Development/Plasticity/Repair

Branch-Specific Plasticity Enables Self-Organization of Nonlinear Computation in Single Neurons

Robert Legenstein and Wolfgang Maass

Institute for Theoretical Computer Science, Graz University of Technology, 8010 Graz, Austria

It has been conjectured that nonlinear processing in dendritic branches endows individual neurons with the capability to perform complex computational operations that are needed to solve for example the binding problem. However, it is not clear how single neurons could acquire such functionality in a self-organized manner, because most theoretical studies of synaptic plasticity and learning concentrate on neuron models without nonlinear dendritic properties. In the meantime, a complex picture of information processing with dendritic spikes and a variety of plasticity mechanisms in single neurons has emerged from experiments. In particular, new experimental data on dendritic branch strength potentiation in rat hippocampus have not yet been incorporated into such models. In this article, we investigate how experimentally observed plasticity mechanisms, such as depolarization-dependent spike-timing-dependent plasticity and branch-strength potentiation, could be integrated to self-organize nonlinear neural computations with dendritic spikes. We provide a mathematical proof that, in a simplified setup, these plasticity mechanisms induce a competition between dendritic branches, a novel concept in the analysis of single neuron adaptivity. We show via computer simulations that such dendritic competition enables a single neuron to become member of several neuronal ensembles and to acquire nonlinear computational capabilities, such as the capability to bind multiple input features. Hence, our results suggest that nonlinear neural computation may self-organize in single neurons through the interaction of local synaptic and dendritic plasticity mechanisms.

Introduction

Dendritic processing in pyramidal neurons is highly nonlinear, exhibiting several types of dendritic spikes (Losonczy and Magee, 2006; Sjöström et al., 2008; Larkum et al., 2009). It has been hypothesized that such dendritic nonlinearities enhance processing capabilities at the single-neuron level (Mel, 1994; Häusser and Mel, 2003) that could be important for the binding of object features, for the participation of single neurons in several ensembles, for various memory processes (Morita, 2008; Wu and Mel, 2009), and for sequence detection (Branco et al., 2010). However, the computational advantage of nonlinear branches for the organism is unclear, because similar functional properties could be achieved by networks of neurons without dendritic nonlinearities. In this article, we investigate one striking advantage of nonlinear branch processing, namely that it enables single neurons to acquire nonlinear functionality with local plasticity mechanisms. Thus, the independence of any one dendritic branch results in the entire dendritic tree behaving like a network, and critically, any one branch is able to optimize responsiveness to a specific input

through plasticity mechanisms that take aspects of the behavior of the whole dendritic arbor into account.

A rich and diverse set of plasticity mechanisms has been identified in the pyramidal neuron (Sjöström et al., 2008). Spike-timing-dependent plasticity (STDP) depends on the precise timing of presynaptic and postsynaptic spikes (Markram et al., 1997; Bi and Poo, 1998). However, pairing frequency and local dendritic depolarization also influence synaptic plasticity (Artola et al., 1990; Ngezahayo et al., 2000; Sjöström et al., 2001). Recent experimental data (Losonczy et al., 2008) also show that the coupling between dendritic branches and the soma (via dendritic spikes) is adapted through a mechanism called "branch-strength potentiation" (BSP), and evidence exists that BSP is relevant *in vivo* (Makara et al., 2009). Finally, synaptic and dendritic plasticity also depend on neuromodulatory input (Reynolds and Wickens, 2002; Gu, 2003; Dringenberg et al., 2007; Losonczy et al., 2008).

It is unclear how these plasticity mechanisms interact to enable robust learning and self-organization in neurons and networks. In this article, we show that synaptic plasticity mechanisms that depend on membrane depolarization enable a neuron model with nonlinear dendritic integration to acquire complex functionality in a self-organized manner. These self-organization capabilities are made possible by an emerging competition between branches for synaptic activation. Specifically, we show in a theoretical analysis of a simplified model that the plasticity mechanisms give rise to a max-operation such that only the most strongly activated dendritic branches specialize to a given input pattern. Thus, learning in our model induces a weight distribution in which stored patterns create clusters of strong

synapses at individual dendritic branches. A neuron can in this way self-organize to solve a binding problem. This self-organization is possible if the presence of object features is coded through increased firing rates of neuronal ensembles or through increased synchrony in such ensembles. BSP can in this context serve as a mechanism for boosting dendritic responses to stored patterns, as suggested by Losonczy et al. (2008), and homeostatically regulate the effectivity of dendritic spikes.

Materials and Methods

Specific implementation of the neuron model for computer simulations
In this article, we use a stochastic spiking neuron model with dendritic nonlinearities as illustrated in Figure 1.

Nonlinear summation of EPSPs at dendritic branches. The local potential $b_k(t)$ at a branch k at time t in our model is constituted by two components, a passive component $p_k(t)$ and an active component $a_k(t)$. The passive component $p_k(t)$ is given by the linear summation of incoming postsynaptic potentials:

$$p_k(t) = \sum_j w_{kj} \sum_{t_{kj}^{(f)} \epsilon x_{kj}} \varepsilon(t - t_{kj}^{(f)}),$$

(1)

where x_{kj} is the set of spike times of input j to branch k, and ε is the EPSP that results from one input spike. This means that all EPSPs arriving at a dendritic branch are added up linearly and propagated passively. w_{kj} is the synaptic efficacy of a synapse from input j to branch k. The postsynaptic potential ε is modeled by an α function (for parameters, see Table 2). In addition to the passive component $p_k(t)$, the membrane potential $b_k(t)$ at branch k includes an active nonlinear component $a_k(t)$ of size β_0, which is activated whenever $p_k(t)$ is above a threshold φ^{dend}:

$$a_k(t) = \beta_0 \Theta(p_k(t) - \varphi^{\text{dend}}),$$

(2)

where Θ denotes the Heaviside step function. The branch potential $b_k(t)$ is given by the sum of the local synaptic activation $p_k(t)$ at the branch and the additional contribution from the dendritic spike $a_k(t)$, leading to nonlinear branch characteristics as illustrated in Figure 2:

$$b_k(t) = p_k(t) + a_k(t).$$

(3)

The passive component of the branch potential $p_k(t)$ is conducted to the soma, and its impact on the somatic potential is scaled by an attenuation factor $u^{\text{passive}} < 1$. The dendritic spike is scaled by a different weighting factor, the branch strength u_k, before it reaches the soma. According to recent experimental findings, this branch strength can vary dramatically between branches and is subject to plasticity (Losonczy et al., 2008) (for the plasticity model used in this study, see below). Thus, the somatic membrane potential before refractory effects are taken into account is given by the sum of these weighted components from all branches k:

$$v(t) = V_{\text{rest}} + \sum_{k=1}^{K} u^{\text{passive}} p_k(t) + u_k a_k(t),$$

(4)

where V_{rest} denotes the resting potential of the neuron.

Model for somatic response. The somatic spiking mechanism and reset behavior of our model is based on the model of Jolivet et al. (2006). To obtain the somatic membrane voltage V_m of the neuron, relative refrac-
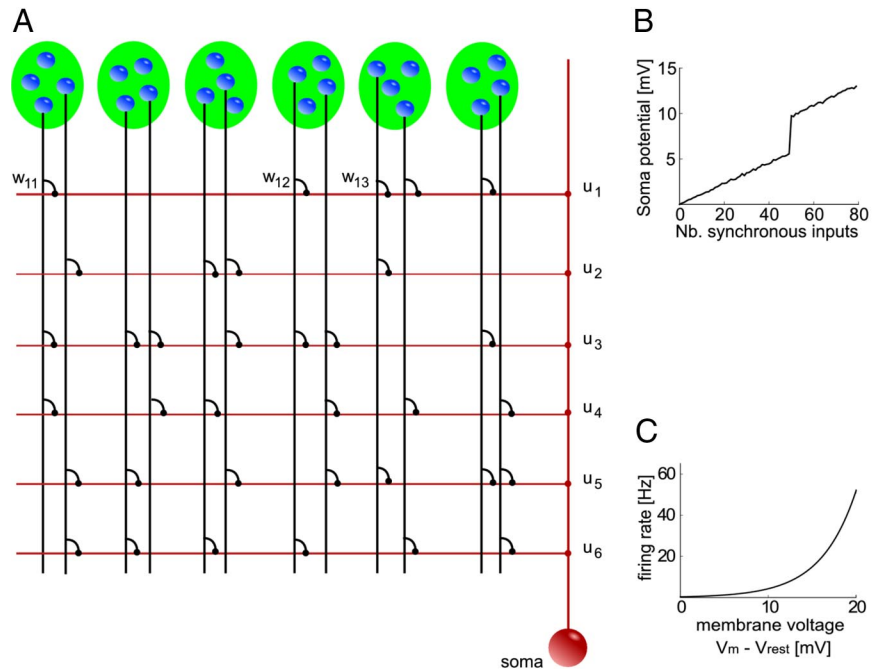

Figure 1. Schema of the neuron model with dendritic branches. **A**, The neuron is schematically illustrated in red. It receives input from presynaptic neurons (blue spheres), which participate in several neuronal ensembles (indicated by green ellipses). The neuron is composed of a set of independent branches (horizontal red lines) and a soma (red sphere). Axons of presynaptic neurons (vertical black lines) connect to the branches via synapses (black dots). Synaptic inputs are weighted by synaptic weights w_{kj}. Dendritic spikes are weighted by branch strengths u_k (red dots). The soma sums the dendritic activations and the weighted dendritic spikes. **B**, Dendritic nonlinearity. At a certain local dendritic depolarization, a dendritic spike is initiated, leading to nonlinear dendritic integration. Shown is the modeled somatic membrane voltage in dependence of the number of synchronously (in a time window of 10 ms) activated synapses. Compare with Losonczy and Magee (2006, their Fig. 1G). **C**, Somatic nonlinearity. The neuron fires stochastically with instantaneous firing rate $\rho(t)$ that depends exponentially on the somatic membrane potential $V_m(t)$.

toriness is included by adding a reset kernel η_{reset} to the membrane voltage for each action potential. This reset kernel effectively adds a large negative contribution to the membrane voltage for times after the postsynaptic spike, which then decays exponentially over time. This can be described mathematically as follows:

$$V_m(t) = v(t) + \sum_{t^{(f)} \epsilon y t} \eta_{\text{reset}}(t - t^{(f)}),$$

(5)

where y_t denotes the set of output spikes of the neuron up to time t. The instantaneous firing rate of the neuron at time t is given by an exponential function of the membrane voltage:

$$\rho(t) = f_{\text{out}}(V_m(t)) = \rho_o \exp\left(\frac{V_m(t) - \theta}{\Delta U}\right),$$

(6)

where θ is the stochastic threshold, ΔU is the width of the spike trigger zone, and ρ_0 is the instantaneous rate at threshold. In other words, the firing rate of the neuron depends exponentially on its membrane potential (Fig. 1C). If the membrane potential is clamped to 20 mV above resting potential, this results in a mean firing rate of 52 Hz, a clamping at 25 mV already results in a mean firing rate of 100 Hz. Because the firing rate of the model depends only on the difference between the membrane voltage V_m and the stochastic firing threshold θ, we can arbitrarily set the resting potential to $V_{\text{rest}} = 0$ mV and the threshold to $\theta = 20$ mV (for other parameter values, see Table 2). Output spikes are generated by a Poisson process with this instantaneous firing rate. After each spike, the neuron enters an absolute refractory period of duration t_{refract} during which no spike can be emitted. This firing mechanism implements a stochastic threshold that has been shown to approximate well the behavior of layer 5 pyramidal neurons of somatosensory cortex of male and female rats with the parameters used in this article (Jolivet et al., 2006).

Simplified model for theoretical analysis. The model described above and further elaborated in Results includes a step-like nonlinearity in
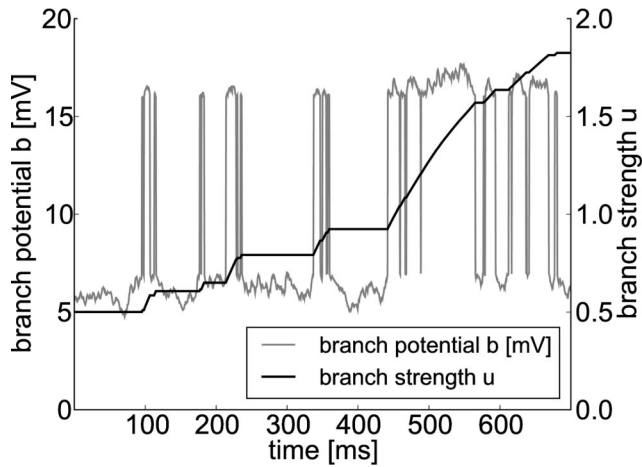
Figure 2. Nonlinear integration of synaptic input at dendritic branches and branch strength potentiation. The local potential at a dendritic branch k (gray line) is given by the sum of a passive component $p_k(t)$ and an active component $a_k(t)$. The value of the passive component is given by the linear sum of incoming EPSPs. Summation of EPSPs from many synapses that receive Poissonian input from active cell ensembles leads to temporal fluctuations of the potential. The active component models dendritic spikes and is nonzero whenever $p_k(t)$ is above a threshold of 7 mV. Such nonlinear events are visible as sharp increases in the dendritic potential. The temporal evolution of the corresponding branch strength u_k is plotted in black. Dendritic spikes at the branch induce potentiation of its branch strength.

dendritic branches. This model was used in our simulations; it is, however, not suited for mathematical analysis. In a simplified model, the nonlinear dependence of the output $b_k(t)$ of branch k is modeled by an exponential branch nonlinearity,

$$ b_k(t) = \beta_0 \exp\left(\frac{p_k(t) - \theta_b}{\Delta\beta}\right), \qquad (7) $$

and the branch outputs are linearly summed at the somatic site,

$$ V_m(t) = \sum_{k=1}^{K} u_K b_k(t), \qquad (8) $$

where u_k now denotes the branch strength of branch k. In other words, in the simplified model, the branch voltage depends exponentially on the summed EPSPs. The membrane voltage at the soma is then given by the sum of all branch potentials. Reasonable firing behavior of this simplified model is achieved with parameters $\beta_0 = 25\,\text{mV}$, $\theta_b = 20\,\text{mV}$, $\Delta\beta = 3\,\text{mV}$. The advantage of this simplified model is that it allows a thorough analysis of the weight dynamics.

Plasticity mechanisms
STDP model. We adopt a standard model for STDP (Abbott and Nelson, 2000) in which the change Δw of the synaptic weight depends on the firing time t^{pre} of the presynaptic neuron and the firing time $t^{\text{post}} = t^{\text{pre}} + \Delta t$ of the postsynaptic neuron (Fig. 3). If the presynaptic action potential precedes the postsynaptic action potential in a time window of a few tens of milliseconds ($\Delta t > 0$), long-term potentiation (LTP) is induced. A reversed sequence of presynaptic and postsynaptic action potentials leads to long-term depression (LTD).

Recent experiments showed that STDP also depends on the depolarization level at the local dendritic site (Nevian and Sakmann, 2006; Sjöström and Häusser, 2006; Larkum and Nevian, 2008). More specifically, these findings suggest that LTP is preferentially induced at large local depolarizations. For example, it was shown by Sjöström and Häusser (2006) that LTD at distal synaptic inputs can be converted to LTP by either stronger stimulation that recruits more inputs (cooperativity) or strong extracellular input. In the same direction, Sjöström et al. (2001) showed that low-frequency pre-before-post pairings lead to LTP only for sufficiently strong EPSPs. Again, LTP could be recovered in their
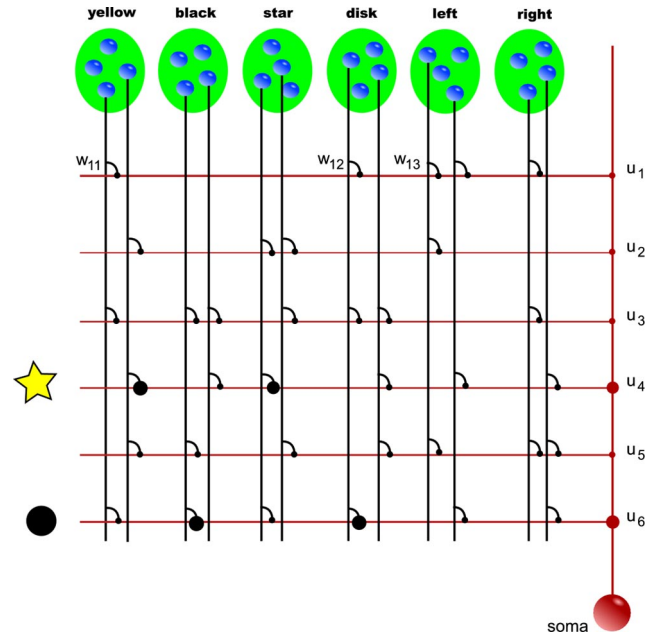

Figure 3. Principle of feature binding in the neuron model. The neuron and presynaptic neuronal ensembles are indicated as in Figure 1. Neuronal ensembles code specific object features, such as color, shape, or movement direction. Neurons of the ensembles yellow and star have strong synaptic connections (the size of black dots indicates the strength of the synaptic connection w_{kj}) to branch 4 of the neurons. Furthermore, strong synapses connect neurons from ensembles black and disk with branch 6. Thus, branch 4 is strongly depolarized by feature combination yellow star and branch 6 by feature combination black disk (indicated by symbols left to the branch). Dendritic spikes induced by this depolarization have a strong impact on the soma as a result of large branch strengths of branches 4 and 6 (the size of red dots indicates the branch strength value u_k). Somatic activation is less strong for a black star or a yellow disk because they partly activate branches 4 and 6.

experiments by strong extracellular stimulation in addition to unitary EPSPs. Interestingly, this dependence on EPSP size was absent for LTD induced by post-before-pre pairings. Therefore, the model is extended by an induction threshold for LTP such that LTP is induced only if the local branch potential $b_k(t)$ exceeds a threshold φ_+. Furthermore, Sjöström et al. (2001) showed that LTP depends approximately linearly on pairing frequency, whereas there is no such dependency of LTD below 40 Hz. In this study, the presynaptic and postsynaptic firing rates were always covaried. The relative contribution of each of these two factors is therefore unclear. We modeled frequency effects by a linear dependency of LTP on the estimated presynaptic firing rate $\hat{\rho}_{kj}(t)$:

$$ \Delta w_{kj}(t, \Delta t) = \begin{cases} A_+ \cdot \dfrac{\hat{\rho}_{kj}(t)}{\rho_o^{\text{LTP}}} \cdot e^{-\Delta t/\tau_+} \Theta(b_k(t) - \varphi_+), & \text{if } \Delta t > 0 \\ -A_- \cdot e^{\Delta t/\tau_-}, & \text{if } \Delta t \leq 0, \end{cases} $$
$$(9)$$

where Θ denotes the Heaviside step function, and ρ_0^{LTP} defines the linear scaling of the frequency dependence of LTP. A_+ (A_-) and τ_+ (τ_-) define the amount of weight change and the width of the exponential learning windows for LTP and (LTD), respectively. Summarizing, pre-before-post spike pairs induce LTP if the branch is sufficiently depolarized. The amount of LTP depends linearly on the presynaptic firing rate and exponentially on the time difference between the presynaptic and the postsynaptic spike. A presynaptic spike that follows a postsynaptic action potential leads to LTD (Fig. 3). Weights were clipped at 0 and a maximum value w_{max}. In the simulations, the presynaptic firing rate ρ_{kj} was estimated by a running average over inverse presynaptic interspike intervals and updated at presynaptic spike times $t_{kj}^{(2)}, t_{kj}^{(3)}, t_{kj}^{(4)}, \ldots$ as

$$ \hat{\rho}_{kj}(t_{kj}^{(f)}) = 0.98\hat{\rho}_{kj}(t_{kj}^{(f-1)}) + 0.02 \max\{5\text{ms}, t_{kj}^{(f)} - t_{kj}^{(f-1)}\}^{-1} $$
$$(10)$$

with initial value $\hat{\rho}_{kj}(t_{kj}^{(1)}) = 10$ Hz. The maximum operator in the denominator was included to avoid temporary large estimates. For times t between presynaptic spikes, $\hat{\rho}_{kj}(t)$ was set to its value at the most recent presynaptic spike time. This means that effectively, $\hat{\rho}_{kj}$ estimates the presynaptic firing rate by averaging over recent inverse interspike intervals.

It is well known that pairing of presynaptic spikes with postsynaptic depolarization is often sufficient for the induction of LTP, i.e., a postsynaptic somatic spike is not necessary (Kelso et al., 1986; Gustafson et al., 1987; Hardie and Spruston, 2009). We model LTP in the absence of postsynaptic somatic spiking by the following weight dynamics:

$$\frac{d}{dt} w_{kj}(t) = \eta \, \frac{\hat{\rho}_{kj}}{\rho_0^{\text{LTP}}} \, (b_k(t) + \kappa) PSP_{kj}(t), \qquad (11)$$

where η is a small learning rate, $\hat{\rho}_{kj}$ is an estimate of the presynaptic firing rate as before, and ρ_0^{LTP} defines the linear scaling of the frequency dependence of LTP. $PSP_{kj}(t)$ is a low-pass filtered version of the presynaptic spike train (i.e., $PSP_{kj}(t) = \Sigma_f \varepsilon(t - t_{kj}^{(f)})$ where $t_{kj}^{(f)}$ denotes the fth spike time of the presynaptic neurons). The constant κ increases learning speed if branch activations are very small initially but is not crucial for the results. In other words, each presynaptic spike potentiates the synapse by an amount that depends linearly on the presynaptic firing rate and the local branch potential. Because the branch potential $b_k(t)$ is the sum of the local synaptic activation $p_k(t)$ and the additional contribution from the dendritic spike $a_k(t)$, synapses can be potentiated in the absence of dendritic spikes, although potentiation in the presences of a dendritic spike is substantially stronger. Some experimental data indicate that this form of plasticity needs dendritic spikes (Golding et al., 2002). Conversely, potentiation without dendritic spikes may be more pronounced in the presence of an increased concentration of ACh (Bröcher et al., 1992; Auerbach and Segal, 1994; Dringenberg et al., 2007). Also, a fairly large number of experimental data (Watt and Desai, 2010) support the existence of mechanisms for the so-called synaptic scaling, i.e., mechanisms that potentiate synapses of neurons whose average firing rate is very low, thereby bringing them back into a normal firing regimen. A mechanism of this type is also needed in the context of our model, and therefore a weak potentiation of synapses without postsynaptic firing or dendritic spikes is included in Equation 11.

As noted above, the contribution of the presynaptic rate in the reported influence of pairing frequency on STDP is still unclear. We therefore performed control simulations in which LTP did not depend on the presynaptic frequency, i.e., in which weight updates are performed without the factor $\dfrac{\hat{\rho}_{kj}}{\rho_0^{\text{LTP}}}$ in the learning rules 9 and 11. Removing this factor changes the learning speed as one side effect. To compensate, we adjusted the learning rates (see below, Simulation parameters). We report in Results that simulations with this reduced version of LTP leads to results quantitatively very similar to those obtained with the rate-dependent version of LTP.

Branch-strength potentiation. Recent experimental data from hippocampal CA1 pyramidal neurons of male rats (Losonczy et al., 2008; Makara et al., 2009) shows that the impact of a dendritic spike at branch k in some dendritic structures on the somatic membrane potential depends on branch strength. We denote the branch strength of a branch k as $u_k(t)$. Furthermore, it was shown that this branch strength is plastic. We model this plastic behavior of branch strength u_k by the following temporal dynamics:

$$\frac{d}{dt} u_k(t) = \eta_{\text{branch}} \Theta(a_k(t))[\varphi_b - (u_k(t)a_k(t) + p_k(t))]. \quad (12)$$

Here, η_{branch} denotes a small learning rate, and φ_b denotes the target membrane voltage at the site of spike potentiation. The term $\Theta(a_k(t))$ indicates that a branch spike has appeared. In other words, a change in the branch strength only appears during branch spikes. The term in square brackets in the dynamics Equation 12 models a saturation of BSP at a given local membrane voltage. According to this term, BSP vanishes when the local potential $(u_k(t)a_k(t) + p_k(t))$ equals a constant φ_b and the branch strength is depressed at a higher local potential. We note that only

Table 1. Overview of the plasticity model

Synaptic: STDP

Pre–post timing	Dendritic depolarization	Presynaptic rate	Effect
Pre before post	Low	*	No effect
Pre before post	High	Linear dependence	Potentiation
Post before pre	*	*	Depression
Pre	Linear dependence	Linear dependence	Potentiation

Branch strength: BSP

Dendritic spike	Dendritic depolarization	Presynaptic rate	Effect
No	*	*	No effect
Yes	Low	*	Potentiation
Yes	High	*	Saturation/depression

We use STDP for synaptic plasticity and BSP for plasticity of branch strengths. An asterisk denotes that the effect does not depend on this variable; "Linear dependence" indicates that the amplitude weight change depends linearly on that variable.

potentiation of branch strengths was observed in the study by Losonczy et al. (2008). Hence, our model for BSP, which leads to depression for excessive branch potentials, is an extension of the experimentally observed effects. The behavior of Equation 12 is illustrated in Figure 2. One can see that the branch strength is potentiated during dendritic spikes and potentiation is weakened as it increases, which illustrates the effect of the saturation term in Equation 12. Alternatively, one can model saturation simply by clipping $u_k(t)$ at some constant upper boundary. This leads to similar results (see Discussion). Although the branch strength is plastic and thus depends on time, we will in the following denote it simply by u_k and skip the dependency on time for simplicity.

For additional discussion of the branch-strength dynamics, see Results. A summary of the plasticity rules in the model is given in Table 1.

Adaptive learning rate. In the intact brain, synaptic plasticity itself is regulated by activity-dependent mechanisms (Abraham, 2008). In particular, it is conjectured that the molecular machinery that implements synaptic plasticity provides a progressive stabilization of changes in the synapse during the transition from short-term to intermediate-term to long-term memory storage (Kandel, 2009). Because of the lack of quantitative experimental data on this mechanism, we model such progressive stabilization of synaptic weights by introducing an adaptive learning rate η_{kj} for each synaptic connection, which starts with initial value 1, and is multiplied by a constant $\eta_{\text{stabilize}} < 1$ each time when the branch potential b_k is above a threshold $\varphi_{\text{stabilize}}$ during a postsynaptic spike. The adaptive learning rate η_{kj} is set to 0 when it falls below 2% of its initial value. This is a quite intuitive mechanism that tends to stabilize synaptic weights at branches that are strongly activated by synaptic input, indicating that an important synaptic pattern has emerged at the branch. To our knowledge, this strategy has not been used before in a similar context, although adaptive learning rates are commonly used in artificial neural network training. Previous studies on weight stability under STDP-like synaptic plasticity focused on the retention of memory under random, i.e., unstructured, stimulation or random weight changes (van Rossum et al., 2000; Fusi and Abbott, 2007). In our study, input patterns are always strongly structured with highly correlated input, in terms of either the input rate or temporally precise spike correlations. Under such conditions, memories are in general quickly erased in these types of models. In networks of spiking neurons, lateral inhibition can lead to prolonged retention of weight patterns even during structured ongoing input, indicating that inhibition could act to stabilize synaptic efficacy patterns (Billings and van Rossum, 2009). However, because stabilization mechanisms are not the focus of this work, we only applied the above mentioned strategy of an adaptive learning rate that works fine for our purposes.

Mathematical analysis of neuron adaptation in a rate formulation

Here we analyze plasticity in the simplified model with exponential branch nonlinearity described by Equations 7 and 8. We show that, in a mean-field rate approximation, the branch with the highest initial activation for a given pattern wins the competition and converges to a stable nonzero branch activation, whereas the activations of other branches

decay to 0. In this section, we disregard branch strengths, setting them to 1 throughout the analysis. Furthermore, we will only consider the synapses that are activated by the pattern—those with high presynaptic firing rate—and disregard other synapses for the sake of simplicity. This can be justified by the frequency dependence of LTP. Because of this dependence, those synapses with low presynaptic activity are subject to weak LTP only, and their weights will decay quite fast as the postsynaptic rate increases. As an additional simplifying assumption, we consider weights without bounds in this analysis, i.e., they are neither clipped at a maximal weight value nor at 0. This strongly simplifies the analysis in continuous time.

We first convert the spike-based formulation of our model into a rate-based formulation. Let $t_{ki}^{(f)}$ denote the fth firing time of the presynaptic neuron of synapse i at branch k where $f = 1, 2, \ldots$. We formally denote the corresponding spike train S_{ki} as a sum of Dirac delta functions (Gerstner and Kistler, 2002) $S_{ki}(t) = \Sigma_f \delta(t - t_{ki}^{(f)})$. Denoting the fth firing time of the postsynaptic neuron as $t_{post}^{(f)}$ for $f = 1, 2, \ldots$, we can write the postsynaptic spike train as $S^{post}(t) = \Sigma_f \delta(t - t_{post}^{(f)})$. The local voltage at branch k at time t is given by the following:

$$b_k(t) = \beta_0 \exp\left(\frac{\sum_i \int_0^\infty w_{ki} S_{ki}(t - s)\varepsilon(s)ds - \theta_b}{\Delta\beta}\right). \quad (13)$$

Assuming a large number of inputs, all with the same constant firing rate ρ and uncorrelated spike times, we approximate the summed input potential by its mean field:

$$b_k(t) \approx \beta_0 \exp\left(\frac{\rho \sum_i w_{ki} \int_0^\infty \varepsilon(s)ds - \theta_b}{\Delta\beta}\right)$$

$$= \beta_0 \exp\left(\frac{\rho w_k \int_0^\infty \varepsilon(s)ds - \theta_b}{\Delta\beta}\right) = \beta_0 \exp\left(\frac{\rho w_k - \theta_b}{\Delta\beta}\right), \quad (14)$$

where $w_k = \Sigma_i w_{ki}$ is the summed weight at branch k and, without loss of generality, we assumed $\int_0^\infty \varepsilon(s)ds = 1$. From the plasticity effects implemented in simulations, we include the minimal set that is essential for dendritic competition. This set consists of voltage-dependent LTP for each presynaptic spike as defined in Equation 11 and LTD as defined in Equation 9 for negative Δt. We can thus write the weight change $w_{ki}'(t)$ at time t as

$$w_{ki}'(t) \approx \eta \frac{\rho}{\rho_0^{LTP}} \int_0^\infty S_{ki}(t - s)\varepsilon(s)(b_k(t) + \kappa)ds$$

$$- A_- S_{ki}(t) \int_0^\infty S^{post}(t - s)e^{-\frac{s}{\tau}}ds. \quad (15)$$

The average weight change for the summed weight at branch k—where the average is taken over realizations of presynaptic and postsynaptic spike trains—is then

$$\langle w_k'(t)\rangle \approx \rho^2 \overline{W}_+ (b_k(t) + \kappa) - \rho\rho^{post}(t)\overline{W}_-, \quad (16)$$

with $\overline{W}_+ = N_{syn} \frac{\eta}{\rho_0^{LTP}} \int_0^\infty \varepsilon(s)ds = N_{syn}\frac{\eta}{\rho_0^{LTP}}$ and $\overline{W}_- = N_{syn}A_-\tau_-$. Here, N_{syn} is the number of active synapses to the branch (we assume that N_{syn} is equal for all branches). Hence, we arrived at a rate-based description of weight dynamics for a single pattern presentation.

We assume that, at time $t = 0$, one branch has higher branch activation than any other branch. Let k denote the index of the branch with the largest branch activation at time 0, i.e., $b_k(0) > b_j(0)$ for all $j \neq k$, which implies that $w_k(0) > w_j(0)$ for all $j \neq k$. We now show that, for $\kappa \geq \frac{\overline{W}_- \rho_0}{\rho\overline{W}_+}$, these weight dynamics asymptotically converge to $\mathbf{b}^\star$, where $b_j^\star = 0$ for $j \neq k$ and $b_k^\star > 0$, i.e., $\lim_{t\to\infty} \mathbf{b}(t) = \mathbf{b}^\star$.

To simplify notation, we will in the following not indicate directly the dependence of dynamic variables on time; these are $b_i(t)$, $w_i(t)$, $\rho^{post}(t)$. First, we note that

$$b_i' = \lambda b_i(b_i + \kappa) - b_i\delta\rho^{post}, \quad (17)$$

with $\lambda = \frac{\rho^3 \overline{W}_+}{\Delta\beta}$ and $\delta = \frac{\rho^2 \overline{W}_-}{\Delta\beta}$. We further find that

$$w_k' - w_j' = \rho^2 \overline{W}_+(b_k - b_j). \quad (18)$$

Observation 1. The local potential at branch k is always greater or equal to the local potential at any other branch, i.e., $b_k \geq b_j \forall t > 0, j \neq k$ (this also implies $w_k \geq w_j \forall t > 0, j \neq k$).

Proof (sketch). By contradiction, a change of order implies a previous change of the order of derivatives. This is a contradiction to the fact that the derivatives are proportional to the branch activations, because they are ordered at any time before the change.

Observation 2. The local potential at branch k does never fall below a minimal value $b_k^{min} > 0$. More precisely, there exists a $b_k^{min} > 0$ and a $t^\star > 0$ such that $b_k > b_k^{min}$ for all $t > t^\star$.

Proof. We can bound the weight change at branch k from below:

$$w_k' = \rho^2 \overline{W}_+(b_k + \kappa) - \rho\overline{W}_-\rho_0 \exp\left(\frac{\sum_i b_i - \theta}{\Delta U}\right)$$

$$> \rho^2 \overline{W}_+(b_k + \kappa) - \rho\overline{W}_-\rho_0 \exp\left(\frac{Nb_k - \theta}{\Delta U}\right), \quad (19)$$

where N is the number of branches. Hence, $w_k' > 0$ independent of other branch activations if

$$\rho^2 \overline{W}_+(b_k + \kappa) > \rho\overline{W}_-\rho_0 \exp\left(\frac{Nb_k - \theta}{\Delta U}\right)$$

$$b_k > \frac{\rho_0 \overline{W}_-}{\rho\overline{W}_+} \exp\left(\frac{Nb_k - \theta}{\Delta U}\right) - \kappa$$

$$b_k > \frac{\rho_0 \overline{W}_-}{\rho\overline{W}_+}\left[\exp\left(\frac{Nb_k - \theta}{\Delta U}\right) - 1\right], \quad (20)$$

for $\kappa \geq \frac{\rho_0 \overline{W}_-}{\rho\overline{W}_+} = \frac{\delta}{\lambda}\rho_0$. Inequality 20 holds true for $b_k < \frac{\theta}{N}$. Hence, b_k will, independently of other dynamic variables, converge in finite time to a value above b_k^{min} and stay above b_k^{min}.

Observation 3. The local potential at branch k does never rise above a maximal value $b_k^{max} > 0$. More precisely, there exists a $b_k^{max} > 0$ and a $t^\star > 0$ such that $b_k < b_k^{max}$ for all $t > t^\star$.

Proof. We can bound the weight change at branch k from above:

$$w_k' = \rho^2 \overline{W}_+(b_k + \kappa) - \rho\overline{W}_-\rho_0 \exp\left(\frac{\sum_i b_i - \theta}{\Delta U}\right)$$

$$< \rho^2 \overline{W}_+(b_k + \kappa) - \rho\overline{W}_-\rho_0 \exp\left(\frac{b_k - \theta}{\Delta U}\right). \quad (21)$$

Hence, $w_k' < 0$ independent of other branch activations if

$$\rho^2 \overline{W}_+(b_k + \kappa) < \rho\overline{W}_-\rho_0 \exp\left(\frac{b_k - \theta}{\Delta U}\right)$$

Table 2. Parameters for computer simulations

Neuron parameters		
θ	Neuron stochastic threshold	20 mV
ρ_0	Neuron firing rate at threshold	52 Hz
ΔU	Width of spike trigger zone	4 mV
t_{refract}	Refractory period	2 ms
η_{reset}	Reset kernel	$\eta_{\text{reset}}(s) = -10mVe^{-\frac{s}{20ms}}\Theta(s)$
Synapse parameters		
$\varepsilon(s)$	EPSP kernel	$\varepsilon(s) = 1.3mV[e^{-\frac{s}{20ms}} - e^{-\frac{s}{0.7ms}}]\Theta(s)$
w_{ini}	Initial weight value	Uniformly distributed in $[0.0025, 0.0225]$
w_{max}	Maximal weight value	$0.15[-]$
Branch parameters		
β_0	Dendritic spike amplitude	9 mV
φ^{dend}	Dendritic spike threshold	7 mV
u_{ini}	Initial branch strength	$0.5[-]$
u^{passive}	Passive EPSP decay constant	$0.8[-]$
Plasticity parameters		
η	Learning rate, pre-spike LTP	$0.004\,\text{mV}^{-2}\,\text{s}^{-1}$
κ	Offset for pre-spike LTP	1 mV
A_+	STDP, LTP factor	$0.02[-]$
A_-	STDP, LTD factor	$0.01[-]$
τ_+	STDP, LTP time constant	16.8 ms
τ_-	STDP, LTD time constant	33.7 ms
ρ_0^{LTP}	STDP, rate dependency factor	$40\,\text{s}^{-1}$
φ_+	Threshold for LTP induction	17 mV
η_{branch}	Learning rate for BSP	$0.25\,\text{mV}^{-1}\,\text{s}^{-1}$
φ_b	Target potential for BSP	30 mV
$\varphi_{\text{stabilize}}$	Threshold for adaptive learn rate	16.5 mV
$\eta_{\text{stabilize}}$	Adaptive learn-rate stabilization factor	$0.97[-]$

$$b_k < \frac{\rho_0 \overline{W}_-}{\rho \overline{W}_+} \exp\left(\frac{b_k - \theta}{\Delta U}\right) - \kappa. \quad (22)$$

For $\kappa \geq \frac{\rho_0 \overline{W}_-}{\rho \overline{W}_+} = \frac{\delta}{\lambda}\rho_0$, this condition is fulfilled for all $b_k > b_k^{\text{max}} > 0$. Because the exponential grows much faster than the linear function, such a b_k^{max} can always be found. For reasonable parameters, b_k^{max} is ~25 mV. Hence, b_k will, independently of other dynamic variables, converge in finite time to a value below b_k^{max} and stay below b_k^{max}.

In the following we assume without loss of generality that $b_k^{\text{min}} < b_k(0) < b_k^{\text{max}}$. Let $\Delta_{kj}(t) \equiv w_k(t) - w_j(t)$ denote the difference between the summed weights at branch k and branch j. We show that weights w_j for $j \neq k$ decay at least linearly with time, i.e.,

$$w_j(t) \leq w_k^{\text{max}} - tC_{kj} + \Delta_{kj}(0) \quad \forall\, j \neq k, \quad (23)$$

for $C_{kj} = \rho^2 \overline{W}_+ b_k^{\text{min}}\left(1 - \exp\left(-\frac{\Delta_{kj}(0)}{\Delta\beta}\right)\right)$. We can bound the change in the weight difference $\Delta'_{kj}(t)$ by

$$\Delta'_{kj}(t) = \rho^2 \overline{W}_+ (b_k(t) - b_j(t))$$
$$= \rho^2 \overline{W}_+ b_k(t)\left(1 - \frac{b_j(t)}{b_k(t)}\right)$$
$$= \rho^2 \overline{W}_+ b_k(t)\left(1 - \exp\left(-\frac{\Delta_{kj}(t)}{\Delta\beta}\right)\right)$$
$$\geq \rho^2 \overline{W}_+ b_k^{\text{min}}\left(1 - \exp\left(-\frac{\Delta_{kj}(0)}{\Delta\beta}\right)\right), \quad (24)$$

and the validity of Equation 23 follows. This also shows that $\lim_{t\to\infty} b_j(t) = 0$ for all $j \neq k$. Hence, for all $\varepsilon > 0$ and all $j \neq k$, there exists a $t^\varepsilon > 0$ such that $b_k(t) \geq b_k^{\text{min}}$ and $b_j(t) < \frac{\varepsilon}{N-1}$ for all $t > t^\varepsilon$. Consider the dynamics for b_k:

$$b'_k = \lambda b_k(b_k + \kappa) - b_k \delta\rho_0 \exp\left(\frac{\sum_i b_i - \theta}{\Delta U}\right). \quad (25)$$

This equation has a fixed point at $b_k = 0$, which is not reached because $b_k(t) \geq b_k^{\text{min}} > 0$ for all t and the fixed point $b_k^\star$ given by the solution to:

$$b_k^\star = \frac{\delta}{\lambda}\rho_0 \exp\left(\frac{\sum_i b_i - \theta}{\Delta U}\right) - \kappa. \quad (26)$$

For $\kappa \geq \frac{\delta}{\lambda}\rho_0$, the fixed point $b_k^\star$ defined by Equation 26 is larger or equal to a fixed point defined by

$$b_k^\star = \frac{\delta}{\lambda}\rho_0\left[\exp\left(\frac{\sum_i b_i - \theta}{\Delta U}\right) - 1\right] \quad (27)$$

and has the same stability properties. Now consider the dynamics after time t^ε. The fixed point $b_k^\star$ defined by Equation 27 is again larger than the fixed point defined by

$$b_k^\varepsilon = \frac{\delta}{\lambda}\rho_0\left[\exp\left(\frac{b_k^\varepsilon + \varepsilon - \theta}{\Delta U}\right) - 1\right]. \quad (28)$$

Inspection of the dynamics of the equation $\bar{b}_k^{\varepsilon\prime} = \lambda\bar{b}_k^\varepsilon(\bar{b}_k^\varepsilon + \kappa) - \bar{b}_k^\varepsilon\delta\rho_0^{post}\exp\left(\frac{b_k^\varepsilon + \varepsilon - \theta}{\Delta U}\right)$ shows that, for sufficiently small ε, the fixed point b_k^ε is globally stable. In the limit of large t, $b_k^\star$ converges to $\lim_{\varepsilon\to o} b_k^\varepsilon = b_k^o$. Thus, $\lim_{t\to\infty} \mathbf{b}(t) = \mathbf{b}^\star = (b_1^\star,\ldots,b_N^\star)^T$ with $b_j^\star = 0$ for $j \neq k$ and $b_k^\star > 0$.

Simulation parameters
Simulation parameters are listed in Table 2. All simulations were based on the same parameter sets except for simulations with the correlation-based coding in which learning rates were adjusted to $\eta = 0.002$ mV^{-2} s^{-1}, $A_+ = A_- = 0.01$, and $\eta_{\text{stabilize}} = 0.995$.

We performed control experiments in which LTP was not dependent on the presynaptic firing rate. In these experiments, learning rates were adjusted. In the control experiments reported below (see Self-organization of nonlinear neural computation), if the rate dependency is

abolished, synapses from neurons with low rates tend to be more strongly potentiated. Hence, these weights are on average larger than in experiments with frequency-dependent LTP. This increases the firing rate of the postsynaptic neuron when a pattern is stored, which in turn leads to stronger depression of synapses from highly active presynaptic inputs. As a result, these synapses are not potentiated strongly enough to ensure robust pattern separation. One possible solution for this problem is to reduce LTD. In our simulations, we achieved this by reducing the STDP scaling constants to $A_+ = 0.01$ and $A_- = 0.006$ but leaving the learning rate for LTP without postspikes unchanged at $\eta = 0.004 \text{ mV}^{-2} \text{ s}^{-1}$. In the control experiments reported below (see Acquisition of nonlinear functionality based on synchrony), all inputs have a firing rate of 10 Hz. Hence, when we ignore the rate-dependent factor, this effectively increases the learning rate. We partially compensated for this increase by halving the learning rate factors that led to values $\eta = 0.001 \text{ mV}^{-2} \text{ s}^{-1}$ and $A_+ = A_- = 0.005$.

The passive EPSP decay constant u^{passive} was chosen to obtain qualitatively similar behavior as reported previously (Poirazi et al., 2003; Losonczy and Magee, 2006). Several studies (Williams and Stuart, 2002; Nevian et al., 2007), however, report very strong attenuation of the EPSPs in distal as well as basal dendrites toward the soma. In our simulations, the value of u^{passive} can be varied in a wide range without a qualitative change of the results. Using, for example, $u^{\text{passive}} = 0.2$ leads to very similar results without any adaptation of other parameters for most simulations. For the simulations presented below (see Acquisition of nonlinear functionality based on synchrony), the target voltage φ_b in Equation 12 has to be increased from 30 to 37 mV such that stronger branches compensate for the smaller passive component.

For simulations, time was discretized with a discretization time of 0.5 ms. The simulation code was written in C++, and simulations were performed on a standard UNIX desktop.

Measurement of spike train correlations and generation of correlated spike trains
To compute correlation coefficients between spike trains for the analysis shown below (see Acquisition of nonlinear functionality based on synchrony), spike trains were convolved with a Gaussian kernel with $\sigma = 5$ ms, and the correlation coefficient of the resulting time series were computed.

To produce spike trains with correlation factor cc and frequency ρ, we proceeded as Gütig et al. (2003) with time bin of size $\Delta t = 0.5$ ms. We constructed a Poisson spike train S_r with frequency f by assigning a spike to each bin with probability $\rho\Delta t$. The spike train S_r was used as a template for the construction of the input spike trains. Each input spike train was generated by assigning a spike to a bin not in S_r with probability $\rho\Delta t(1 - \sqrt{cc})$ and assigning a spike to a bin in S_r with probability $\rho\Delta t(1 - \sqrt{cc}) + \sqrt{cc}$.

Results

Nonlinear dendritic responses
According to Losonczy and Magee (2006), radial oblique branches in CA1 pyramidal neurons act as single integrative compartments that integrate synaptic inputs without regard to the spatial distribution within the branch. The view of oblique branches as single integrative compartments proposed by Losonczy and Magee (2006) was relativized by subsequent findings. Branco et al. (2010) showed that single oblique and basal dendritic branches are able to differentially read out spatiotemporal synaptic activation sequences. This suggests additional compartmentalizaton of individual branches. Multiple integrative compartments were also identified in individual basal branches (Major et al., 2008). Although additional compartmentalization of single branches increases the number of nonlinear integrative subunits per neuron and thus it may increase its capabilities for nonlinear computation, our model described below and illustrated in Figure 1 is based on the single-compartment view for the

sake of simplicity and clarity. The branches linearly sum asynchronous postsynaptic potentials from their synapses. We denote the passive/linear component of the dendritic potential at time t as $p_k(t)$ (for a detailed equation, see Materials and Methods). Strong synchronous synaptic input at oblique dendrites elicits dendritic spikes. We model such a spike at branch k by an additional active contribution to the branch voltage $a_k(t)$ whenever $p_k(t)$ is above a certain threshold. Note that this is a purely phenomenological model of dendritic spikes without the explicit inclusion of any positive feedback mechanism that underlies the generation of the event. The branch potential $b_k(t)$ is given by the sum of the local synaptic activation $p_k(t)$ at the branch and the additional contribution from the dendritic spike $a_k(t)$, leading to nonlinear branch characteristics (Fig. 2).

The passive component of the branch potential $p_k(t)$ is attenuated by a factor $u^{\text{passive}} < 1$ while it is conducted to the soma. According to recent experimental findings, the impact of a dendritic spike on the somatic potential can vary dramatically between branches and is subject to branch strength potentiation, a form of dendritic plasticity as detailed below (Losonczy et al., 2008). Dendritic spikes from different branches are thus scaled by branch-specific weighting factors—the branch strengths u_k—in our model before they reach the soma. Thus, the somatic membrane potential before refractory effects are taken into account is given by the sum of these passive and active weighted components from all branches (for detailed equations, see Materials and Methods).

The behavior of this model of nonlinear dendritic integration is illustrated in Figure 1B. It is consistent with neurophysiological recordings (Losonczy and Magee, 2006) and a study by Poirazi et al. (2003) in which a mixture of a sigmoid function with a linear contribution was found to provide the best fit to experimental data on the impact of dendritic input on the somatic membrane voltage.

This dendritic model exhibits two important behaviors. First, dendritic spikes serve as a basis for nonlinear neural computations as discussed in detail below. Second, the local branch potential $b_k(t)$ is a measure for the total synaptic input at this branch. This variable can potentially be read out by synaptic plasticity mechanisms—for example, by voltage-dependent STDP—such that weight updates can be based on the success of synapses in driving the dendritic branch (see below for the voltage-dependent plasticity mechanisms used in our model).

The somatic spiking mechanism and reset behavior of our model was based on the model of Jolivet et al. (2006). It is a stochastic spiking neuron model with an instantaneous firing rate that depends exponentially on the somatic membrane potential (Fig. 1C). The model parameters were fitted by Jolivet et al. (2006) to obtain an optimal fit to the behavior of layer 5 pyramidal neurons in rat somatosensory cortex. More details on the soma model can be found in Materials and Methods.

Nonlinear neural computation
Nonlinear dendritic responses facilitate the processing capabilities of individual neurons. Several important hypotheses on neural processing in cortex demand nonlinear computations. Two examples are the binding problem (von der Malsburg, 1981; Roskies, 1999) and the participation of neurons in neuronal ensembles (Hebb, 1949). Suppose that the occurrence of a specific feature in the input, such as a shape (square, triangular, circular, etc.), a color, or motion direction, is coded by the activation of a neuronal ensemble (Fig. 3). Binding problems are in general

problems in which several features of an object have to be bound for additional processing or storage. One class of binding problems is about relating a concept to a percept (Roskies, 1999). It demands neurons or populations of neurons to store combinations of feature and attributes. A neuron that binds—i.e., responds to—for example, a black disk and a yellow star, should however not respond to a yellow disk or a black star. This problem can be solved with superlinear dendritic integration (Fig. 3).

The same example can be viewed from the perspective of a neuron that participates in neuronal ensembles. Suppose a neuron participates in two neuronal ensembles. In the previous example, the neuron participates in ensembles "black disk" and "yellow star." However, the neuron should not be activated if half of the neurons in each ensemble are active. In the previous example, that would be the ensembles "yellow disk" and "black star." More abstractly, the binding problem calls for learning of concepts that have the logical form of an OR of ANDs. In other words, to solve a binding problem, one has to perform first a logical AND operation for each feature combination and then has to detect whether one such combination is active by a logical OR over these ANDs. The situation described above is a nonlinear pattern separation problem that we refer to as the "feature binding problem" (FBP) in the following. Given four input features, the neuron should respond to two disjunct combinations consisting of two features each—black disk and yellow star in the previously described example—but not to any other two-feature combination, such as a yellow disk and black star. This function cannot be computed by a neuron model that linearly sums individual synaptic inputs, even if weights can be chosen arbitrarily to be excitatory or inhibitory. To see this, consider our neuron model without the dendritic nonlinearity. This means that the somatic membrane potential is given by the summed EPSPs from all branches. Now consider the mean somatic potential v_{mean} if a single input ensemble is active. That is, record the somatic potential v_{yellow} when only the ensemble coding for "yellow" is active. Then record the somatic potential v_{black} when only the ensemble coding for "black" is active, and so on. Finally, compute the mean over all these somatic potentials, that is $v_{mean} = \frac{1}{4}(v_{yellow} + v_{black} + v_{star} + v_{disk})$. Note that, for a given neuron with given synaptic efficacies and branch strengths, we obtain a unique v_{mean}. Because we demand that the neuron is active for yellow star and black disk, we demand that, for these combinations, the somatic membrane potential is above threshold θ. It follows that the mean somatic membrane potential over single features v_{mean} is above half of the threshold $v_{mean} > \frac{\theta}{2}$. Conversely, we demand that, for combinations yellow disk and black star, the somatic membrane potential is below threshold. This constraint implies that the mean somatic membrane potential over single features v_{mean} is below half of the threshold $v_{mean} < \frac{\theta}{2}$. Because for a given neuron, v_{mean} cannot be above and below half of the threshold, the function cannot be computed by such a neuron. Hence, nonlinear dendritic operations are essential to solve the tasks with single neurons, because a linear model would always classify at least one of the four patterns incorrectly regardless of its weight configuration. Note however that the traditionally often considered exclusive or (XOR) function (Minsky and Papert, 1988) cannot be computed by our neuron model even with nonlinear branches, because this function requires that the neuron responds to a pattern like yellow but

does not respond to a pattern like yellow disk, which is a superset of the stored pattern. The classical XOR can thus only be solved with the help of inhibitory synapses, which are not modeled in this article. A possible role of inhibitory circuits could thus be to enable pyramidal neurons to store such XOR-like pattern configurations.

Synaptic plasticity and branch-strength potentiation

To examine whether such nonlinear functionality could in principle be acquired by a neuron in a self-organized manner, we used a plasticity model based on several experimentally observed plasticity effects in pyramidal neurons, including depolarization-dependent STDP (Ngezahayo et al., 2000; Nevian and Sakmann, 2006; Sjöström and Häusser, 2006), dependence of STDP pairing frequency (Sjöström et al., 2001), and the potentiation of branch strengths (Losonczy et al., 2008). Acetylcholine (ACh) has a strong effect on synaptic plasticity and branch-strength potentiation. Specifically, our model for branch-strength potentiation discussed below is based on findings of Losonczy et al. (2008) for elevated cholinergic modulation levels. Furthermore, ACh can lower the threshold for LTP induction (Auerbach and Segal, 1994; Dringenberg et al., 2007). We thus assume that learning is enabled when local acetylcholine levels are elevated, which might signal exploratory behavior, saliency of the stimulus, or alertness. Conversely, we hypothesize that ACh levels are low during recall, which reduces plasticity effects. These assumptions are supported by experimental data that indicate that ACh can gate cortical plasticity (Weinberger, 2008) and enhance memory storage in hippocampus (Ragozzino et al., 1996). For the sake of simplicity, we do not explicitly model the ACh signal in our simulations but disable all plasticity mechanisms in the testing phase.

Spike-timing-dependent plasticity is a Hebbian plasticity mechanism in which the weight change of a synapse depends on the precise timing of the presynaptic and postsynaptic action potentials (Markram et al., 1997; Bi and Poo, 1998). Recent experiments showed that STDP also depends on the local dendritic depolarization and Ca^{2+} level at the local dendritic site (Nevian and Sakmann, 2006; Sjöström and Häusser, 2006; Larkum and Nevian, 2008). These findings are of special importance for neuron models with dendritic arborizations because dendritic depolarization may be an important regulating factor that differentiates plasticity at different dendritic branches for the same input pattern. We therefore extended a standard model for STDP (Abbott and Nelson, 2000) to include depolarization dependence (a related model for depolarization-dependent STDP was studied by Clopath et al., 2010). Pre-before-post spike pairs within a certain time window induce LTP if the depolarization at the branch exceeds a certain threshold φ_+. Conversely, post-before-pre spike pairs induce LTD (Fig. 4). Sjöström et al. (2001) showed that LTP depends approximately linearly on pairing frequency, whereas there is no such dependency of LTD below 40 Hz. Because the presynaptic and postsynaptic firing rates were always covaried in this study, the relative contribution of each of these two factors remains unclear. From a theoretical perspective, the presynaptic firing rate is especially interesting. If low presynaptic rates reduce LTP but not LTD, weakly activated inputs can be depressed selectively. We therefore modeled frequency dependence by a linear dependence of LTP on the presynaptic firing rate (Fig. 4). Because the contribution of the presynaptic rate in the reported influence of pairing frequency on STDP is still unclear, we performed control simulations in which LTP did not depend on the presynaptic frequency. As reported below, this
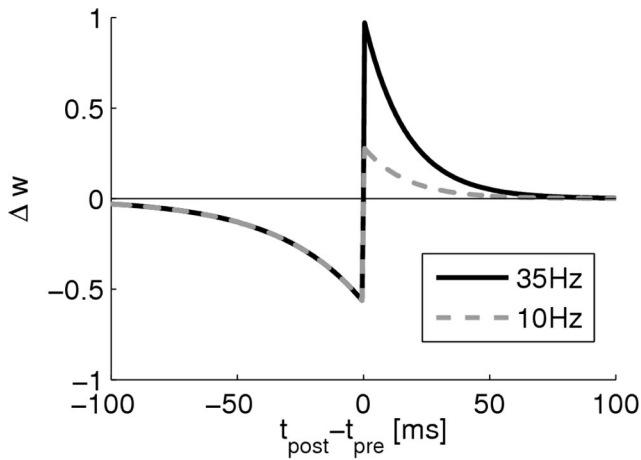
Figure 4. Rate-dependent STDP. The model includes linear dependence of LTP on presynaptic firing rate and an induction threshold for LTP. Shown is the weight change for a presynaptic rate of 35 Hz (black curve) and 10 Hz (gray dashed curve) scaled to the maximum possible weight change at 35 Hz in dependence of the time difference $\Delta t = t_{post} - t_{pre}$ of a pre–post spike pair.

Figure 5. Emergent competition between dendritic branches. **A**, Black dots denote input spikes to the neuron (every 10th neuron is shown). Ensembles *a* and *c* are highly active. **B**, Voltages at individual branches. Colors code for branch identity; the voltage of branch 6 (the winning branch) is plotted in red. **C**, Bars denote output spikes of the trained neuron. Once the neuron becomes responsive, its firing rate increases to ~50 Hz. **D**, Weights and branch strengths for branch 1 (left panels) and branch 6 (right panels; the winning branch). The top six panels show average weights for synapses from the six input ensembles to these two branches. The bottom plot shows the evolution of branch strengths for these two branches. The branch strength of the winning branch (right panel) increases, whereas others stay approximately constant because no dendritic spikes are elicited in these branches.

induction of LTP (Kelso et al., 1986; Gustafsson et al., 1987; Hardie and Spruston, 2009). In our model, synapses can also be somewhat potentiated in the absence of postsynaptic firing and dendritic spikes, although potentiation in the presences of a dendritic spike is substantially stronger because the amount of potentiation increases linearly with the local branch depolarization b_k (see Eq. 11 and discussion in Materials and Methods). The plasticity mechanism therefore preferentially potentiates groups of synapses at a single branch that tend to be coactivated and thus depolarize the branch. We will see in the analysis below that this feature of the plasticity dynamics contributes to synaptic clustering and dendritic competition. Furthermore, as in the STDP model described above, we model a linear dependence on presynaptic firing rate (for detailed equations, see Materials and Methods).

Recent experimental data (Losonczy et al., 2008; Makara et al., 2009) show that the impact of a spike in oblique dendrites on the somatic membrane potential depends on properties of the dendritic branch giving rise to the notion of a branch strength. Furthermore, it was shown that this branch strength is plastic. More precisely, the strength of the branch–soma coupling can be increased through pairing of dendritic spikes with ACh. The phenomenon was termed BSP. We model this plastic behavior of branch strength u_k by the temporal dynamics given in Equation 12 in Materials and Methods. Furthermore, potentiation was only reported for weak branches, indicating a saturating mechanism. We therefore implemented a saturation of BSP if the local potential equals a constant φ_b (see Materials and Methods).

Finally, to stabilize weights, every synapse has an individual adaptive learning rate. This learning rate decreases whenever the local membrane voltage is above a threshold $\varphi_{stabilize}$ (set to 0.5 mV above the voltage that occurs during a dendritic spike) in coincidence with a postsynaptic somatic spike (for details, see Materials and Methods). A summary of the plasticity rules in the model is given in Table 1.
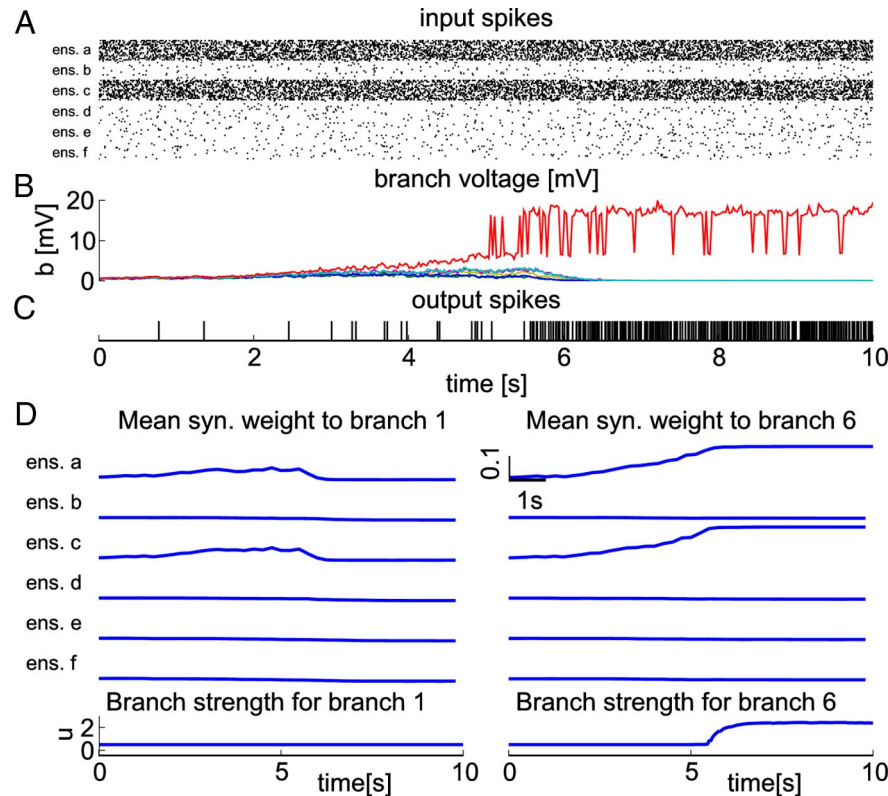
Dendritic competition

We first tested how the model behaves when a single pattern of presynaptic ensemble activity is presented to the neuron. In these simulations, we considered six input ensembles, each consisting of 144 neurons. Each neuron in the input ensembles was synaptically connected to a randomly chosen branch of the adapting neuron (using a uniform distribution over branches) as schematically indicated in Figure 3. This resulted in an average of 24 synaptic connections from each ensemble to each of the six branches. Neurons in an active ensemble were emitting 35 Hz Poisson spike trains in which spike–spike correlations between neurons in the ensemble were not enforced in the generation of these spike trains. Other neurons were spiking at a background rate of 2 Hz. We presented one combination of two active ensembles for 10 s to the neuron.

Figure 5 shows that the neuron becomes responsive for this

leads to quantitatively very similar results (for details on the STDP learning rule used, see Materials and Methods).

By definition, induction of STDP requires a postsynaptic spike. It is well known, however, that pairing of presynaptic spikes with postsynaptic depolarization can be sufficient for the

input pattern and that a single branch stores the pattern as a result of an emergent competition between branches. Until $t = 5$ s, all branch activities rise because presynaptic activity potentiates synaptic efficacies (Eq. 11). One can see, however, that the voltage difference between branch 6 (Fig. 5B, red trace) and other branches is consistently growing starting at time $t = 2.5$ s. This effect arises because higher local branch potential leads to stronger LTP. At time $t = 5$ s, the threshold for dendritic spikes is first reached at branch 6. When the neuron starts to fire at high rate (starting at $t = 5.5$ s), less strongly active branches become less and less responsive to the pattern. In this case, a single branch wins and specializes to the given combination of active ensembles.

The weight dynamics at two branches are shown in Figure 5D and follow a similar pattern: in a first phase, the weight updates at all branches favor synapses originating from active ensembles because of the linear dependence of LTP on presynaptic frequency. Thus, all branches increase their response to this pattern. However, because the updates depend on the local membrane voltage at the branch,

Figure 6. A neuron with STDP and BSP can store three different patterns on individual branches. **A**, Black dots denote input spikes to the neuron. Each line corresponds to one input neuron (every 30th neuron is shown). Input spike trains were sorted according to their input ensemble. Different patterns are presented sequentially. **B**, Voltages at individual branches. Colors code for branch identity (green, branch 2; blue, branch 4; red, branch 6). **C**, Firing rate of the trained neuron during learning.

the most active branch—branch 6—is favored and increases its response most strongly. As dendritic spikes are robustly evoked in branch 6 at $t = 5.5$ s, the neuron starts to fire at high rate and STDP leads to LTD. At this time, the local branch voltage at branch 6 has a high value as a result of dendritic spikes. Therefore, depolarization-dependent LTP can compensate for the depression at this branch. The local branch voltage at other branches is much lower (Fig. 5B) and the mean synaptic weight decreases at these branches. As dendritic spikes are robustly evoked in branch 6 after $t = 5$ s, the corresponding branch strength u_6 also increases (Fig. 5D, bottom plots). This leads to an increase of the firing rate of the neuron until the branch strength reaches a steady state (see below). Note that the backpropagating postsynaptic spike plays an important part in the competitive dynamics through its critical role in LTD induction. It indicates that the neuron reached threshold and thus that some branch was strongly activated. This information is used to weaken competing branches to avoid redundant information storage.

We wanted to better understand how and under which conditions this competitive behavior arises from the interplay between LTP and LTD. To answer these questions, we have analyzed the weight dynamics theoretically. To make a theoretical analysis possible, we used a simplified model in which the discontinuous branch behavior is replaced by an exponential nonlinearity. From the learning rules, we included the minimal set from our model that is essential for dendritic competition. This set consists of voltage-dependent LTP for each presynaptic spike and LTD for post-before-pre spike pairs. The full analysis can be found in Materials and Methods. Assuming a large number of synaptic inputs, we have then approximated the voltage at dendritic branches by their mean field, which yields a rate formulation of neuronal dynamics. Accordingly, the weight changes attributable to spike-based synaptic learning rules were approximated by the average change over different realizations of the
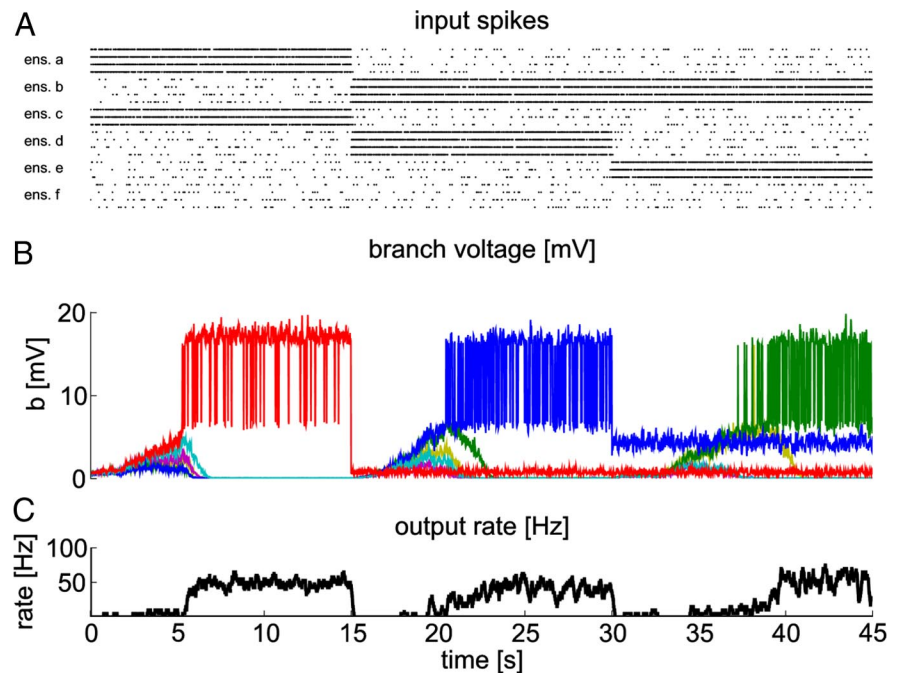
Poissonian input statistics, again yielding a rate-based formulation of the weight dynamics. In this simplified model, we proved that the weight dynamics in fact lead to dendritic competition such that only the most strongly activated dendritic branch stores the presented pattern. More precisely, assume a pattern of constant input rates and assume that, at time $t = 0$, branch k has higher branch activation than any other branch, i.e., $b_k(0) > b_j(0)$ for all $j \neq k$. We then showed that, for a wide range of parameters, the weight dynamics imply changes in the branch potentials such that the potential $b_k(t)$ of the most strongly activated branch will converge to a positive value, whereas the potentials of other branches will converge to 0. The proof is quite instructive for the understanding of the basic mechanism of dendritic competition and sketched in the following. First, we observe that there is a minimal and a maximal value of the branch potential $b_k(t)$ such that, after some time, b_k will always stay within these bounds. This happens because, at low postsynaptic firing rates, LTP dominates LTD, which increases the synaptic weights and as a consequence also increases the branch potential b_k. At high firing rates, on the contrary, LTD dominates, which decreases the synaptic weights and therefore the branch potential b_k. It follows that the sum of weights at branch k is also bounded. Consider the difference between the sum of weights at branch k and the sum of weights at another branch $j \neq k$. Because of voltage-dependent LTP, this difference increases linearly with time. These observations imply that b_j converges to 0 for $j \neq k$, whereas the initially highest branch potential b_k converges to some positive value.

Self-organization of nonlinear neural computation

With such a competitive mechanism, a neuron can dedicate a small number of dendritic compartments to the storage of a given input pattern. This endows a neuron with the capability to store many different patterns in its dendritic arborizations and thus to self-organize nonlinear behavior. We tested the self-organization
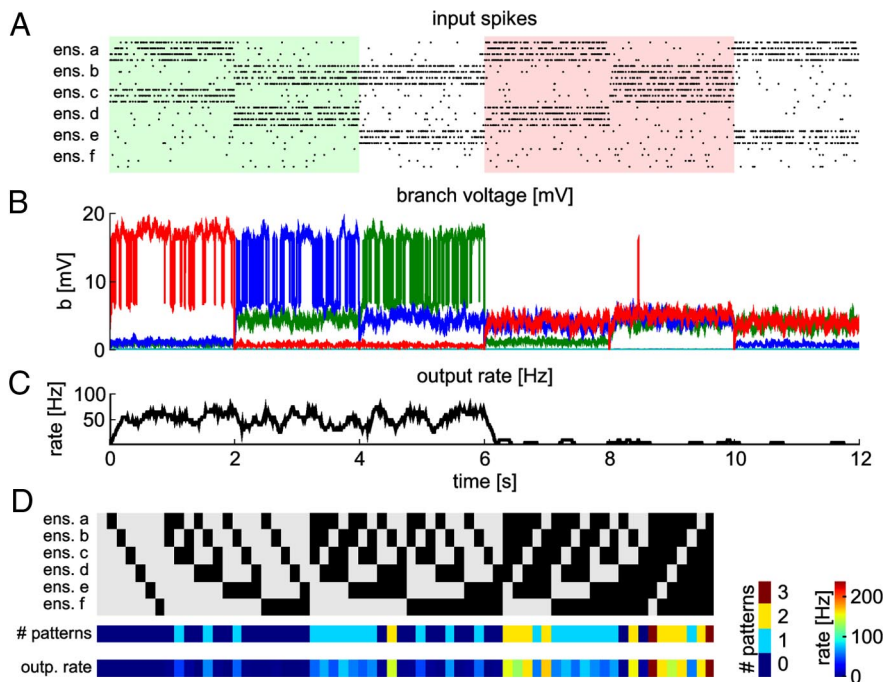
Figure 7. A neuron with STDP and BSP can acquire nonlinear functionality through STDP and BSP. **A**, Black dots denote input spikes to the neuron (every 30th neuron is shown). Different combinations of active ensembles are presented sequentially. The first three patterns were previously entrained by STDP and BSP. Color shading indicates patterns corresponding to the FBP problem discussed in Results. **B**, Voltages at individual branches. Colors code for branch identity (green, branch 2; blue, branch 4; red, branch 6; data smoothed). **C**, Firing rate of the trained neuron during testing. The neuron has high firing rate for all trained patterns and low firing rate for other patterns. It has high firing rate for patterns 1 and 2 (light green shading) and is silent for patterns 4 and 5 (light red shading), which define an FBP problem. This shows that the neuron has acquired nonlinear functionality. **D**, Summary of neuron responses after training. Each row shows the result for 1 of the 64 possible combinations of active ensembles. Top, Input patterns; active ensembles are indicated in black. Middle, Number of trained patterns that are contained in this input (left color bar). Bottom, Neuron firing rate (right color bar) for the corresponding input pattern. The neuron is activated only if at least one trained pattern is contained in the current input. If an input contains several trained patterns, the neuron is more strongly activated, because each pattern activates one branch.

capabilities of the neuron in two difficult situations of the binding problem. First, consider the case in which two input patterns should be stored, where we define a pattern as a set of coactive input ensembles. The first pattern is given by active ensembles *a* and *c*—we abbreviate such patterns as (*a*,*c*) in the following—and the second combination is indicated by active ensembles (*b*,*d*) (Fig. 6*A*, *t* = 0−15 s and *t* = 15−30 s, respectively). This situation corresponds to the FBP problem discussed above and illustrated in Figure 3. Using again the example illustrated in Figure 3, ensemble *a* codes the feature yellow and ensemble *c* the feature star. The first pattern thus represents a yellow star. The second pattern represents a black disk with ensemble *b* coding for black and *d* for disk. After training, the neuron should respond to these positive patterns, but it should not respond to the patterns yellow disk indicated by active ensembles *a* and *d* and black star indicated by active ensembles *b* and *c*. The situation is summarized in Figure 7*A*, where the patterns yellow star and black disk are emphasized by light green shading, and the patterns yellow disk and black star are indicated by light red shading.

The second situation is again characterized by two patterns, one in which ensembles (*b*,*d*) are active and another pattern with active ensembles (*b*,*e*) (Fig. 6*A*, *t* = 15−30 s and *t* = 30−45 s, respectively). Hence, there is an ∼50% overlap between the patterns. Combining these two situations, the neuron should respond to the patterns (*a*,*c*), (*b*,*d*), and (*b*,*e*) and stay unresponsive to any other combination of two active ensembles, especially for other combinations with ensemble *b*. We refer to this problem as

the extended feature binding problem (eFBP) in the following. Note that this problem is at least as hard as the FBP. For reasons discussed below (see The role of branch strength potentiation), it is actually harder from the practical point of view.

We tested the self-organization capabilities of a neuron in this situation, i.e., in the eFBP, in the following simulation: the three patterns were presented sequentially, each for 15 s of simulated biological time (Fig. 6*A*). We refer to patterns that are presented during training as trained patterns in the following. The behavior during learning is shown in Figure 6, *B* and *C*. Individual branches specialize to specific patterns as can be seen in *B*. This behavior results from dendritic competition as described above.

Figure 7 shows the response of the neuron after training to the three trained patterns and to three nontrained patterns. Synaptic input during presentation of nontrained patterns does not depolarize dendritic branches strongly enough to cause dendritic spikes. Therefore, the firing rate of the neuron for these patterns is quite low. Figure 7*D* summarizes the responses of the trained neuron to all possible combinations of active ensembles (each presented for 2 s after training). It shows that the neuron responds to an input pattern only if it contains at least one of the trained patterns. Formally, we say that an input pattern P_{in} contains a trained pattern P_{tr} if all coactive ensembles of P_{tr} are also in the set of coactive ensembles of P_{in}, i.e., if the set of coactive ensembles in P_{tr} is a subset of the set of coactive ensembles in P_{in}. For examples, the input pattern (*a*,*b*,*c*) (the 23rd pattern in Fig. 7*D*) contains the trained pattern (*a*,*c*) but not the pattern (*b*,*d*) because ensemble *d* is not active in this input pattern. Because we do not model inhibitory inputs, the model is monotone. That is, the addition of an active ensemble to an input pattern cannot lead to a decrease of the neuron response (see also the discussion of the XOR problem above). Hence, every input pattern that contains a trained pattern activates the neuron (Fig. 7*D*). The firing rate of the neuron increases with the number of trained patterns contained in the current input, because each trained ensemble combination activates at least one distinct branch. In summary, Figure 7*D* shows that the neuron has learned to separate inputs that contain trained patterns from those that do not contain any trained pattern.

These capabilities of the neuron have been tested in 20 independent simulations with different random connectivity from the input ensembles to the neuron, different random initial weights, and different realizations of the Poissonian spike trains emitted by the ensemble neurons. The neuron was trained for 40 s simulated biological time on each of the three patterns. To assess the ability of the neuron to separate trained from nontrained patterns, we tested the response of the neuron to the presentation of any possible combination of two active ensemble for 500 ms after training. The results are summarized in Figure 8*A* together with experiments in which neuron parameters were varied ran-

domly to investigate the robustness to parameter perturbations. In these control experiments, the stochastic threshold θ, the firing rate at threshold ρ_0, the width of the spike-trigger zone ΔU, and the absolute refractory time $t_{refract}$ were concurrently perturbed. The parameters were concurrently varied by various perturbation levels α of up to $\pm100\%$. That is, for a simulation with perturbation level α (in percentage), each parameter value was multiplied in each run by a random number between $1 - \alpha/100$ and $1 + \alpha/100$, for the threshold θ, the firing rate at threshold ρ_0, and ΔU. The refractory period was at the same time varied by $\pm50\%$. As can be seen in Figure 8A, the patterns are well separated even for large parameter perturbations. We note that BSP has only been described in CA1 pyramidal neurons that have also extensive oblique dendrites, whereas we used a soma model that has been fit to a layer 5 pyramidal neuron in somatosensory cortex. We used this soma model because it provides a good tradeoff between physiological plausibility and model complexity. The control experiments described above, in which soma parameters were randomly varied, show that the details of the soma model are not important for the main results reported here.

These results were obtained with a plasticity model in which LTP depends linearly on the presynaptic firing rate (see update rules Eq. 6 and 8). To assess whether this dependence is necessary for neuronal self-organization, we performed control simulations with a model in which this dependency was dropped (see Materials and Methods). In 20 independent simulations, the response to trained patterns was 52 ± 10 Hz, whereas for other patterns, it was 3 ± 3 Hz. This shows that the rate dependence is not essential for pattern storage in this model.

The role of branch strength potentiation
Potentiation of a branch strength through BSP increases the influence of dendritic spikes originating from this branch on the somatic membrane potential. Hence, the probability of postsynaptic spikes is increased, which contributes to the competitive mechanism described above. Consistent with a hypothesis expressed by Losonczy et al. (2008), the branch strength stores a reference to the input pattern, and a subsequent pattern presentation will elicit reliable spiking of the neuron. Figure 9A (top) summarizes the results of the eFBP experiment described above and illustrated in Figure 7, in which we trained a neuron on three feature combinations (a,c), (b,d), and (b,e) and tested on any other combination of two features in 20 independent trials. Figure 9A (top) shows the minimal response to any of the trained patterns (left) and the maximal response to any of the other patterns (right; mean and SD over 20 independent trials). One can see that the patterns are well separated. To illustrate the role of BSP in the separation of patterns, we performed the same experiments but without BSP. The branch strengths thus stay at their initial value of 0.5 during the learning experiment. We first tested this setup on the FBP, i.e., when only the patterns (a,c) and (b,c) are trained (Fig. 9A, middle). The neuron is activated by the trained patterns, although this activation is weaker because branch strengths are not potentiated. The impact of the dendritic spike onto the soma is weak, such that trained patterns are not well separated from other patterns. As one can see in the bottom
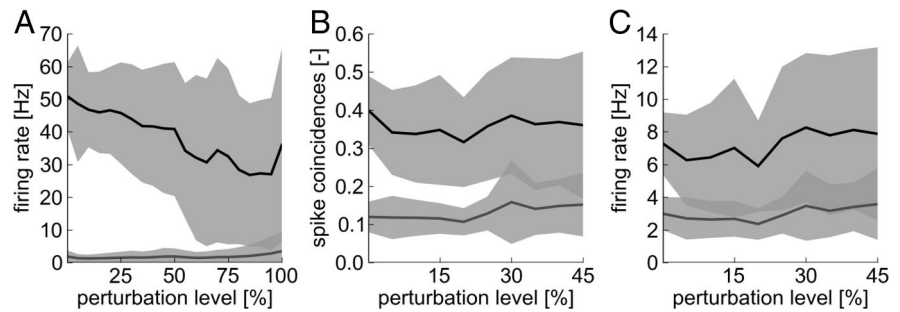

Figure 8. Pattern separation is possible under a wide range of neuron parameters. Several somatic parameters were perturbed by up to 100% of their standard values (x-axis, perturbation level). Full lines show means (over 20 experiments) and shaded areas show STDs. Dark gray shaded areas indicate overlapping SEs. ***A***, Mean firing rate response to trained patterns (black line) and mean response to other patterns (gray line) in the pattern separation experiment based on firing rates of input ensembles. Pattern separation is possible even for parameter perturbations of up to 75%. ***B, C***, Mean spike coincidences (***B***) and firing rates (***C***) for trained patterns (black line) and for other patterns (gray line) in the pattern separation experiment based on synchrony of input ensembles. Although firing rates can overlap, the patterns can well be separated based on spike coincidences for parameter perturbations of up to 25%.
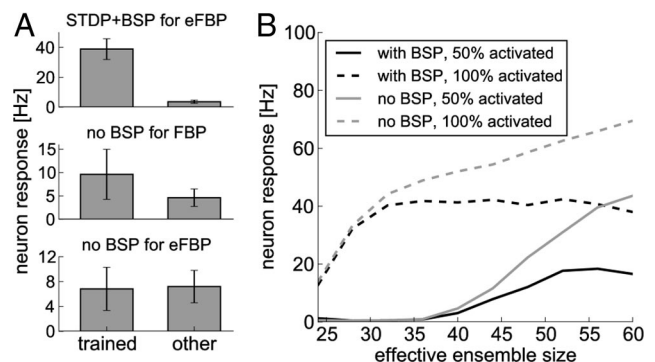

Figure 9. BSP improves the pattern separation capabilities of a neuron. ***A***, Comparison of learning with STDP and BSP (top) to learning without BSP for small initial branch strengths (middle, bottom). Shown is the minimum response to a trained pattern (left; mean and SD over 20 trials) and the maximum response to a nontrained pattern (right; mean and SD over 20 trials). With BSP, patterns are clearly separated in the extended feature binding problem (top). Without BSP, patterns are not well separated in the feature binding problem (middle) and not separated in the extended feature binding problem (bottom). ***B***, A comparison of neuron responses with BSP during training (black lines) to the responses of a neuron with constant high branch strength (gray lines) as a function of the number of synapses projected from an ensemble to each branch (effective ensemble size). Shown are responses after training to the full pattern (dashed lines) and to a partial activation in which only 50% of the neurons in these ensembles were activated (full lines). Without BSP, neuron responses increase steadily with increasing ensemble size, whereas BSP stabilizes responses over a wide range of effective ensemble sizes.

of Figure 9A, in the case of the eFBP, nontrained patterns activate the neuron as strongly as trained patterns. The problem becomes apparent when we consider the pattern (b,c). Because this pattern was not presented to the neuron during training, no branch is activated strongly enough to elicit branch spikes reliably. Still, the soma is activated by the linear voltage components of three branches that store patterns that partially overlap with (b,c). Those are the branches that store the patterns (a,c), (b,d), and (b,e) respectively. This linear activation outweighs the weak nonlinear component that is added to trained patterns and the two cannot be separated anymore. Hence, increasing the branch strength through mechanisms such as BSP is essential for these types of feature binding problems.

The weak activation of trained patterns in the absence of BSP could be compensated by increasing the maximum possible synaptic efficacy w_{max}. In that process, however, the nonlinear com-
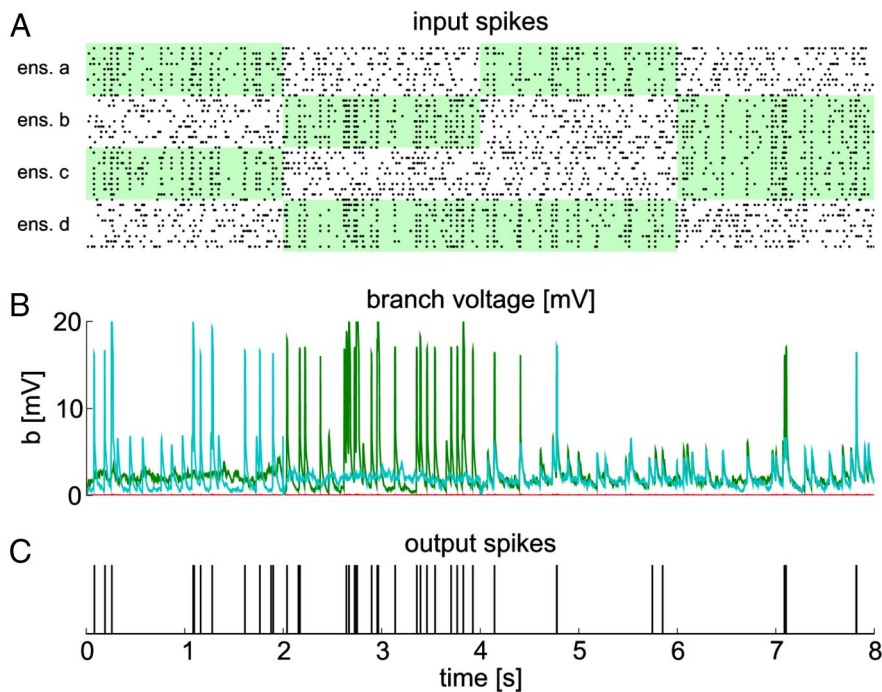
Figure 10. A neuron with STDP and BSP can acquire nonlinear functionality through STDP and BSP in a correlation-based coding scheme. **A**, Black dots denote input spikes to the neuron (every 10th neuron is shown). Different patterns are presented sequentially. Green areas indicate correlated input ensembles. The first two patterns were previously entrained by STDP and BSP. **B**, Voltages at individual branches. Colors code for branch identity (green, branch 2; cyan, branch 4). **C**, Spike times of the neuron during testing.

ponent of branches is even less pronounced compared with the linear component, and the neuron totally fails to solve even the less demanding FBP.

In the above analysis, we have considered the case in which the branch strength is weak initially. This seems to be the general situation in pyramidal neurons, in which initially weak branch strengths are potentiated by BSP (Losonczy et al., 2008; Makara et al., 2009). We then asked the question what would happen if potentiation with BSP would not occur. We now consider the hypothetical case in which branch strengths are already large initially and patterns are stored by synaptic plasticity only.

In this case, feature binding would be possible with synaptic plasticity alone if branch strengths were set to correct values. However, as we show in the following, adaptive branch strengths are advantageous even in this case. More specifically, the postulated dependence of BSP on the local membrane potential in the dynamics Equation 12 can serve a homeostatic role. Because the number of synapses that connect a given ensemble with a specific branch—referred to here informally as effective ensemble size—varies between ensembles and branches, different patterns would lead to quite different neural responses if branch strengths were fixed. Our model for BSP leads to an approximate equalization of the somatic voltage across patterns of different sizes. Figure 9B summarizes several simulations in which a neuron was trained on a single pattern consisting of two active ensembles as before. To test the influence of effective ensemble size on the neural response, we varied the number of synapses that each ensemble projected onto each of the six branches between 24 and 60. We then compared the responses of two neurons, one neuron trained with BSP as above and the other one with constant branch strength. This constant value was set such that the neurons had the same firing rate of 40 Hz at 26 synapses per ensemble and branch after training. The responses of these neurons after train-

ing to the full pattern and to a presentation in which 50% of the synapses onto each branch were activated were then compared. Figure 9B shows that, with BSP, the neuron stabilizes its response to the full pattern at ~40 Hz when the effective ensemble size is large enough to reliably elicit dendritic spikes after training. For constant branch strength, the branch cannot compensate for the stronger input drive and the response increases with effective ensemble size. A similar observation can be made for the half-activated pattern. In this case, the rate of the neuron equipped with BSP stabilizes at ~20 Hz. With constant branch strength, the response increases more strongly and steadily with increasing effective ensemble size, making it impossible to distinguish between a fully activated pattern of low effective ensemble size and a partly activated pattern of large effective ensemble size. With BSP, however, patterns can be stored over a wide range of effective ensemble sizes, although lower ensemble sizes may be advantageous because of weaker responses to partial activations. Thus, we have shown that, for both cases of small and large initial values, adaptation of branch strengths significantly improves the feature binding capabilities of single neurons.

Acquisition of nonlinear functionality based on synchrony
We then considered the case when information about input patterns is coded by synchronous activation of neurons instead of by firing rate. All neurons were activated at a constant rate of 10 Hz. The firing times of neurons in active ensembles were correlated while other neurons emitted uncorrelated Poissonian spike trains. Again, competition between branches emerges like in the rate-based coding scheme. However, because all inputs have the same firing rate, the rate dependency of LTP cannot be exploited to exclusively strengthen synapses that contribute to the currently presented pattern. Instead, when the threshold for LTP is reached at an individual branch, correlated inputs are subject to considerably more LTP than noncorrelated inputs because they produce postsynaptic spikes and therefore presynaptic spikes fall more frequently in the LTP part of the STDP learning window. As the postsynaptic firing rate increases, the net-negative weight changes attributable to STDP at uncorrelated inputs exceeds the amount of potentiation, and these synapses are depressed (for detailed analysis of this effect of STDP for correlated inputs, see Kempter et al., 1999, 2001; Song et al., 2000).

We simulated one neuron with four branches and presented two patterns representing the FBP as in the rate-based scheme, but now with 10Hz Poisson input spike trains at all inputs. These input spike trains were correlated with a correlation coefficient of 0.5 at active ensembles and uncorrelated at other ensembles. Each ensemble consisted of 196 neurons, and each neuron was synaptically connected to a randomly chosen branch of the adapting neuron. This resulted in an average of 24 synaptic connections from each ensemble to each of the four branches.

Again, we tested the response of the neuron to trained patterns and other combinations of active ensembles. Figure 10 shows a simulation in which four patterns were presented after training.
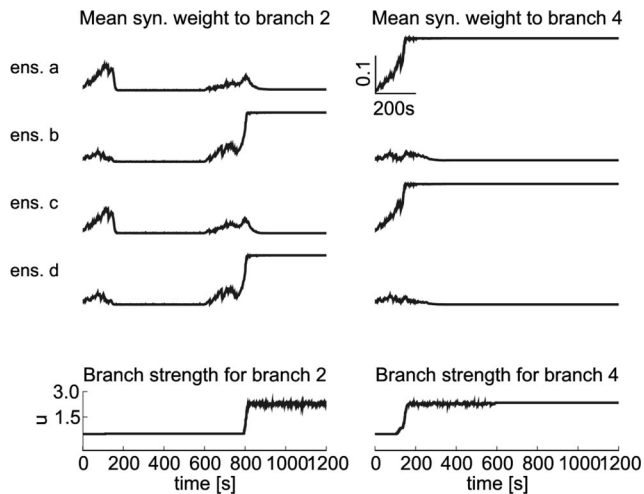
Figure 11. Weights and branch strengths during neuronal self-organization in a synchrony-based coding scheme. The panels show average weights for the two winning branches 1 (left) and 2 (right). The top four panels show average weights for synapses from the four input ensembles to these two branches. The two correlated ensembles are *a* and *c* for the first 600 s and *b* and *d* for the remaining time. The bottom plots show the evolution of branch strengths for these two branches. Because all ensembles have the same average firing rate, all weights increase initially. However, as soon as postsynaptic spikes appear, the synaptic strengths of uncorrelated inputs as well as synaptic strengths of non-winning branches decay quickly.

For the trained patterns, the neuron produces action potentials that are correlated with the input while spikes occur only occasionally for other patterns. The evolution of weights and branch strengths during training is shown in Figure 11.

To assess the influence of random initial conditions, we performed 20 independent simulations with different random connectivity from the input ensembles to the neuron, different random initial weights, and different realizations of the Poissonian spike trains emitted by the ensemble neurons. In each trial, every pattern was presented for 20 s after training. The firing rate for the trained patterns was 7.3 ± 1.9 Hz, whereas for the other patterns, the neuron spiked with a firing rate of 3 ± 1 Hz. The mean correlation coefficient between the output spike train and the input spike trains of correlated ensembles was 0.38 ± 0.06 for trained patterns and 0.21 ± 0.06 for other patterns. Downstream neurons should thus be able to discriminate whether the pattern has been stored by the neuron not only on the basis of its firing rate but also based on its spike correlation with the input ensemble. For the membrane potential of a downstream neuron, however, the correlation coefficient itself is probably less relevant than the number of coincident spikes between the input ensemble and the trained neuron. We therefore analyzed the fraction of spikes from an input ensemble neuron that coincide with a spike from the trained neuron. Here, two spikes were said to coincide when the time difference between them was <5 ms. Averaged over all neurons of the correlated ensembles, this yielded fractions of 0.4 ± 0.09 for the trained patterns and 0.12 ± 0.04 for the other patterns. In other words, from 100 spikes of a given input ensemble neuron, a downstream neuron experiences on average 40, which coincides with spikes of the trained neuron if the pattern was presented during training. If the pattern was not trained, this number reduces to 12. We conclude that a discrimination based on spike coincidences is quite plausible.

We also performed a set of control simulations to test whether perturbations of soma parameters have an influence on these results. Here, we performed 20 independent simulations in which the parameters were concurrently varied as described above (see

Self-organization of nonlinear neural computation). We found, however, that the stochastic firing threshold cannot be much lowered without increasing the neurons firing rate significantly. For a perturbation level of α, the threshold θ was thus varied in the range between 100% and $(100 + \alpha)$% of its initial value. The results are shown in Figure 8*B* for the fraction of coincident spikes and in Figure 8*C* for the firing rate response of the neuron. Although the firing rate responses for trained and nontrained patterns shows some overlap for weak parameter changes, the fraction of spike coincidences is well separable for parameter variations of up to 25%. This shows that the basic mechanism works despite considerable variation of somatic parameters.

Again, these results were obtained with a plasticity model in which LTP depends linearly on the presynaptic firing rate (see update rules Eq. 6 and 8). Because all input neurons fire with the same average rate, this dependence should not be essential for the results. Control simulations with a model in which this dependency of LTP was dropped showed that this intuition is correct. In 20 independent simulations, the firing rate was 7.4 ± 2.5 and 3.4 ± 1 Hz for trained and other patterns, respectively. The mean correlation coefficient between the output spike train and the input spike trains of correlated ensembles was 0.38 ± 0.1 for the trained patterns and 0.24 ± 0.05 for other patterns. Finally, the fraction of coincident input output spike pairs was 0.4 ± 0.13 for the trained patterns and 0.14 ± 0.04 for other patterns.

Discussion

Nonlinear dendritic behavior has been proposed as a plausible basis for single-neuron computation (Häusser and Mel, 2003; Mel, 2007; Morita, 2008; Wu and Mel, 2009). However, it is not clear how such functionality could emerge in single neurons in a self-organized manner. The hypothesis that synaptic and dendritic plasticity rules provide the substrate for such self-organization of neuronal behavior is strengthened by recent findings that show, on the one hand, the dependence of synaptic plasticity on local dendritic variables and, on the other hand, plasticity of dendritic excitability (for review, see Sjöström et al., 2008). To our knowledge, the present study provides the first model for how emergence of nonlinear computation could take place in single pyramidal neurons based on experimentally observed plasticity mechanisms. We show that single neurons can in this way learn to solve difficult problems that arise for example in the context of feature binding. Which general principles can be extracted from our results? First, LTD mechanisms that depend on postsynaptic activity via the backpropagating action potential (bAP) can lead to competition between dendritic branches of pyramidal neurons. Such mechanisms use global information conveyed by somatic action potentials that indicate that some parts of the neuron were activated strongly. Second, LTP, which depends linearly on the local dendritic depolarization, supports this competitive mechanism because it consistently favors strongly activated branches over less strongly activated ones. Third, the LTP part of STDP at above-threshold depolarization results (1) in the stabilization of potentiated synapses at highly activated branches despite strong LTD for high postsynaptic firing rates and (2) in the preference of correlated over uncorrelated inputs, although (3) this does not lead to unnecessary potentiation of synapses at weakly activated branches. Finally, BSP can act as a homeostatic mechanism that adapts the branch strength to regulate somatic depolarization for the stored synaptic input pattern.

It has been hypothesized that synaptic clustering, i.e., the clustering of correlated synapses onto dendritic branches, could be important for single-neuron computation (Iannella and Tanaka,

2006; Larkum and Nevian, 2008; Migliore et al., 2008; Morita, 2008). We show that synaptic clustering is an emergent property of the plasticity mechanisms in our model. The clustering results from dendritic competition because correlated input ensembles are potentiated at the winning branch whereas correlated synapses at other branches are depressed. Recent findings indicate that LTP at individual synapses can reduce the threshold for potentiation at neighboring synapses (Harvey and Svoboda, 2007). Although such local interaction between plasticity events may further favor synaptic clustering, our study shows that it is in principle not necessary. An intriguing possibility is that synaptic clusters in our model may interact with micro-rewiring processes, i.e., with local changes in synaptic connectivity (DeBello, 2008). We have assumed in this study a static connectivity pattern. Additional micro-rewiring processes could eliminate weak synapses that contribute little to the branch activation and establish new synaptic connections that contribute to branch depolarization. In contrast, Jia et al. (2010) found through calcium imaging that synaptic inputs onto dendrites of layer 2/3 spiny neurons of mouse visual cortex were not clustered with respect to their orientation selectivity. However, this leaves open the possibility that these inputs were clustered with respect to other task-relevant features.

Relation to other plasticity models
Iannella and Tanaka (2006) showed in a modeling study that, in the presence of two weakly correlated groups of inputs, local dendritic compartments specialize to one of the groups if a variant of STDP is used that is driven by the local dendritic spike instead of the postsynaptic somatic spike. The study also showed that all parts of the dendrite specialize to one of the input groups. This property may, however, lead to an inefficient use of synaptic resources. If a neuron becomes responsive to a certain input pattern, it may be sufficient that a single or a few branches produce dendritic spikes when the pattern is present, whereas other branches are still free to store other patterns. In general, efficient pattern storage requires a competitive mechanism between branches such as the one identified in this study.

The influence of the bAP on STDP was studied theoretically by Tamosiunaite et al. (2007). They also report a form a winner-take-all competition between branches. In contrast to our study, they consider only the case in which different presynaptic ensembles project to different branches.

The competitive nature of STDP on the synapse level has been emphasized by several theoretical and modeling studies (Kempter et al., 1999, 2001; Song et al., 2000). These studies were based on point-neuron models. Taking spatial structure through dendritic compartments into account leads to important new properties. For example, in our correlation-based coding scheme, we can see two levels of competition. On the level of branches, the branches compete for synaptic activation. At the winning branch, we also observe a competition among synapses, because only the correlated synapses are potentiated whereas the strength of other synapses decay.

Assumptions for the plasticity model
Experimentally observed plasticity mechanisms provided the basis for our model for synaptic and dendritic plasticity. Our model for branch strength potentiation (Table 1) is matched to the findings of Losonczy et al. (2008) for elevated ACh levels. In our model for BSP, we assume an additional dependence on the local membrane potential that leads to a favorable homeostatic effect. For the competitive effect demonstrated in this study, however, the voltage dependence of BSP is not necessary. Omitting the

term in the square brackets in Equation 9 and clipping u_k at an upper bound of 2.3 leads to similar results in all simulations except for the homeostasis experiment (Fig. 9B). Losonczy et al. (2008) also reported induction of BSP in the absence of ACh when dendritic spikes coincide with bAPs. Because we assumed elevated ACh levels during learning, we did not model this form of BSP. One can, however, speculate that such a mechanism could contribute to the stabilization of branch strengths in a second phase in which high branch activations lead to increased probability of postsynaptic spiking. Also, interactions between BSP and the bAP are possible, because they share a mechanism for active propagation based on sodium channels. As Remy et al. (2009) showed, this may lead to a cross-desensitization between dendritic spikes and bAPs. Therefore, strong dendritic spiking activity could block subsequent action potential propagation into strong branches but not into weak ones and reduce bAP-dependent plasticity in strong branches, further stabilizing the stored patterns.

We disabled plasticity during testing in our simulations. This is, however, not necessary if the recall stimulus is brief. In our model, plasticity during recall of stored patterns has no effect on the behavior of the neuron. Naturally, other patterns are stored when the recall stimulus is presented long enough. In control simulations identical to those summarized in Figure 8, but with plasticity during testing, we observed very similar performances.

Predictions of the model
Our modeling study provides several predictions that can be experimentally tested. The model predicts that neurons that participate in a neuronal ensemble should excite each other only through one or a few dendritic branches and preferentially lead to dendritic spikes at these branches. This prediction is hard to test with current techniques. Alternatively, one could strongly activate in an experiment several dendritic branches, for example, through glutamate uncaging. Our model predicts that these branches will compete such that, after repeated stimulation, synapses on only a few branches will be potentiated. Furthermore, according to our model, blocking of the bAP should abolish competition leading to potentiated synapses at all stimulated branches. Finally, our simulations showed that a dependence of BSP on the local membrane potential stabilizes neural responses over a wide range of ensemble sizes. It is not yet clear whether the quantitative change in branch strengths depends on dendritic voltages. By measuring dendritic depolarizations during BSP induction, experimental studies could clarify whether BSP depends on the amount of local dendritic depolarization before or during branch spikes.

Conclusions
This study provides for the first time an integrated model for nonlinear dendritic computation with synaptic plasticity and branch-strength potentiation. It shows that single pyramidal neurons can acquire through their synapses nonlinear functionality that is needed for example in the context of feature binding. We have analyzed the role of dendritic competition in the learning dynamics, a novel concept in the analysis of single neuron learning, that emerges from the interplay of local variables and the bAP in the plasticity model. Our model is self-organizing in the sense that the neuron organizes the use of dendritic nonlinearities during pattern presentation without the need of external teacher signals that guide this process. For example, we did not assume that different branches receive input from different input ensembles but rather allowed that all ensembles project to all branches. Still, synaptic efficacies can self-organize on the den-

dritic arbor to cluster stored patterns on individual dendritic branches. This study thus provides evidence that complex nonlinear functions can be acquired by single neurons in a self-organizing manner through biologically plausible local plasticity mechanisms.

References

Abbott LF, Nelson SB (2000) Synaptic plasticity: taming the beast. Nat Neurosci 3:1178–1183.

Abraham WC (2008) Metaplasiticity: tuning synapses and networks for plasticity. Nat Rev Neurosci 9:387–399.

Artola A, Bröcher S, Singer W (1990) Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature 347:69–72.

Auerbach JM, Segal M (1994) A novel cholinergic induction of long-term potentiation in rat hippocampus. J Neurophysiol 72:2034–2040.

Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472.

Billings G, van Rossum MC (2009) Memory retention and spike-timing-dependent plasticity. J Neurophysiol 101:2775–2788.

Branco T, Clark BA, Häusser M (2010) Dendritic discrimination of temporal input sequences in cortical neurons. Science 329:1671–1675.

Bröcher S, Artola A, Singer W (1992) Agonists of cholinergic and noradrenergic receptors facilitate synergistically the induction of long-term potentiation in slices of rat visual cortex. Brain Res 573:27–36.

Clopath C, Büsing L, Vasilaki E, Gerstner W (2010) Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nat Neurosci 13:344–352.

DeBello WM (2008) Micro-rewiring as a substrate for learning. Trends Neurosci 31:577–584.

Dringenberg HC, Hamze B, Wilson A, Speechley W, Kuo MC (2007) Heterosynaptic facilitation of in vivo thalamocortical long-term potentiation in the adult rat visual cortex by acetylcholine. Cereb Cortex 17:839–848.

Fusi S, Abbott LF (2007) Limits on the memory storage capacity of bounded synapses. Nat Neurosci 10:485–493.

Gerstner W, Kistler WM (2002) Spiking neuron models. Cambridge, UK: Cambridge UP.

Golding NL, Staff NP, Spruston N (2002) Dendritic spikes as a mechanism for cooperative long-term potentiation. Nature 418:326–331.

Gu Q (2003) Contribution of acetylcholine to visual cortex plasticity. Neurobiol Learn Mem 80:291–301.

Gustafsson B, Wigström H, Abraham WC, Huang YY (1987) Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. J Neurosci 7:774–780.

Gütig R, Aharonov R, Rotter S, Sompolinsky H (2003) Learning input correlations through non-linear temporally asymmetric Hebbian plasticity. J Neurosci 23:3697–3714.

Hardie J, Spruston N (2009) Synaptic depolarization is more effective than back-propagating action potentials during induction of associative long-term potentiation in hippocampal pyramidal neurons. J Neurosci 29:3233–3241.

Harvey CD, Svoboda K (2007) Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature 450:1195–1200.

Häusser M, Mel B (2003) Dendrites: bug or feature? Curr Opin Neurobiol 13:372–383.

Hebb DO (1949) The organization of behavior. New York: Wiley.

Iannella N, Tanaka S (2006) Synaptic efficacy cluster formation across the dendrite via STDP. Neurosci Lett 403:24–29.

Jia H, Rochefort NL, Chen X, Konnerth A (2010) Dendritic organization of sensory input to cortical neurons in vivo. Nature 464:1307–1312.

Jolivet R, Rauch A, Lüscher HR, Gerstner W (2006) Predicting spike timing of neocortical pyramidal neurons by simple threshold models. J Comput Neurosci 21:35–49.

Kandel ER (2009) The biology of memory: a forty-year perspective. J Neurosci 29:12748–12756.

Kelso SR, Ganong AH, Brown TH (1986) Hebbian synapses in hippocampus. Proc Natl Acad Sci U S A 83:5326–5330.

Kempter R, Gerstner W, van Hemmen JL (1999) Hebbian learning and spiking neurons. Phys Rev E Stat Nonlin Soft Matter Phys 59:4498–4514.

Kempter R, Gerstner W, van Hemmen JL (2001) Intrinsic stabilization of output rates by spike-based Hebbian learning. Neural Comput 13:2709–2741.

Larkum ME, Nevian T (2008) Synaptic clustering by dendritic signalling mechanisms. Curr Opin Neurobiol 18:321–331.

Larkum ME, Nevian T, Sandler M, Polsky A, Schiller J (2009) Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. Science 325:756–760.

Losonczy A, Magee JC (2006) Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. Neuron 50:291–307.

Losonczy A, Makara JK, Magee JC (2008) Compartmentalized dendritic plasticity and input feature storage in neurons. Nature 452:436–441.

Major G, Polsky A, Denk W, Schiller J, Tank DW (2008) Spatiotemporally graded NMDA spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. J Neurophysiol 99:2584–2601.

Makara JK, Losonczy A, Wen Q, Magee JC (2009) Experience-dependent compartmentalized dendritic plasticity in rat hippocampal CA1 pyramidal neurons. Nat Neurosci 12:1485–1487.

Markram H, Lübke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275:213–215.

Mel BW (1994) Information processing in dendritic trees. Neural Comput 6:1031–1085.

Mel BW (2007) Why have dendrites? A computational perspective. In: Dendrites, Ed 2, pp 351–399. New York: Oxford UP.

Migliore M, Novara G, Tegolo D (2008) Single neuron binding properties and the magic number 7. Hippocampus 18:1122–1130.

Minsky M, Papert S (1988) Perceptrons: an introduction to computational geometry. Cambridge, MA: Massachusetts Institute of Technology.

Morita K (2008) Possible role of dendritic compartmentalization in the spatial working memory circuit. J Neurosci 28:7699–7724.

Nevian T, Sakmann B (2006) Spine Ca^{2+} signaling in spike-timing-dependent plasticity. J Neurosci 26:11001–11013.

Nevian T, Larkum ME, Polsky A, Schiller J (2007) Properties of basal dendrites of layer 5 pyramidal neurons: a direct patch-clamp recording study. Nat Neurosci 10:206–214.

Ngezahayo A, Schachner M, Artola A (2000) Synaptic activity modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. J Neurosci 20:2451–2458.

Poirazi P, Brannon T, Mel BW (2003) Pyramidal neuron as two-layer neural network. Neuron 37:989–999.

Ragozzino ME, Unick KE, Gold PE (1996) Hippocampal acetylcholine release during memory testing in rats: augmentation by glucose. Proc Natl Acad Sci U S A 93:4693–4698.

Remy S, Csicsvari J, Beck H (2009) Activity-dependent control of neuronal output by local and global dendritic spike attenuation. Neuron 61:906–916.

Reynolds JN, Wickens JR (2002) Dopamine-dependent plasticity of corticostriatal synapses. Neural Netw 15:507–521.

Roskies AL (1999) The binding problem. Neuron 24:7–9, 111–125.

Sjöström PJ, Häusser M (2006) A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. Neuron 51:227–238.

Sjöström PJ, Turrigiano GG, Nelson SB (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron 32:1149–1164.

Sjöström PJ, Rancz EA, Roth A, Häusser M (2008) Dendritic excitability and synaptic plasticity. Physiol Rev 88:769–840.

Song S, Miller KD, Abbott LF (2000) Competitive Hebbian learning through spike-timing dependent synaptic plasticity. Nat Neurosci 3:919–926.

Tamosiunaite M, Porr B, Wörgötter F (2007) Self-influencing synaptic plasticity: Recurrent changes of synaptic weights can lead to specific functional properties. J Comput Neurosci 23:113–127.

van Rossum MC, Bi GQ, Turrigiano GG (2000) Stable Hebbian learning from spike timing-dependent plasticity. J Neurosci 20:8812–8821.

von der Malsburg C (1981) The correlation theory of brain function. MPI Biophysical Chemistry, Internal Report 81-2. In: Models of neural networks II (1994) (Domany E, van Hemmen JL, Shulten K, eds). Berlin: Springer.

Watt AJ, Desai NS (2010) Homeostatic plasticity and STDP: keeping a neuron's cool in a fluctuating world. Front Synaptic Neurosci 2:5.

Weinberger NM (2008) Cortiacal plasticity in associative learning and memory. In: Cognitive psychology of memory (Learning and memory: a comprehensive reference) (Byrne JH, ed), pp 187–218. Oxford: Elsevier.

Williams SR, Stuart GJ (2002) Dependence of EPSP efficacy on synapse location in neocortical pyramidal neurons. Science 295:1907–1910.

Wu XE, Mel BW (2009) Capacity-enhancing synaptic learning rules in a medial temporal lobe online learning model. Neuron 62:31–41.