

LETTER 

---

 Communicated by Terrence Sejnowski

## Learning to Discriminate Through Long-Term Changes of Dynamical Synaptic Transmission

**Christian Leibold**

*leibold@bio.lmu.de*

**Michael H. K. Bendels**

*bendels@bio.lmu.de*

*Division of Neurobiology, University of Munich, 82152 Planegg-Martinsried, Germany, and Bernstein Center for Computational Neuroscience Munich, 82152 Planegg-Martinsried, Germany*

Short-term synaptic plasticity is modulated by long-term synaptic changes. There is, however, no general agreement on the computational role of this interaction. Here, we derive a learning rule for the release probability and the maximal synaptic conductance in a circuit model with combined recurrent and feedforward connections that allows learning to discriminate among natural inputs. Short-term synaptic plasticity thereby provides a nonlinear expansion of the input space of a linear classifier, whereas the random recurrent network serves to decorrelate the expanded input space. Computer simulations reveal that the twofold increase in the number of input dimensions through short-term synaptic plasticity improves the performance of a standard perceptron up to 100%. The distributions of release probabilities and maximal synaptic conductances at the capacity limit strongly depend on the balance between excitation and inhibition. The model also suggests a new computational interpretation of spikes evoked by stimuli outside the classical receptive field. These neuronal activities may reflect decorrelation of the expanded stimulus space by intracortical synaptic connections.

### 1 Introduction ---

Synapses exhibit reversible activity-induced changes of their responses that decay faster than 1 minute. These changes are summarized as dynamical synaptic transmission or short-term synaptic plasticity (STP) (Zucker & Regehr, 2002). The functional role of STP has mainly been attributed to temporal information processing (Abbott & Regehr, 2004) such as filtering (Fortune & Rose, 2001), gain control (Abbott, Varela, Sen, & Nelson, 1997), and providing a short-term memory buffer for the subsequent readout of temporal activity patterns (Maass, Natschläger, & Markram, 2002; Leibold et al., 2008). Moreover, such dynamical computations via STP can themselves undergo changes that last at least for many tens of minutes (Markram

& Tsodyks, 1996; Gundlfinger et al., 2007) and are usually considered to reflect changes in the probability of release (Tsodyks & Markram, 1997). These changes are considered one type of long-term synaptic plasticity. Quantitative functional hypotheses for how such an interplay between short- and long-term plasticity can be made use of are, however, absent.

On the cognitive and behavioral levels, synaptic plasticity is seen as a basis for learning and memory. However, it is difficult to bridge the gap between cognition and specific features of synaptic physiology from both the theoretical and experimental perspectives. Here, we present a theoretical approach that addresses the ability of sensory systems to learn to discriminate among natural stimuli, which is one of the key components for behavioral performance (Wilson & Stevenson, 2003; Moore, Rosenberg, & Coleman, 2005). A combined recurrent and feedforward (RFF) circuit model (see Figure 1A) is used to demonstrate how the interplay between short- and long-term synaptic plasticity can improve such discrimination learning (Hertz, Krogh, & Palmer, 1991). The synaptic changes induced by the applied learning paradigm can be interpreted in terms of long-term plasticity of the biophysical synaptic properties probability of release and maximum conductance, which are essential for short-term synaptic plasticity.

## 2 Results

---

**2.1 Circuit Model.** To fulfill the minimum requirement for STP, we consider processing of sensory information in a neuronal network during two discrete time steps. The network is organized in  $N$  neuron pools, where a pool is the group of neurons that receive the same sensory input and thus exhibit the same average sensory-driven dendritic depolarization  $x$  (see Figure 1A). We organize these depolarizations in a vector  $\mathbf{x} = (x_1, \dots, x_N)^T$  representing the inputs to the whole circuit. To obtain a simple example for natural stimulus statistics, inputs  $x$  are derived from natural images (van Hateren & van der Schaaf, 1998) (see the appendix) by means of independent component analysis (ICA) (Bingham & Hyvärinen, 2000). The depolarization  $x_n$  of the  $n$ th pool thus indicates the strength of the  $n$ th independent component in the present (image) stimulus and assigns each pool to one independent component. The decorrelation by ICA can be interpreted biologically as the computation performed by the feedforward synapses to the recurrent network. These feedforward weights may have been learned during early experience but remain fixed during discrimination learning (adulthood). Instead of ICA, these synapses might as well implement any other algorithm of decorrelation (Olshausen & Field, 1997).

In response to  $\mathbf{x}$ , the pools exhibit a feedforward-driven activity at time step 1, which is denoted by the vector

$$\mathbf{y}(1) = (y_1(1), \dots, y_N(1))^T = (\sigma(x_1), \dots, \sigma(x_N))^T. \quad (2.1)$$

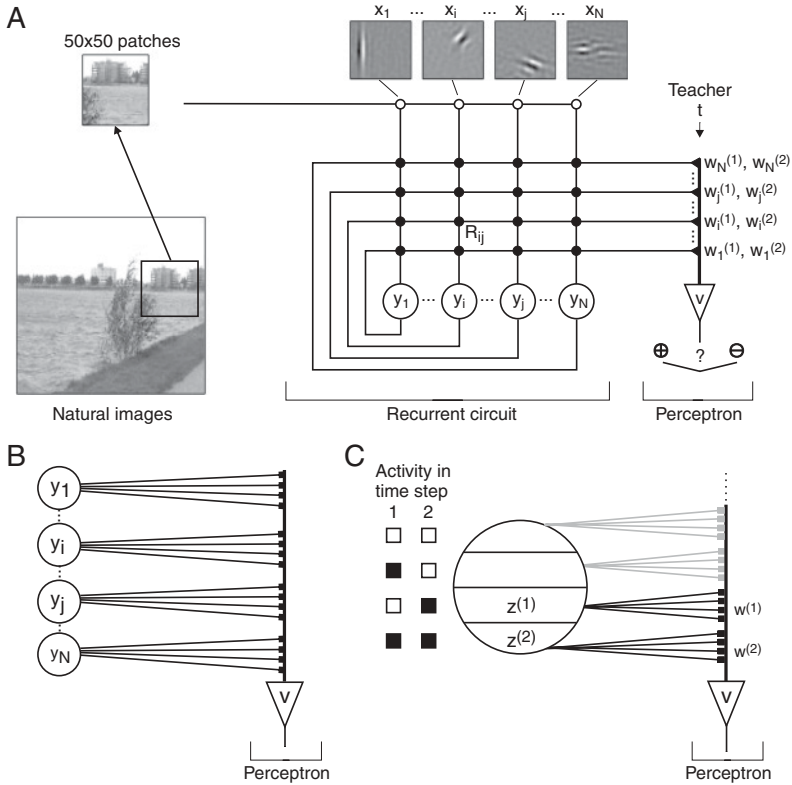


Figure 1: Hybrid recurrent and feedforward (RFF) model for discriminating natural stimuli. (A) RFF network that receives  $50 \times 50$  input patches derived from a natural image database (van Hateren & van der Schaaf, 1998). Feedforward synapses (white circles) are fixed and perform decorrelation via independent component analysis (ICA). The depolarization  $x_i$  of neuron pool  $i$  corresponds to the contribution of one specific independent component. The percentage of neurons firing in response to this depolarization is denoted by  $y_i$ . Recurrent synapses  $R_{ij}$  (black dots) between pools  $j$  and  $i$  transform the sensory-driven output pattern  $\mathbf{y}(1)$  at time step 1 to a recurrently computed output pattern  $\mathbf{y}(2)$  at time step 2. Teacher inputs  $t$  are used for learning to classify images into  $\oplus$  and  $\ominus$  image by changing the probability of release and the maximal synaptic conductance of the synapses (black triangles) to a perceptron-like readout neuron. (B) A pool  $i = 1, \dots, N$  contains many neurons, each of which sends synapses to the readout neuron. (C) A fraction  $z^{(1)} = y(2) [1 - y(1)]$  of neurons has fired only at time step 2. The effective weight of these synapses is denoted by  $w^{(1)}$ . The fraction  $z^{(2)} = y(2) y(1)$  of neurons that have fired at both time steps is connected to the readout neuron with an effective weight  $w^{(2)}$ . Neurons that do not fire at time step 2 are assumed not to contribute to the discrimination task (gray connections).

Here  $\sigma$  is a sigmoid activation function that measures the percentage of activated neurons in a pool for a given mean depolarization  $x$ . As such, it captures all the biophysical dynamics arising from the different sorts of dynamics of interneurons, intrinsic recurrent couplings, and so forth. Throughout this letter we use  $\sigma(x) = [\tanh(\beta x) + 1]/2$ , in which  $\beta$  determines the steepness of the activation function of the pools: for large  $\beta$ , the sigmoid activation function is more steplike, and for small  $\beta$ ,  $\sigma$  becomes more linear. To adjust the sensitive range of  $\sigma$  to the variability of the inputs, we normalize each stimulus vector to  $|\mathbf{x}| = 1$ , which implies the assumption of a preprocessing step that achieves intensity invariance.

The depolarizations  $\xi$  of the neurons in the next time step are assumed to be determined solely by the recurrent connections. Assuming linear summation of the recurrent activities, we have

$$\xi = \mathbf{R} [\mathbf{y}(1) - \mathbf{y}_0], \quad (2.2)$$

in which  $\mathbf{R}$  is a zero-mean random matrix of recurrent weights with diagonal elements set to zero since self-couplings within the pools are considered to be accounted for by the activation function  $\sigma$ . A model variant in which the input  $\mathbf{x}$  is also present at time step 2 is discussed further below. The percentage of activated neurons due to the recurrent computation at time step 2 is denoted by the activity vector  $\mathbf{y}(2) = (\sigma(\xi_1), \dots, \sigma(\xi_N))$ . To this end, the constant vector  $\mathbf{y}_0 = \frac{1}{2}(1, \dots, 1)^T$  accounts for  $\xi$  to be adjusted to the sensitive range of  $\sigma$ . The latter may be achieved by a homeostatic process that tunes the balance between excitation and inhibition.

The last stage of the circuit is a feedforward projection to a readout neuron (or pool), whose depolarization  $v$  is imparted by synaptic weights that exhibit short-term plasticity. To also allow for supervised learning of these synapses, the readout unit must have additional access to a top-down teacher signal  $t$  (see Figure 1A). For now, we consider the readout to be driven only by the activity pattern  $\mathbf{y}(2)$  at time step 2. The effect of temporally integrating the activity pattern  $\mathbf{y}(1)$  is studied further below. Thus, a pool  $n$  that was activated only at time step 2 contributes a postsynaptic depolarization  $w_n^{(1)} y_n(2)$ , whereas a pool  $m$  that was activated at both time steps contributes a postsynaptic depolarization  $w_m^{(2)} y_m(2)$ , where, due to STP, the respective weights  $w_m^{(1)}$  and  $w_m^{(2)}$  are generally different. The depolarization  $v$  at time step 2 is then constructed via a linear combination of the two contributions,

$$v(\mathbf{x}) = \sum_{n=1}^N y_n(2) [w_n^{(1)} (1 - y_n(1)) + w_n^{(2)} y_n(1)], \quad (2.3)$$

in which the weights are multiplied with the respective fractions of neurons that were or were not active at time step 1,  $y_n(1)$  and  $(1 - y_n(1))$ , respectively

(see Figures 1B and 1C). Equation 2.3 can be simplified by introducing

$$z_n^{(1)} = y_n(2) [1 - y_n(1)] \quad (2.4)$$

as the percentage of active neurons in pool  $n$  that have fired at time step 2 but did not fire at time step 1, and

$$z_n^{(2)} = y_n(2) y_n(1) \quad (2.5)$$

as the percentage of active neurons in pool  $n$  that fired at both time steps. Combining equations 2.3 to 2.5 then yields

$$v(\mathbf{x}) = \sum_{n=1}^N [w_n^{(1)} z_n^{(1)} + w_n^{(2)} z_n^{(2)}]. \quad (2.6)$$

The mapping from the activation variables  $\mathbf{y}(1)$  and  $\mathbf{y}(2)$  to the synaptic state vectors  $\mathbf{z}^{(1)}$  and  $\mathbf{z}^{(2)}$  thus provides a nonlinear expansion of the  $N$ -dimensional inputs  $\mathbf{x}$  to the space  $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}\}$  with  $2N$  dimensions.

It is one of the hallmarks of the theory of neural networks that expanding the inputs into a higher-dimensional space generally improves discrimination learning (Hertz et al., 1991; Gardner, 1987; Vapnik, 1998; see Figure 2A). However, since correlations in the inputs deteriorate discrimination learning (Shinzato & Kabashima, 2008), such an improvement additionally requires that the data in the expanded input space be decorrelated (see Figure 2B). This decorrelation can be visualized via the spectrum of eigenvalues of the  $2N \times 2N$  covariance matrix of the expanded inputs. Figures 2C to 2F show these eigenvalue spectra obtained with several two-dimensional expansions of the activities  $y_n$ . The distributions have two peaks: a pronounced one close to zero and a second more dispersed peak at higher values. Both peaks contain roughly half the eigenvalues (not shown). Zero eigenvalues can be interpreted as correlated or “lost” dimensions. The data can thus be considered less correlated the farther this first peak is away from zero. As a quantitative measure, we therefore introduce the center of gravity  $r$  of the first peak. For a purely feedforward nonlinear expansion,  $r$  is very small; thus, the “effective” input space is close to  $N$ -dimensional. When a random recurrent coupling matrix  $\mathbf{R}$  is employed via equations 2.4 and 2.5, the first peak moves toward more positive values, and  $r$  increases. The examples shown further suggest that the stronger the variance  $\kappa^2$  of  $\mathbf{R}$ , the more the data are decorrelated (see section 2.2).

To understand the dynamics of the recurrent circuit, we assessed the receptive fields of the neuron pools at the different time steps by a reverse correlation analysis of activities  $\mathbf{y}(1)$  and  $\mathbf{y}(2)$  using white noise images (see Figure 3). As expected, activities  $\mathbf{y}(1)$  at time step 1 very much resemble

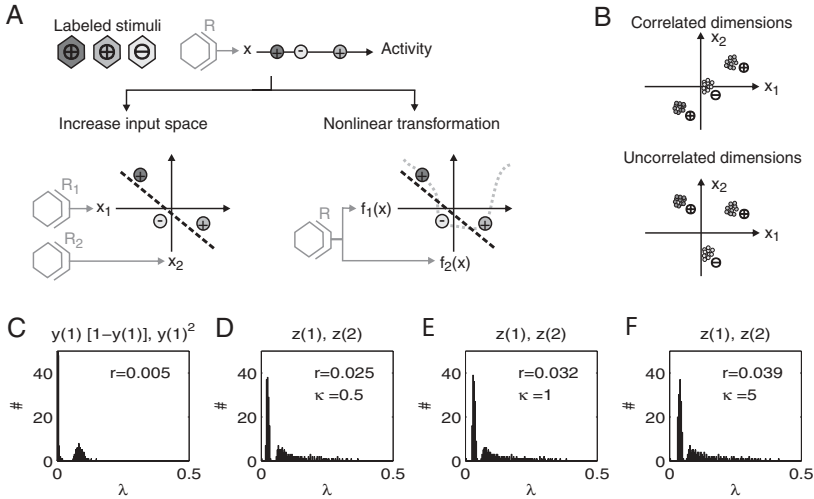


Figure 2: Nonlinear expansion and decorrelation. (A) Top: With one stimulus dimension  $x$  (e.g., activity of a receptor  $R$ ), there is no general solution for a linear separation problem of three arbitrarily chosen inputs (different gray levels), that carry labels ( $\oplus, \ominus$ ), that is, it is impossible to define a threshold value that is exceeded for all  $\oplus$  patterns but not for  $\ominus$  patterns. Left: Separability can be improved by a higher sensory resolution (e.g., two receptors  $R_1$  and  $R_2$ ). A possible linear classification boundary is depicted by a dashed line. Right: A different strategy to achieve linear separability is to introduce a nonlinear mapping  $f_1, f_2$  (dotted line) into a higher-dimensional space. (B) Linear separation of (noisy) stimuli (clusters with different gray levels) is hampered if the sensory activity is highly correlated between input dimensions (top). Uncorrelated input dimensions (bottom) allow linear separation. (C) Distribution of eigenvalues  $\lambda$  for the covariance matrix of the 2  $N$ -dimensional inputs obtained by the nonlinear expansion  $\{y_n(1)[1 - y_n(1)], y_n(1)^2\}$ , which would result from feedforward short-term synaptic plasticity. Here it is assumed that the same activity  $y(1)$  is present at the two consecutive time steps. Thus,  $y_n(1)[1 - y_n(1)]$  accounts for the fraction of synapses that are activated only at time step 2 and  $y_n(1)^2$  is the fraction of synapses activated at both time steps. (D) Distribution of eigenvalues for the expansion into the space  $z^{(1)}, z^{(2)}$  from equations 2.4 and 2.5. The recurrent weight matrix  $\mathbf{R}$  was generated from a zero-mean gaussian random process with standard deviation  $\kappa = 0.5$ . (E) Same as in D with  $\kappa = 1$ . (F) Same as in D with  $\kappa = 5$ . All data were obtained for  $\beta = 1$ .

the independent components that are represented by the input depolarizations  $x$  (correlations coefficients  $> 0.98$ ). At time step 2, the receptive fields are decorrelated (correlation coefficients  $< 0.1$ ). Experimentally measured receptive fields would correspond to the (temporal) average over both time

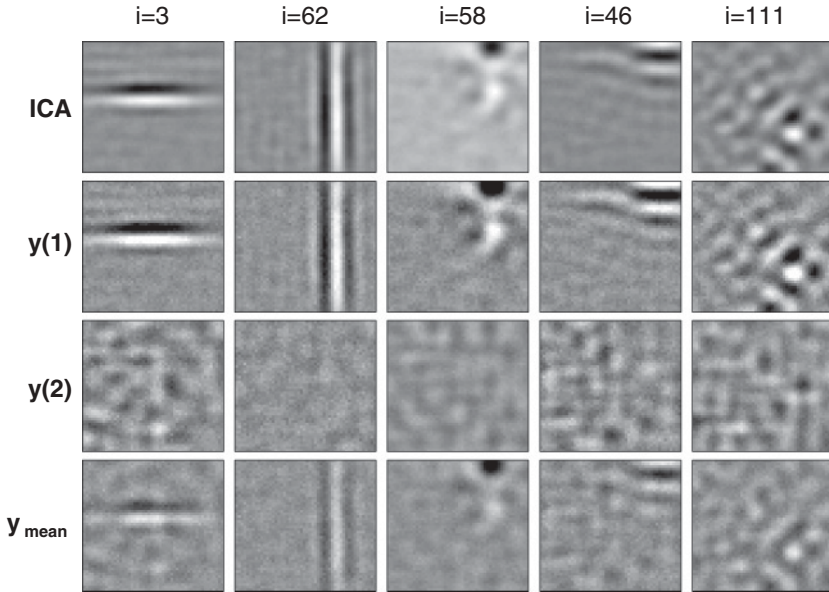


Figure 3: Receptive fields of the RFF circuit. (Top) Independent components of five exemplary pools. (Two middle rows) Receptive fields obtained through reverse correlation of the activities at time steps 1 and 2, respectively. Averaging is done over 200,000 white noise stimuli; the variance of  $R$  was  $\kappa = 2$ . (Bottom) Temporal average  $\frac{y(1)+y(2)}{2}$  of the receptive fields. Bright colors depict high and dark colors low responsiveness. Data were obtained for  $\beta = 1$ .

steps, which still largely reflect the original receptive field (correlation coefficients  $> 0.8$ ). With respect to cortical information processing, this scheme can be interpreted such that early spikes that are evoked by stimuli in the classical receptive field are useful to reliably transmit information to higher centers. In contrast, later spikes reflect complex computations such as decorrelation.

**2.2 Discrimination of Natural Images.** To test the function and evaluate the capacity of the RFF circuit (see Figure 1) for discriminating among natural images, we measured the discrimination performance of the RFF circuit and compared it to that of other related networks (see Figure 4A). The depolarization  $v$  of the readout unit at the second time step was trained to be a decision variable for a discrimination task in the input space of coefficients  $\mathbf{z}^{(i)}$  such that input images can be classified into a  $\oplus$  and a  $\ominus$  category (corresponding to the teacher signals  $t \in \{+1, -1\}$ , respectively). Such a setting is the standard discrimination paradigm for a perceptron.

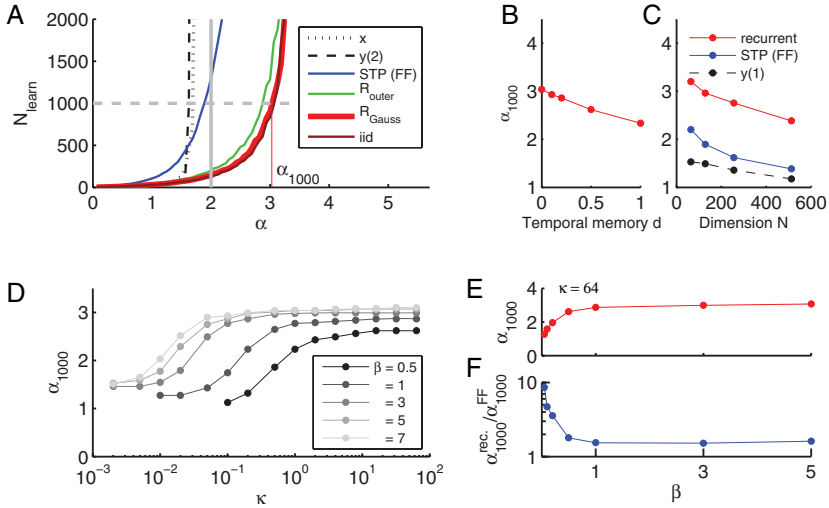


Figure 4: Discrimination learning in the RFF circuit. (A) The number  $N_{\text{learn}}$  of learning epochs required to achieve 100% performance on a discrimination task as a function of the load  $\alpha = P/N$ . Black dots: Perceptron rule (PR) applied to the sensory inputs  $x$  ( $N$  dimensions). Black dashed line: PR applied to the activity vector  $y(2)$  at time step 2 ( $N$  dimensions). Blue solid line: PR applied to the input space  $\{y(1)[1 - y(1)], y(1)^2\}$  from feedforward short-term synaptic plasticity ( $2N$  dimensions). Red line: PR applied to the input space ( $2N$  dimensions) obtained by recurrent decorrelation with gaussian  $R$ . Green line: Same as red line for recurrent decorrelation with the the outer product matrix  $R$  from equation 2.7. Brown line: Same as red line for independent and identically distributed random inputs instead of natural images. Data were obtained with  $\beta = 5$  and  $\kappa = 5$  and averaged over  $200/\alpha$  repetitions (see the appendix). The gray vertical line at  $\alpha = 2$  indicates the capacity limit of a perceptron for  $N \rightarrow \infty$ . (B) Capacity criterion  $\alpha_{1000}$  decreases with increasing memory  $d$  of the activity  $y(1)$  that remains from time step 1. (C) Capacity criterion  $\alpha_{1000}$  at which 1000 learning epochs are required to achieve a performance of 100% correct decisions as a function of the number  $N$  of dimensions. (D) Capacity criterion  $\alpha_{1000}$  increases with both the variance  $\kappa^2$  of the random recurrent weights and the steepness  $\beta$  of the activation function  $\sigma$ . (E) Capacity criterion  $\alpha_{1000}$  for  $\kappa = 64$  increases with  $\beta$ . (F) Improvement due to recurrence. The ratio between  $\alpha_{1000}$  obtained with recurrent couplings ( $\kappa = 64$ ) and  $\alpha_{1000}$  obtained within a feedforward setting: The improvement with respect to the STP-mediated expansion (blue line) is always greater than about 1.5.

The depolarization  $v$  of the readout unit determines its output  $O$  in that  $O = +1$  if  $v \geq w_0$  and  $O = -1$  if  $v < w_0$ . In response to  $O$ , the weights  $w^{(k)}$  as well as the threshold  $w_0$  were changed according to the standard perceptron learning rule (Hertz et al., 1991): if upon presentation



of the input image  $\mathbf{x}^\mu$  the readout unit gave a correct response ( $O^\mu = t^\mu$ ), the weights remained unchanged. If the  $O^\mu$  was different from  $t^\mu$ , the weights were updated by  $\Delta w_n^{(k)} = \eta t^\mu z_n^{(k)\mu}$ , in which  $\mathbf{z}^{(k)\mu}$  were the coefficients derived from the input  $\mathbf{x}^\mu$  via equations 2.1, 2.4, and 2.5. The constant  $\eta > 0$  is called the *learning rate* (see the appendix). A biologically more plausible implementation of such a learning rule is discussed in section 2.3.

To measure the performance of the learning rule, we define a learning epoch as one consecutive presentation of all  $P$  input images and the corresponding weight changes. The number  $N_{\text{learn}}$  of learning epochs that are required to achieve a performance of 100% correct decisions is an increasing function of the number  $P$  of learned inputs and therefore also the load  $\alpha = P/N$  (see Figure 4A).

The performance of the learning algorithm is compared for different networks employing the capacity criterion  $\alpha_{1000}$  at which, on average, 1000 learning epochs are required to achieve a performance of 100% correct decisions. For input spaces with  $N = 128$  dimensions as the ICA input vectors  $\mathbf{x}$  or the network activities  $\mathbf{y}(2)$ , the capacity criterion  $\alpha_{1000}$  ranges between 1 and 2 (see Figure 4A). Note that the theoretical capacity limit of  $\alpha = 2$  cannot be reached since this is an asymptotic value for  $N \rightarrow \infty$  (Hertz et al., 1991).

If we construct a  $2N$ -dimensional input space by nonlinear expansions via STP, the capacity criterion increases beyond 2. In particular, if we combine the nonlinear expansion via STP and the decorrelation by recurrent synaptic connections, the criterion increases to about 3 (see the red line in Figure 4A). To check for the generality of the obtained results, we also conducted simulations with independent and identically distributed (i.i.d.) random inputs that were uniformly distributed with zero mean and variance 1. Learning with these i.i.d. inputs exhibits a similar time course as for the ICA inputs; it is even slightly faster (see the brown line in Figure 4A). Hence, discrimination learning of the RFF circuit on natural images is virtually as good as for ideally uncorrelated noise inputs.

In equation 2.2, we assumed the feedforward input  $\mathbf{x}$  to occur only at time step 1. For static inputs, however, it is a reasonable assumption that the input is present at both time steps. In terms of the depolarization  $\xi$  at time step 2, this would require changing equation 2.2 to

$$\xi = \mathbf{R}[\mathbf{y}(1) - \mathbf{y}_0] + \mathbf{x}.$$

Such a static input pattern could easily induce correlations between activity patterns  $\mathbf{y}(1)$  and  $\mathbf{y}(2)$ . Using this model variant, we therefore tested whether these correlations can be removed by an appropriately chosen recurrent coupling matrix. Assuming the activity patterns  $\mathbf{y}(1)$  to be approximately

uncorrelated,

$$\sum_{j=1}^N [y_j^\mu(1) - y_0] [y_j^\nu(1) - y_0] \approx \delta_{\mu\nu} |\mathbf{y}^\mu(1) - \mathbf{y}_0|^2,$$

a natural choice for such a coupling matrix is the outer product,

$$(\mathbf{R})_{ij} = \kappa \sum_{\mu=1}^P [\Xi_i^\mu - x_i^\mu] \frac{y^\mu(1)_j - y_0}{|\mathbf{y}^\mu(1) - \mathbf{y}_0|^2}. \quad (2.7)$$

Here,  $\Xi^\mu$  ( $\mu = 1, \dots, P$ ) denote  $P$  given random patterns (mean: 0, variance 1) that are uncorrelated to the inputs  $\mathbf{x}^\mu$ . As shown in Figure 4A (green trace), the criterion  $\alpha_{1000} = 2.85$  obtained with the matrix from equation 2.7 is only slightly worse than that obtained by decorrelating with a gaussian random matrix. This means that at least in principle, one can find a learning rule for the recurrent connections that decorrelates the activity patterns  $\mathbf{y}(1)$  and  $\mathbf{y}(2)$  even if the input is present at both time steps.

To evaluate the effect of subthreshold temporal integration at the readout unit, we next investigated the nonlinear expansion  $z_n^{(1)} = y_n(1)d + y_n(2)[1 - y_n(1)]$ , in which  $d$  measures the amount of depolarization that remains from time step 1. We find that the capacity criterion  $\alpha_{1000}$  decreases with increasing  $d$  (see Figure 4B). This means that if the activity pattern of the first time step is still partly present at the second time step owing to slow synaptic integration, discrimination learning is less efficient.

Next, we tested the generality of the used capacity criterion  $\alpha_{1000}$  by running simulations for different input dimensions  $N$  and found that  $\alpha_{1000}$  decreases with the number of dimensions (see Figure 4C). This is because the time to convergence increases with the number of input dimensions and the learning rule has not fully converged after 1000 epochs. Nevertheless,  $\alpha_{1000}$  is good enough to clearly discriminate between the learning curves for the investigated models. A capacity estimate that is closer to the theoretical limit might be obtained by increasing the number of permitted learning epochs beyond 1000. This, however, would heavily increase the required computing time. Moreover, since we considered only a small, finite  $N$ , not all learning problems may converge as  $\alpha$  gets closer to the capacity limit. Thus, with a higher number of permitted epochs, one would have to give up the 100% correct criterion. To conclude,  $\alpha_{1000}$  is a reasonable choice for a capacity criterion for the practical purposes of this study; however, it does not allow for comparison between different input dimensions.

In a further series of simulations, we studied how discrimination learning depends on the circuit parameters  $\kappa$  and  $\beta$ . We found that  $\alpha_{1000}$  increases with the variance  $\kappa^2$  of the random weights in  $\mathbf{R}$ , but also with the steepness  $\beta$  of the activation function  $\sigma$  (see Figures 4D and 4E). The improvement

of the discrimination performance due to recurrent couplings was measured via the ratio between the criteria  $\alpha_{1000}$  with and without recurrent couplings (see Figure 4F). As compared to the  $2N$ -dimensional input space  $\{y(1)[1 - y(1)], y(1)^2\}$  motivated by feedforward short-term synaptic plasticity, additional recurrence yielded at least a 1.5-fold improvement.

**2.3 Interconnecting Short- and Long-Term Synaptic Plasticity.** The synaptic weights  $\mathbf{w}^{(1)}$  and  $\mathbf{w}^{(2)}$  are properties of the same synapse. For training the depolarization  $v$  of the readout unit on a discrimination task, we have to make sure that these weights can be altered independently by long-term modifications. We therefore employ a biophysically motivated model in which the weight  $w_n^{(i)}$  is a product of the following three quantities: the probability  $P_n$  of release, the (relative) number  $R_n$  of readily releasable vesicles, and the postsynaptic amplitude  $Q_n$  on release of a vesicle. The last captures both the quantal size and the amount and sensitivity of postsynaptic receptors. Then the (effective) weights can be written as

$$w_n^{(k)} = P_n^{(k)} R_n^{(k)} Q_n^{(k)} - g, \quad (2.8)$$

in which the  $g > 0$  introduces a negative contribution  $-g y_n(2)$  to the membrane potential  $v(\mathbf{x})$  (see equation 2.3). This inhibition is proportional to the activation  $y_n(2)$  of the pool at time step 2 and can therefore be interpreted as a constant, pool-specific, feedforward inhibition.

Thus, the weights  $\mathbf{w}^{(1)}$  upon transmission of the first pulse are  $w_n^{(1)} = P_n^{(1)} R_n^{(1)} Q_n^{(1)} - g = p_n a_n - g$ , in which  $p_n$  denotes the baseline probability of release  $P_n^{(1)}$  and the maximal synaptic conductance  $a_n$  is the product of the baseline values  $R_n^{(1)}$  and  $Q_n^{(1)}$ .

If the synapse transmits a second pulse, the probability of release is enhanced to  $P_n^{(2)} = F(P_n^{(1)})$  due to facilitory processes such as residual calcium (Zucker & Regehr, 2002; Felmy, Neher, & Schneggenburger, 2003). Figure 5A depicts the facilitation model  $F(p)$  (see the appendix). The second factor that determines the synaptic amplitude at the second pulse is the size  $R^{(2)}$  of the pool of vesicles that are releasable. Assuming that the replenishment of vesicles is much slower than the interpulse interval, the pool size is reduced to  $R_n^{(2)} = R_n^{(1)} - p_n R_n^{(1)}$  owing to the depletion of vesicles by the first pulse. This depletion reduces the amount of released transmitter and accounts for short-term synaptic depression. For simplicity, we furthermore assume the amplitude  $Q$  to be unchanged:  $Q^{(2)} = Q^{(1)}$  (see section 3). Thus the weights  $\mathbf{w}^{(2)}$  on transmission of the second pulse are  $w_n^{(2)} = P_n^{(2)} R_n^{(2)} Q_n^{(2)} - g = F(p_n)(1 - p_n)a_n - g$ .

Combining the expressions for  $w^{(1)}$  and  $w^{(2)}$ , we can derive the excitatory paired-pulse ratio as  $\rho(p) = F(p)(1 - p)/p$  (see Figure 5B). Since  $\rho$  covers a range between 0 (for  $p = 1$ ) and its maximum (for  $p \rightarrow 0$ ), both weights  $w_n^{(1)}$  and  $w_n^{(2)}$  can be adjusted independently in this range by manipulating

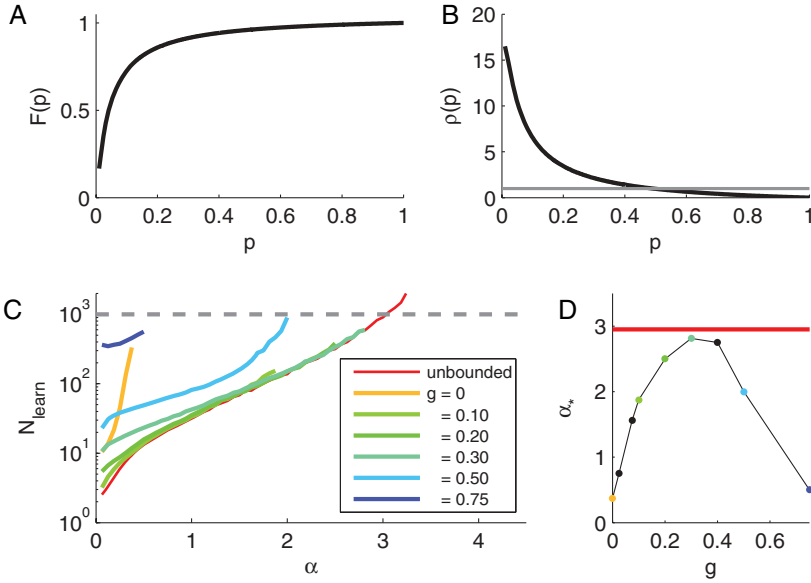


Figure 5: Discrimination learning and short-term synaptic plasticity. (A) Facilitation  $F$  as a function of the initial release probability  $p$  from equation A.1. (B) Excitatory paired-pulse ratio  $\rho(p) = [w^{(2)} + g]/[w^{(1)} + g] = F(p)(1 - p)/p$  as a function of the initial release probability  $p$ . The gray line indicates  $\rho = 1$ . (C) Number  $N_{\text{learn}}$  of learning epochs required for 100% correct performance as a function of the load ( $\alpha = P/N$ ) for the standard perceptron rule from Figure 4A (red line) and the learning rules for  $a$  and  $p$  from equations 2.9 and 2.10 (yellow to blue) for different values  $g$  of the feedforward inhibition. Simulations were terminated if more than 5% of the repetitions failed to converge. (D) The load  $\alpha_*$  at which the fraction of nonconvergent repetitions exceeds 5% as a function of  $g$ . The red line depicts the capacity criterion  $\alpha_{1000}$  obtained with the standard perceptron rule. Colors correspond to  $g$  values from C.

the synaptic parameters  $p_n$  and  $a_n$ . For small  $p$ ,  $\rho$  is larger than 1, and the synapse exhibits short-term facilitation; for large  $p$ ,  $\rho$  is smaller than 1, and the synapse exhibits short-term depression.

Synaptic learning requires implementing long-term changes of the two parameters  $p_n$  and  $a_n$ . Here we derive update rules for these two quantities from the standard perceptron learning rule that provide the weight change  $\Delta w_n^{(k)}$  on presentation of a stimulus teacher pair  $(\mathbf{x}^\mu, t^\mu)$ . All results of the letter are derived for the standard learning rule (Hertz et al., 1991), that is,  $\Delta w_n^{(k)} = 0$  if the output unit responds correctly and  $\Delta w_n^{(k)} = \eta t^\mu z^{(k)\mu}$  if the unit responds incorrectly.

The basic idea is to map the weight changes differentially to parameter changes. From  $w_n^{(1)} = p_n a_n$  and  $w_n^{(2)} = p_n \rho(p_n) a_n$ , we obtain the differential

changes

$$\begin{aligned}\Delta w^{(1)} &= a \Delta p + \Delta a p \\ \Delta w^{(2)} &= [a \rho + a p \rho'(p)] \Delta p + \Delta a \rho p.\end{aligned}$$

These two equations connect the two unknowns  $\Delta a$  and  $\Delta p$  to the two weight changes  $\Delta w_n^{(k)}$ . The solution of this two-dimensional system of equations provides the learning rules

$$\Delta a_n = \eta t^\mu \frac{z_n^{(1)\mu} [\rho(p_n) + p_n \rho'(p_n)] - z_n^{(2)\mu}}{\rho'(p_n) p_n^2} \quad (2.9)$$

$$\Delta p_n = \eta t^\mu \frac{z_n^{(2)\mu} - \rho(p_n) z_n^{(1)\mu}}{\rho'(p_n) p_n a_n}. \quad (2.10)$$

In equations 2.9 and 2.10,  $\rho'(p)$  denotes the derivative of the paired-pulse ratio (see Figure 5B). Since  $p$  is a probability, it is restricted to values between 0 and 1. The maximum conductance  $a$  is constrained to values between 0 and 1 to meet Dale's law and to introduce an upper limit for the maximum synaptic input.

To avoid divergences in the update rules of equations 2.9 and 2.10, we restrict the probabilities  $p$  to values even larger than 0.01 since then (for the present choice of  $F(p)$ ) the derivative  $\rho'(p)$  is nonzero. Moreover, we avoid a further divergence in equation 2.10 by not updating  $p$  if  $a = 0$ . This is reasonable, since for  $a_n = 0$ , all weights  $w_n^{(k)}$  are zero anyway.

We note that all quantities in these learning rules can be interpreted as local—specific to the presynaptic pool for which the synaptic weight is about to change. Moreover, since  $z_n^{(k)} \propto y_n(2)$ , all changes  $\Delta a_n$  and  $\Delta p_n$  from equations 2.9 and 2.10 are proportional to the pool's activity  $y_n(2)$  at the second time step. This means that one can assume only those synapses to change at which the neuron was active at time step 2. In this sense, the update rules can also be considered local to a specific synapse.

To assess the performance of the learning rules for  $a$  and  $p$ , we have run the same simulation as for the standard learning rule shown in Figure 4A. The obtained learning curves exhibit two major differences (see Figure 5C). First, the learning rule no longer always converges. This can be understood since  $p_n$  as a probability is bounded and  $a_n$  is limited to positive values. As a consequence,  $w^{(1)}$  and  $w^{(2)}$  can no longer take all possible values, and hence not all problems can be separated. We thus require introducing a new capacity criterion, the terminal load  $\alpha_*$ , at which the error rate defined as the fraction of nonconverging repetitions increases beyond 5%. Second, the capacity criterion depends on the magnitude  $g$  of feedforward inhibition. However, there exists an optimal  $g \approx 0.3$  (see Figure 5D) for which the

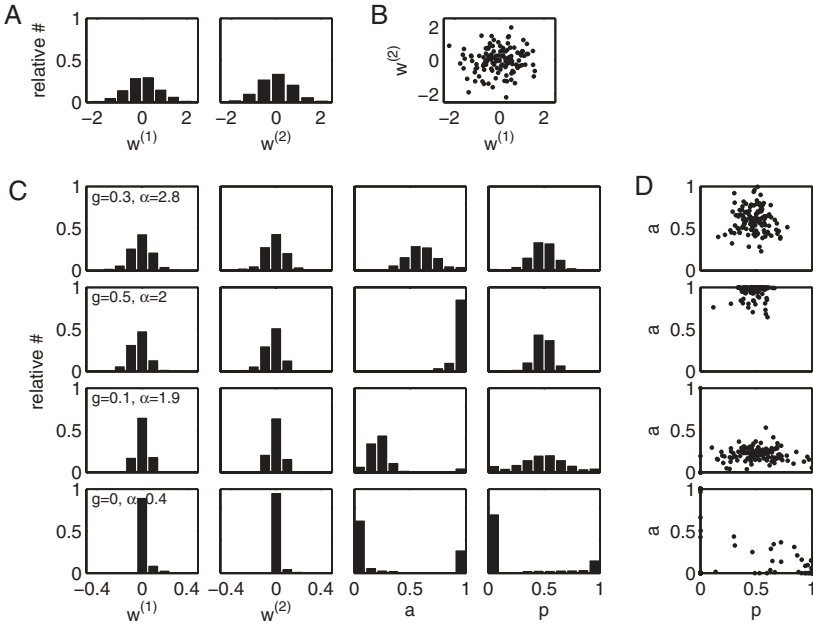


Figure 6: Distributions of synaptic parameters. (A) Distribution of the synaptic weights  $w^{(1)}$ , and  $w^{(2)}$  for the standard learning rule from Figure 4. (B) Scatter plot from one example simulation. (C) Distributions of the synaptic weights  $w^{(1)}$  and  $w^{(2)}$  and the respective synaptic parameters  $p$  and  $a$  at  $\alpha = \alpha_*$  for four values of  $g$ . (D) Scatter plots from one example simulation each. Distributions in A and C were derived from 20 converging repetitions.

learning curve  $N_{\text{learn}}(\alpha)$  is virtually identical to the standard case. The terminal load  $\alpha_*$  at this optimal  $g$  is 95% of the capacity criterion  $\alpha_{1000}$  obtained with the standard learning rule from Figure 4. The terminal load for  $g = 0$  is about 0.4, which is similar to values reported for a sign-constrained perceptron with binary inputs (Amit, Compbell, & Wong, 1989; Brunel, Hakim, Isope, Nadal, & Barbour, 2004). More rigorous upper bounds for  $\alpha_*$  are hard to find, since they strongly depend on specific properties of the input set (Legenstein & Maass, 2008).

Next we studied the distribution of synaptic parameters close to the capacity limit. For the standard perceptron rule, both weights  $w^{(1)}$  and  $w^{(2)}$  are distributed around zero (see Figures 6A and 6B). With the learning rules for  $a$  and  $p$  and values of  $g$  at which  $\alpha_*$  is larger than 1.75 (e.g.,  $g = 0.1, 0.3, 0.5$ ), the weight distributions are similar to the ones from the standard rule (see Figures 6C and 6D). The realization of these distributions in terms of synaptic parameters  $p$  and  $a$ , however, is different. Whereas the distribution of  $p$  remains clustered around 0.5, the distribution of  $a$  peaks at low values for small inhibition  $g$  or clusters around 1 for high values of

*g*. If the inhibition becomes too small (e.g.,  $g = 0$ ) to realize a reasonable terminal load  $\alpha_*$ ,  $p$  and  $a$  can no longer realize weight distributions that are as broad as observed in the standard perceptron case; instead,  $\mathbf{w}^{(1)}$  and  $\mathbf{w}^{(2)}$  are tightly clustered around zero and take only positive values (see Figures 6C and 6D bottom row; see also Köhler & Widmaier, 1991; Brunel et al. (2004)). The corresponding distributions of  $p$  and  $a$  are bimodal, with most values close to the bounds of the permitted parameter range: 0 and 1. The peak at  $w = 0$  for sign-constrained perceptrons has recently been interpreted as to arise from the population of the so-called silent synapses (Brunel et al., 2004), which lack postsynaptic AMPA receptors. In our model, silent synapses would be only those that constitute the peak at  $a = 0$ .

### 3 Discussion

---

We have presented a neuronal network model that utilizes long-term changes of short-term synaptic plasticity to improve the performance of learning to discriminate between natural stimuli (see Figure 1A). A neuronal system's capability of learning to discriminate stimuli is generally limited by the dimension of the input space of the learning problem. There, Gardner's bound (Gardner, 1987) sets the theoretical limit for the number of discriminable stimuli to twice the dimension of the problem. A sufficient discriminability of the sensory code can thus be achieved by simply providing a large enough sensory resolution. However, the relative gain of represented information becomes smaller with the increasing number of sensory receptors. A different strategy to improve discriminability is to increase the dimension of the learning problem through an additional nonlinear transformation. Such a nonlinear expansion is widely used in the theory of statistical learning (Vapnik, 1998) and has recently been adopted by computational neuroscience, where it is known as a liquid state machine (Maass et al., 2002) or echo state network (Jaeger & Haas, 2004).

This letter shows that short-term synaptic plasticity can provide such an expansion of the input space to twice the original dimension. Here, this idea has been illustrated considering two consecutive time steps. At the first time step, sensory information primes the feedforward synapses, adjusting their state of facilitation and partly depleting their vesicle pool. At the second time step, computations via recurrent synaptic connections generate a decorrelated activity pattern (see Figure 3) to read out the STP state at those synapses. Such a mechanism can discriminate between synapses that were active at both time steps and those that were active only at the second time step and thereby expands the input space to twice its original dimension.

The discrimination performance of the proposed circuit is up to 100% better as compared to a standard perceptron (see Figure 4A). If synaptic integration at the readout unit bridges the gap between the two time steps and thus parts of the activity pattern of time step 1 are still present in  $v$  at the

second time step, discrimination performance is reduced (see Figure 4C). Such a removal of the initial input pattern might be realized by strong feedforward inhibition at time step 1.

Finally, we have derived a learning rule for the biophysical synaptic parameters probability  $p$  of release at baseline and the maximal synaptic conductance  $a$ . This learning rule achieves similarly high discrimination performance as the standard perceptron rule (see Figure 5). Close to the capacity limit and for balanced feedforward inhibition, the distributions of the two synaptic parameters  $p$  and  $a$  are highly clustered (see Figures 6C and 6D). Without feedforward inhibition (see Figure 6), the distributions become bimodal, with most values close to the bounds of the permitted parameter range.

The distributions of synaptic parameters can, at least in principle, be compared with distributions derived from physiological experiments. These distributions are, however, rather variable across different systems. For example, at neonatal hippocampal Shaffer collateral synapses, probabilities of release are distributed rather uniformly (Hanse & Gustafsson, 2001), which, in the context of our model, would account for a naive system loaded far below the capacity limit. In the visual and prefrontal (Wang et al., 2006) as well as the somatosensory cortex (Le Bè, Silberberg, Wang, & Markram, 2007), distributions of release probabilities are more unimodal, which is closer to what the model presented here predicts close to the capacity limit; however, the mean probability of release strongly differs across these systems, ranging between 0.2 and 0.5. In our model, the exact value for the mean probability of release is expected to depend on many factors. Besides the specific model  $F(p)$  of short-term synaptic facilitation, these factors may also include sparseness in the input and output (Brunel et al., 2004).

The mathematical structure and biophysical meaning of the update rules for the parameters  $a$  and  $p$  can be further elucidated if equations 2.9 and 2.10 are rewritten by introducing the following rate functions:

$$f_a(p) = -\frac{\rho(p) + p \rho'(p)}{p \rho'(p)} \quad (3.1)$$

$$f_p(p) = -\frac{\rho(p)}{p \rho'(p)} \quad (3.2)$$

$$f_x(p) = -\frac{1}{p \rho'(p)}. \quad (3.3)$$

Using these rate functions, expressing  $z$  via the activities  $y$ , and setting  $t^\mu = -O^\mu$  (indicating an incorrect response), the update rules read

$$\Delta a_n = \eta O^\mu y_n^\mu(2) \frac{1}{p_n} [f_a(p_n) - y_n^\mu(1) (f_a(p_n) + f_x(p_n))] \quad (3.4)$$

$$\Delta p_n = -\eta O^\mu y_n^\mu(2) \frac{1}{a_n} [f_p(p_n) - y_n^\mu(1) (f_p(p_n) + f_x(p_n))]. \quad (3.5)$$



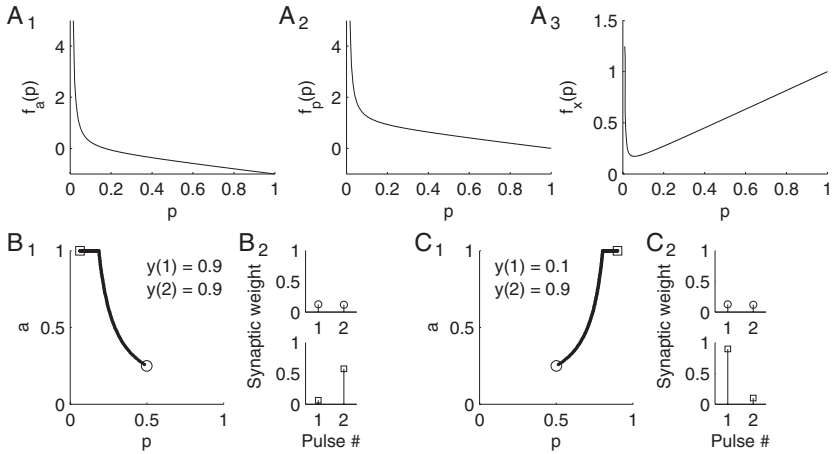


Figure 7: ( $A_1$ – $A_3$ ) Rate functions of synaptic change from equations 3.1 to 3.3. ( $B_1$ ) Trajectory of the learning rule for  $y(1) = 0.9$ ,  $y(2) = 0.9$ . The starting configuration ( $a = 0.25$ ,  $p = 0.5$ ) is marked with a circle. The learning rule converges to a low- $p$  configuration (square). ( $B_2$ ) Predicted synaptic responses  $w^{(1)} + g$  and  $w^{(2)} + g$  on a paired pulse stimulation. Top: Derived from initial conditions (circle in  $B_1$ ). Bottom: Derived from point of convergence (square in  $B_1$ ). ( $C_1$ ) Same as  $B_1$  for  $y(1) = 0.1$ ,  $y(2) = 0.9$ . The learning rule converges to a high- $p$  configuration (square). ( $C_2$ ) Same as  $B_2$  for  $y(1) = 0.1$ ,  $y(2) = 0.9$ .

This rewriting reveals that the update rules are constructed from three components. First is a Hebbian component  $O y(2)$  correlating pre- and postsynaptic activity at time step 2. Second is a state dependence expressed by the rate functions and the prefactors  $1/p$  and  $1/a$ , respectively. Third, rate functions depend on the history of the presynaptic activation via activity  $y(1)$  at the first time step. The rate functions as they follow from equations 3.1 to 3.3 are plotted in Figure 7A.

The functional consequences of the plasticity rule strongly depend on the activities  $y(1)$  and  $y(2)$  at the two time steps. Figures 7B and 7C illustrate two examples for which the iteration of equations 2.9 and 2.10 results in either a facilitating (see Figure 7B) or a depressing (see Figure 7C) synapse. Considering only the first response to a paired-pulse stimulus, the former situation would generally be considered long-term depression, whereas the latter would be considered long-term potentiation. Of course, such a paradigm is functionally not very helpful because for the synapse to change, one always has to assume an incorrect response. Nevertheless, these simulations predict that the nature of synaptic change should be different depending on whether the presynaptic stimulation is delivered in pairs.

The model of short-term synaptic plasticity is still highly simplified. Further parameters incorporating, for example, multiple vesicle pools,

postsynaptic desensitization, and spike broadening (Zucker & Regehr, 2002) have been omitted. They might provide additional degrees of freedom that could be exploited by third or fourth spikes or complex temporal patterns. If these additional parameters undergo long-term changes, they provide an expansion to an even higher-dimensional input space and, hence, might further increase the performance of discrimination learning.

Temporal integration of complex patterns has been found to increase storage capacity without short-term plasticity (Gütig & Sompolinsky, 2006). Here, we show that for STP and fixed interspike intervals, temporal integration at the readout unit is detrimental to the discrimination performance (see Figure 4C). This temporal integration could be prohibited by a fast feedforward inhibition onto the readout cells that is slightly delayed with respect to excitation. Such a dynamical pattern has been reported in the piriform cortex (Ketchum & Haberly, 1993).

The RFF network is inspired by analogies to biological circuits. As a first example, one can consider thalamic activity that is fed into the recurrent network of cortical layer IV, the output of which is then projected to layer II/III (Lund, Angelucci, & Bressloff, 2003). A second example is provided by the entorhinal cortex that projects to the hippocampal CA3, in which the axons form extensive recurrent connections as well as collaterals to CA1. The third example we mention is the olfactory system. There, the olfactory bulb projects to the anterior piriform cortex via the lateral olfactory tract, and the axons arising from the anterior piriform cortex make recurrent connections as well as feedforward projections to the posterior piriform cortex.

Although the presentation of our model makes use of analogies to the visual system, we stress that the RFF circuit is not primarily intended as a model of the primary visual cortex. Rather, it provides a general principle for pattern discrimination. The (decorrelated) receptive fields of the recurrently connected pools might thus also be auditory or olfactory (Fried, Fuss, & Korsching, 2002). In particular, learning to discriminate stimuli is commonly considered a problem in olfactory processing (Wilson & Stevenson, 2003). In contrast, models for visual categorization usually focus on generalization rather than discrimination abilities. These generalization and invariance properties are often modeled as to arise from a series of feedforward-connected layers (Riesenhuber & Poggio, 2002). For the visual system, the RFF may thus provide an additional computational advantage for discrimination tasks in each one of these layers. Although the RFF model is formally restricted to static inputs  $x$ , the generalization to dynamic input regimes (as in the auditory domain) is straightforward by simply regarding  $x$  itself as resulting from a temporal receptive field and assuming appropriate integration time constants.

The RFF model is formulated in terms of neuronal pools (Gerstner, 2000; Leibold, 2004), because a direct cellular interpretation would not allow graded activation functions  $\sigma$  and, moreover, would also violate Dale's

law. To understand how such a computational model can be realized biophysically in a neuronal microcircuit thus requires a detailed model of a cortical column (Haeusler & Maass, 2007) that may take into account different classes of interneurons and perhaps even cell morphologies. The model presented here can thus provide a computational framework for interpreting spatiotemporal population patterns in more detailed models.

## Appendix: Methods

---

**A.1 Natural Images.** Sensory input was emulated by 83,340 different random patches from the natural image selection from van Hateren and van der Schaaf (1998). The physical dimension of the patches was 2500 ( $50 \times 50$  pixels). From those patches, feedforward inputs  $\mathbf{x} = (x_1, \dots, x_N)^T$  were derived via ICA (Bingham & Hyvärinen, 2000) using the  $N$  most significant dimensions. For discrimination learning, in a given subset of  $P$  inputs  $\mathbf{x}^\mu$ ,  $\mu = 1, \dots, P$  is associated randomly with a teacher signal  $t^\mu = \pm 1$ .

**A.2 Facilitation Model.** Short-term synaptic facilitation is considered to increase the probability  $P^{(2)}$  of release on a second presynaptic spike as compared to the probability  $P^{(1)}$  of release on a first spike. In this letter, this process is modeled by  $P^{(2)} = F(P^{(1)})$  via the Michaelis-Menten-like function,

$$F(p) = \left( \frac{p(1+K)}{1+Kp} \right)^{5/4}, \quad (\text{A.1})$$

in which, for all simulations, the saturation constant is chosen  $K = 30$ . The power  $5/4$  reflects calcium-mediated interrelation of facilitation and release probability by assuming, for example, the following power laws:  $P_n^{(1)} \propto [\text{Ca}^{2+}]^4$  and  $P_n^{(2)}/P_n^{(1)} \propto [\text{Ca}^{2+}]$  (Zucker & Regehr, 2002; Felmy et al., 2003).

**A.3 Simulation of the Learning Rule.** For the simulations of the learning rules for  $a$  and  $p$  from equations 2.9 and 2.10, we used a learning rate of  $\eta = 2/P$ . Here the synaptic parameters were initialized at  $a_n = 0.5$  and  $p_n = 0.25$ . Learning curves  $N_{\text{learn}}(\alpha)$  are obtained by averaging over the heuristically defined number of  $200/\alpha$  trials, which takes into account that for low  $\alpha$ , fluctuations have a larger effect than for high  $\alpha$ , where the number  $N_{\text{learn}}(\alpha)$  of learning epochs increases more steeply.

## Acknowledgments

---

We thank Nick Lesica and Taro Toyozumi for comments on an earlier version of the manuscript, Felix Felmy for discussions on short-term synaptic

plasticity and comments on the manuscript, and Alex Loebel for discussions. This work was supported by the Bundesministerium für Bildung und Forschung under grant 01GQ0440 (Bernstein Center for Computational Neuroscience, Munich).

## References

---

- Abbott, L., & Regehr, W. (2004). Synaptic computation. *Nature*, *431*, 796–803.
- Abbott, L., Varela, J., Sen, K., & Nelson, S. (1997). Synaptic depression and cortical gain control. *Science*, *275*, 220–224.
- Amit, D., Campbell, C., & Wong, K. (1989). The interaction space of neural networks with sign-constrained synapses. *J. Phys. A: Math. Gen.*, *22*, 4687–4693.
- Bingham, E., & Hyvärinen, A. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural. Syst.*, *10*, 1–8.
- Brunel, N., Hakim, V., Isope, P., Nadal, J., & Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: Perceptron versus Purkinje cell. *Neuron*, *43*, 745–757.
- Felmy, F., Neher, E., & Schneggenburger, R. (2003). Probing the intracellular calcium sensitivity of transmitter release during synaptic facilitation. *Neuron*, *37*, 801–811.
- Fortune, E., & Rose, G. (2001). Short-term synaptic plasticity as a temporal filter. *Trends Neurosci.*, *24*, 381–385.
- Fried, H., Fuss, S., & Korsching, S. (2002). Selective imaging of presynaptic activity in the mouse olfactory bulb shows concentration and structure dependence of odor responses in identified glomeruli. *Proc. Natl. Acad. Sci. U.S.A.*, *99*, 3222–3227.
- Gardner, E. (1987). Maximum storage capacity in neural networks. *Europhys. Lett.*, *4*, 481–485.
- Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Comput.*, *12*, 43–89.
- Gundlfinger, A., Leibold, C., Gebert, K., Moisel, M., Schmitz, D., & Kempter, R. (2007). Differential modulation of short-term synaptic dynamics by long-term potentiation at mouse hippocampal mossy fibre synapses. *J. Physiol. (Lond.)*, *585*, 853–865.
- Gütig, R., & Sompolinsky, H. (2006). The tempotron: A neuron that learns spike timing-based decisions. *Nat. Neurosci.*, *9*, 420–428.
- Hanse, E., & Gustafsson, B. (2001). Vesicle release probability and pre-primed pool at glutamatergic synapses in area CA1 of the rat neonatal hippocampus. *J. Physiol. (Lond.)*, *531*, 481–493.
- Haeusler, S., & Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb. Cortex.*, *17*, 149–162.
- Hertz, J., Krogh, A., & Palmer, R. (1991). *Introduction to the theory of neural computation*. Boulder, CO: Westview Press.
- Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*, 78–80.
- Ketchum, K., & Haberly, L. (1993). Membrane currents evoked by afferent fiber stimulation in rat piriform cortex. I. Current source-density analysis. *J. Neurophysiol.*, *69*, 248–260.

- Köhler, H., & Widmaier, D. (1991). Sign-constrained linear learning and diluting in neural networks. *J. Phys. A: Math. Gen.*, *24*, L495–L502.
- Le Bè, J., Silberberg, G., Wang, Y., & Markram, H. (2007). Morphological, electrophysiological, and synaptic properties of corticoallosoal pyramidal cells in the neonatal rat neocortex. *Cereb. Cortex*, *17*, 2204–2213.
- Legenstein, R., & Maass, W. (2008). On the classification capability of sign-constrained perceptrons. *Neural Comput.*, *20*, 288–309.
- Leibold, C. (2004). Stability analysis of asynchronous states in neuronal networks with conductance-based inhibition. *Phys. Rev. Lett.*, *93*, 208104.
- Leibold, C., Gundlfinger, A., Schmidt, R., Thurley, K., Schmitz, D., & Kempter, R. (2008). Temporal compression mediated by short-term synaptic plasticity. *Proc. Natl. Acad. Sci. U.S.A.*, *105*, 4417–4422.
- Lund, J., Angelucci, A., & Bressloff, P. (2003). Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cereb. Cortex.*, *13*, 15–24.
- Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.*, *14*, 2531–2560.
- Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, *382*, 807–810.
- Moore, D., Rosenberg, J., & Coleman, J. (2005). Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. *Brain Lang.*, *94*, 72–85.
- Olshausen, B., & Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, *37*, 3311–3325.
- Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.*, *12*, 162–168.
- Shinzato, T., & Kabashima, Y. (2008). Perceptron capacity revisited: Classification ability for correlated patterns. *J. Phys. A: Math. Theor.*, *41*, 324013.
- Tsodyks, M., & Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. U.S.A.*, *94*, 719–723.
- van Hateren, J., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci.*, *265*, 359–366.
- Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.
- Wang, Y., Markram, H., Goodman, P., Berger, T., Ma, J., & Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat. Neurosci.*, *9*, 534–542.
- Wilson, D., & Stevenson, R. (2003). The fundamental role of memory in olfactory perception. *Trends Neurosci.*, *26*, 243–247.
- Zucker, R., & Regehr, W. (2002). Short-term synaptic plasticity. *Annu. Rev. Physiol.*, *64*, 355–405.

Copyright of *Neural Computation* is the property of MIT Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.