Neural Networks 48 (2013) 109–124

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Categorization and decision-making in a neurobiologically plausible spiking network using a STDP-like learning rule

Michael Beyeler [a,*], Nikil D. Dutt [a], Jeffrey L. Krichmar [a,b]

[a] Department of Computer Science, University of California Irvine, Irvine, CA 92697-3435, United States
[b] Department of Cognitive Sciences, 3151 Social Science Plaza, University of California Irvine, Irvine, CA 92697-5100, United States

ARTICLE INFO

Article history:
Received 24 February 2013
Received in revised form 28 July 2013
Accepted 31 July 2013

Keywords:
Spiking neural network
Synaptic dynamics
STDP
Supervised learning
Object recognition
Decision-making

ABSTRACT

Understanding how the human brain is able to efficiently perceive and understand a visual scene is still a field of ongoing research. Although many studies have focused on the design and optimization of neural networks to solve visual recognition tasks, most of them either lack neurobiologically plausible learning rules or decision-making processes. Here we present a large-scale model of a hierarchical spiking neural network (SNN) that integrates a low-level memory encoding mechanism with a higher-level decision process to perform a visual classification task in real-time. The model consists of Izhikevich neurons and conductance-based synapses for realistic approximation of neuronal dynamics, a spike-timing-dependent plasticity (STDP) synaptic learning rule with additional synaptic dynamics for memory encoding, and an accumulator model for memory retrieval and categorization. The full network, which comprised 71,026 neurons and approximately 133 million synapses, ran in real-time on a single off-the-shelf graphics processing unit (GPU). The network was constructed on a publicly available SNN simulator that supports general-purpose neuromorphic computer chips. The network achieved 92% correct classifications on MNIST in 100 rounds of random sub-sampling, which is comparable to other SNN approaches and provides a conservative and reliable performance metric. Additionally, the model correctly predicted reaction times from psychophysical experiments. Because of the scalability of the approach and its neurobiological fidelity, the current model can be extended to an efficient neuromorphic implementation that supports more generalized object recognition and decision-making architectures found in the brain.

1. Introduction

Object recognition in monkeys has traditionally been associated with an anatomically distinct pathway termed the "what" (or ventral) visual stream (Ungerleider & Haxby, 1994), which consists of at least V1, V2, V4, and various regions in the inferior and anterior temporal cortices (e.g., TEO, TE1–TE3, TEa, and TEm) (Rolls, 2012; Rolls & Deco, 2002). While traveling along this pathway, the characteristics of the stimuli to which neurons respond become more complex (Rolls, 2012; Ungerleider & Haxby, 1994), ranging from rather simple stimuli with small receptive fields such as oriented bars in V1 (Hubel & Wiesel, 1965) to relatively large and more abstract objects such as faces in the inferotemporal cortex (IT) (Bruce, Desimone, & Gross, 1981). These empirical observations have led to a number of classic studies modeling the ventral stream as a hierarchical feed-forward network, such as the Neocognitron (Fukushima, 1980), HMAX (Riesenhuber & Poggio, 1999), or VisNet (Rolls, 2012; Wallis & Rolls, 1997)—although it should be noted that the notion of a strictly hierarchical or feed-forward network has been questioned by recent anatomical studies that reserve a more important functional role for bi-directional and non-hierarchical connections (Markov et al., 2012, 2011). Inspired by these classic models, a variety of more conventional machine learning algorithms have emerged that demonstrate the extraordinary performance in certain recognition tasks, such as convolutional neural networks (CNNs) in handwriting recognition (Ciresan, Meier, Masci, & Schmidhuber, 2011; LeCun, Bottou, Bengio, & Haffner, 1998; Simard, Steinkraus, & Platt, 2003)—or for that matter, adaptive boosting in face recognition (Viola & Jones, 2001). Although CNNs implement a network topology that is biologically-inspired, they often rely on the error backpropagation (gradient descent), which has been criticized for being biologically unrealistic because it involves variables that cannot be computed locally (Rolls & Deco, 2002). Part of the challenge is thus to discover how comparably hard problems can be solved by more biologically plausible networks relying on local learning rules that operate on the abstraction level of a synapse.

* Correspondence to: Department of Cognitive Sciences, 2236 Social & Behavioral Sciences Gateway, University of California Irvine, Irvine, CA 92697-5100, United States.
E-mail address: mbeyeler@uci.edu (M. Beyeler).
URL: http://sites.uci.edu/mbeyeler (M. Beyeler).

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.07.012

A potential candidate for such a mechanism is spike-timing-dependent plasticity (STDP), (Bi & Poo, 2001; Sjöström, Turrigiano, & Nelson, 2001; Song, Miller, & Abbott, 2000), a paradigm which modulates the weight of synapses according to their degree of causality. Many different variants of STDP seem to exist in the brain, and many different models to explain them have emerged over the years (Morrison, Diesmann, & Gerstner, 2008). In an effort to implement STDP-like learning rules using only information locally available at the synapse without algorithmically storing the spike timings, several models have proposed to pair presynaptic spiking with postsynaptic voltage, determining the weight change by using either the temporal change of postsynaptic voltage (Porr, Saudargiene, & Worgotter, 2004), a piece-wise linear function to approximate postsynaptic voltage (Gorchetchnikov, Versace, & Hasselmo, 2005) or postsynaptic Calcium concentration (Brader, Senn, & Fusi, 2007; Graupner & Brunel, 2012). Networks with STDP have been shown to be able to learn precise spike times through supervised learning (Legenstein, Naeger, & Maass, 2005; Pfister, Toyoizumi, Barber, & Gerstner, 2006), to implement reinforcement learning (Florian, 2007; Izhikevich, 2007b; O'Brien & Srinivasa, 2013), to develop localized receptive fields (Clopath, Busing, Vasilaki, & Gerstner, 2010), or to classify highly correlated patterns of neuronal activity (Brader et al., 2007).

Once an internal representation of a visual object is built in the brain, the question then remains how this memory can be retrieved from the system in order to make a perceptual decision. A general mechanism has been suggested to involve the temporal integration and comparison of the outputs of different pools of sensory neurons in order to compute a decision variable (Heekeren, Marrett, Bandettini, & Ungerleider, 2004). This temporal integration might be performed in one of several regions such as the dorsolateral prefrontal cortex (dlPFC) (Heekeren et al., 2004; Kim & Shadlen, 1999), lateral intraparietal area (LIP) (Shadlen & Newsome, 2001), superior colliculus (SC) (Horwitz & Newsome, 1999), frontal eye fields (FEF) (Schall, 2002; Schall & Thompson, 1999; Thompson, Hanes, Bichot, & Schall, 1996) or intraparietal sulcus (IPS) (Colby & Goldberg, 1999), which all cooperate in order to translate the accumulated evidence into an action (Heekeren, Marrett, & Ungerleider, 2008; Rorie & Newsome, 2005). Neuronal activity in integrator areas gradually increases and then remains elevated until a response is given, with the rate of increase being slower during more difficult trials. A successful approach to explaining these kinds of neurophysiological data has been through the means of drift–diffusion or race models (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Schall & Thompson, 1999; Smith & Ratcliff, 2004), in which the noisy sensory information is integrated over time until a decision threshold is reached.

Here we present a large-scale model of a hierarchical spiking neural network (SNN) that integrates a low-level memory encoding mechanism with a higher-level decision process to perform a visual classification task in real-time. The model consists of Izhikevich neurons and conductance-based synapses for realistic approximation of neuronal dynamics (Dayan & Abbott, 2001; Izhikevich, 2003; Izhikevich, Gally, & Edelman, 2004), a STDP synaptic learning rule with additional synaptic dynamics for memory encoding (Brader et al., 2007), and an accumulator model for memory retrieval and categorization (Smith & Ratcliff, 2004). Grayscale input images were fed through a feed-forward network consisting of V1 and V2, which then projected to a layer of downstream classifier neurons through plastic synapses that implement the STDP-like learning rule mentioned above. Population responses of these classifier neurons were then integrated over time to make a perceptual decision about the presented stimulus. The full network, which comprised 71,026 neurons and approximately 133 million synapses, ran in real-time on a single off-the-shelf graphics processing unit (GPU).

In order to evaluate the feasibility of our model, we applied it to the extensively studied MNIST database of handwritten digits (LeCun et al., 1998). Due to the large variability within a given class of digits and a high level of correlation between members of different classes, the database provides stimuli whose categorization might span a wide range of difficulty levels, and as such is well-suited as a first benchmark for our model. However, it should be noted that MNIST does not pose many of the challenges of biological vision, such as distractors, occluders or translation invariance. Moreover, all the images are static and isolated in their receptive field. The network achieved 92% correct classifications, which is comparable to other SNN approaches (Brader et al., 2007; Querlioz, Bichler, & Gamrat, 2011) and simple machine learning algorithms (such as linear classifiers, k-Nearest Neighbor classifiers and simple artificial neural networks LeCun et al., 1998), but not to state-of-the-art models whose performance is close to 99.8% (Ciresan et al., 2011; Niu & Suen, 2012).

Additionally, our network produces reaction time (RT) distributions that are comparable to the behavioral RT distributions reported in psychophysical experiments. For example, we show that when the network makes an error, its RT is significantly slower than when making a correct class prediction; and that RTs do not decrease when the target stimulus has become familiar, which has also been observed in a rapid categorization study (Fabre-Thorpe, Richard, & Thorpe, 1998).

Although the present model does not reach the performance of specialized classification systems (Ciresan et al., 2011; Niu & Suen, 2012), our model represents a first step towards the construction of a general-purpose neurobiologically inspired model of visual recognition and perceptual decision-making. The model includes many neurobiologically inspired details not found in the algorithms described above. The present network was constructed on a publicly available SNN simulator that uses design principles, data structures, and process flows that are in compliance with general-purpose neuromorphic computer chips, and that allows for real-time execution on off-the-shelf GPUs (Richert, Nageswaran, Dutt, & Krichmar, 2011); its neuron model, synapse model, and address-event representation (AER) are compatible with recent neuromorphic hardware (Srinivasa & Cruz-Albrecht, 2012). Because of the scalability of our approach, the current model can readily be extended to an efficient neuromorphic implementation that supports the simulation of more generalized object recognition and decision-making regions found in the brain. Ultimately, understanding the neural mechanisms that mediate perceptual decision-making based on sensory evidence will further our understanding of how the brain is able to make more complex decisions we encounter in everyday life (Lieberman, 2007), and could shed light on phenomena like the misperception of objects in neuropsychiatric disorders such as schizophrenia (Persaud & Cutting, 1991; Summerfield, Egner, Mangels, & Hirsch, 2006).

2. Methods

We performed all simulations in a large-scale SNN simulator which allows the execution on both generic x86 central processing units (CPUs) and standard off-the-shelf GPUs (Richert et al., 2011). The simulator provides a PyNN-like environment (PyNN is a common programming interface developed by the neuronal simulation community) in C/C++ and is publicly available at http://www.socsci.uci.edu/~jkrichma/Richert-FrontNeuroinf-SourceCode.zip. The simulator's API allows for details and parameters to be specified at the synapse, neuron, and network levels. In our experiments we used an Intel Xeon X5675 CPU operating at 3.07 GHz and a single NVIDIA Tesla M2090 GPU card.

All parameter values can be found in Table 1.

Table 1
All parameter values used in our simulations.

Simulation parameters:		
t_{stim}	Stimulation period	500 ms
t_{delay}	Delay period	1000 ms
n_{train}	Total number of training patterns	10/100/1000/2000
n_{test}	Total number of test patterns	1000
n_{cycles}	Number of training cycles	200/20/2/1
Single-cell parameters:		
V_{cutoff}	Spike cutoff	+30
V_{reset}	Reset potential RS/FS	−65
θ_V	Voltage threshold for X up-/down-regulations	−62.5
Network parameters:		
N_{inp}	# of spike generators in the input layer	65,856
$N_{\text{inp},i}$	# of neurons in the input inhibitory layer	784
N_{orient}	# of neurons in the orientation layer	3136
$N_{\text{orient},i}$	# of neurons in the orientation inhibitory layer	800
N_{dec}	# of neurons in the decision layer	150
$N_{\text{dec},i}$	# of neurons in the decision inhibitory layer	150
N_{teach}	# of spike generators in the teacher population	150
Synaptic parameters:		
w^{+}_{plastic}	Potentiated synaptic efficacy (binary synapse)	0.002
w^{-}_{plastic}	Depressed synaptic efficacy (binary synapse)	0
τ_{AMPA}	Time constant (decay) for AMPA channels	5 ms
τ_{NMDA}	Time constant (decay) for NMDA channels	150 ms
τ_{GABA_a}	Time constant (decay) for GABA_a channels	6 ms
τ_{GABA_b}	Time constant (decay) for GABA_b channels	150 ms
Synaptic dynamics:		
θ_X	Threshold for synaptic transition	0.5
α	Upwards drift rate (towards one)	0.1 s^{-1}
β	Downwards drift rate (towards zero)	0.1 s^{-1}
γ	Amplitude of X upwards jump	0.1
δ	Amplitude of X downwards jump	0.1
Calcium variable:		
J_C	Amplitude of Calcium up-regulation	3.4
τ_C	Calcium decay time constant	60 ms
θ^{l}_{up}	Lower Ca-bound for up-regulation of X	3
θ^{h}_{up}	Upper Ca-bound for up-regulation of X	12
θ^{l}_{down}	Lower Ca-bound for down-regulation of X	3
θ^{h}_{down}	Lower Ca-bound for down-regulation of X	4

2.1. Model specifications

Spiking neurons in the network were modeled using the four parameter Izhikevich model (Izhikevich, 2003), which aims to reduce Hodgkin–Huxley-type neuronal models to a two-dimensional system of ordinary differential equations,

$$\frac{dV(t)}{dt} = 0.04V^2(t) + 5V(t) + 140 - U(t) + I_{\text{syn}}(t) \tag{1}$$

$$\frac{dU(t)}{dt} = a\left(bV(t) - U(t)\right), \tag{2}$$

where (1) describes the membrane potential V for a given external current I_{syn}, (2) describes a recovery variable U, the parameter a is the rate constant of the recovery variable, and the parameter b describes the sensitivity of the recovery variable to the subthreshold fluctuations of the membrane potential. All parameters in (1) and (2) are dimensionless; however, the right-hand side of (1) is in a form so that the membrane potential V has mV scale and the time t has ms scale (Izhikevich, 2003). The Izhikevich model is well-suited for large-scale simulations, because it is computationally inexpensive yet capable of spiking, bursting, and being either an integrator or a resonator (Izhikevich, 2004, 2007a).

In contrast to other simple models such as the leaky integrate-and-fire (LIF) neuron, the Izhikevich neuron is able to generate the upstroke of the spike itself. Thus the voltage reset occurs not at the threshold, but at the peak ($V_{\text{cutoff}} = +30$), of the spike. The action potential downstroke is modeled using an instantaneous reset of the membrane potential whenever V reaches the spike cutoff, plus a stepping of the recovery variable:

$$V(V > V_{\text{cutoff}}) = c \quad \text{and} \quad U(V > V_{\text{cutoff}}) = U + d. \tag{3}$$

The inclusion of U in the model allows for the simulation of typical spike patterns observed in biological neurons. The four parameters a, b, c, and d can be set to simulate different types of neurons. All excitatory neurons were regular-spiking (RS) neurons ($a = 0.02, b = 0.2, c = -65, d = 8$), and all inhibitory neurons were fast-spiking (FS) neurons ($a = 0.1, b = 0.2, c = -65, d = 2$) (Izhikevich, 2003, 2004).

Synaptic conductances were modeled as dynamic synaptic channels with exponential decay:

$$\frac{dg_r(t)}{dt} = -\frac{1}{\tau_r} g_r(t) + w \sum_i \delta\left(t - t_i\right) \tag{4}$$

where δ is the Dirac delta, the sum is over all presynaptic spikes arriving at times t_i, w is the weight of that synapse, τ_r is its time constant, and the subscript r denotes the receptor type; that is, AMPA (fast decay), NMDA (slow decay and voltage-dependent), GABAa (fast decay), or GABAb (slow decay). A spike arriving at a synapse that is postsynaptically connected to an excitatory (inhibitory) neuron increases both g_{AMPA} and g_{NMDA} (g_{GABA_a} and g_{GABA_b}). In our simulations we set the time constants for the decay of the conductances to $\tau_{\text{AMPA}} = 5$ ms, $\tau_{\text{NMDA}} = 150$ ms, $\tau_{\text{GABA}_a} = 6$ ms, and $\tau_{\text{GABA}_b} = 150$ ms (Dayan & Abbott, 2001; Izhikevich et al., 2004). Rise time of these conductances was modeled as instantaneous, which is a reasonable assumption in the case of AMPA, NMDA, and GABA_a (Dayan & Abbott, 2001), but a simplification in the case of GABA_b, which has a rise time on the order of 10 ms (Koch, 1999).

Then the total synaptic current I_{syn} in (1) for each neuron is given by:

$$\begin{aligned} I_{\text{syn}} = & -g_{\text{AMPA}}\left(V - 0\right) \\ & -g_{\text{NMDA}} \frac{\left(\frac{V+80}{60}\right)^2}{1 + \left(\frac{V+80}{60}\right)^2} \left(V - 0\right) \\ & -g_{\text{GABA}_a}\left(V + 70\right) \\ & -g_{\text{GABA}_b}\left(V + 90\right), \end{aligned} \tag{5}$$

where V is the membrane potential of the neuron, and the subscript indicates the receptor type. This equation is equal to the one described in Izhikevich et al. (2004).

2.2. Synaptic dynamics

The model of synaptic plasticity used in our study was previously described in Brader et al. (2007), henceforth referred to as the original model implementation.

We justify our choice by the fact that the model (i) is among the first STDP-like learning rules shown to be able to classify complex patterns (Brader et al., 2007), that it (ii) reproduces a series of empirical observations that the classic STDP cannot (Brader et al., 2007; Graupner & Brunel, 2012; Sjöström et al., 2001), and that it (iii) is well-suited for neuromorphic and large-scale implementations (Brader et al., 2007; Indiveri & Fusi, 2007; Mitra, Fusi, & Indiveri, 2009). However, it is certainly not the only available model that meets these criteria, and it would be interesting to compare its performance to other models of spike-driven synaptic plasticity.

To the best of our knowledge, our implementation is the first to study this plasticity rule in the framework of Izhikevich spiking neurons and conductance-based synapses. In this section, we reproduce some of the plots from the original model implementation in order to identify any influence this change of model might

have on the tuning properties of the plasticity rule. It was our intention to stay as close to the original model as possible in order to verify the correctness of our implementation. However, we have not systematically explored the parameter space, and it is possible that other sets of parameter values exist that would be equally appropriate for a given learning task or model application.

The present synaptic plasticity rule differs from the basic STDP model in that its update rule does not depend on the precise timing between pre and postsynaptic spikes, but is rather governed by (i) the postsynaptic depolarization at the time of a presynaptic spike and (ii) an auxiliary variable that integrates the postsynaptic action potentials. The model incorporates these two variables to account for experimental findings, for example, that the postsynaptic neuron needs to be sufficiently depolarized for long-term potentiation (LTP) to occur, and that LTP dominates long-term depression (LTD), independent of relative spike timing, when both pre and postsynaptic neurons fire at elevated frequencies (Sjöström et al., 2001). The auxiliary variable implements a regulatory mechanism that suppresses all synaptic modifications if the firing rate of the postsynaptic neuron is either too low or too high. A natural candidate for such a variable would be an internal Calcium concentration (henceforth denoted as $C(t)$), as it is assumed that a relatively slow variable acting on a timescale on the order of 100 ms is measuring the postsynaptic mean firing rate (Brader et al., 2007). The Calcium variable was implemented as

$$\frac{dC(t)}{dt} = -\frac{1}{\tau_C} C(t) + J_C \sum_i \delta(t - t_i), \tag{6}$$

where δ is the Kronecker delta, and the sum is over all postsynaptic spikes arriving at times t_i. J_C is the contribution of a single postsynaptic spike, and τ_C is the time constant.

Fig. 1 shows the empirical probability distribution (relative frequency) over Calcium concentrations for different values of postsynaptic activity ($\tau_C = 60$ ms and $J_C = 3.4$). Here, a single neuron at rest ($V(0) = -65$ mV, $C(0) = 0$) was stimulated for 500 ms with Poisson spike trains of a given mean firing rate. Each curve is the empirical probability distribution based on the observed $C(t)$ in 10^5 repetitions of this protocol. The induced mean postsynaptic firing rate is indicated in the plot legend. For low mean activity ($f_{\text{mean}} < 20$ Hz) the shape of the distribution is dominated by the exponential decay in (6). As the activity increases, the distribution is approximately Gaussian. The shaded regions correspond to regions where LTP transitions (red) and LTD transitions (yellow) are possible, as explained below. Calcium concentration falling in a shaded region will co-determine how likely LTP and LTD transitions are at that point, as will be explained in the next paragraph. The value for J_C was chosen such that the shape of the Calcium distribution approximated the one reported in the original model implementation (see Fig. 2(a) of Brader et al., 2007).

The plastic synapses are modeled with a hidden variable $X(t)$ that is confined between values zero and one. If the hidden variable has a value above a certain threshold θ_X, it drifts towards one (potentiated state); otherwise it drifts towards zero (depressed state) (Fusi, Annunziato, Badoni, Salamon, & Amit, 2000). The state of a synapse can be altered only if a series of modifications is made to the hidden variable. In turn, a modification to the hidden variable (a "jump") can only be made upon the arrival of a presynaptic spike, given that the following conditions on the postsynaptic membrane potential and on the postsynaptic Calcium are met:

$$\begin{aligned} X &\leftarrow X + \gamma \quad \text{if } V(t_{\text{pre}}) > \theta_V \text{ and } \theta^l_{\text{up}} < C(t_{\text{pre}}) < \theta^h_{\text{up}} \\ X &\leftarrow X - \delta \quad \text{if } V(t_{\text{pre}}) \le \theta_V \text{ and } \theta^l_{\text{down}} < C(t_{\text{pre}}) < \theta^h_{\text{down}}, \end{aligned} \tag{7}$$

where t_{pre} is the time of an arriving presynaptic spike, θ_V is a threshold on the postsynaptic membrane potential, and the parameters γ and δ correspond to the parameters a and b in the original model, respectively. Up-jumps are only allowed in the Calcium interval $(\theta^l_{\text{up}}, \theta^h_{\text{up}})$, which corresponds to the red-shaded region (labeled "LTP") in Fig. 1. Analogously, down-jumps are only allowed in the interval $(\theta^l_{\text{down}}, \theta^h_{\text{down}})$, which corresponds to the yellow-shaded region (labeled "LTD") in Fig. 1. Because in our case θ^l_{down} is equal to θ^l_{up}, the two regions overlap (indicated in orange), which is in agreement with the original model implementation. Note that in this region both up- and down-jumps are possible, depending on the relative position of $V(t_{\text{pre}})$ to θ_V, as explained next. Refer to Table 1 for the exact parameter values.

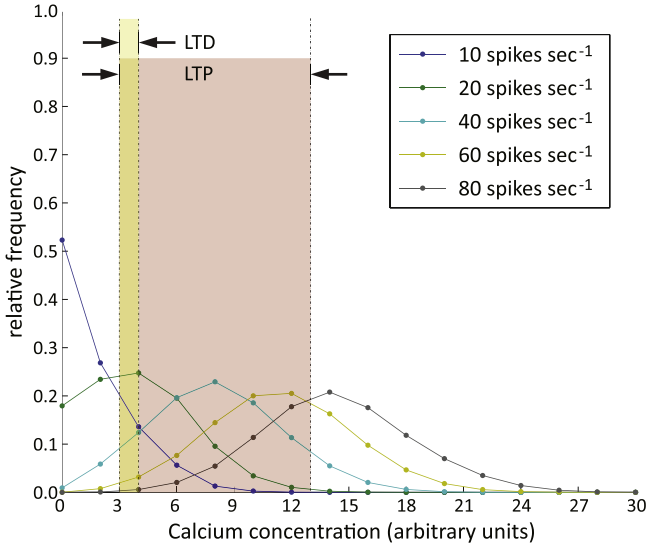

Fig. 1. Empirical probability distribution (relative frequency) of Calcium concentration (arbitrary units) as a function of mean postsynaptic activity for different values of postsynaptic activity ($J_C = 3.4$ and $\tau_C = 60$ ms). For low postsynaptic activity ($f < 20$ Hz) the shape of the distribution is dominated by the exponential decay. As presynaptic activity increases, the distribution is approximately Gaussian. Vertical dashed lines are thresholds on $C(t)$ which define LTP- (red) and LTD-relevant (yellow) regimes, in which X up- and down-jumps are allowed (see (7)). Note that these two regimes overlap (orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Jumps and state transitions are related as follows. In order to make an up-jump of magnitude γ the postsynaptic membrane potential needs to be sufficiently depolarized ($V(t_{\text{pre}}) > \theta_V$) and the postsynaptic neuron must have had an elevated mean firing rate in the recent past, such that $\theta^l_{\text{up}} < C(t_{\text{pre}}) < \theta^h_{\text{up}}$. The hidden variable may eventually cross the threshold θ_X through an accumulation of up-jumps. Then an LTP transition is defined as the event in which X crosses θ_X from below to above. Analogously, down-jumps of magnitude δ are allowed only if the postsynaptic membrane potential is hyperpolarized ($V(t_{\text{pre}}) \le \theta_V$) and only if the postsynaptic neuron has had a relatively low firing rate in the recent past ($\theta^l_{\text{down}} < C(t_{\text{pre}}) < \theta^h_{\text{down}}$); and the event in which X crosses θ_X from above to below is called an LTD transition.

As mentioned above, in the absence of stimulation the internal synaptic variable drifts towards values zero or one (intrinsically stable fixed points) according to

$$\begin{aligned} \frac{dX}{dt} &= \alpha \quad \text{if } X > \theta_X \\ \frac{dX}{dt} &= -\beta \quad \text{if } X \le \theta_X \end{aligned} \tag{8}$$

where θ_X is the transition threshold. The actual weight w of the synapse, which is being integrated over time according to (4), is different from $X(t)$: At each time step $X(t)$ is compared to θ_X, and the weight w is set to $w^+_{\text{plastic}} = 0.002$ whenever $X(t) > \theta_X$, and it is set to $w^-_{\text{plastic}} = 0$ whenever $X(t) \le \theta_X$. Refer to Morrison et al. (2008) for a discussion on the biological evidence for and against

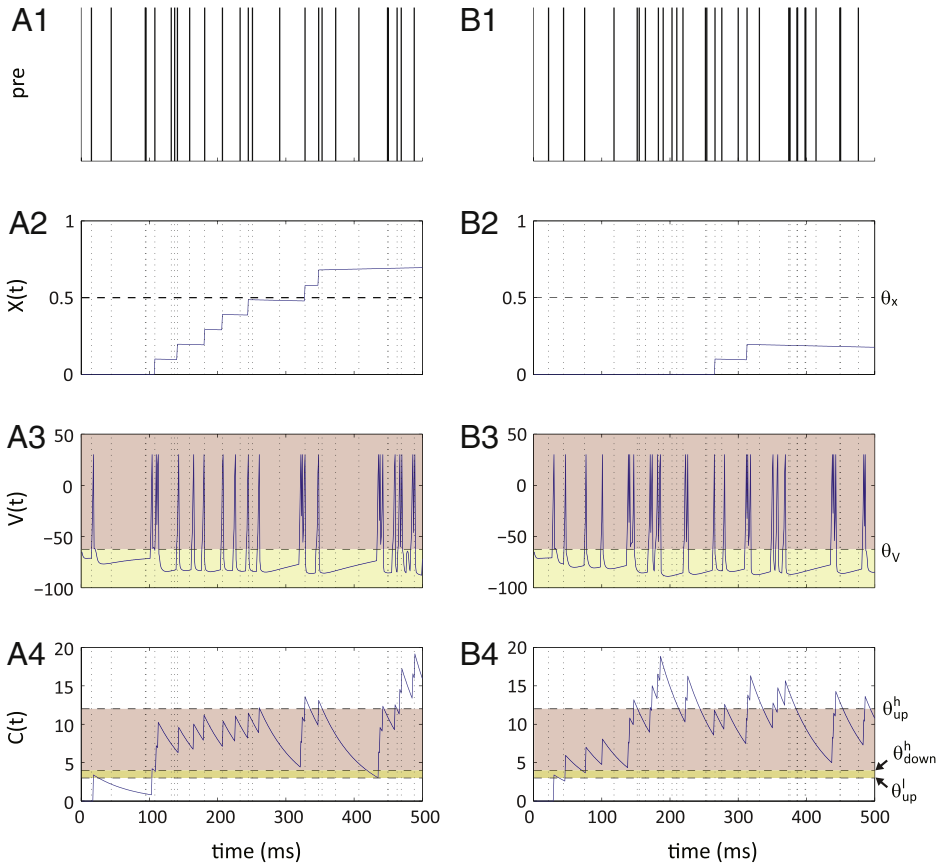

Fig. 2. Illustration of the stochastic nature of synaptic modifications. In two sample runs A and B, a single synapse was both pre and postsynaptically connected to a spike generator that fired at a mean rate of 50 Hz. The protocol of the synaptic learning rule was executed for 500 ms. Panels A1–4 illustrate a sample run for which the accumulation of jumps caused X to cross the threshold θ_X (horizontal dashed line in Panel A2), that is, an LTP transition. Upon a presynaptic spike (Panel A1) the postsynaptic membrane potential V (Panel A3) and the Calcium variable C (Panel A4) are read out. If V exceeded the threshold θ_V (horizontal dashed line in Panel A3) and C fell within the LTP-relevant region, an up-jump in X was registered (Panel A2; see (7)). In the absence of stimulation X drifted towards one of the two stable states (see (8)). In another sample run (Panels B1–4) no LTP transition occurred. Analogous to Fig. 1, shaded regions indicate where up-jumps and LTP (red) or down-jumps and LTD (yellow) are possible. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

binary synapses. The amplitude of the weight w^+_{plastic} should be set such that several LTP transitions are required to affect the postsynaptic firing rate. Other than that the model is fairly robust against the amplitude changes of w^+_{plastic}. For a thorough analysis of the stability of the model with respect to parameter variations please refer to Section 5.7 in Brader et al. (2007).

Fig. 2 illustrates the stochastic nature of a synaptic modification. In two sample runs A and B, a single synapse was both presynaptically and postsynaptically connected to a Poisson spike generator which fired at a mean rate of $f_{\text{pre}} = f_{\text{post}} = 50$ Hz. The update protocol was executed for $t_{\text{stim}} = 500$ ms. Panels A1–4 illustrate a sample run for which the accumulation of jumps caused X to cross the threshold θ_X (horizontal dashed line in Panel A2), that is, an LTP transition. In another sample run under the same conditions (B1–4) no LTP transition occurred. Analogous to Fig. 1, shaded regions indicate where up-jumps and LTP (red) or down-jumps and LTD (yellow) are possible.

If presynaptic spikes arrive at random times, then consolidation is activated with some probability (Brader et al., 2007; Fusi et al., 2000). Thus the stochastic nature of a synaptic transition arises from the Poisson statistics of the input signal, not from the update protocol itself (which is in fact deterministic). Fig. 3 plots the empirical probability of an LTP transition (Panel A) and of an LTD transition (Panel B) for a single synapse as a function of mean postsynaptic activity. The protocol was the same as in Fig. 2, but the mean firing rates were varied between 0 Hz and 100 Hz. The probabilities of LTP and LTD transitions shown in Fig. 3 roughly reproduced the distributions reported in Fig. 2(b)–(d) of the original model implementation. In Panel A, the synapse is initialized with $X(0) = 0$, then stimulated for $t_{\text{stim}} = 500$ ms. If $X(t_{\text{stim}}) > \theta_X$, an LTP transition is registered. Analogously in Panel B, the synapse is initialized with $X(0) = 1$, and an LTD transition is registered if $X(t_{\text{stim}}) < \theta_X$. Every data point is the mean value of $n = 10^5$ trials, and the vertical bars are the standard error. The parameter values suggested in the original model implementation qualitatively reproduced the shape of their reported distributions, but the peak probabilities were significantly lower. We thus set $\theta_V = -62.5$ mV and gradually reduced the drift rates α and β described in (8) until the peak probabilities roughly matched the original model implementation. Differences in peak values are probably due to differences in the subthreshold dynamics of the two employed neuron models. For a more elaborate discussion of the implementation differences refer to Section 4.1 below.

The peak LTP transition probability is observed when both pre and postsynaptic neurons fire at approximately 50 Hz. The probability drops off above 50 Hz, because it is increasingly less likely that $C(t)$ falls in the LTP-relevant regime. This regulatory mechanism is used as a stop-learning condition as follows. If the frequency of the postsynaptic neuron is too high – or for that matter, too low – no long-term modification should be induced. Below 50 Hz the constraint on the postsynaptic membrane potential plays an increasingly important role: at low frequencies $V(t)$ is likely to fluctuate near the resting potential, which does not exceed θ_V, and thus favors LTD (peak probability near 20 Hz).

In the case of $f_{\text{pre}} = 50$ Hz, the LTP and LTD transition probabilities intersect at roughly 30 Hz. This is an important

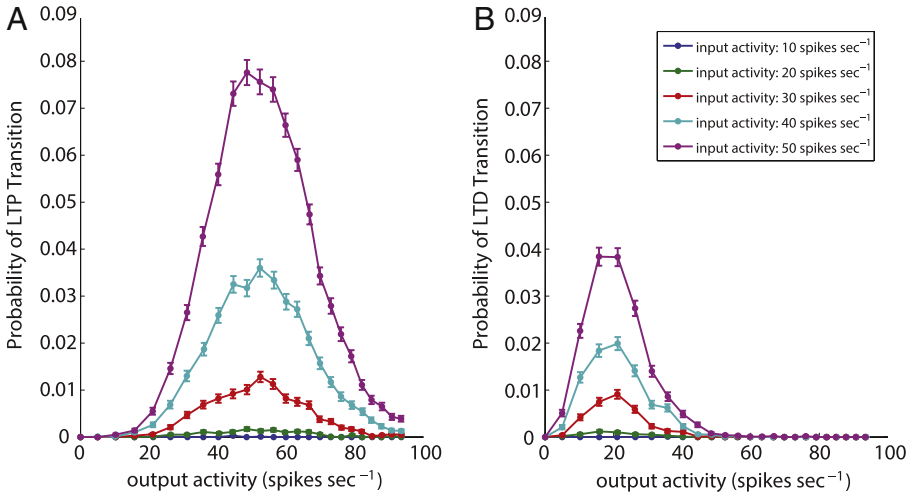

Fig. 3. LTP (left) and LTD (right) transition probabilities. The transition probability distribution of a single synapse is plotted as a function of the mean output activity for different values of input activity. For the left panel, a single synapse was initialized with $X(0) = 0$, and then repeatedly stimulated ($n = 10^5$ trials) for $t_{\text{stim}} = 500$ ms. If $X(t_{\text{stim}}) > \theta_X$ an LTP transition was registered. Analogously for LTD, the synapse was initialized with $X(0) = 1$, and an LTD transition was registered if $X(t_{\text{stim}}) < \theta_X$. Vertical bars are the standard error.

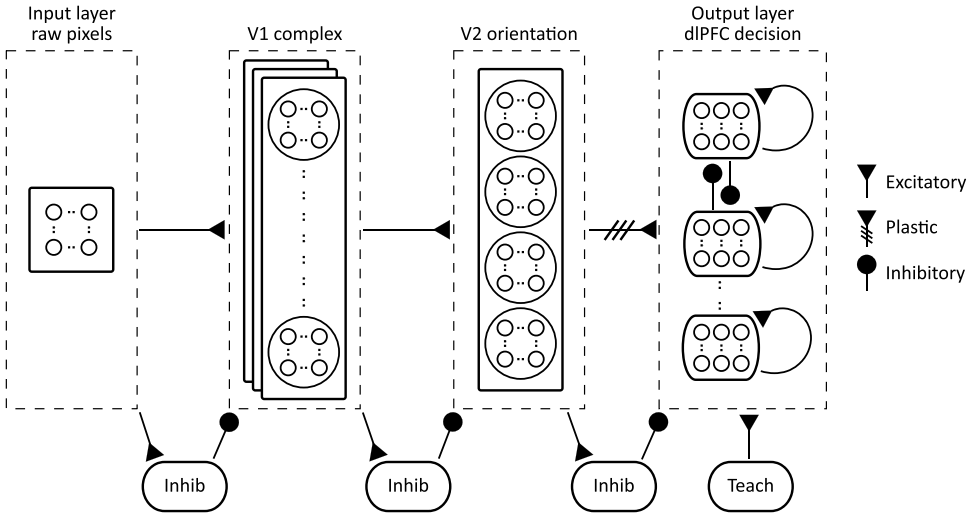

Fig. 4. Network architecture. The network consisted of 71,026 neurons (small circles) organized in four layers (dotted boxes), and approximately 133 million synapses (denoted by arrows; triangles: excitatory synapses, filled circles: inhibitory synapses). MNIST grayscale images (input layer) were fed through a feed-forward network consisting of V1 and V2, which then projected to a downstream population of decision neurons (output layer).

characteristic of the learning mechanism, as it implies that postsynaptic firing rates above 30 Hz will most likely induce LTP transitions, while rates below that value will most likely induce LTD transitions. Also, it accounts for the fact that LTP always dominates LTD when both pre and postsynaptic neurons fire at elevated frequencies (Sjöström et al., 2001).

The memory span is given by the learning rate (or the probability of a synaptic transition), which can be adjusted by changing the parameters α, β, γ, and δ in (7)–(8). The learning rate should be set sufficiently low in order to allow gradual learning (French, 1999). The challenge is to make the learning rule simultaneously sensitive to, but not radically disrupted by, new information (the so-called "stability–plasticity" dilemma) (French, 1999; Grossberg, 1980). In the here employed model, presynaptic activity acts as a trigger for synaptic modifications. Low presynaptic activity significantly reduces the probability of a synaptic transition, which makes the formation of a memory robust to disruptions by spontaneous activity (in the case of $f_{\text{pre}} = 50$ Hz we observe $\text{Pr}_{\text{LTP}}\left(f_{\text{post}} < 5 \text{ Hz}\right) < 10^{-5}$ and $\text{Pr}_{\text{LTD}}(f_{\text{post}} < 5 \text{ Hz}) \cong 10^{-4}$). Moreover, it theoretically allows a memory to be preserved indefinitely in the absence of stimulation, as $\text{Pr}_{\text{LTP}}\left(f_{\text{post}} = 0\right) = \text{Pr}_{\text{LTD}}\left(f_{\text{post}} = 0\right) = 0$. On the other hand, new stimuli can be learned only gradually after experiencing them several times (also known as "interleaved learning") (French, 1999).

2.3. The MNIST database

To evaluate the feasibility of our model, we conducted experiments on the extensively studied MNIST dataset of handwritten digits (LeCun et al., 1998). The dataset contains a total of 70,000 grayscale images, which are size-normalized to fit in a 20×20 pixel box and centered (by center of mass) in 28×28 fields. The MNIST dataset is freely available at http://yann.lecun.com/exdb/mnist/, which also lists a large number of classification results achieved by more conventional machine learning algorithms.

2.4. Network architecture

The network architecture is shown in Fig. 4. The network consisted of 71,026 neurons (small circles) organized in four layers (dotted boxes), and approximately 133 million synapses (denoted

by arrows; triangles: excitatory synapses, filled circles: inhibitory synapses). MNIST grayscale images (input layer) were fed through a feed-forward network consisting of V1 and V2, which then projected to a downstream population of decision neurons (output layer). V1 and V2 populations were previously implemented and tested in a Compute Unified Device Architecture (CUDA) environment (Richert et al., 2011). We thus extended the existing implementation to train a layer of downstream decision neurons on V2 representations of the MNIST database using the synaptic plasticity rule described above in combination with an excitatory teacher signal (see below). Parameter values such as population size, connection probability, and connection strength are listed in Table 1.

V1 simple and complex cell responses were constructed by using the first stage of the (rate-based) motion energy model (Simoncelli & Heeger, 1998). In this model, a simulated V1 simple cell computes an inner product of the image contrast with one of 28 space–time oriented receptive fields (third derivatives of a Gaussian, which are very similar to a Gabor filter), which is then half-rectified, squared, and normalized within a Gaussian envelope (Gaussian half-width $\sigma = 3.2$ pixels). V1 complex cell responses were computed as a weighted sum of simple cell afferents that have the same space–time orientation and phase, but are distributed over a local spatial region (Gaussian half-width $\sigma = 1.6$ pixels). Please refer to Simoncelli and Heeger (1998) for a complete model description and a justification of model parameters. The 28×28 pixel images in the MNIST dataset were processed at three different spatial scales (indicated by the three banks of neuron populations in the V1 layer of Fig. 4) resulting in 84 filter responses per pixel, where the first scale was equivalent to processing at the original image resolution and each subsequent scale reduced the resolution by a factor of $\sqrt{2}$. Filter responses were then interpreted as the mean firing rates of Poisson spike trains ($f_{\max} = 50$ Hz, as will be explained in Section 2.5), and make up the full V1 layer of 65,856 Poisson spike generators.

Simulated spiking neurons in V2 received input only from V1 complex cells that were more selective to orientation than motion, across all three spatial scales. V2 neurons preferentially responded to one of four spatial orientations in 45° increments (horizontal "H", right-diagonal "RD", vertical "V", and left-diagonal "LD"). The orientation layer thus consisted of four pools of 28×28 neurons (which correspond to the four pools illustrated in Fig. 4), or a total of 3136 neurons. The neurons were broadly tuned such that, for example, an LD-tuned neuron responded strongest to LD stimuli but to a lesser degree also to V and H stimuli. Inhibitory neurons in the orientation layer pooled over the activity of neurons in a Gaussian neighborhood selective to a specific orientation (Gaussian half-width $\sigma = 6.4$ pixels) and sent inhibitory projections to neurons selective to the anti-preferred direction. For more information about the implementation details please refer to Richert et al. (2011).

The V2 orientation layer was fully connected to a decision layer of 150 excitatory neurons, responsible for integrating the sensory information over time. A good candidate for performing this integration in humans might be dlPFC, as this region of the brain is believed to have general decision-making functions, independent of stimulus and response modalities (Heekeren et al., 2004). In the macaque several regions have been found to have integrator functions, such as dlPFC (Kim & Shadlen, 1999), LIP (Shadlen & Newsome, 2001), or SC (Horwitz & Newsome, 1999). Future iterations of our model should be extended to include higher-order visual areas, such as areas in the inferotemporal gyrus (Shum et al., 2013), which would then project to an integrator area such as the dlPFC.

In addition to these plastic connections, divisive normalization was used to account for large variations in V2 layer activity. This normalization was implemented as an inhibitory population of 800 neurons that was uniformly connected to all neurons in both the orientation and the decision layer (20% connection probability). The weight of these projections was set such that the inhibitory population had a non-zero firing rate for all presented input patterns.

Neurons in the decision layer were equally divided into ten pools of 15 neurons each (see output layer of Fig. 4). Each neuron sent excitatory projections to other neurons in the same pool, and inhibitory projections to all other pools. Upon each stimulus presentation in the training phase, one of the ten pools received an additional excitatory signal from a teacher population of 15 Poisson spike generators that drove neurons towards a firing rate of 50 Hz, at which learning should occur. The signal was sent to the jth pool whenever a stimulus of class j was presented (e.g., the first decision pool received a teaching signal whenever a zero was presented to the network). As a result of this specific connectivity, each pool became selective after training to one class of input stimuli from the MNIST dataset; that is, one of the ten digits.

Inhibitory populations were approximately four times smaller than their excitatory counterpart (784 inhibitory versus 3136 excitatory neurons in the orientation layer; 800 versus 3136 neurons in the decision layer), which is thought to be the ratio of excitatory to inhibitory neurons found in the mammalian cortex (Braitenberg & Schuz, 1998).

2.5. Simulation procedure

The simulation consisted of the numerical integration of the equations for the Izhikevich neuron model (1)–(3), for the synaptic conductances (4)–(5), and for the variables involved in the synaptic learning rule (6)–(8) using the forward-Euler method. Whereas the basic integration step was 1 ms for (4)–(8), the integration step was set to 0.5 ms for (1)–(3) in order to avoid numerical instabilities associated with fast spiking activity (Izhikevich et al., 2004). Although this approach yields adequate accuracy in our simulations, other scenarios might exist where restricting the spike times to multiples of the integration step will lead to significant integration errors. These errors may be avoided by interpolating the exact spike times (Morrison, Straube, Plesser, & Diesmann, 2007), or by using a more accurate numerical integration method (Stewart & Bair, 2009).

The simulation comprised three stages: a pre-processing stage in which the MNIST dataset was converted into orientation responses, a training phase, and a testing phase.

The pre-processing stage was performed only once initially. Our implementation of the motion energy model directly takes grayscale movies as an input, and returns the total of 84 space–time-oriented filter responses per pixel as an output (28 filters at three spatial scales). Because the motion energy model expects movies rather than still images as an input, we presented each MNIST image for 100 frames and averaged the response of each cell in the V2 orientation layer to get a quasi-steady-state response for each image. In order to make these responses compatible with the protocol of the synaptic plasticity rule, we linearly mapped the response magnitudes onto the interval $[f_{\min} = 2$ Hz, $f_{\max} = 50$ Hz$]$. This was to ensure that all firing rates in the orientation layer stayed within reasonable boundaries, that each pattern had at least one element with firing rate $f_{\max}$, and that the minimum response was a non-zero "spontaneous" firing rate $f_{\min}$. The upper bound was set to 50 Hz because this is the mean rate at which it is most likely to get LTP (see Fig. 3). However, this procedure is merely a technical measure of precaution, as the motion energy model is perfectly capable of producing V1 (and V2) responses within a few milliseconds. The average orientation layer responses were converted to Poisson spike trains. These activity patterns were taken

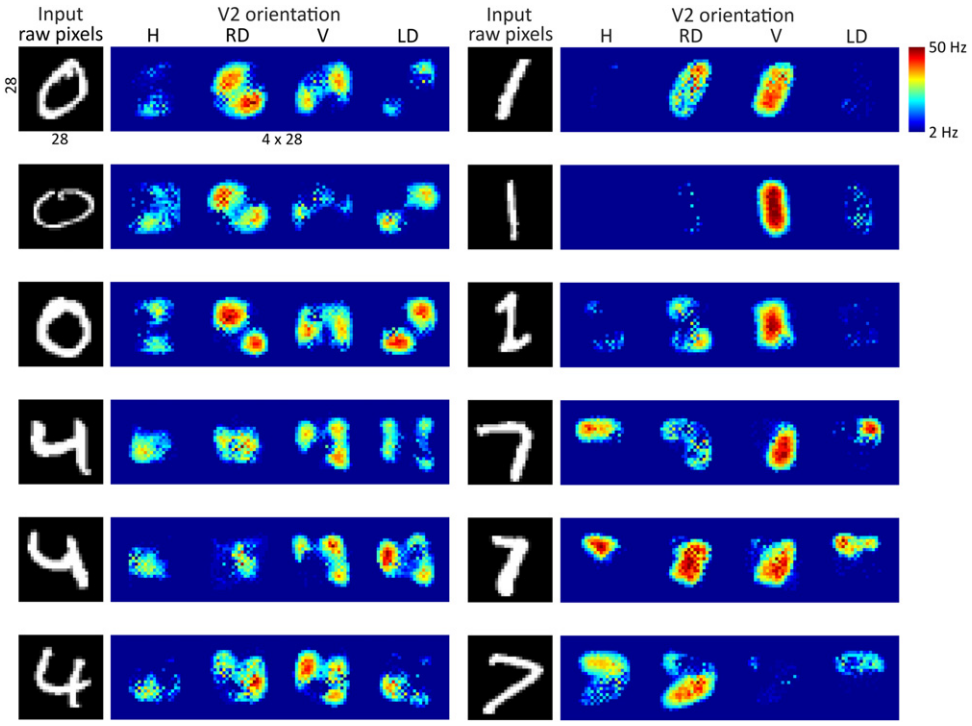

Fig. 5. Some MNIST exemplars (input layer) and their corresponding orientation-selective responses (orientation layer). The orientation layer consisted of four pools of 28×28 neurons responding to one of four spatial orientations (horizontal "H", right-diagonal "RD", vertical "V", and left-diagonal "LD"). Firing rates are color-coded ranging from 2 Hz (blue) to 50 Hz (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

randomly from the dataset, under the one restriction that the training and testing sets be mutually exclusive.

Fig. 5 illustrates some MNIST exemplars and their corresponding orientation-selective responses. Firing rates are color-coded ranging from 2 Hz (blue) to 50 Hz (red).

At the beginning of an experiment, all synaptic internal variables and plastic weights were set to zero. In order to avoid synchronous bursting of network activity upon the first stimulus presentation, we uniformly distributed the membrane potential of all excitatory neurons between −95 mV and −65 mV, and the Calcium variable between zero and three.

During training, a total number of n_{train} activity patterns belonging to ten different stimulus classes (the digits zero through nine) were presented sequentially to the output layer for n_{cycles} training cycles. A variety of values were tested for these parameters, all of which can be found in Table 1. Blocks of ten stimulus presentations or trials were set up such that in each block the ten stimuli (one exemplar of each class) were presented in random order and without repetition. This rather restrictive protocol was introduced to ensure that, even for the relatively small training sets, after k blocks all ten digits had been presented exactly k times. However, it would be interesting to see whether this restriction can be relaxed to a more natural presentation order when dealing exclusively with relatively large training sets.

In each trial, the stimulus was presented for a period of $t_{\text{stim}} = 500$ ms, which is a common value in object recognition tasks (Mongillo, Curti, Romani, & Amit, 2005; Rolls, 2012). However, other values are possible, where longer stimulation periods tend to slightly increase the likelihood of synaptic transitions (Brader et al., 2007).

The teacher population sent an additional excitatory signal to those output neurons that should be selective for the presented stimulus. These connections were set such that at the beginning of the training phase the output neuron pool fired at approximately 50 Hz, which is the firing rate that corresponds to the highest LTP probability (compare Fig. 3). The synaptic learning rule would then select a number of synapses for potentiation based on the activity of their presynaptic and postsynaptic cells. As training progressed an increasing number of synapses got potentiated, which caused the firing rate of the decision neurons to rise. Recall that at firing rates above 50 Hz the LTP transition rate drops off; because it is increasingly less likely that $C(t)$ falls in the LTP-relevant regime (see Fig. 1). Thus in the second half of the training phase a synapse potentiation got increasingly rare. We ended the training after a total of 2000 stimulus presentations.

Every stimulus presentation was followed by a delay period of $t_{\text{delay}} = 1$ s, in which all the input neurons fired at $f_{\text{min}} = 2$ Hz mean rate (in order to simulate spontaneous input activity), and the teacher remained silent. In our model, the delay activity in the decision layer started out to be zero at the beginning of training, but as learning progressed activity tended to gradually increase from 0 Hz at the beginning to roughly 10 Hz at the end of training. Towards the end of the training phase, this activity persisted throughout the entire delay period. Because the network learns gradually, this activity is similar but not equivalent to the network response that was evoked by the last presented stimulus. Interestingly, similar selective delay activity has been observed in inferotemporal, prefrontal, and entorhinal cortices in monkeys when performing a delay-response task, and is believed to be mediated by slow NMDA currents (Erickson & Desimone, 1999; Nakamura & Kubota, 1995).

Following the training, the teacher signal was removed, and the learning mechanism was turned off. The network activity was then purely driven by the input activity and the learned weight matrix. The training and testing sets were presented to the network ($t_{\text{stim}} = 500$ ms, $t_{\text{delay}} = 1$ s) to evaluate the error rates.

For classification we adopted a computational framework from a motion discrimination task, where the decision-making was modeled as the integration of sensory evidence over time (Shadlen & Newsome, 2001). In this model, each pool of output neurons accumulates evidence for a particular stimulus class separately. The first pool to emit 75 spikes (on average five spikes per neuron)

wins the race, and thus signals a choice for that stimulus class. The time it takes the first pool to reach the decision threshold is termed the RT. Note that, the mutual inhibition of decision pools is a fundamental modification of the conventional race model (Smith & Ratcliff, 2004). For more information see Section 4.3.

Based on this rule, three different outcomes are possible:

(i) *correctly classified*: the correct stimulus pool of output neurons wins the vote.
(ii) *misclassified*: an incorrect stimulus pool wins the vote.
(iii) *non-classified* ("forced choice"): if no decision was made after the stimulation period of 500 ms, we forced the network to make a choice by selecting the most active pool as the winner of the vote.

Due to the stochastic nature of synaptic transitions it is possible that the network performance changes significantly from run to run. Thus, in order to get a robust performance metric, we ran each experiment 100-fold on randomly sampled patterns and report the average result over all experimental runs.

3. Results

We addressed the question of how many training samples are needed to allow good classification by varying the size of the training set between ten patterns (one per digit) and 2000 patterns (200 per digit). The testing set always consisted of 1000 patterns the network had not seen before. We ran a total of four experiments, where each experiment featured a set of n_{train} training samples, and each experiment was run 100 times. The number of training cycles was adjusted such that overall 2000 stimuli were presented in the training phase of each experimental run (e.g., if the training set had $n_{\text{train}} = 100$ training patterns, they were presented for $n_{\text{cycles}} = 20$ training cycles, such that $n_{\text{train}} * n_{\text{cycles}} = 2000$). The full network ran in real-time on the GPU, and was roughly eight times slower than real-time when run on the CPU.

The resulting misclassification rates are plotted in Fig. 6, where each bar represents the mean misclassification rate in 100 runs of the same experiment and the error bars are the standard deviation. Whereas the training and testing set sizes were the same in each experiment, the actual patterns were drawn randomly from MNIST in each run. If only ten training patterns (navy blue) were used, the network was able to "memorize" them (achieve zero error) in 98 out of 100 runs. Even 100 training patterns (cyan) could be memorized in two out of 100 runs. However, a small training set is clearly underrepresenting MNIST, which is reflected in the large testing error.

It is worth noting that the network was able to achieve its best testing error after a one-time experience of 2000 patterns (maroon). Generally speaking, the network was able to generalize best to new data if the training set was large. This was, however, only true up to a certain point. In fact, the distributions of the misclassification rate for training set sizes $n_{\text{train}} = 1000$ (gold) and $n_{\text{train}} = 2000$ (maroon) were not significantly different (two-sample t-test, $p \cong 0.75$), indicating that these training sets were equally representative of the data. Thus we did not study larger training sets. However, other studies argued that MNIST is too small for some classifiers to infer generalization properly, and thus vastly increased the training set through a variety of distortion techniques (Ciresan et al., 2011; Simard et al., 2003). Thus, it is possible that applying similar techniques to our model would aid the classification performance.

The results for the experiment with 2000 training and 1000 testing patterns are plotted again in Fig. 7 for further investigation. On average, the network scored 91.88% and 91.64% correct classifications (green) on the training (Panel A) and testing set (Panel B), respectively. Analogously, the misclassification rates (red) were

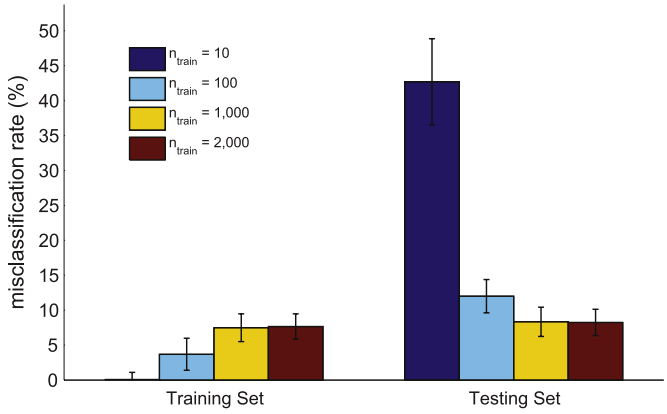

Fig. 6. Misclassification rate on the training (left) and testing set (right). Data was collected from experiments with varying training set sizes and 1000 testing patterns (100 simulation runs each). As the training set size increased, the generalization error (right panel) decreased. The distributions of the misclassification rate for training set sizes 1000 and 2000 were not significantly different (two-sample t-test, $p \cong 0.75$).

7.66% and 8.24%, and the non-classification rates (blue) were 0.46% and 0.12%. Similar performance on the training and testing sets indicates that our model is unlikely to overfit the training data. The relatively low non-classification rates (as compared to misclassification rates) are due to the network being able to reach the decision threshold within 500 ms even for stimuli that evoked relatively low responses. This was necessary in order to simulate psychophysical RT distributions (see Fig. 10). However, other thresholds may exist that would allow for most misclassifications to be turned into non-classifications (as was done in Brader et al. (2007)). Even though the network did not reach the decision threshold for 0.58% of all patterns, the forced choice in these non-classified trials was correct 78% of the time. The best observed result in a single round was 95% correct classifications (observed in roughly 10% of all cross-validation rounds, see Panel D). According to the Jacques–Bera test of composite normality all distributions in Fig. 7 were approximately Gaussian, which indicates that the classification result is robust, repeatable, and as such does not allow a significant fraction of outliers. Error bars in Panels A and B indicate the standard deviation. Misclassification rates are plotted again as a normalized histogram (relative frequency) for the training set (Panel C) and for the testing set (Panel D). Vertical dotted lines indicate the mean.

This result is comparable to other SNN models of similar complexity, such as a memristor-based SNN that achieved 93.5% accuracy on the full MNIST set (Querlioz et al., 2011), but is slightly worse than the 96.5% accuracy reported in the original study (Brader et al., 2007), which was achieved on 10,000 randomly selected patterns (no information is given about the reproducibility of this result). Needless to say, the average performance of our network compares unfavorably with state-of-the-art machine learning methods (Ciresan et al., 2011; Niu & Suen, 2012; Simard et al., 2003), of which a hybrid CNN-support vector machine (SVM) classifier was able to achieve an error rate of 0.19% (Niu & Suen, 2012). Refer to Section 4.2 for a more exhaustive performance assessment.

The network made a range of different predictive errors, all of which are shown in the confusion matrix in Fig. 8. In this grayscale map, misclassifications are grouped according to the actual label of the presented pattern (rows) and the class prediction (columns). Number of occurrences are indicated on the right, and ranged from zero (black) to 1200 (white). For example, the matrix element in the lower left corner shows the number of occurrences where the digit nine had been misclassified as a zero. Correct classifications (which would lie on the diagonal) were not included in the plot. The network most easily recognized the number one, followed by

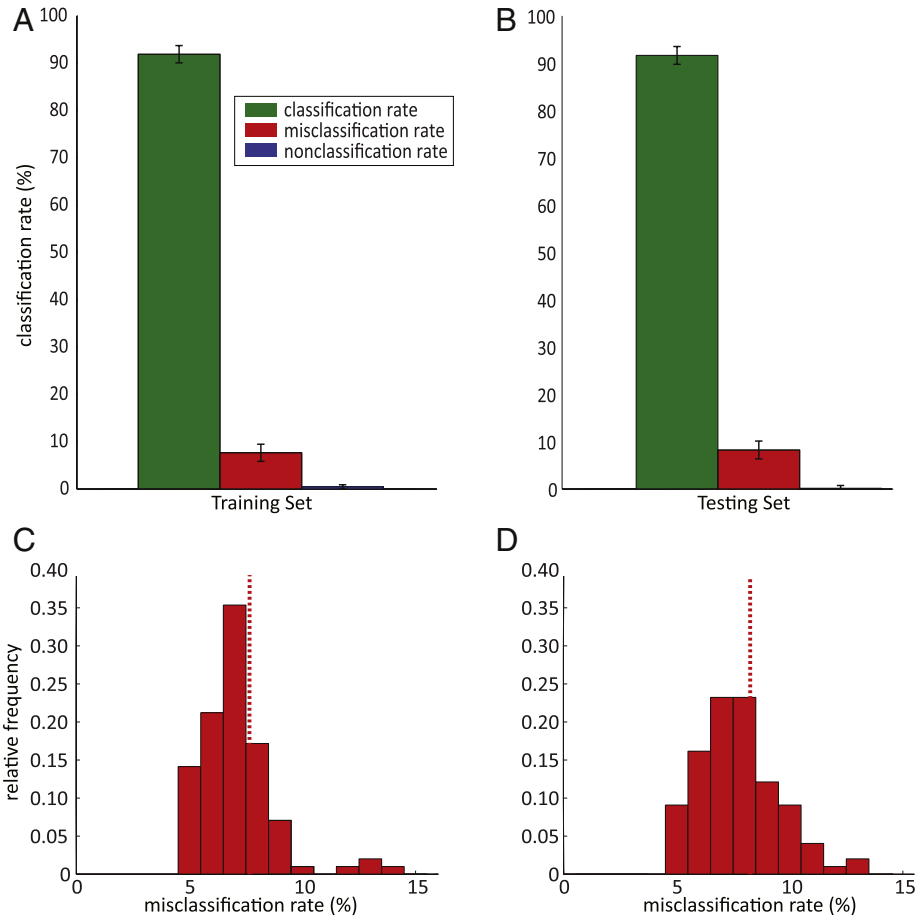

Fig. 7. Classification performance for 100 simulation runs with 2000 training and 1000 testing patterns. Classification rates are given for correctly classified samples (green), misclassified samples (red), and non-classified samples (blue). Error bars indicate the standard deviation. The network scored 91.88% and 91.64% correct classifications on the training (Panel A) and testing set (Panel B). Analogously, the misclassification rates were 7.66% and 8.24%, and the non-classification rates were 0.46% and 0.12%. Panels C and D: normalized histogram of the misclassification rate on the training set (C) and testing set (D). Vertical dotted line indicate the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

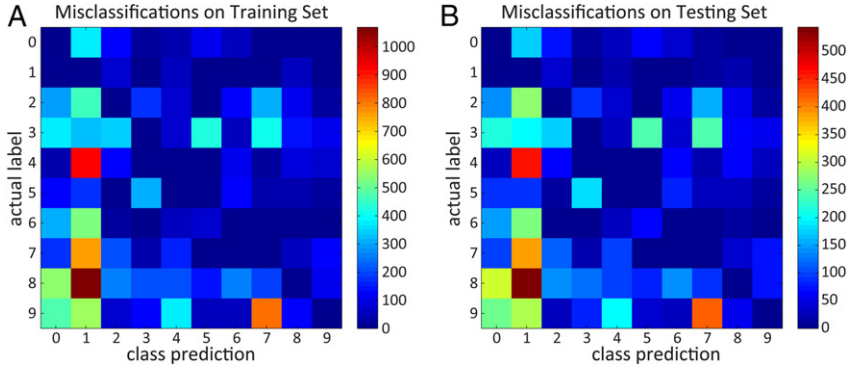

Fig. 8. Confusion matrix from training (left) and testing (right) for 100 simulation runs with 2000 training and 1000 testing patterns. The plot shows all misclassifications grouped according to the actual label of the presented pattern (rows) and the class prediction (columns) in a grayscale map, where the number of occurrences ranges from zero (black) to 1200 (white). For instance, the matrix element in the lower left corner shows the number of occurrences where the digit nine had been misclassified as a zero. Correct classifications (which would lie on the diagonal) were not included in the plot. The most common error was mistaking other digits for the number one.

the number zero. The most predictive errors were made on the digits eight and nine. Not entirely unrelated, the network was prone to mistaking other digits for a one, especially the number four ("4 → 1"), seven ("7 → 1") and eight ("8 → 1"). This particular mistake was usually made on narrowly drawn digits, whose spatial frequency posed problems on the pre-processing filters (see Section 4.4). Other frequently confused pairs were "8 → 0", "9 → 7", "3 → 5", and "3 → 8", which is not surprising as their pixel representations are highly correlated. In contrast, other pairs were never confused, such as "1 → 9", "6 → 7", and "6 → 9".

In order to find out which MNIST exemplars were the most troublesome, we counted the number of errors the network made on each pattern in the set, divided by the number of times a given pattern had been presented to the network. The resulting 50 most often confused patterns are shown in Fig. 9 in decreasing order from left to right and from top to bottom. Their actual and

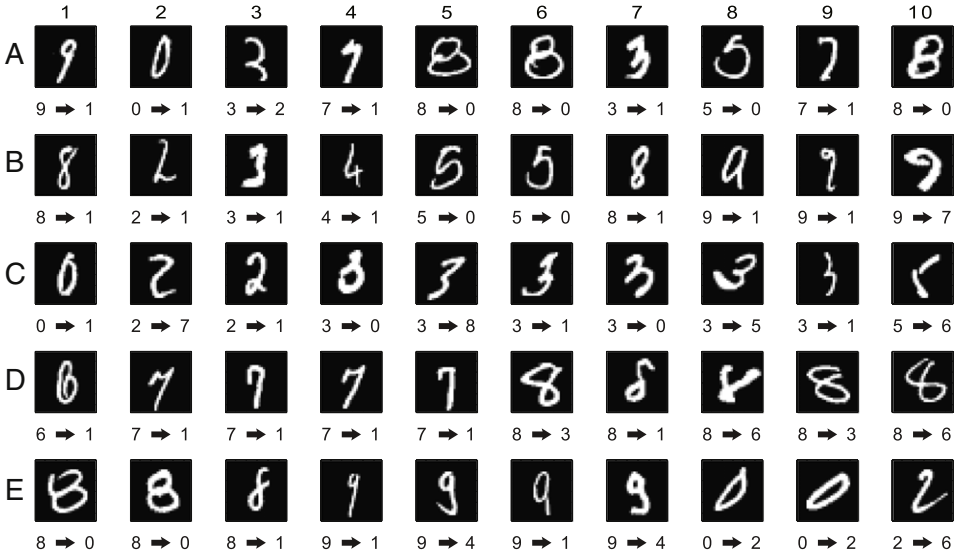

Fig. 9. The 50 most often misclassified patterns in decreasing order from left to right and from top to bottom, collected in 100 simulation runs (2000 training patterns, 1000 testing patterns). Under each image the actual label and the most common class prediction is indicated as label → prediction.

predicted labels are indicated below the image. From the pairs mentioned above the most prevalent mistakes were to confuse a seven or a nine with a one (e.g., patterns A1, A4, and A9). This is not surprising, as all three digits are very similar in shape. In addition, the network often mistook digits that have a distinct circle-like feature for a zero (e.g., patterns A5, A6, and A8). Other patterns were of degraded quality (e.g., missing stroke in patterns D7 and D8), or were at least ambiguous (e.g., pattern A3) making them hard to categorize even for humans.

We tested if the network was faster at reaching a decision when it correctly classified digits, as opposed to when it made an error. Fig. 10 shows the histogram of the RT in training (Panel A) and testing (Panel C), which are comparable to RT distributions reported in Fig. 2 of Fabre-Thorpe et al. (1998). One might prefer to explore the data from Panel A and Panel C in box plot format (see Panel B and Panel D, respectively). RTs involved in a correct decision and in a wrong decision are shown in green and red, respectively. Vertical dashed lines indicate the median RT for correct (green) and wrong (red) decisions. All distributions were positively skewed, which is a direct consequence of the race model (Smith & Ratcliff, 2004). The median response time of the network for a correct decision was 289 ms on both the training and testing set. For misclassifications the median response time was 321 ms and 320 ms on the training and testing set, respectively. Overall the shortest RT was 178 ms. Even though predictive errors took longer than correct responses on average, the network still made fast errors; a quality that cannot be reproduced by the conventional race model (Smith & Ratcliff, 2004).

Using the Kolmogorov–Smirnov test we verified that the distributions of RTs of correct and wrong decisions were significantly different in both panels ($\alpha = 0.00167, p < 0.0001$). However, the RT distributions for the correct classification of either familiar stimuli (the ones in the training set) or new stimuli (the ones in the testing set), and analogously for misclassifications, were not ($\alpha = 0.00167, p < 0.005$), which is in agreement with the behavioral evidence from psychophysical experiments (Fabre-Thorpe et al., 1998).

4. Discussion

The main contributions of the present study are as follows. First, we modified the original model (Brader et al., 2007) to be more biologically plausible most notably by (i) implementing a SNN using Izhikevich spiking neurons and conductance-based synapses, (ii) implementing the different dynamics seen in excitatory and inhibitory neurons, (iii) incorporating a pre-processing stage that approximates the spatiotemporal tuning properties of simple and complex cells in the primary visual cortex, and (iv) modeling the classification rule as a decision-making process based on the race model. Second, the network achieved 92% correct classifications in 100 rounds of random sub-sampling, which provides a conservative and reliable performance metric, and is comparable to a number of SNN and simple machine learning approaches. Third, our network correctly predicts a number of properties related to behavioral RTs reported in psychophysical experiments, namely (i) the shape of these RT distributions, (ii) shorter mean RTs for correct responses than for errors, (iii) similar RTs for both unfamiliar and familiar stimuli, and (iv) RT values that are comparable to those reported in a rapid categorization study. Overall we demonstrate how a STDP-like learning rule can be utilized to store object information in a SNN, and how a simple decision-making paradigm is able to retrieve this memory in a way that allows categorization of highly correlated patterns of neural activity. By using a SNN simulator that supports large-scale instantiations and neurobiologically plausible brain architectures (Richert et al., 2011), our model represents a first step towards the construction of a general-purpose neurobiologically inspired model of visual object recognition and perceptual decision-making.

4.1. Changes to the original model implementation

Table 2 lists all major changes made to the original model implementation by Brader et al. (2007). The most prominent differences are changes in the neuron and synapse model, in the pre-processing stage, in the simulation procedure, in the decision (classification) rule, and in the cross-validation. We will briefly discuss each of these modifications.

Employing Izhikevich spiking neurons and conductance-based synapses instead of leaky integrate and fire (LIF) neurons allows for better approximation of the subthreshold dynamics of a neuronal membrane potential while still being computationally efficient (Izhikevich, 2003). For example, the Izhikevich model is able to capture the qualitative differences between the response properties of excitatory (RS) and inhibitory (FS) neurons (e.g., the

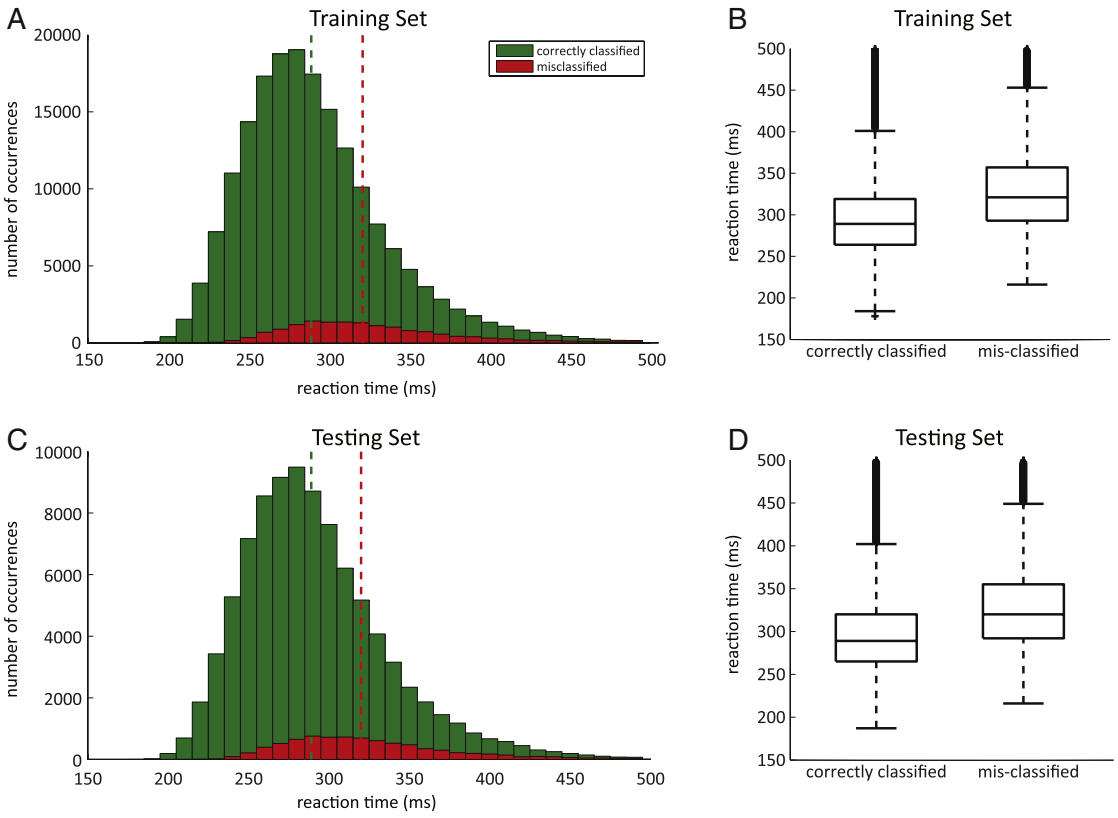

Fig. 10. Histogram and box plots of reaction times in the training (Panel A and C) and testing phase (Panel B and D). The reaction times for correct (green) and wrong decisions (red) were collected in 100 simulation runs (2000 training patterns, 1000 testing patterns). Vertical dashed lines in Panel A and C are the median of each distribution. The median response time of the network for a correct decision was 289 ms on both the training and testing set. For misclassifications the median response time was 321 ms and 320 ms on the training and testing set, respectively. Overall the shortest RT was 178 ms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison between the original model and the current implementation.

	Original model[a]	Current implementation
Neuron model:	Leaky I&F neurons with a floor	Izhikevich spiking neurons (RS & FS)
Synapse model:	Current-based	Conductance-based
Pre-processing:	None/edge detection[b]	Spatial orientation[c]
Recognition layer:	Full connectivity, no competition	Full connectivity, competition
Decision rule:	Threshold on firing rate (per neuron), majority decision rule	Accumulator model (group spiking)
# Plastic synapses:	784	$4 \times 784 = 3136$
# Classifier neurons:	$N_{\text{class}}^{\text{out}} = 15$	$N_{\text{class}}^{\text{out}} = 15$
Stimulation period:	$t_{\text{stim}} = 300$ ms	$t_{\text{stim}} = 500$ ms
Delay period:	Unspecified	$t_{\text{delay}} = 1\,s$
Drift rates:	$a = b = 1/0.35\ \text{s}^{-1}$	$a = b = 0.1\ \text{s}^{-1}$
Calcium up-jump:	$J_C = 1$	$J_C = 3.4$
# Simulation runs:	1	100
# Training patterns:	20,000	10/100/1000/2000
# Test patterns:	10,000	1000

[a] Brader et al. (2007).
[b] 8 orientations, binary responses, directional.
[c] 4 orientations, analog responses, no polarity.

firing rate of a FS neuron is a poor predictor of the strength of stimulation) which the LIF model cannot (Izhikevich, 2004). Another difference (and problem) is that the Izhikevich spiking neurons do not have a constant instantaneous threshold potential; that is, a constant level of postsynaptic depolarization above which an action potential is very likely to be triggered (Izhikevich, 2007b). This has direct implications on the learning rule under study, which requires the membrane potential to be sufficiently depolarized in order for an LTP transition to occur (compare (7)). In both the original model and the present implementation this sufficiency is defined as θ_V, a constant potential whose range of values is bounded by the resting potential and the spiking threshold. Our results indicate that adequate performance is possible with a constant θ_V. However, a more exhaustive analysis of the subthreshold path in the Izhikevich model might add to the understanding of model limitations.

Conductance-based synapses allow for distinct temporal dynamics per synaptic channel, such as the nonlinear dependence of NMDA currents on the membrane potential (see (5)). Slow NMDA currents proved to be crucial especially to prevent the excessive synaptic depression immediately following the removal of a stimulus. Moreover, slow NMDA currents allowed for the generation of persisting stimulus-selective activity throughout the delay period, which might be similar to activity observed in inferotemporal, prefrontal, and entorhinal cortices in monkeys when performing a delay-response task (Erickson & Desimone, 1999; Nakamura & Kubota, 1995). This phenomenon would be in agreement with recent theories highlighting the importance of NMDA receptors in the functioning of selective delay activity (Amit & Mongillo, 2003; Mongillo et al., 2005).

The synaptic learning rule was studied before on a dataset that had been pre-processed with binary edge filters (Amit & Mascaro, 2001; Brader et al., 2007). Two weaknesses of this transformation were mentioned by the authors: first, the filters had to be very coarsely tuned in order to model some degree of location invariance, resulting in a tuning curve that is not bell-shaped but more flat; and second, the filters were binary. Our approach overcomes both limitations by modeling the spatiotemporal tuning of simple and complex cells according to the V1 model of Simoncelli and Heeger (1998). Additionally, the classification problem benefits

from two other properties of the pre-processing transformation. First, the set of all orientation responses of a certain digit shows higher intraclass correlation than the set of all corresponding grayscale images, which is mainly due to the spatial pooling of simple cell responses. And second, the transformation increases the dimensionality of the input space from 784 to 3136 input features, which in turn increases the capacity of the classification algorithm.

Another change was made to the way a stored memory is retrieved from the system. In the original model implementation a threshold on the output activity decided whether a neuron was considered active or not, and a majority decision rule classified the data (Brader et al., 2007). While this approach bears resemblance to an ensemble of perceptron classifiers, it brings with it a set of new problems. A minor issue is that a different threshold may be required for each neuron in the ensemble. More importantly, in order to retrieve a memory it is often necessary to discriminate between very close synaptic currents, to a degree that may be biologically implausible (Fusi & Senn, 2006). We resolve this issue by introducing a decision variable, whose behavior is based on the well-founded drift–diffusion model (Bogacz et al., 2006; Smith & Ratcliff, 2004). In order to decide which stimulus is being presented to the network, sensory evidence for each possible choice is integrated separately over time, until a decision threshold is reached. The decision is then to pick the alternative which "wins the race", that is, which reaches the threshold first. This method is in agreement with a number of studies that have identified small populations of neurons in frontal cortical areas of monkeys and humans which are not only predictive of the behavioral response in a perceptual decision task, but whose behavior can be explained by drift–diffusion or race models (Philiastides & Sajda, 2006; Schall & Thompson, 1999; Shadlen & Newsome, 2001).

Other modifications include the adaptation of the stimulation period to a more common interval of 500 ms (Mongillo et al., 2005; Rolls, 2012), and the introduction of a delay period. The latter serves two purposes. First, as mentioned above it allows for the generation of persistent selective delay activity. Second, it serves as a regulatory mechanism to return the network activity to a "near-resting" state: synaptic internal variables are guaranteed to drift to one of their two stable states, thus consolidating the memory; synaptic currents as well as the postsynaptic depolarization rebound; and the Calcium variable is likely to return to values around zero. Although the original model implementation has demonstrated functionality of the synaptic learning rule in the absence of a delay period, it remains to be shown how critical this parameter is for the success of the model in more natural or less controlled environments (e.g., when embodied on a robot).

All these changes arguably improve the neurobiological plausibility of the model, while still allowing the network to perform the classification task at hand.

4.2. Performance assessment

The network demonstrates robust classification performance averaging 92% correct (with 95% as its best performance) classifications over 100 rounds of random sub-sampling. Employing this sub-sampling technique is an effort to provide a relatively conservative and reliable performance metric. In contrast to a full (k-fold) cross-validation procedure, our technique was shown to be asymptotically consistent; that is, an asymptotically consistent method selects the best prediction model with probability one as sample size $n \rightarrow \infty$, which results in more pessimistic predictions of the test data (Shao, 1993). In other words, the network is likely to score even better in a full cross-validation procedure. An advantage of random sub-sampling is that the size of the sub-samples does not depend on the number of runs. A disadvantage of this method is that some observations may never be selected in the validation sub-samples, whereas others may be selected more than once.

Whereas this result is comparable to other SNNs of similar complexity, such as a memristor-based SNN that achieved 93.5% accuracy on the full test set (Querlioz et al., 2011), the introduction of new components leads to slightly worse performance than the 96.5% accuracy reported in the original study (Brader et al., 2007). It is possible that the classification performance could be improved by choosing a different decision threshold or by further limiting the allowed duration of a response. However, since errors happen over a wide range of RT values (see Fig. 10), it is not straight-forward to find suitable parameter values that both maximize performance and reproduce the desired RT distributions. Thus more work is needed in this direction.

Similar performance could be achieved using non-spiking neuron models and neurobiologically plausible learning paradigms. For example, Huerta and Nowotny demonstrated that a network of McCulloch–Pitts neurons modeled after the insect olfactory system was able to achieve 93% accuracy on MNIST using 50,000 hidden units and a reinforcement learning rule (Huerta & Nowotny, 2009). Amit and Mascaro achieved 94% accuracy on a downscaled, pre-processed version of MNIST using perceptrons in an attractor network with 200 attractor states (Amit & Mascaro, 2001). Note that, the number of hidden units in the first study was an order of magnitude larger than in our approach (50,000 hidden units versus 3136 V2 orientation-selective cells); and that the second study reduced the dimensionality of the problem by classifying a 16×16 version of MNIST.

Not surprisingly, our network showed worse performance than machine learning algorithms of higher representational power (or "model complexity"), such as deep-hierarchical CNNs (Ciresan et al., 2011; LeCun et al., 1998; Simard et al., 2003) and a hybrid CNN-SVM classifier (Niu & Suen, 2012). The latter achieved a best error rate of 0.19%, outperforming all current CNN approaches as well as the human recognition error, which they estimated to be 0.2%. Also, these studies vastly expanded the (with 60,000 samples already strong) training set using a variety of distortion effects (e.g., affine, elastic, scaling, and rotation).

4.3. Decision-making and reaction time

In the present study the decision process was based on an accumulation model, where totals of sensory evidence race each other to a decision threshold, and the RT is determined by the first total to reach criterion (Bogacz et al., 2006; Smith & Ratcliff, 2004). The authors of these studies have pointed out that the addition of self-excitation and mutual inhibition of decision pools is a fundamental modification of the conventional race model. Self-excitatory recurrent connections are essential to enable neurons, whose membrane potential decays on the order of milliseconds, to integrate information on the timescale of hundreds of milliseconds (Smith & Ratcliff, 2004). Mutual inhibition between decision pools means that evidence for one response is evidence against all others, which allows the model to be mapped back to the diffusion model.

The RT distributions shown in Fig. 10 are comparable to the behavioral RTs reported in psychophysical experiments (Schall, 2002; Shadlen & Newsome, 2001; Thompson et al., 1996). These studies have identified populations of neurons in frontal cortical areas of monkeys, such as direction-sensitive neurons in LIP, whose firing rates are predictive of the behavioral RT. Our network correctly predicts (i) the shape of these RT distributions, (ii) shorter mean RTs for correct responses than errors, (iii) an increase in RT for "difficult" stimuli, (iv) similar RTs for both unfamiliar and familiar stimuli, and (v) RT values that are comparable to those reported in a rapid categorization study (Fabre-Thorpe et al., 1998). We will briefly discuss each of these qualities.

Fast behavioral responses are associated with a rapid rise in firing rate, whereas slow responses are associated with a slow rise (Smith & Ratcliff, 2004). According to these authors, the resulting RT distributions are right-skewed because of the geometry of the diffusion process paths: equal size differences in accumulation rate between pairs of sample paths lead to unequal size differences on the decision boundaries.

Correct responses are typically faster than errors in easy, speed-stress conditions (Smith & Ratcliff, 2004). Our network correctly predicts the ordering of the median RTs for correct responses (289 ms) and errors (roughly 320 ms), while still allowing fast errors.

When the difficulty of a decision is high, information favoring a particular response grows more slowly (Smith & Ratcliff, 2004), and the probability of making an error is higher (Shadlen & Newsome, 2001). The definition of difficulty is usually task-dependent. For instance, a difficult stimulus in a motion discrimination task would have a low level of coherent motion. Similarly, in the present study difficulty could be interpreted as a measure of the quality of the visual input. The stronger a presented stimulus overlaps with the learned weight matrices of multiple decision pools, the lower the difference in firing rates between selective and nonselective pools will be. Under these conditions, the network not only takes longer to reach the decision threshold, but it also is more likely to make a predictive error. The network was likely to reach the decision threshold within 500 ms even for stimuli that evoked relatively low responses (and thus might be viewed as relatively "difficult" stimuli to classify); thus favoring misclassifications over non-classifications (see Fig. 7). However, as stated in Section 4.2, other thresholds may exist that optimize classification performance.

The network categorizes handwritten digits in roughly 300 ms, with the fastest observed RT being 178 ms. Pre-processing is not included in this measure, but would only add a few milliseconds (see Section 2.5). In order to enable the network to reach these values, we chose a decision threshold (75 spikes per pool, 5 spikes per neuron on average) that yielded good classification performance. The resulting RT values are similar to the ones reported in a rapid categorization study (Fabre-Thorpe et al., 1998). In this study previously unseen images were presented to two rhesus monkeys for only 80 ms. If the picture contained one or more examples of food – or, in a second experiment, an animal – the monkey had a second to signal a "go" response by releasing a button and touching the middle of the screen. The two monkeys were able to perform go trials in as little as 250 ms and 350 ms median response time for the food and animal task, respectively. Humans took roughly 100 ms longer to perform the same task. Excluding the time it takes to execute a motor response, the median visual response time could be ranging somewhere between 150 ms and 350 ms.

Another interesting finding is that the RT of the two monkeys did not decrease when the targets had become familiar. Although misclassifications were punished, the monkeys would get some targets repeatedly wrong; despite the fact that there would have been ample opportunity to learn the stimulus features (Fabre-Thorpe et al., 1998). We compared the RT distributions for correct classification of either familiar stimuli (the ones in the training set) or new stimuli (the ones in the testing set), and indeed did not find any statistical difference.

4.4. Model capabilities and limitations

A notable implication of the synaptic learning rule is that it enables a decision neuron to learn stimulus features which are prevalent among the members of a certain stimulus class; simply because the neuron will experience more common features more often. For example, a classifier neuron which is trying to learn the shape of the number one in the MNIST database will tend to imprint the vertical stroke, or the right-diagonal stroke in cursive handwritings, in its weight matrix, not necessarily the rarer (and thus probably less defining) horizontal serif at the lower end of the digit.

Furthermore, the synaptic learning rule natively implements a local stop-learning condition. With each training cycle a number of new synaptic connections should be learned, ideally leading to a monotonously decreasing learning rate. However, unlike in other methods, the learning rate does not simply depend on the number of training cycles, but is adapted based on the correlation between the state of the synaptic weight matrix and each presented stimulus. The stop-learning criterion is biologically plausible, because it only depends on information that is readily available at the site of the synapse (Brader et al., 2007).

We would like to stress the fact that in the current configuration the learning mechanism must be deactivated for testing. This requirement is not necessarily biologically implausible, because in the brain metaplasticity mechanisms might be able to quickly and effectively alter LTP and LTD transition probabilities (Huang et al., 2012).

Judging from the confusion matrix in Fig. 8, the most common error that the network made is mistaking another digit for the number one. In fact, if we were able to remove all mistakes of this type, the average misclassification rate would shrink to 5.5%. This particular mistake was usually made on narrowly drawn images, whose spatial frequency posed problems on the pre-processing filters. In some cases weakly pronounced features, such as the horizontal stroke at the upper end of the number seven, might not have been detected by the simple cells. In other cases, the spatial pooling of linear cell responses might have washed out a small and weakly pronounced feature, such as the two circles in a narrowly drawn eight. Other pairs that have similar shape and structure, such as "5 $\rightarrow$ 3" and "9 $\rightarrow$ 7", were frequently confused. Some images are of degraded quality, making correct predictions extremely difficult even for humans.

A significant drawback of the present approach and pattern-matching in general is that the classifier neuron will not recognize an "atypical" (in terms of correlation) member of its preferred stimulus class. It has been argued that a neuron which computes the dot product of the input pattern with its synaptic weight vector is unable to implement translation invariance (Rolls, 2012). Although a neuron may be able to generalize to other patterns, its response will simply depend on the similarity of the new stimulus with what it has already learned. Any notion of translation invariance must therefore come from another mechanism, and cannot be caused by a summation of weights, however exceptional these weights might be. This is evident in Fig. 9, where some digits are deformed, tilted or shifted in a way that they strongly overlap with the learned weight matrix of another decision pool, causing the network to make predictive errors.

Nevertheless, our current approach demonstrates that the visual information encoded in simulated V2 cell responses was sufficient to categorize a large number of MNIST exemplars the network had not seen before. At least in this case, increasing the network complexity (e.g., by extending the feature hierarchy) could only marginally improve the classification performance.

5. Conclusion

We have shown the experimental results from a neurobiologically plausible spiking network that is able to rapidly categorize highly correlated patterns of neural activity. Our approach demonstrates how a STDP-like learning rule (previously described in Brader et al. (2007)) can be utilized to store object information in a SNN, and how a simple decision-making paradigm is able to

retrieve this memory in a way that allows the network to generalize to a large number of MNIST exemplars. Additionally, the RT of the network is comparable to behavioral RTs from psychophysical experiments, in that it predicts both qualitative and quantitative properties of the reported RT distributions. Our implementation is fast, efficient, and scalable; it is optimized to run on off-the-shelf GPUs, but is also well-suited to be run on neuromorphic hardware once available.

Our study is a first step into modeling large-scale cortical networks in more biological detail, with the ultimate goal of understanding whole brain dynamics such as the integration of lower-level sensory information with higher-level decision processes. Studying these processes is important, as perceptual decision-making is not only believed to be the basis of more complex decision-making routines, but could shed light on phenomena like the misperception of objects in neuropsychiatric disorders such as schizophrenia.

Acknowledgments

We thank four anonymous reviewers for their feedback, which has greatly improved the manuscript. This work was supported by the Defense Advanced Research Projects Agency (DARPA) subcontract 801888-BS.

References

Amit, Y., & Mascaro, M. (2001). Attractor networks for shape recognition. *Neural Computation*, *13*, 1415–1442.
Amit, D. J., & Mongillo, G. (2003). Selective delay activity in the cortex: phenomena and interpretation. *Cerebral Cortex*, *13*, 1139–1150.
Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, *24*, 139–166.
Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.
Brader, J. M., Senn, W., & Fusi, S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Computation*, *19*, 2881–2912.
Braitenberg, V., & Schuz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity* (2nd thoroughly revised ed.). Berlin, New York: Springer.
Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 369–384.
Ciresan, D. C., Meier, U., Masci, J., & Schmidhuber, J. (2011). Flexible, high-performance convolutional neural networks for image classification. In *IJCAI* (pp. 1237–1242).
Clopath, C., Busing, L., Vasilaki, E., & Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neuroscience*, *13*, 344–352.
Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319–349.
Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, Mass: Massachusetts Institute of Technology Press.
Erickson, C. A., & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *Journal of Neuroscience*, *19*, 10404–10416.
Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, *9*, 303–308.
Florian, R. V. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation*, *19*, 1468–1502.
French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*, 128–135.
Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202.
Fusi, S., Annunziato, M., Badoni, D., Salamon, A., & Amit, D. J. (2000). Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Computation*, *12*, 2227–2258.
Fusi, S., & Senn, W. (2006). Eluding oblivion with smart stochastic selection of synaptic updates. *Chaos*, *16*.
Gorchetchnikov, A., Versace, M., & Hasselmo, M. E. (2005). A model of STDP based on spatially and temporally local information: derivation and combination with gated decay. *Neural Networks*, *18*, 458–466.
Graupner, M., & Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 3991–3996.
Grossberg, S. (1980). How does a brain build a cognitive code. *Psychological Review*, *87*, 1–51.
Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*, 859–862.
Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, *9*, 467–479.
Horwitz, G. D., & Newsome, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, *284*, 1158–1161.
Huang, S. Y., Trevino, M., He, K. W., Ardiles, A., de Pasquale, R., Guo, Y. T., et al. (2012). Pull–push neuromodulation of LTP and LTD enables bidirectional experience-induced synaptic scaling in visual cortex. *Neuron*, *73*, 497–510.
Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional archichitecture in 2 nonstriate visual areas (18 and 19) of cat. *Journal of Neurophysiology*, *28*, 229–289.
Huerta, R., & Nowotny, T. (2009). Fast and robust learning by reinforcement signals: explorations in the insect brain (vol. 21, p. 8, 2009). *Neural Computation*, *21*, 2123–2151.
Indiveri, G., & Fusi, S. (2007). Spike-based learning in VLSI networks of spiking neurons. In *IEEE international symposium on circuits and systems*.
Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*, 1569–1572.
Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, *15*, 1063–1070.
Izhikevich, E. M. (2007a). *Dynamical systems in neuroscience: the geometry of excitability and bursting*. Cambridge, Mass: MIT Press.
Izhikevich, E. M. (2007b). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, *17*, 2443–2452.
Izhikevich, E. M., Gally, J. A., & Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex*, *14*, 933–944.
Kim, J. N., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience*, *2*, 176–185.
Koch, C. (1999). *Biophysics of computation: information processing in single neurons*. New York: Oxford University Press.
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*, 2278–2324.
Legenstein, R., Naeger, C., & Maass, W. (2005). What can a neuron learn with spike-timing-dependent plasticity? *Neural Computation*, *17*, 2337–2382.
Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*, 259–289.
Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., et al. (2012). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral Cortex*.
Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., et al. (2011). Weight consistency specifies regularities of macaque cortical networks. *Cerebral Cortex*, *21*, 1254–1272.
Mitra, S., Fusi, S., & Indiveri, G. (2009). Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI. *IEEE Transactions on Biomedical Circuits and Systems*, *3*, 32–42.
Mongillo, G., Curti, E., Romani, S., & Amit, D. J. (2005). Learning in realistic networks of spiking neurons and spike-driven plastic synapses. *European Journal of Neuroscience*, *21*, 3143–3160.
Morrison, A., Diesmann, M., & Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, *98*, 459–478.
Morrison, A., Straube, S., Plesser, H. E., & Diesmann, M. (2007). Exact subthreshold integration with continuous spike times in discrete-time neural network simulations. *Neural Computation*, *19*, 47–79.
Nakamura, K., & Kubota, K. (1995). Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *Journal of Neurophysiology*, *74*, 162–178.
Niu, X. X., & Suen, C. Y. (2012). A novel hybrid CNN-SVM classifier for recognizing handwritten digits. *Pattern Recognition*, *45*, 1318–1325.
O'Brien, M. J., & Srinivasa, N. (2013). A spiking neural model for stable reinforcement of synapses based on multiple distal rewards. *Neural Computation*, *25*, 123–156.
Persaud, R., & Cutting, J. (1991). Lateralized anomalous perceptual experiences in schizophrenia. *Psychopathology*, *24*, 365–368.
Pfister, J. P., Toyoizumi, T., Barber, D., & Gerstner, W. (2006). Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Computation*, *18*, 1318–1348.
Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, *16*, 509–518.
Porr, B., Saudargiene, A., & Worgotter, F. (2004). Analytical solution of spike-timing dependent plasticity based on synaptic biophysics. In S. Thrun, L. Saul, & B. Scholkopf (Eds.), *Advances in neural information processing systems, Vol. 16*. Cambridge, MA: MIT Press.
Querlioz, D., Bichler, O., & Gamrat, C. (2011). Simulation of a memristor-based spiking neural network immune to device variations. In *IJCNN* (p. 1775).
Richert, M., Nageswaran, J. M., Dutt, N., & Krichmar, J. L. (2011). An efficient simulation environment for modeling large-scale cortical processing. *Frontiers in Neuroinformatics*, *5*, 19.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Rolls, E. T. (2012). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Frontiers in Computational Neuroscience*, *6*, 35.
Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford, New York: Oxford University Press.
Rorie, A. E., & Newsome, W. T. (2005). A general mechanism for decision-making in the human brain? *Trends in Cognitive Sciences*, *9*, 41–43.
Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *357*, 1073–1082.
Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience*, *22*, 241–259.
Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.
Shao, J. (1993). Linear-model selection by cross-validation. *Journal of the American Statistical Association*, *88*, 486–494.
Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., et al. (2013). A brain area for visual numerals. *Journal of Neuroscience*, *33*, 6709–6715.
Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *International conference on document analysis and recognition*.
Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*, 743–761.
Sjöström, P. J., Turrigiano, G. G., & Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, *32*, 1149–1164.
Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.
Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, *3*, 919–926.
Srinivasa, N., & Cruz-Albrecht, J. M. (2012). Neuromorphic adaptive plastic scalable electronics analog learning systems. *IEEE Pulse*, *3*, 51–56.
Stewart, R. D., & Bair, W. (2009). Spiking neural network simulation: numerical integration with the Parker–Sochacki method. *Journal of Computational Neuroscience*, *27*, 115–133.
Summerfield, C., Egner, T., Mangels, J., & Hirsch, J. (2006). Mistaking a house for a face: neural correlates of misperception in healthy humans. *Cerebral Cortex*, *16*, 500–508.
Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology*, *76*, 4040–4055.
Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165.
Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer vision and pattern recognition conference, Vol. 1* (pp. 511–518).
Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167–194.