

Author's Proof

Before checking your proof, please see the instructions below.

- Carefully read the entire proof and mark all corrections in the appropriate place, using the Adobe Reader commenting tools ([Adobe Help](#)).
- Provide your corrections in a single PDF file or post your comments in the Production Forum making sure to reference the relevant query/line number. Upload or post all your corrections directly in the Production Forum to avoid any comments being missed.
- We do not accept corrections in the form of edited manuscripts nor via email.
- Before you submit your corrections, please make sure that you have checked your proof carefully as once you approve it, you won't be able to make any further corrections.
- To ensure the timely publication of your article, please submit the corrections within 48 hours. After submitting, do not email or query asking for confirmation of receipt.

Do you need help? Visit our [Production Help Center](#) for more information. If you can't find an answer to your question, contact your Production team directly by posting in the Production Forum.

Quick Check-List

- Author names** - Complete, accurate and consistent with your previous publications.
- Affiliations** - Complete and accurate. Follow this style when applicable: Department, Institute, University, City, Country.
- Tables** - Make sure our formatting style did not change the meaning/alignment of your Tables.
- Figures** - Make sure we are using the latest versions.
- Funding and Acknowledgments** - List all relevant funders and acknowledgments.
- Conflict of Interest** - Ensure any relevant conflicts are declared.
- Supplementary files** - Ensure the latest files are published and that no line numbers and tracked changes are visible.
Also, the supplementary files should be cited in the article body text.
- Queries** - Reply to all typesetters queries below.
- Content** - Read all content carefully and ensure any necessary corrections are made.

Author Queries Form

Query No.	Details Required	Author's Response
Q1	The citation and surnames of all of the authors have been highlighted. Check that they are correct and consistent with the authors' previous publications, and correct if need be. Please note that this may affect the indexing of your article in repositories such as PubMed.	
Q2	Confirm whether the insertion of the article title is correct.	
Q3	Confirm that all author affiliations are correctly listed. Note that affiliations are listed sequentially as per journal style and requests for non-sequential listing will not be applied. Note that affiliations should reflect those at the time during which the work was undertaken.	
Q4	Confirm that the email address in your correspondence section is accurate.	

Query No.	Details Required	Author's Response
Q5	Confirm that the short running title is correct, making sure to keep it to a maximum of five words.	
Q6	Confirm that the keywords are correct and keep them to a maximum of eight and a minimum of five. (Note: a keyword can be comprised of one or more words.) Note that we have used the keywords provided at Submission. If this is not the latest version, please let us know.	
Q7	Check if the section headers (i.e., section leveling) were correctly captured.	
Q8	If you decide to use previously published, copyrighted figures in your article, please keep in mind that it is your responsibility, as the author, to obtain the appropriate permissions and licenses and to follow any citation instructions requested by third-party rights holders. If obtaining the reproduction rights involves the payment of a fee, these charges are to be paid by the authors.	
Q9	Ensure that all the figures, tables and captions are correct, and that all figures are of the highest quality/resolution. Please note that Figures and Tables must be cited sequentially.	
Q10	Verify that all the equations and special characters are displayed correctly.	
Q11	Confirm that the Data Availability statement is accurate. Note that we have used the statement provided at Submission. If this is not the latest version, please let us know.	
Q12	Confirm that the details in the “Author Contributions” section are correct and note that we have added the sentence “Both authors contributed to the article and approved the submitted version.”	
Q13	Ensure to add all grant numbers and funding information, as after publication this will no longer be possible. All funders should be credited and all grant numbers should be correctly included in this section.	
Q14	Confirm if the text included in the Conflict of Interest statement is correct.	
Q15	We have replaced the section head “Acknowledgments” with “Funding.” Please confirm that this is correct.	
Q16	We received the Supplementary information section inside the manuscript, please provide the Appendix as a separate supplementary file.	



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77

Nonlinear Dendritic Coincidence Detection for Supervised Learning

Fabian Schubert* and Claudius Gros

Institute for Theoretical Physics, Goethe University Frankfurt am Main, Frankfurt am Main, Germany

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

OPEN ACCESS

Edited by:

Joaquín J. Torres,
University of Granada, Spain

Reviewed by:

Guy Elston,
Centre for Cognitive Neuroscience
Ltd., Australia
Thomas Wennekers,
University of Plymouth,
United Kingdom

*Correspondence:

Fabian Schubert
fschubert@itp.uni-frankfurt.de

Received: 31 May 2021

Accepted: 13 July 2021

Published: xx July 2021

Citation:

Schubert F and Gros C (2021)
Nonlinear Dendritic Coincidence
Detection for Supervised Learning.
Front. Comput. Neurosci. 15:718020.
doi: 10.3389/fncom.2021.718020

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 toward embracing the entire dendritic structure (Larkum et al., 2009; Poirazi, 2009; Shai et al., 2015). 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 Ca^{2+} spike can be elicited that propagates toward the soma, causing rapid, bursting spiking activity. One of the cases in which dendritic spiking can occur was termed ‘backpropagation-activated Ca^{2+} spike firing’ (“BAC firing”): A single somatic spike can backpropagate toward 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 toward 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 of whether the nonlinear 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. Compartmental Neuron

The neuron model used throughout this study is a discrete-time rate encoding model that contains two separate input variables, representing 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. (2015). 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 θ_d are threshold variables with respect to proximal and distal inputs. Equation (1) defines the firing rate y as a function of I_p and I_d . Note that the firing rate is normalized to take values within $y \in [0, 1]$. In the publication by Shai et al. (2015), firing rates varied between 0 and 150 Hz. High firing rates typically appear in the form of bursts of action potentials, lasting on the order of 50–100 ms (Larkum et al., 1999; Shai et al., 2015). Therefore, since our model represents “instantaneous” firing rate responses to a discrete set of static input patterns, we conservatively estimate the time scale of our model to be on the order of tenths of seconds.

In general, the input currents I_p and I_d are meant to comprise both excitatory and potential inhibitory currents. Therefore, we did not restrict the sign of I_p and I_d to positive values. Moreover, since we chose the thresholds θ_{p0} and θ_d to be zero, I_p and I_d should be rather seen as a total external input relative to intrinsic firing thresholds.

Note that the original form of this phenomenological model by Shai et al. (2015) is of the form

$$y(I_p, I_d) = \sigma(I_p - A\sigma(I_d)) [1 + B\sigma(I_d)] , \quad (3)$$

where σ denotes the same sigmoidal activation function. This equation illustrates that I_d has two effects: It shifts the basal activation threshold by a certain amount (here controlled by the parameter A) and also multiplicatively increases the maximal firing rate (to an extent controlled by B). Our equation mimics these effects by means of the two thresholds θ_{p0} and θ_{p1} , as well as the value of α relative to the maximal value of y (which is 1 in our case).

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 α , where $0 < \alpha < 1$. A plot of Equation (1) is shown in **Figure 1**.

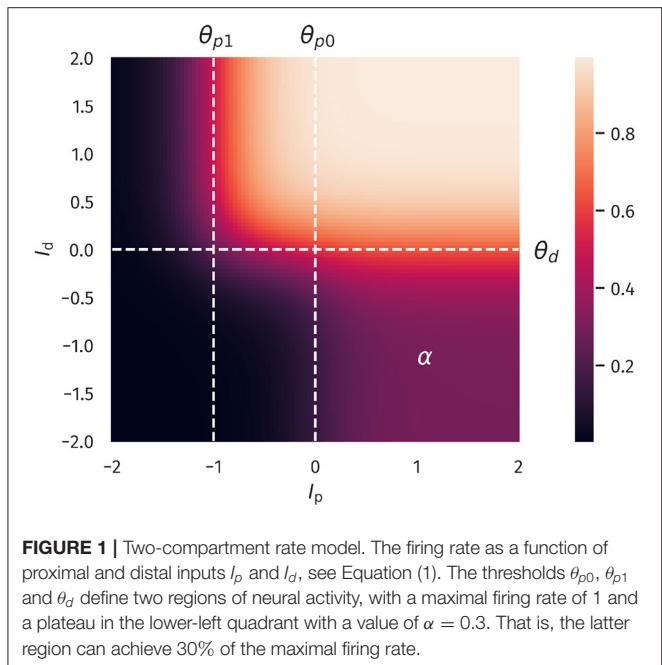
When both input currents I_d and I_p are large, that is, larger than the thresholds θ_d and θ_{p1} , the second term in Equation (1) dominates, which leads to $y \approx 1$. An intermediate activity plateau, of strength α emerges in addition when $I_p > \theta_{p0}$ and $I_d < \theta_d$. As such, the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

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

$$y(t) = \sigma(I_p(t) + I_d(t) - \theta) . \quad (4)$$

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

$$I_d(t) = n_d(t)x_d(t) - b_d(t) , \quad (5)$$



where $n_d(t)$ is a scaling factor, $x_d(t)$ a discrete time sequence, which represents the target signal to be predicted by the proximal input, and $b_d(t)$ a bias. In our experiments, we chose x_d according to the prediction task at hand, see Equations (17), (19), and (20).

Note that n_d and b_d are time dependent since they are subject to adaptation processes, which will be described in the next section. Similarly, the proximal input $I_p(t)$ is given by

$$I_p(t) = n_p(t) \sum_{i=1}^N x_{p,i}(t) w_i(t) - b_p(t), \quad (6)$$

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

2.2. Homeostatic Parameter Regulation

The bias variables entering the definitions (Equations 5, 6) of the distal proximal current, I_d and I_p , are assumed to adapt according to

$$b_p(t+1) = b_p(t) + \mu_b [I_p(t) - I_p^t] \quad (7)$$

$$b_d(t+1) = b_d(t) + \mu_b [I_d(t) - I_d^t], \quad (8)$$

where $I_p^t = 0$ and, $I_d^t = 0$ are preset targets and $1/\mu_b = 10^3$ is the timescale for the adaption. Since this is a slow process, over time, both the distal and the proximal currents, I_d and I_p , will approach a temporal mean equal to I_d^t and I_p^t , respectively, while still allowing the input to fluctuate. The reason for choosing the targets to be zero lies in the fact that we expect a neuron to operate

TABLE 1 | Model parameters, as defined in sections 2.1, 2.3.

θ_{p0}	0	V_d^t	0.25
θ_{p1}	-1	μ_b	10^{-3}
θ_d	0	μ_n	10^{-4}
α	0.3	μ_{av}	$5 \cdot 10^{-3}$
μ_w	$5 \cdot 10^{-5}$	I_p^t	0
ϵ	0.1	I_d^t	0
V_p^t	0.25		

in a dynamical regime that can reliably encode information from its inputs. In the case of our model, this implies that neural input should be distributed close to the threshold (which was set to zero in our case), such that fluctuations in the can have an effect on the resulting neural activity. See e.g., Bell and Sejnowski (1995) and Triesch (2007) for theoretical approaches to optimizing gains and biases based on input and output statistics. Hence, while we chose the mean targets of the input to be the same as the thresholds, this is not a strict condition, as relevant information in the input could also be present in parts of the input statistics that significantly differ from its actual mean (for example in the case of a heavily skewed distribution).

Adaptation rules for the bias entering a transfer function, such as Equations (8), (7), 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 Equations (5), (6), will regulate in contrast the variance of the neural activity, and not the average level (Schubert and Gros, 2021). In this spirit we consider

$$n_d(t+1) = n_d(t) + \mu_n \left[V_d^t - (I_d(t) - \tilde{I}_d(t))^2 \right] \quad (9)$$

$$n_p(t+1) = n_p(t) + \mu_n \left[V_p^t - (I_p(t) - \tilde{I}_p(t))^2 \right] \quad (10)$$

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

$$\tilde{I}_p(t+1) = (1 - \mu_{av})\tilde{I}_p(t) + \mu_{av}I_p(t). \quad (12)$$

Here, V_p^t and V_d^t define targets for the temporally 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).

Mapping the control of the mean input current to the biases and the control of variance to the gains is, in a sense, an idealized case of the more general notion of dual homeostasis. As shown by Cannon and Miller (2017), the conditions for a successful control of mean and variance by means of gains and biases are relatively loose: Under certain stability conditions, a combination of two nonlinear functions of the variable that is to be controlled can yield a dynamic fixed point associated with a certain mean and variance. In fact, a possible variant of dual homeostasis could potentially be achieved by coupling the input gains to a certain firing rate (which is a non-linear function of the input), while biases are still adjusted to a certain mean input. This, of course,

would make it harder to predict the variance of the input resulting from such an adaptation, since it would not enter the equations as a simple parameter that can be chosen a priori (as it is the case for Equations 11, 12).

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 toward 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.

2.3. Synaptic Plasticity

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

$$w_i(t+1) = w_i(t) + \mu_w [(x_{p,i}(t) - \tilde{x}_{p,i}(t)) (y(t) - \tilde{y}) - \epsilon w_i(t)] \quad (13)$$

$$\tilde{x}_{p,i}(t+1) = (1 - \mu_{av})\tilde{x}_{p,i}(t) + \mu_{av}x_{p,i}(t) \quad (14)$$

$$\tilde{y}(t+1) = (1 - \mu_{av})\tilde{y}(t) + \mu_{av}y(t) \quad (15)$$

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 (13) as reference levels. Pre- and post-synaptic neurons are considered to be active/inactive when being above/below the respective trailing averages. This is a realization of the Hebbian rule proposed by Linsker (1986). The timescale of the averaging, $1/\mu_{av}$, is 200 time steps, see **Table 1**. As discussed in section 2.1, a time step can be considered to be on the order of 100 ms, which equates to an averaging time of about 20 s. Generally, this is much faster than the timescales on which metaplasticity, i.e., adaptation processes affecting the dynamics of synaptic plasticity itself, are believed to take place, which are on the order of days (Yger and Gilson, 2015). However, it should be noted that our choice of the timescale of the averaging process used in our plasticity model is motivated mostly by considerations regarding the overall simulation time: Given enough update steps, the same results could be achieved by an arbitrarily slow averaging process.

Since classical Hebbian learning does not keep weights bounded, we use an additional proportional decay term ϵw_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 (Equation 4) is equipped with the same plasticity rule for the proximal weights as Equation (13).

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

$$w_i(t+1) = w_i(t) + \mu_w [y(y - \theta_M) x_i - \epsilon w_i], \quad (16)$$

where θ_M is a threshold defining a transition from long-term potentiation (LTP) to long-term depression (LTD) and, again, ϵ 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 is present at the same time. Obviously, for the point model, the reasoning behind our choice of θ_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 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 $i \in [1, N]$.

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

$$x_d(t) = \sum_{i=1}^N a_i x_{p,i}(t) \quad (17)$$

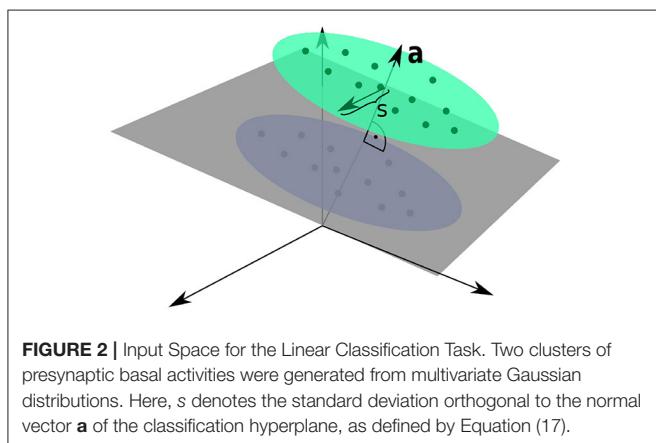
of the $x_{p,i}(t)$, where the a_i are the components of a random vector a of unit length.

Given that we use with Equation (13) 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' 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_{dist} of the distracting subspace within the basal input space. The N_{dist} randomly generated basis vectors are orthogonal to the superposition vector a , as defined by Equation (17), and to each others. Within this N_{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 protocol 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 synaptic weights are adapted via Hebbian plasticity, see Equation (13). The error $(I_p - I_d)^2$ does not enter the adaption rules employed. Results are presented in **Figure 3** as a function of the distraction parameters s and $N_{\text{dist}} \in [0, N - 1]$. The total number of basal inputs is $N = 100$.

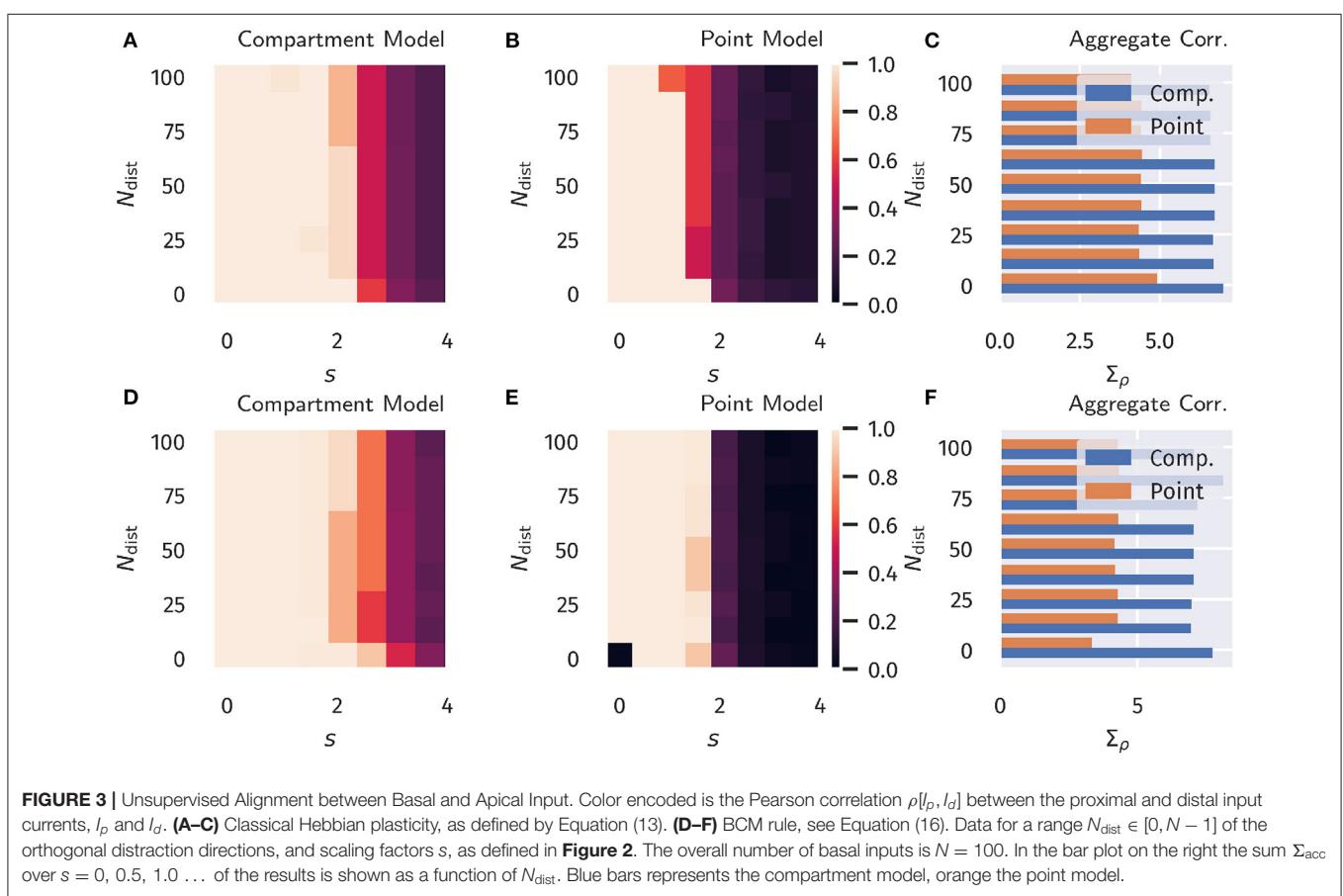


For comparison, in **Figure 3** data for both the compartment model and for a point neuron are presented (as defined, respectively by Equations 1, 4), as well as results for both classical Hebbian and BCM learning rules. A decorrelation transition as a function of the distraction scaling parameter s is observed for both models and plasticity rules. In terms of the learning rules, only marginal differences are present. However, the compartment model is able to handle a significantly stronger distraction as compared to the point model. These findings support the hypothesis examined here, namely that nonlinear interactions between basal and apical input improve learning guided by top-down signals.

3.2. Supervised Learning in a Linear Classification Task

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

$$\mathbf{x}_p(t) = \mathbf{b} + \mathbf{a}[c(t) + \sigma_a \zeta_a(t)] + s \cdot \sum_{i=1}^{N_{\text{dist}}} \zeta_{\text{dist},i}(t) \mathbf{v}_{\text{dist},i}, \quad (18)$$



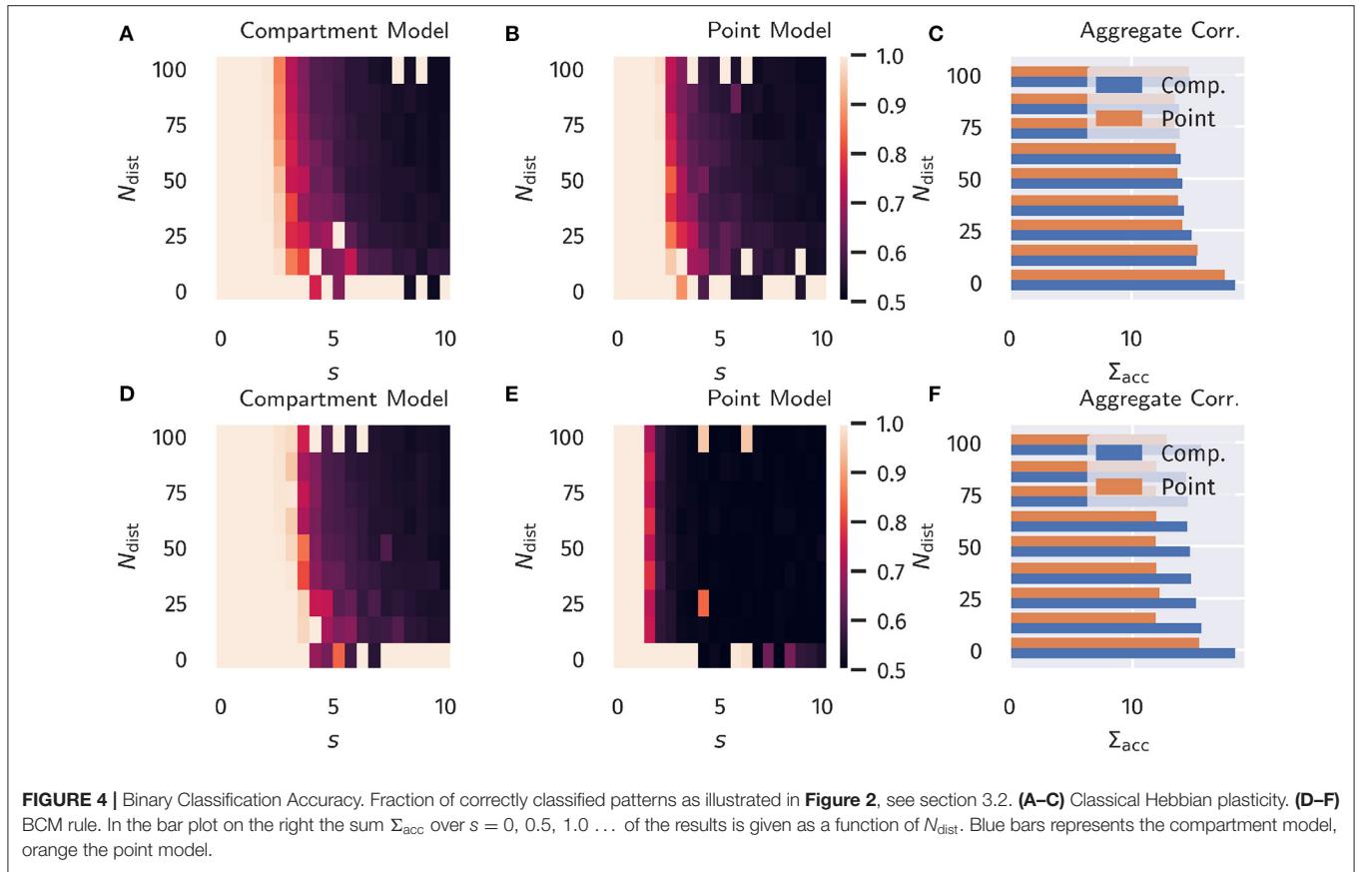


FIGURE 4 | Binary Classification Accuracy. Fraction of correctly classified patterns as illustrated in **Figure 2**, see section 3.2. **(A–C)** Classical Hebbian plasticity. **(D–F)** BCM rule. In the bar plot on the right the sum Σ_{acc} over $s = 0, 0.5, 1.0 \dots$ of the results is given as a function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

where \mathbf{b} is a random vector, where each entry is drawn uniformly from $[0, 1]$, \mathbf{a} is random unit vector as introduced in section 3.1, $c(t)$ is a binary variable drawn from $\{-0.5, 0.5\}$ with equal probability and $\zeta_a(t)$ and the $\zeta_{\text{dist},i}(t)$ are independent Gaussian random variables with zero mean and unit variance. Hence, σ_a simply denotes the standard deviation of each Gaussian cluster along the direction of the normal vector \mathbf{a} and was set to $\sigma_a = 0.25$. Finally, the set of $\mathbf{v}_{\text{dist},i}$ forms a randomly generated orthogonal basis of N_{dist} unit vectors which are—as in section 3.1—also orthogonal to \mathbf{a} . The free parameter s parameterizes the standard deviation along this subspace orthogonal to \mathbf{a} . As indicated by the time dependence, the Gaussian 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 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,

$$x_{d,0}(t) = 1 - \Theta \left((\mathbf{x}_p(t) - \mathbf{b})^T \mathbf{a} \right) \quad (19)$$

$$x_{d,1}(t) = \Theta \left((\mathbf{x}_p(t) - \mathbf{b})^T \mathbf{a} \right), \quad (20)$$

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

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

The resulting accuracy as a function of N_{dist} and s is shown in **Figure 4**, again for all four combinations of neuron models and learning rules.

For classical Hebbian plasticity, the differences between compartmental and point neuron are small. Interestingly, the compartment model performs measurably better in the case of the BCM rule (16), in particular when the overall accuracies for the tested parameter range are compared, see **Figure 4D**. This indicates that during learning, the compartmental neuron makes better use, of the three distinct activity plateaus at 0, α and 1, when the BCM rule is at work. Compare **Figure 1**. We point out in this respect that the sliding threshold θ_M in (16) has been set to the point halfway between the two non-trivial activity levels, α and 1.

It should be noted that the advantage of the compartment model is also reflected in the actual correlation between proximal and distal input as a measure of successful learning (as done in the previous section), see Figure 5 in the Appendix. Interestingly, the discrepancies are more pronounced when measuring the

correlation as compared to the accuracy. Moreover, it appears that above-chance accuracy is still present for parameter values where alignment is almost zero. We attribute this effect to the fact that the classification procedure predicts the class by choosing the node that has the higher activity, independent of the actual “confidence” of this prediction, i.e., how strong activities differ relative to their actual activity levels. Therefore, marginal differences can still yield the correct classification in this isolated setup, but it would be easily disrupted by finite levels of noise or additional external input.

4. DISCUSSION

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

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

4.1. Learning Targets by Maximizing Correlation

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

The proximal input current I_p is a linear projection of the proximal input space. Maximizing the correlation between I_p and I_d (the distal current), can therefore be regarded as a form of canonical correlation analysis (CCA) (Härdle and Simar, 2007). The idea of using CCA as a possible mode of synaptic learning has previously been investigated by Haga and Fukai (2018). Interestingly, according to 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 toward 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.

4.2. Generalizability of the Model to Neuroanatomical Variability

While some research on cortical circuits suggests the possibility of generic and scalable principles that apply to different cortical regions and their functionality (Douglas and Martin, 2007; George and Hawkins, 2009; Larkum, 2013), it is also well-known that the anatomical properties of pyramidal neurons, in particular the dendritic structure, varies significantly across cortical regions (Fuster, 1973; Funahashi et al., 1989). More specifically, going from lower to higher areas of the visual pathway, one can observe a significant increase of spines in the basal dendritic tree (Elston and Rosa, 1997; Elston, 2000), which can be associated with the fact that neurons in higher cortical areas generally encode more complex or even multi-sensory information, requiring the integration of activity from a higher number and potentially more distal neurons (Elston, 2003; Luebke, 2017).

With respect to a varying amount of basal synaptic inputs, it is interesting to note that the dimensionality N of the basal input patterns did not have a large effect on the results of our model, see Figures 3, 4, Figure 5 in the Appendix, as long as the homeostatic processes provided weight normalization.

Apart from variations in the number of spines, variability can also be observed within the dendritic structure itself (Spruston, 2008; Ramaswamy and Markram, 2015). Such differences obviously affect the internal dynamics of the integration of synaptic inputs. Given the phenomenological nature of our neuron model, it is hard to predict how such differences would be reflected, given the diverse dynamical properties that can arise from the dendritic structure (Häusser et al., 2000). The two models tested in our study can be regarded as two extreme cases, where the point neuron represents a completely linear superposition of inputs and the compartment model being strongly nonlinear with respect to proximal and distal inputs. In principle, pyramidal structures could also exhibit properties in between, where the resulting plasticity processes would show a mixture between the classical point neuron behavior (e.g., if a dimensionality reduction of the input via PCA is the main task) and a regime dominated by proximal-distal input correlations if top-down signals should be predicted.

4.3. Outlook

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). Therefore, we expect that our results can be also transferred to deep network structures, for which plasticity is classically guided by local errors between top-down and bottom-up signals.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Bell, A. J., and Sejnowski, T. J. (1995). An Information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129
- Bengio, Y. (2014). How auto-encoders could provide credit assignment in deep networks via target propagation. *CoRR, abs/1407.7*.
- Bi, G. Q., and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472. doi: 10.1523/JNEUROSCI.18-24-10464.1998
- Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* 2, 32–48. doi: 10.1523/JNEUROSCI.02-01-00032.1982
- Branco, T., and Häusser, M. (2011). Synaptic integration gradients in single cortical pyramidal cell dendrites. *Neuron* 69, 885–892. doi: 10.1016/j.neuron.2011.02.006
- Cannon, J., and Miller, P. (2017). Stable control of firing rate mean and variance by dual homeostatic mechanisms. *J. Math. Neurosci.* 7, 1. doi: 10.1186/s13408-017-0043-7
- Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat. Neurosci.* 13, 344–352. doi: 10.1038/nn.2479
- Debanne, D., Gähwiler, B. H., and Thompson, S. M. (1994). Asynchronous pre- and postsynaptic activity induces associative long-term depression in area CA1 of the rat hippocampus *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1148–1152. doi: 10.1073/pnas.91.3.1148
- Douglas, R. J., and Martin, K. A. (2007). Recurrent neuronal circuits in the neocortex. *Curr. Biol.* 17, R496–R500. doi: 10.1016/j.cub.2007.04.024
- Ebner, C., Clopath, C., Jedlicka, P., and Cuntz, H. (2019). Unifying long-term plasticity rules for excitatory synapses by modeling dendrites of cortical pyramidal neurons. *Cell Rep.* 29, 4295.e6–4307.e6. doi: 10.1016/j.celrep.2019.11.068
- Elston, G. N. (2000). Pyramidal cells of the frontal lobe: all the more spinous to think with. *J. Neurosci.* 20:RC95. doi: 10.1523/JNEUROSCI.20-18-j0002.2000
- Elston, G. N. (2003). Cortex, cognition and the cell: new insights into the pyramidal neuron and prefrontal function. *Cereb. Cortex* 13, 1124–1138. doi: 10.1093/cercor/bhg093
- Elston, G. N., and Rosa, M. G. (1997). The occipitoparietal pathway of the macaque monkey: comparison of pyramidal cell morphology in layer III of functionally related cortical visual areas. *Cereb. Cortex* 7, 432–452. doi: 10.1093/cercor/7.5.432
- Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349. doi: 10.1152/jn.1989.61.2.331
- Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *J. Neurophysiol.* 36, 61–78. doi: 10.1152/jn.1973.36.1.61
- George, D., and Hawkins, J. (2009). Towards a Mathematical Theory of Cortical Micro-circuits. *PLoS Comput. Biol.* 5:e1000532. doi: 10.1371/journal.pcbi.1000532
- Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites. *eLife* 6:e22901. doi: 10.7554/eLife.22901.027
- Gustafsson, B., Wigstrom, H., Abraham, W. C., and Huang, Y. Y. (1987). Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J. Neurosci.* 7, 774–780. doi: 10.1523/JNEUROSCI.07-03-00774.1987
- Haga, T., and Fukai, T. (2018). Dendritic processing of spontaneous neuronal sequences for single-trial learning. *Sci. Rep.* 8, 15166. doi: 10.1038/s41598-018-33513-9
- Härdle, W., and Simar, L. (2007). “Canonical correlation analysis,” in *Applied Multivariate Statistical Analysis* (Berlin; Heidelberg: Springer Berlin Heidelberg), 321–330.
- Häusser, M., Spruston, N., and Stuart, G. J. (2000). Diversity and dynamics of dendritic signaling. *Science* 290, 739–744. doi: 10.1126/science.290.5492.739
- Hay, E., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2011). Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Comput. Biol.* 7:e1002107. doi: 10.1371/journal.pcbi.1002107
- Intrator, N., and Cooper, L. N. (1992). Objective function formulation of the BCM theory of visual cortical plasticity: statistical connections, stability conditions. *Neural Netw.* 5, 3–17. doi: 10.1016/S0893-6080(05)80003-6
- Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Larkum, M. E., Kaiser, K. M. M., and Sakmann, B. (1999). Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14600–14604. doi: 10.1073/pnas.96.25.14600
- Larkum, M. E., Nevanian, T., Sandier, M., Polksy, A., and Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* 325, 756–760. doi: 10.1126/science.1171958
- Law, C. C., and Cooper, L. N. (1994). Formation of receptive fields in realistic visual environments according to the Bienenstock, Cooper, and Munro (BCM) theory. *Proc. Natl. Acad. Sci. U.S.A.* 91, 7797–7801. doi: 10.1073/pnas.91.16.7797
- Lee, D. H., Zhang, S., Fischer, A., and Bengio, Y. (2015). “Difference target propagation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9284 (Verlag: Springer), 498–515.
- Letzkus, J. J., Kampa, B. M., and Stuart, G. J. (2006). Learning rules for spike timing-dependent plasticity depend on dendritic synapse location. *J. Neurosci.* 26, 10420–10429. doi: 10.1523/JNEUROSCI.2650-06.2006
- Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nat. Commun.* 7, 1–10. doi: 10.1038/ncomms13276
- Linsker, R. (1986). From basic network principles to neural architecture: emergence of orientation-selective cells. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8390. doi: 10.1073/pnas.83.21.8390
- Luebke, J. I. (2017). Pyramidal Neurons Are Not Generalizable Building Blocks of Cortical Networks. *Front. Neuroanat.* 11, 11. doi: 10.3389/fnana.2017.00011
- Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic

AUTHOR CONTRIBUTIONS

FS and CG contributed equally to the writing and review of the manuscript. FS provided the code, ran the simulations, and prepared the figures. Both authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge the financial support of the German Research Foundation (DFG).

- 913 APs and EPSPs. *Science* 275, 213–215. doi: 10.1126/science.275.52
914 97.213
- 915 Park, J., Papoutsis, A., Ash, R. T., Marin, M. A., Poirazi, P., and Smirnakis,
916 S. M. (2019). Contribution of apical and basal dendrites to orientation
917 encoding in mouse V1 L2/3 pyramidal neurons. *Nat. Commun.* 10, 1–11.
918 doi: 10.1038/s41467-019-13029-0
- 919 Poirazi, P. (2009). Information processing in single cells and small networks:
920 insights from compartmental models. *AIP Conf. Proc.* 1108, 158–167.
921 doi: 10.1063/1.3117124
- 922 Ramaswamy, S., and Markram, H. (2015). Anatomy and physiology of
923 the thick-tufted layer 5 pyramidal neuron. *Front. Cell Neurosci.* 9:233.
924 doi: 10.3389/fncel.2015.00233
- 925 Schiess, M., Urbanczik, R., and Senn, W. (2016). Somato-dendritic synaptic
926 plasticity and error-backpropagation in active dendrites. *PLoS Comput. Biol.*
927 12:1004638. doi: 10.1371/journal.pcbi.1004638
- 928 Schubert, F., and Gros, C. (2021). Local homeostatic regulation of the
929 spectral radius of echo-state networks. *Front. Comput. Neurosci.* 15:12.
930 doi: 10.3389/fncom.2021.587721
- 931 Shai, A. S., Anastassiou, C. A., Larkum, M. E., and Koch, C. (2015).
932 Physiology of layer 5 pyramidal neurons in mouse primary visual cortex:
933 coincidence detection through bursting. *PLoS Comput. Biol.* 11:e1004090.
934 doi: 10.1371/journal.pcbi.1004090
- 935 Sjöström, P. J., and Häusser, M. (2006). A cooperative switch determines the sign of
936 synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron*
937 51, 227–238. doi: 10.1016/j.neuron.2006.06.017
- 938 Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic
939 integration. *Nat. Rev. Neurosci.* 9, 206–221. doi: 10.1038/nrn2286
- 940
- 941
- 942
- 943
- 944
- 945
- 946
- 947
- 948
- 949
- 950
- 951
- 952
- 953
- 954
- 955
- 956
- 957
- 958
- 959
- 960
- 961
- 962
- 963
- 964
- 965
- 966
- 967
- 968
- 969
- 970 Spruston, N., Schiller, Y., Stuart, G., and Sakmann, B. (1995). Activity-
971 dependent action potential invasion and calcium influx into hippocampal
972 CA1 dendrites. *Science* 268, 297–300. doi: 10.1126/science.77
973 16524
- 974 Stuart, G. J., and Häusser, M. (2001). Dendritic coincidence detection of EPSPs and
975 action potentials. *Nat. Neurosci.* 4, 63–71. doi: 10.1038/82910
- 976 Triesch, J. (2007). Synergies between intrinsic and synaptic plasticity mechanisms.
977 *Neural Comput.* 19, 885–909. doi: 10.1162/neco.2007.19.4.885
- 978 Urbanczik, R., and Senn, W. (2014). Learning by the dendritic prediction of
979 somatic spiking. *Neuron* 81, 521–528. doi: 10.1016/j.neuron.2013.11.030
- 980 Weissenberger, F., Gauy, M. M., Lengler, J., Meier, F., and Steger, A.
981 (2018). Voltage dependence of synaptic plasticity is essential for rate based
982 learning with short stimuli. *Sci. Rep.* 8, 4609. doi: 10.1038/s41598-018-
983 22781-0
- 984 Yger, P., and Gilson, M. (2015). Models of metaplasticity: a review of concepts.
985 *Front. Comput. Neurosci.* 9:138. doi: 10.3389/fncom.2015.00138
- 986
- 987
- 988
- 989
- 990
- 991
- 992
- 993
- 994
- 995
- 996
- 997
- 998
- 999
- 1000
- 1001
- 1002
- 1003
- 1004
- 1005
- 1006
- 1007
- 1008
- 1009
- 1010
- 1011
- 1012
- 1013
- 1014
- 1015
- 1016
- 1017
- 1018
- 1019
- 1020
- 1021
- 1022
- 1023
- 1024
- 1025
- 1026