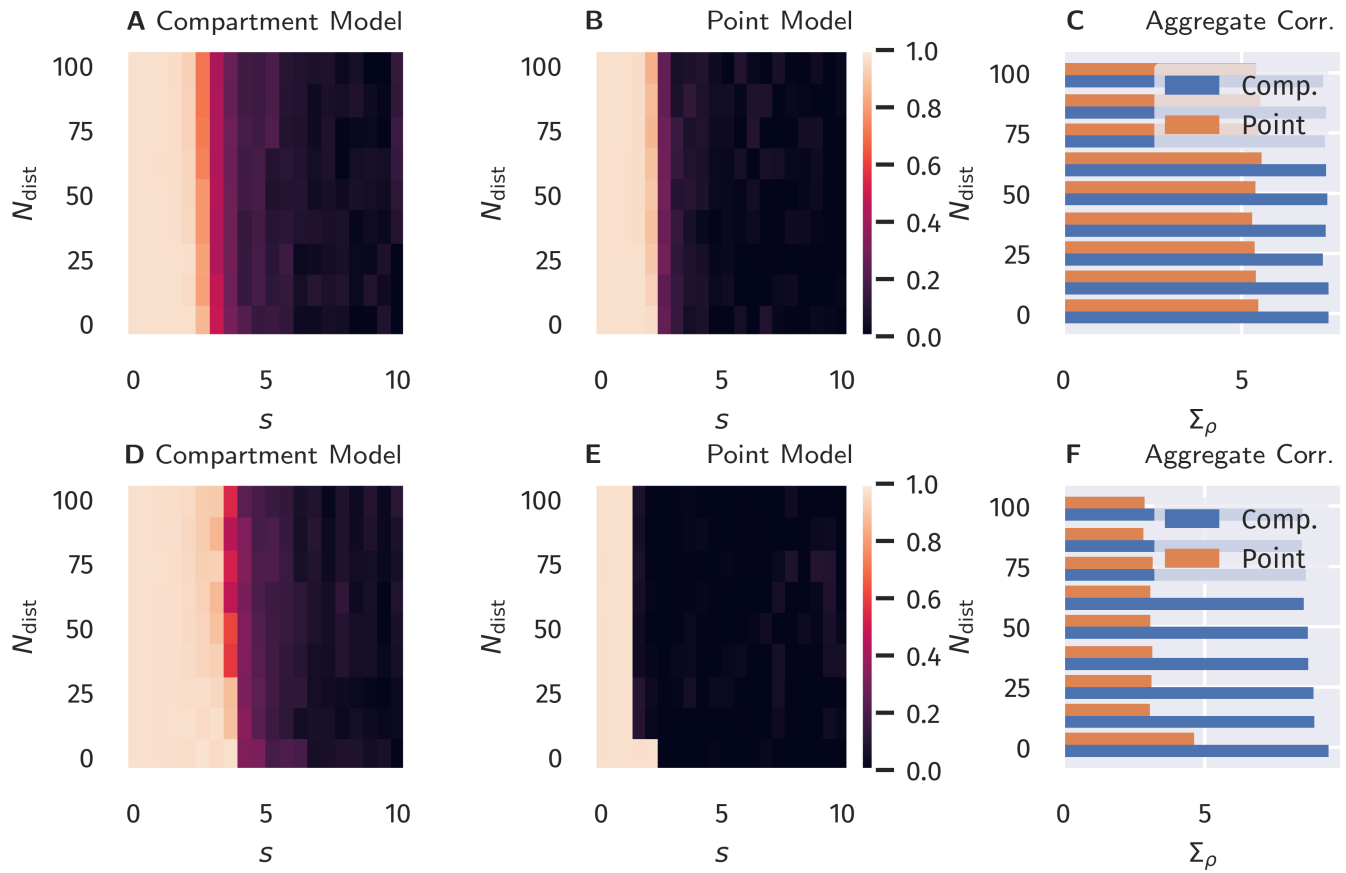
in the input space, while CCA requires an additional multiplicative term between local basal and apical activity. In contrast, our results indicate that such a multiplicative term is not required to drive basal synaptic plasticity towards a maximal alignment between basal and apical input, even in the presence of distracting principal components. Apart from the advantage that this avoids the necessity of giving a biophysical interpretation of such cross-terms, it is also in line with the view that synaptic plasticity should be formulated in terms of local membrane voltage traces (Clopath et al., 2010; Weissenberger et al., 2018). According to this principle, distal compartments should therefore only implicitly affect plasticity in basal synapses, e.g. by facilitating spike initiation.

Here we concentrated on one-dimensional distal inputs. For the case of higher-dimensional distal input patterns, as for structured multi-layered networks, it thus remains to be investigated how target signals are formed. However, as previous works have indicated, random top-down weights are generically sufficient for successful credit assignment and learning tasks (Lillicrap et al., 2016; Guerguiev et al., 2017). We therefore expect that our results can be transferred also to deep network structures, for which plasticity is classically guided by local errors between top-down and bottom-up signals.

## 5 APPENDIX

### 5.1 Alignment in the Classification Task

Instead of measuring the model performance in the classification task presented in Sect. 3.2 by the fraction of correctly classified patterns, as shown in Fig. 4, one can also use the correlation between  $I_p$  and  $I_d$ , as done in Sect. 3.1. This is shown in Fig. 5. One observes a more pronounced difference between the point model and the compartment model, where the latter results in an overall better alignment for the tested parameter space.



**Figure 5. Alignment between Basal and Apical Input after Binary Classification Learning.** Correlation between proximal and distal inputs after training, as described in Sect. 3.2. A–C: Classical Hebbian plasticity. D–F: BCM rule. In the bar plot on the right the sum  $\Sigma_{\text{acc}}$  over  $s = 0, 0.5, 1.0$  of the results is shown as a function of  $N_{\text{dist}}$ . Blue bars represents the compartment model, orange the point model.

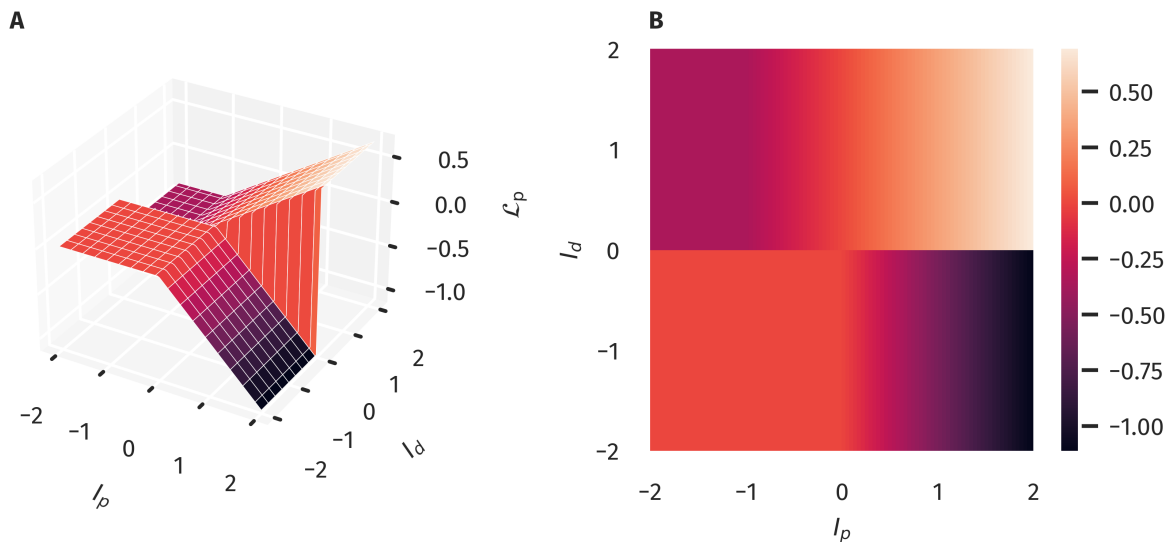
## 5.2 Objective Function of BCM Learning in the Compartment Model

To gain a better understanding of why the BCM-type learning rule in combination with the implemented compartment model drives the neuron towards the temporal alignment between  $I_p$  and  $I_d$ , we can formalize the learning rule for the proximal weights in terms of an objective function. For this purpose, we further simplify (1) by replacing the sigmoid functions  $\sigma(x)$  by a simple step function  $\Theta(x)$ . This does not change the overall shape or topology of the activation in the  $(I_p, I_d)$  space but merely makes the smooth transitions sharp and instantaneous. Using  $\Delta w_i \propto y(y - \theta_M)x_i$ , we find in this case

$$\Delta w_i \propto \left[ (1 - \alpha)\Theta(I_d - \theta_d)\Theta(p - \theta_{p1}) + \alpha(\alpha - 1)\Theta(\theta_d - I_d)\Theta(p - \theta_{p0}) \right] x_i. \quad (20)$$

Noting that  $\Theta(x)$  is the first derivative of the ReLu function  $[x]^+ \equiv \max(0, x)$ , we find that this update rule can be written as

$$\begin{aligned} \Delta w_i &\propto \frac{\partial \mathcal{L}_p}{\partial w_i} \\ \mathcal{L}_p &= (1 - \alpha)\Theta(I_d - \theta_d)[p - \theta_{p1}]^+ + \alpha(\alpha - 1)\Theta(\theta_d - I_d)[p - \theta_{p0}]^+. \end{aligned} \quad (21)$$



**Figure 6. Objective Function for the Proximal Weight Update.** The approximate objective function for the proximal weights as given in (21) as a 3d-plot (A) and color-coded (B). This corresponds to a combination of using (1) together with (15). Note the ridge-like structure along the  $I_p$ - $I_d$  diagonal, which supports the alignment between proximal and distal input.

229 The objective function  $\mathcal{L}_p$  is shown in Fig. 6. One observes that states closer to the  $I_p$ - $I_d$  diagonal are  
 230 preferred since they tend to yield higher values of  $\mathcal{L}_p$ , while the opposite is the case for off-diagonal states.

231 It should be noted, though, that the objective function is not scale-invariant (as would be e.g. if the  
 232 squared error was used) in the sense that the prior distributions of both proximal and distal inputs need  
 233 a certain mean and variance to cover a region of input states for which the described effects can take  
 234 place. As a counterexample, one could imagine that the input samples only covered a flat area of  $\mathcal{L}_p$ , as  
 235 for example in Fig. 6B in the lower left quadrant, leading to a zero average gradient. This is prevented,  
 236 however, by the homeostatic processes acting simultaneously on the gains and biases, making sure that  
 237 the marginal distributions of  $I_p$  and  $I_d$  are such that higher correlations are preferred. For example, if we  
 238 assume a Gaussian marginal distribution for both  $I_p$  and  $I_d$  with zero means and a standard deviation of 0.5  
 239 (which is used as a homeostatic target in the simulations), the expected value of  $\mathcal{L}(I_p, I_d)$  is  $-0.055$  if  $I_p$   
 240 and  $I_d$  are completely uncorrelated, and  $0.07$  in the perfectly correlated case.

## CONFLICT OF INTEREST STATEMENT

241 The authors declare that the research was conducted in the absence of any commercial or financial  
 242 relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

243 Both authors, F.S. and C.G., contributed equally to the writing and review of the manuscript. F.S. provided  
 244 the code, ran the simulations and prepared the figures.

## ACKNOWLEDGMENTS

245 The authors acknowledge the financial support of the German research foundation (DFG)

## DATA AVAILABILITY STATEMENT

246 The simulation datasets for this study can be found under <https://cloud.itp.uni-frankfurt.de/s/mSRJ6BPXjwwHmfq>. The simulation and plotting code for this project can be found under  
247 [https://github.com/FabianSchubert/frontiers\\_dendritic\\_coincidence\\_detection](https://github.com/FabianSchubert/frontiers_dendritic_coincidence_detection).  
248

## REFERENCES

- 249 Bengio, Y. (2014). How Auto-Encoders Could Provide Credit Assignment in Deep Networks via Target  
250 Propagation
- 251 Bi, G. Q. and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on  
252 spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* 18, 10464–10472.  
253 doi:10.1523/jneurosci.18-24-10464.1998
- 254 Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron  
255 selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience* 2,  
256 32–48. doi:10.1523/jneurosci.02-01-00032.1982
- 257 Branco, T. and Häusser, M. (2011). Synaptic Integration Gradients in Single Cortical Pyramidal Cell  
258 Dendrites. *Neuron* 69, 885–892. doi:10.1016/j.neuron.2011.02.006
- 259 Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: A model of  
260 voltage-based STDP with homeostasis. *Nature Neuroscience* 13, 344–352. doi:10.1038/nn.2479
- 261 Debanne, D., Gähwiler, B. H., and Thompson, S. M. (1994). Asynchronous pre- and postsynaptic activity  
262 induces associative long-term depression in area CA1 of the rat hippocampus in vitro. *Proceedings of*  
263 *the National Academy of Sciences of the United States of America* 91, 1148–1152. doi:10.1073/pnas.91.  
264 3.1148
- 265 Ebner, C., Clopath, C., Jedlicka, P., and Cuntz, H. (2019). Unifying Long-Term Plasticity Rules for  
266 Excitatory Synapses by Modeling Dendrites of Cortical Pyramidal Neurons. *Cell Reports* 29, 4295–  
267 4307.e6. doi:10.1016/j.celrep.2019.11.068
- 268 Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites.  
269 *eLife* 6. doi:10.7554/eLife.22901
- 270 Gustafsson, B., Wigstrom, H., Abraham, W. C., and Huang, Y. Y. (1987). Long-term potentiation in the  
271 hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic  
272 potentials. *Journal of Neuroscience* 7, 774–780. doi:10.1523/jneurosci.07-03-00774.1987
- 273 Haga, T. and Fukai, T. (2018). Dendritic processing of spontaneous neuronal sequences for single-trial  
274 learning. *Scientific Reports* 8, 15166. doi:10.1038/s41598-018-33513-9
- 275 Härdle, W. and Simar, L. (2007). Canonical Correlation Analysis. In *Applied Multivariate Statistical*  
276 *Analysis* (Berlin, Heidelberg: Springer Berlin Heidelberg). 321–330. doi:10.1007/978-3-540-72244-1\  
277 \_14
- 278 [Dataset] Häusser, M., Spruston, N., and Stuart, G. J. (2000). Diversity and dynamics of dendritic signaling.  
279 doi:10.1126/science.290.5492.739
- 280 Hay, E., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2011). Models of Neocortical Layer  
281 5b Pyramidal Cells Capturing a Wide Range of Dendritic and Perisomatic Active Properties. *PLoS*  
282 *Computational Biology* 7, e1002107. doi:10.1371/journal.pcbi.1002107
- 283 Intrator, N. and Cooper, L. N. (1992). Objective function formulation of the BCM theory of visual  
284 cortical plasticity: Statistical connections, stability conditions. *Neural Networks* 5, 3–17. doi:10.1016/  
285 S0893-6080(05)80003-6
- 286 [Dataset] Larkum, M. (2013). A cellular mechanism for cortical associations: An organizing principle for  
287 the cerebral cortex. doi:10.1016/j.tins.2012.11.006

- 288 Larkum, M. E., Nevian, T., Sandier, M., Polsky, A., and Schiller, J. (2009). Synaptic integration  
 289 in tuft dendrites of layer 5 pyramidal neurons: A new unifying principle. *Science* 325, 756–760.  
 290 doi:10.1126/science.1171958
- 291 Law, C. C. and Cooper, L. N. (1994). Formation of receptive fields in realistic visual environments  
 292 according to the Bienenstock, Cooper, and Munro (BCM) theory. *Proceedings of the National Academy*  
 293 *of Sciences of the United States of America* 91, 7797–7801. doi:10.1073/pnas.91.16.7797
- 294 Lee, D. H., Zhang, S., Fischer, A., and Bengio, Y. (2015). Difference target propagation. In *Lecture Notes*  
 295 *in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*  
 296 *Bioinformatics)* (Springer Verlag), vol. 9284, 498–515. doi:10.1007/978-3-319-23528-8\\_31
- 297 Letzkus, J. J., Kampa, B. M., and Stuart, G. J. (2006). Learning Rules for Spike Timing-Dependent  
 298 Plasticity Depend on Dendritic Synapse Location. *Journal of Neuroscience* 26, 10420–10429. doi:10.  
 299 1523/JNEUROSCI.2650-06.2006
- 300 Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback  
 301 weights support error backpropagation for deep learning. *Nature Communications* 7, 1–10. doi:10.1038/  
 302 ncomms13276
- 303 Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by  
 304 coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215. doi:10.1126/science.275.5297.213
- 305 Park, J., Papoutsi, A., Ash, R. T., Marin, M. A., Poirazi, P., and Smirnakis, S. M. (2019). Contribution  
 306 of apical and basal dendrites to orientation encoding in mouse v1 l2/3 pyramidal neurons. *Nature*  
 307 *communications* 10, 1–11
- 308 Poirazi, P. (2009). Information processing in single cells and small networks: Insights from compartmental  
 309 models. In *AIP Conference Proceedings* (American Institute of Physics), vol. 1108, 158–167. doi:10.  
 310 1063/1.3117124
- 311 [Dataset] Ramaswamy, S. and Markram, H. (2015). Anatomy and physiology of the thick-tufted layer 5  
 312 pyramidal neuron. doi:10.3389/fncel.2015.00233
- 313 Schiess, M., Urbanczik, R., and Senn, W. (2016). Somato-dendritic Synaptic Plasticity and Error-  
 314 backpropagation in Active Dendrites. *PLoS Computational Biology* 12, 1004638. doi:10.1371/journal.  
 315 pcbi.1004638
- 316 Schubert, F. and Gros, C. (2021). Local homeostatic regulation of the spectral radius of echo-state networks.  
 317 *Frontiers in computational neuroscience* 15, 12
- 318 Shai, A. S., Anastassiou, C. A., Larkum, M. E., and Koch, C. (2015a). Physiology of Layer 5  
 319 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. *PLOS*  
 320 *Computational Biology* 11
- 321 Shai, A. S., Anastassiou, C. A., Larkum, M. E., and Koch, C. (2015b). Physiology of Layer 5  
 322 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. *PLOS*  
 323 *Computational Biology* 11
- 324 Sjöström, P. J. and Häusser, M. (2006). A Cooperative Switch Determines the Sign of Synaptic Plasticity  
 325 in Distal Dendrites of Neocortical Pyramidal Neurons. *Neuron* 51, 227–238. doi:10.1016/j.neuron.2006.  
 326 06.017
- 327 Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews*  
 328 *Neuroscience* 9, 206–221. doi:10.1038/nrn2286
- 329 Spruston, N., Schiller, Y., Stuart, G., and Sakmann, B. (1995). Activity-dependent action potential invasion  
 330 and calcium influx into hippocampal CA1 dendrites. *Science* 268, 297–300. doi:10.1126/science.  
 331 7716524

- 332 Stuart, G. J. and Häusser, M. (2001). Dendritic coincidence detection of EPSPs and action potentials.  
333 *Nature Neuroscience* 4, 63–71. doi:10.1038/82910
- 334 Urbanczik, R. and Senn, W. (2014). Learning by the Dendritic Prediction of Somatic Spiking. *Neuron* 81,  
335 521–528. doi:10.1016/j.neuron.2013.11.030
- 336 Weissenberger, F., Gauy, M. M., Lengler, J., Meier, F., and Steger, A. (2018). Voltage dependence of  
337 synaptic plasticity is essential for rate based learning with short stimuli. *Scientific Reports* 8, 4609.  
338 doi:10.1038/s41598-018-22781-0