RESEARCH ARTICLE

Somato-dendritic Synaptic Plasticity and Error-backpropagation in Active Dendrites

Mathieu Schiess[1]*, **Robert Urbanczik**[1]*, **Walter Senn**[1,2]*

1 Department of Physiology, University of Bern, Bern, Switzerland, **2** Center for Cognition, Learning and Memory, University of Bern, Bern, Switzerland

* mathieu.schiess@hesge.ch (MS); urbanczik@pyl.unibe.ch (RU); senn@pyl.unibe.ch (WS)

Citation: Schiess M, Urbanczik R, Senn W (2016) Somato-dendritic Synaptic Plasticity and Error-backpropagation in Active Dendrites. PLoS Comput Biol 12(2): e1004638. doi:10.1371/journal.pcbi.1004638

Editor: Boris S. Gutkin, École Normale Supérieure, College de France, CNRS, FRANCE

Received: January 27, 2015

Accepted: October 26, 2015

Published: February 3, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The work was supported by the Swiss National Science Foundation (personal grant of WS, 31003A_133094 and 310030L_156863) and the Human Brain Project funded by the European Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In the last decade dendrites of cortical neurons have been shown to nonlinearly combine synaptic inputs by evoking local dendritic spikes. It has been suggested that these nonlinearities raise the computational power of a single neuron, making it comparable to a 2-layer network of point neurons. But how these nonlinearities can be incorporated into the synaptic plasticity to optimally support learning remains unclear. We present a theoretically derived synaptic plasticity rule for supervised and reinforcement learning that depends on the timing of the presynaptic, the dendritic and the postsynaptic spikes. For supervised learning, the rule can be seen as a biological version of the classical error-backpropagation algorithm applied to the dendritic case. When modulated by a delayed reward signal, the same plasticity is shown to maximize the expected reward in reinforcement learning for various coding scenarios. Our framework makes specific experimental predictions and highlights the unique advantage of active dendrites for implementing powerful synaptic plasticity rules that have access to downstream information via backpropagation of action potentials.

Author Summary

Error-backpropagation is a successful algorithm for supervised learning in neural networks. Whether and how this technical algorithm is implemented in cortical structures, however, remains elusive. Here we show that this algorithm may be implemented within a single neuron equipped with nonlinear dendritic processing. An error expressed as mismatch between somatic firing and membrane potential may be backpropagated to the active dendritic branches where it modulates synaptic plasticity. This changes the classical view that learning in the brain is realized by rewiring simple processing units as formalized by the neural network theory. Instead, these processing units can themselves learn to implement much more complex input-output functions as previously thought. While the original algorithm only considered firing rates, the biological implementation enables learning for both a firing rate and a spike-timing code. Moreover, when modulated by a reward signal, the synaptic plasticity rule maximizes the expected reward in a reinforcement learning framework.

Introduction

One of the fascinating and still enigmatic aspects of cortical organization is the widespread dendritic arborization of neurons. These dendrites have been shown to generate dendritic spikes [1–3] that support local dendritic processing [4–7], but the nature of this computation remains elusive. An interesting view is that the dendritic nonlinearities endow the neuron with the structure of a 2-layer neural network of point neurons, in particular if the dendrites show themselves step-like dendritic spikes, but also if the dendritic nonlinearities remain continuous [8–11]. Here we show that the dendritic morphology actually offers a substantial additional benefit over the 2-layer network. This is because it allows for the implementation of powerful learning algorithms that rely on the backpropagation of the somatic information along the dendrite that, in a network of point neurons, would not be possible in this form.

Error-backpropagation has become the classical algorithm for adapting the connection strengths in artificial neural networks [12, 13]. In this algorithm, an error at an output unit is assessed by comparing the self-generated activity with a target activity. Plasticity in hidden units is driven by this error that propagates backwards along the connections of the network. Synapses, however, transmit information just in one direction, making it difficult to implement error-backpropagation in biological neuronal circuitries. But this is different for dendritic trees. In the 2-layer structure of a dendritic tree information at the output site may be physically backpropagated across the intermediate computational layer to the synapses targeting the tree.

While the suggested dendritic error-backpropagation is a plasticity rule for supervised learning, it is also suitable for reinforcement learning. Instead of imposing the somatic spiking to learn pre-assigned target spike timings, the somatic spikes can be generated by the dendritic inputs alone, while learning is driven by a delayed reward signal. The synapse itself can be agnostic about the coding and the learning scenario; it learns by continuously adapting synaptic strength according to molecular mechanisms that are identical in the different scenarios.

Various experimental work revealed that synaptic plasticity depends on the precise timing between pre- and postsynaptic action potentials [14, 15] and the postsynaptic voltage [16]. It has further been shown that the specific form of this spike-timing-dependent plasticity (STDP) may vary with the synaptic location on the dendritic tree [17–19], and that synaptic plasticity in general is modulated by dendritic spikes [20–22]. Yet, no coherent view on the impact of dendritic nonlinearities on plasticity has emerged. Correspondingly, beside an early attempt to assign a fitness score to dendritic synapses [23] and the suggestion of a Hebbian-type plasticity rule for synapses on active dendrites [24], no computational framework for synaptic plasticity with regenerative dendritic events exists that would guide its experimental exploration. In our previous study, we derived a reward-maximizing plasticity rule that incorporates dendritic spikes, but no online implementation was presented [25]. Here, starting from biophysical properties of NMDA conductances [26], we consider an integrated somato-dendritic spiking model that captures the main biological ingredients of dendritic spikes and that is simple enough to derive an online plasticity rule for different coding schemes in the context of both supervised and reinforcement learning.

Results

Neuron model

We model a multi-compartment neuron with several active dendritic branches, each directly linked to a somatic compartment (Fig 1A1). The subthreshold dendritic voltage in branch *d* is the weighted sum of normalized postsynaptic potentials (PSPs) triggered by the presynaptic

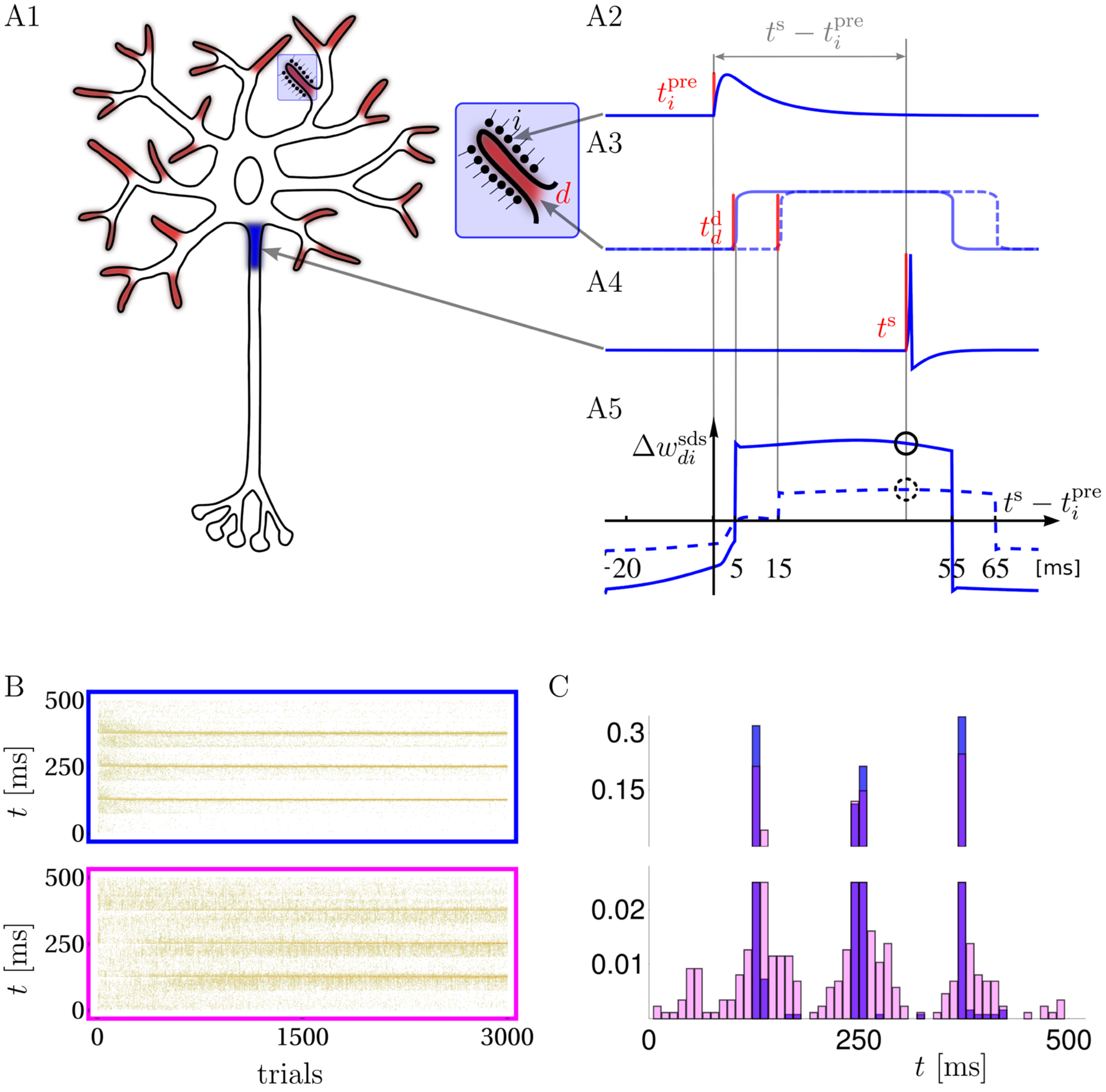

Fig 1. Neuron model, synaptic plasticity rule and learning of spike timings. A: Synaptic inputs targeting dendritic NMDA activation zones (A1, red endings with enlargement) propagate, together with possible NMDA-spikes, to the somatic spike trigger zone (A1, blue). Individual postsynaptic potentials in a dendritic branch (PSPs, arriving e.g. at time t_i^{pre}, A2), may trigger NMDA-spikes, e.g. at time $t_d^{d} = 5\,ms$ (solid) or 15 ms (dashed) after t_i^{pre}, forming a local dendritic plateau potential of 50 ms duration (A3). A somatic spike triggered at t^{s} during the ongoing NMDA-spike (A4) causes a synaptic weight change Δw_{di}^{sds} that is large/small depending on whether the NMDA-spike was triggered 5/15 ms after the presynaptic spike (A5, solid/dashed circle, respectively). A5: Δw_{di}^{sds} as a function of $t^{s} - t_i^{pre}$ for a NMDA-spike at 5 (solid) and 15 ms (dashed). **B**: Raster plots of freely generated somatic spikes from test trials that are interleaved with learning trials. For the full somato-dendritic synaptic plasticity rule (sdSP) the somatic spikes converge to the 3 target times with a precision of ±3 ms (top), while the rule neglecting the dendritic spikes (i.e. suppressing the term $\dot{w}_{di}^{sds}$) achieves a precision of only ±14 ms (bottom). **C**: The two spike distributions from C after 3000 presentations.

doi:10.1371/journal.pcbi.1004638.g001

spikes in the afferents projecting to that branch, $u_d^{d}(t) = \sum_i w_{di}\,\mathrm{PSP}_i(t)$. Here, w_{di} represents the synaptic strength of the synapse from afferent i onto branch d that scales the PSP amplitude. The dendritic branches can generate temporally extended NMDA-spikes of a fixed amplitude, similar to experimental observations *in vitro* [1, 3, 5] and *in vivo* [7]. In our model an NMDA-spike is represented by a square voltage pulse of amplitude a and duration Δ = 50 ms (Fig 1A2–1A3). It is stochastically elicited with a rate that is an increasing function of the local subthreshold membrane potential u_d^{d} and, implicitly, of the local glutamate level. In fact, in an *in vivo* scenario the joint voltage and glutamate condition for triggering an NMDA-spike effectively reduces to a single condition on the local voltage alone. This is because the depolarization required to activate the NMDA receptors is only reached when enough glutamate was released, making the glutamate condition automatically satisfied at high enough voltages (see S1 Text).

The subthreshold dendritic voltage $u_d^{\mathrm{d}}(t)$ and the dendritic spike train $\mathrm{NMDA}_d(t)$ in branch d propagate with some attenuation factor α to the soma where they add up with inputs from other branches to form the somatic voltage $u^{\mathrm{s}} = \sum_d \alpha(u_d^{\mathrm{d}} + \mathrm{NMDA}_d) - \kappa$. This voltage is also modulated by a spike reset kernel $\kappa(t)$ incorporating the transient hyperpolarisation caused by each somatic spike (Fig 1A4). For supervised learning, the somatic spikes $S(t)$ are imposed by an external input, whereas in reinforcement learning they are stochastically triggered with an instantaneous rate $\rho^{\mathrm{s}}(t)$ that is an increasing function of the somatic potential u^{s} (Online Methods).

Learning rule

We first consider a supervised learning scenario where somatic spikes S are enforced by one modality (e.g. a visual stimulus) while the synaptic inputs to the dendritic branches are caused by another modality (e.g. representing an auditory stimulus [27]). The strengths of the synapses on the dendrites, w_{di}, are adapted in order to reproduce the somatic spike train $S(t)$ from just the dendritic input alone, without direct somatic drive. This can be achieved by ongoing synaptic weight changes, $\dot{w}_{di}$, that together maximize the likelihood of observing S in response to this dendritic input. According to the two types of contributions to the somatic voltage, the sub- and supra-threshold dendritic voltages, u_d^{d} and NMDA_d, the synaptic weight change can also be decomposed into a sub- and suprathreshold contribution, $\dot{w}_{di} = \dot{w}_{di}^{\mathrm{ss}} + \dot{w}_{di}^{\mathrm{sds}}$, that take into account the subthreshold somato-synaptic (ss) and the suprathreshold somato-dendro-synaptic (sds) drive. We also refer to $\dot{w}_{di}$ as somato-dendritic synaptic plasticity (sdSP).

The somato-synaptic contribution is proportional to the postsynaptic error term $(S - \rho^{\mathrm{s}})$ times the local postsynaptic potential PSP_i induced by synapse i on that dendritic branch,

$$\dot{w}_{di}^{\mathrm{ss}} \propto (S - \rho^{\mathrm{s}}) \cdot \mathrm{PSP}_i\,. \tag{1}$$

This corresponds to the gradient learning rule that was previously derived for a single compartment neuron [28] and that was shown to be consistent with the experimentally observed STDP (see e.g. [29]). The error term in the rule ensures that if the rate ρ^{s} is too small for generating S, the weight is increased, and if the rate is too high, the weight is decreased, eventually leading in average to $\langle S \rangle = \rho^{\mathrm{s}}$.

The main sdSP-effect stems from the somato-dendro-synaptic contribution $\dot{w}_{di}^{\mathrm{sds}}$. The instantaneous synaptic weight change at time t is induced by the dendritic activity Den_d in branch d during the interval Δ prior to t. Any NMDA-spike elicited in this interval will affect the somatic voltage at time t, and the likelihood of a dendritic spike is itself influenced by the local synaptic potentials PSP_i arriving in this interval and a few milliseconds before (Fig 1A2–1A5). Overall, we obtain an expression of the form

$$\dot{w}_{di}^{\mathrm{sds}} \propto (S - \rho_{\setminus d}^{\mathrm{s}}) \cdot \mathrm{Den}_d * \mathrm{PSP}_i\,, \tag{2}$$

where $\mathrm{Den}_d * \mathrm{PSP}_i$ captures the impact of synapse i on the triggering of an NMDA-spike in the preceding interval Δ, and $\rho_{\setminus d}^{\mathrm{s}}$ represents the instantaneous somatic firing rate in the absence of a dendritic spike in branch d (see Online Methods). A positive error term $(S - \rho_{\setminus d}^{\mathrm{s}})$ tells the synapses on branch d how worth it is to increase their weights in order to trigger a local

NMDA-spike; a negative error term suggests to rather decrease the weights since even without NMDA-spike from that branch the somatic firing rate, in average, is too high. When only dendritic nonlinearities without spiking are present, the rule Eq (2) simplifies to a pure 3-factor rule composed of a somatic difference factor, a dendritic factor, and a presynaptic factor that can be applied to dendrites showing supra- or sublinear dendritic summations (see Eq (9) in Online Methods).

The learning rule of Eq (2) can be interpreted as error-backpropagation for spiking neurons where a somatic error signal is propagated back to the dendrites that represent the nonlinear hidden units. These hidden units further modulate the error signal depending on their impact on the output unit. Classical error-backpropagation would also adapt the weights from the hidden units to the output unit. This would correspond to adapting the impact of NMDA spikes on the somatic voltage and could be modeled as dendritic branch strength plasticity [24, 30]. For conceptual clarity we discard from this extension, but the gradient calculations could also be applied to infer an optimal learning rule for these branch strengths.

Supervised learning with active dendrites

The overall synaptic modification, Δw_{di}, induced by sdSP is obtained by integrating the instantaneous changes $\dot{w}_{di}$ over the stimulus duration, $\Delta w_{di} = \int_0^T \dot{w}_{di}(t)\,dt$. Using the decomposition $\dot{w}_{di} = \dot{w}^{\text{ss}}_{di} + \dot{w}^{\text{sds}}_{di}$ we may also write $\Delta w_{di} = \Delta w^{\text{ss}}_{di} + \Delta w^{\text{sds}}_{di}$ and have a closer look to the somato-dendro-synaptic contribution $\Delta w^{\text{sds}}_{di}$ (Fig 1A2–1A5). We fixed a presynaptic spike at time $t_i^{\text{pre}} = 0$ and plotted $\Delta w^{\text{sds}}_{di}$ as a function of the somatic spike time t^s for the case of a NMDA-spike at $t^d = 5$ ms and 15 ms. The dendritic spike immediately after a presynaptic spike considerably extends the classical time window for causal 'pre-post' potentiation to a 'pre-dend-post' potentiation. In fact, a presynaptic spike that was taking part in triggering a NMDA-spike may indirectly contribute also to a postsynaptic spike more than 50 ms later. In turn, synaptic depression is induced in an a-causal configuration where the somatic spike comes either before the presynaptic spike or after the NMDA-spike has already decayed.

Endowed with sdSP a neuron is able to learn precise output spike-timings as shown in Fig 1B and 1C; blue) where 3 somatic spike times were imposed during the learning. The dendritic input consisted of 100 frozen presynaptic Poisson spike trains with frequency 6Hz and duration $T = 500$ ms. The dendritic tree had 20 branches, each being targeted by a random subset of the 100 afferents with a connection probability of 0.5. After repeated pattern presentations with somatic output clamped to the target spikes, the neuron learned to generate the target output from the synaptic input alone with a precision of a few milliseconds. The high spike-time precision is lost when synapses are modified only by the somato-synaptic contribution $\dot{w}^{\text{ss}}_{di}$ (Fig 1B and 1C; pink). Because this plasticity contribution is blind to dendritic activity, small synaptic weight changes may cause undesired appearance or disappearance of NMDA-spikes. In this case dendritic spikes arise as unpredicted knock-on effects of synaptic plasticity. Note that the somato-synaptic contribution alone, being identical to the gradient rule [28], would be able to learn the precise spiking (as would also the rules in [29, 31]) if the neuron were note endowed with the dendritic spiking mechanism and instead would only show linear summation with passive voltage propagation.

Reward-modulated somato-dendritic plasticity

We next considered a reinforcement learning scenario where synaptic modifications are modulated by a binary feedback signal $R = \pm 1$ that is applied at the end of the stimulus presentation and that assesses the appropriateness of the somatic firing pattern. While this feedback is itself an external quantity, it is assumed to induce an internal signal, e.g. in the form of a neuromodulator, that globally modulates the previously induced synaptic changes. To control the balance between reward and punishment, the internal feedback modulates the past plasticity induction by a factor $(R - R_\circ)$ with a constant reward bias $R_\circ$.

When deriving a plasticity rule that maximizes the expected reward we again obtain the same sdSP (see Eqs (1) and (2)), but now integrated across the stimulus interval and then modulated by the feedback signal,

$$\Delta w_{di} \propto (R - R_\circ) \int_0^T (\dot{w}_{di}^{\mathrm{ss}} + \dot{w}_{di}^{\mathrm{sds}})\, dt\,, \tag{3}$$

see S1C Text. We refer to this rule as reward-modulated somato-dendritic synaptic plasticity (R-sdSP). Due to the term $\dot{w}_{di}^{\mathrm{sds}}$ it is effectively a 4-factor rule of the form '$\Delta w = Rwrd \cdot som \cdot dend \cdot pre$'. The intuition is that the intrinsic neuronal stochasticities generate fluctuations in the somatic spiking that deviate from the prediction made by the dendritic input and cause an 'error' expressed in the somatic factor $(S - \rho^{\mathrm{s}}_{\setminus d})$ of the rule. These fluctuations will be reinforced or suppressed by the feedback signal. As before, the synaptic modification will be strengthened if a presynaptic spike contributed to a dendritic NMDA-spike that in turn affects the somatic voltage.

Reinforcement learning with active dendrites

We tested R-sdSP for various coding schemes. First, we considered a standard binary classification of frozen Poisson spike patterns by a postsynaptic spike- / no-spike code (Fig 2A). Each input pattern is defined as above (6Hz in 100 afferents for 500 ms) and belongs to one of two classes. For one class the soma is required to fire at least one spike while for the other class it should be silent. After repeated presentations followed by a reward signal, R-sdSP perfectly learned the correct classification of 4 random patterns. In contrast, reward-modulated STDP (R-STDP, [32]) implemented in its best performing version (see Online Methods and [33]) did not (Fig 2B). To achieve an appropriate alignment of dendritic spikes (Fig 2C and 2D), any successful learning rule needs to take account of the causal chain linking presynaptic spikes to dendritic and somatic spikes, the latter deciding upon reward or punishment. R-sdSP derived from maximizing the expected reward captures this causal relationship, but R-STDP does not, neither with a 10 ms (Fig 2B) nor with a 50 ms learning window (S2 Fig), and hence fails. Interestingly, R-STDP improves when the NMDA-spike generation is suppressed (Fig 2B, dashed). This shows that the increased flexibility in neuronal information processing provided by dendritic nonlinearities will in fact impede learning when a rule is used that does not take the nonlinearities into account.

R-sdSP is still able to correctly learn the classification even when the spike timings were noisy with a jitter up to 100 ms, or when the somatic voltage modulating the synaptic plasticity (via ρ^{s} and $\rho^{\mathrm{s}}_{\setminus d}$) was low-pass filtered to mimic the dilution of information back-propagating to the synaptic site (S2 Fig).

Incidentally, the same task from Fig 2 can also be solved in a supervised scenario e.g. with the tempotron where, beside telling a neuron whether it should spike or not spike in response to a stimulus, the neuron is supposed to have access to the time of the voltage maximum within the stimulus interval [0, T], see e.g. [34, 35]. Although with these additional assumptions learning in principle becomes faster, the rules will again suffer from the ignorance about NMDA spikes and the possible acausality between a presynaptic spike and an immediately following somatic spike.

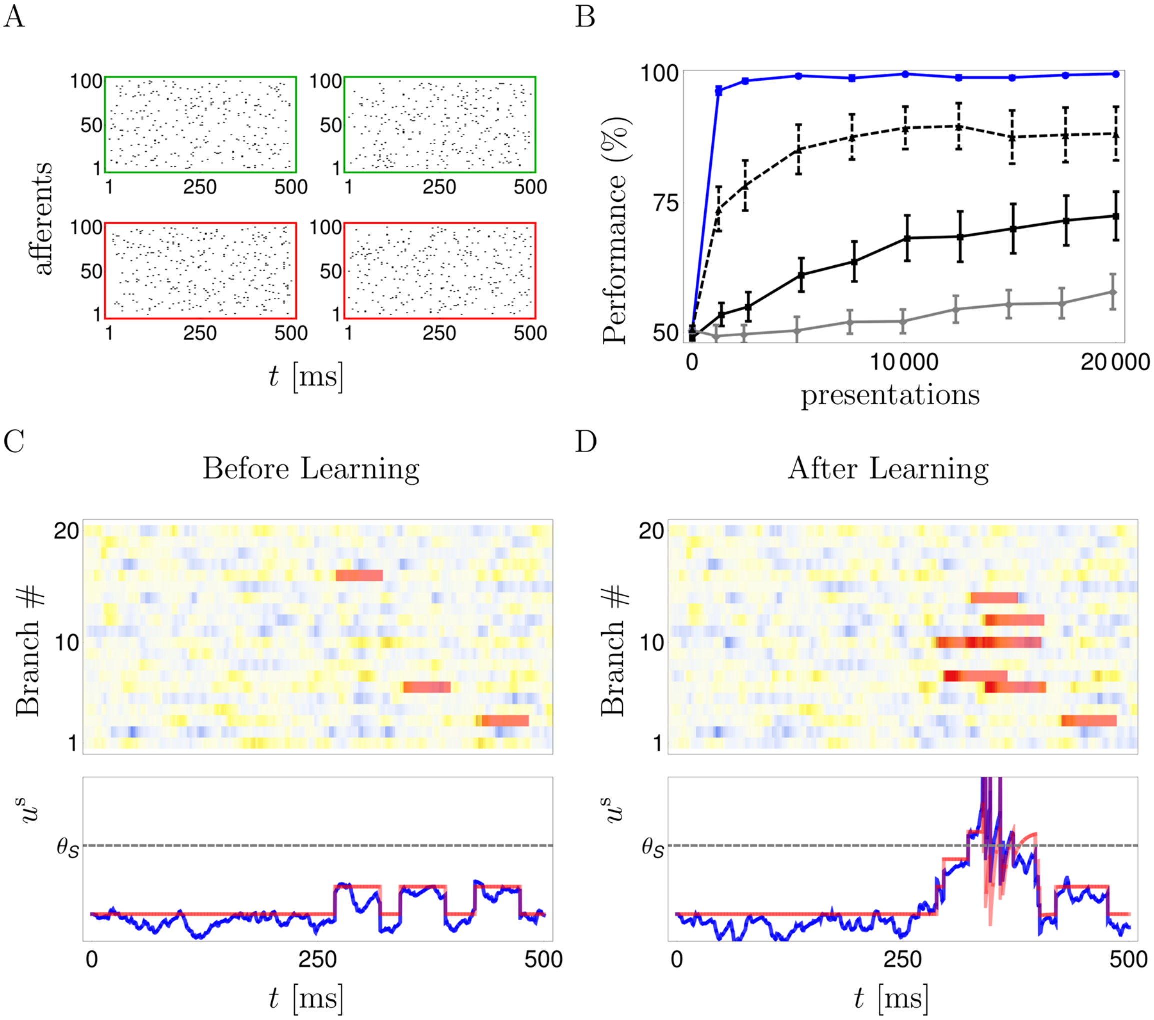

Fig 2. Binary classification of frozen input spike patterns by a somatic spike / no-spike code for the reward-modulated somato-dendritic synaptic plasticity (R-sdSP). A: Four input patterns, the two patterns in the top row should elicit no somatic spikes; the patterns in the bottom row should. **B**: R-sdSP perfectly learns the classification after roughly 1000 presentations (blue solid). In contrast, classical R-STDP fails when applied to the presynaptic–somatic ('pre-som', solid black) or the presynaptic–dendritic ('pre-den', gray) spike pairs. R-STDP improves when the dendritic spike generation is suppressed (black dashed), although it does not reach the high performance of R-sdSP. **C,D**: Dendritic and somatic voltages in response to an input pattern that requires spiking, before (**C**) and after (**D**) learning. The initially sparse dendritic spikes (NMDA$_d(t)$, red bars, overlaid on a $u_d^{\mathrm{d}}(t)$ intensity plot) become more numerous, co-align and sum up in the soma to enable the somatic firing. Yellow indicates depolarization. Bottom: Time course of the somatic voltage $u^{\mathrm{s}}(t)$ (blue) with the contribution of the NMDA-spikes and the somatic spike reset kernel (red).

doi:10.1371/journal.pcbi.1004638.g002

Learning direction selectivity and nonlinear separation

To apply the dendritic learning to a biological example we consider the direction selectivity of pyramidal neurons that was found to be mediated by nonlinear dendritic processing *in vitro* [5] and *in vivo* [7]. To mimick directional inputs moving in the stimulus space from right to left and left to right, we randomly enumerated the synapses across the whole dendritic tree and stochastically activated these synapses once in increasing and once in decreasing order (Fig 3A). After the stimulation, a positive reward signal was applied to the synapses when at least one somatic spike was elicited during the left-to-right patterns, or no somatic spike was elicited

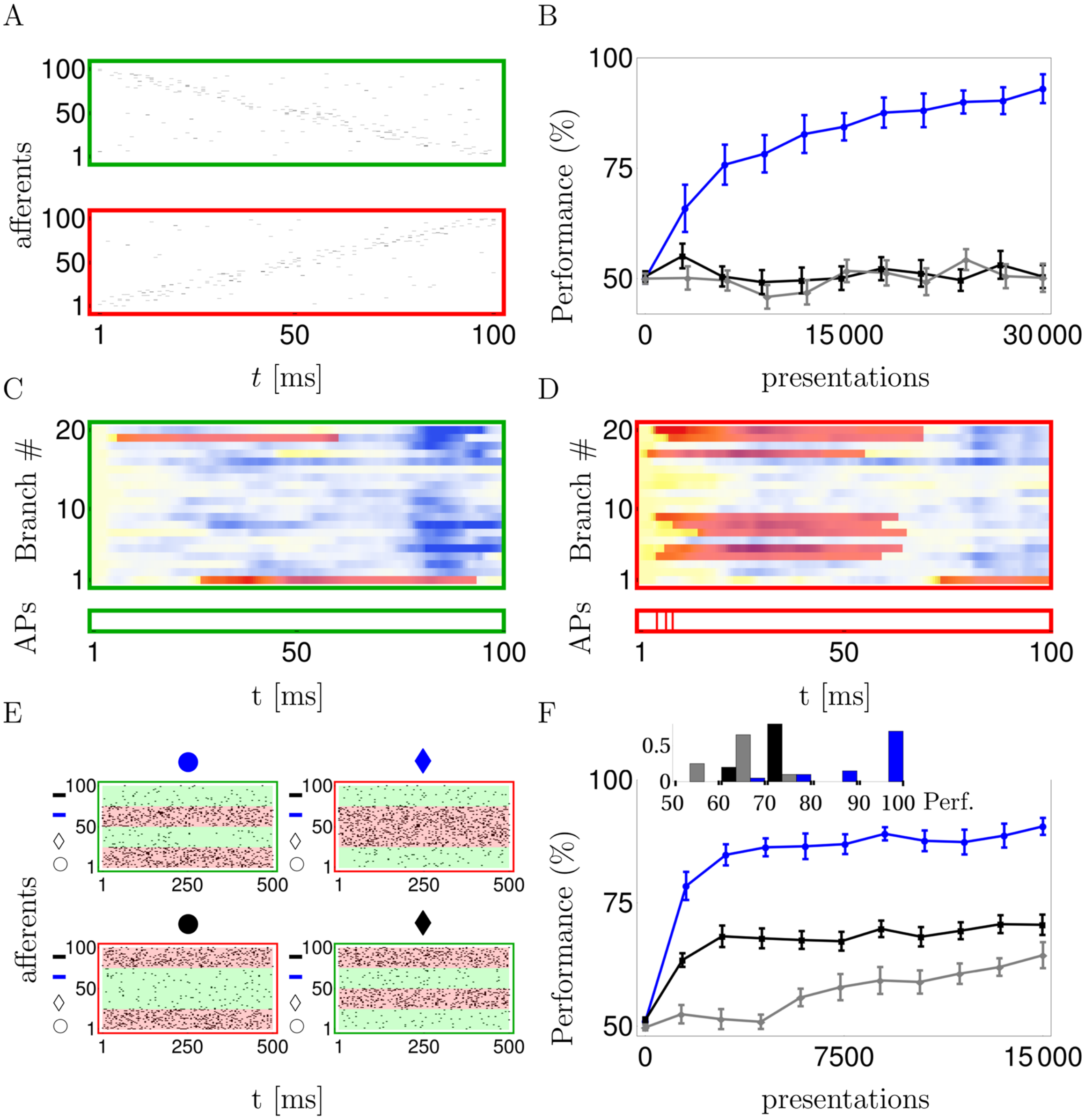

Fig 3. R-sdSP can exploit the representational power endowed by active dendrites. A: Example of presynaptic firing pattern that requires the neuron to be silent (green) or to elicit at least one somatic spike (red). **B**: R-sdSP (blue), but not R-STDP, learns to become direction selective (black: 'pre-som'; grey: 'pre-den'). **C, D**: The subthreshold dendritic voltages $u_d^{\mathrm{d}}(t)$ and NMDA traces NMDA$_d(t)$ in response to the two input patterns shown in A (color code as in Fig 2). Individual branches developed direction selectivity (green). Bottom: action potentials are only generated for one direction. **E**: The 4 input patterns of the linearly non-separable feature-binding problem combine one of two shapes ('circle' or 'diamond') with one of two fill colors ('blue' or 'black'). Each of the four features is represented by 25 afferents (next to the corresponding symbol on the y-axes) that encode its presence or absence by a high (40Hz) or low (5Hz) Poisson firing rate, respectively. **F**: R-sdSP learns the correct response to the 4 inputs, R-STDP does not (line code as above). Inset: average performance of each run after learning.

doi:10.1371/journal.pcbi.1004638.g003

during a right-to-left pattern. A negative reward signal was applied in the other cases. R-sdSP, but not R-STDP, could learn such direction selectivity (Fig 3B). Individual dendritic branches may become selective to the synaptic activation order and learn to generate NMDA-spikes that, after summation in the soma, eventually trigger somatic action potentials (Fig 3C and 3D). Hence, the neuron learned to employ the dendritic nonlinearities to achieve direction selectivity, even though solving the task does not require them.

A classical task that exceeds the representational power of a point neuron is the XOR (exclusive-or) problem that is equivalent to the linearly non-separating feature binding problem [24]. In this task, the neuron has to respond exclusively to two disjoint pairs of features (e.g. to black & circle and to blue & diamond), but not to the cross combinations of these features (black & diamond and blue & circle). The presence and absence of a feature was encoded in a high and low Poisson firing rate, respectively, of a subpopulation of afferents projecting to our classifying neuron (Fig 3E). R-sdSP on the active dendrites could learn the correct responses, although due to the intrinsic stochasticity failures occurred in some cases. Classical R-STDP failed also in solving the feature binding problem problem on the dendritic tree, whether applied to *pre-dend* or to *pre-som* spike pairings (Fig 3F).

Learning spike timings with delayed reward

Besides learning a spike / no-spike code or a firing rate code, R-sdSP can also learn a spike timing code, i.e. to fire only at specific times. In a first task showing this, the neuron had to learn to spike at a target time $t^{\mathrm{targ}} = 250$ ms in response to a frozen Poisson spike pattern. Deviations from this time were punished at the stimulus ending, using a graded feedback signal that increases with the magnitude of the deviation (Online Methods). During repeated pattern presentations, while applying R-sdSP and the delayed punishing signal, the postsynaptic spiking becomes concentrated in a narrow time window around the target spike (Fig 4A–4C). To understand the role of the active dendrites we separated the time course of the somatic voltage into the contribution from the subthreshold dendritic potentials and the NMDA-spikes (Fig 4D and 4E). After successful learning, the averaged NMDA-spikes form a broad ridge around t^{targ} on top of which the subthreshold dendritic voltages act as 'scorers'. Before and immediately after t^{targ} the subthreshold voltage is hyperpolarized to prevent somatic spikes from coming too early or too late. Notice that in an individual run the summed NMDA-spikes can form plateaus that are much shorter than the NMDA-spike duration of 50 ms. In the example shown, this arises because just 5 ms after the initiation of an NMDA-spike in one branch another NMDA-spike ends in a second branch, cutting the somatic plateau short to 5 ms (Fig 4D). By virtue of the backpropagated somatic activity, R-sdSP learns to coordinate the timing of the NMDA-spikes in the different branches, creating a narrow window for a somatic spike around the target time.

Learning a spike-timing code is also possible if the rewarding / punishing signal is binary and is potentially delayed by several stimulus durations. We conceived a spatial navigation task where 7 positions on a circle are encoded each by a frozen 500 ms spike pattern in 100 afferents projecting to the dendritic branches of the model neuron as before. The task is to jump to position 0 when being in one of the other 6 positions and, after reaching position 0, staying there (Fig 5A). Actions consisted in either no jump or jumps of 1, 2 or 3 steps clock or counter clockwise. No jump is encoded by no somatic spikes, and a jump of *n* steps in the clock or counterclockwise direction is encoded by the first somatic spike arising in the *n*'th time bin to the left or right from the center (Fig 5A, inset). A positive reward signal $R = 1$ is delivered when the agent, being in a non-target position, directly jumps to the target, or when it is at the target position and stays there; else $R = -1$. After an initial average of 20 random actions needed to

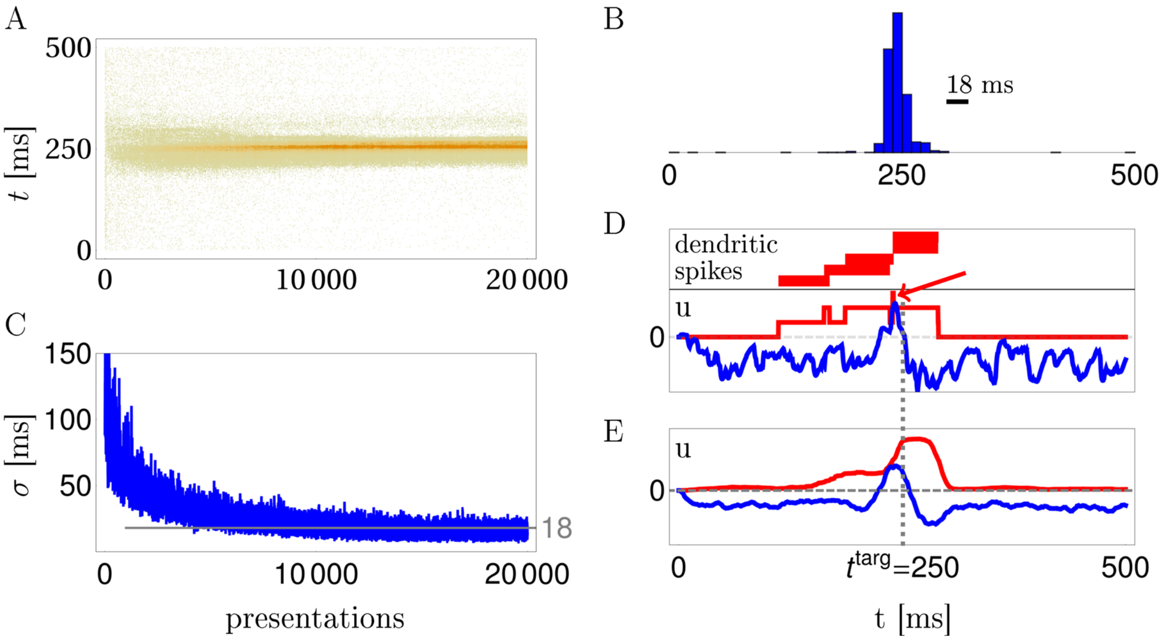

Fig 4. R-sdSP learns exact somatic spike timing. A: Somatic spike trains during 20000 trials in a reward based scenario. **B**: The distribution of somatic spikes after learning of the target time at 250 ms has a Gaussian half-width of 18 ms. **C**: Evolution of the width (*σ*) of the spike-time distribution during training. **D,E**: Separation of the somatic voltage into a contribution from the NMDA-spikes (red) and the subthreshold dendritic potentials (blue) for a single run (D) and averaged across 20 runs (E). Note that after learning the summed NMDA-spikes can form a narrow depolarizations at the target time beyond the duration of an individual spike (arrow in D).

doi:10.1371/journal.pcbi.1004638.g004

reach the target position, the R-sdSP modulated agents learned to eventually reach the target with a single action and stay there (Fig 5B and 5C). While initially the first somatic spike times of our model neuron were uniformly distributed across the 500 ms stimulus interval, the neuron eventually learned to respond in the appropriate time bin of 83 ms duration that encoded the correct jumps (Fig 5D). During learning, the dendritic branches develop a shared selectivity for the patterns and until the first NMDA-spikes become properly aligned in the correct time bin (Fig 5E).

Discussion

We derived a synaptic plasticity rule for synapses on active dendrites that minimizes errors in the supervised and maximizes reward in the reinforcement learning scenario. More precisely, the rule follows the gradient of (a lower bound of) the log-likelihood of reproducing a given spike train for supervised learning, and the gradient of the expected reward for reinforcement learning. The rule specifies the optimal timing between the presynaptic, dendritic and somatic spikes, including the time course of the postsynaptic voltages. We showed that neurons can only exploit the increased representational power of active dendrites when synaptic plasticity is modulated by both the somatic and the dendritic spiking. The suggested somato-dendritic spike-dependent synaptic plasticity (sdSP) learns to correctly respond to synaptic input patterns coding by either frozen spikes times or firing rates, while classical STDP fails. It is remarkable that the same plasticity induction that supports the learning of precise spiking in the supervised learning scenario also maximizes the expected reward when modulated by an

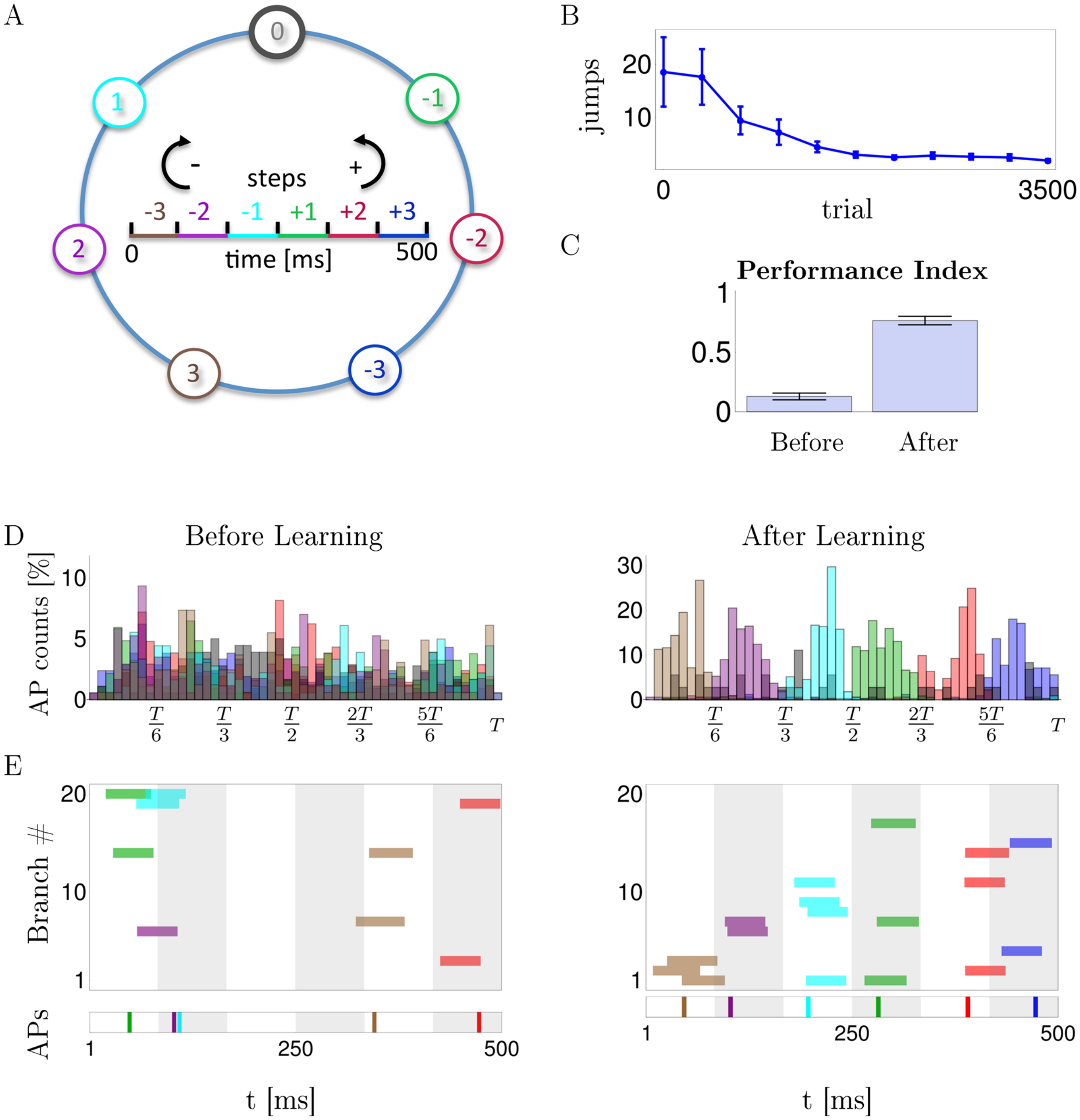

Fig 5. R-sdSP learns the correct spike-timing in a navigational task with binary and delayed feedback. A: At each position a fixed spike pattern is presented, and the timing of the first somatic spike tells how many steps in the clock (–) or counter clock (+) direction are taken. Color code of the time bin indicates the preferred spike timing for directly jumping to target position 0 when being at the correspondingly colored circle position (see text). **B**: Evolution of the mean number of jumps needed from a randomly chosen circle position until 0 is reached. **C**: Performance Index defined as the probability of directly jumping from any of the 6 circle positions to the target, and staying there if already at 0. Before learning this probability is 0.13, after learning it is 0.78. **D**: Histogram of first somatic spikes at the various positions before and after learning, averaged across patterns and learning runs (color code as in A). (**E**) Timing of the first NMDA spike in each of the 20 branches (upper panel) and the first somatic spike (lower panel) when stimulated with the patterns associated to the 6 circle positions (colors encode positions as in A). After learning, NMDA-spikes in 2-4 branches co-align and trigger somatic spike timing the appropriate time bin.

doi:10.1371/journal.pcbi.1004638.g005

internal and possibly delayed reward signal, irrespective of whether the postsynaptic code is based on spike times or firing rates.

The neuron model for which we derived the gradient rules considered dendritic spikes as saturating square-shaped depolarizations triggered by the crossing of a dendritic voltage threshold. We showed that the dual voltage-glutamate criterion for NMDA spikes reduces in the presence of balanced excitation and inhibition to a pure voltage criterion. This is because the glutamate condition is always satisfied when reaching the voltage threshold. This leads to a dendritic spike scenario that also includes dendritic sodium [36] or calcium spikes [37] differing in their voltage threshold, duration and amplitude. In the supervised learning scenario, the general plasticity rule we derived consists of a somatic error term that measures the difference between the actual spiking and the instantaneous spike rate, a dendritic rate- and spike-term, and a presynaptic spike term. Potentiation is triggered if the presynaptic spike is followed by a postsynaptic spike within roughly 10 ms, and this time window is stretched to roughly 50 ms if between the pre- and postsynaptic spike an additional dendritic spike occurs. Plasticity, be it potentiation or depression, can also be boosted by a mere nonlinear dendritic depolarizations without dendritic spikes, linking the rule also to computational models considering nonlinear but continuous dendritic processing [8–11, 24]. In the reinforcement learning scenario, the same plasticity rule is modulated by a global reward signal.

As learning is driven by a somatic error term, the synapses must be able to readout this error by disentangling the backpropagating spike and the somatic voltage (or at least a low-pass filtered version of it, see S2C–S2F Fig). Synapses must also read out the local dendritic spike and potential, and the synapse-specific postsynaptic potential (PSP). The PSP may be inferred from the concentration of the local glutamate released by the presynaptic bouton. The somatic and dendritic spikes may be determined from their characteristic voltage upstrokes and sustained depolarizations, and the NMDA spike can be further detected by a rapid increase in the local calcium concentration. Finally, the (subthreshold) somatic and dendritic depolarization may be distinguished by co-sensing local ion concentrations. In fact, the synaptically induced dendritic depolarization goes together with an increased local sodium concentration while the backpropagating somatic voltage does not cause such a ion influx. We assume that synapses developed a molecular machinery to extract these quantities and infer approximate estimates for the terms occuring in our plasticity rules.

Our computational framework for active dendrites contributes to the debate whether plasticity on dendritic branches should depend on the dendritic rather than the somatic spike [38], or whether it subserves synaptic clustering [23, 39] or a homeostatic adaptation [40]. In fact, when seen in the light of learning, synaptic plasticity is predicted to depend on all the postsynaptic quantities. Based on the model of dendritic NMDA receptor conductances in an *in vivo* stimulation scenario, the learning rule yields a guideline for experimental testings. For instance, it is in line with the observed synaptic depression induced by a synaptically generated dendritic spike alone ([41], but see [42]), or with the extended time window for plasticity induction involving NMDA-spikes [21]. It predicts that an NMDA-spike within roughly 50 ms after an excitatory synaptic input always enhances the synaptic modification. While the sign of the synaptic modification is determined by the presence or absence of a somatic spike following the synaptic input, an additional synaptically evoked NMDA-spike will only enhance it, never revert this sign.

Dendritic structures that have been suggested to form a 2-layer network [8] offer the additional advantage of easily backpropagating the information of the output to the synaptic sites 2 layers upstream. From a computational perspective it is interesting to note that, one the one hand, 2-layer networks represent an universal function approximator [43] while, on the other hand, networks with more than two layers are difficult to be trained [13]. For the dendritic implementation this suggests to limit the internal nonlinearities to a single layer of active dendritic branches. Because stacking dendritic nonlinearities across multiple layers would cause additional cross-talk, the restriction to a single dendritic nonlinearity may just be nature's solution to the trade-off between achieving more representational power and paying the associated signaling costs required for efficient learning.

Online Methods

Neuron parameters. The postsynaptic potentials are defined by $\mathrm{PSP}_i(t) = \sum \epsilon(t - t_i^{\mathrm{pre}})$, where the sum extends across all presynaptic spike times t_i^{pre} of afferent i. The spike reset is defined by $\kappa(t) = \sum_{t^s} \kappa_\circ(t - t^s)$, where the sum extends across all somatic spike times t^s. The two kernels are defined as

$$\epsilon(t) = \frac{\Theta(t)}{\tau_\mathrm{m} - \tau_\mathrm{s}} \left(e^{-t/\tau_\mathrm{m}} - e^{-t/\tau_\mathrm{s}}\right) , \quad \kappa_\circ(t) = \Theta(t) e^{-t/\tau_\mathrm{m}} ,$$

with $\tau_\mathrm{m} = 10$ ms and $\tau_\mathrm{s} = 1.5$ ms. Here, $\Theta(t) = 1$ for $t \geq 0$ and $\Theta(t) = 0$ for $t < 0$. The instantaneous rate for generating a NMDA-spike in dendrite d is $\rho_d^\mathrm{d}(t) \equiv \rho^\mathrm{d}(u_d^\mathrm{d}(t))$, and the instantaneous rate of generating a somatic spike is $\rho^\mathrm{s}(t) \equiv \rho^\mathrm{s}(u^\mathrm{s}(t))$ with

$$\rho^\mathrm{d}(u) = r_\mathrm{D}/(1 + \exp(-\beta_\mathrm{D}(u - \theta_\mathrm{D}))) , \quad \rho^\mathrm{s}(u) = \exp(\beta_\mathrm{s}(u - \theta_\mathrm{s})) . \tag{4}$$

This model of NMDA-spike generation can be deduced from the biophysical properties of NMDA receptors in a roughly balanced input scenario where the glutamate level required to activate the NMDA receptors is always reached for those voltages that also relieve their magnesium block (see S1A Text). The choice of the saturating rate function for the NMDA generation was motivated both by stability reasons, and also because the dendritic nonlinearities are saturating [11].

The neuronal parameters are $r_\mathrm{D} = 5$, $\beta_\mathrm{D} = \beta_\mathrm{s} = 5$, $\theta_\mathrm{s} = 2$ and $\theta_\mathrm{D} = 2.4$. We considered $n = 20$ branches and a dendritic attenuation factor $\alpha = 0.06$. The probability that one out of the 100 afferents is connected to a given branch was $p = 0.5$. The amplitude of the dendritic NMDA-spike is $a = 6$, its duration $\Delta = 50$ ms. Different NMDA-spikes in the same branch add in time but not in amplitude, yielding a dendritic plateau potential in branch d of the form $\mathrm{NMDA}_d(t) = a$ if at least one NMDA-spike was triggered in the interval Δ before t and $\mathrm{NMDA}_d(t) = 0$ else. For simplicity we assumed the two parameters α and a to be identical for all branches, but they may vary across branches or even be treated as adaptable dendritic 'coupling strengths' that could be learned by analogous gradient rules as suggested by experimental work [30].

The learning rule. To obtain an online rule that is identical in the supervised and reinforcement scenarios up to reward modulation, we consider an additional low-pass filtering of the instantaneous synaptic changes. Plasticity is then triggered when either the stimulus ends or when reward is applied. We introduce the instantaneous synaptic eligibility for somato-synaptic and the somato-dendro-synaptic contribution, respectively, by

$$e_{di}^{\mathrm{ss}}(t) \quad = \quad (S(t) - \rho^\mathrm{s}(t))\mathrm{PSP}_i(t) \tag{5}$$

$$e_{di}^{\mathrm{sds}}(t) \quad = \quad (S(t) - \rho_{\backslash d}^\mathrm{s}(t))\mathrm{Den}_d {*} \mathrm{PSP}_i(t) \tag{6}$$

with $S(t) = \sum_{t^s} \delta(t - t^s)$ representing the somatic spike train and $\rho_{\backslash d}^\mathrm{s}(t)$ the instantaneous somatic escape rate without contribution of the putative NMDA-spike from branch d at time t,

$$\rho_{\backslash d}^\mathrm{s}(t) = c\, \rho^\mathrm{s}(u^\mathrm{s}(t) - \alpha \mathrm{NMDA}_d(t)) . \tag{7}$$

Here, $c = (e^{\alpha a \beta_s} - 1)/(\alpha a \beta_s)$. Note that a low-pass filtered version of ρ^{s} and $\rho^{\mathrm{s}}_{\backslash d}$ could be extracted at the synaptic site by using the local ionic concentrations to disentangle the local, synaptically generated voltage and NMDA-spike from the backpropagated somatic voltage (see also Discussion). The factor $\mathrm{Den}_d{*}\mathrm{PSP}_i$ expresses a modulation of the synaptic signal PSP_i by the presence or absence of an NMDA-spike in branch d before t, i.e. within the time interval $t - \Delta$ to t. If there is such a spike, the last NMDA-spike initiation time at branch d in the interval $[t - \Delta, t]$ is denoted by t^{d}_d. We then set

$$\mathrm{Den}_d{*}\mathrm{PSP}_i(t) = \begin{cases} \dfrac{1}{\rho^{\mathrm{d}}_d(t^{\mathrm{d}}_d)} \rho^{\mathrm{d}'}_d(t^{\mathrm{d}}_d)\,\mathrm{PSP}_i(t^{\mathrm{d}}_d) & \text{if } t \text{ within NMDA–spike triggered at } t^{\mathrm{d}}_d, \\ \int_{t-\Delta}^{t} \rho^{\mathrm{d}'}_d(t')\mathrm{PSP}_i(t')dt' & \text{else}, \end{cases} \tag{8}$$

where $\rho^{\mathrm{d}'}_d = \frac{\beta_{\mathrm{D}}}{r_{\mathrm{D}}} \rho^{\mathrm{d}}_d (r_{\mathrm{D}} - \rho^{\mathrm{d}}_d)$ represents the derivative of $\rho^{\mathrm{d}}_d(u)$ with respect to $u = u^{\mathrm{d}}_d$.

The upper line in Eq (8) can be understood as a sampling version of the lower line: Let $D_d(t)$ be the sum of delta-functions centered at the triggering times of NMDA-spikes in branch d. In the case that the NMDA-spikes in the same dendritic branch would add up, the upper line becomes $\int_{t-\Delta}^{t} \frac{D_d(t')}{\rho^{\mathrm{d}}_d(t')} \rho^{\mathrm{d}'}_d(t')\mathrm{PSP}_i(t')dt'$, and this averages out to yield the lower line. Since in our case the NMDA-spike triggerings are rare, the two versions for the upper line are roughly equal. We further approximated the integral in the lower line by ς_{di} defined as low-pass filtered version of the integrand, $\dot{\varsigma}_{di} = -\varsigma_{di}/\overline{\Delta} + \rho^{\mathrm{d}'}_d \mathrm{PSP}_i$ with filtering time constant $\overline{\Delta} = \Delta/2$. It is also possible to define $\mathrm{Den}_d{*}\mathrm{PSP}_i(t)$ by any convex combination of the two lines in Eq (8), but an equal weighting of them yielded best performances in our simulations.

It is illustrative to deduce from the above formulas the limit when the dendritic spiking disappears and merely a dendritic nonlinearity remains. Remember that in deriving these formulas we allowed the NMDA spikes to add up, and since for biological frequences NMDA spikes rarely overlap, this assumption is justified. If NMDA spikes are still allowed to add up (although not to infinity), we may formally scale the NMDA amplitude, the NMDA duration and the NMDA rate function by a factor $\lambda \to 0$, i.e. replacing $a \to \lambda a$, $\Delta \to \lambda\Delta$ and $\rho^{\mathrm{d}}_d \to \rho^{\mathrm{d}}_d/\lambda^2$ and take the limit of λ converging to 0. The time course of the dendritic spikes in branch d, $\mathrm{NMDA}_d(t)$, is then replaced by the nonlinearly summed dendritic voltage $\rho^{\mathrm{d}}_d(t) = \rho^{\mathrm{d}}_d(u^{\mathrm{d}}_d(t))$, and the somato-dendro-synaptic contribution for synapse i on branch d (Eqs (2) and (6), respectively) becomes the 3-factor rule

$$e^{\mathrm{sds}}_{di}(t) = (S(t) - \rho^{\mathrm{s}}(t))\,\rho^{\mathrm{d}'}_d(t)\,\mathrm{PSP}_i(t)\,. \tag{9}$$

This rule could be seen as a gradient version of the pair-based rule in [24] and applies to the dendritic nonlinearities considered in [8–10]. An alternative derivation of the rule Eq (9) is to consider the deterministic somatic voltage $u^{\mathrm{s}} = \sum_d \alpha(u^{\mathrm{d}}_d + \rho^{\mathrm{d}}_d(u^{\mathrm{d}}_d)) - \kappa$ and derive the learning rule as in [28] for the case of the exponential somatic spike rate $\rho^{\mathrm{s}}(u^{\mathrm{s}})$ defined in Eq (4). The direct gradient calculation leads to the two plasticity components in Eqs (5) and (9), respectively, corresponding to the linear and nonlinear summation of the dendritic potentials u^{d}_d.

Coming back to the case with dendritic spikes, the two instantaneous eligibilities for sdSP (Eqs (5) and (6)) are again weighted and low-pass filtered,

$$\dot{E}_{di}(t) = -E_{di}(t)/\tau_{\mathrm{E}} + \bar{a}\, e^{\mathrm{sds}}_{di}(t) + e^{\mathrm{ss}}_{di}(t)\,, \tag{10}$$

with $\bar{a} = a/2$ and filtering time constant $\tau_{\mathrm{E}} = T/2$, $T = 500$ ms. The supervised learning rule (sdSP) is obtained by clamping the somatic output to the target spike train and updating the weights after each stimulus by the synaptic eligibility trace at that time,

$$\Delta w_{di}(T) = \eta\, E_{di}(T)\,,$$

with optimized learning rate η (see below). In reinforcement learning the synaptic weights are updated at the times t^{Rew} when a reward signal is applied, i.e. when a stimulus ends ($t^{\mathrm{Rew}} = T$ in Figs 2–5). For R-sdSP we obtain

$$\Delta w_{di}(t^{\mathrm{Rew}}) = \eta (R - R_0)\, E_{di}(t^{\mathrm{Rew}})\,.$$

The reward signal R depends on the input spike pattern and somatic spike train S, and R_0 is a baseline set to $R_0 = 1$ for all simulations with R-sdSP. Setting R_0 to a running average of the reward would speed up learning, but for simplicity we refrained from this optimization (see, however, the implementation of R-STDP in Eq (11), where R_0 must even depend on the stimulus). For a derivation of the sdSP and R-sdSP as gradient of a target function see S1C Text and [25].

Simulation details. For all tasks and learning rules we optimized the learning rate η such that the performance for the optimized value $\eta = \eta_\circ$ is better than for the adjacent values $\eta = 1.5\,\eta_\circ$ and $\eta = \eta_\circ/1.5$. In all the simulations involving R-sdSP we used the binary reward signal $R = \pm 1$, except for learning the very precise spike timing in Fig 4 that required a graded reward. There, $R = 1 - \sum_{t^{\mathrm{som}}} g(t^{\mathrm{som}} - t^{\mathrm{targ}})$ with $t^{\mathrm{targ}} = 250$ ms, $g(\delta) = 1 -\ \exp\left(-\frac{|\delta|}{T/2}\right)$, and the sum extending across all somatic spike times t^{som} (setting $t^{\mathrm{som}} = 0$ when no somatic spike was emitted).

The R-STDP plasticity rule was implemented in its best performing version found in [33]. More precisely, Frémaux et al. defined the eligibility

$$e_{di}^{\mathrm{STDP}}(t) = A_+\, S(t) \int_{-\infty}^{t} x_i(s)\, e^{(s-t)/\tau_+} ds + A_-\, x_i(t) \int_{-\infty}^{t} S(s)\, e^{(s-t)/\tau_-} ds$$

with $x_i(s) = \sum_{t_i^{\mathrm{pre}}} \delta(s - t_i^{\mathrm{pre}})$ being the presynaptic spike train in afferent i. They set $A_- = 0$, $A_+ = 1$, and τ_+ matched the synaptic time constant $\tau_s = 10$ ms. In our case the postsynaptic signal *post*(s) is either the somatic spike train $S(t)$ or the local dendritic spike train $D_d(t)$ in branch d. As pointed out in [33], the constant baseline reward R_0 used in gradient rules must be replaced by a running mean across pattern presentations that depends on the identity of the input pattern x, $\tilde{R}(x) = \frac{1}{\tau_p} R(x) + \left(1 - \frac{1}{\tau_p}\right) \tilde{R}(x)$. Here, the history length constant is here set to $\tau_p = 5$. The weight update after each stimulus presentation is

$$\Delta w_{di}^{\mathrm{STDP}}(t^{\mathrm{Rew}}) = \eta \left(R - \tilde{R}(x)\right) E_{di}^{\mathrm{STDP}}(t^{\mathrm{Rew}})\,. \tag{11}$$

where E_{di}^{STDP} is a low-pass version of the eligibility e_{di}^{STDP} with time constant τ_{E} (see Eq (10)).

Supporting Information

S1 Text. Supplementary Information. A: From the biophysics to a stochastic NMDA-spike model. **B**: Additional analysis and simulation results. **C**: Mathematical derivations.**D**: Appendix. (PDF)

S1 Fig. For balanced synaptic inputs, the NMDA-spike probability becomes a function of the voltage alone. A: Top: AMPA (full line), NMDA (dashed) and GABA$_A$ (dotted) currents, at the peak conductance level, as a function of u defined in Eqns S1 and S2 of the Supporting Information ($\alpha = \beta = 0.05$; excitatory currents with positive sign). Bottom: Voltage traces $u(t)$ for 6 different synaptic drives $g_\circ$ = 0,25,50,75,100,125nS (curves from light to dark, Eq S3), with NMDA-spikes elicited by the 2 strongest $g_\circ$. **B**: $I(u)$ ('I–V curves') defined by the right-

hand-side of Eq S4 for the 6 synaptic drives $g_\circ$ used in A. **C**: Zero crossings $I(u) = 0$ of the family of curves parametrized by $g_\circ$ and 6 of with shown in B, for different inhibitory-excitatory balancing ratios $\beta = 0.05$ (red), $\beta = 0.10$ (blue) and $\beta = 0.15$ (green); AMPA/NMDA ratio: $\alpha = 0.05$. **D**: The 6 voltage traces $u(t)$ from A plotted against the glutamate time course at the NMDA receptors, $g_\circ\, \varepsilon^{\mathrm{NMDA}}(t)$, overlaid on the red zero-crossing curve shown in C. **E**: Whenever the Gaussian noise (red cloud) added to the mean $(g_\circ, u)$ on the red line (center of cloud) drops into the green area, a NMDA-spike is elicited. **F**: The probability of eliciting a NMDA-spike at a given voltage ($P(\text{spike}|u)$, top) is almost the same for the three different balancing ratios β that vary by a factor of 3; it is therefore roughly proportional to the instantaneous spike rate $\varphi(u) \propto P(\text{spike}|u)$ that is a function of u alone. Yet, because u as a function of $g_\circ$ saturates (C), plotting $P(\text{spike}|u)$ versus $g_\circ$ may still give deviating curves (bottom).
(EPS)

S2 Fig. Robustness of R-sdSP against noise and imperfect voltage readout. A, **B**: When introducing a Gaussian jitter in the spike timings of the 4 frozen 6 Hz Poisson spike patterns (A) their classification into a spike / no spike code only smoothly degrades (B). Standard deviation of spike jitter: 10 ms (blue), 20 ms (red), 50 ms (green) and 100 ms (brown). **C**: The classification is still learnable by R-sdSP when the somatic voltage $u^{s}(t)$ is low pass filtered with different time constants: 5 ms (blue), 10 ms (red), 20 ms (green) and 40 ms (brown). **D**: The performance barely changes when only considering the somato-dendro-synaptic contribution $\dot{w}_{di}^{\mathrm{sds}}$ of the rule (Eq (5) in Online Methods, blue dashed). On the other hand, when learning is only based on the somo-synaptic contribution ($\dot{w}_{di}^{\mathrm{ss}}$, Eq (4) in Online Methods) the performance degrades (magenta). Inset: performances over the first 1000 presentations. **E**, **F**: Learning curves for R-STDP when the time constant τ_+ matches the NMDA-spike duration $\Delta = 50$ ms. **E**: Still, R-STDP cannot learn a binary classification of 4 randomized spatio-temporal spike patterns, both when applied to the presynaptic–somatic spikes (solid black; dashed: performance when the NMDA-spikes are suppressed) or the presynaptic–dendritic spikes (gray). **F**: Similarly, R-STDP is not able to learn the XOR-problem (curve legend as in E). Inset: average performance after each of the 20 runs.
(EPS)

Acknowledgments

We thank J.P. Pfister for discussions and feedback to an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: MS RU WS. Performed the experiments: MS. Analyzed the data: MS RU WS. Contributed reagents/materials/analysis tools: MS RU WS. Wrote the paper: MS WS.

References

1. Schiller J, Major G, Koester H.J, and Schiller Y. NMDA spikes in basal dendrites of cortical pyramidal neurons. *Nature*, 404(6775):285–289, 2000. doi: 10.1038/35005094 PMID: 10749211
2. Polsky A, Mel B.W, and Schiller J. Computational subunits in thin dendrites of pyramidal cells. *Nat. Neurosci.*, 7(6):621–627, 2004. doi: 10.1038/nn1253 PMID: 15156147
3. Nevian T, Larkum M.E, Polsky A, and Schiller J. Properties of basal dendrites of layer 5 pyramidal neurons: a direct patch-clamp recording study. *Nat. Neurosci.*, 10:206–214, Feb 2007. doi: 10.1038/nn1826 PMID: 17206140

4. Larkum M.E, Nevian T, Sandler M, Polsky A, and Schiller J. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science*, 325(5941):756–760, 2009. doi: 10.1126/science.1171958 PMID: 19661433

5. Branco T, Clark B.A, and Hausser M. Dendritic discrimination of temporal input sequences in cortical neurons. *Science*, 329(5999):1671–1675, Sep 2010. doi: 10.1126/science.1189664 PMID: 20705816

6. Jia H, Rochefort N.L, Chen X, and Konnerth A. Dendritic organization of sensory input to cortical neurons in vivo. *Nature*, 464(7293):1307–1312, 2010. doi: 10.1038/nature08947 PMID: 20428163

7. Lavzin M, Rapoport S, Polsky A, Garion L, and Schiller J. Nonlinear dendritic processing determines angular tuning of barrel cortex neurons in vivo. *Nature*, 2012. doi: 10.1038/nature11451 PMID: 22940864

8. Poirazi P, Brannon T, and Mel B.W. Pyramidal neuron as two-layer neural network. *Neuron*, 37:989–999, 2003. doi: 10.1016/S0896-6273(03)00149-1 PMID: 12670427

9. Caze R.D, Humphries M, and Gutkin B. Passive dendrites enable single neurons to compute linearly non-separable functions. *PLoS Comput. Biol.*, 9(2):e1002867, 2013. doi: 10.1371/journal.pcbi.1002867 PMID: 23468600

10. Breuer D, Timme M, and Memmesheimer R.M. Statistical Physics of Neural Systems with Nonadditive Dendritic Coupling. *Physical Review X*, 4(011053):1–23, 2014.

11. Tran-Van-Minh A, Caze R.D, Abrahamsson T, Cathala L, Gutkin B.S, and DiGregorio D.A. Contribution of sublinear and supralinear dendritic integration to neuronal computations. *Front Cell Neurosci*, 9:67, 2015. doi: 10.3389/fncel.2015.00067 PMID: 25852470

12. Rumelhart D.E, Hinton G.E, and Williams R.J. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. doi: 10.1038/323533a0

13. Hinton G.E and Salakhutdinov R.R. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006. doi: 10.1126/science.1127647 PMID: 16873662

14. Markram H, Lübke J, Frotscher M, and Sakmann B. Regulation of synaptic efficacy by concidence of postsynaptic APs and EPSPs. *Science*, 275:213–215, 1997. doi: 10.1126/science.275.5297.213 PMID: 8985014

15. Bi G.Q and Poo M.M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, 18(24):10464–10472, Dec 1998. PMID: 9852584

16. Sjostrom P.J, Turrigiano G.G, and Nelson S.B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6):1149–1164, 2001. doi: 10.1016/S0896-6273(01)00542-6 PMID: 11754844

17. Froemke R.C, Poo M.M, and Dan Y. Spike-timing-dependent synaptic plasticity depends on dendritic location. *Nature*, 434(7030):221–225, 2005. doi: 10.1038/nature03366 PMID: 15759002

18. Letzkus J.J, Kampa B.M, and Stuart G.J. Learning rules for spike timing-dependent plasticity depend on dendritic synapse location. *J. Neurosci.*, 26(41):10420–10429, 2006. doi: 10.1523/JNEUROSCI.2650-06.2006 PMID: 17035526

19. Sjostrom P.J and Hausser M. A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron*, 51(2):227–238, 2006. doi: 10.1016/j.neuron.2006.06.017 PMID: 16846857

20. Golding N.L, Staff N.P, and Spruston N. Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature*, 418(6895):326–331, 2002. doi: 10.1038/nature00854 PMID: 12124625

21. Gordon U, Polsky A, and Schiller J. Plasticity compartments in basal dendrites of neocortical pyramidal neurons. *J. Neurosci.*, 26(49):12717–12726, 2006. doi: 10.1523/JNEUROSCI.3502-06.2006 PMID: 17151275

22. Remy J and Spruston N. Dendritic spikes induce single-burst long-term potentiation. *Proc. Natl. Acad. Sci. U.S.A.*, 104(43):17192–17197, 2007. doi: 10.1073/pnas.0707919104 PMID: 17940015

23. Poirazi P and Mel B.W. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron*, 29(3):779–796, 2001. doi: 10.1016/S0896-6273(01)00252-5 PMID: 11301036

24. Legenstein R and Maass W. Branch-specific plasticity enables self-organization of nonlinear computation in single neurons. *J. Neurosci.*, 31(30):10787–10802, 2011. doi: 10.1523/JNEUROSCI.5684-10.2011 PMID: 21795531

25. Schiess M.E, Urbanczik R, and Senn W. Gradient estimation in dendritic reinforcement learning. *The Journal of Mathematical Neuroscience*, 2(2), 2012. doi: 10.1186/2190-8567-2-2 PMID: 22657827

26. Major G, Polsky A, Denk W, Schiller J, and Tank D.W. Spatiotemporally graded NMDA spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. *J. Neurophysiol.*, 99:2584–2601, May 2008. doi: 10.1152/jn.00011.2008 PMID: 18337370

27. Gerstner W, Kempter R, van Hemmen J.L, and Wagner H. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–81, Sep 1996. doi: 10.1038/383076a0 PMID: 8779718

28. Pfister J, Toyoizumi T, Barber D, and Gerstner W. Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Computation*, 18:1318–1348, 2006. doi: 10.1162/neco.2006.18.6.1318 PMID: 16764506

29. Brea J, Senn W, and Pfister J.P. Matching recall and storage in sequence learning with spiking neural networks. *J. Neurosci.*, 33(23):9565–9575, Jun 2013. doi: 10.1523/JNEUROSCI.4098-12.2013 PMID: 23739954

30. Losonczy A, Makara J.K, and Magee J.C. Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature*, 452(7186):436–441, Mar 2008. doi: 10.1038/nature06725 PMID: 18368112

31. Memmesheimer R.M, Rubin R, Olveczky B.P, and Sompolinsky H. Learning precisely timed spikes. *Neuron*, 82(4):925–938, May 2014. doi: 10.1016/j.neuron.2014.03.026 PMID: 24768299

32. Izhikevich E. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17:2443–2452, 2007. doi: 10.1093/cercor/bhl152 PMID: 17220510

33. Frémaux N, Sprekeler H, and Gerstner W. Functional requirements for reward-modulated spike-timing-dependent plasticity. *J. Neurosci.*, 30:13326–13337, Oct 2010. doi: 10.1523/JNEUROSCI.6249-09.2010 PMID: 20926659

34. Gütig R and Sompolinsky H. The tempotron: a neuron that learns spike timing-based decision. *Nature Neurosci.*, 9:420–428, 2006. doi: 10.1038/nn1643 PMID: 16474393

35. Urbanczik R and Senn W. A gradient learning rule for the tempotron. *Neural Comput.*, 21:340–352, 2009. doi: 10.1162/neco.2008.09-07-605 PMID: 19431262

36. Golding N.L and Spruston N. Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. *Neuron*, 21(5):1189–1200, Nov 1998. doi: 10.1016/S0896-6273(00)80635-2 PMID: 9856473

37. Cichon J and Gan W.B. Branch-specific dendritic Ca(2+) spikes cause persistent synaptic plasticity. *Nature*, 520(7546):180–185, Apr 2015. doi: 10.1038/nature14251 PMID: 25822789

38. Kampa B.M, Letzkus J.J, and Stuart G.J. Dendritic mechanisms controlling spike-timing-dependent synaptic plasticity. *Trends Neurosci.*, 30(9):456–463, Sep 2007. doi: 10.1016/j.tins.2007.06.010 PMID: 17765330

39. Larkum M.E and Nevian T. Synaptic clustering by dendritic signalling mechanisms. *Curr. Opin. Neurobiol.*, 18(3):321–331, Jun 2008. doi: 10.1016/j.conb.2008.08.013 PMID: 18804167

40. Goldberg J, Holthoff K, and Yuste R. A problem with Hebb and local spikes. *Trends Neurosci.*, 25(9):433–435, 2002. doi: 10.1016/S0166-2236(02)02200-2 PMID: 12183194

41. Holthoff K, Kovalchuk Y, Yuste R, and Konnerth A. Single-shock LTD by local dendritic spikes in pyramidal neurons of mouse visual cortex. *J. Physiol. (Lond.)*, 560(Pt 1):27–36, 2004. doi: 10.1113/jphysiol.2004.072678

42. Gambino F, Pages S, Kehayas V, Baptista D, Tatti R, Carleton A, and Holtmaat A. Sensory-evoked LTP driven by dendritic plateau potentials in vivo. *Nature*, 515(7525):116–119, Nov 2014. doi: 10.1038/nature13664 PMID: 25174710

43. Hornik K, Stinchcombe M, and White H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989. doi: 10.1016/0893-6080(89)90020-8

Somato-dendritic synaptic plasticity and error-backpropagation in active dendrites

Supplementary Information

Mathieu Schiess[1], Robert Urbanczik[1], Walter Senn[1,2]
1 Department of Physiology, University of Bern, Bern, Switzerland
2 Center for Cognition, Learning and Memory, University of Bern

Contents

S1A From the biophysics to a stochastic NMDA-spike model

This Section shows that the simplified NMDA-spike model described in the main text represents a viable approximation of the full conductance-based NMDA model in the presence of an *in vivo*-like input scenario. In this scenario the AMPA and GABA$_A$ conductances are assumed to be roughly balanced, say with GABA/AMPA ratios varying

Email address: `schiess,urbanczik,senn@pyl.unibe.ch` (Mathieu Schiess[1], Robert Urbanczik[1], Walter Senn[1,2]).

between 1 to 3. This implies that the high voltages where the NMDA-receptors are unblocked from the magnesium can only be reached when there is also enough glutamate present to activate them. As a consequence, the voltage alone becomes the criterium for triggering an NMDA-spike (see Fig. S1F, top).

S1A.1 Biophysical model of NMDA-spikes

Here we describe the state-of-the art biophysical model of NMDA-spike generation [1, 2, 3]. The NMDA conductance g^{N} depends on the peak conductance of a unit NMDA receptor $\bar{g}^{\mathrm{N}}$ ($= 3.9\,$[nS], see [4]), the released glutamate, and the postsynaptic voltage u. The voltage dependence is modeled by the sigmoidal function

$$\sigma(u) = \frac{1}{1 + \exp\left(-\frac{u - V^{\mathrm{N}}_{1/2}}{V^{\mathrm{N}}_{\mathrm{spread}}}\right)}$$

with $V^{\mathrm{N}}_{1/2} = -20\,\mathrm{mV}$ and $V^{\mathrm{N}}_{\mathrm{spread}} = 12.5\,\mathrm{mV}$ [1]. The time dependence of the NMDA conductance on the glutamate released at $t = 0$, is modeled by the kernel function

$$\varepsilon^{\mathrm{N}}(t) = \Theta(t) B^{\mathrm{N}} \left(e^{-t/\tau^{\mathrm{N}}_1} - e^{-t/\tau^{\mathrm{N}}_2}\right)$$

where $\Theta(t)$ is the Heaviside step function ($= 0$ for $t < 0$ and 1 else), the constants $\tau^{\mathrm{N}}_1 = 40$ms and $\tau^{\mathrm{N}}_2 =$3ms determine the rise and fall of the kernel, and the factor $B^{\mathrm{N}} = 1.33$ that normalizes the peak amplitude of ε^{N} [5]. The NMDA conductance induced by the glutamate release becomes

$$g^{\mathrm{N}} = g_{\circ}\, \varepsilon^{\mathrm{N}}(t)\, \sigma(u)\,.$$

Since $g_{\circ}$ is proportional to the peak glutamate level and as such will also scale the AMPA currents, and since in a balanced input scenario it will further be proportional to the peak inhibitory current, we will term $g_{\circ}$ below as synaptic drive.

Glutamate is also assumed to activate AMPA receptors that generate a total AMPA conductance proportional to the synaptic drive, $g^{\mathrm{A}} = \alpha\, g_{\circ}$, with proportionality factor $\alpha = 0.05$. The AMPA kernel is given by the alpha-function

$$\varepsilon^{\mathrm{A}}(t) = \Theta(t)\, \frac{t}{\tau^{\mathrm{A}}\, e} e^{-t/\tau^{\mathrm{A}}}$$

with time constant τ^{A} =5ms. The total excitatory synaptic input current to the dendritic compartment for a given peak glutamate level then becomes the sum of the

AMPA and NMDA current,

$$I^{\mathrm{E}}_{\mathrm{syn}}(t) = g_\circ\, \alpha\, \varepsilon^{\mathrm{A}}(t)(E_{\mathrm{A}} - u) + g_\circ \varepsilon^{\mathrm{N}}(t)\sigma(u)(E_{\mathrm{N}} - u) \tag{S1}$$

where $E_{\mathrm{A}} = E_{\mathrm{N}} = 0$ represent the reversal potentials for the AMPA and NMDA receptors.

We assume that the excitatory synaptic input to some degree is balanced by inhibitory synaptic input $I^{\mathrm{I}}_{\mathrm{syn}}$. The GABAergic conductance strength g^{G} is proportional to the synaptic drive for a specific glutamate level, $g^{\mathrm{G}} = \beta g_\circ$, and some balancing factor $\beta = 0.05$. The GABA kernel is given by an alpha function

$$\varepsilon^{\mathrm{G}}(t) = \Theta(t)\, \frac{t}{\tau^{\mathrm{G}}\, e} e^{-t/\tau^{\mathrm{G}}}$$

with τ^{G} =5ms [6]. The inhibitory current then becomes

$$I^{\mathrm{I}}_{\mathrm{syn}}(t) = g_\circ\, \beta\, \varepsilon^{\mathrm{G}}(t)\, (E_{\mathrm{G}} - u)\ , \tag{S2}$$

where $E_{\mathrm{G}} = -70$ is the reversal potential of the GABA$_A$ conductance. Note that for a AMPA/NMDA ratio $\alpha = 0.05$ and GABA/NMDA ratio $\beta = 0.05$ the AMPA/GABA ration becomes 1.

Besides the synaptic input to the dendritic compartment, its membrane potential is modulated by a constant leak conductance, $\bar{g}_{\mathrm{L}}$, and an additional voltage-dependent potassium conductance resulting in the K^+ inward rectifying (KIR) current [2, 3]. The KIR voltage-dependence is modeled by a sigmoidal function that monotonically decreases with increasing voltage, with half activation at $V^{\mathrm{KIR}}_{1/2} = -70$mV and spread $V^{\mathrm{KIR}}_{\mathrm{spread}}$ =12.5 [3],

$$\kappa(u) = \frac{1}{1 + \exp\left(\frac{u - V^{\mathrm{KIR}}_{1/2}}{V^{\mathrm{KIR}}_{\mathrm{spread}}}\right)}\ .$$

Overall, the membrane potential u of the dendritic compartment is governed by the dynamics

$$C_m \dot{u} = \bar{g}_{\mathrm{L}}\, (E_{\mathrm{L}} - u) + \bar{g}_{\mathrm{KIR}} \kappa(u)\, (E_{\mathrm{K}} - u) + I^{\mathrm{E}}_{\mathrm{syn}}(t) + I^{\mathrm{I}}_{\mathrm{syn}}(t)\,, \tag{S3}$$

where $E_{\mathrm{L}} = -65$mV and $E_{\mathrm{K}} = -80$mV denote the leak and potassium reversal potentials, $\bar{g}_{\mathrm{L}} = 7$nS is the leak conductance, $\bar{g}_{\mathrm{KIR}} = 8\,\bar{g}_{\mathrm{L}}$ is the KIR peak conductance [3], and $C_m = 70$nF is the membrane capacitance (yielding a time constant of $\tau_m = C_m/\bar{g}_{\mathrm{L}} = 10\,\mathrm{ms}$).

S1A.2 Reduced model of the NMDA-spike generation

We next show how the biophysical model described above can be reduced to a model in which the generation of an NMDA-spike only depends on voltage (Fig. S1), with the glutamate dependence being negligible. To justify this simplification we note that for balanced input the NMDA-spikes are triggered at roughly the same voltage independently of the glutamate level. In fact, an NMDA-spike is triggered if the voltage is high enough to unblock the magnesium, provided enough glutamate is present. Crucially, for balanced excitation and inhibition this minimal glutamate level is always reached at the unblocking voltage, and more glutamate only marginally increases the amplitude of the NMDA-spike. This limited amplitude is due to the saturation of the driving force at high voltages.

To formalize the reasoning we insert the expressions for the excitatory (S1) and inhibitory current (S2) into the dynamics for the voltage (S3). We assume that the synaptically driven input currents are all proportional to the same synaptic drive $g_\circ$ and consider the stationary solutions of

$$\begin{aligned} C_m\dot{u} = &\, \bar{g}_{\mathrm{L}}\,(E_{\mathrm{L}} - u) + \bar{g}_{\mathrm{KIR}}\kappa(u)\,(E_{\mathrm{K}} - u) + g_\circ\beta(E_{\mathrm{G}} - u) + \\ &+ g_\circ\alpha(E_{\mathrm{A}} - u) + g_\circ\sigma(u)(E_{\mathrm{N}} - u)\,. \end{aligned} \tag{S4}$$

Abbreviating the right-hand-side of S4 by I this translates to $C_m\dot{u} = I$ (with a positive I leading to a depolarization). For each value of u plugged into the right-hand-side of S4 this gives a total current $I(u)$. When identifying the voltage with the symbol $V \equiv u$ we obtain the classical I–V curves for different values of synaptic drives $g_\circ$. The I–V curves for the individual, synaptically driven currents AMPA, NMDA and GABA$_A$ currents are displayed in Fig. S1A, top. Together with the leak and KIR current the form a N-shaped the overall I–V curve (Fig. S1B) that underlies the generation of the NMDA-spikes (Fig. S1A, bottom; Eq. S3). The zero-crossings of these curves, $I(u) = 0$, give the sustained voltage u for a given drive $g_\circ$ (i.e. for which $\dot{u} = 0$). These stationary voltages as a function of $g_\circ$ form the S-shaped curves in Fig. S1C, with colors indicating different balancing ratios β of excitation and inhibition. For low and high synaptic drives there is a unique stable u, but for intermediate values of $g_\circ$ two stable solutions with an intermediate unstable solution coexist.

When plotting the voltage trajectories $u(t)$ of panel A (bottom) against the time-dependent synaptic NMDA drive, $g_\circ\,\varepsilon^{\mathrm{N}}(t)$, into the $(g_\circ, u)$ phase plane, the trajectories showing an NMDA-spike make the turn around the S-shaped steady-state curve (Fig. S1, D). For a given pair synaptic drive and stationary voltage, $(g_\circ, u)$, we may ask for the likelihood that a NMDA-spike is elicited, given some Gaussian noise $\xi_{g_\circ}$ and

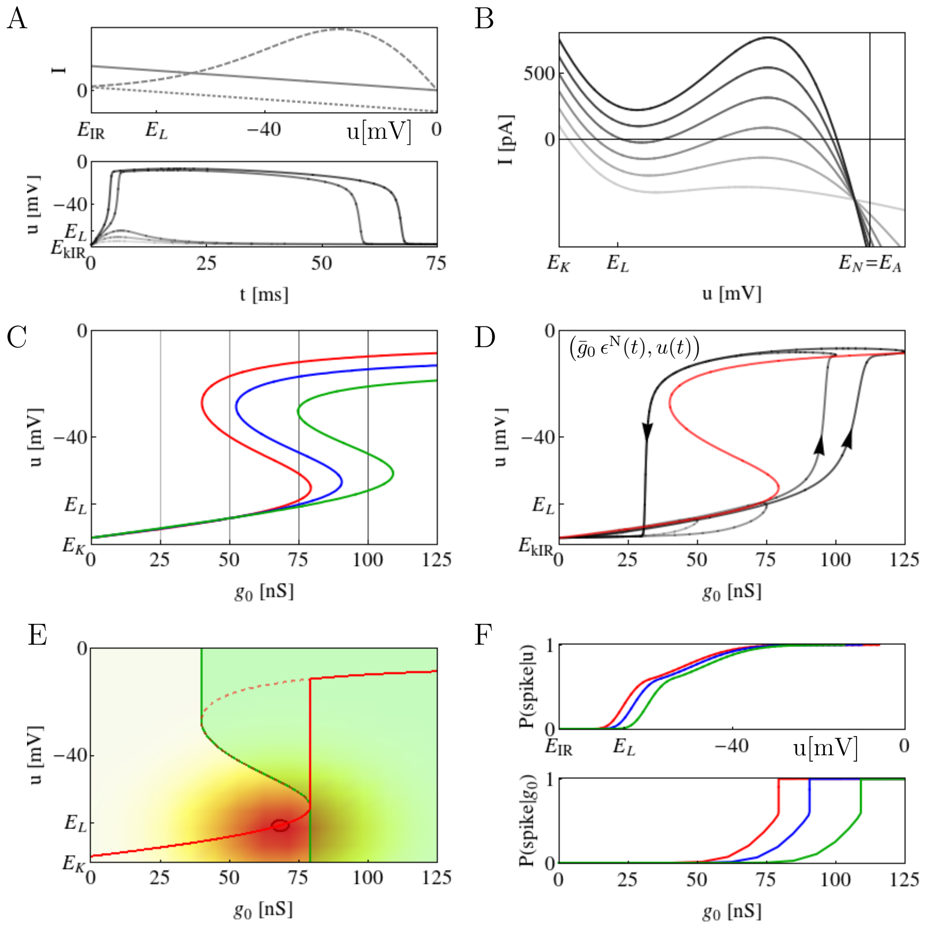

Figure S1. For balanced synaptic inputs, the NMDA-spike probability becomes a function of the voltage alone. **A**: Top: AMPA (full line), NMDA (dashed) and GABA$_A$ (dotted) currents, at the peak conductance level, as a function of u defined in Eqns S1 and S2 ($\alpha = \beta =$0.05; excitatory currents with positive sign). Bottom: Voltage traces $u(t)$ for 6 different synaptic drives $g_\circ = 0,\ 25,\ 50,\ 75, 100, 125\,\mathrm{nS}$ (curves from light to dark, Eq. S3), with NMDA-spikes elicited by the 2 strongest $g_\circ$. **B**: $I(u)$ ('I–V curves') defined by the right-hand-side of Eq. S4 for the 6 synaptic drives $g_\circ$ used in A. **C**: Zero crossings $I(u) = 0$ of the family of curves parametrized by $g_\circ$ and 6 of with shown in B, for different inhibitory-excitatory balancing ratios $\beta = 0.05$ (red), $\beta = 0.10$ (blue) and $\beta = 0.15$ (green); AMPA/NMDA ratio: $\alpha =$0.05 . **D**: The 6 voltage traces $u(t)$ from A plotted against the glutamate time course at the NMDA receptors, $g_\circ \varepsilon^{\mathrm{N}}(t)$, overlaid on the red zero-crossing curve shown in C. **E**: Whenever the Gaussian noise (red cloud) added to the mean $(g_\circ, u)$ on the red line (center of cloud) drops into the green area, a NMDA-spike is elicited. **F**: The probability of eliciting a NMDA-spike at a given voltage ($P(\text{spike}|u)$, top) is almost the same for the three different balancing ratios β that vary by a factor of 3; it is therefore roughly proportional to the instantaneous spike rate $\phi(u) \propto P(\text{spike}|u)$ that is a function of u alone. Yet, because u as a function of $g_\circ$ saturates (C), plotting $P(\text{spike}|u)$ versus $g_\circ$ may still give deviating curves (bottom).

ξ_u) added to $g_\circ$ and u, respectively (Fig. S1, E; with standard deviation $\sigma_u = 8$ and $\sigma_{g_\circ} = 3$). This likelihood is given by the probability that a point $(g_\circ + \xi_{g_\circ}, u + \xi_u)$ of the red cloud falls into the green area on the right part in panel E. When plotting the likelihood for an NMDA-spike as a function of the mean voltage that moves along the stable branch (and jumps up at the lower knee along the red line in Fig. S1E) we obtain a sigmoidal function that is roughly independent of the balancing factor β (Fig. S1F, top). Nevertheless, the same likelihood as a function of the synaptic drive (reflecting the total glutamate) yield strongly differing curves (Fig. S1F, bottom). Hence, while different balancing ratios lead to different glutamate concentrations that are required to trigger an NMDA-spike, these spikes are triggered with roughly the same likelihood at the same voltages. This justifies the stochastic NMDA-spike generation model that produces NMDA-spikes with instantaneous rate $\phi(u)$, see Fig. S1F top, as a function of the membrane potential, independently of the glutamate level.

S1B Additional analysis and simulation results

S1B.1 Robustness against noise and errors in the voltage readout

We further analyzed the robustness of the suggested reward-modulated synapto-denritic synaptic plasticity (R-sdSP) based on the classification of the 4 spatio-temporal spike patterns (as presented in Fig. 2 of the Main Text). As we have shown, the learning rule is able to classify spike patterns with frozen presynaptic spike timings and random frozen spike timings which were generated by Poisson processes with specific rates. To interpolate between these two extreme coding scenarios we also considered presynaptic spike-patterns that show stochastic spike-timings of varying degrees of stochasticity. Starting with the 4 frozen spike patterns generated once with a 6 Hz Poisson process, we perturbed each of these spike-times by a Gaussian of mean 0 and standard deviation σ (Fig. S2A, B). The learning performance shows a high robustness against these perturbations. The mean inter-spike interval for the original and perturbed spike trains are 167 ms. Even when the spike-time jitter has a width of $2\sigma = 200\,\text{ms}$ was the learning rule able to classify the patterns with an average performance of $\sim 90\%$ (Fig. S2B).

To explore the robustness against a dilution of the backpropagated voltage we low-pass filtered the somatic voltage $u^{\text{s}}(t)$ with different time constants up to 40 ms. Learning is still possible, although it slows down with increasing filtering time constant (Fig. S2C). Note that from the low-pass filtered version $\tilde{u}^{\text{s}}$ of the somatic voltage the synapse could calculate $\rho^{\text{s}}_{\setminus d}(t)$ since it has access to the local NMDA-spike in branch d and hence could subtract the contribution from the own branch. Moreover, since the passive backpropagation of the somatic voltage, the synaptic input currents and the NMDA-spikes involve different changes in ionic concentrations, a synapse sensing these concentrations may in principle disentangle the various contributions to the local voltage.

S1B.2 Dendritic contribution to R-sdSP and comparison with R-STDP

Next, we investigated the learning based on the individual components of R-sdSP. Recall that the weight change Δw_{di} of the reward gradient rule R-sdSP is composed of two components, a somato-synaptic contribution $R\,\Delta w^{\mathrm{ss}}_{di}$ originating from the forward propagated subthreshold dendritic potential, and a somato-dendro-synaptic contribution $R\,\Delta w^{\mathrm{sds}}_{di}$ originating from the supra-threshold dendritic plateau potentials sustained by the NMDA-spikes (Eq. 3 in the Main Text). As expected, learning based on the supra-threshold component $R\,\Delta w^{\mathrm{sds}}_{di}$ alone is equally fast as learning based on the full R-sdSP, but the subthreshold component $R\,\Delta w^{\mathrm{ss}}_{di}$ alone is considerably slower as it does not take account of the crucial dendritic spiking (Fig. S2D).

In the Main Text we have shown that 'classical' reward-modulated spike-timing dependent plasticity (R-STDP) [7, 8] does not reach the performance of R-sdSP (Fig. 2B and 3B,C). Here we further show that R-STDP does not perform better in the classification of the frozen spike patterns when the time constant τ_+ matches the duration of a NMDA-spike (Δ =50ms, Fig. S2E). In contrast to the gradient rule, R-STDP is not able to learn more than 75% in the presence of the dendritic spikes. The performance improves but remains below the gradient rules when the dendritic spikes are suppressed in the neuronal processing. The wider learning window in R-STDP is neither helping to improve learning for the XOR-problem that is encoded in mean firing rates (Fig. S2F).

S1B.3 Additional simulation details

In all simulations initial weights were picked independently from a Gaussian distribution with mean 0. The variance was set such that at least one somatic spike was elicited for half of the pattern presentations.

Input patterns were defined for 100 afferents. For the tasks involving temporal codes, a pattern was generated once with a constant Poisson rate of 6Hz for each afferent and the spike timings were then frozen. For the rate tasks (XOR and direction selectivity, Fig. 3 of the Main Text) each presentation was using a new realization of the pattern. For

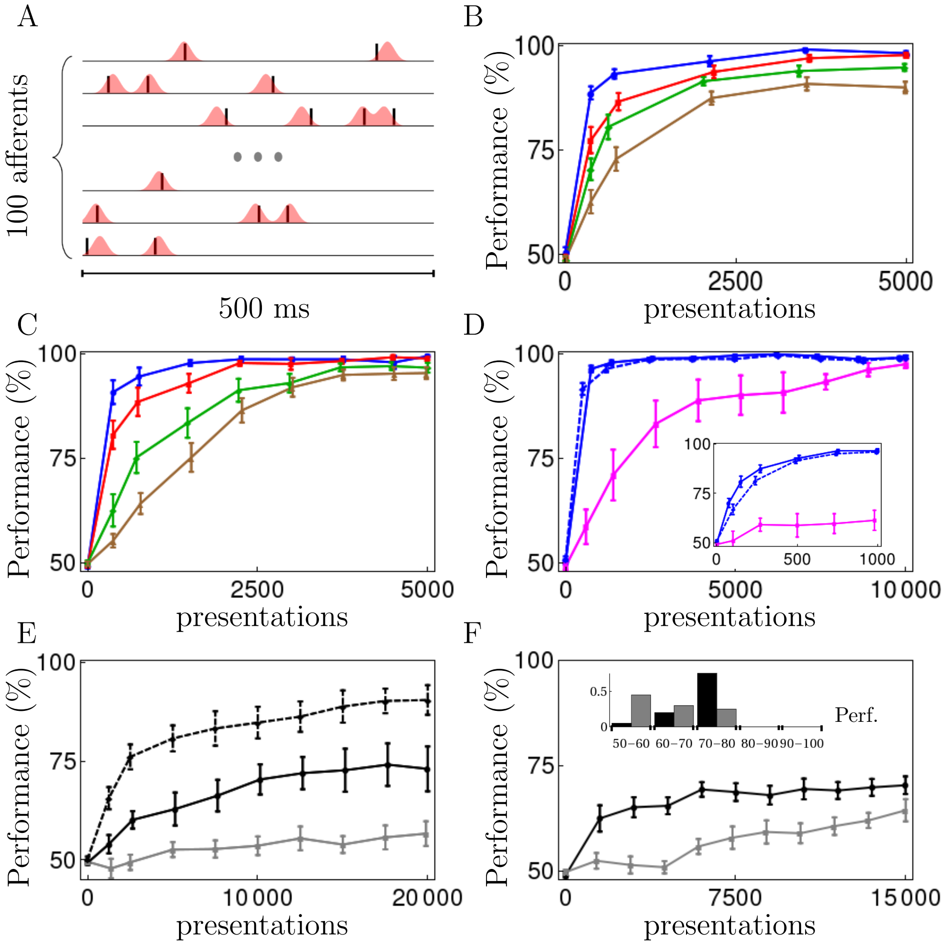

Figure S2. Robustness of R-sdSP against noise and imperfect voltage readout. **A**, **B**: When introducing a Gaussian jitter in the spike timings of the 4 frozen 6 Hz Poisson spike patterns (A) their classification into a spike / no spike code only smoothly degrades (B). Standard deviation of spike jitter: 10ms (blue), 20ms (red), 50ms (green) and 100 ms (brown). **C**: The classification is still learnable by R-sdSP when the somatic voltage $u^{\mathrm{s}}(t)$ is low pass filtered with different time constants: 5ms (blue), 10ms (red), 20ms (green) and 40 ms (brown). **D**: The performance barely changes when only considering the somato-dendro-synaptic contribution $\dot{w}_{di}^{\mathrm{sds}}$ of the rule (Eq. 5 in Online Methods, blue dashed). On the other hand, when learning is only based on the somo-synaptic contribution ($\dot{w}_{di}^{\mathrm{ss}}$, Eq. 4 in Online Methods) the performance degrades (magenta). Inset: performances over the first 1000 presentations. **E**, **F**: Learning curves for R-STDP when the time constant τ_+ matches the NMDA-spike duration Δ =50ms. **E**: Still, R-STDP cannot learn a binary classification of 4 randomized spatio-temporal spike patterns, both when applied to the presynaptic–somatic spikes (solid black; dashed: performance when the NMDA-spikes are suppressed) or the presynaptic–dendritic spikes (gray). **F**: Similarly, R-STDP is not able to learn the XOR-problem (curve legend as in E). Inset: average performance after each of the 20 runs.

the XOR problem the afferents had low (5Hz) or high (40Hz) Poisson firing rates that were again constant during the whole stimulus duration. For the direction selectivity task afferents had a low background firing rate (5Hz) replaced by a moving high firing rate interval (100Hz) of 15ms duration. An input pattern had a duration of 500ms for all tasks except the direction selectivity task which learns patterns with a duration of 100ms.

To obtain a learning curve, a running mean of the performance across presentations was computed with exponential decay constant 0.2/p, where p denotes the number of patterns to be learned. These running means were again averaged across 20 runs of the full learning for different weight and pattern initializations.

S1C Mathematical derivation of the learning rules

S1C.1 Derivation of the error-minimizing supervised learning rule (sdSP)

The aim of the supervised plasticity rule is to learn stimulus-response pairs $(\mathbf{x}, z)$ where $\mathbf{x}$ denotes a full set of presynaptic spike trains and z is the somatic spike train as a sum of delta functions ($z(t) = \sum_{t^{\mathrm{s}}} \delta(t - t^{\mathrm{s}})$, denoted as $S(t)$ in the Online Methods). Each pair $(\mathbf{x}, z)$ is drawn from a target distribution $P^*(z, \mathbf{x})$. Here we show that learning with the supervised plasticity rule maximizes a cost function. This cost function is a lower bound on the log-likelihood

$$\mathcal{L}(\mathbf{w}) = \langle \log P_{\mathbf{w}}(z|\mathbf{x}) \rangle_{P^*(z,\mathbf{x})} = \int \mathrm{d}\mathbf{x} \mathrm{d}z \, P^*(z, \mathbf{x}) \, \log P_{\mathbf{w}}(z|\mathbf{x}).$$

Note that maximizing $\mathcal{L}(\mathbf{w})$ is equivalent to minimizing the Kullback-Leibler divergence of the learned distribution P to the target distribution P^*.

In our 2-layer architecture, the conditional probability $P_{\mathbf{w}}(z|\mathbf{x})$ is not analytically tractable since the activity of dendritic branches acts as hidden variables. We denote by y_i the NMDA-spike timings of the i-th branch. In addition, the entire set of NMDA-spikes trains is denoted by $\mathbf{y} = (y_1, \ldots, y_N)$. To compute the log-likelihood, we marginalize out the hidden variables $\mathbf{y}$ in the expression,

$$\mathcal{L}(\mathbf{w}) = \left\langle \log \int \mathrm{d}\mathbf{y} \, P_{\mathbf{w}}(z|\mathbf{x}, \mathbf{y}) \, P_{\mathbf{w}}(\mathbf{y}|\mathbf{x}) \right\rangle_{P^*(z,\mathbf{x})} .$$

We apply Jensen's inequality to show that the cost function

$$\begin{aligned}\mathcal{C}(\mathbf{w}) &= \left\langle \int \mathrm{d}\mathbf{y}\, P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})\, \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y}) \right\rangle_{P^*(z,\mathbf{x})} \\ &= \langle \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\rangle_{P^*(z,\mathbf{x})\, P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})}\,.\end{aligned}$$

bounds the log-likelihood from below ($\mathcal{L}(\mathbf{w}) \geqslant \mathcal{C}(\mathbf{w})$). In the sequel, the notation $\langle\cdot\rangle$ alone means that the expression is averaged over $P^*(z,\mathbf{x})\, P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})$. The cost function $\mathcal{C}(\mathbf{w})$ is maximized via a stochastic gradient algorithm. The derivative of $\mathcal{C}(\mathbf{w})$ with respect to the synaptic weight w_{di} is

$$\tfrac{\partial}{\partial w_{di}}\mathcal{C}(\mathbf{w}) = \left\langle \tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\right\rangle + \left\langle \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\, \tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})\right\rangle. \tag{S5}$$

As computed in [9], the gradient of the first term on the RHS is expressed as (see Eq. (4) in Online Methods)

$$\begin{aligned}\tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y}) &= \int_0^T \mathrm{d}t\, \beta_{\mathrm{s}}\, \left(\alpha\, \mathrm{PSP}_i(s)\right)\left(z(t) - \rho^{\mathrm{s}}(t)\right) \\ &= \int_0^T \mathrm{d}t\, \beta_{\mathrm{s}}\, \alpha\, e^{\mathrm{ss}}_{di}(t).\end{aligned} \tag{S6}$$

In addition, we can manipulate the second term of the RHS to exhibit a efficient gradient estimator [10]. The procedure consists in averaging the term that accounts for the neuronal output $\log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})$ over the hidden variable y_d at each point in time.

Let $\mathbf{y}^{\backslash d}$ denote the vector of all NMDA-spike trains but the d-th and $\mathbf{w}^{\backslash d}$ the collection of synaptic weights in all but the d-th dendritic branch. Conditioned on the input stimulus $\mathbf{x}$, each dendritic spike train is generated independently ($P_{\mathbf{w}}(\mathbf{y}|\mathbf{x}) = P_{\mathbf{w}^{\backslash d}}(\mathbf{y}^{\backslash d}|\mathbf{x})\, P_{w_{\cdot d}}(y_d|\mathbf{x})$), thus we write

$$\begin{aligned}\left\langle \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\, \tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})\right\rangle &= \int \mathrm{d}\mathbf{x}\mathrm{d}z\mathrm{d}\mathbf{y}^{\backslash d}\, P^*(z,\mathbf{x}) P_{\mathbf{w}^{\backslash d}}(\mathbf{y}^{\backslash d}|\mathbf{x})\, c'_d(w_{\cdot,d}) \\ \text{with}\quad c'_d(w_{\cdot d}) &= \int \mathrm{d}y_d\, \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y}^{\backslash d}, y_d)\, \tfrac{\partial}{\partial w_{di}} P_{w_{\cdot d}}(y_d|\mathbf{x})\,.\end{aligned} \tag{S7}$$

In the definition of c'_d, we can regard $\mathbf{x}$ and $\mathbf{y}^{\backslash d}$ as fixed and suppress them in the notation. To shorten the notation we use y, $w_{\cdot}$ and w instead of y_d, $w_{\cdot d}$ and w_{di}respectively. We now replace the Poisson process generating y ($= y_d$) by a discrete time process with step-size $\delta > 0$. The time interval $[0, T]$ is divided into K intervals of length δ. The probability to trigger a spike in interval k is

$$P_{w_{\cdot}}(Y_k = 1) = 1 - e^{-\delta \rho^{\mathrm{d}}(t_k)} \tag{S8}$$

where $t_k = k\,\delta$. Here the bold notation $\mathbf{Y} = (Y_1, \ldots, Y_K)$ denotes the full series of discrete binary events in the dendritic branch. With this definition, we can recover the original Poisson process by taking the limit $\delta \to 0^+$. We denote by $\tilde{\mathbf{Y}}$ the set of NMDA-spike timings in $\mathbf{Y}$, i.e. $\tilde{\mathbf{Y}} = \{t_k | Y_k = 1\}$. Therefore, we can regard $c_d'(w)$ as the limit

$$c_d'(w) = \lim_{\delta \to 0^+} \sum_{\mathbf{Y}} \log P_\mathbf{w}(z|\tilde{\mathbf{Y}}) \, \tfrac{\partial}{\partial w} P_{w.}(\mathbf{Y})$$

where the sum runs over the set $\{0,1\}^K$. Since the local firing rate $\rho^\mathrm{d}(t_k)$ in eq. (S8) (see the Online Methods) depends only on the input $\mathbf{x}$ (y is generated by an inhomogeneous Poisson process), each Y_k are independently generated. We can express $P_{w.}(\mathbf{Y})$ as the product $P_{w.}(\mathbf{Y}^{\backslash k}) P_{w.}(Y_k)$ where $\mathbf{Y}^{\backslash k}$ denotes the full set of discrete events (spikes) but the k-th ($k = 1, \ldots, K$). Therefore, we can express the function $c_d'(w)$ as

$$c_d'(w) = \lim_{\delta \to 0^+} \sum_{k=1}^{K} \mathrm{grad}_k,$$

with

$$\mathrm{grad}_k = \sum_{\mathbf{Y}} \log P_\mathbf{w}(z|\tilde{\mathbf{Y}}) \, P_{w.}(\mathbf{Y}^{\backslash k}) \tfrac{\partial}{\partial w} P_{w.}(Y_k) \, .$$

We analytically compute the average of grad_k over the two outcomes $Y_k = 1$, spike at time bin k, and $Y_k = 0$, stay quiescent at time bin k. We obtain

$$\begin{aligned}
\mathrm{grad}_k &= \sum_{\mathbf{Y}^{\backslash k}} P_{w.}(\mathbf{Y}^{\backslash k}) \sum_{Y_k} \log P_\mathbf{w}(z|\tilde{\mathbf{Y}}) \, \tfrac{\partial}{\partial w} P_{w.}(Y_k) \\
&= \sum_{\mathbf{Y}^{\backslash k}} P_{w.}(\mathbf{Y}^{\backslash k}) \Big[\log P_\mathbf{w}(z|\tilde{\mathbf{Y}} \cup \{t_k\}] \tfrac{\partial}{\partial w} P_{w.}(Y_k = 1) \\
&\qquad\qquad + \log P_\mathbf{w}(z|\tilde{\mathbf{Y}} \setminus \{t_k\}) \tfrac{\partial}{\partial w} P_{w.}(Y_k = 0) \Big] \\
&= \sum_{\mathbf{Y}^{\backslash k}} P_{w.}(\mathbf{Y}^{\backslash k}) \Big[\log P_\mathbf{w}(z|\tilde{\mathbf{Y}} \cup \{t_k\}] \tfrac{\partial}{\partial w} P_{w.}(Y_k = 1) \\
&\qquad\qquad - \log P_\mathbf{w}(z|\tilde{\mathbf{Y}} \setminus \{t_k\}) \tfrac{\partial}{\partial w} P_{w.}(Y_k = 1) \Big] \, ,
\end{aligned}$$

where the last line follows from the identity $\frac{\partial}{\partial w} P_{w.}(Y_k = 1) = -\frac{\partial}{\partial w} P_{w.}(Y_k = 0)$. We introduce the notation

$$\gamma_{\tilde{\mathbf{Y}}}(t_k) = \log P_\mathbf{w}(z|\tilde{\mathbf{Y}} \cup \{t_k\}) - \log P_\mathbf{w}(z|\tilde{\mathbf{Y}} \setminus \{t_k\}). \tag{S9}$$

The function $\gamma_{\tilde{\mathbf{Y}}}(t_k)$ quantifies the impact that the initiation of a NMDA spike at t_k would lean on the somatic output. Since Y_k is a binary variable, the identity $\frac{\partial}{\partial w} P_{w.}(Y_k =$

$1) = (2\,Y_k - 1)\,\frac{\partial}{\partial w} P_{w.}(Y_k)$ holds independently of the value of Y_k. We deduce

$$\begin{aligned}
\mathrm{grad}_k &= \sum_{\mathbf{Y}^{\setminus k}} P_{w.}(\mathbf{Y}^{\setminus k})\gamma_{\tilde{\mathbf{Y}}}(t)\,(2Y_k - 1)\,\tfrac{\partial}{\partial w} P_{w.}(Y_k) \\
&= \sum_{\mathbf{Y}^{\setminus k}} P_{w.}(\mathbf{Y}^{\setminus k})\gamma_{\tilde{\mathbf{Y}}}(t)\,\frac{1}{2}\sum_{Y_k}(2Y_k - 1)\,\tfrac{\partial}{\partial w} P_{w.}(Y_k) \\
&= \sum_{\mathbf{Y}} P_{w.}(\mathbf{Y})\frac{\gamma_{\tilde{\mathbf{Y}}}(t)}{2}(2Y_k - 1)\,\tfrac{\partial}{\partial w}\log P_{w.}(Y_k),
\end{aligned}$$

and

$$\sum_{k=1}^{K} \mathrm{grad}_k = \sum_{\mathbf{Y}} P_{w.}(\mathbf{Y}) \sum_{k=1}^{K} \frac{\gamma_{\tilde{\mathbf{Y}}}(t)}{2}(2Y_k - 1)\,\tfrac{\partial}{\partial w}\log P_{w.}(Y_k).$$

From the definition (S8), we have

$$\begin{aligned}
\tfrac{\partial}{\partial w}\log P_{w.}(Y_k = 1) &= \frac{\frac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}}\rho^{\mathrm{d}}(t_k)}{\rho^{\mathrm{d}}(t_k)}\mathrm{PSP}(t_k) + \mathcal{O}(\delta) \\
\tfrac{\partial}{\partial w}\log P_{w.}(Y_k = 0) &= -\delta\,\tfrac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}}\rho^{\mathrm{d}}(t_k)\mathrm{PSP}(t_k).
\end{aligned}$$

So, taking the limit $\delta \to 0^+$ and the original notation (the calculation is for the i-th synapse located in the d-th dendritic branch, see Eq. S7), we obtain

$$c'_d(w, z) = \int \mathrm{d}y_d\, P_{w.d}(y_d|\mathbf{x}) \int_0^T \mathrm{d}t\, \underbrace{\left[\tfrac{1}{2}\gamma_{y_d}(t)\tfrac{\mathrm{d}}{\mathrm{d}u_d^{\mathrm{d}}}\log\rho_d^{\mathrm{d}}(t)\left(y_d(t) + \rho_d^{\mathrm{d}}(t)\right)\,\mathrm{PSP}_i(t)\right]}_{e_{di}^{\mathrm{SL}}(t):=}, \qquad \text{(S10)}$$

where $y_d(t)$ denotes the δ-function representation of the set y_d, $y_d(t) = \sum_{s\in y_d}\delta(t-s)$. We obtain a gradient estimate where hidden variables are partially averaged. In particular, the second term of the RHS in the equation (S5) is (see Eq. S7)

$$\left\langle \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\,\tfrac{\partial}{\partial w_{di}}\log P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})\right\rangle = \left\langle \int_0^T \mathrm{d}t\, e_{di}^{\mathrm{SL}}(t)\right\rangle$$

and it follows that

$$\tfrac{\partial}{\partial w_{di}}\mathcal{C}(\mathbf{w}) = \left\langle \int_0^T \mathrm{d}t\, \left(\beta_{\mathrm{s}}\,\alpha\, e_{di}^{\mathrm{ss}}(t) + e_{di}^{\mathrm{SL}}(t)\right)\right\rangle. \qquad \text{(S11)}$$

This term in the brackets is our unbiased gradient estimate for the cost function $\mathcal{C}(\mathbf{w})$.

S1C.2 Online version of the gradient estimate

Here we show how an approximated version of the gradient estimate (S11) can be computed online. The central idea is to rewrite the exact gradient estimate (Eq. (S11)) with integrals that could then be implemented by low-pass filtered version. We therefore replace the rectangular integration window in the Eq. S11 by an exponential one. First, we introduce the inactivation function $\Psi_{y_d}(t)$ that depends on the dendritic spike timings y_d and that is 0 during an ongoing NMDA-spike and 1 elsewhere. Note that $\Psi_{y_d}(t)$ is related to the NMDA-spikes response function $\textsc{nmda}_d(t)$ via $\textsc{nmda}_d(t) = a\,(1 - \Psi_{y_d}(t))$. As computed in [10], the function $\gamma_{y_d}(t)$ (Eq. S9) is given by

$$\gamma_{y_d}(t) = a\,\alpha\,\beta_{\mathrm{s}} \int_{t}^{\min(T,t+\Delta)} \mathrm{d}s\, \Psi_{y_d^{\setminus t}}(s) \left(z(s) - \rho^{\mathrm{s}}_{\setminus d}(s)\right)$$

where $y_d^{\setminus t}$ is the set y_d with no spike timing at t ($y_d^{\setminus t} = \{s \in y_d | s \neq t\}$). In its current form, it is impossible to compute $e^{\mathrm{SL}}_{di}(t)$ (Eq. S10) online, since the integration of $\gamma_{y_d}(t)$ extends from the current time t into the future up to $t + \Delta$. We therefore permute the integration order to turn the integration into the future to an integration across the past (see Appendix),

$$\begin{aligned}\int_0^T dt\, e^{\mathrm{SL}}_{di}(t) &= \int_0^T \mathrm{d}t\, \tfrac{\mathrm{d}}{\mathrm{d}u_d^{\mathrm{d}}} \log \rho_d^{\mathrm{d}}(t)\, \left(y_d(t) + \rho_d^{\mathrm{d}}(t)\right)\, \mathrm{PSP}_i(t) \int_t^{\min(T,t+\Delta)} \mathrm{d}s\, \Psi_{y_d^{\setminus t}}(s)\, f_d(s) \\ &= \int_0^T \mathrm{d}s\, f_d(s) \underbrace{\int_{\max(0,s-\Delta)}^{s} \mathrm{d}t\, \Psi_{y_d^{\setminus t}}(s) \tfrac{\mathrm{d}}{\mathrm{d}u_d^{\mathrm{d}}} \log \rho_d^{\mathrm{d}}(t)\, \left(y_d(t) + \rho_d^{\mathrm{d}}(t)\right)\, \mathrm{PSP}_i(t)\,,}_{\xi^{\mathrm{N}}_{di}(s):=}\end{aligned} \qquad \text{(S12)}$$

with

$$f_d(t) = \tfrac{a\,\alpha\,\beta_{\mathrm{s}}}{2} \left(z(s) - \rho^{\mathrm{s}}_{\setminus d}(t)\right).$$

Here the stimulus started at 0, therefore the synaptic signal $\textsc{psp}_i(t)$ vanishes for $t < 0$ and we can set $s - \Delta$ instead of $\max(0, s - \Delta)$ as a lower bound for the second integral. Our aim is to encode each integral by a low pass filter (see below). Since the function $\Psi_{y_d^{\setminus t}}(s)$ depends on the variables t and s, the function $\xi^{\mathrm{N}}_{di}(s)$ is generally not computable by an online procedure. In the sequel, we will see that we can drop the dependence with respect to t in the function $\Psi_{y_d^{\setminus t}}(s)$. We start to decompose $\xi^{\mathrm{N}}_{di}(s)$ as follows:

$$\begin{aligned} \xi^{\mathrm{N}}_{di}(s) = & \int\limits_{s-\Delta}^{s} \mathrm{d}t\, \Psi_{y_d^{\setminus t}}(s) \tfrac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \rho^{\mathrm{d}}_d(t)\, \mathrm{PSP}_i(t) \\ & + \int\limits_{s-\Delta}^{s} \mathrm{d}t\, \Psi_{y_d^{\setminus t}}(s)\, y_d(t) \tfrac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \log \rho^{\mathrm{d}}_d(t)\, \mathrm{PSP}_i(t)\,. \end{aligned} \tag{S13}$$

For a given s, the functions $\Psi_{y_d^{\setminus t}}(s)$ and $\Psi_{y_d}(s)$ are equal on the set $[s-\Delta, s]\setminus y_d$. Since y_d is a set of zero measure, we can replace $\Psi_{y_d^{\setminus t}}(s)$ by $\Psi_{y_d}(s)$ in the first integral in Eq. S13. The inactivation function $\Psi_{y_d}(s)$ vanishes if there is an ongoing NMDA spike at time s and so does the first integral in Eq. S13. Otherwise we have $\Psi_{y_d}(s) = 1$ which implies that the second integral in Eq. S13 vanishes since this integral runs only over the different NMDA-spike timings in the interval $[s-\Delta, s]$. We introduced the function $y_d(t)$ as the δ-function representation constructed from the set of individual spike times y_d and hence, if the inactivation function is 1, no NMDA-spike was initiated in $[s-\Delta, s]$. These two observations allow us to rewrite $\xi^{\mathrm{N}}_{di}(s)$ as

$$\xi^{\mathrm{N}}_{di}(s) = \begin{cases} \int_{s-\Delta}^{s} \Psi_{y_d^{\setminus t}}(s)\, \frac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \log \rho^{\mathrm{d}}_d(t)\, y_d(t)\, \textsc{psp}_i(t) dt & \text{if s within a } \textsc{nmda}\text{-spike,} \\ \int_{s-\Delta}^{s} \frac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \rho^{\mathrm{d}}_d(t)\, \textsc{psp}_i(t) dt & \text{else}\,. \end{cases} \tag{S14}$$

In our model, spikes are triggered by point processes, a point event is called a spike timing. When two NMDA-spikes are triggered in a short interval they do not add up in amplitude but instead the second one extends the duration of the first one (see Online Methods). This renders the evaluation of $\Psi_{y_d^{\setminus t}}(s)$ complicated. In order to simplify the calculation, we assume that NMDA-spike timings are sparse. More precisely, we assume that each rectangular NMDA spike is triggered by a unique point event. Therefore, if we assume the presence of a NMDA-spike at time s which was initiated at t^{d}_d then the top integral in (S14) is $\frac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \log \rho^{\mathrm{d}}_d(t^{\mathrm{d}}_d)\, \textsc{psp}_i(t^{\mathrm{d}}_d)$ since the inactivation function $\Psi_{y_d^{\setminus t^{\mathrm{d}}_d}}(s)$ is 1 when the unique point event which causes the current dendritic spike is removed. To summarize, we showed that

$$\xi^{\mathrm{N}}_{di}(s) = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \log \rho^{\mathrm{d}}_d(t^{\mathrm{d}}_d)\, \textsc{psp}_i(t^{\mathrm{d}}_d) & \text{if s within a } \textsc{nmda}\text{-spike triggered at } t^{\mathrm{d}}_d, \\ \int_{s-\Delta}^{s} \frac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \rho^{\mathrm{d}}_d(t)\, \textsc{psp}_i(t) dt & \text{else .} \end{cases} \tag{S15}$$

Note that in the Online Methods we put $\mathrm{Den}_d * \mathrm{PSP}_i(s) = \xi^{\mathrm{N}}_{di}(s)$. Therefore, Eq. (S11) becomes (see Eq. (S12))

$$\begin{aligned}
\tfrac{\partial}{\partial w_{di}}\mathcal{C}(\mathbf{w}) &= \left\langle \int_0^T \mathrm{d}t \left[\tfrac{a\,\alpha\,\beta_{\mathrm{s}}}{2}\left(z(s) - \rho^{\mathrm{s}}_{\backslash d}(t)\right) \mathrm{Den}_d * \mathrm{PSP}_i(t) + \beta_{\mathrm{s}}\,\alpha\, e^{\mathrm{ss}}_{di}(t)\right]\right\rangle \\
&= \left\langle \alpha\,\beta_{\mathrm{s}} \int_0^T \mathrm{d}t \left[\tfrac{a}{2} e^{\mathrm{sds}}_{di}(t)(t) + e^{\mathrm{ss}}_{di}(t)\right]\right\rangle .
\end{aligned}$$

In a stochastic ascent algorithm, the term in brackets defines the synaptic update

$$\Delta w_{di} = \eta \int_0^T \mathrm{d}t \left[\tfrac{a}{2} e^{\mathrm{sds}}_{di}(t)(t) + e^{\mathrm{ss}}_{di}(t)\right],$$

where η denotes the learning rate. We could eliminate the constants α and β_{s} in the plasticity rule since as a multiplicative constant it can be absorbed by the learning rate η. The previous update is roughly equivalent to the sdSP plasticity rule in the Main Text since the low-pass filter $E_{di}(t)$ (Eq. 8 in Online Methods) at time t represents the integration of $\frac{a}{2} e^{\mathrm{sds}}_{di}(t) + \alpha\, e^{\mathrm{ss}}_{di}(t)$ from the past to t with an exponential window (the time constant is τ_{E}).

S1C.3 Derivation of the gradient-based reinforcement learning rule (R-sdSP)

Here we show that the rule R-sdSP approximates an online estimate of the gradient of the expected reward

$$\bar{R}(\mathbf{w}) = \int \mathrm{d}\mathbf{x}\mathrm{d}z\, P_{\mathbf{w}}(z,\mathbf{x})\, R(z,\mathbf{x})$$

We maximize $\bar{R}$ through a stochastic gradient ascent algorithm. We start to marginalize out $\mathbf{y}$

$$\begin{aligned}
\bar{R}(\mathbf{w}) &= \int \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\mathrm{d}z\, P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})\, R(z,\mathbf{x}) \\
&= \int \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\mathrm{d}z\, P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\, P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})\, P(\mathbf{x})\, R(z,\mathbf{x})\,.
\end{aligned}$$

The derivative of $\bar{R}$ with respect to the synaptic weight w_{di} is

$$\tfrac{\partial}{\partial w_{di}}\bar{R}(\mathbf{w}) = \left\langle R(z,\mathbf{x}) \tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(z|\mathbf{x},\mathbf{y})\right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})} + \left\langle R(z,\mathbf{x}) \tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(\mathbf{y}|\mathbf{x})\right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})}. \tag{S16}$$

The first term on the RHS of (S16) is computed in Eq. (S6) and represents the classical reward maximizing rule [9]. As previously introduced [10], we consider an alternative gradient estimator based on the following identity

$$\left\langle R(z,\mathbf{x}) \tfrac{\partial}{\partial w_{di}} \log P_{\mathbf{w}}(\mathbf{y}|\mathbf{x}) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})} = \left\langle R(z,\mathbf{x}) \int\limits_0^T \mathrm{d}t \, e^{\mathrm{R}}_{di}(t) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})}$$

where

$$e^{\mathrm{R}}_{di}(t) = \tanh\!\left(\tfrac{1}{2}\gamma_{y_d}(t)\right) \tfrac{\mathrm{d}}{\mathrm{d}u^{\mathrm{d}}_d} \log \rho^{\mathrm{d}}_d(t) \, \left(y_d(t) + \rho^{\mathrm{d}}_d(t)\right) \, \mathrm{PSP}_i(t).$$

More precisely, we have shown in [10] that $e^{\mathrm{R}}_{di}(t)$ is an appropriate gradient estimator for the second term on the RHS of (S16). Thus, the gradient of the expected reward can be written as

$$\tfrac{\partial}{\partial w_{di}} \bar{R} = \left\langle R(z,\mathbf{x}) \int_0^T \mathrm{d}t \, \left(\alpha \, \beta_{\mathrm{s}} \, e^{\mathrm{ss}}_{di}(t) + e^{\mathrm{R}}_{di}(t)\right) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})},$$

We observe that if we perform a linear approximation of the hyperbolic tangent then we obtain $e^{\mathrm{R}}_{di}(t) \approx e^{\mathrm{SL}}_{di}(t)$. As a result, we can apply the calculation of the previous section (see *Online version of the gradient estimate*), it leads to

$$\begin{aligned} \tfrac{\partial}{\partial w_{di}} \bar{R} &\approx \left\langle R(z,\mathbf{x}) \int_0^T \mathrm{d}t \, \left(e^{\mathrm{SL}}_{di}(t) + \alpha \, \beta_{\mathrm{s}} \, e^{\mathrm{ss}}_{di}(t)\right) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})} \\ &= \left\langle \alpha \, \beta_{\mathrm{s}} R(z,\mathbf{x}) \int_0^T \mathrm{d}t \, \left(\tfrac{a}{2} e^{\mathrm{sds}}_{di}(t)(t) + e^{\mathrm{ss}}_{di}(t)\right) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})}. \end{aligned} \tag{S17}$$

and the update rule

$$\Delta w_{di} = \eta \, (R - R_0) \int_0^T \mathrm{d}t \, \left[\tfrac{a}{2} e^{\mathrm{sds}}_{di}(t)(t) + e^{\mathrm{ss}}_{di}(t)\right],$$

where η is the learning rate and R_0 is a baseline. The constant R_0 can be introduced since the update is roughly a gradient rule [11] and so the following identity holds

$$\begin{aligned} \left\langle R_0 \, \alpha \, \beta_{\mathrm{s}} \int_0^T \mathrm{d}t \, \left(\tfrac{a}{2} e^{\mathrm{sds}}_{di}(t)(t) + e^{\mathrm{ss}}_{di}(t)\right) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})} &\approx R_0 \left\langle \int_0^T \mathrm{d}t \, \left(e^{\mathrm{R}}_{di}(t) + \alpha \, \beta_{\mathrm{s}} \, e^{\mathrm{ss}}_{di}(t)\right) \right\rangle_{P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})} \\ &= R_0 \int \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{y} \mathrm{d}z \, \tfrac{\partial}{\partial w_{di}} P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y}) \\ &= R_0 \tfrac{\partial}{\partial w_{di}} \underbrace{\int \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{y} \mathrm{d}z \, P_{\mathbf{w}}(z,\mathbf{x},\mathbf{y})}_{=1} \\ &= 0. \end{aligned}$$

where the approximation sign accounts for the linear approximation of tanh in the definition of $e^{\mathrm{R}}_{di}(t)$. As observed at the end of the previous section, the integral of the update rule is roughly equal to the value of the low pass-filter $E_{di}(t)$ (Eq. 8 in Online Methods) at time T, yielding the plasticity rule cited in the Main Text.

Overall, we made two approximations: (1) The NMDA spikes were supposed to be sparse in time such that they do not overlap and we could neglect the voltage saturation. With an NMDA spike duration of 50 ms the approximation error is small if the NMDA spike rate is smaller than 20 Hz. (2) The symmetric tanh has been linearized around 0. Importantly, both approximations never change the sign of any component of the true gradient vector. Hence, although after the approximations the learning rule may deviate from the true gradient, it will still be hill climbing. Note that both target functions, the lower bound of the log-likelihood for supervised learning and the expected reward for reinforcement learning, are everywhere continuous (in fact differentiable, but in general not convex), so that learning with these approximations still smoothly maximizes these target functions (locally).

S1D Appendix

Here we detail the steps from the first to the second line in the formula (S12).

We rewrite the equation (S12) in a form that allows the permutation of the integration order,

$$\int_0^T \mathrm{d}t\, e_{di}^{\mathrm{SL}}(t) = \int_0^T \mathrm{d}t\, \zeta_{di}(t) \int_0^T \mathrm{d}s\, \chi_{[t,t+\Delta]}(s)\, \Psi_{y_d^{\setminus t}}(s)\, f_d(s),$$

with

$$\zeta_{di}(t) = \tfrac{\mathrm{d}}{\mathrm{d}u_d^{\mathrm{d}}} \rho_d^{\mathrm{d}}(t)\, \left(y_d(t) + \rho_d^{\mathrm{d}}(t)\right)\, \mathrm{PSP}_i(t),$$

and where $\chi_{[t,t+\Delta]}(s)$ denotes the indicator function

$$\chi_{[t,t+\Delta]}(s) = \begin{cases} 1 & \text{if } s \in [t, t+\Delta] \\ 0 & \text{else} \end{cases}.$$

Now we change the integration order

$$e_{di}^{\mathrm{Caus}}(z, \mathbf{y}, \mathbf{x}) = \int_0^T \mathrm{d}s\, f_d(s) \int_0^T \mathrm{d}t\, \chi_{[t,t+\Delta]}(s)\, \Psi_{y_d^{\setminus t}}(s)\, \zeta_{di}(t).$$

The inequalities $s \leqslant t \leqslant s+\Delta$ is equivalent to $t-\Delta \leqslant s \leqslant t$, therefore the two functions $\chi_{[t,t+\Delta]}(s)$ and $\chi_{[s-\Delta,s]}(t)$ too. As initially, the action of $\chi_{[s-\Delta,s]}(t)$ is equivalent to change the boundaries of the second integral, we deduce

$$\begin{aligned}\int_0^T \mathrm{d}t\, e_{di}^{\mathrm{SL}}(t) &= \int_0^T \mathrm{d}s\, f_d(s) \int_0^T \mathrm{d}t\, \chi_{[s-\Delta,s]}(t)\, \Psi_{y_d^{\backslash t}}(s)\, \zeta_{di}(t) \\ &= \int_0^T \mathrm{d}s\, f_d(s) \int_{\max(0,s-\Delta)}^{s} \mathrm{d}t\ \Psi_{y_d^{\backslash t}}(s)\, \zeta_{di}(t).\end{aligned}$$

References

[1] G. Major, A. Polsky, W. Denk, J. Schiller, and D. W. Tank. Spatiotemporally graded NMDA spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. *J. Neurophysiol.*, 99:2584–2601, May 2008.

[2] H. Sanders, M. Berends, G. Major, M. S. Goldman, and J. E. Lisman. NMDA and GABAB (KIR) conductances: the "perfect couple" for bistability. *J. Neurosci.*, 33(2):424–429, Jan 2013.

[3] G. Major, M. E. Larkum, and J. Schiller. Active properties of neocortical pyramidal neuron dendrites. *Annu. Rev. Neurosci.*, 36:1–24, Jul 2013.

[4] B. F. Behabadi and B. W. Mel. Mechanisms underlying subunit independence in pyramidal neuron dendrites. *Proc. Natl. Acad. Sci. U.S.A.*, 111(1):498–503, Jan 2014.

[5] F. Gabbiani, J. Midtgaard, and T. Knopfel. Synaptic integration in a model of cerebellar granule cells. *J. Neurophysiol.*, 72(2):999–1009, Aug 1994.

[6] P. Rhodes. The properties and implications of NMDA spikes in neocortical pyramidal cells. *J. Neurosci.*, 26(25):6704–6715, Jun 2006.

[7] E. Izhikevich. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17:2443–2452, 2007.

[8] N. Frémaux, H. Sprekeler, and W. Gerstner. Functional requirements for reward-modulated spike-timing-dependent plasticity. *J. Neurosci.*, 30:13326–13337, Oct 2010.

[9] J. Pfister, T. Toyoizumi, D. Barber, and W. Gerstner. Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Computation*, 18:1318–1348, 2006.

[10] M. E. Schiess, R. Urbanczik, and W. Senn. Gradient estimation in dendritic reinforcement learning. *The Journal of Mathematical Neuroscience*, 2(2), 2012.

[11] R. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.