Synaptic Plasticity Across Different Time Scales and its Functional Implications

PAR

Claudia CLOPATH

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2009

Resumaziun

GLIEUD ad animals emprenden tras modifitgar la forza synaptica tranter neurons, en fenomen conuschent sco plasticitad sinaptica. Quests midaments pon essar inducis tras stimulis curts (tge duran per exempel me pocas secundas), ma, par essar nizzeval par la memorgia a la lunga, els ston esser stabils sur mains ad ons.

Experimentalists studegan la plasticidad sinaptica tras applicaziun d'ena granda quantitad da protocols. En questa tesa nus focussagn sin protocols tge crudan sut duas categorias principalas: (i) Quellas tge inducessan modificaciuns synapticas tge duran me pocas uras (fasa tempriva da plasticitad) (ii) quellas tge permetten synapticas en ena sequenza da zaps par transformar las midadas rapidas tge succedan durant la "fasa tempriva" en en fastiz d'ena memorgia stabila (la "fase tardiva" da plasticitad).

La mira da questa tesa e da meglier capir la plasticitad sinaptica tras quellas fasas differentas, tempriva e tardiva, tras crear models mathematics compacts par descriver igl mecanissem da la plasticitad. La nossa proposta permetta ena vista syntetica d'igl champ e l'exploraziun da consequenzas funzionalas dal emprender. En questa direcziun nus proponeschan en model par la inducziun da plasticitad sinaptica tge dependa dal piz d'igl impuls presinaptic e nonlinear da la tensiun plastizica. Quest model e capabel da reproducir en grand sectur da protocols experimentals sco par exempel experiments digl clupper da tensiun u experiments da temp d'impuls. Perquai che l'impuls es en element da clav in quest model, nus descrivan la activitad neuronala tras utilisar en model compact da neurons tge reproduzessa fidaivel igl curs d'igl temp dagl impuls da neurons pyramidals. En pli, quest model dagl inducziun da plasticitads sinapticas e cumbino cun en process scludider par la sintesa da protein ed igl mecanissem per la stabilsaziun finala par descriver la "fasa tardiva". En questa forma cumbinada, igl model e capabel da declarar fenomens experimentels conuschent sco experiments d'identifitgar e da far predicziuns controllabels. Studis da consequenzas funzionalas d'igl model d'inducziun scuvrischan selectivitads d'igls inputs, analisa da computaziun da components independents ed ena storta relaziun tranter connectividad e codaziun.

En parallel ena proposta da sura en bassa e utiliso par chattar components independents par deducir ena regla dad emprender tge se funda sin ena qual-

ificaziun tge mussa correlaziuns cun igl model d'inducziun. Quest model unid tras differentas dimensiuns da temp tge permetta la stabilisaziun da sinapsas e fetg important par capir igls process dad emprendar e da memoria en animals e glieud e funda ena ingredienza basegnaivla par mintga model da granda dimensiun digl tscharvi.

Pleds da clav: Neuroscienza fundada sin calculaziun, Plasticitad Sinaptica, Emprender e memoria/regurdanza, "Fasa tempriva" da potenzaziun a lunga vista, "Fasa tardiva" da potenzaziun a lunga vista, Identificaziun sinaptica, Analisa da components independents, Codaziun.

ABSTRACT

Humans and animals learn by modifying the synaptic strength between neurons, a phenomenon known as synaptic plasticity. These changes can be induced by rather short stimuli (lasting, for instance, only a few seconds), yet, in order to be useful for long-term memory, they should remain stable for months or years.

Experimentalists study synaptic plasticity by applying a vast variety of protocols. In the present thesis we focus on protocols that fall under two main categories: (i) Those that induce synaptic modifications that last for only a few hours ("early phase" of plasticity) (ii) Those that allow synapses to undergo a sequence of steps that transforms the rapid changes occuring during the "early phase" into a stable memory trace ("late phase" of plasticity).

The goal of this thesis is to better understand synaptic plasticity across these different phases, early and late, by creating compact mathematic models to describe the plasticity mechanisms. Our approach allows for a synthetic view of the field as well as the exploration of functional consequences of learning. In this direction, we propose a model for the induction of synaptic plasticity that depends on the presynaptic spike time and nonlinearly on the postsynaptic voltage. The model is able to reproduce a broad range of experimental protocols such as voltage-clamp experiments and spike-timing experiments. Since the voltage is a key element in the model, we describe the neuronal activity by using a compact neuron model that faithfully reproduces the voltage time course of pyramidal neurons. In addition, this model of the induction of synaptic plasticity is combined with a trigger process for protein synthesis, and the final stabilization mechanism in order to describe the "late phase". In this combinatory form, it is able to explain experimental phenomena known as tagging experiments and to make testable predictions. A study of functional consequences of the induction model reveals selectivity in the inputs, independent component analysis computation and a tight relation between connectivity and coding.

In parallel a top-down approach finding independent components is used to derive a rate-based learning rule which shows structural correlations with the induction model. This unified model across different time scales allowing the stabilization of synapses is crucial to understand learning and memory processes

in animals and humans, and a necessary ingredient for any large-scale model of the brain.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

INTRODUCTION

1.1 EXPERIMENTAL BACKGROUND

A NIMALS have an incredible ability to learn and memorize experiences during their life. Memory, or at least a part of memory, is believed to be stored in the connections between the neurons, the synapses. They potentially offer a very large memory capacity since there are about 100 to 500 trillions, i.e. 10^{14} synapses in a human brain. Moreover, the synapses have been shown to be plastic, i.e. their strength is variable. The basic idea of how synapses change was already proposed by Hebb in 1949 in his postulate "When an axon of cell j repeatedly or persistently takes part in firing cell i, then j's efficiency as one of the cells firing i is increased" (Hebb 1949). It is only about 25 years later that synaptic potentiation was measured experimentally in anesthetized rabbits (Bliss and Gardner-Medwin 1973). A few years later synaptic depression was measured (Lynch, Dunwiddie, and Gribkoff 1977).

Two neurons can be connected via chemical and/or electrical (gap junctions) synapses. A chemical synapse is placed between the axon of a presynaptic neuron and a dendrite of the postsynaptic neuron forming an unidirectional connection. This type of synapse is the main focus of the thesis; the electrical one is not considered here. A simplified view of the synaptic communication mechanism is the following: When a presynaptic spike arrives at the presynaptic terminal, neurotransmitters, typically glutamate for excitatory synapses, are released. They can bind at the postsynaptic side to different receptors like N-methyl D-aspartate (NMDA) or α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors. AMPA receptors then open and let sodium ions flow into the cell resulting in a depolarization of the postsynaptic membrane potential. The NMDA receptor opens only if at the same time (a) the glutamate binds to the receptors and (b) the postsynaptic cell is depolarized freeing the magnesium block (see Fig 1.1). This depolarization of the postsynaptic cell can typically come from the back-propagating action potential. The NMDA receptor therefore is seen as a coincidence detector between presynaptic and postsynaptic activities. The opening of these receptors allows calcium to enter the synapse. Despite the detailed description of the mechanisms that allow synaptic commu-
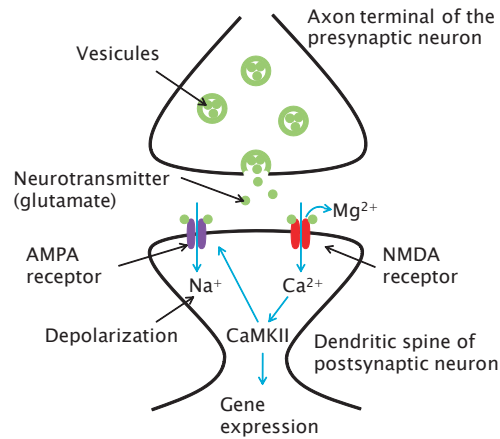
Figure 1.1: Cartoon of a glutamatergic synapse. When a presynaptic spike arrives at the synapse, neurotransmitters are released in the synaptic cleft. They can bind to the NMDA receptor and if at the same time the postsynaptic cell is depolarized, the channel opens. Calcium enters the cell which induces a molecular cascade that phosphorylates the Ca2+/calmodulin-dependent protein kinases (CaMKII), which in turn acts on the AMPA receptor activation. Neurotransmitters can also directly bind to AMPA receptor, in which case the channel opens allowing sodium to enter, leading to a depolarization which is called excitatory postsynaptic potential (EPSP).

nication (Rubin, Gerkin, Bi, and Chow 2005; Lisman and Zhabotinsky 2001), the mechanism that leads to changes in the synaptic strength is not completely clear. Calcium seems to play an important role for further cascade signalling which acts on AMPA receptors activation through kinases and phosphatases (Lisman, Schulman, and Cline 2002). Moreover retrograde messengers like endocannabinoids seem to be important, at least for depression of the synapses (Sjöström, Turrigiano, and Nelson 2003; Piomelli 2003; Sjöström, Turrigiano, and Nelson 2004; Nevian and Sakmann 2006).

Synaptic plasticity measurements

The typical measurement of synaptic strength is the amplitude or the slope of the excitatory postsynaptic potential (EPSP), i.e. the potential response to a single (or a group of coincident) presynaptic spike(s). Synaptic plasticity can be separated in two distinct phenomena: short-term plasticity where changes

persist during few hundreds of milliseconds and long-term plasticity that lasts more than 30 minutes. Short-term plasticity is believed to be presynaptically expressed (Markram, Wu, and Tosdyks 1998; Gupta, Wang, and Markram 2000) and is partly caused by the limited number of neurotransmitters at the synapse. It is well modeled by a reservoir that partially empties with some probability at the time of a presynaptic spike and recovers with a certain time constant (Markram and Tsodyks 1996; Abbott, Varela, Sen, and Nelson 1997). This however is not the interest of this thesis. Here the focus is on long-term plasticity, which can be induced by different types of protocols.

(a) *Simultaneous voltage clamp and presynaptic stimulations* (Ngezahayo, Schachner, and Artola 2000; Ling, Benardo, Serrano, Blace, Kelly, Crary, and Sacktor 2002) (Fig. 1.2A). When the cell is slightly depolarized the synaptic weight is depressed, whereas it is potentiated when the cell is highly depolarized.

(b) *Extracellular presynaptic stimulations at different frequencies* (Kelso, Ganong, and Brown 1986; Dudek and Bear 1993; O'Connor, Wittenberg, and Wang. 2005) (Fig. 1.2B). Low frequency stimulation leads to depression, whereas high frequency leads to potentiation .

(c) *Pairing of presynaptic and postsynaptic spikes at different time lags* (Fig. 1.2C). Typically in pyramidal cells pre-post pairing results in potentiation whereas post-pre in depression (Markram, Lübke, Frotscher, and Sakmann 1997; Bi and Poo 1998). However, it seems that in spiny-stellate neurons (Egger, Feldmeyer, and Sakmann 1999) and in synapses from pyramidal cells onto fast spiking interneurons (Lu, Li, Zhao, ming Poo, and Zhang 2007), pre-post and post-pre pairing both leads to depression; in a cerebellum-like structure of the electrical fish, the temporal order seems to be reversed, i.e. pre-post pairing leads to depression and post-pre to potentiation (Bell, Han, Sugawara, and Grant 1997); and in synapses of parallel fibers onto cartwheel cells, only pre-post pairing leads to depression (Tzounopoulos, Kim, Oertel, and Trussell 2004).

(d) *Pairing at different frequencies* (Markram, Lübke, Frotscher, and Sakmann 1997; Sjöström, Turrigiano, and Nelson 2001) (Fig. 1.2D). Pre-post pairing at low frequency does not change the synaptic weight, increasing the frequency leads to potentiation. Synaptic plasticity can also be induced with different patterns like triplets of spikes (Froemke and Dan 2002), bursts (Nevian and Sakmann 2006; Gustafsson, Wigstrom, Abraham, and Huang 1987), quadruplets (Wang, Gerkin, Nauen, and Bi 2005) or even natural spike trains (Froemke and Dan 2002). There is also a difference if the presynaptic stimulation is done extracellularly (many inputs at the same time, slices more active) or intracellularly.

(e) *Synaptic tagging experiments* (Frey and Morris 1997). These experiments provide evidence for another separation of time scales for long-term plasticity. The early phase of long-term plasticity lasts 2 to 3 hours and is induced by tetanic stimulation for potentiation (Fig. 1.2E). The late phase, or consolidation phase, however lasts more than 10 hours (i.e. the time of those experiments)

3

and is induced by a stronger extracellular tetanus that also stimulates dopaminergic fibers (Fig. 1.2F). Neuromodulation is thus important in the process of consolidation (Reymann and Frey 2007).

1.2 Theoretical Background

1.2.1 Rate Models

The first models of synaptic plasticity in line with Hebb's principle ("Who fires together wires together") depend on the correlation of presynaptic and postsynaptic activities, typically the firing rates. The next improvements were to subtract a baseline so that the weight can also be depressed (covariance rule (Sejnowski and Tesauro 1989)), add weight dependency, like hard bounds, soft bounds, add normalization to induce competition between weights (Miller and MacKay 1994). Multiplicative normalization was introduced in Oja's rule (Oja 1982) which performs Principle Component Analysis (PCA). At the same time the Bienenstock, Cooper, Munro (BCM) rule became influential (Bienenstock, Cooper, and Munroe 1982). It has a nonlinearity in the postsynaptic firing rate and a sliding threshold as homeostasis. It exhibits properties of selectivity in the inputs. Those rate based rules were used in artificial neuron networks like Hopfield networks (Hopfield 1982) and Boltzmann machines (Hinton and Sejnowski 1983), for map formation (von der Malsburg 1973; Kohonen 1990; Bednar and Miikkulainen 2000), receptive field development (Linsker 1986; Bienenstock, Cooper, and Munroe 1982), among others.

1.2.2 Spike Models

In 1996 a theoretical work looking at precise temporal coding (Gerstner, Kempter, van Hemmen, and Wagner 1996) suggested that synaptic plasticity should depend on the time between the presynaptic and the postsynaptic spikes. In parallel Markram et al. were able to show the spike timing dependence experimentally in neocortical slices (Markram, Lübke, Frotscher, and Sakmann 1997). From then on, Spike-Timing-Dependent Plasticity (STDP) models became influential (Gerstner, Kempter, van Hemmen, and Wagner 1996; Kempter, Gerstner, and van Hemmen 1999; Senn, Tsodyks, and Markram 2001; Song, Miller, and Abbott 2000; van Rossum, Bi, and Turrigiano 2000; Rubin, Lee, and Sompolinsky 2001; Gütig, Aharonov, Rotter, and Sompolinsky 2003a; Karmarkar and Buonomano 2002). These models typically use local variables, i.e. the weight change depends on a presynaptic and postsynaptic trace. Every time a presynaptic spike occurs, the weight decreases by an amount corresponding to
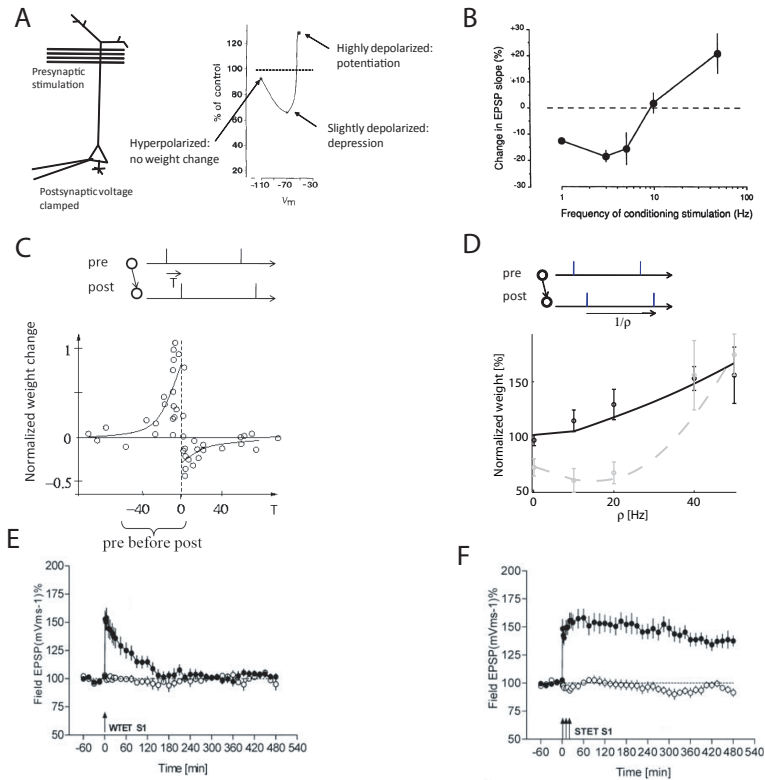
Figure 1.2: Different experimental protocols. A. Voltage clamp experiment. The postsynaptic voltage is clamped at the soma during extracellular presynaptic stimulation. If the voltage is hyperpolarized, no weight change is recorded; for slight depolarization, depression is observed; for strong depolarization, potentiation occurs. Figure redrawn from (Artola, Bröcher, and Singer 1990). B. Presynaptic frequency dependence. Extracellular presynaptic spike trains at different frequencies are induced. Low frequency stimulation yields depression whereas high frequency stimulation results in potentiation. Figure redrawn from (Dudek and Bear 1993) C. STDP experiment. Pairs of presynaptic and postsynaptic spikes are elicited. The lag between the presynaptic spike and the postsynaptic spike varies. Pre before post pairing induces Long Term Potentiation (LTP) whereas post before pre leads to Long Term Depression (LTD). Figure redrawn from (Bi and Poo 1998). D. Pairing frequency experiment. Here the lag between the pre and postsynaptic spike is constant (pre-post (black), post-pre(grey)) but the frequency between the pairing varies. Pre-post at low frequency does not lead to any weight change whereas increasing frequency allows more potentiation. Figure redrawn from (Sjöström, Turrigiano, and Nelson 2001). E. Weak tetanus stimulation. Few extracellular high frequency trains stimulated presynaptically yield LTP that lasts 2 to 3 hours. F. Strong tetanus stimulation. However, if more spikes are induced, the potentiation is stable for longer than 10 hours. Figure redrawn from (Sajikumar and Frey 2004a).

5

the postsynaptic trace (see section on "Standard Pair-Based Models" and Fig. 1.3B). If a postsynaptic spike occurs, the synapse is potentiated proportionally to the presynaptic trace (see section on "Standard Pair-Based Models" and Fig. 1.3A). We can assume the traces are such that all the spikes are considered, i.e. all-to-all interactions (Fig. 1.3E) or only the nearest neighbor spike (Fig. 1.3F). However the weights can not grow indefinitely, i.e. some kind of bounds should exist. In addition, a dependence of plasticity on the actual strength of the synapses was measured experimentally (Bi and Poo 1998; Turrigiano and Nelson 2004). Theoretically also, some STDP models explored different possibilities for weight dependencies (van Rossum, Bi, and Turrigiano 2000; Gütig, Aharonov, Rotter, and Sompolinsky 2003b) and homeostasis (Turrigiano and Nelson 2004). Great effort has also been put into the study of computational consequences of STDP, for example studying the implications for plastic networks (Roberts and Bell 2000; Mehta, Quirk, and Wilson 2000; Song, Miller, and Abbott 2000; Izhikevich 2004; Legenstein, Naeger, and Maass 2005; Guyonneau, VanRullen, and Thorpe 2005; Iglesiasa, Erikssonb, Grize, Tomassini, and Villa 2005; Morrison, Aertsen, and Diesmann 2007; Izhikevich and Edelman 2008; Kozloski and Cecchi 2008). Another step leads to exploration beyond spike pair interactions (Senn, Tsodyks, and Markram 2001; Froemke and Dan 2002; Pfister and Gerstner 2006; Gütig and Sompolinsky 2006). Experimental evidences show that intracellular stimulation of pre-post pairing at low frequency does not induce any weight change. Moreover, if the frequency between the pairings is increased, potentiation increases (Markram, Lübke, Frotscher, and Sakmann 1997; Sjöström, Turrigiano, and Nelson 2001). This is not consistent with the traditional view of STDP models where building blocks of plasticity are composed of pairs of pre-post and post-pre spikes. Therefore non-linear models were developed to describe those experiments where triplet interaction of spikes are considered (Senn, Tsodyks, and Markram 2001; Pfister and Gerstner 2006) (see section on "Triplet Model" and Fig. 1.3C,D) or discount factors on the "efficacy" of successive spikes (similar to including a model of short-term plasticity). This last model is based on extracellular inductions of triplets of spikes (Froemke and Dan 2002).

Standard Pair-Based Models

For the LTD part, standard pair-based models assume that presynaptic spike arrival at synapse i induces depression of the synaptic weight w_i by an amount that is proportional to $\bar{y}$, an exponential low-pass filtered version of the postsynaptic spike train $Y(t)$ with a time constant τ_- (see Fig 1.3B, trace post):

$$\tau_- \frac{d}{dt}\bar{y}(t) = -\bar{y}(t) + Y(t).$$

where $Y(t)$ is expressed as the series of short pulses at time t^n with n an index that counts the spike, $Y(t) = \sum_n \delta(t - t^n)$. The variable $\bar{y}$ is an abstract variable
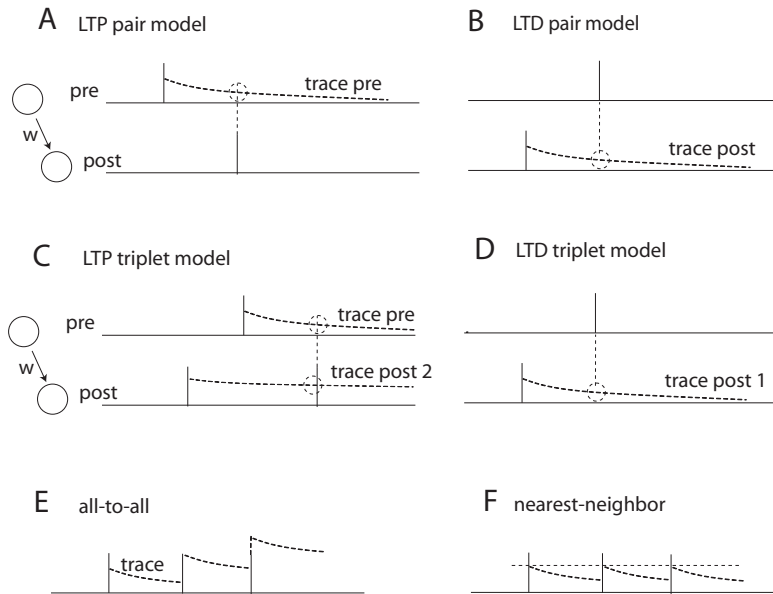
Figure 1.3: Different types of models. A. LTP with standard STDP model. The synaptic weight is potentiated at the time of the postsynaptic spike by an amount corresponding to the presynaptic trace. B. LTD with standard STDP model. The synaptic weight is depressed at the time of the presynaptic spike by an amount corresponding to the postsynaptic trace. C. LTP with triplet rule from (Pfister and Gerstner 2006) minimal model. The synaptic weight is potentiated at the time of the postsynaptic spike by an amount corresponding to the product of a presynaptic and a postsynaptic trace. D. LTD with triplet rule. Same than standard STDP model. E. All-to-all interaction of spikes. The trace jumps from a fix amount when a spike occurs and decays otherwise, leading to cumulative effect from all the previous spikes. F. Nearest-neighbor interaction of spikes. The trace jumps to a fix value when a spike occurs and decays otherwise. Only the previous spike affects the trace.

which could, for instance, reflect the level of calcium concentration (Lisman 1989) or the release of endocannabinoids (Sjöström, Turrigiano, and Nelson 2004), the back-propagating action potential, though such an interpretation is not necessary for this type of phenomenological rules. Similarly, the presynaptic spike train is described as a series of short pulses at time t_i^n with i is the index of the synapse, $X_i(t) = \sum_n \delta(t - t_i^n)$. The depression is then modeled as the following update rule (see Fig 1.3B):

$$\frac{d}{dt} w_i^- = -A_{\text{LTD}}(w_i^-)\, X_i(t)\, \bar{y}, \qquad (1.1)$$

where A_{LTD} is the amplitude for depression.

For the LTP part, the temporal evolution of the presynaptic low pass filter $\bar{x}_i(t)$ is described by (see Fig 1.3A, trace pre):

$$\tau_+ \frac{d}{dt} \bar{x}_i(t) = -\bar{x}_i(t) + X_i(t),$$

where X_i is the spike train defined above. The quantity $\bar{x}_i(t)$ could for example represent the amount of glutamate bound to postsynaptic receptors (Karmarkar and Buonomano 2002; Pfister and Gerstner 2006) or the number of NMDA receptors in an activated state (Senn, Tsodyks, and Markram 2001). The potentiation is then described by the following equation (see Fig 1.3A):

$$\frac{d}{dt} w_i^+ = +A_{\text{LTP}}(w_i^+)\, Y(t)\, \bar{x}_i(t). \qquad (1.2)$$

where A_{LTP} is the amplitude for potentiation.

These types of model can reproduce the spike-timing-dependent learning window (see 1.2C) but not the pairing frequency dependence (see 1.2D) nor the voltage clamp experiment (see 1.2A).

Triplet Model

The minimal triplet model (Pfister and Gerstner 2006) describes the depression the same way as the standard pair-based models. However, the potentiation takes into account triplet interactions of spike, 2 postsynaptic spikes and one presynaptic spike (see Fig 1.3C). The model defines a second type of postsynaptic trace $\bar{y}_2$ that decays with a time constant τ_2 that is typically in the order of 100ms. It can for example represent calcium concentration in the cell. The synapse is potentiated at a time of a postsynaptic spike from an amount that is proportional of the presynaptic spike trace $\bar{x}_i(t)$ (see Fig 1.3C, trace pre) and proportional also to this second postsynaptic spike trace $\bar{y}_2$ (see Fig 1.3C, trace post 2). The potentiation is written:

8

$$\frac{d}{dt} w_i^+ = +A_{\mathrm{LTP}}(w_i^+)\, Y(t)\, \bar{x}_i(t)\, \bar{y}_2.$$ (1.3)

where A_{LTP} is the amplitude for potentiation.

This model is able to reproduce the frequency experiment (see 1.2D) but not the voltage clamp experiment (see 1.2A) since it depends only on the time of the spike and not on the postsynaptic membrane potential.

1.2.3 BIOPHYSICAL MODELS

There have been a few attempts to describe the plasticity with its biophysical quantities such as (i) the voltage (Abarbanel, Huerta, and Rabinovich 2002), (ii) the Calcium/Calmodulin-Dependent Protein Kinase II (CaMKII) phosphorylation and bistability (Lisman 1985; Lisman 1989; Zhabotinsky 2000; Okamoto and Ichikawa 2000; Miller, Zhabotinsky, Lisman, and Wang 2005; Graupner and Brunel 2007), (iii) the calcium concentration (Karmarkar, Najarian, and Buonomano 2002; Karmarkar and Buonomano 2002; Shouval, Bear, and Cooper 2002; Abarbanel, Gibb, Huerta, and Rabinovich 2003; Rubin, Gerkin, Bi, and Chow 2005; Cai, Gavornik, Cooper, Yeung, and Shouval 2007) (see section "Calcium Model"), glutamate binding, AMPA receptors (Saudargiene, Porr, and Wörgötter 2003), NMDA receptors (Senn, Tsodyks, and Markram 2001) (see section "STM model") etc. For a detailed description of the biophysical models, please read Chapter 2.7.2-4 of (Graupner 2008).

STM Model

The STM model (Senn, Tsodyks, and Markram 2001) takes into account the dynamics of the NMDA receptor. Those receptors can be in 3 different states: rest, up or down. In absence of spikes, NMDA receptors are in the rest state, but they can be up-regulated when a presynaptic spike occurs or down-regulated with a postsynaptic spike. A notion of two types of second messengers is introduced in the model so that when a postsynaptic spike occurs, second messengers type 1 can be up-regulated only if the NMDA receptors are in the up states already. Inversely, the second messengers type 2 can be down-regulated if there is a presynaptic spike and if the NMDA receptors are already in the down state. Finally LTP appears when there is a postsynaptic spike and the second messengers type 1 are in the up state, LTD occurs at the time of a presynaptic spike if the second messengers type 2 are down regulated (see Fig 1.4A). This model takes into account pair interaction of spikes and also triplet interactions of spikes, i.e., 1 presynaptic spike and 2 postsynaptic ones for potentiation and 2 pre- and 1 postsynaptic spike for depression. It reproduces frequency dependence experiment (see experiment 1.2D, model 1.4C) as well as STDP experiment (see experiment 1.2 C, model 1.4B) but not the voltage clamp

A

second messenger 1 up $\xrightarrow{\text{post}}$ LTP

NMDA up $\longrightarrow$ post

second messenger 1 rest

pre

NMDA rest

post

second messenger 2rest

NMDA down $\longrightarrow$ pre

second messenger 2 down $\xrightarrow{\text{pre}}$ LTD

B

P^{∞}_{dis} (% of init. val.)

102
100
98
96

−50   0   50

$t^{sp}_{pre} - t^{sp}_{post}$ (ms)

C

P^{∞}_{dis} (% of init.val.)

150
140
130
120
110
100
90

2 5 10   20   30   40

AP−EPSP frequency (Hz)
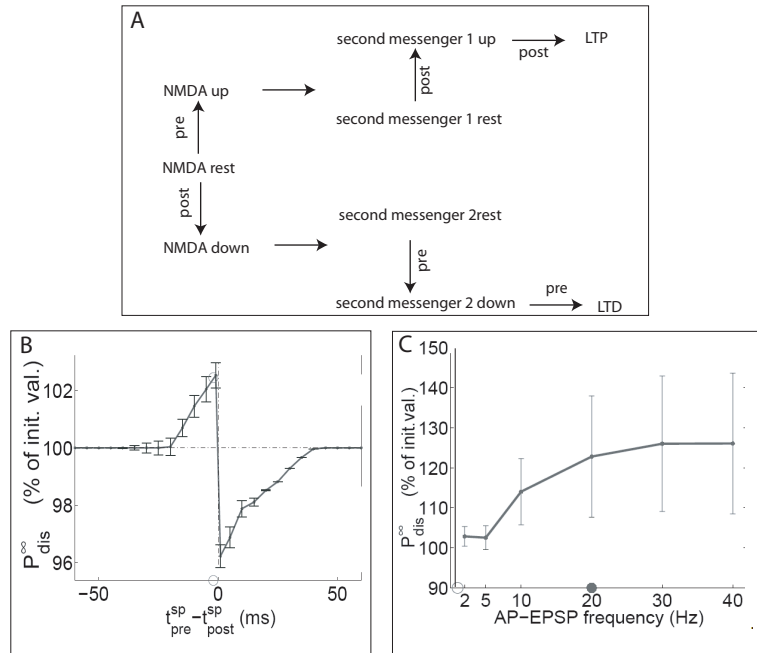
Figure 1.4: A. STM model. Simulation of the model reproducing B. STDP experiment, C. pairing frequency experiment. Figure redrawn from (Senn, Tsodyks, and Markram 2001).

experiment (see experiment 1.2 A). It offers a parallel to the BCM rule although the depression term in not linear in the presynaptic term.

Calcium Model

The calcium model by Shouval et al. (Shouval, Bear, and Cooper 2002) describes concentration of calcium in the postsynaptic cell as a measure for plasticity. Indeed, low calcium of concentration is not affecting the synapse but intermediate concentration leads to LTD whereas high concentration leads to LTP (see 1.5A). In order to compute the calcium concentration in the cell, the calcium current flows through the NMDA receptors only if the presynaptic spike is paired with a back propagating action potential.

It reproduces the voltage clamp experiment (see experiment 1.2A, model 1.5B), as well as the presynaptic stimulation frequency (see experiment 1.2B, model 1.5C) and the STDP experiment (see experiment 1.2C, model 1.5D). However, it predicts a LTD part in the pre-post side of the STDP curve due to the shape of the back propagating action potential.

Figure 1.5: Calcium model from Shouval et al. A. Weight change as a function of calcium concentration. Simulation of the model reproducing B. the voltage clamp experiment, C. presynaptic frequency stimulation experiment, D. STDP experiment. Figure redrawn from (Shouval, Bear, and Cooper 2002).

1.2.4 Consolidation Models

As mentioned above, synaptic tagging experiments (Frey and Morris 1997) revealed two phases of long-term synaptic plasticity: the early phase, which induces a change that lasts 2 to 3 hours (Fig. 1.2E) and the late phase, wh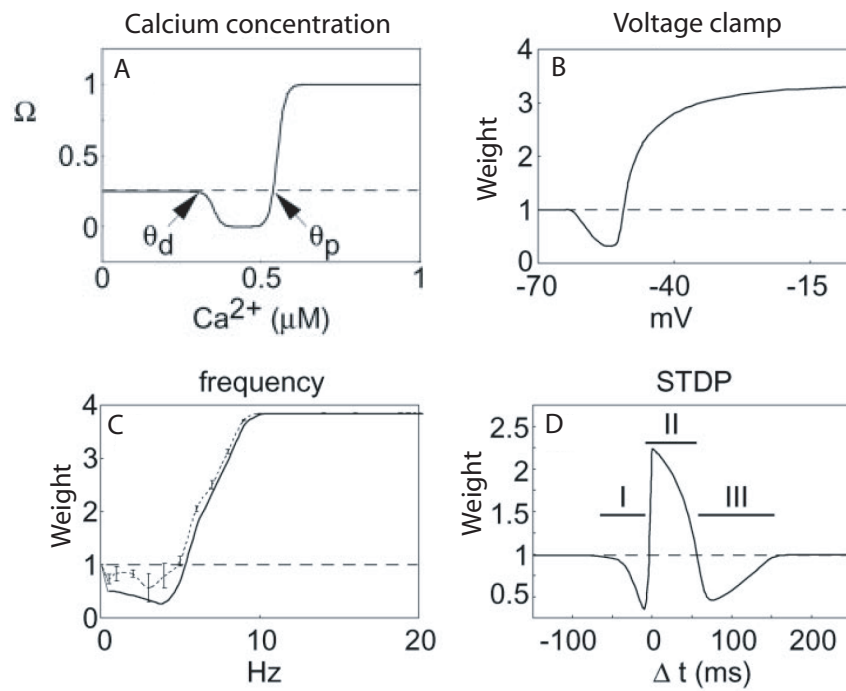ich lasts more than 10 hours (Fig. 1.2F). However, standard STDP models as well as the more detailed biophysical models typically describe only the early phase of long term plasticity and assume the changes to be long lasting. As an aside note we will mention the cascade model (Fusi, Drew, and Abbott 2005) (see section "Cascade Model"), which has different degrees of plasticity associated with different states, one of which could be interpreted as the maintenance phase described by the tagging experiments. Moreover, in the model of (Graupner and Brunel 2007) describing the CaMKII bistability, it is not clear if this bistablity reflects the maintenance or the CaMKII is only part of the early phase of plasticity.

Cascade Model

The cascade model (Fusi, Drew, and Abbott 2005) is design to optimize the memory capacity in a network. It proposes a bistable synapse that can take a weak value or a strong value. However the synapse can be in different plastic states for each value, called metastates. For example, if the synapse is already weak and undergoes a LTD protocol, it will keep the same weak efficacy but will go to a lower metastate where the synaptic weight is harder to change, i.e less plastic (see Fig 1.6).

1.2.5 Optimal Models

A completely different approach in developing learning models is the so-called top-down or optimality approach. In this framework, models are designed to perform a given task (e.g. Independent Component Analysis), maximize some quantities such as reward (Florian 2007), precision of spike timing (Pfister, Toyoizumi, Barber, and Gerstner 2006), transmission of information (Bell and Sejnowski 1995; Toyoizumi, Pfister, Aihara, and Gerstner 2005), sparseness (Olshausen and Field 1996), slowness (Wiskott and Sejnowski 2002; Sprekeler, Michaelis, and Wiskott 2007) etc. They can have the constraints of being online and local in order to be biologically plausible. They thus ideally take the form of a typical Hebbian learning rule or an STDP model. The obvious goal is that the top down and the bottom up models (inspired directly from the experiments) are consistent with each other.
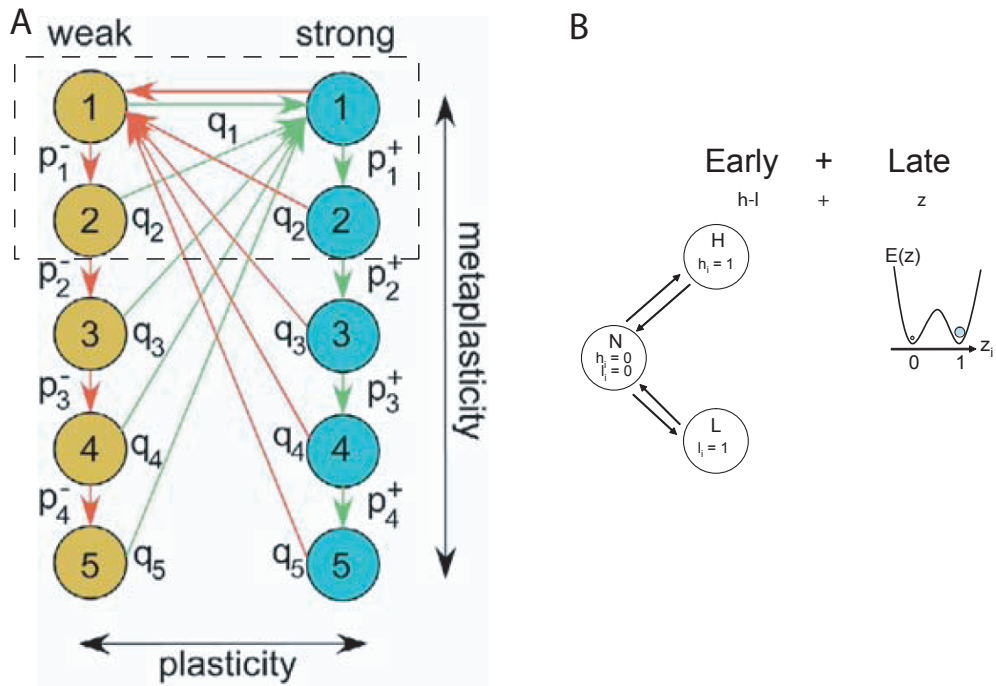
Figure 1.6: A cascade model. Bistable synapse (weak or strong) in different metastates that reflect different levels of plasticity (increasing numbers means less plastic). Dashed box corresponds to a cascade model with two levels. Figure redrawn from (Fusi, Drew, and Abbott 2005). B. For a comparison we show the TagTriC model. The total weight is the addition of the early weight which is a 3-states value and the late weight, a bistable value. For details of the model see Chapter 4 and (Clopath, Ziegler, Vasilaki, Buesing, and Gerstner 2008).

A different type of models has the feature of selecting which "synaptic experiences" should be remembered and which not. It is the reward modulated learning or reinforcement learning (Sutton and Barto 1998). These models have been derived from a top-down reward maximization approach. Learning typically occurs if a coincidence of pre- and postsynaptic activity is paired with a reward signal. Later on Schultz et al. (Schultz, Dayan, and Montague 1997; Schultz and Dickinson 2000) suggested a candidate for encoding the reward prediction error, the neuromodulator dopamine. Moreover, dopamine was shown in some synapses to be necessary for STDP (Pawlak and Kerr 2008). Rate-based models were shown to be functionally useful for learning (Williams 1992; Foster, Morris, and Dayan 2000; Sheynikhovich, Chavarriaga, Strosslin, Arleo, and Gerstner 2009). Recently, spiking versions, i.e. reward modulated STDP became fashionable, derived from an optimal framework (Seung 2003; Xie and Seung 2004; Pfister, Toyoizumi, Barber, and Gerstner 2006; Florian 2007) or from a phenomenological approach (Izhikevich 2007).

1.3   AIM OF THIS THESIS

The goal of the thesis is to better understand learning and memory. We used a theoretical approach. Aiming at a synthetic view of synaptic plasticity, we constructed mathematical models validated by experimental data. The models developed are so-called minimal ones, i.e. they are complex enough to describe the phenomena studied but not more complex. The biophysical details are not taken into account for the following reason: (i) to keep flexibility across systems, (ii) to avoid overfitting and (iii) to have intuitive, understandable and if possible analytically tractable models. However, we do associate abstract variables to possible biophysical candidates, without being bound to it. This thesis presents a compact model of synaptic plasticity across different time scales, faithfully reproducing the different types of experimental data described above. The early and the late phase of long term plasticity is taken into account. Short term plasticity effects are neglected.

The model for the early phase of plasticity (Clopath, Vasilaki, Buesing, and Gerstner xxxx) depends on the voltage of the postsynaptic neuron to reproduce voltage clamp experiments (Artola, Bröcher, and Singer 1990; Ngezahayo, Schachner, and Artola 2000). Indeed no postsynaptic spike is required for synaptic potentiation whereas the time course of the voltage seems to be relevant. Thus a proper description of the voltage is essential and thus we propose an adequate neuron model that fits the biological data properly (Clopath, Jolivet, Rauch, Luescher, and Gerstner 2007). Note that by construction standard

STDP models (Gerstner, Kempter, van Hemmen, and Wagner 1996; Kempter, Gerstner, and van Hemmen 1999; Senn, Tsodyks, and Markram 2001; Song, Miller, and Abbott 2000; van Rossum, Bi, and Turrigiano 2000; Gütig, Aharonov, Rotter, and Sompolinsky 2003a; Karmarkar and Buonomano 2002) cannot reproduce the voltage dependence os plasticity experiment because they consider only spike times. In addition our model incorporates triplet interactions of spikes (Senn, Tsodyks, and Markram 2001; Pfister and Gerstner 2006) in order to reproduce the frequency dependence (Markram, Lübke, Frotscher, and Sakmann 1997; Sjöström, Turrigiano, and Nelson 2001) (as described above). Finally, functional implications of this model are studied such as network connectivity under different coding schemes and Independent Component Analysis.

Synaptic plasticity is such a fascinating topic to study because it is considered as the basis of learning and memory. Long lasting changes are thus relevant to study. However standard STDP models (Gerstner, Kempter, van Hemmen, and Wagner 1996; Kempter, Gerstner, and van Hemmen 1999; Senn, Tsodyks, and Markram 2001; Song, Miller, and Abbott 2000; van Rossum, Bi, and Turrigiano 2000; Gütig, Aharonov, Rotter, and Sompolinsky 2003a; Karmarkar and Buonomano 2002) and even biophysical models looking at bistability of CAMKII (Lisman 1989; Miller, Zhabotinsky, Lisman, and Wang 2005; Graupner and Brunel 2007) do not tackle the problem of late phase plasticity and maintenance. This thesis presents a minimal rule describing the long lasting changes measured by synaptic tagging experiments. Another type of model considering maintenance is the cascade model (Fusi, Drew, and Abbott 2005). We therefore expect to find some structural similarities between ours and the cascade model. In addition, reinforcement learning models, or reward modulated learning have also the property of selecting what should be learned. We thus expect our model of synaptic tagging to depend on neuromodulation. The model presented in this thesis (Clopath, Ziegler, Vasilaki, Buesing, and Gerstner 2008) contains a triggering of plasticity related proteins and a bistable maintenance variable. It describes faithfully the synaptic tagging experiments and captures cross tagging as well.

As introduced in the "theoretical background" section, there are two main approaches for modeling: (i) Either the model is derived from the biology, i.e. the phenomenological model or the biophysical model, which might be exhibiting some functions. (ii) Or the model is derived from a given function, i.e. the optimal or top-down model and then structural similarities to biology can be drawn. Ideally, the two methods should give rise to similar models. In the last part of the thesis, we derive a top-down model performing one of the functions solved by the induction model, that is Independent Component Analysis. This last model is a rate-based learning rule (Clopath, Longtin, and Gerstner 2008) and the link to the model for the early phase of plasticity is discussed in the

"future work" section.

1.4 ROAD-MAP OF THIS DISSERTATION

In this dissertation, the following four chapters contain the different papers published during the thesis. Each chapter starts with an introduction and a link to the general aim of the thesis. The last chapter summarizes the results and offers some opening for future work.

Chapter 2: SINGLE CELL MODELING - THE ADAPTIVE EXPONENTIAL INTEGRATE AND FIRE MODEL

This chapter presents a neuron model, the Adaptive Exponential Integrate and Fire model, and compares it to voltage traces of layer V pyramidal cells under random current injections.
Paper published in Neurocomputing 2007

Chapter 3: MODEL OF THE INDUCTION OF LONG-TERM SYNAPTIC PLASTICITY AND ITS FUNCTIONAL IMPLICATIONS

This chapter proposes a model for the early phase of plasticity that depends nonlinearly on the postsynaptic voltage. It reproduces various plasticity experiments and is used to explore different computational roles.
Paper under review in Nature Neuroscience 2009

Chapter 4: A MODEL FOR THE LATE PHASE OF LONG TERM SYNAPTIC PLASTICITY

This chapter presents a model for synaptic tagging experiments that includes the model for the early phase of plasticity presented above (Chapter 3) and describes the maintenance phase. It faithfully reproduces synaptic tagging and cross tagging experiments.
Paper published in PLoS Computational Biology 2008

Chapter 5: A TOP DOWN APPROACH TO A RULE PERFORMING INDEPENDENT COMPONENT ANALYSIS

This chapter shows the derivation of a top down model performing Independent Component Analysis (ICA). It is a Hebbian rate-based rule and it is tested against standard ICA benchmarks. Noteworthy this model shares common functional properties with the induction model presented in Chapter 3.
Paper published in NIPS 2008

Chapter 6: CONCLUSIONS AND FUTURE WORK

Single cell modeling - the Adaptive Exponential Integrate and Fire Model

To study synaptic plasticity between neurons, we need an accurate model describing their activity. Indeed the synaptic plasticity models presented in this thesis not only depend on the exact spike timing but also on the whole voltage time course. Therefore, we present in the following paper (Clopath, Jolivet, Rauch, Luescher, and Gerstner 2007) how well the Adaptive Exponential Integrate and Fire (AdEx) model reproduces voltage traces of layer V pyramidal cells under random current injection.

The AdEx model (Brette and Gerstner 2005) is a simple Integrate and Fire type model augmented by an exponential term (Fourcaud-Trocme, Hansel, van Vreeswijk, and Brunel 2003) which describes the activation of a rapid sodium current. It allows an accurate prediction of the spike times. Actually, the exact shape of the neuron nonlinearity was measured from experimental data where an exponential was found to be the best fit (Badel, Lefort, Brette, Petersen, Gerstner, and Richardson 2008). Moreover, the model has an additional variable which describes the spike-triggered adaptation and the subthreshold adaptation. It was shown to fit accurately a more detailed Hodgkin-and-Huxley model (Hodgkin and Huxley 1952; McCormick, Wang, and Huguenard 1993), i.e. prediction of 96% of spike times (Brette and Gerstner 2005). We were wondering how well this model is in reproducing voltage traces of real neurons. It turns out that the model reproduces up to 96% (average 60%) of the spikes and matches accurately the subthreshold voltage. Later on, we showed that this model can account for different spiking regimes (Naud, Marcille, Clopath, and Gerstner 2008), like irregular spiking, bursting etc.

This model has 2 variables and 7 parameters, it therefore is a good trade-off between complexity and accuracy. Overfitting is avoided and the computational time is reasonable when used to test functional implications in long plasticity experiments.

ELSEVIER

Predicting neuronal activity with simple models of the threshold type: Adaptive Exponential Integrate-and-Fire model with two compartments

Claudia Clopath[a,*], Renaud Jolivet[a], Alexander Rauch[b,c],
Hans-Rudolf Lüscher[c], Wulfram Gerstner[a]

[a]*School of Computer and Communication Sciences and Brain Mind Institute, EPFL, CH-1015 Lausanne, Switzerland*
[b]*Max-Plank-Institute for Biological Cybernetics, D-72012 Tübingen, Germany*
[c]*Institute of Physiology, University of Bern, CH-3012 Bern, Switzerland*

Abstract

An adaptive Exponential Integrate-and-Fire (aEIF) model was used to predict the activity of layer-V-pyramidal neurons of rat neocortex under random current injection. A new protocol has been developed to extract the parameters of the aEIF model using an optimal filtering technique combined with a black-box numerical optimization. We found that the aEIF model is able to accurately predict both subthreshold fluctuations and the exact timing of spikes, reasonably close to the limits imposed by the intrinsic reliability of pyramidal neurons.
© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Electrophysiological data can be described by detailed conductance-based models (Hodgkin–Huxley-type models [6]). However, those models are rather complex, which implies that they are difficult to analyze and costly to implement numerically. Moreover, it is unclear how many details of conductance-based models are really necessary for the reproduction of experimental spike patterns [4,13]. For those reasons, simple phenomenological spiking neurons such as Integrate-and-Fire models are highly popular.

The adaptive Exponential Integrate-and-Fire (aEIF) model used in this paper generalizes the standard leaky Integrate-and-Fire model in several directions: the strict threshold is replaced by a more realistic smooth threshold zone as in the Exponential Integrate-and-Fire neuron [2]. Furthermore, addition of a second variable captures subthreshold resonance or adaptation [7,15]. The aEIF

model showed convincing performances when compared to more detailed models [1], but so far, has never been tested on recordings of real neuron.

In this report, we will test the performances of the aEIF model on layer-V neocortical pyramidal neurons under random current injection.

2. Model

The aEIF is defined by [1]

$$C\frac{\mathrm{d}u(t)}{\mathrm{d}t} = -g_{\mathrm{L}}(u(t) - E_{\mathrm{L}}) + g_{\mathrm{L}}\Delta_{\mathrm{T}}\exp\left(\frac{u(t) - V_{\mathrm{T}}}{\Delta_{\mathrm{T}}}\right) - w + I, \tag{1}$$

$$\tau_w \frac{\mathrm{d}w(t)}{\mathrm{d}t} = a(u(t) - E_{\mathrm{L}}) - w(t), \tag{2}$$

where C is the membrane capacitance, g_{L} the leak conductance, E_{L} the resting potential, Δ_{T} the slope factor and V_{T} the threshold potential (Fig. 1). Note that formally, E_{L} is not exactly the resting potential because of the

---

*Corresponding author.
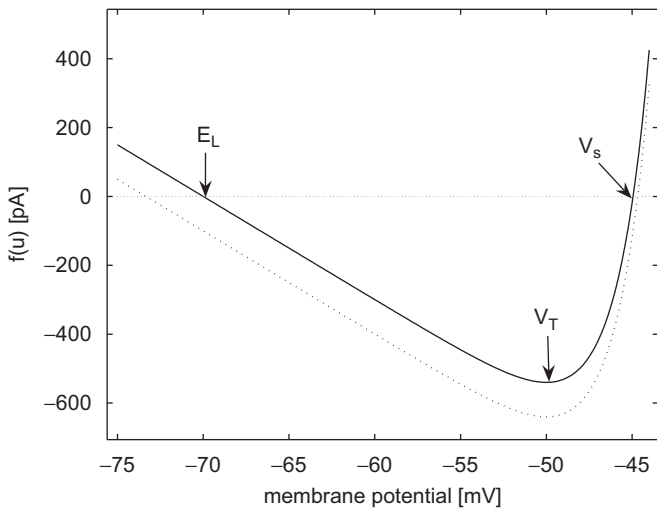
E-mail address: claudia.clopath@epfl.ch (C. Clopath).

Fig. 1. Spike function $f(u) = C\,\mathrm{d}u/\mathrm{d}t - I$ of the aEIF model in the non-adapted state ($w = 0$; black line). The left intersection of the spike function with the horizontal axis is the resting potential E_{L}, the right intersection gives the potential V_{s} above which the spike is generated. The minimum of $f(u)$, V_{T}, gives the maximum subthreshold potential that can be reached by constant current injection. In the adapted state ($w > 0$), the spike function $f(u) = C\,\mathrm{d}u/\mathrm{d}t - I$ is shifted vertically downward (dotted line).

exponential term. The variable w describes the level of adaptation of the neuron and a represents the relevance of subthreshold adaptation. The exponential term describes the early activation of voltage-gated sodium channels.

Formally the model is said to generate a spike if the potential u grows rapidly to infinity. In practice, a spike event is recorded when the voltage reaches a threshold $V_{peak} = 20\,\mathrm{mV}$. The exact value is not critical because V_{peak} only shifts spike times by a fraction of millisecond. After the spike has been triggered, u is reset to the resting potential E_{L} and the variable w is increased by an amount b, which accounts for spike-triggered adaptation.

The original aEIF model is a point neuron model i.e. without spatial structure. However, in this study, we decided to take into account the coupling of the soma with the dendrites. Therefore, we used a two-compartment model (one somatic compartment coupled to a passive dendritic compartment) defined by

$$C\frac{\mathrm{d}u_s}{\mathrm{d}t} = -g_{\mathrm{L}}(u_{\mathrm{s}} - E_{\mathrm{L}}) - \frac{g_{\mathrm{c}}}{p}(u_{\mathrm{s}} - u_{\mathrm{d}})$$
$$+ g_{\mathrm{L}}\varDelta_{\mathrm{T}}\exp\left(\frac{u_{\mathrm{s}} - V_{\mathrm{T}}}{\varDelta_{\mathrm{T}}}\right) - w + I, \qquad (3)$$

$$C\frac{\mathrm{d}u_{\mathrm{d}}}{\mathrm{d}t} = -g_{\mathrm{L}}(u_{\mathrm{d}} - E_{\mathrm{L}}) - \frac{g_{\mathrm{c}}}{1 - p}(u_{\mathrm{d}} - u_{\mathrm{s}}), \qquad (4)$$

$$\tau_w\frac{\mathrm{d}w}{\mathrm{d}t} = a(u_{\mathrm{s}} - E_{\mathrm{L}}) - w, \qquad (5)$$

where u_{s} is the membrane voltage in the somatic compartment, u_{d} the membrane voltage in the dendritic compartment, g_{c} the coupling conductance and $p = $ somatic area/

total area. The two-compartment model is motivated by experimental results. Indeed, the linear response kernel is best fitted by a double exponential (see below point 3(i)), suggesting a coupling between soma and a passive dendrite acting as current sink [8].

3. Parameter fitting

Recordings of layer-V pyramidal neurons of rat neocortex were used to determine parameters of the model. The neurons were recorded intracellularly in vitro while stimulated at the soma by a randomly fluctuating current generated by an Ornstein–Uhlenbeck (OU) process (autocorrelation time 1 ms). Both mean and variance of the OU process were varied in order to sample the response of the neurons to various levels of tonic and time-dependent inputs. Details of the experimental procedure can be found in [14].

Our method to extract the parameters of the aEIF model is based on the following steps:

(i) Passive membrane properties (C, g_{L}, g_{c}, p, E_{L}): In subthreshold regime (where the exponential term can be neglected), Eqs. (3) and (4) can be integrated [3],

$$u(t) = \int_0^{+\infty} \kappa_\infty(s)\,I(t - s)\,\mathrm{d}s. \qquad (6)$$

In the non-adapted state $w = 0$, we find

$$\kappa_\infty(s) = \frac{1}{C}[p\mathrm{e}^{-s/\tau_{\mathrm{s}}} + (1 - p)\mathrm{e}^{-s/\tau_{\mathrm{c}}}], \qquad (7)$$

where $\tau_{\mathrm{s}} = C/g_{\mathrm{L}}$ is the somatic membrane time constant and $\tau_{\mathrm{c}} = [p(p-1)C]/[p(p-1)g_{\mathrm{L}} - g_{\mathrm{c}}]$ is the coupling time constant. The kernel κ_∞ is extracted by the Wiener–Hopf optimal filtering technique [8]. This step involves a comparison of the subthreshold fluctuations with the corresponding input current. This yields a "raw" filter κ_{exp} (Fig. 2). The filter κ_{exp} is well fitted by the double exponential κ_∞ derived from our two-compartment model. C, g_{L}, g_{c}, p were extracted from the double exponential fit κ_∞ (Eq. (7)) of κ_{exp}, E_{L} from the resting value at the beginning of the recording.

(ii) Slope factor: The slope factor determines the sharpness of the threshold. In the limit $\varDelta_{\mathrm{T}} \to 0$, the model becomes a standard leaky Integrate-and-Fire model. As the threshold has a region of fuzziness, we decided to fix the slope factor at $\varDelta_{\mathrm{T}} = 2\,\mathrm{mV}$ so as to restrict the number of parameters to be optimized. This value seems reasonable and is close to the value found for the Wang–Buszaki model [2].

(iii) Subthreshold adaptation: According to systems theory, it is not possible to extract the subthreshold adaptation a from our data set. Therefore, we set a to zero. Indeed, the Laplace transformed system has one pole and one zero that masque each other (i.e. the determinant of the identifiability matrix is
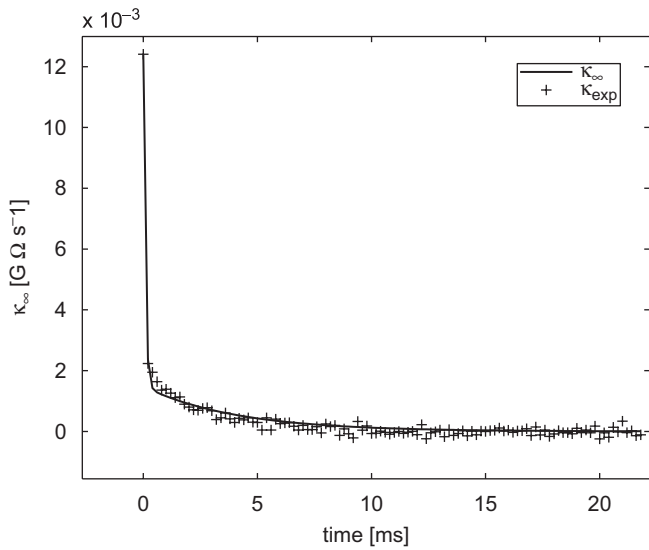
Fig. 2. Raw data of the kernel κ_{exp} extracted by optimal filtering technique (symbols) have been fitted by the double exponential κ_{∞} (solid line).

close to zero), preventing the system to be fully characterizable [10].

(iv) Voltage reset: After a spike has been triggered, the voltage is simply reset to the resting potential E_L.

(v) Optimization: Finally, the remaining parameters, V_T, τ_w, b were optimized using the simulated annealing technique optimizing the firing rate and maximizing the coincidence factor $\Gamma_{n\to m}$. We minimized the following expression:

$$2\left|\frac{v_{data} - v_{sim}}{v_{data}}\right| - \Gamma_{n\to m},$$

where v_{data} is the firing rate of the neuron data and v_{sim} is the firing rate of the simulated data. $\Gamma_{n\to m}$ is defined by [9]

$$\Gamma_{n\to m} = \frac{N_{coinc} - \langle N_{coinc}\rangle}{\frac{1}{2}(N_{data} + N_{aEIF})}\frac{1}{N}, \qquad (8)$$

where N_{data} is the number of spikes in the reference spike train (recordings of pyramidal cells), N_{aEIF} is the number of spikes in the predicted spike train (generated with the aEIF model with the same driving current). N_{coinc} is the number of coincident spikes with precision $\Delta = 2\,\text{ms}$ and $\langle N_{coinc}\rangle$ is the number of coincidences generated by a homogeneous Poisson process with the same rate v_{sim} as the spike train generated with the aEIF model. Finally, the normalization $N = 1-2v_{sim}\Delta$ allows $\Gamma_{n\to m}$ to reach 1 only if the spike train of the aEIF model reproduces exactly the spike train of the cell. $\Gamma_{n\to m}$ will be 0 if the similarity between the two spike trains is not better than between that two random spike trains generated by homogeneous Poisson processes at the same rate. In order to test the robustness of the method, we picked one cell and repeated the parameter optimization by simulated annealing 10

times. We found that the V_T is very robust within errors less than 3%. The parameters b and τ_w are strongly correlated. While individual variance is high their product $b\tau_w$ is stable.

4. Results

The data set consists of four different neurons. For each cell, a set of 10 different input currents with different means and variances are injected. Each input trace is repeated four times. Fig. 3 shows a direct comparison between predicted and recorded spike trains for a typical neuron. Both spike trains are almost indistinguishable (Fig. 3A; for clarity reasons, the predicted spike train has been shifted upward). Even when zooming in the subthreshold regime, differences are in the range of a few millivolts only (Fig. 3B). The spike dynamics is correctly predicted apart from a short period of time just after the spike is emitted (Fig. 3C). This is due to the reset value of the voltage which is set to the resting potential.

The model performances were evaluated using the coincidence factor $\Gamma_{n\to m}$ (Eq. (8)). Our model is facing natural limits of prediction because cortical pyramidal neurons respond with very different reliability depending on the type of stimulation they receive [11]. As we cannot expect our model to yield better predictions than the intrinsic reliability of the neuron, we consider the intrinsic reliability of the neuron as an upper bound. The intrinsic reliability can be easily measured since the same input has been injected four times in the same cell. The reliability of neurons does not vary significantly with the mean of the injected current. However, it strongly depends on the variance of the current [8,11]. In the case of low variance, the spike timing is not controlled by the stimulus anymore. Therefore, we abandon the data with low variance ($\sigma < 150\,\text{pA}$). Intrinsic reliability is characterized by the factor $\Gamma_{n\to n}$ in analogy to Eq. (8). The subscript $n \to n$ means that the neuron is compared to itself across two different trials with the same realization of the input. We remark that data with high variance input are more likely to resemble an in vivo situation than low variance input data. For data used below, the intrinsic reliability varies from $\Gamma_{n\to n} = 75\%$ to as low as $\Gamma_{n\to n} = 20\%$.

We found that the aEIF model predicts up to $\Gamma_{eff} = 96\%$ of the spikes that can be predicted ($\Gamma_{eff} = \Gamma_{n\to m}/\Gamma_{n\to n}$, $m \to n$ means model compared to neuron) and on average $\Gamma_{eff} = 60\%$ (Fig. 4).

Fig. 5 shows the experimental spike trains during four repetitions with the same driving current (bottom traces) as well as the simulated spike train (top trace) for different reliability and performance cases.

5. Discussion and conclusion

We tested the aEIF model on experimental electrophysiology recordings and found a prediction of the spike

A

Fig. 4. The performance of the aEIF ($\Gamma_{m \to n}$) is plotted versus the intrinsic reliability ($\Gamma_{n \to n}$) for each data set. The diagonal yields the upper bound of the model. The average prediction is shown by the dashed line ($\Gamma_{\text{eff}} = \Gamma_{n \to m}/\Gamma_{n \to n} = 0.6$). More details on the specific cases A, B, C are shown in Fig. 5.
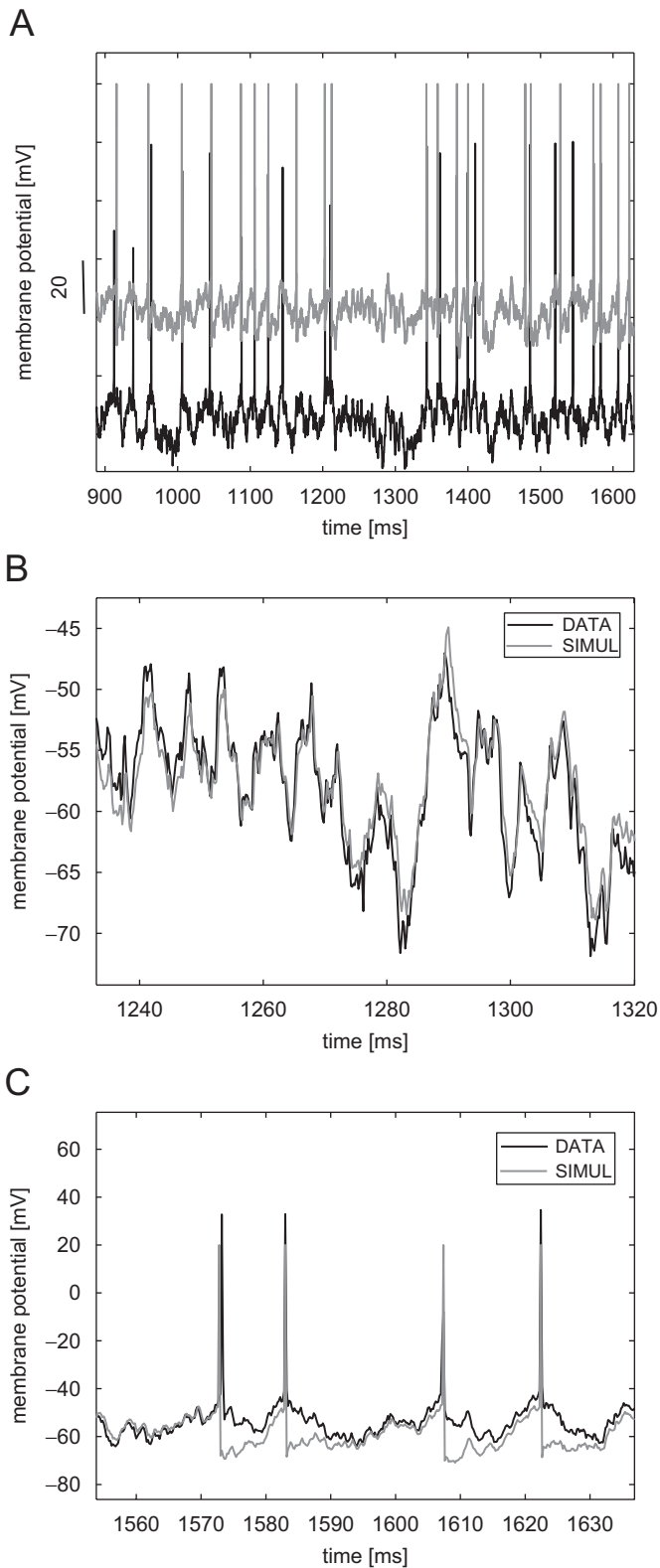
B

C

Fig. 3. Performances of the aEIF model. A. Predictions of the model (grey line) are compared to the spike train of the corresponding neuron (black line). For this graph only, the simulated trace has been shifted by 40 mV upward for reasons of visual clarity. B. Zoom on the subthreshold regime. C. Zoom on four correctly predicted spikes.
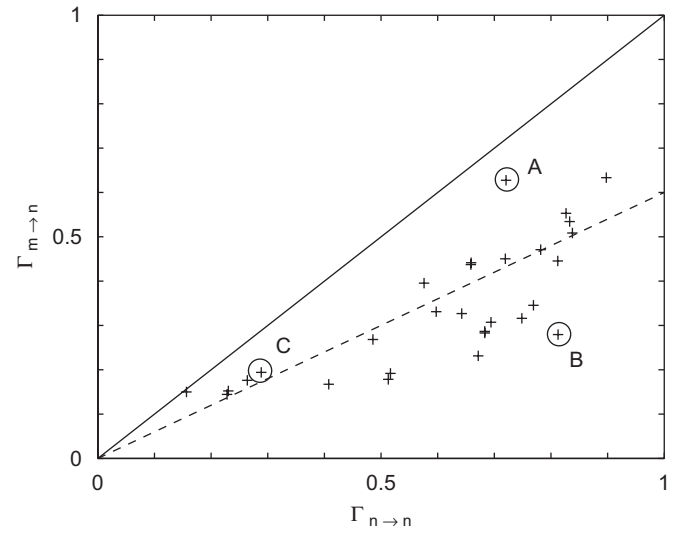
times Γ_{eff} up to 96% (average of 60%). With the same data set, a Spike Response Model with dynamic threshold has been evaluated and the performances were Γ_{eff} up to 75% (average 65%) [8].

We remark that the protocol used for the recordings is not the most suitable for characterizing our model: in our extraction method, we had to set the subthreshold adaptation a to 0. In addition, data generated purely by simulation of the aEIF model were characterized very badly with our method (average of $\Gamma = 85\%$). A completely different protocol to extract the parameters of the aEIF model has been proposed recently by Brette and Gestner [1]. This protocol contains a series of standard electrophysiological paradigms (injection of current pulses, slow current ramps and random conductance injections). It has been tested with data generated by a detailed model and yielded excellent performances ($\Gamma = 96\%$). In addition, this protocol allows to reduce noise (averaging over several recordings), so that the subthreshold adaption a could, in principal, be extracted from pyramidal cell recordings. The latter protocol is under study at the moment using a new data set recorded following the methodology proposed by Brette and Gerstner [1].

In the aEIF model, adaptation is a useful component since it allows the model to account for different driving regimes. We found as well that it is relatively easy to correctly predict the subthreshold dynamics even with a simple leaky integrator but it is difficult to find an efficient threshold criterion for spike initiation. This problem is solved by the aEIF model which includes an additional exponential term to describe early activation of voltage-gated sodium channels. This last addition allows to model

specific behaviors like delayed spike initiation and offers flexibility at the level of the threshold mechanism. It was recently suggested by Naundorf and colleagues that the rapid dynamics of action potential initiation in cortical neurons are outside the range of behaviors described by the classical Hodgkin–Huxley theory [12]. In the aEIF model, the exponential term allows a fast activation of the action potential. Thus, on a phenomenological level, the aEIF model could possibly account for rapid spike initiation in real neurons.

References

[1] R. Brette, W. Gerstner, Adaptive exponential integrate-and-fire model as an effective description of neuronal activity, J. Neurophysiol. 94 (2005) 3637–3642.

[2] N. Fourcaud-Trocmé, D. Hansel, C. van Vreeswijk, N. Brunel, How spike generation mechanisms determine the neuronal response to fluctuating inputs, J. Neurosci. 23 (2003) 11628–11640.

[3] W. Gerstner, W. Kistler, Spiking neuron Models: Single Neurons, Populations, Plasticity, Cambridge University Press, Cambridge, 2002.

[4] M.S. Goldman, J. Golowasch, E. Marder, L.F. Abbott, Global structure, robustness and modulation of neuronal models, J. Neurosci. 21 (2001) 5229–5238.

[6] A. Hodgkin, A. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J. Physiol. 117 (1952) 500–544.

[7] E. Izhikevich, Which model to use for cortical spiking neurons IEEE Trans. Neural Networks 15 (2004) 1063–1070.

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

[9] W. Kistler, W. Gerstner, J. van Hemmen, Reduction of Hodgkin–Huxley equations to a single-variable threshold model, Neural. Comput. 9 (1997) 1015–1045.

[10] L. Ljung, System Identification: Theory for the User, second ed., Prentice-Hall, Englewood Cliffs, NJ, 1999.

[11] Z. Mainen, T. Sejnowski, Reliability of spike timing in neocortical neurons, Science 268 (1995) 1503–1506.

[12] B. Naundorf, F. Wolf, M. Volgushev, Unique features of action potential initiation in cortical neurons, Nature 440 (2006) 1060–1063.

[13] A.A. Prinz, C.P. Billimoria, E. Marder, Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons, J. Neurophysiol. 90 (2003) 3998–4015.

[14] A. Rauch, G. La Camera, H.-R. Lüscher, W. Senn, S. Fusi, Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo-like input currents, J. Neurophysiol. 90 (2003) 1598–1612.

[15] M.J. Richardson, N. Brunel, V. Hakim, From subthreshold to firing-rate resonance, J. Neurophysiol. 89 (2003) 2538–2554.
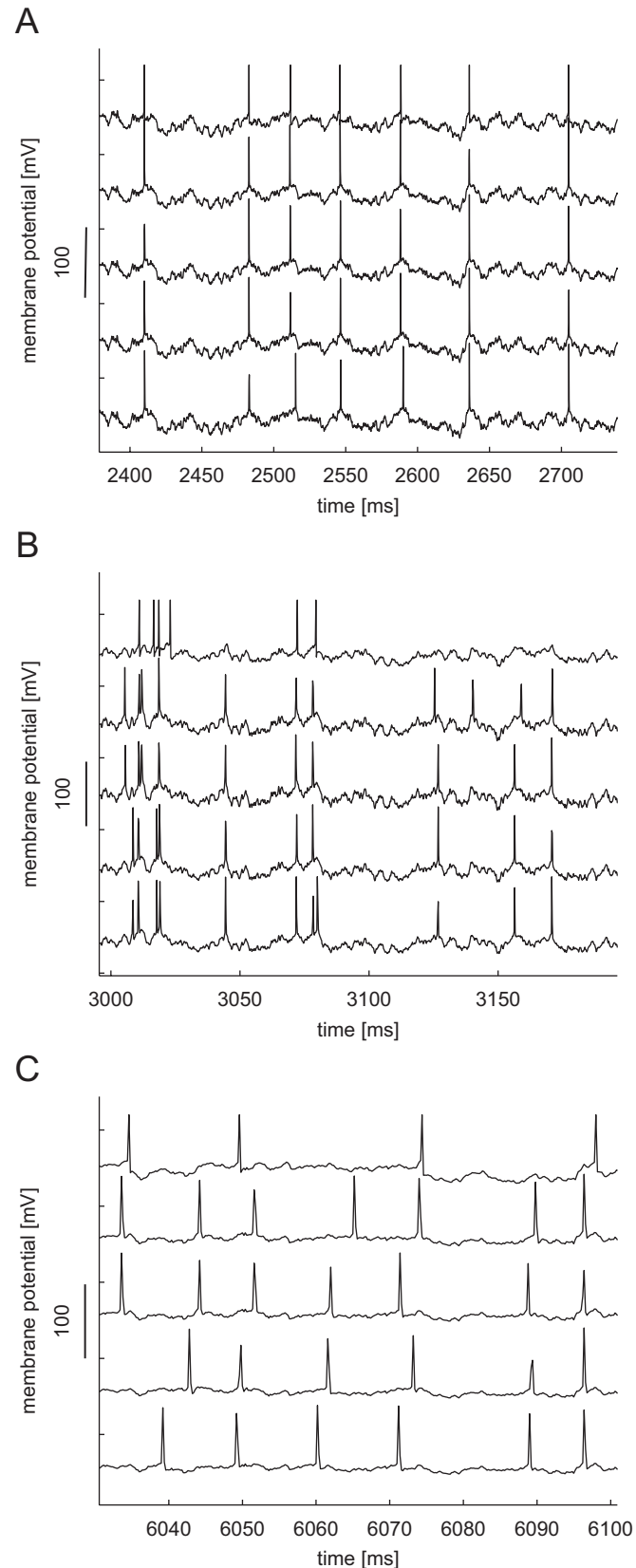
Fig. 5. The lower four spike trains of each graph are experimental voltage traces recorded in response to the same driving current. The top trace is the simulated spike train with this driving current. The traces shown in *A* have a high intrinsic reliability and high prediction, i.e. the model is good. *B* has high reliability but a bad prediction, i.e. the model is insufficient. In *C* the reliability is low and the prediction is low, but the model is good since $\Gamma_{eff} > 0.65$. See Fig. 4 for the Γ values of the different cases *A*, *B*, *C*.

Claudia Clopath has obtained her M.Sc. in physics from the EPFL, in April 2005. She is now a Ph.D. in the Laboratory of Computational Neuroscience at the EPFL and her current research interest is to describe neuronal activity with simple Integrate-and-Fire type models.

Renaud Jolivet obtained his Ph.D. in theoretical neuroscience at EPFL, in 2005, after studies in biophysics at the University of Lausanne (M.Sc.). He is now a postdoc in the laboratory of Prof. Pierre Magistretti at the University of Lausanne where his research is focused on neuro-astrocytic interactions in the context of brain metabolism and signal processing.

Alexander Rauch obtained his M.D., in 2003, at the University of Bern after studies in medicine at the University of Basel. His clinical experience includes surgery and neurorehabilitation. Between 1999 and 2003, he has worked as an electrophysiologist at the University of Bern focusing his research on links between cortical pyramidal neurons and simple models of neuronal activity. He is now a research assistant in the laboratory of Prof. Nikos Logothetis at the Max-Planck-Institute where his research is focused on BOLD signal and brain imaging techniques.

Hans-Rudolf Lüscher is Director of the Department of Physiology, University of Bern, Switzerland. He has a Medical Degree from the University of Zürich, Switzerland. He did his postdoctoral training with Elwood Henneman at the Department of Physiology and Biophysics, Harvard Medical School, Boston, USA. He spent extended time periods at the Division of Neuroscience, John Curtin School of Medical Research, Canberra, Australia. Hans-Rudolf Lüscher served as Dean and Vice-Dean of the Medical Faculty, University of Bern. He is a member of the Science Council of the Swiss National Science Foundation and of the Swiss Academy of Medical Sciences.

Wulfram Gerstner received his Ph.D. degree in theoretical physics from the TU Munich, Germany, in 1993, after studies in Tübingen, Berkeley, and Munich. He is a Professor and Head of Laboratory of Computational Neuroscience, EPFL, Switzerland.

MODEL OF THE INDUCTION OF LONG-TERM SYNAPTIC PLASTICITY AND ITS FUNCTIONAL IMPLICATIONS

S TANDARD models for Spike-Timing-Dependent Plasticity (STDP) take into account pair interactions of presynaptic and postsynaptic spike times (Gerstner, Kempter, van Hemmen, and Wagner 1996; Kempter, Gerstner, and van Hemmen 1999; Senn, Tsodyks, and Markram 2001; Song, Miller, and Abbott 2000; van Rossum, Bi, and Turrigiano 2000; Gütig, Aharonov, Rotter, and Sompolinsky 2003a; Karmarkar and Buonomano 2002). However, a number of pairing experiments (Wang, Gerkin, Nauen, and Bi 2005; Sjöström, Turrigiano, and Nelson 2001) show that pair interactions are not sufficient to fully describe Long-Term Potentation and Depression (LTP/LTD) (Pfister and Gerstner 2006). Moreover, by construction, simple spike-based models fail in voltage clamp experiments, where LTP or LTD can be induced by coincidence of presynaptic spike arrival and depolarization of the postsynaptic membrane (Artola, Bröcher, and Singer 1990; Ngezahayo, Schachner, and Artola 2000). The following paper presents a triplet model which takes into account the presynaptic spike times and the postsynaptic membrane potential, filtered with three different time constants. For spike induced experiments, this model can formally be reduced to the triplet rule proposed by Pfister et al. (Pfister and Gerstner 2006), and yields similar results to, for example, frequency dependent pairing experiments by Sjöström et al. (Sjöström, Turrigiano, and Nelson 2001). Moreover, it also reproduces the behavior of the ABS rule (Artola, Bröcher, and Singer 1990), i.e. no synaptic changes are observed under presynaptic stimulation when the postsynaptic potential is hyperpolarized; while small depolarization leads to LTD and strong depolarization to LTP (Fig 1.2A). Additionally, this model can describe the hybrid experiment by Sjöström et al. where low-frequency potentiation is rescued by depolarization (Sjöström, Turrigiano, and Nelson 2001). It offers testable predictions as to how other protocols may change the weights. This model therefore closes the debate whether STDP is more fundamental than voltage dependence (Lisman and Spruston 2005) since it shows that most if not all existing experimental data on STDP can be derived from a model with voltage dependence.

This paper also explores the functional consequences of this model. Due to its similarity to the well-known rate-based Bienenstock-Cooper-Munro model (Bienenstock, Cooper, and Munro 1982), the model exhibits selectivity in the

inputs, which can be seen in receptive field development scenarios. In addition, the model performs ICA-like computation. For example when presented with natural scenes the weights develop Gabor-like oriented filters. Finally, due to the frequency dependence of the model (Sjöström, Turrigiano, and Nelson 2001), a plastic network under this model exhibits a tight relation between connectivity and coding. Under rate coding the network supports a few strong bidirectional connections in a sea of weak connections as measured in visual cortex (Song, Sjöström, Reigl, Nelson, and Chklovskii 2005). On the contrary standard STDP models (Gerstner, Kempter, van Hemmen, and Wagner 1996; Kempter, Gerstner, and van Hemmen 1999; Senn, Tsodyks, and Markram 2001; Song, Miller, and Abbott 2000; van Rossum, Bi, and Turrigiano 2000; Gütig, Aharonov, Rotter, and Sompolinsky 2003a) do not support stable bidirectional connections. Interestingly a network under temporal code leads, with our model, to stable unidirectional connections as seen in the barrel cortex (Lefort, Tomm, Sarria, and Petersen 2009).

Connectivity reflects Coding: A Model of Voltage-based Spike-Timing-Dependent-Plasticity with Homeostasis

Claudia Clopath, Lars Büsing*, Eleni Vasilaki, Wulfram Gerstner

Laboratory of Computational Neuroscience

Brain-Mind Institute and School of Computer and Communication Sciences

Ecole Polytechnique Fédérale de Lausanne

1015 Lausanne EPFL, Switzerland

* permanent address: Institut für Grundlagen der Informationsverarbeitung, TU Graz, Austria

June 3, 2009

Abstract

Electrophysiological connectivity patterns in cortex often show a few strong connections in a sea of weak connections. In some brain areas a large fraction of strong connections are bidirectional, in others they are mainly unidirectional. In order to explain these connectivity patterns, we use a model of Spike-Timing-Dependent Plasticity where synaptic changes depend on presynaptic spike arrival and the postsynaptic membrane potential, filtered with two different time constants. The model describes several nonlinear effects in STDP experiments, as well as the voltage dependence of plasticity under voltage clamp and classical paradigms of LTP/LTD induction. We show that in a simulated recurrent network of spiking neurons our plasticity rule leads not only to receptive field development, but also to connectivity patterns that reflect the neural code: for temporal coding paradigms strong connections are predominantly unidirectional, whereas they are bidirectional under rate coding. Thus variable connectivity patterns in the brain could reflect different coding principles across brain areas; moreover our simulations suggest that rewiring the network can be surprisingly fast.

1 Introduction

Experience-dependent changes in receptive fields [1, 2, 3] or in learned behavior [4] may occur through changes in synaptic strength. Thus, electrophysiological measurements of functional connectivity patterns in slices of

neural tissue [5, 6] or anatomical connectivity measures [7] can only present a snapshot of the momentary connectivity – which may change with the next set of stimuli. Indeed, modern imaging methods show that spine motility can lead to a rapid rewiring of the connectivity pattern [8, 9] by formation of new synapses or by strengthening or weakening of existing synapses. The question then arises whether the connectivity patterns and changes that are found in experiments can be connected to basic rules of synaptic plasticity, in particular to modern or traditional forms of Hebbian plasticity [10] such as Long-Term Potentiation and Depression [11].

Long-term potentiation LTP and depression LTD of synapses depends on the exact timing of pre- and postsynaptic action potentials [12, 13], but also on postsynaptic voltage [14, 15], and presynaptic stimulation frequency [16]. Spike-Timing-Dependent Plasticity (STDP) has attracted particular interest in recent years, since temporal coding schemes where information is contained in the exact timing of spikes rather than mean frequency could be learned by a neural system using STDP [17, 18, 19, 20, 21]. The question, however, whether STDP is more fundamental than frequency dependent plasticity or voltage dependent plasticity rules has not been resolved, despite an intense debate [22]. Moreover it is unclear how the interplay of coding and plasticity yield the functional connectivity patterns seen in experiments. In particular, the presence or absence of bidirectional connectivity between cortical pyramidal neurons seems to be contradictory across experimental preparations in visual [5] or somatosensory cortex [6].

Recent experiments have shown that STDP is strongly influenced by postsynaptic voltage before action potential firing [23], but could not answer the question whether spike timing dependence is a direct consequence of voltage dependence, or the manifestation of an independent process. In addition, STDP depends on stimulation frequency [23] suggesting an interaction between timing and frequency dependent processes — or this interaction could be the manifestation of a single process in different experimental paradigms. We show that a simple Hebbian plasticity rule that pairs presynaptic spike arrival with the postsynaptic membrane potential is sufficient to explain STDP and the dependence of plasticity upon presynaptic stimulation frequency. Moreover, the intricate interplay of voltage and spike-timing dependence seen in experiments [23] as well as the frequency dependence of STDP can be explained in our model from one single principle. In contrast of earlier attempts towards a unified description of synaptic plasticity rule that focused on detailed biophysical descriptions [24, 25], our model is a mechanistic one (phenomenological model). It does not give an explicit interpretation in terms of biophysical quantities such a Calcium concentration [24], CaMKII [25], glutamate binding, NMDA receptors etc. Rather it aims at a minimal description of the major phenomena observed in electrophysiology experiments.

The advantage of such a minimal model is that it allows us to discuss functional consequences in small [26, 27], and possibly even large [28, 29], networks. We show that in small networks of up to 10 neurons the learning rule leads to input specificity, necessary for receptive field development - similar to earlier models of STDP [17, 26] or rate-based plasticity rules [30, 31]. Going significantly beyond earlier studies we explicitly address the question of whether functional connectivity patterns of cortical pyramidal neurons measured in recent electrophysiological studies [5, 6] could be the result of plasticity during continued stimulation of neuronal model networks. We found that connectivity patterns strongly depend on the underlying coding hypothesis: With a temporal coding hypothesis, where input spikes arrive in a fixed temporal order, the recurrent network develops a connectivity

pattern with a few strong unidirecitional connections. However, under a rate coding paradigm, where stimuli are stationary during a few hundred milliseconds the same network exhibits sustained and strong bidirectional connections. This is in striking contrast to standard STDP rules where bidirectional connections are impossible [26].

The mathematical simplicity of the model enables us to identify conditions under which it becomes equivalent to the well-known Bienenstock-Cooper-Munro model [30] used in classical rate-based descriptions of developmental learning; and equivalent to some earlier models of STDP [32] — and why our model is fundamentally different from classical STDP models [17, 26, 21], widely used for temporal coding.

2 Results

In order to study how connectivity patterns in cortex can emerge from an interplay of plasticity rules and coding, we need a plasticity rule that is consistent with a large body of experiments, not just a single paradigm such as STDP. Since synaptic depression and potentiation take place through different pathways [33] our model uses separate additive contributions to the plasticity rule, one for LTD and another one for LTP (see Fig. 1 and methods).

2.1 Fitting the Plasticity Model to Experimental Data

Consistent with voltage clamp [15] and stationary depolarization experiments [14] LTD is triggered in our model if presynaptic spike arrival occurs while the membrane potential of the postsynaptic neuron is slightly depolarized (above a threshold θ_-) whereas LTP occurs if depolarization is big (above a second threshold θ_+ (see Fig. 1). The mathematical formulation of the plasticity rule makes a distinction between the momentary voltage u and the low-pass filtered voltage variables $\bar{u}_-$ or $\bar{u}_+$ which denote temporal averages of the voltage over the recent past (the symbols $\bar{u}_-$ and $\bar{u}_+$ indicate filtering of u with two different time constants). Similarly, the event x of presynaptic spike arrival needs to be distinguished from the trace $\bar{x}(t)$ that is left at the synapse after stimulation by neurotransmitter. Potentiation occurs only if the momentary voltage is above θ_+ (this condition is fulfilled during action potential firing) AND the average voltage $\bar{u}_+$ above θ_- (this is fulfilled if there has a been a depolarization in the recent past) AND the trace $\bar{x}$ left by a previous presynaptic spike event is nonzero (this condition holds if a presynaptic spike arrived a few milliseconds earlier at the synapse); these conditions for plasticity are illustrated in Fig. 1B. LTD occurs if the average voltage $\bar{u}_-$ is above rest at the moment of a presynaptic spike arrival (see Fig. 1A). The amount of LTD in our model depends on homeostatic process on a slower time scale [34]. Low-pass filtering of the voltage by the variable ($\bar{u}_-$ or $\bar{u}_+$) refers to some unidentified intracellular processes triggered by depolarization, e.g., increase in calcium concentration or second messengers messenger chains. Similarly, the biophysical nature of the trace $\bar{x}$ is irrelevant for the functionality of the model, but a good candidate process is the fraction of glutamate bound to postsynaptic receptors.

We checked the performance of the model on a simulated STDP protocol, where presynaptic spikes arrive

3

a few milliseconds before or after a postsynaptic spike that is triggered by a strong depolarizing current pulse. If a post-pre pairing with a timing difference of 10 millisecond is repeated 60 times at frequencies below 35Hz, LTD occurs in our model (Fig. 2 A, B), consistent with experiments [23]. Repeated pre-post pairings (with 10 millisecond timing difference) at frequencies above 10Hz yield LTP, but pairings at 0.1Hz do not show any significant change in the model or in experiments [23]. In the model these results can be explained by the fact that at 0.1Hz repetition frequency, the low-pass filtered voltage $\bar{u}_+$ which increases abruptly during postsynaptic spiking decays back to zero before the next impulse arrives, so that LTP can not be triggered. However, since LTD in the model requires only a weak depolarization of $\bar{u}_-$ at the moment of presynaptic spike arrival, post-pre pairings give rise to depression, even at very low frequency. At repetition frequencies of 50Hz, the post-pre paradigm is nearly indistinguishable from a pre-post timing, and LTP dominates.

Since spike-timing dependence in our model is induced only indirectly via voltage dependence of the model, we wondered whether our model would also be able to account for the intricate interactions of voltage and spike timing found by Sjöström et al. [23]. If a pre-post protocol at 0.1Hz, that normally does not induce LTP, is combined with a depolarizing current pulse (lasting from 50ms before to 50ms after the postsynaptic firing event), then potentiation is observed in the experiments [23], as well as in our model (Fig. 2 C, F, I). Due to the injected current, the low-pass filtered voltage variable $\bar{u}_+$ is depolarized before the pairing. Thus at the the moment of the postsynaptic spike, the average voltage $\bar{u}_+$ is above the threshold θ_- leading to potentiation. Similarly, a pre-post protocol that normally leads to LTP can be blocked if the postsynaptic spikes are triggered on the background of a hyperpolarizing current (Fig. 2 E, H, I).

In order to study some nonlinear aspects of STDP, we simulate a protocol of burst-timing-dependent plasticity where presynaptic spikes are paired with 1, 2 or 3 postsynaptic spikes [35] (see Methods). We observe that 60 pre-post pairs at 0.1Hz do not change the synaptic weight, as discussed above. However, repeated triplets pre-post-post generate potentiation in our model because the first postsynaptic spike induces a depolarizing spike after potential so that $\bar{u}_+$ is depolarized. Adding a third postsynaptic spike to the protocol (i.e., quadruplets pre-post-post-post) does not lead to stronger LTP (Fig. 3A). Our model also describes the dependence of LTP upon the intra-burst frequency (Fig. 3B). At an intra-burst frequency of 20Hz, no LTP occurs, because the second spike in the burst comes so late that the presynaptic trace $\bar{x}$ has decayed back to zero. At higher intra-burst frequencies, the three conditions for LTP ($u(t) > \theta_+$ and $\bar{u}_+ > \theta_-$ and $\bar{x} > 0$) are fulfilled. The burst timing dependence (Fig. 3C) is qualitatively similar to that found in experiments [35], but only four of the six experimental data points are quantitatively reproduced by the model.

2.2 Functional implications

Connectivity patterns in a local cortical circuit have been shown to be non-random, i.e. the majority of connections are weak and the rare strong ones have a high probability of being bidirectional [5]. However, standard models of STDP do not exhibit stable bidirectional connections [36]. Intuitively, if the cell A fires before the cell B, a pre-post pairing for the 'AB' connection is formed so that the connection is strengthened. The post-pre

pairing occurring at the same time in the 'BA' connection leads to depression. Therefore it is impossible to strengthen both connections at the same time. Moreover, in order to assure long-term stability of firing rates parameters in standard STDP rules are typically chosen such that inhibition slightly dominates excitation [17] which implies that under purely random spike firing connections decrease, rather than increase. However, the non-linearity aspects of plasticity in our model change such a simple picture. If we simulate two neurons with bidirectional connections at low firing rates, the plasticity model behaves like standard STDP and only unidirectional connections emerge. However, from Fig. 3B we expect that at higher neuronal firing rates, our model could develop a stable bidirectional connection, in striking contrast to standard STDP rules [21].

Since bidirectional connections require neurons to fire at a high rate, we wondered how coding and connectivity relate to each other. We hypothesized that bidirectional connections are supported by rate-coding as opposed to temporal-coding. To test this idea we first simulated a small network of 10 all-to-all connected neurons in a simplified rate-coding scheme where each neuron fires at a fixed frequency, but the frequency varies across neurons. We find that bidirectional connections are formed only between those neurons that both fire at a high rate, but not if one or both of the neurons fire at low frequencies (Fig. 4A). In a second paradigm, the neurons in the same network are stimulated such that they are firing in a distinct order (1, 2 , 3,..) mimicking an extreme form of temporal coding [37]. In that case, the weights form a loop where strong connections from 1 to 2, 2 to 3, ... develop, but no bidirectional connections (Fig. 4B). These results are in striking contrast to simulation experiment with a standard STDP rule, where connections are always unidirectional, independently of coding (Fig. 4C, D).

We wondered whether the same results would emerge in a more realistic network of excitatory and inhibitory neurons driven by feedforward input. We simulated a network of 10 excitatory neurons and 3 inhibitory neurons. Each inhibitory neuron receives input from 8 randomly selected excitatory neurons and randomly projects back to 6 excitatory neurons. In addition to the recurrent input, each excitatory neuron receives feedforward spike input from 500 presynaptic neurons j that generate stochastic Poisson input at a rate ν_j. The rates of neighboring input neurons are correlated, mimicking the presence or absence of spatially extended objects. In a rate-coding scheme, the location of the stimulus is switched every 100ms to a new random position. In case of retinal input, this would correspond to a situation where the subject fixates every 100ms on a new stationary stimulus. In a temporal-coding paradigm, the model input is shifted every 20ms to a neighboring location, mimicking movement of an object across an array of sensory receptors. For both scenarios the network is identical. Feedforward connections and lateral connections between model pyramidal neurons are plastic whereas connections to and from inhibitory neurons are fixed.

After 1000s of stimulation with the rate-coding paradigm, the excitatory neurons developed localized receptive fields and a structured pattern of synaptic connections (Fig. 5B). While the labeling of the excitatory neurons at the beginning of the experiment was randomly assigned, we can relabel the neurons after the formation of lateral connectivity patterns so that neurons with strong reciprocal connections have similar indices, reflecting the neighborhood relation of the network topology. After reordering we can clearly distinguish that three groups of neurons have been formed, characterized by similar receptive fields and strong bidirectional

connectivity within the group, and different receptive fields and no lateral connectivity between groups (Fig. 5C). If the overall amplitude of plastic changes is small (compared to that found in the experiments) the pattern of lateral connectivity is stable and shows a few strong bidirectional connections in a sea of weak lateral connectivity. Unidirectional strong connections are nearly absent. If the amplitude and rate of plasticity is more realistic and in agreement with the data of Fig. 2, then the pattern of lateral connectivity changes between one snapshot and another one 5 seconds later, but the overall pattern is stable when averaged over 100s. In each snapshot, about half of the strong connections are bidirectional (Fig. 5H).

This is in striking contrast with the temporal coding paradigm. Neurons develop receptive fields similar to those seen with the rate-coding paradigm. As expected for temporal Hebbian learning rate [21] the receptive field slowly shifts over time. More importantly, amongst the lateral connections, strong reciprocal links are completely absent (Fig. 6). This suggests that temporal coding paradigms are reflected in the functional connectivity pattern by strong uni-directional connections whereas rate coding leads to strong bidirectional connections.

3 Discussion

Plasticity models over the last decades have primarily focused on questions of development of receptive fields and cortical maps [30], or memory formation [38]. Because traditional plasticity rules are rate models, the relation between coding and connectivity could not be studied. Our plasticity rule is formulated on the level of postsynaptic voltage. Since action potentials present large and narrow voltage peaks, they act as singular events in a voltage rule so that in the presence of spike our rule turns automatically into spike-timing dependent rule. Indeed, for spike coding (and in the absence of significant subthreshold voltage manipulations) our plasticity rule behaves like a STDP rule where triplets of spikes with pre-post-post or post-pre-post timing evoke LTP, whereas pairs with post-pre timing evoke LTD. Moreover, for rate coding where pre- and postsynaptic neurons fire with Poisson firing statistics, our plasticity rule presents structural similarities to the model of Bienenstock, Cooper, and Munro (BCM-model, [30]). Both our spiking rule and the rate-based BCM model require presynaptic activity in order to induce a change. Furthermore for our rule as well as for the simplest BCM rule (see [30]), the depression terms are linear and the potentiation terms are quadratic in the postsynaptic variables (i.e., the postsynaptic potential or the postsynaptic firing rate). Beyond these qualitative similarities, an approximate quantitative relation between the BCM model and our model can be constructed under appropriate assumptions. In this case the total weight change Δw in our model is proportional to $\nu^{\mathrm{pre}}\nu^{\mathrm{post}}(\nu^{\mathrm{post}} - \vartheta)$ where ν^{pre} and ν^{post} denotes the firing rate of a pre- and postsynaptic neurons, respectively and ϑ is a sliding threshold related to the ratio between the LTP and LTD inducing processes (see methods).

Due to its similarities to BCM, it is not surprising that our spike-based learning rule with sliding threshold is able to support independent component analysis (ICA) that has been hypothesized to underly receptive field development [30, 39]. In our experiments, the input consists of small patches of natural images using standard preprocessing [40]. Image patches are selected randomly and presented to the neuron for $T = 200$ms, which

is on the order of a fixation time between saccades [41]. Pixel intensities above an average grey value are converted to spike trains of ON-cells and and those below reference intensity to spikes in OFF-cells, using the relative intensity as the rate of a Poisson process. The spike trains from ON- and OFF-cells are the input to a cortical neuron. The synaptic weights undergo plasticity following our learning rule (Eq. 3). After learning, the weights exhibit a spatial structure that can be interpreted as a receptive field (Fig. 7). In contrast to the principal component analysis of the image patches (as for example implemented by Hebbian learning in linear neurons [42]), the receptive fields are *localized* (i.e. the region with significant weights does not stretch across the whole image patch). Development of localized receptive fields can be interpreted as a signature of ICA [40]. In contrast to most other ICA algorithms [43] our rule is biologically more plausible since it is consistent with a large body of plasticity experiments.

For a comparison of our model with experiments we have mainly focused on experiments in slices of visual cortex, but some of the results can also be related to work in hippocampus. First, as the model explicitly takes into account the postsynaptic membrane potential it can successfully reproduce the voltage dependence of LTP/LTD seen in experiments under depolarization of the postsynaptic membrane [14, 15]. Second, for classical STDP experiments such as [13, 23, 44], which have a stimulation protocol unambiguously defined in terms of pre- and postsynaptic spike times, the model gives a timing dependence reminiscent of the typical STDP function [13]. Moreover in contrast to standard STDP rules [21], more complicated effects such as the pairing frequency dependence [23] and burst-timing dependence plasticity [35] are qualitatively described. In addition the rule is expected to reproduce the triplet and quadruplet experiments in hippocampal slices [44] (data not shown), because for all STDP protocols the plasticity rule in this paper is similar to an earlier nonlinear STDP rule [32]. Deriving STDP rules from voltage dependence has been attempted before [45, 46]. However, since these earlier models use the momentary voltage [46] or its derivative [45], rather than a combination of momentary and averaged voltage as in our model, these earlier models cannot account for the broad range of nonlinear effects in STDP experiments or interaction of voltage and spike-timing. Our model shows similarities with LTP induction in the TagTriC model [47], but the TagTriC model focuses on the long-term stability of synapses, rather than spike timing dependence of the induction mechanism.

Our plasticity rule allows to explain experiments from two different laboratories by one single principle. Both the "potentiation is rescued by depolarization" [23] scenario (Fig. 2F) and that of burst-timing dependent LTP [35] (Fig. 3) show that LTP at low frequency is induced when the membrane is depolarized before the pre-post pairing. This depolarization can be due to a previous spike during a postsynaptic burst [35] or to a depolarization current. Our model is also consistent with results that LTP can be induced in distal synapses only if additional cooperative input or dendritic depolarization prevent failure of backpropagating action potentials [48]. A further unexpected result is that, with the set of parameters derived from visual cortex slice experiments, synapses fluctuate between strong and weak weights. This aspect is interesting in view of synapse mobility reported in imaging experiment [8].

There are, however, certain limitations to our plasticity rule. First, we did not address the problem of weight dependence of synaptic plasticity and simply assumed that weights can grow to a hard upper bound. Neverthe-

less, the rule can be easily changed to soft bounds [21] by changing the prefactors A_{LTP}, A_{LTD} accordingly [47]. Second, short term plasticity [49] could be added for a better description of the plasticity phenomena occurring especially during high frequency protocols. Third, our plasticity rule describes only induction of potentiation or depression during the early phase of LTP/LTD [50]. Additional mechanisms need to be implemented in the model to describe the transition from early to late LTP/LTD [47, 51]. Finally, in modeling voltage-clamp experiments, we assume in our model a unique voltage throughout the whole neuron. In particular the dendrite is assumed to be equipotential to the soma. Yet, experiments controlling the voltage at the soma do not guarantee an equal or even fixed voltage at the synapse with respect to the soma. An obvious and promising improvement would be to use a multi-compartment neuron model (e.g. distinct compartments for the soma and dendrites). In the presented work we did not use a more sophisticated multi-compartment model as this would introduce a considerable number of new parameters making overfitting more likely to occur.

Our plasticity model leads to several predictions that could be tested in slice experiments. First, under the assumption of voltage clamp, our rule is linear in the presynaptic activities (see Methods). Thus the model predicts that in voltage clamp experiments the weight change is only dependent on the voltage and the *number* of presynaptic spikes but not on their exact timing (e.g., low frequency, tetanus, burst input should give the same result). Second, in the scenario where potentiation is rescued by depolarization, the amount of weight change should be the same whether a depolarizing current of amplitude B stops precisely when the postsynaptic spike is triggered or whether a current of slightly bigger amplitude B' stops a few milliseconds earlier. Third, multiple STDP experiments have shown that pre-post pairing (with 10 millisecond timing difference) repeated at 10Hz leads to potentiation [23]. In our plasticity model, LTP occurs in that case because the depolarizing spike-afterpotential of the last postsynaptic spike leads to an increase of the filtered membrane voltage just before the next postsynaptic spike. If this interpretation is correct, a hyperpolarizing current sufficient to cancel the spike afterpotential during 40 milliseconds should block LTP (note that this is different from blocking LTP by a hyperpolarizing current a few milliseconds *before* the next spike [23]). Alternatively cutting dendrites, i.e. dendrotomy [52] would sharpen the spike after potential.

The influence of STDP on temporal coding has been studied in the past primarily with respect to changes in the feedforward connections [21]. The effect of STDP on lateral connectivity has been studied much less [28, 29, 27]. We have shown in this paper that, because of STDP, coding influences the network topology, because different codes give different patterns of lateral connectivity. Our results are in contrast to standard STDP rules which always suppress short loops, and in particular bidirectional connections [36]. Our more realistic plasticity model shows that under a rate coding paradigm bidirectional connectivity and highly connected clusters with multiple loops are not only possible, but even dominant. It is only for temporal coding, that our biologically plausible rule leads to dominant unilateral directions. Our model also predicts that for a code consisting of synchronous firing events at low frequencies synapses decrease, consistent with earlier findings [27]. We speculate that the differences in coding between different brain areas could lead, even if the learning rule were exactly the same, to different network topologies. Our model predicts that experiments where cells in a recurrent network are repeatedly stimulated in a fixed order would decrease the fraction of strong bidirectional connections, whereas

a stimulation pattern where clusters of neuron fire at high rate during episodes of a few hundred milliseconds would increase this fraction. In this views it is tempting to connect the low degree of bidirectional connectivity in barrel cortex [6] to the bigger importance of temporal structure in whisker input [37], compared to visual input.

4 Acknowledgments

5 Figure Captions

Figure 1: Illustration of the model. Synaptic weights react to presynaptic events (top) and postsynaptic membrane potential (bottom) A. The synaptic weight is decreased if a presynaptic spike x (green) arrives when the low pass filtered value $\bar{u}_-$ (magenta) of the membrane potential is above θ_- (dashed horizontal line) B. The synaptic weight is increased if the membrane potential u (black) is above a threshold θ^+ and the low pass filtered value of the membrane potential $\bar{u}_+$ (blue) higher than a threshold θ^- as well as the presynaptic low pass filter $\bar{x}$ (orange) non zero. C. Step current injection makes the postsynaptic neuron fire at 50Hz in the absence of presynaptic stimulation (membrane potential u in black). No weight change is observed. Note the depolarizing spike-afterpotential consistent with experimental data D., reproduced from [23]. E-H. Voltage clamp experiment. A neuron receives weak presynaptic stimulation of 2Hz during 50s while the postsynaptic voltage is clamped to values between -60mV and 0mV. E-G. Schematic drawing of the trace $\bar{x}$ (orange) of the presynaptic spike train (green) as well as the voltage (black) and the synaptic weight (blue) for the experimental conditions E. Hyperpolarization F. Slight depolarization and G. Large depolarization. H. The weight change as a function of clamped voltage using the standard set of parameters for visual cortex data (blue line, voltage paired with 25 spikes at the synapse). With a different set of parameters the model fits experimental data (red circles) in hippocampal slices [15], see methods for details.

Figure 2: A-B. Simulated STDP experiments. A. Spike-timing dependent learning window. The change of the synaptic weight is shown for different time intervals T between the presynaptic and the postsynaptic spike using 60 presynaptic/postsynaptic spike pairs at 20Hz. B. Weight change as a function of repetition frequency for 5 spike pairs at frequency ρ with a time delay of +10ms (pre-post, blue) and -10ms (post-pre, red), repeated 15 times at 0.1Hz (only 10 times for frequency of ρ=0.1Hz). Weight changes are shown as a function of the frequency, dots represent the data taken from Sjöström et al. [23] and lines the plasticity model simulation. C-I. Interaction of voltage and STDP. C-E. Schematic induction protocols (green: presynaptic input, black: postsynaptic current, blue: evolution of synaptic weight). C. Low-Frequency Potentiation is rescued by depolarization [23]. Low frequency (0.1Hz) pre-post spike pairs yield LTP if a 100ms-long depolarized current is injected around the pairing. D. LTP fails in the previous scenario if an additional brief hyperpolarized pulse is applied 14-ms before postsynaptic spike so that voltage is brought to rest. E. Hyperpolarization preceding action potential prevents potentiation. Sjöström et al. [23] show that high frequency (40Hz) pairing leads to LTP. However, when a constant hyperpolarizing current is applied on top of the short pulses inducing the spikes, no weight change is measured. F. The simulated postsynaptic voltage u (black) following protocol A. is shown as well as the temporal averages $\bar{u}_-$ (magenta) and $\bar{u}_+$ (blue). The presynaptic spike time is indicated by the green arrow. Using the model Eq. 3 this setting results in potentiation. G. Same as F, but following protocol D. No weight change is measured. H. Same as F., but following protocol E. No weight change is measured. I. Histogram summarizing the normalized synaptic weight of the simulation (bar) and the experimental data [23] (dot, blue bar=variance) 0.1Hz pairing (control 1); 0.1Hz pairing with the depolarization (protocol C.); 0.1Hz pairing with the depolarization and brief hyperpolarization (protocol D.); 40Hz pairing (control 2); 40Hz pairing with the constant hyperpolarization (protocol E.). The parameters are summarized in Table 1B.

Figure 3: Burst-timing-dependent plasticity. One presynaptic spike is paired with a burst of postsynaptic spikes. This pairing is repeated 60 times at 0.1Hz. A. Normalized weight is shown as a function of the number of postsynaptic spikes (1,2,3) at 50Hz. (dots: data from [35], crosses: simulation). The presynaptic spike is paired +10ms before the first postsynaptic spike (blue) or -10ms after (red). B. Normalized weight as a function of the frequency between the three postsynaptic action potentials (dot: data, line: simulation; blue: pre-post, red: post-pre). C. Normalized weight as a function of the timing between the presynaptic spike and the first postsynaptic spike of a 3-spike burst at 50Hz (dot: data, line: simulation). A hard upper bound has been set to 250% normalized weight.

Figure 4: Weight evolution in a all-to-all connected network of 10 neurons. A. Rate code: Neurons fire at different frequencies, neuron 1 at 2Hz, neuron 2 at 4Hz... neuron 10 at 20Hz. The weights (bottom) averaged over 100s show that neurons with high firing rates develop strong bidirectional connections (light blue: weak connections (under 2/3 of the maximal value); yellow: strong unidirectional connections (above 2/3 of the maximal value); brown: strong bidirectional connections). The cluster is schematically represented on top ("after"). B. Temporal code: Neurons fire successively every 20ms (neuron 1 then 20ms later neuron 2, then 3..). Connections (bottom) are unidirectional with strong connections from presynaptic neuron with index n (vertical axis) to postsynaptic neuron with index n+1, n+2 and n+3 leading to a ring-like topology (top: schematic). C. D. Same but with standard STDP rule [17, 26, 21]. Bidirectional connections are impossible.

Figure 5: Plasticity during rate coding. A network of 10 excitatory neurons is connected to 3 inhibitory neurons and receives feedforward inputs from 500 Poisson spike trains with a Gaussian profile of firing rates. The center of the Gaussian is shifted randomly every 100ms A. The schematic figure shows the network before and after the plasticity experiment. B-E. Learning with small amplitudes. Model parameters are taken from table 1B (visual cortex data) except for the amplitudes A_{LTP} and A_{LTD} which are reduced by a factor 100. B. Mean feedforward weights (left) and recurrent excitatory weights (right) averaged over 100s. The grey level graph for the feedforward weights (left) indicates that neurons develop receptive fields that are localized in the input space. The recurrent weights (right) are classified into: light blue - weak (less than 2/3 of the maximal weight), yellow - strong (more than 2/3 of the maximal weight) unidirectional, brown - strong reciprocal connections. The diagonal is white, since self-connections do not exist in the model. C. Same as (B) but for the sake of visual clarity the index of neurons is reordered so that neurons with similar receptive fields have adjacent numbers, highlighting that neurons with similar receptive fields (e.g., neurons 1 to 4) have strong bilateral connections. D. Three snap shots of the recurrent connections taken 5s apart indicating that recurrent connections are stable. E. Histogram of reciprocal, unidirectional and weak connections in the recurrent network averaged over 100s as in (B). The total number of weight fluctuations during 100s is 79 (noted on the figure). The histogram shows an average of 10 repetitions (errorbars are the standard deviation). F-I. Rate code during learning with normal amplitudes. Same network as before but standard set of parameters (table 1B, visual cortex). F. Receptive fields are localized; G. Reordering allows to visualize that the strong bidirectional give rise to clusters of neurons. These clusters are stable when averaged over 100 seconds, but H connections can change from one time step to the next. I. The percentage of reciprocal connections is high, but because of fluctuations (fluc) more than 1000 transitions between strong unidirectional to strong bidirectional or back occur during 100 seconds.

Figure 6: Temporal coding paradigm. The setting is the same as in Fig. 5 (parameters from table 1B, visual cortex) but the input patterns are moved successively every 20ms, corresponding to a step-wise motion of the Gaussian stimulus profile across the input neurons. A. The schematic figure shows the network before and after the plasticity experiment. B. Receptive fields are localized, but in the recurrent network no reciprocal connections appear. C. Reordering of neurons shows that the network develops a ring-like structure with strong unidirectional connections from neuron 8 (vertical axis) to neuron 7 and 6 (horizontal axis); from neuron 7 to neuron 6, 5, and 4; from neuron 4 to neuron 3, 2, and 1 etc. D. Some of the strong unilateral connections appear or disappear from one time step to the next, but the ring-like network structure persists, since the lines just below the diagonal are much more populated than the line above the diagonal. E. Reciprocal connections are completely absent, but unidirectional connections fluctuate several times between 'weak' and 'strong' during 100s.

Figure 7: A small patch of 16x16 pixels is chosen from the whitened natural images benchmark [40]. The patch is selected randomly and is presented as input to 512 neurons for 200ms. The positive part of the image is used as the firing rate to generate Poisson spike trains of the 256 "ON" inputs and the negative one for the 256 "OFF" inputs. B. The weights after convergence are shown for the "ON" inputs and the "OFF" inputs rearranged on a 16x16 image. The filter is calculated by subtracting the "OFF" weights from the "ON" weights. The filter is localized and bimodal, corresponding to an oriented receptive field.

Table 1: A. Parameters for the neuron model. B. Plasticity rule parameters for the various experiments. VC stands for Visual Cortex cells (for experimental details see [23], * standard set of parameters), SC for Somatosensory Cortex cells (see [35]) and HP for Hippocampal cells (see [15]). Bold numbers indicate the free parameters fitted to experimental data. Other parameters are set in advance to values based on the literature.

6 Methods

6.1 Neuron Model

In contrast to standard models of STDP, the plasticity model presented in this paper involves the postsynaptic membrane potential $u(t)$. Hence, predicting the weight change in a given experimental paradigm requires a neuron model that describes the temporal evolution of $u(t)$. For this purpose we chose the adaptive Exponential Integrate-and-Fire (AdEx) model [53] with an additional current describing the depolarizing spike after potential [54]. The neuron model is described by a voltage equation:

$$C\frac{d}{dt}u = -g_L(u - E_L) + g_L\Delta_T \exp\left(\frac{u - V_T}{\Delta_T}\right) - w_{ad} + z + I$$

where C is the membrane capacitance, g_L the leak conductance, E_L the resting potential and I the stimulating current. The exponential term describes the activation of a rapid sodium current. The parameter Δ_T is called the slope factor and V_T the threshold potential [53]. A hyperpolarizing adaptation current is described by the variable w_{ad} with dynamics

$$\tau_{w_{ad}}\frac{d}{dt}w_{ad} = a(u - E_L) - w_{ad},$$

where $\tau_{w_{ad}}$ is the time constant of the adaption of the neuron. Upon firing the variable u is reset to a fixed value V_{reset} whereas w_{ad} is increased by an amount b. The main difference to the Izhikevich model [55] is that

11

the voltage is exponential rather than quadratic allowing a better fit to data [54]. The spike afterpotential of the cells used in typical STDP experiments [23] have a long depolarizing spike after potential. We therefore add an additional current z which is set to a value I_{sp} immediately after a spike occurs and decays otherwise with a time constant τ_z.

$$\tau_z \frac{d}{dt} z = -z,$$

Finally, refractoriness is shown in pyramidal cells [54] and therefore is modeled with the adaptive threshold V_T. Therefore V_T is set to $V_{T_{max}}$ after a spike and decays to $V_{T_{rest}}$ with a time constant τ_{V_T} as measured in [54], i.e.

$$\tau_{V_T} \frac{d}{dt} V_T = -(V_T - V_{T_{rest}}).$$

Parameters for the neuron model are taken from [53] for the AdEx, τ_z is set to 40ms in agreement with [23, 54] and kept fixed throughout all simulations (see table 1A).

6.2 Plasticity Model

Since synaptic depression and potentiation take place through different pathways [33] our model exhibits separate additive contributions to the plasticity rule, one for LTD and another one for LTP.

For the LTD part, we assume that presynaptic spike arrival at synapse i induces depression of the synaptic weight w_i by an amount $-A_{\mathrm{LTD}}[\overline{u}_-(t) - \theta_-]_+$ that is proportional to the average postsynaptic depolarization $\overline{u}_-$. The brackets $[\,]_+$ indicate rectification, i.e. any value $\overline{u}_- < \theta_-$ does not lead to a change and implement experimental findings showing that postsynaptic depolarization should exceed a certain value θ^- to establish depression of the synapse [14] (see Fig. 1H). The quantity $\overline{u}_-(t)$ is an exponential low-pass filtered version of the postsynaptic membrane potential $u(t)$ with a time constant τ_-:

$$\tau_- \frac{d}{dt} \overline{u}_-(t) = -\overline{u}_-(t) + u(t).$$

The variable $\overline{u}_-$ is an abstract variable which could, for instance, reflect the level of calcium concentration [24] or the release of endocannabinoids [56], though such an interpretation is not necessary for our rule. Since the presynaptic spike train is described as a series of short pulses at time t_i^n where i is the index of the synapse and n an index that counts the spike, $X_i(t) = \sum_n \delta(t - t_i^n)$, depression is modeled as the following update rule, see also Fig. 1:

$$\frac{d}{dt} w_i^- = -A_{\mathrm{LTD}}(\bar{\bar{u}}) \, X_i(t) \, [\overline{u}_-(t) - \theta_-]_+ \qquad \text{if } w_i > w_{\min}, \tag{1}$$

where $A_{\mathrm{LTD}}(\bar{\bar{u}})$ is an amplitude parameter that is under the control of homeostatic processes [34]. For slice experiment the parameter has a fixed value extracted from experiment. For network simulations, we make it depend on the mean depolarization $\bar{\bar{u}}$ of the postsynaptic neuron, averaged over a time scale of 1 second. Eq. 1

is a simple method to implement homeostasis; other methods such as weight rescaling would also be possible [34].

For the LTP part, we assume that each presynaptic spike at the synapse w_i increases the trace $\bar{x}_i(t)$ of some biophysical quantity, which decays exponentially with a time constant τ_x in the absence of presynaptic spikes, similar to previous work [17, 32]. The temporal evolution of $\bar{x}_i(t)$ is described by:

$$\tau_x \frac{d}{dt}\bar{x}_i(t) = -\bar{x}_i(t) + X_i(t),$$

where X_i is the spike train defined above. The quantity $\bar{x}_i(t)$ could for example represent the amount of glutamate bound to postsynaptic receptors [32] or the number of NMDA receptors in an activated state. The potentiation of w_i is modeled by the following expression, which is proportional to the trace $\bar{x}_i(t)$ (see also Fig. 1):

$$\frac{d}{dt}w_i^+ = +A_{\mathrm{LTP}}\,\bar{x}_i(t)\,[u(t) - \theta_+]_+\,[\overline{u}_+(t) - \theta_-]_+ \qquad \text{if } w_i < w_{\max}. \tag{2}$$

Here, A_{LTP} is a free amplitude parameter fitted to the data and $\overline{u}_+(t)$ is another low-pass filtered version of $u(t)$ similar to $\overline{u}_-(t)$ but with a shorter time constant τ_+ around 10ms. Thus positive weight changes can occur if the momentary voltage $u(t)$ surpasses a threshold θ_+ and, at the same time the average value $\overline{u}_+(t)$ is above θ_-.

The final rule used in the simulation is described by the equation

$$\frac{d}{dt}w_i = -A_{\mathrm{LTD}}(\bar{\bar{u}})\,X_i(t)\,[\overline{u}_-(t) - \theta_-]_+ + A_{\mathrm{LTP}}\,\bar{x}_i(t)\,[u(t) - \theta_+]_+\,[\overline{u}_+(t) - \theta_-]_+, \tag{3}$$

combined with hard bounds $0 \leq w_i \leq w_{max}$. For network simulation, $A_{\mathrm{LTD}}(\bar{\bar{u}}) = A_{\mathrm{LTD}}\,\frac{\bar{\bar{u}}^2}{u_{ref}^2}$ where u_{ref}^2 is a reference value.

6.3 Parameters and Data Fitting

For the plasticity experiments in slices, we take $\bar{\bar{u}} = u_{ref}$ as fixed and fit the parameters A_{LTD}. The total number of parameters of the plasticity model is then 7. For all data sets, except the one taken from [15], the threshold θ^- is set to the resting potential and θ^+ to the firing threshold of the AdEx model, i. e. $\theta^- = -70.6$mV and $\theta^+ = -45.3$mV. The remaining five parameters τ_x, τ_-, τ_+, A_{LTD} and A_{LTP} are fitted to each data set individually by the following procedure. We calculate the theoretically predicted weight change $\Delta w_i^{\mathrm{th},j}$ by integrating (analytically or numerically) Eq. (**3**), for a given experimental protocol j, as a function of the free parameters. We then estimate the free parameters by minimizing the mean-square error E between the theoretical calculations and the experimental data $\Delta w_i^{\mathrm{exp},j}$:

$$E = \sum_j \left(\Delta w_i^{\mathrm{th},j} - \Delta w_i^{\mathrm{exp},j}\right)^2.$$

For the data set in hippocampus [15], we also fit the two parameters θ^- and θ^+ since completely different preparations and cell type were used. Moreover for this data set, the time constant τ_x is taken from physiological measurements given in [13] and fixed to the values of 16ms. The parameters for the various experiments are summarized in table 1B.

6.4   Protocols and mathematical methods

Voltage clamp experiment. (Fig. 1H) The postsynaptic membrane potential was switched in the simulations to a constant value u_{clamp} chosen from -80mV to 0mV while presynaptic fibers were stimulated with either 25 (blue line) or 100 pulses (red line) at 50Hz. Due to voltage clamping, the actual value of the voltage u itself and the low-pass filtered versions $\bar{u}$ are constant and equal to u_{clamp}. Hence, the synaptic plasticity rule becomes $\frac{d}{dt}w_i = -A_{\text{LTD}} X_i(t) [u_{\text{clamp}} - \theta_-]_+ + A_{\text{LTP}} \bar{x}_i(t) [(u_{\text{clamp}} - \theta_-)(u_{\text{clamp}} - \theta_+)]_+$.

Frequency dependence experiment. (Fig. 2B) Presynaptic spikes in the simulation were paired with postsynaptic spikes that were either advanced by +10ms or delayed by -10ms with respect to the presynaptic spike. This pairing was repeated 5 times with different frequencies ranging from 0.1 to 50Hz. These 5 pairings were repeated 15 times at 0.1Hz. However, the 5 pairing at 0.1Hz were repeated only 10 times to mimic the experimental protocol [23].

Burst-timing-dependent plasticity. (Fig. 3A) The presynaptic spike is paired $\Delta t =+10$ms before (or $\Delta t =-10$ms after) 1, 2 or 3 postsynaptic spikes. The frequency of the burst is 50Hz. The neuron receives 60 pairings at a frequency of 0.1Hz. Fig. 3B: The presynaptic spike is paired with a burst of 3 action potentials ($\Delta t =+10$ms and -10ms), while the burst frequency varies from 20 to 100Hz. Fig. 3C: A presynaptic spike is paired with a burst of 3 postsynaptic action potentials with burst frequency of 50Hz. The time Δt between the presynaptic spike and the first postsynaptic action potential varies from -80 to 40 ms. For a detailed description of the experiments see [35].

Poisson input for functional scenarios. (Fig. 4-7) Poisson inputs are used in all the following experiments. They are generated by a stochastic process where the spike is elicited with a stochastic intensity ν.

Relation between connectivity and coding: Toy model. (Fig. 4) Weights of ten all-to-all connected neurons are initialized at 1, bounded between 0 and 3. Weights evolve with the voltage-based rule with homeostasis (Eq. 3) for 100s. The model is compared to a canonical pair-based STDP model written as $\frac{d}{dt}w_i = -A_{\text{LTD}}^{\text{pair}} X_i \bar{y} + A_{\text{LTP}}^{\text{pair}} \bar{x}_i Y$, where Y is the postsynaptic spike train defined the same way as the presynaptic spike train X_i with a filter of the postsynaptic spikes $\bar{y}$ similar to $\bar{x}_i$. The parameters are chosen $A_{\text{LTD}}^{\text{pair}} = A_{\text{LTP}}^{\text{pair}} = 1e^{-5}$ for the amplitudes and τ_x for the time constant of $\bar{x}_i$ as well as for the time constant of the postsynaptic low-pass filter $\bar{y}$. Rate code: Neuron 1 fire at 2Hz, neuron 2 at 4Hz... neuron 10 at 20Hz

14

following a Poisson statistics, i.e. short current pulses are injected to make the neuron fire with Poisson statistics at this frequency. The neurons have different reference values from $u_{ref}^2 = 60$ to 600mV^2. Temporal code: Neurons fire successively every 20ms, first neuron 1 fires then 20ms later neuron 2 then... 10 then 1 etc, in a loop. The neurons have a reference value set to $u_{ref}^2 = 60\text{mV}^2$.

Rate coding in network simulation. (Fig. 5) Five hundred presynaptic Poisson neurons with firing rates ν_i^{pre} ($1 \le i \le 500$) are connected to 10 postsynaptic excitatory neurons. The inputs rates ν_i^{pre} follow a Gaussian profile, i. e. $\nu_i^{\text{pre}} = A \cdot \exp(-(i - \mu)^2/(2\sigma^2))$, with variance $\sigma = 10$ and amplitude $A = 30\text{Hz}$. The center μ of the Gaussian shifts randomly every 100ms between 10 different positions equally distributed. Circular boundary conditions are assumed, i.e. neuron $i = 500$ is considered as neighbor of $i = 1$. Synaptic weights of the feedforward connections are initialized randomly (uniformly in [0.5,2]) and hard bound are set to 0 and 3. The 10 excitatory neurons are all to all recurrently connected with a starting synaptic weight of 0.25 (hard bounds set to 0 and 0.75). In addition, 3 inhibitory neurons are randomly driven by 8 excitatory neurons and project on 6 excitatory neurons, also chosen randomly. Those random connections are fixed and have a weight equal to 1. The reference value is set to $u_{ref}^2 = 60\text{mV}^2$ and the simulation time to 1000s. Parameters are normally chosen as in table 1B, visual cortex data, except for Fig. 5 B-E, where A_{LTP} and A_{LTD} where reduced by a factor 100.

Temporal coding in network simulation. (Fig. 6) Same setting than rate code but the patterns are presented for 20ms successively (from center position 500, to 450, to 400 etc in a circular manner). The reference value has been set to $u_{ref}^2 = 80\text{mV}^2$.

ICA-like computation - Orientation selectivity with natural images. (Fig. 7) Ten natural images have been taken from the benchmark of Olshausen et al. [40]. A small patch of 16 by 16 pixels from any of the images is randomly chosen every 200ms. After prewhitening, the inputs for the "ON" ("OFF") image are Poisson spike trains generated by the positive (negative) part of the patch (with respect to a reference grey value reflecting the ensemble mean) with maximum frequency of 50Hz. The 2x16x16 inputs are connected to one postsynaptic neuron. The initial weights are set randomly between 0 and 2 and hard bounds are set between 0 and 3. The connections follow the synaptic rule (Eq. 3), where the reference value is set to $u_{ref}^2 = 50\text{mV}^2$. Parameters are chosen as in table 1B (visual cortex data) but A_{LTP} and A_{LTD} where reduced by a factor 10. Every 20 s an extra normalization is applied to equalize the norm of the "ON" weights to the one of the "OFF" weights [31].

References

[1] D.V Buonomano and M.M. Merzenich. Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, 21:149–186, 1998.

[2] Y. Fregnac and D. Shulz. Activity-dependent regulation of receptive field properties of cat area 17 by supervised hebbian learning. *J. Neurobiol*, 41:69–82, 1999.

[3] Robert C. Froemke, Michael M. Merzenich, and Christoph E. Schreiner. A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450:425–429, 2007.

[4] G. H. Recanzone, C. E. Schreiner, and M. M. Merzenich. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *The Journal of Neuroscience*, 13:87–103, 1993.

[5] S. Song, P.J. Sjöström, M. Reigl, S. Nelson, and D.B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3:507–519, 2005.

[6] S. Lefort, C. Tomm, J.C.F. Sarria, and C.C.H. Petersen. The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61:301–316, 2009.

[7] W Denk and H Horstmann. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol*, 2(11):e329. doi:10.1371/journal.pbio.0020329, 2004.

[8] R. Yuste and T. Bonhoeffer. Genesis of dendritic spines: insights from ultrastructural and imaging studies. *Nat Rev Neurosci*, 5(1):24–34, 2004.

[9] J. T. Trachtenberg, B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Welker, and K. Svoboda. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420:788–794, 2002.

[10] D. O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.

[11] Robert C. Malenka and Mark F. Bear. LTP and LTD: An embarassment of riches. *Neuron*, 44:5–21, 2004.

[12] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postysnaptic AP and EPSP. *Science*, 275:213–215, 1997.

[13] G.-q. Bi and M.-m. Poo. Synaptic modification of correlated activity: Hebb's postulate revisited. *Ann. Rev. Neurosci.*, 24:139–166, 2001.

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

[15] A. Ngezahayo, M. Schachner, and A. Artola. Synaptic activation modulates the induction of bidirectional synaptic changes in adult mouse hippocamus. *J. Neuroscience*, 20:2451–2458, 2000.

[16] S. M. Dudek and M. F. Bear. Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. *J. Neuroscience*, 13:2910–2918, 1993.

[17] W. Gerstner, R. Kempter, J.L. van Hemmen, and H. Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–78, 1996.

[18] P.D. Roberts and C.C. Bell. Computational consequences of temporally asymmetric learning rules: II. Sensory image cancellation. *Computational Neuroscience*, 9:67–83, 2000.

[19] R. Legenstein, C. Naeger, and W. Maass. What can a neuron learn with spike-timing dependent plasticity. *Neural Computation*, 17:2337–2382, 2005.

[20] R. Guyonneau, R. VanRullen, and S.J. Thorpe. Neurons tune to the earliest spikes through stdp. *Neural Computation*, 17(4):859–879, 2005.

[21] W. Gerstner and W. K. Kistler. *Spiking Neuron Models*. Cambridge University Press, Cambridge UK, 2002.

[22] John Lisman and Nelson Spruston. Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. *Nature Neuroscience*, 8(7):839–841, 2005.

[23] P.J. Sjöström, G.G. Turrigiano, and S.B. Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32:1149–1164, 2001.

[24] H. Z. Shouval, M. F. Bear, and L. N. Cooper. A unified model of nmda receptor dependent bidirectional synaptic plasticity. *Proc. Natl. Acad. Sci. USA*, 99:10831–10836, 2002.

[25] J.E. Lisman and A.M. Zhabotinsky. A model of synaptic memory: A CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron*, 31:191–201, 2001.

[26] S. Song and L. F. Abbott. Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32:339–350, 2001.

[27] E. V. Lubenov and A. G. Siapas. Decoupling through synchrony in neuronal circuits with propagation delays. *Neuron*, 58:118–131, 2008.

[28] A. Morrison, A. Aertsen, and M. Diesmann. Spike-timing dependent plasticity in balanced random networks. *Neural Computation*, 19:1437–1467, 2007.

[29] Eugene M. Izhikevich and Gerald M. Edelman. Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences*, 105:3593–3598, 2008.

[30] L.N. Cooper, N. Intrator, B.S. Blais, and H. Z. Shouval. *Theory of cortical plasticity*. World Scientific, Singapore, 2004.

[31] K. D. Miller. A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity dependent competition between ON- and OFF-center inputs. *J. Neurosci.*, 14:409–441, 1994.

[32] J. P. Pfister and W. Gerstner. Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neursci.*, 26:9673–9682, 2007.

[33] D.H. O'Connor, G.M. Wittenberg, and S.S.H. Wang. Dissection of Bidirectional Synaptic Plasticity Into Saturable Unidirectional Processes. *Journal of Neurophysiology*, 94:1565–1573, 2005.

[34] G.G. Turrigiano and S.B. Nelson. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5:97–107, 2004.

[35] Thomas Nevian and Bert Sakmann. Spine ca2+ signaling in spike-timing-dependent plasticity. *J. Neurosci.*, 26(43):11001–11013, 2006.

[36] J. Kozloski and G. A. Cecchi. Topological effects of spike timing-dependent plasticity. *arxiv.org*, abs:0810.0029, 2008.

[37] S. P. Jadhav, J. Wolfe, and D. E. Feldman. Sparse temporal coding of elementary tactile features during active whisker sensation. *Nature Neuroscience*, page doi:10.1038/nn.2328, 2009.

[38] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79:2554–2558, 1982.

[39] B. Blais, H. Shouval, and L. Cooper. Receptive field formation in natural scene environments: comparison of single-cell learning rules. *Neural Computation*, 10:1797–1813, 1998.

[40] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[41] S. Martinez-Conde, S. Macknik, and D. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5:229–240, 2004.

[42] E. Oja. A simplified neuron model as a principal component analyzer. *J. Mathematical Biology*, 15:267–273, 1982.

[43] A. Hyvaerinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001.

[44] H.-X. Wang, R.C. Gerkin, D.W. Nauen, and G.-Q. Wang. Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience*, 8:187–193, 2005.

[45] A. Saudargiene, B. Porr, and F. Wörgötter. How the shape of pre- and postsynaptic signals can influence stdp: A biophysical model. *Neural Computation*, 16:595–626, 2003.

[46] J.M. Brader, W. Senn, and S. Fusi. Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Computation*, 19:2881–2912, 2007.

[47] Claudia Clopath, Lorric Ziegler, Eleni Vasilaki, Lars Bsing, and Wulfram Gerstner. Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression. *PLoS Comput Biol*, 4(12), Dec 2008.

[48] P.J. Sjöström and M. Häusser. A Cooperative Switch Determines the Sign of Synaptic Plasticity in Distal Dendrites of Neocortical Pyramidal Neurons. *Neuron*, 51(2):227–238, 2006.

[49] M. Tsodyks and H. Markram. The neural code between neocortical pyramidal neurons depends on neuro-transmitter release probability. *Proc. Natl. Academy of Sci., USA*, 94:719–723, 1997.

[50] U. Frey and R.G.M. Morris. Synaptic tagging and long-term potentiation. *Nature*, 385:533 – 536, 1997.

[51] A.B. Barrett, G.O. Billings, R.G.M. Morris, and M.C.W. van Rossum. State based model of long-term potentiation and synaptic tagging and capture. *Plos Comp Biol*, 5(1):e1000259. doi:10.1371/journal.pcbi.1000259, 2009.

[52] JM. Bekkers and M. Häusser. Targeted dendrotomy reveals active and passive contributions of the dendritic tree to synaptic integration and neuronal output. *Proc. Natl. Acad. Sci. USA*, 104(27):11447–11452, 2007.

[53] R. Brette and W. Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.*, 94:3637 – 3642, 2005.

[54] L Badel, S Lefort, R Brette, CC Petersen, W Gerstner, and MJ Richardson. Dynamic i-v curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J Neurophysiol*, 99:656 – 666, 2008.

[55] E.M. Izhikevich. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15:1063–1070, 2004.

[56] P.J. Sjöström, G.G. Turrigiano, and S.B. Nelson. Endocannabinoid-dependent neocortical layer-5 ltd in the absence of postsynaptic spiking. *J. Neurophysiol.*, 92:3338–3343, 2004.
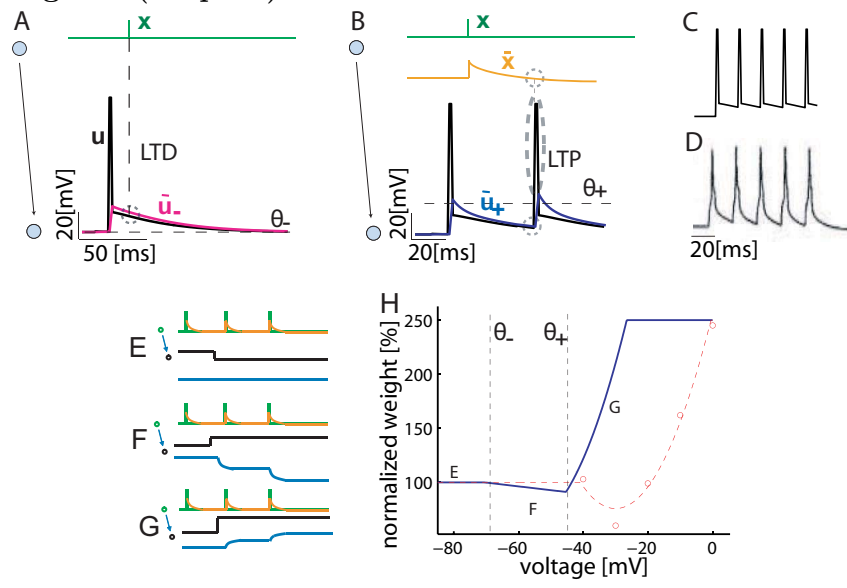
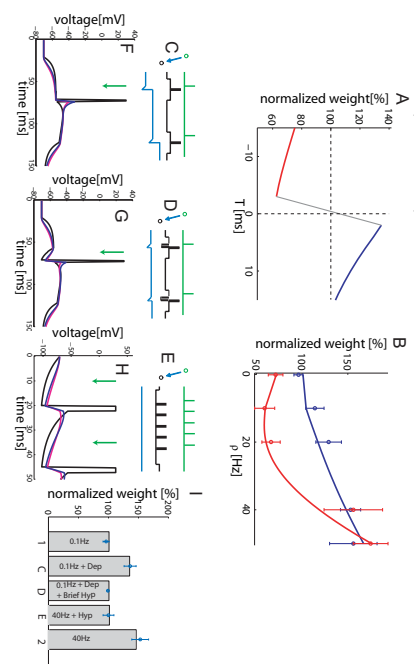**Figure-1(Clopath)**

Figure-2 (Clopath)

Figure-3(Clopath)
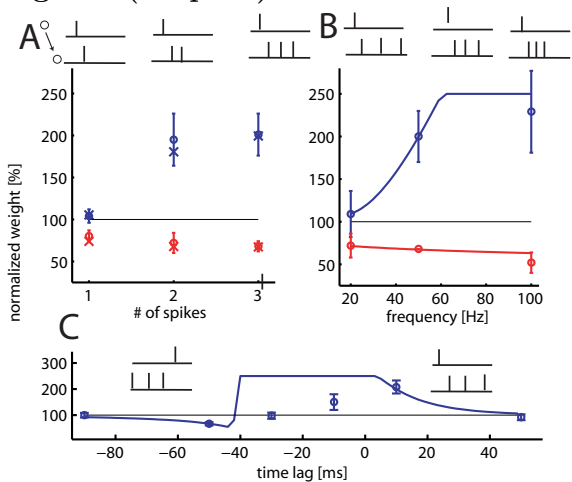
Figure-4(Clopath)

A Rate Code

40Hz 36Hz

4Hz

8Hz

before   after

B Temporal Code

before   after

C

D

Figure-5(Clopath)

A

before   after

B   **C**

D   **E**

fluc = 79

F   **G**

H   **I**

fluc = 1.8e4

Figure-6(Clopath)

A

before   after

B   **C**

D   **E**

fluc = 1.2e5
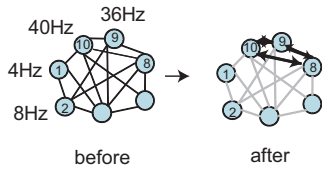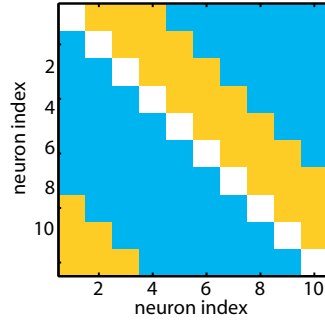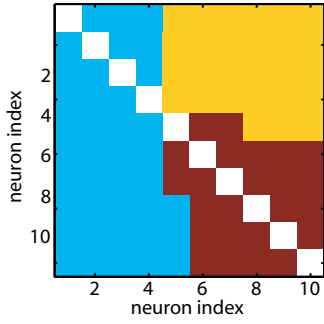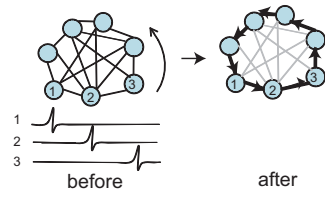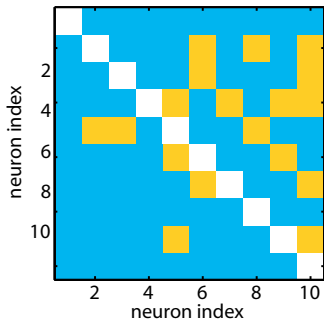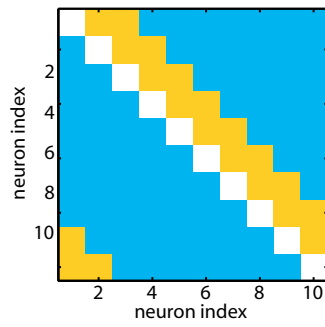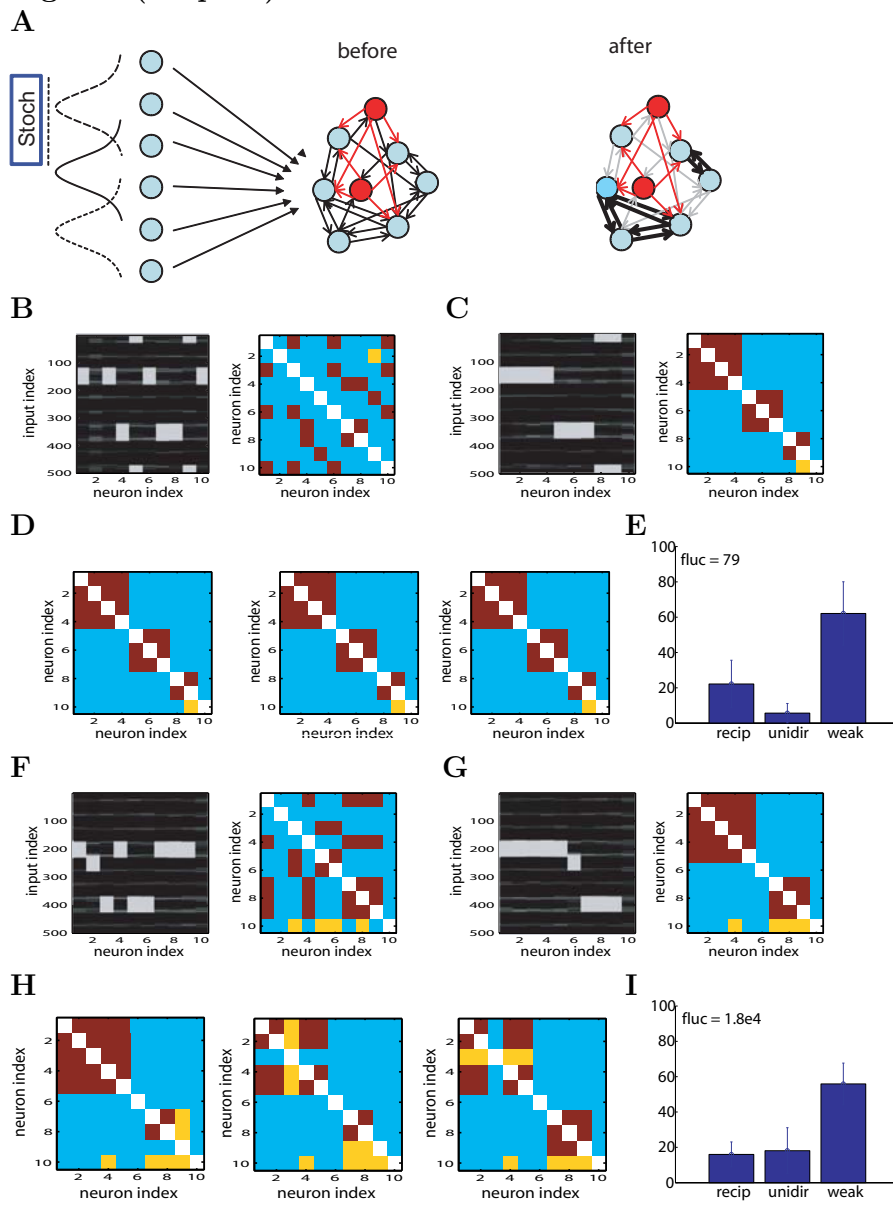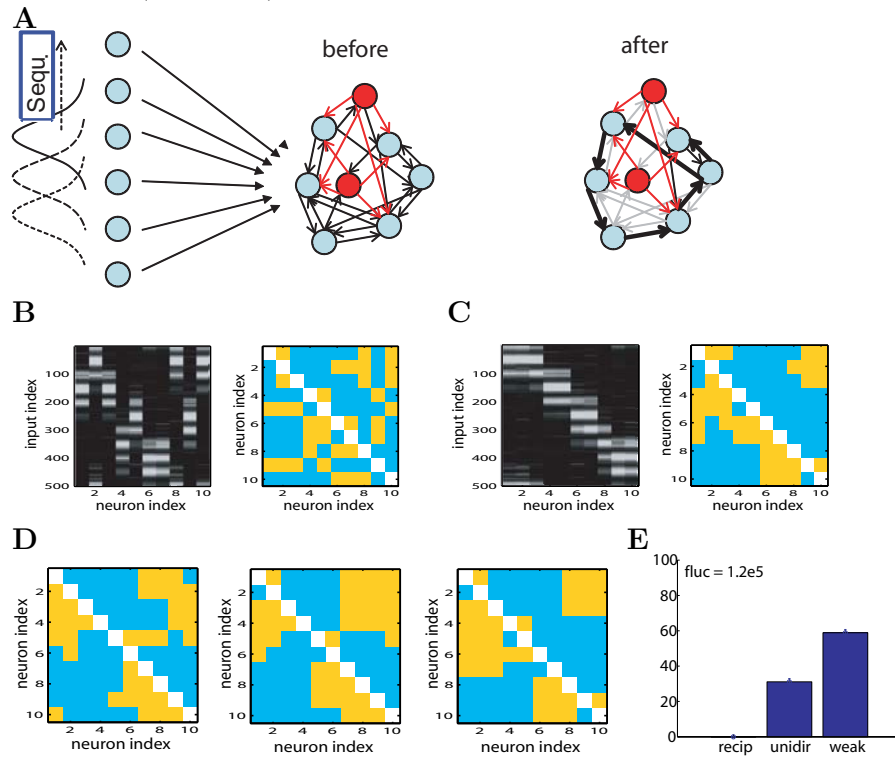
Figure-7(Clopath)

A
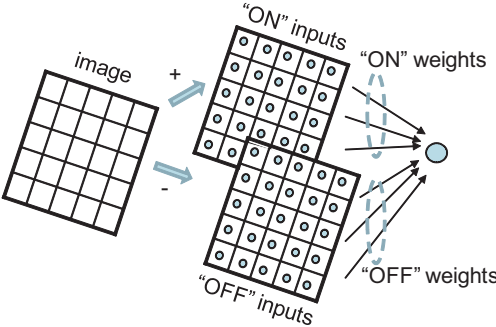
B
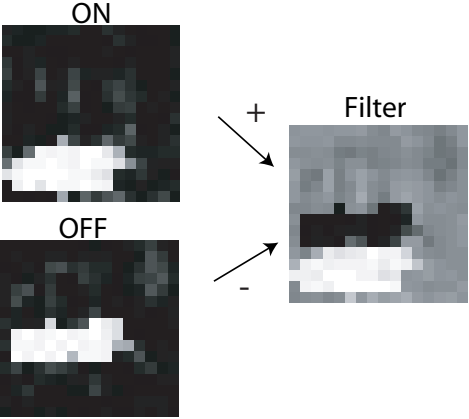
Table-1(Clopath)

A

Parameters	Value
C - membrane capacitance	281pF
g_L - leak conductance	30nS
E_L - resting potential	-70.6mV
Δ_T - slope factor	2mV
$V_{T_{rest}}$ - threshold potential at rest	-50.4mV
$\tau_{w_{ad}}$ - adaptation time constant	144ms
a - subthreshold adaptation	4nS
b - spike triggered adaptation	80.5pA
I_{sp} - spike current after a spike	400nA
τ_z - spike current time constant	40ms
τ_{V_T} - threshold potential time constant	50ms
$V_{T_{max}}$ - threshold potential after a spike	$-30.4mV$

B

Exper.	$\theta_-(mV)$	$\theta_+(mV)$	$A_{\mathrm{LTD}}(mV)^{-1}$	$A_{\mathrm{LTP}}(mV)^{-2}$	$\tau_x(ms)$	$\tau_-(ms)$	$\tau_+(ms)$
VC*	-70.6	-45.3	**14e^{-5}**	**8e^{-5}**	**15**	**10**	**7**
SC	-70.6	-45.3	**21e^{-5}**	**67e^{-5}**	**15**	**8**	**5**
HP	-41	-38	**38e^{-5}**	**2e^{-5}**	16		

A MODEL FOR THE LATE PHASE OF LONG TERM SYNAPTIC PLASTICITY

THE key element in understanding memory is the stable changes of the synapses. It is therefore essential to describe the long lasting changes which are for example measured by synaptic tagging experiments (Frey and Morris 1997). These experiments exhibit three phases leading to a maintenance of synaptic plasticity (1) the induction of long-term potentiation and depression (LTP/LTD) during the early phase of synaptic plasticity and the setting of synaptic tags, (2) a trigger process for protein synthesis, and a slow transition leading to synaptic consolidation during (3) the late phase of synaptic plasticity. In the following paper (Clopath, Ziegler, Vasilaki, Buesing, and Gerstner 2008) we present a minimal model that describes these three different phases of synaptic plasticity. The early phase of plasticity is modeled with the induction model presented in the previous paper (Clopath, Vasilaki, Buesing, and Gerstner xxxx) and leads directly to setting the tag. The synapses are considered discrete, consistent to some experimental evidence that LTP under minimal stimulation (Petersen, Malenka, Nicoll, and Hopfield 1998; O'Connor, Wittenberg, and Wang. 2005) or glutamate encaging (Bagal, Kao, Tang, and Thompson 2005) is a switch-like process. This model explains a large body of experimental data on synaptic tagging and capture, cross-tagging, and the late phases of LTP and LTD. It offers structural similarities to reinforcement learning models such as a dependence on neuromodulation.

Tag-Trigger-Consolidation: A Model of Early and Late Long-Term-Potentiation and Depression

Claudia Clopath[◗], Lorric Ziegler[◗], Eleni Vasilaki, Lars Büsing[¤], Wulfram Gerstner*

Laboratory of Computational Neuroscience, Brain-Mind Institute and School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Abstract

Changes in synaptic efficacies need to be long-lasting in order to serve as a substrate for memory. Experimentally, synaptic plasticity exhibits phases covering the induction of long-term potentiation and depression (LTP/LTD) during the early phase of synaptic plasticity, the setting of synaptic tags, a trigger process for protein synthesis, and a slow transition leading to synaptic consolidation during the late phase of synaptic plasticity. We present a mathematical model that describes these different phases of synaptic plasticity. The model explains a large body of experimental data on synaptic tagging and capture, cross-tagging, and the late phases of LTP and LTD. Moreover, the model accounts for the dependence of LTP and LTD induction on voltage and presynaptic stimulation frequency. The stabilization of potentiated synapses during the transition from early to late LTP occurs by protein synthesis dynamics that are shared by groups of synapses. The functional consequence of this shared process is that previously stabilized patterns of strong or weak synapses onto the same postsynaptic neuron are well protected against later changes induced by LTP/LTD protocols at individual synapses.

Introduction

Changes in the connection strength between neurons in response to appropriate stimulation are thought to be the physiological basis for learning and memory formation [1,2]. A minimal requirement for proper memory function is that these changes, once they are induced, persist for a long time. For several decades, experimentalists have therefore focused on Long-Term Potentiation (LTP) and Long-Term Depression (LTD) of synapses in hippocampus [3,4] and cortical areas [5,6]. LTP can be induced at groups of synapses by strong 'tetanic' high-frequency stimulation of the presynaptic pathway [3] while stimulation at lower frequency leads to LTD Dudek92. Both LTP and LTD can also be induced at a single synapse or a small number of synaptic contacts if presynaptic activity is paired with either a depolarization of the postsynaptic membrane [5,7] or tightly timed postsynaptic spikes [8,9].

While the induction protocol for LTP and LTD is often as short as a few seconds, the changes in synaptic efficacy persist for much longer [9]. In typical slice experiments on LTP [and similarly for LTD or Spike-Timing Dependent Plasticity (STDP)] the persistence of the change is monitored for 30 minutes to 1 hour. Accumulating evidence suggests, however, that after this early phase of LTP (E-LTP) different biochemical processes set in that are necessary for the further maintenance of potentiated synapses during the late phase of LTP (L-LTP) [10,11]. For an understanding of the transition from early to late LTP, the concept of 'synaptic tagging and capture' has become influential [12,13]. During induction of the early phase of LTP, each potentiated synapse sets a tag that marks that it has received a specific afferent signal. A candidate molecule, involved in the tag signaling LTP induction in apical dendrites of hippocampal neurons, is the calcium-calmodulin dependent kinase II (CaMKII) [13]. Newly synthesized plasticity-related proteins are 'captured' by the tagged synapse and transform E-LTP into L-LTP that can be maintained over hours or days. A candidate protein involved in the maintenance of potentiated hippocampal synapses is the protein kinase Mζ (PKMζ) [11,14].

The stabilization and maintenance of potentiated synapses poses a number of theoretical challenges. First, on the level of single synapses we must require synaptic strength to remain stable, despite the fact that AMPA channels in the postsynaptic membrane are continuously exchanged and recycled [15–17]. Thus the synapse is not 'frozen' but part of a dynamic loop. Second, on the level of neuronal representation in cortical areas, one finds representations of input features that are stable but at the same time sufficiently plastic to adjust to new situations [18]. In the theoretical community, this paradox has been termed the stability-plasticity dilemma in unsupervised learning [19]. Third, humans keep the ability to memorize events during adulthood, but can also remember earlier episodes years back. However, continued learning of new patterns in theoretical models of associative memory networks forces the erasure or 'overwriting' of old ones, the so-called palimpsest property [20,21]. In the context of

Author Summary

Humans and animals learn by changing the strength of connections between neurons, a phenomenon called synaptic plasticity. These changes can be induced by rather short stimuli (lasting sometimes only a few seconds) but should then be stable for months or years in order to be useful for long-term memory. Experimentalists have shown that synapses undergo a sequence of steps that transforms the rapid change during the early phase of synaptic plasticity into a stable memory trace in the late phase. In this paper we introduce a model with a small number of equations that can describe the phenomena of induction of synaptic changes during the early phase of synaptic plasticity, the trigger process for protein synthesis, and the final stabilization. The model covers a broad range of experimental phenomena known as tagging experiments and makes testable predictions. The ability to model the stabilization of synapses is crucial to understand learning and memory processes in animals and humans and a necessary ingredient for any large-scale model of the brain.

continued learning, theoretical arguments show that synaptic plasticity on multiple time scales cannot prevent, but at most delay the erasure of memories in the presence of ongoing synaptic activity [22]. This suggests that additional mechanisms are necessary to further protect existing memories and 'gate' the learning of new ones.

Despite these challenges for the long-term stability of synapses, most classical models of synaptic plasticity focus on the induction and early phase of LTP or LTD and completely ignore the question of maintenance. Traditional models of associative memories separate the learning phase from the retrieval phase [23] and the same holds for standard models of STDP [24–26]. Detailed biophysical models of LTP and LTD describe calcium dynamics and Calcium/Calmodulin-Dependent Protein Kinase II (CaMKII) phosphorylation during the induction and early phase of LTP [27–29]. While these models show that switches built of CaMKII proteins can be stable for years, they do not address aspects of tagging leading to heterosynaptic interaction during L-LTP and L-LTD. Moreover, while CaMKII phosphorylation is necessary for induction of LTP and mediate tags in the apical dendrites of hippocampal CA1 neurons [30], it is less clear whether it is necessary for its maintenance [31]. On the other hand protein kinase Mζ is essential for maintenance of some synapse types [11,13,14] but the same molecule is potentially relevant for *induction* in others [30].

We wondered whether a simple model that connects the process of LTP induction with that of maintenance would account for experimental results on tagging and 'cross-tagging' [11–13,32] without specific assumptions about the (partially unknown) molecular pathways involved in the maintenance process. If so, the model should allow us to discuss functional consequences that are generic to the tagging hypothesis independent of the details of a biophysical implementation in the cell. Even though we believe that the model principles are more general, we focus on synapses from the Schaffer-Collaterals onto the CA1 neurons in hippocampus as an experimentally well-studied reference system for synaptic plasticity. Since typical tagging experiments involve the extracellular stimulation of one or several *groups* of synapses (rather than single synapses), our model of early and late LTP/LTD is developed in the context of a neuron model with hundreds of synapses. The application of the principles of synaptic consolida-

tion to experiments inducing E-LTP/E-LTD at *single* synapses is considered in the discussion section.

Results

We study a model with a large number of synapses i onto a single postsynaptic neuron. To be specific, we think of a pyramidal neuron in the CA1 area of hippocampus. Our model combines features of traditional models for the *induction* of potentiation [24–26,33–36] with a simple description of tagging and synthesis of plasticity related proteins that finally lead to the *maintenance* of the induced changes. The section is organized as follows: We first introduce the essential components of the model step by step ('Constructing the Model'). We then test the performance of the model with a set of stimuli typically used to induce long-term changes of synapses ('Testing the Model').

Constructing the Model

Our model contains three elements, Figure 1. The first one sets the *tag* during the induction of E-LTP or E-LTD. A tag is indicated by a value $h = 1$ for LTP or $l = 1$ for LTD. In the absence of tags we have $h = l = 0$. The second one describes the process that *triggers* the synthesis of plasticity related proteins. The final component describes the up-regulation of a maintenance-related process from a low value ($z = 0$) to a high value ($z \approx 1$). The dynamics of this component is intrinsically bistable and leads to a *consolidation* of the previously induced change at the labeled synapses upon interaction with the protein p ('protein capture'). The total change Δw of the synaptic strength reported in experiments contains contributions [13] of the early components l and h as well as the late component z. Since the model describes a sequence of three steps 'Tag-Trigger-Consolidation' we call it in the following the TagTriC-Model (Figure 1).

Tag and Induction of LTP/LTD

Results from minimal stimulation protocols which putatively activate only a single synapse suggest that the induction of LTP is a switch-like process [7,37]. We therefore model individual synapses as discrete quantities that can switch, during the induction of LTP, from an initial 'non-tagged state' (N) to a 'high state' (H) with a transition rate ρ_H that depends on the induction protocol. Similarly, induction of LTD moves the synapse from the initial non-tagged state (N) to a 'low state' (L) at a rate ρ_L. If synapse i is in the high state, the synaptic variable h_i is equal to one. If it is in the low state, another local variable l_i is set to one. These local variables h_i and l_i do not only control the weight of the synapse during E-LTP and E-LTD, but also serve as 'tags' for up- or down-regulation of the synapse. Tags reset to zero stochastically with a rate k_h and k_l, respectively. If both tags are zero, the synapse is in the non-tagged state N. Since the synapse is either up-regulated OR down-regulated, at most one of the tags can be non-zero (Figure 1A).

The stochastic transitions from the initial state N with $h_i = 0$ and $l_i = 0$ to the down-regulated state $l_i = 1$ or an upregulated state $h_i = 1$ depend in a Hebbian manner on presynaptic activity and the state of the postsynaptic neuron. In the absence of presynaptic activity, the LTD rate ρ_L vanishes. Presynaptic activity combined with a time-averaged membrane potential $\bar{u}$ above a critical value ϑ_{LTD} leads in the TagTriC model to a LTD transition rate ρ_L proportional to $[\bar{u}(t) - \vartheta_{\mathrm{LTD}}]$. For a transition from the initial state to the high state, we require in addition that the *momentary* membrane potential is above a second threshold ϑ_{LTP}. Hence the transition rate ρ_H is proportional to $[\bar{u}(t) - \vartheta_{\mathrm{LTD}}][u - \vartheta_{\mathrm{LTP}}]$
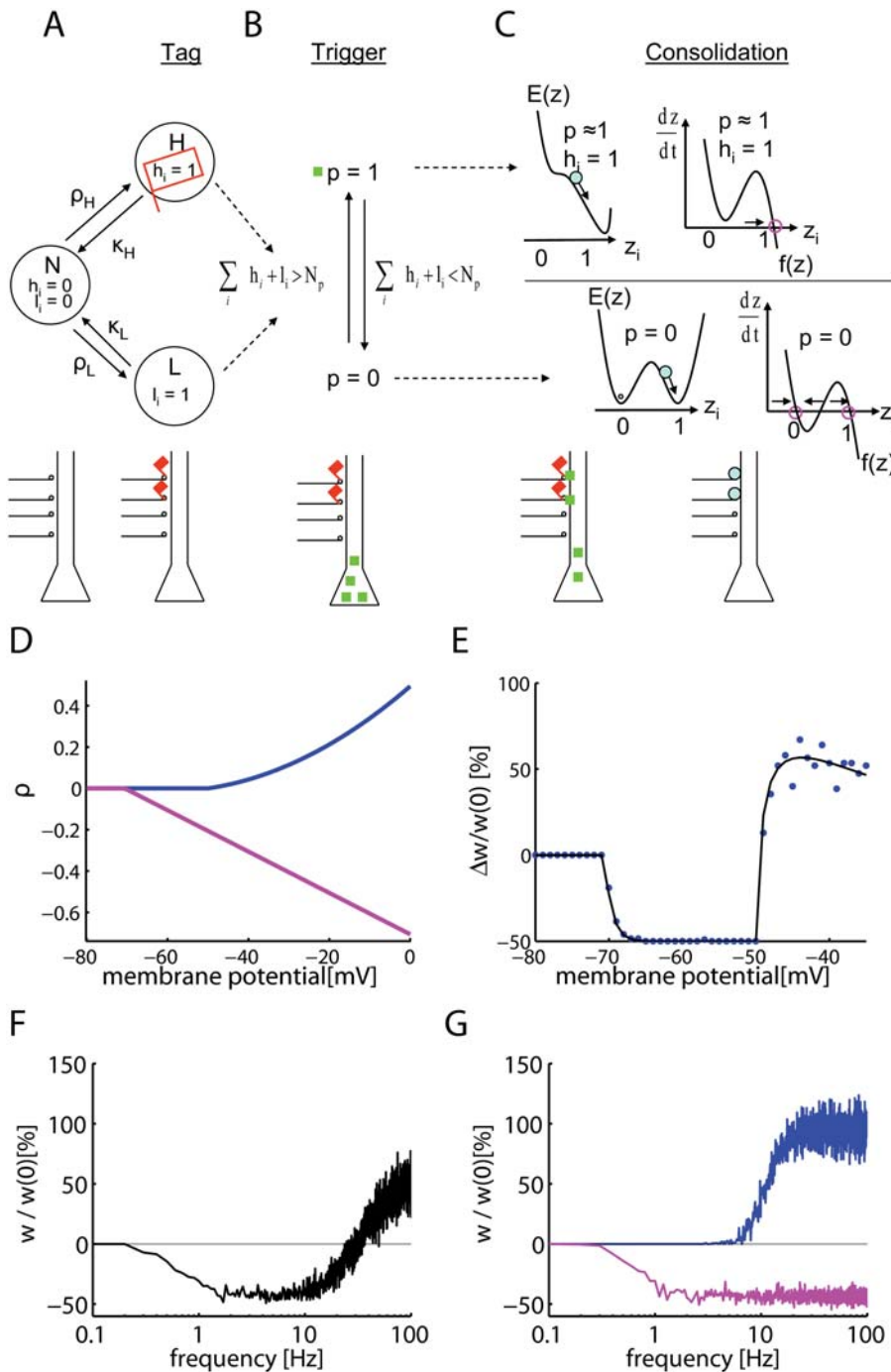
Figure 1. The three components of the Tag-Trigger-Consolidation (TagTriC) model. (A) A synapse can be in the non-tagged state N, the high state H or the low state L. A synapse i in H (or L) has a tag $h_i = 1$ (or $l_i = 1$, respectively). Transitions to a tagged state occur with rates ρ_H for potentiation and ρ_L for depression. The tag $h_i = 1$ is indicated by a red flag in both the flow graph and the schematic drawing below. (B) Synthesis of plasticity related proteins p (green squares) is triggered if the total number of set tags is larger than a critical number N_p. If the trigger threshold N_p is not reached, the protein concentration decays back to zero. (C) The consolidation dynamics can be visualized as downward motion in a potential surface $E(z)$. The function $f(z)$ (shown to the right) is the derivative of E and characterizes the dynamics $dz/dt = f(z)$. If a tag is set at the synapse ($h_i = 1$) and protein synthesis has been triggered ($p \approx 1$), the dynamics can be imagined as downward motion into the right well of the potential $E(z)$. In this case, $z \approx 1$ is the only fixed point of the dynamics (magenta circle). In the absence of tags ($h_i = l_i = 0$, below) the consolidation variable z_i of synapse i is bistable and approaches (direction of flow indicated by arrows) stable fixed points at $z_i = 0$ or $z_i = 1$ (magenta circles). The steps of synaptic tagging and capture are indicated immediately below the flow diagram. (D) The tagging rates for depression ($-\rho_L$(magenta)) and for potentiation ρ_H (blue) are shown as a function of the clamped voltage under the assumption that a presynaptic spike has arrived less than 1 millisecond before. Note that for depression we plot the negative rate $-\rho_L$ rather than ρ_L to emphasize the fact that depression leads to a down-scaling of the synapse. (E) Voltage dependence of early LTP and LTD. The weight change $\Delta w/w(0)$ induced by a stimulation of 100 synapses at 2 Hz during 50 s while the postsynaptic voltage is clamped is shown as a function of voltage. The percent change $\Delta w/\hat{w}$ in simulations (circles) of LTP/LTD induction experiments can be predicted from a theory (solid line) based on the difference in transition rates $\rho_H - \rho_L$. The simulation reflects the voltage dependence seen in experiments [5,39]. (F,G) Frequency dependence of early LTP and LTD. Simultaneous stimulation of 100 synapses by 3 trains (separated by 5 min) of 100 pulses at rates ranging 0.03 to 100 Hz shows LTD at low frequencies and LTP at frequencies above 30 Hz. (G) If LTP is blocked in the model, LTD (pink line) occurs up to high frequencies as in experiments [7]. Blue line: LTP with blocked of LTD.
doi:10.1371/journal.pcbi.1000248.g001

whenever these threshold conditions are satisfied; see Methods for details.

Our assumptions regarding the transition rates essentially summarize the qualitative voltage dependence seen in the Artola-Bröcher-Singer experiments [5]. Indeed, when 100 synapses in the TagTriC model are stimulated at low frequency during 50 seconds while the membrane voltage is kept fixed at different values (Figure 1D), the total weight change summed across all synapses exhibits LTD at low voltage and LTP at high voltage [38,39]. As expected, the resulting weight changes in the simulations of Figure 1E reflect the voltage dependence of the transition rates in Figure 1D.

Trigger for Protein Synthesis

Previously induced LTP or LTD needs to be consolidated in order to last for more than one hour. Consolidation requires that protein synthesis is triggered. Experimental evidence indicates that triggering of protein synthesis needs the presence of neuromodulators such as dopamine (in the apical CA1 region) or other modulators (in other regions). In typical tagging experiments, extracellular stimulation co-stimulates dopaminergic input leading to a phasic dopamine signal [13,40]. In our model, induction of E-LTP or E-LTD through appropriate stimulation protocols changes the synaptic efficacy and sets tags at the modified synapses, both described by the variables $h_i = 1$ or $l_i = 1$. Protein synthesis in the model is triggered (see methods for details) if the total number of tags $\Sigma_i(h_i + l_i)$ (which indirectly reflects the phasic dopamine signal) reaches a threshold $\mathcal{N}_p$ which depends on the level of background dopamine (and other neuromodulators). More specifically, $\mathcal{N}_p$ decreases with the concentration of background dopamine so that the presence of dopamine facilitates the trigger process [32].

If the trigger criterion is satisfied, the concentration p of synthesized plasticity related proteins approaches with rate k_p a value close to one. If the number of tags falls below the threshold $\mathcal{N}_p$, the protein concentration p decays with a time constant τ_p back to zero. Further details on the role of the trigger threshold and its relation to neuromodulators can be found in the discussion section.

Consolidation and Late LTP

The total weight w_i of a synapse i depends on the present value of the tags h_i or l_i as well as on its long-term value z_i. The slow variable z_i is a continuous variable with one or two stable states described by a generic model of bistable switches, that could be implemented by suitable auto-catalytic processes [16]. While the concentration p of plasticity related proteins is zero, the variable z_i has two stable states at $z_i = 0$ and $z_i = 1$, respectively. If the protein concentration takes a value of $p \approx 1$, one of the stable states disappears and, depending on the tag that was set, the long term-value of the synapse can be up- or down-regulated; see methods and Figure 1C for details.

In order to illustrate the mechanism of induction of L-LTP, let us suppose that the synapse has been initially close to the state $z_i = 0$. The dynamics of the synapse can be imagined as downward motion in a 'potential' E. The current stable state of the synapse is at the bottom of the left well in the potential pictured in Figure 1C. We assume that during a subsequent LTP induction protocol the synapse has been tagged with $h_i = 1$ and that the total number of tags set during the LTP induction protocol surpasses the trigger threshold $\mathcal{N}_p$. If the protein concentration p approaches one, the potential surface is tilted so that the synapse now moves towards the remaining minimum at $z \approx 1$. After decay of the tags, p returns to zero, and we are back to the original potential, but now with the synapse trapped in the state $z = 1$. It can be maintained in this state for a long time, until another strong tagging event occurs during

which the synapse is tagged with $l_i = 1$ as a result of LTD induction. In this case the potential surface can be tilted towards the left so that the only equilibrium point is at $z = 0$. Since consolidation is typically studied in animals that are more than 20 days old [13], we assume that before the beginning of the experiment 30 percent of the synapses are already in the upregulated state $z = 1$ and the remaining 70 percent in the state $z = 0$; see also [7]. Because of the bistable dynamics of consolidation, only synapses that are initially in the upregulated state $z = 1$ can undergo L-LTD and only synapses that start from $z = 0$ can undergo L-LTP; compare [7]. Note, however, that tags for potentiation and depression can be set independently of the value of z. We may speculate that the variable z is related to the activity of PKMζ [11,14], or to the self-sustained clustering of AMPA receptors [41], but the exact biochemical signaling chain is irrelevant for the functional consequences of the model discussed in the results section. In our model, the bistable dynamics of the z-variable captures the essence of synaptic persistence despite molecular turnover [15,16,28] and mobility of AMPA receptors [41].

Tests of the Model

The TagTriC model has been tested on a series of stimulation protocols that reflect induction of LTP and LTD as well as the consolidation of plasticity events.

Induction of Synaptic Changes

A typical LTP induction experiment starts with extracellular stimulation of a bundle of presynaptic fibers (i.e., the Schaffer collaterals leading from CA3 to CA1) that activate a large number (typically hundreds [13]) of presynaptic terminals. With an extracellular probe electrode placed close to one of the postsynaptic neurons, a change in synaptic efficacy is measured via the amplitude (or initial slope) of the evoked postsynaptic potential, representing the total response summed across all the stimulated synapses. In our simulations, we mimic these experiments by simultaneous stimulation of 100 synapses. The state of the postsynaptic neuron is described by the adaptive exponential integrate-and-fire model [42] and can be manipulated by current injection.

In a preliminary set of simulation experiments done with presynaptic stimulation alone (no manipulation of the postsynaptic neuron), the TagTriC model exhibits LTD or LTP depending on the frequency of the presynaptic stimulation (Figure 1F) in agreement with experimental results [4,43]. Moreover, under the assumption that LTP has been blocked pharmacologically ($\rho_H = 0$ in the model), our model shows LTD even for high stimulation frequencies (Figure 1G). This stems from the fact that LTD and LTP are represented in the TagTriC model by two independent pathways (Figure 1A) which are under control condition in competition with each other, but show up individually if one of the paths is blocked [43]. Together with the voltage dependence of Figure 1E, the above simulation results indicate that our model of LTP and LTD induction can account for a range of experiments on excitatory synapses in the hippocampal CA1 region, in particular, voltage and frequency dependence.

Consolidation of Synaptic Changes

In order to study whether consolidation of synaptic changes in our model follows the time course seen in experiments, we simulate standard experimental stimulation protocols [12,13]. A weak tetanus consisting of a stimulation of 100 synapses at 100 Hz for 0.2 seconds (21 pulses) leads in our model to the induction of LTP (change by +15 percent) which decays back to baseline over

the time course of two hours (Figure 2A). Thus, after the early phase of LTP the synapses are not consolidated. A stronger stimulus consisting of stimulating the same group of hundred synapses by 100 pulses at 100 Hz (repeated 3 times every 10 minutes) yields stronger LTP that consolidates and remains elevated (weight change by 22 ± 5 percent) for as long as the simulations are continued (more than 10 hours, only the first 5 hours are shown in Figure 2B). Thus our model exhibits a transition from early to late LTP if E-LTP is induced by the strong tetanic stimulation protocol, but not the weak one, consistent with results in experiments [12,13]. If, however, the weak tetanus at a first group of 100 synapses is given 30 minutes before or after a strong tetanus at a second group of 100 synapses, the synapses in both the weakly and strongly stimulated groups are consolidated (Figure 2C and 2D). If the weak tetanus in group one is given 120 minutes after the strong tetanus in group two, then consolidation of the synapses in the weakly stimulated group does not occur (Figure 2E). Thus our model exhibits a time course of heterosynaptic interaction between the two groups of synapses as reported in classical tagging experiments [12,13].

An advantage of a modeling approach is that we can study the dependence of the heterosynaptic interaction between the two groups of synapses upon model parameters. A critical parameter in the model is the trigger threshold $\mathcal{N}_p$ that needs to be reached in order to start protein synthesis (Figure 1B). With our standard choice of parameters, where $\mathcal{N}_p = 40$, we can plot the consolidated weight change $\Delta w/w(0)$ in the weakly stimulated group (measured 10 hours after the induction) as a function of the time difference between the stimulation of the group receiving the strong tetanus and that receiving the weak tetanus. The curve in Figure 2F shows that for a time difference up to 1 hour there is significant interaction between the two groups of synapses leading to synaptic consolidation, whereas for time differences beyond 2 hours this is no longer the case. If the trigger threshold is increased to $\mathcal{N}_p = 60$ (corresponding to less available neuromodulator), then the maximal time difference that still yields L-LTP in the weakly stimulated group of synapses is reduced to about 20 minutes (Figure 2F) whereas a reduction of $\mathcal{N}_p$ yields an increased time window of interaction (data not shown). If $\mathcal{N}_p$ is reduced much further, the weak tetanus alone will be sufficient to allow a transition from the early to the late phase of LTP. We speculate that $\mathcal{N}_p$ could depend on the age of the animal as well as on the background level of dopamine or other neuromodulators so as to enable a tuning of the degree of plasticity (see discussion for details).

LTD and Cross-Tagging

We consider two experimental protocols known to induce LTD—a weak low-frequency protocol consisting of 900 pulses at 1 Hz and a strong low-frequency protocol consisting of 900 repetitions at 1 Hz of a short burst of three pulses at 20 Hz. This strong low-frequency protocol applied to 100 model synapses leads to a significant level of LTD (reduction of weights to 70 ± 4 percent of initial value) which is consolidated 5 hours later at a level of 83 ± 3 percent of initial value. If a group of 100 synapses is stimulated with the weak low-frequency protocol, an early phase of LTD is induced that is not consolidated but decays over the time course of 3 hours (Figure 3A and 3B). However, if the weak low-frequency stimulation occurs after another group of 100 synapses had been stimulated by the strong low-frequency protocol, then the group that has received the weak stimulation shows consolidated synapses (at 90 ± 2 percent 5 hours after stimulus induction, Figure 3C). Moreover, consolidation of LTD (at 92 ± 3 percent 5 hours after stimulus induction) in the group of synapses

receiving the weak low-frequency protocol also occurs if it was stimulated thirty minutes after the stimulation of a second group of synapses by a strong tetanus, leading to LTP (Figure 3D). Thus, the TagTriC model exhibits cross-tagging consistent with experiments [11,32]. In our model, cross-tagging occurs because the tags for LTP and LTD (h_i and l_i, respectively) enter in a symmetric fashion into the trigger criterion for the synthesis of plasticity-related proteins (see Figure 1 and Methods).

Model Mechanism for Tagging, Cross-Tagging, and Consolidation

In order to elucidate how the model gives rise to the series of results discussed in the preceding paragraphs, we have analyzed the evolution of the model variables during and after induction of LTP (Figure 4). Critical for consolidation is the synthesis of plasticity related proteins, characterized by the variable p in the model. Synthesis is only possible while the total number of tags $\sum_i^N h_i + l_i$ is above the protein triggering threshold $\mathcal{N}_p$. For the strong tetanic stimulus this criterion is met for about 90 minutes (shaded region in Figure 4A) leading to high levels of plasticity related proteins. After 90 minutes the concentration of proteins starts to decay back to baseline. While the level of proteins is sufficiently elevated the consolidation variable z_i of each tagged synapse moves towards $z_i \approx 1$ since this is the only stable fixed point of the dynamics (Figure 1C). This leads to a consolidation time of about 2 hours, enough to switch a large fraction of synapses into the up-regulated state $z \approx 1$ (green line, Figure 4A). Hence the average weight of the stimulated synapses stabilizes at a value above baseline, indicating L-LTP (Figure 4A, solid line).

If, in a different experiment, 100 synapses are stimulated by the weak tetanus, the synthesis of plasticity related proteins is only possible during a few minutes (Figure 4B, red line), which is not sufficient to switch tagged synapses from $z = 0$ into the upregulated state $z \approx 1$. Hence the weights (Figure 4B, black line) decay together with the tags (Figure 4B, magenta line) back to baseline and the transition from early to late LTP does not occur. The decay of the weights is controlled by the rate k_H at which tags stochastically return to zero. The evolution of the protein concentration p and the consolidation variable z after a strong tetanus that leads to 90 minutes of protein synthesis and a weaker tetanus that only leads to 40 minutes of protein synthesis has been illustrated in (Figure 5A).

The total amount of available protein that is synthesized depends in our model on the time that the total number of tags stays above the protein triggering threshold $\mathcal{N}_p$. Even though always 100 synapses are stimulated in our model, not all receive tags in each experiment; moreover because of the competition for potentiation tags ($h_i = 1$) and depression tags ($l_i = 1$) during induction of plasticity, different synapses can receive different tags in the same experiment. With our strong tetanus protocol, on average 70 (out of 100) synapses receive a potentiation tag and 30 a depression tag while with the weak tetanus the numbers are 30 and 10, respectively. For the depression protocols, on average 10 synapses receive a potentiation tag and 90 a depression tag under strong low-frequency stimulation, and typically zero a potentiation tag and 40 a depression tag under the weak low-frequency protocol. These numbers vary from one trial to the next so that sometimes the protein trigger threshold $\mathcal{N}_p = 40$ is reached with the weak protocols and sometimes not. The important aspect is that even if the threshold is reached for a short time, the duration of protein synthesis is not long enough to provide a sufficient protein concentration p for consolidation of the tagged synapses; see Figure 4B and Figure 5A.
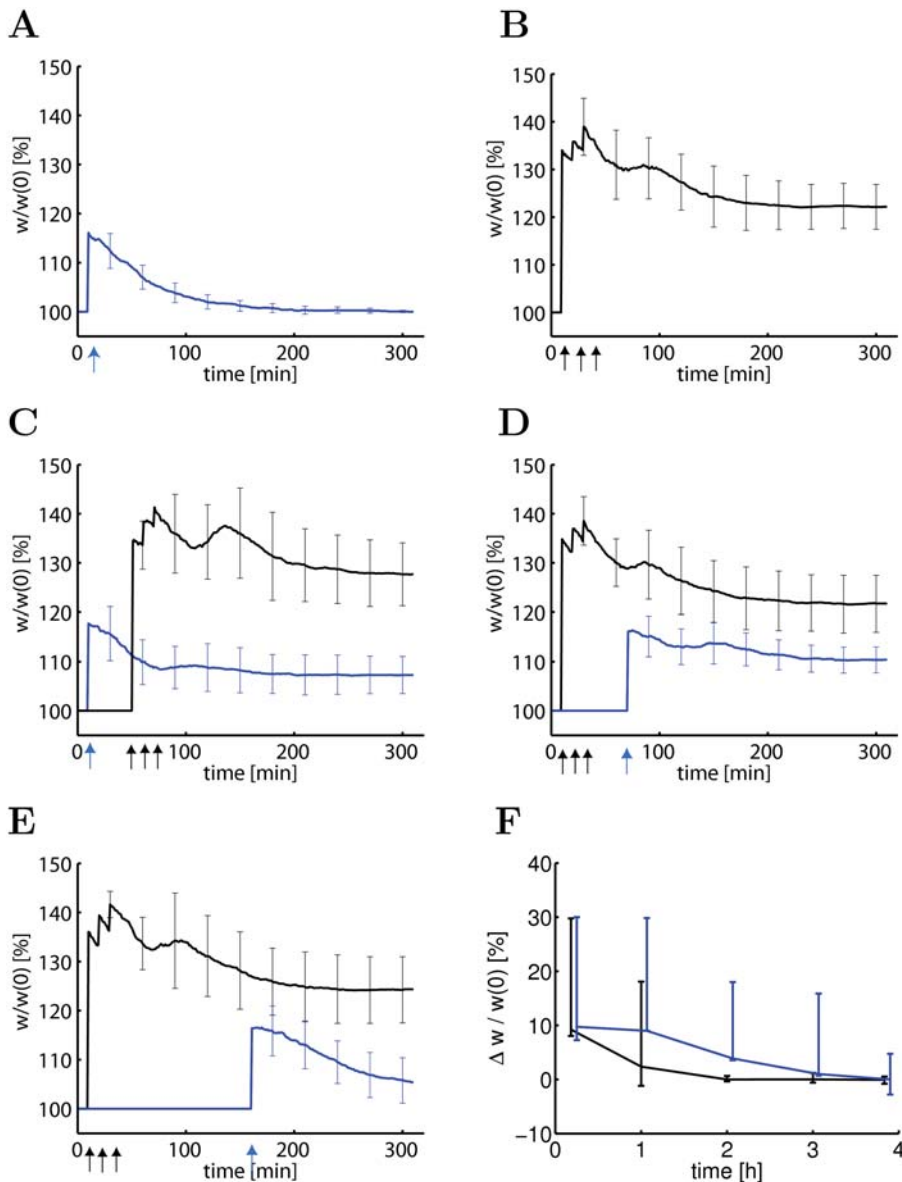
Figure 2. The model accounts for tagging paradigms. (A) A weak tetanus (21 pulses at 100 Hz) applied at a group of 100 synapses at $t = 10$ min (arrow) leads to an increased connection weight ($w/w(0)$, blue line) that decays back to baseline. (B) A strong tetanus (100 pulses at 100 Hz repeated three times, arrows) leads to late LTP that is sustained for 5 hours (black line). (C) If the weak tetanus (blue arrow) in a first group of synapses is followed thirty minutes later by a strong tetanus (black arrows) in a second group of synapses, the weights in the first group (blue line) and the second group (black line) are stabilized above baseline. (D) Stimulating a group of synapses by a weak tetanus (blue arrow) 30 minutes *after* the end of the strong tetanic stimulation of a second group also leads to stabilization of the weights in both groups above baseline. (E) If the weak tetanic stimulation occurs 2 hours after the strong tetanic stimulation of the other group, only synapses in the strongly stimulated group will be stabilized (black line), but not those in the weakly stimulated group (blue line). (F) Fraction of stabilized weights $\Delta w/w(0)$ in the weakly stimulated group measured 10 hours after induction of LTP as a function of the time difference between the weak stimulation and the end of the strong tetanic stimulation in the second group. Blue line: normal set of parameters ($N_p = 40$). Black line: protein trigger threshold increased to $N_p = 60$. In panels A–E, lines indicate the result averaged over 10 repetitions of the simulation experiments and bars standard deviation. In panel F, line indicates the result averaged over 100 repetitions. 90 of the 100 individual trials stayed within the bounds indicated by the error bars.
doi:10.1371/journal.pcbi.1000248.g002

Since the concentration p of plasticity related proteins is crucial for the transition from early to late LTP we wondered how a block of protein synthesis would interfere with the consolidation of weights in the TagTriC model. Application of a protein synthesis inhibitor (modeled by setting the rate k_p of protein synthesis to zero) during 1 hour starting thirty minutes before a strong tetanus is given to a group of 100 synapses that would normally lead to L-LTP, induced E-LTP but prevented consolidation into L-LTP

(data not shown). However, if the same simulation experiment was repeated after a second group of synapses had received a strong tetanic stimulation 35 minutes prior to the application of protein synthesis blocker, then both groups of synapses showed consolidation of weights (Figure 4D), consistent with experiments [12]. Closer inspection of the lower panel in Figure 4D shows that two components contribute to consolidation: Firstly, the concentration of plasticity related proteins (red line) that has increased because of
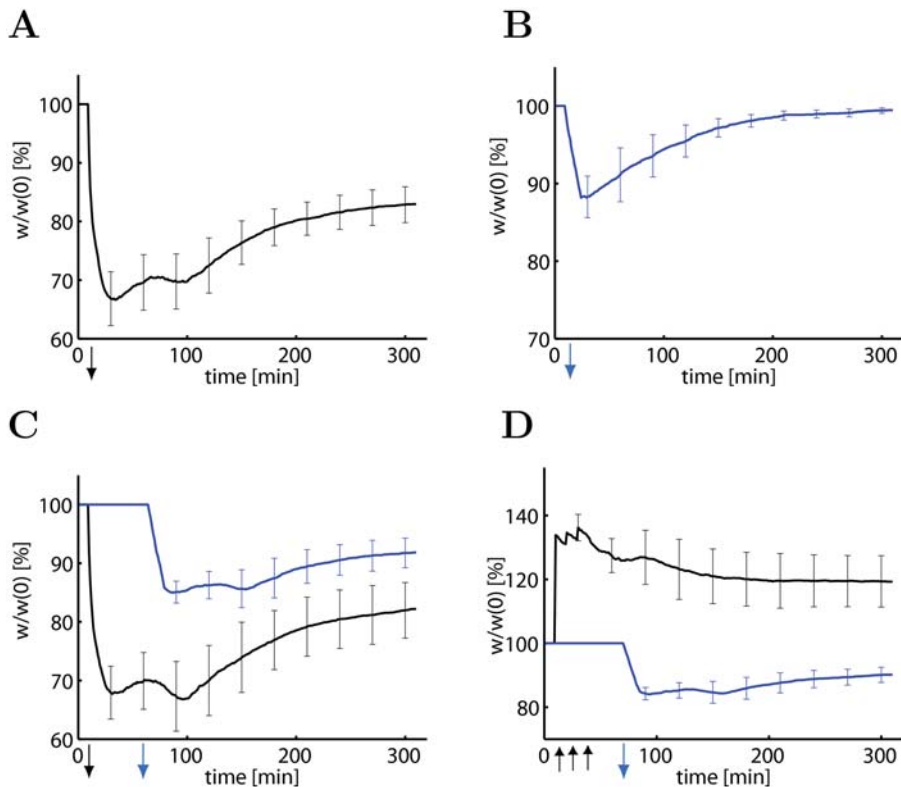
Figure 3. The model accounts for cross-tagging between LTP and LTD. (A) A strong low-frequency stimulus (3 pulses at 20 Hz, repeated 900 times every second) applied to a group of $N = 100$ synapses induces LTD with mean weights ($w/w(0)$) stabilized at $83 \pm 3\%$ of initial value after 5 hours (black line). (B) A weak low-frequency stimulus (1 pulse repeated 900 times at 1 Hz) induces early LTD, which is not consolidated. (C) If the weak low-frequency stimulus is applied 30 minutes after a second group of synapses has received the strong low-frequency protocol, the weights in both groups (blue, weak stimulus; black, strong stimulus) are consolidated at values below baseline. (D) Consolidation of LTD in the group receiving weak low-frequency stimulation (blue line) also happens if induction occurs 30 minutes after stimulating a second group of synapses with a strong tetanic protocol (see Figure 2) inducing LTP (black line). Downward arrows indicated the period of weak (blue arrow) or strong (black arrow) low-frequency protocols. The black upward arrows indicate strong tetanic stimulation. Lines show mean results, averaged over 10 repetitions of the simulation experiment. Error bars are standard deviation.
doi:10.1371/journal.pcbi.1000248.g003

the first strong tetanic stimulus decreases only slowly back to baseline enabling the switching of the slow components (variable z, green line) even in the presence of protein synthesis blocker. Secondly, even after the end of the application of the blocker, the total number of tags that has been set by LTP induction is still above the critical value $\mathcal{N}_p$ (shaded region in Figure 4D) so that protein synthesis can be resumed after the end of the blocking period. In summary, the detailed analysis of the TagTriC model allows to account for many aspects of tagging experiment in terms of a limited number of variables.

Discussion

Relation of Models to Experiments

Synaptic plasticity is based on intricate signal transduction chains involving numerous processing steps and a large number of different molecules [2,13,17]. Despite the complexity of the molecular processes, synaptic plasticity has experimentally been characterized by a small set of distinct phenomena such as short-term plasticity [44] as well as early and late phases of LTP and LTD [13].

Existing models of synaptic plasticity have focused on the description of short-term plasticity [44] and on the induction of LTP and LTD [24–26,33–36]. The question of maintenance has received much less attention and was mainly addressed in the

context of bistability of the CaMKII auto-phosphorylation process [27–29], AMPA receptor aggregation [41], or four identified kinase pathways [45]. While CaMKII is necessary for induction of long-term potentiation [46], it is probably too narrow to focus modeling studies only on a single or a few kinases such as CaMKII and neglect other proteins and signaling cascades that are involved in synaptic maintenance [13]. For example, there is strong evidence that PKMζ is involved in synaptic maintenance and necessary for the late phase of LTP in vitro [11] and in vivo [14]. However, the actual processes are complex and the molecules involved in setting tags may differ between different parts of the dendrite. For example PKMζ is involved in setting tags during E-LTP in the basal dendrite, whereas CaMKII (or MAPK for E-LTD) plays a similar role in apical dendrites [30].

Instead of focusing on specific signaling cascades, the TagTriC model presented in this papers aims at describing the essential ingredients of any possible functional model of L-LTP and tagging. These ingredients include (i) a bistable switch (described by the dynamics of the z_i-variable) for each synapse that guarantees long-term stability in the presence of molecular turn-over [16]; (ii) a global triggering signal for protein synthesis (described by the dynamics of the p variable); a formalism to (iii) induce early forms of LTP and LTD and (iv) set synaptic tags. Since we aimed for the simplest possible model, we have identified the synaptic tags h_i and l_i for potentiation and depression with the
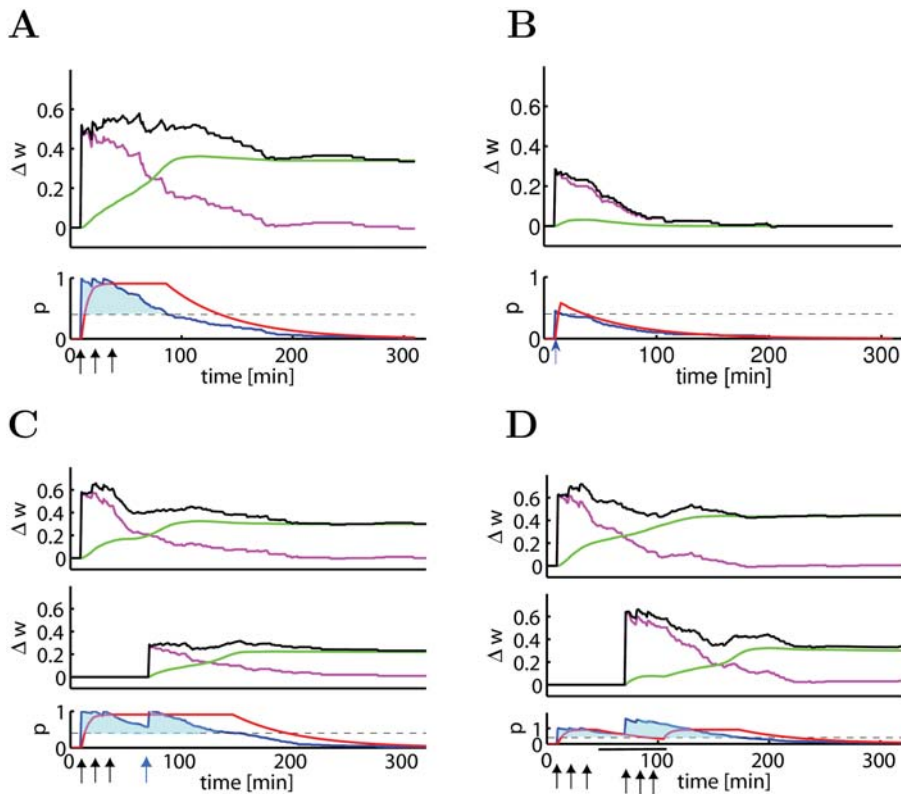
Figure 4. Dynamics of the TagTriC Model during different tagging protocols and protein synthesis blocking. The change of the total synaptic weight (top panels, black line $\Delta w = \sum_{i=1}^{N}[w_i(t)-w_i(0)/N]$) has contribution from early LTP (top panels, magenta line represents $\sum_{i=1}^{N}(h_i - \alpha l_i/N)$) and from late LTP (top panels, green line represents $\sum_{i=1}^{N}\beta(z_i - z_i(0))/N$). The protein variable p (red line, bottom panels) grows as long as the average number of tags ($\sum_{i=1}^{N}(h_i + l_i)/N$, blue line) is above the protein synthesis trigger threshold (N_p/N, dashed horizontal line). For better visibility, the regions where the blue line is above the trigger threshold is shaded. (A) A strong tetanus ($N=100$ synapses, stimulated by 100 pulses at 100 Hz, repeated three times every ten minutes) leads to a sustained period of about 90 minutes where the number of tagged synapses is above the protein synthesis triggering threshold (lower panel, blue shaded). During this time the protein synthesis variable p is close to one (red line, lower panel), causing an increase in the fraction of consolidated weights (green line, top panel). (B) During a weak tetanus ($N=100$ synapses, stimulated by 21 pulses at 100 Hz) the number of tags surpasses the protein triggering threshold only for a short time which does not enable switching of the z variable (top panel, green line) to the up-regulated state. (C) If the weak tetanus is given 30 minutes after the strong one, the number of tags set by the strong tetanus is still above the threshold, which allows protein synthesis stabilizing both the group of 100 synapses receiving the strong tetanus (top panel) and the group of 100 synapses receiving the weak tetanus (middle panel). (D) Protein synthesis is blocked for 1 hour (indicated by black bar at bottom of panel) starting 35 minutes after a first group of 100 synapses has been stimulated by a strong tetanus. Despite protein synthesis blocking, both the first group of synapses (top panel) and a second group of 100 synapses that received a strong tetanus during the blocking period (middle panel) develop late LTP because proteins synthesized during the induction of early LTP in the first group decay only slowly (bottom panel).
doi:10.1371/journal.pcbi.1000248.g004

synaptic weights during the early phase of LTP and LTD, respectively, so that points (iii) and (iv) are described by the same transition of the synapse from an initial non-tagged state to the high or low state, respectively. Variants of the model where the weight during the early phase of LTP and LTD is not directly proportional to the value of the tags are conceivable.

Even though we do not want to identify the synaptic variables h_i, l_i, z_i with specific biochemical signals, a couple of candidate molecules and signaling chains should be mentioned. The setting of the tag for LTP under normal physiological conditions involves NMDA receptor activation and elevated levels of calcium which in turn trigger a signaling chain involving Calmodulin and CaMKII. We therefore think that the h_i variable (representing both the tag for LTP induction and the weight increase during the early phase of LTP) should be related to the activation of CaMKII [13,46]. The molecular interpretation of the tag l_i for LTD is less clear [13]. In our model we have taken the tags as discrete quantities that decay stochastically, but a model with continuous tags that decrease exponentially gives qualitatively the same results (data not

shown). The reason is that triggering protein synthesis in our model requires a large number of tags to be set, so that even in the stochastic model only the *mean* number of tags is relevant–and the mean (more precisely, its expectation value) is a continuous variable. Nevertheless, we prefer the model with discrete values over the continuous one in view of the switch-like transitions of synapses after induction of LTP and LTD [7,37]. Maintenance of enhanced synaptic weights is probably implemented by an increased number of AMPA receptors in the postsynaptic membrane. Whether the stability arises from a self-organization process of receptors [41] or from interaction with persistently activated CaMKII molecules [46] or from additional kinases such as PKMζ [11,14], is an open problem of experimental investigation. Similarly, the exact identity of many plasticity related proteins is still unknown [13]. In our model we assume that recently synthesized plasticity related proteins are accessible to all synapses onto the same postsynaptic neuron. However, a distinction between proteins synthesized in, say, basal dendrites and that synthesized in apical dendrites would be possible by
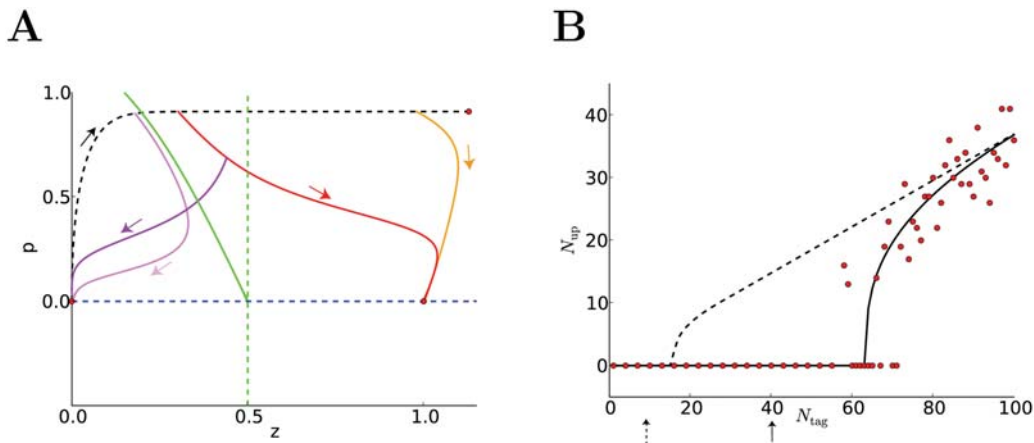
Figure 5. Theory and predictions. (A) Evolution of the variables p and z during tagging. If protein synthesis is 'ON' and the synapse tagged, p and z move along the black dashed line towards the stable fixed point on the upper right ($p \approx 1$, $z \approx 1$) (red filled circle). If protein synthesis stops after some time (yellow line, after 90 min; orange line, after 40 minutes) but the synapse remains tagged, the dynamics converges towards the fixed point $p = 0$, $z = 1$ (red filled circle) indicating that the synapse is consolidated (yellow and orange trajectories). However, if protein synthesis stops too early (after 25 min, pink line), or if the synaptic tag is lost too early (after 60 min, magenta line), the synapse is not consolidated and the trajectories converge towards the non-tagged initial state $p = 0$, $z = 0$ (red filled circle). The green dashed vertical line at $z = 0.5$ indicates the threshold beyond which a loss of the tag does not affect consolidation; the green solid line indicates the separatrix between the stable fixed points at $z = 0$ and $z = 1$. The minimal duration of protein synthesis to allow any consolidation is given by the intersection of the black dashed line with the separatrix. (B) Number of consolidated synapses (N_{up}, vertical axis) as a function of the number of initially tagged synapses (N_{tag}, horizontal axis) in simulations (red filled circles) and theory (solid line). Some of the initially tagged synapses fail to be consolidated because either they lose their tag or protein synthesis stops too early (see A). With a protein synthesis threshold $N_p = 40$ (arrow) we need about 60 initially tagged synapses to achieve any consolidation (solid line). If the protein synthesis threshold is reduced to $N_p = 10$ (dashed arrow), we need at least 15 tagged synapses to see any consolidation (dashed line).

doi:10.1371/journal.pcbi.1000248.g005

replacing the variable p by two or more distinct variables p_k with similar dynamics (but potentially different trigger thresholds N_p), allowing for a compartmentalization of tagging [13].

Experimental cross-tagging results clearly indicate that there are two different types of synaptic tags, one for LTP and one for LTD [13,32], which we called h_i for LTP and l_i for LTD, leading to three different states during tagging (Figure 1A). Since we have identified the tagging with the early phase of LTP and LTD, our model of E-LTP and E-LTD also has three different states (whereas our model of late LTP/LTD has only two states characterized by $z_i = 0$ and $z_- = 1$). The three-state model of early LTP/LTD presented in this paper would predict that all non-tagged synapses can undergo a transition to E-LTP or E-LTD depending on the induction protocol—whereas experiments suggest that about 70 percent of synapses show LTP but not LTD and the remaining 30 percent LTD but not LTP [7]. Moreover, only those synapses that are initially weak can be potentiated and only those that are initially strong can be depressed [7]. This aspect can be included in our model if we replace the induction rates ρ_H for LTP by $\rho_H(1-z_i)$ and ρ_L for LTD by $\rho_l z_i$ so LTP is only possible from a state with $z_i = 0$ and LTD only from an initial state $z_i = 1$ — in agreement with a two-state model of early LTP/LTD [7]. For the tagging and induction experiments presented in this paper, the results do not change significantly when we implement this extension of the induction model.

Functional Consequences and Predictions

One of the advantages of a simple phenomenological model is that it should be capable of illustrating the functional consequences of tagging and L-LTP or L-LTD in a transparent manner. What are these functional consequences?

A characteristic feature that is made transparent in our model (and which we expect to be present in any model of tagging) is

that, under typical experimental conditions, the transition from early to late LTP is only possible if a sizable group of synapses have undergone E-LTP or E-LTD. Hence, while induction of E-LTP is a local Hebbian process that is likely to take place at the postsynaptic site of the synapse (e.g., the dendritic spine), the transition from the early to the late phase of LTP requires a minimum number of synapses to be activated by appropriate stimulation including co-activation of neuromodulatory input so as to trigger synthesis of plasticity related proteins. A direct consequence of this is that synapses cannot be considered as independent. In order to predict whether a synapse memorizes an item for a long time or forgets it and re-learns some other item, it is not sufficient to consider a 'Hebbian' induction model, where synaptic changes depend only on the activity of pre- and postsynaptic neurons. For maintenance, it is not the synapse which decides individually, but it is the neuron as a whole (or a large functional compartment sharing the same site of synthesis of plasticity-related proteins [13,30,47]) which 'decides' whether it is going to store the present information, or not. Hence, classical [20,21,34] and recent [22] theoretical models which studied memory maintenance in the presence of ongoing neuronal activity on the level of *single* synapses need to be reconsidered, since the assumption of independent synapses does not hold (Figure 5A and 5B). In particular, our model predicts that, after an ensemble of identical neurons have received the same stimulus, some neurons learn (adapt a *large* fraction of their synapses to the stimulus) and others don't (keep all their synapses unchanged). With our choice of parameters, this happens in the TagTriC model if the number of synapses that have been tagged during the induction protocol is between 55 and 70 (Figure 5B). This neuronal, rather than synaptic, decision about memorizing an input (see also [48]) is potentially attractive for prototype learning–a standard paradigm in neuronal clustering and categorization algorithms, e.g., [19]. In contrast to traditional neuronal clustering models where learned

memories need to be protected against overwriting by completely different memory items [19], a model based on tagging would have an intrinsic vigilance threshold via the trigger threshold $\mathcal{N}_p$. Hence it is resistant to changes at a single synapse.

In our view, the protein synthesis trigger threshold $\mathcal{N}_P$ is an important control parameter in the model. The results of Figure 2F show that an increase of the trigger threshold reduces the maximal delay after which a weak tetanus leads to L-LTP after a strong tetanic stimulation in a different group of synapses. With our normal value of $\mathcal{N}_p = 40$ we need around 60 synapses to be initially tagged in order to retain any memory. If we decrease the trigger threshold to $\mathcal{N}_p = 10$ and keep all other parameters of the model unchanged, then we need at least a group of 15 synapses tagged during the induction protocol to get any consolidation since some of the initially tagged synapses loose their tag too early to get consolidated (Figure 5B). Only for a very small trigger threshold, say $\mathcal{N}_p = 1$, (which could occur at high concentration of neuromodulators) synapses become (nearly) independent, since a tag at a single synapse would be sufficient to trigger the synthesis of proteins which would then become available at that synapse. Repeated stimulation of the synapse alone would then be sufficient to transform E-LTP into L-LTP.

In our opinion, the trigger threshold $\mathcal{N}_p$ is significantly lower in the presence of neuromodulators such as, for example, dopamine (for synapses from Schaffer collaterals onto CA1 pyramidal neurons) or noradrenaline (for synapses in the dentate gyrus). A simple model for the dependence of $\mathcal{N}_p$ on dopamine would be $\mathcal{N}_p = n_0/(DA_{bg} + c_0)$ where n_0 is some arbitrary number (say $n_0 = 1$), c_0 a small number (say 0.001) and DA denotes the stationary 'background' concentration of dopamine (that is, before the start of the experiment), normalized to $0 < DA_{bg} < 1$. The phasic dopamine signal caused by co-stimulation of dopaminergic input during tagging experiments is assumed to be proportional to the number of tags $\sum_i^N h_i + l_i$. The trigger condition $\sum_i^N h_i + l_i > N_p$ becomes then equivalent to the condition $\left(\sum_i^N h_i + l_i \right)\left(DA_{bg} + c_0 \right) > n_0$ which shows a trade-off between the phasic dopamine signal and the stationary background level of dopamine. In particular in the presence of a large concentration of dopamine ($DA \approx 1$), single synapses can be consolidated. With the assumption that standard tagging experiments in a large group of synapses are performed at a low dopamine concentration of $DA = 0.024$ before stimulation, we retrieve the value of $\mathcal{N}_p = 40$ used in the main part of the results section. The dependence of the trigger criterion on the number of tags $\sum_i^N h_i + l_i$ takes implicitly the co-activation of neuromodulatory input during the experimental stimulation protocol into account: the larger the number of stimulated neurons and the stronger the stimulus, the higher the probability of co-activation of dopaminergic fibers. Blocking dopamine receptors amounts in the model to setting both the background and the phasic dopamine signal to zero. In this case, protein synthesis is not possible.

Our model of LTP/LTD induction does not only account for voltage and frequency dependence of LTP/LTD induction, but also for spike timing dependence. In fact, for a stimulation paradigm where postsynaptic spikes are induced by short current pulses of large amplitude either a few milliseconds before or after presynaptic spike arrival, the model of LTP/LTD induction used in the TagTriC model becomes formally equivalent to a recent model of spike-timing dependent plasticity [35] which can be seen as an extension of classical models of STDP [24–26]. In the case of stochastic spiking of pre- and postsynaptic neurons our model shares important features with the Bienenstock-Cooper-Munro model [33], in particular the quadratic dependence upon the postsynaptic variables. In addition, our model also accounts for the voltage dependence of the Artola-Bröcher-Singer model [38]. Thus, the model of LTP/LTD induction shares features with numerous established theoretical models and covers a large range of experimental paradigms known to induce LTP or LTD [3–6,8].

Since the subsequent steps of protein synthesis trigger and stabilization are independent of the way early phase of LTP is induced, our model predicts that tagging experiments repeated with different stimulation paradigms, but otherwise identical experimental preparation and age of animal, should give similar results as standard tagging protocols. In particular we propose to stimulate a group of synapses in hippocampal slices by 40–60 extracellular current pulses at 10 Hz while the postsynaptic neuron is receiving intracellular current injection that triggers action potential firing either a few milliseconds before or after presynaptic spike arrival and keeps the membrane potential at a depolarized level between postsynaptic action potential firing. Our model predicts that this will induce early LTD or LTP depending on spike timing and depolarization level that is not maintained beyond 1 or 2 hours. However, if the same stimulation occurs after a second group of synapses has received a strong tetanus, then stabilization of synapses at potentiated or depressed levels should occur, similar to standard tagging and cross-tagging experiments. In our opinion, these predictions should not depend on model details, but hold for a broad class of models that combine a mathematical description of induction of synaptic plasticity with a mechanism of consolidation.

Another finding—which is somewhat unexpected and in contrast to other conceptual models of synaptic tagging and capture [12,13,47]—is that during a strong tetanic stimulation a fraction of synapses receives tags for depression (while most, but not all, receive tags for potentiation). This is due to the fact that during induction of plasticity, transition to E-LTP and E-LTD act in parallel [7]. The prediction is that after consolidation (say 2 hours after the strong tetanic stimulation) a small fraction of synapses would show L-LTD, rather than L-LTP.

An essential ingredient of our model that allows long-term stability of consolidated synapses is the bistable dynamics of the variable z. In our opinion, such bistability (or possibly multistability [49] with three or four stable states) is necessary for synaptic maintenance in the presence of molecular turn-over, as recognized in earlier theoretical work [15,16,34]. Our model therefore predicts that L-LTP and L-LTD should have bistable, switch-like properties. While there is evidence for switch like transitions during the induction of E-LTP and E-LTD [7,37], the bistability of the late phase of synaptic plasticity has so far not been shown. A possible experiment would be to combine a minimal stimulation protocol (e.g., a weak tetanus) at a single synapse [7,37] with a medium to strong stimulus at a group of other synapses (e.g., tetanic stimulus varying between 30 and 100 pulses). The prediction is that the weight of the single synapse shows an all-or-none phenomenon with transition probabilities that depend on the stimulation of the group of other synapses. In particular, as the number of pulses of the tetanic stimulation is reduced (covering a continuum from strong to weak tetanic stimulation), the maintenance in the potentiated state should become less likely (averages across many experiments decrease) whereas the results of individual experiments show either full potentiation or none, which should give rise to a bimodal distribution of normalized synaptic weights.

Open Questions and Perspectives

A lot of questions remain open and need to be addressed in future studies. First, can a synapse that has been potentiated in the past and is maintained after a transition to late LTP undergo a

further potentiation step [13]? In our current model this is not possible since the consolidation variable z has only two stable fixed points. If we replace the function $f(z)$ depicted in Figure 1 by another one with more than two stable fixed points, then the answer to the above question would be positive. Indeed, there have been suggestions that self-organization of receptors into stable sub-groups could lead to multiple stable states [49].

Second, induction of LTP or LTD is not only possible by strong extracellular stimulation of groups of synapses, but also at single synapses if presynaptic activity is paired with either a depolarization of the postsynaptic membrane [5,7] or tightly timed postsynaptic spikes as in STDP experiments [6,8]. How can it be that the change induced by STDP seems to be maintained over one hour without visible degradation? [6,7]. Are synapses in these experiments consolidated, and if so what is the concentration of neuromodulators? In the TagTriC model with the choice of parameters used in the present paper, consolidation would not be possible, since the minimum number of synapses that have undergone E-LTP or LTD is $N_p = 40$ in order to trigger protein synthesis, but, as explained above, an increased neuromodulator concentration would make consolidation possible.

Third, what is the role of NMDA receptor activation during synaptic consolidation? In our present model, protein synthesis is triggered by appropriate induction protocols, but is independent of synaptic activity during the consolidation process. However, recent experimental results suggest that protein synthesis blocker needs synaptic stimulation during the consolidation period to become effective [50], suggesting a subtle interplay between protein synthesis and synaptic activation that cannot be captured by our model.

Fourth, has each neuron a single protein synthesis unit or is protein synthesis a local process confined to each dendritic branch? In the first case, there is a single neuron-wide protein synthesis trigger threshold [12] and the neuron as a whole 'decides' whether early forms of synaptic potentiation and depression will be consolidated or not. This is the paradigm posited in the TagTriC model. In the alternative model of local protein synthesis [13,47], the critical unit for consolidation are local groups of synapses on the same dendritic branch. Thus, for the same number of tagged synapses, a local group of synapses on the same dendritic branch is more likely to undergo consolidation than a distributed set of tagged synapses, leading to a form of clustered plasticity [47]. The TagTriC model can be easily adapted to the case of clustered plasticity by (i) replacing the point-neuron model by a neuron model with spatially distributed synapses and (ii) replacing the neuron-wide trigger equation (see 4 and Figure 1B) by a finite number of analogous, but dendrite-specific equations.

Fifth, how can tags be reset? Experiments show that a depotentiating stimulus given 5 minutes after a weak tetanus erases the trace of E-LTP (resets the tag) whereas depotentiation 10 or 15 minutes after the strong tetanus only transiently suppresses the E-LTP, making the consolidation of the synapse by protein capture possible [51]. We have checked in additional simulations that our present model cannot account for these experiments. In our opinion, the above tag-reset experiments show that the synapse has additional hidden states currently not included in the TagTriC model. Additional states would allow to (i) separate the measured early LTP during the first 5 minutes from setting the tag; and (ii) distinguish between depotentiation and depression of synapses. One interpretation of the tag-reset experiments [51] is that during the first five minutes the tag is not yet set whereas early LTP is already visible. The tag would be set

only with a delay of 5–10 minutes. Application of a depotentiating stimulus more than 10 minutes later would then leave the potentiation tag intact, but move the synapse to a transiently depotentiated state.

The final and potentially most interesting question is that of functional relevance: Can the TagTriC model be used to simulate reward-based learning in experiments in vivo [13]? The formal theory of reinforcement learning makes use of an eligibility trace [52] which can be interpreted as a synapse specific tag. In the future we want to check whether the TagTriC model can be linked to reinforcement learning models [53–56] under the assumption that reward prediction errors are represented by a dopamine signal [57] which influences the protein synthesis dynamics in our model. This open link to reward-based learning is of fundamental functional importance.

Methods

Model of Early LTP/LTD and Tagging

In our model we assume that presynaptic spike arrival needs to be combined with a depolarization of the postsynaptic membrane (e.g., [5]) in order to induce a change of the synapse. In voltage clamp experiments (e.g., [39]) the postsynaptic voltage would be constant. However, in general the voltage is time-dependent and described by a variable $u(t)$. In the TagTriC model, we assume that the low-pass-filtered voltage

$$\bar{u}(t) = \frac{1}{\tau_{\text{lowP}}} \int_0^\infty \exp\left(-\frac{s}{\tau_{\text{lowP}}}\right) u(t-s-\varepsilon)ds.$$

needs to be above a critical value ϑ_{LTD} to make a change of the synapse possible. τ_{lowP} is the time constant of the low-pass filter and $\varepsilon = 1$ ms is a short delay twice the width of a spike (see Table 1). This short delay ensures that $\bar{u}$ includes effects of previous presynaptic inputs and postsynaptic spikes, but not of an ongoing postsynaptic action potential.

Table 1. Parameter values used throughout all simulations, except Figure 1E–G where $N_p = 10$ and initial percentage of $z_i = 1$ was 10%, because these simulations refer to experiments with younger animals.

Tag	Trigger	Consolidation
$N = 100$	$k_p = 1/(6 \text{ min})$	$N = 100$
$A_{\text{LTD}} = 0.01$	$\tau_p = 60$ min	$\gamma = 0.1$
$A_{\text{LTP}} = 0.014$	$N_p = 40$	$\tau_z = 6$ min
$\tau_x = 100$ ms		$\beta = 2$
$\tau_{\text{lowP}}^{\text{LTP}} = 100$ ms		Initialisation: $N(z_i = 1) = 30$
$\tau_{\text{lowP}}^{\text{LTD}} = 1$ s		
$\varepsilon = 1$ ms		
$k_h = 1/h$		
$k_l = 1/(1.5 \text{ h})$		
$\Theta_{\text{LTD}} = -70.6$ mV		
$\Theta_{\text{LTP}} = -50$ mV		
$\alpha = 0.5$		
Initialisation: $l_i = h_i = 0$		

doi:10.1371/journal.pcbi.1000248.t001

Combining presynaptic spike arrival at synapse i (represented by x_i) with a depolarization $\bar{u}$ of the postsynaptic neuron above a threshold ϑ_{LTD} we get a rate of LTD

$$\rho_L = A_{\text{LTD}} x_i(t) [\bar{u}(t) - \vartheta_{\text{LTD}}]^+ \qquad (1)$$

where $A_{\text{LTD}} > 0$ is a parameter and $[.]^+$ denotes rectification, i.e., $[y]^+ = y$ if $y > 0$ and zero otherwise. Here $x_i(t) = \sum_f \delta\left(t - t_i^f\right)$ denotes the presynaptic spike train with pulses at time t_i^f and δ the Dirac-delta function. Formally, ρ_L describes the rate of stochastic transitions from the non-tagged state $h = 0$, $l = 0$ to the low state $l = 1$, Figure 1. In simulations we work with discrete time steps of $\Delta = 1$ ms. Eq. 1 indicates that the probability $P_{l=0 \to l=1}$ of a transition to the low-state during the time step Δ vanishes in the absence of presynaptic spike arrival and takes a value of $P_{l=0 \to l=1} = 1 - \exp(-A_{\text{LTD}}[\bar{u}(t) - \vartheta_{\text{LTD}}]^+ \Delta) \approx A_{\text{LTD}}[\bar{u}(t) - \vartheta_{\text{LTD}}]^+ \Delta$ if a presynaptic spike arrives at the synapse i during the time step Δ. Note that the transition from $l = 0$ to $l = 1$ is only possible if $h = 0$ and h remains zero during the transition.

Similarly, a switch from the non-tagged state $h = 0$, $l = 0$ to the high state $h = 1$ occurs at a rate ρ_H which also depends on postsynaptic voltage and presynaptic spike arrival. We assume that each presynaptic spike at synapse i leaves a trace $\bar{x}_i$ that decays exponentially with time constant τ_x. The exact biophysical nature of the trace is irrelevant, but could, for example, represent the amount of glutamate bound to the postsynaptic receptor. The value of the trace at time t caused by earlier spike arrivals at time t_i^f is then $\bar{x}_i(t) = (1/\tau_x) \sum_f \exp\left[-\left(t - t_i^f\right)/\tau_x\right]$ where the sum runs over all firing times $t_i^f < t$. With the trace $\bar{x}_i$ we write

$$\rho_H = A_{\text{LTP}} \bar{x}_i(t) [\bar{u}(t) - \vartheta_{\text{LTD}}]^+ [u(t) - \vartheta_{\text{LTP}}]^+ \qquad (2)$$

which indicates that, in addition to the conditions for LTD induction we also require the *momentary* membrane potential $u(t)$ to be above a second threshold ϑ_{LTP}. This threshold could change on the time scale of minutes or hours as a function of homeostatic processes. To summarize, the rate of LTP transition ρ_H is different from ρ_L in five aspects. First, the constant A_{LTP} is not the same as A_{LTD}. Second, LTP is caused by the *trace* $\bar{x}_i$ left by presynaptic spikes, rather than the spikes themselves. This trace-formulation ensures that presynaptic spikes can interact with later postsynaptic spikes as in classical models of STDP [24–26]. Third, the time constant of the low-pass filter in $\bar{u}$ is different; fourth, the momentary voltage needs to be above a threshold ϑ_{LTP}; and fifth, the total dependence upon the postsynaptic voltage is quadratic, rather than linear. The quadratic dependence ensures that for large depolarization LTP dominates over LTD [39]. Tagged synapses with $h_i = 1$ decay with probability $P_{h=1 \to h=0} = k_H \Delta$ back to the non-tagged state (and analogously, but with rate k_L for the transition $l_i = 1 \to l_i = 0$).

In the TagTriC model, the local synaptic values $h = 1$ for potentiation or $l = 1$ for depression act as tags indicating potential sites for further consolidation, but are also directly proportional to the weight of the synapse after induction of LTP or LTD. Since in minimal stimulation experiments LTD leads to a reduction of about 50 percent of the synaptic efficacy whereas LTP leads to an increase by up to 100 percent [7], we model the weight change during the early phase of LTP as $\Delta w_i = (h_i - \alpha l_i)\hat{w}$ where $\hat{w}$ is the weight of the non-tagged synapse and $\alpha = 0.5$. The total weight change $\Delta w / \hat{w}$ measured shortly after induction of LTP or LTD with extracellular protocols corresponds to the fraction of synapses in the high or low states, respectively, hence, if all synapses start

from the non-tagged state the measured weight change is $\Delta w / \hat{w} = \sum_{i=1}^{N} (h_i - \alpha l_i)/N = \langle h \rangle - \alpha \langle l \rangle$ where N is the number of synapses stimulated by the protocol. The set of parameters of LTP/LTD induction and tagging is given in table 1.

Trigger

The triggering process is controlled by the dynamics of a variable p which describes the amount of plasticity related proteins synthesized in the postsynaptic neuron. Protein synthesis is triggered and the variable p increases while the concentration of dopamine exceeds a critical level ϑ_p [58]. If the dopamine concentration DA falls below ϑ_p, the protein concentration decays with a time constant τ_p. Assuming standard first-order kinetics we have

$$\frac{dp}{dt} = k_p (1-p) \Theta[\text{DA} - \vartheta_p] - \frac{p}{\tau_p} \qquad (3)$$

Protein synthesis has a maximum rate dp/dt of k_p and saturates if the amount of protein approaches a value one. $\Theta[y]$ denotes the unit step function with $\Theta[y] = 1$ for $y > 0$ and zero otherwise.

Dopamine is present at a low stationary background value. In addition a phasic dopamine component is induced in standard tagging experiments in hippocampal slices, because of co-stimulation of dopaminergic inputs during extracellular stimulation of presynaptic fibers [40]. To describe the time course of the phasic dopamine component in our model, we assume that the dopamine is proportional to the total number of tags $\sum_i (h_i + l_i)$ induced by the stimulation protocol. The stationary background level of dopamine DA_{bg} is included in the threshold $\vartheta_p = N_p(\text{DA}_{\text{bg}})$ for protein synthesis. Hence Eq. 3 can be rewritten in the form

$$\frac{dp}{dt} = k_p (1-p) \Theta\left[\sum_i (h_i + l_i) - N_p(\text{DA}_{\text{bg}})\right] - \frac{p}{\tau_p} \qquad (4)$$

Note that we have chosen units so that the threshold for protein synthesis N_p can be interpreted as the minimal number of tags necessary to stimulate protein synthesis. This interpretation is important for the discussion of the model results, in particular Figures 4 and 5.

A suitable model for dependence of the protein synthesis threshold on the background level of dopamine is $N_p(\text{DA}_{\text{bg}}) = n_0/(\text{DA}_{\text{bg}} + c_0)$ where $n_0 = 1$ is a scaling factor, $c_0 = 0.001$ a constant and $0 \leq \text{DA}_{\text{bg}} \leq 1$ is the normalized dopamine concentration. We note that the trigger condition $[\sum_i (h_i + l_i) - N_p(\text{DA}_{\text{bg}})] > 0$ is then equivalent to the condition $(\text{DA}_{\text{bg}} + 0.001)[\sum_i (h_i + l_i)] > 1$. This formulation shows that there is a trade-off between background levels and phasic dopamine. Unless stated otherwise we always use in the simulation a fixed dopamine level $\text{DA}_{\text{bg}} = 0.024$ so that $N_p = 40$. The specific model $N_p(\text{DA}_{\text{bg}})$ of the dependence upon background dopamine levels is therefore irrelevant.

We assume that the plasticity related protein p synthesized in the postsynaptic neuron is diffused in the dendrite of the postsynaptic neuron and hence available to all the synapses under consideration. Hence, the tags h_i and l_i have indices, since they are synapse-specific, whereas p in Eq. 4 does not.

Consolidation and Late LTP

The consolidation variable z describes the late phase of LTP and follows the dynamics

$$\tau_z \frac{dz_i}{dt} = f(z_i) + \gamma(\text{DA})(h_i - l_i)p. \qquad (5)$$

The scaling factor γ is a function of the dopamine level DA. In the simulations we always assumed a fixed dopamine level and set $\gamma(DA) = 0.1$.

In the absence of plasticity related proteins ($p = 0$), or if no tags are set ($h_i = l_i = 0$), the function $f(z) = z(1-z)(z-0.5)$ generates a bistable dynamics with stable fixed points at $z = 0$ and $z = 1$ and an unstable fixed point at $z = 0.5$ marked by the zero crossings of the function f, Figure 1C. In the presence of a finite amount of proteins $p > 0$ and a non-zero tag, the location of the fixed points changes and for $p > 0.47$, only one of the stable fixed points remains. The potential shown in Figure 1C is a function E with $dE/dz = -f(z)$ so that $dz/dt = -dE/dz$. We note that a synapse i can change its consolidated value only if both a tag ($h_i = 1$ or $l_i = 1$) and protein $p > 0.47$ is present–summarizing the essence of 'synaptic tagging and capture' [12,13].

Synaptic Weight

The synaptic weights have contributions from early and late LTP and LTD. The total synaptic weight of a synapse i is $w_i = \hat{w}(1 + h_i - \alpha l_i + \beta z_i)$ where $\hat{w}$ is the value of a non-tagged synapse, $\alpha = 0.5$ and $\beta = 2$ are parameters, h_i and l_i are binary values indicating E-LTP and E-LTD, respectively, and z_i is the value of the L-LTP trace of synapse i. Since we model slice experiments in animals older than 20 days, we assume that 30 percent of the synapses have undergone previous potentiation and have $z = 1$ while the remaining 70 percent of synapses are in the state $z = 0$ [7]. In all simulation experiments we stimulate one or several groups of $N = 100$ synapses each. Assuming that no tags have been set in the recent past ($h_i = l_i = 0$), the initial value of the average weight in a group of N synapses is then $w(0) = \hat{w}\left[\sum_{i=1}^{N} 1 + \beta z_i\right]/N = 1.6\hat{w}$.

Neuron Model

For all simulations in this paper we use the adaptive exponential integrate-and-fire model [42] as a compact description of neuronal firing dynamics. Briefly, it consists of two equations. The voltage equation has an exponential and a linear term as measured in experiments [59]. The second equation describes adaptation. Although firing rate adaptation is not important for the present study, it would be relevant in the context of other stimulation paradigms. Parameters for the neuron model are as in [42] and are kept fixed for all simulations presented in this paper. The voltage threshold V_s of spike initiation by a short current pulse is 25 mV above the resting potential of -70.6 mV [42]. Synaptic input is simulated as a short current pulse. The initial connection weight $\hat{w}$ was adjusted so that simultaneous activation of 40 or more synapses triggers spike firing in the postsynaptic neuron. Hence the amplitude of a single EPSP is about 0.6 mV.

The adaptive exponential integrate-and-fire model is defined in continuous time. If a spike is triggered by a strong current pulse, the voltage rises within less than 0.5 millisecond to a value of 20 mV where integration is stopped. The voltage is then reset to resting level, and integration restarted after a refractory time of 1 ms. In order to enable us to perform simulations of plasticity experiments with a time step of $\Delta = 1$ ms, the voltage equation during the rising slope of the action potential was integrated once at a much higher resolution (time step 0.02 ms), so as to determine the exact contribution of each postsynaptic spike to the probability of LTP induction. Every postsynaptic spike was then treated as an event in the plasticity simulations that contributed a probability $P_{h=0 \to h=1}$ of flipping the tag from $h = 0$ to $h = 1$ in a time step $\Delta = 1$ ms which we can write as $P_{h=0 \to h=1} = a_\Delta \bar{x}(t)[\bar{u}(t) - \vartheta_{\text{LTD}}]^+$ with a numerical conversion factor $a_\Delta = A_{\text{LTP}} 5$ ms mV derived by the above procedure; see Eq. 2.

Number of Consolidated Synapses

In Figure 5 we plot the number of synapses that have been consolidated as a function of the number N_{tag} of initially tagged ($h_i = 1$) synapses. Since the number of tags decays exponentially with rate k_H, the expected duration T_P^{ON} of protein synthesis is $T_P^{\text{ON}} = (1/k_H) ln(N_{\text{tag}}/N_p)$ where N_p is the protein trigger threshold. While protein synthesis is 'ON' the variables p and z move along the black dashed line in Figure 5A which crosses after a time t_1 the separatrix (green line in Figure 5A) and at a time t_2 the line $z = 0.5$ (vertical dashed green line). Different cases have to be distinguished. (i) $T_P^{\text{ON}} < t_1$, no consolidation takes place (see pink trajectory), hence $N_{\text{up}} = 0$. (ii) $T_P^{\text{ON}} > t_2$, consolidation is guaranteed for all synapses that are still tagged at time t_2, hence $N_{\text{up}} = N_{\text{tag}}\exp(-kt_2)$. (iii) In the case of $t_1 < T_P^{\text{ON}} \leq t_2$, the time t_{cross} needed to cross the vertical line $z = 0.5$ is numerically calculated by integrating the equations $dp/dt = -p/(\tau_p)$ and $dz/dt = f(z) + \gamma \, p$ starting at $t = T_P^{\text{ON}}$ at the point $p(T_P^{\text{ON}}), z(T_P^{\text{ON}})$ on the black-dashed line (see orange line in Figure 5A for a sample trajectory). The number of consolidated synapses is then $N_{\text{up}} = N_{\text{tag}}\exp(-kt_{\text{cross}})$. The solid line in Figure 5B represents N_{up} as a function of N_{tag} calculated for the cases (i)–(iii). With our standard set of parameters, we have $t_1 \approx 28$ min and $t_2 \approx 60$ min.

Author Contributions

Conceived and designed the experiments: CC WG. Performed the experiments: CC. Analyzed the data: CC. Wrote the paper: WG. Designed the model of late LTP and performed research: LZ. Participated in research on a precursor model of early LTP/LTD: EV LB.

References

1. Bliss TVP, Collingridge GL (1993) A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361: 31–39.
2. Malenka RC, Bear MF (2004) LTP and LTD: an embarrassment of riches. Neuron 44: 5–21.
3. Bliss T, Gardner-Medwin A (1973) Long-lasting potentiation of synaptic transmission in the dendate area of unanaesthetized rabbit following stimulation of the perforant path. J Physiol 232: 357–374.
4. Dudek SM, Bear MF (1992) Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. Proc Natl Acad Sci U S A 89: 4363–4367.
5. Artola A, Bröcher S, Singer W (1990) Different voltage dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature 347: 69–72.
6. Markram H, Lübke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic AP and EPSP. Science 275: 213–215.
7. O'Connor D, Wittenberg G, Wang SH (2005) Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proc Natl Acad Sci USA 102: 9679–9684.
8. Bi G, Poo M (2001) Synaptic modification of correlated activity: Hebb's postulate revisited. Annu Rev Neurosci 24: 139–166.
9. Abraham W (2003) How long will long-term potentiation last? Philos Trans R Soc Lond B Biol Sci 358: 735–744.
10. Krug M, Lössner B, Ott T (1984) Anisomycin blocks the late phase of long-term potentiation in the dentate gyrus of freely moving rats. Brain Res Bull 13: 39–42.
11. Sajikumar S, Navakkode S, Sacktor T, Frey J (2005) Synaptic tagging and cross-tagging: the role of protein kinase Mζ in maintaining long-term potentiation but not long-term depression. J Neurosci 25: 5750–5756.
12. Frey U, Morris R (1997) Synaptic tagging and long-term potentiation. Nature 385: 533–536.

13. Reymann K, Frey J (2007) The late maintenance of hippocampal ltp: requirements, phases,synaptic tagging, late associativity and implications. Neuropharmacology 52: 24–40.
14. Pastalkova E, Serrano P, Pinkhasova D, Wallace E, Fenton A, et al. (2006) Storage of spatial information by the maintenance mechanism of LTP. Science 313: 1141–1144.
15. Crick F (1984) Memory and molecular turnover. Nature 312: 101.
16. Lisman J (1985) A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. Proc Natl Acad Sci U S A 82: 3055–3057.
17. Newpher TM, Ehlers MD (2008) Glutamate receptor dynamics in dendritic microdomains. Neuron 58: 472–497.
18. Buonomano D, Merzenich M (1998) Cortical plasticity: from synapses to maps. Annu Rev Neurosci 21: 149–186.
19. Carpenter G, Grossberg S (1987) ART 2: self-organization of stable category recognition codes for analog input patterns. Appl Opt 26: 4919–4930.
20. Nadal JP, Toulouse G, Changeux JP, Dehaene S (1986) Networks of formal neurons and memory palimpsests. Europhys Lett 1: 349–381.
21. Amit D, Fusi S (1994) Learning in neural networks with material synapses. Neural Comput 6: 957–982.
22. Fusi S, Drew P, Abbott L (2005) Cascade models of synaptically stored memories. Neuron 45: 599–611.
23. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 79: 2554–2558.
24. Gerstner W, Kempter R, van Hemmen J, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature 383: 76–78.
25. Kempter R, Gerstner W, van Hemmen JL (1999) Hebbian learning and spiking neurons. Phys Rev E 59: 4498–4514.
26. Song S, Miller K, Abbott L (2000) Competitive Hebbian learning through spike-time-dependent synaptic plasticity. Nat Neurosci 3: 919–926.
27. Lisman J (1989) A mechanism for Hebb and anti-Hebb processes underlying learning and memory. Proc Natl Acad Sci U S A 86: 9574–9578.
28. Miller P, Zhabotinsky A, Lisman J, Wang X (2005) The stability of s stochastic CaMKII switch: dependenc on the number of enzyme molecules and protein turnover. PLoS Biol 3: e107. doi:10.1371/journal.pbio.0030107.
29. Graupner M, Brunel N (2007) Stdp in a bistable synapse model based on CaMKII and associate signaling pathways. PLoS Comput Biol 3: e 221. doi:10.1371/journal.pcbi.0030221.
30. Sajikumar S, Navakkode S, Frey J (2007) Identification of compartment- and process-specific molecules required for 'synaptic tagging' during long-term potentiation and long-term depression in hippocampal CA1. J Neurosci 27: 5068–5080.
31. Othmakhov N, Griffith L, Lisman J (1997) Postsynaptic inhibitors of calcium/calmodulin-dependent protein kinase type II block induction but not maintenance of pairing induced long-term potentiation. J Neurosci 17: 5357–5365.
32. Sajikumar S, Frey J (2004) Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. Neurobiol Learn Mem 82: 12–25.
33. Bienenstock E, Cooper L, Munroe P (1982) Theory of the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J Neurosci 2: 32–48.
34. Fusi S (2002) Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. Biol Cybern 87: 459–470.
35. Pfister JP, Gerstner W (2006) Triplets of spikes in a model of spike timing-dependent plasticity. J Neurosci 26: 9673–9682.
36. Gerstner W, Kistler WK (2002) Spiking Neuron Models. Cambridge, UK: Cambridge University Press.
37. Petersen C, Malenka R, Nicoll R, Hopfield J (1998) All-or-none potentiation of ca3-ca1 synapses. Proc Natl Acad Sci U S A 95: 4732–4737.
38. Artola A, Singer W (1993) Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. Trends Neurosci 16: 480–487.
39. Ngezahayo A, Schachner M, Artola A (2000) Synaptic activation modulates the induction of bidirectional synaptic changes in adult mouse hippocamus. J Neurosci 20: 2451–2458.
40. Frey U, Schroeder H, Matthies H (1990) Dopaminergic antagonists prevent long-term maintenance of posttetanic LTP in the CA1 region of rat hippocampal slices. Brain Res 522: 69–75.
41. Hayer A, Bhalla US (2005) Molecular switches at the synapse emerge from receptor and kinase traffic. PLoS Comput Biol 1: e20. doi:10.1371/journal.pcbi.0010020.
42. Brette R, Gerstner W (2005) Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. J Neurophysiol 94: 3637–3642.
43. O'Connor D, Wittenberg G, Wang S (2005) Dissection of bidirectional synaptic plasticity into saturable unidirectional processes. J Neurophysiol 94: 1565–1573.
44. Markram H, Wu Y, Tosdyks M (1998) Differential signaling via the same axon of neocortical pyramidal neurons. Proc Natl Acad Sci U S A 95: 5323–5328.
45. Smolen P, Baxter D, Byrne J (2006) A model of the roles of essential kinases in the induction and expression of late long-term potentiation. Biophys J 90: 2760–2775.
46. Lisman J, Schulman H, Cline H (2002) The molecular basis of CaMKII function in synaptic and behavioural memory. Nat Rev Neurosci 3: 175–190.
47. Govindarajan A, Kelleher R, Tonegawa S (2006) A clustered plasticity model of long-term memory engrams. Nat Rev Neurosci 7: 575–583.
48. Toyoizumi T, Pfister JP, Aihara K, Gerstner W (2007) Optimality model of unsupervised spike-timing dependent plasticity: synaptic memory and weight distribution. Neural Comput 19: 639–671.
49. Lisman J (2003) Long-term potentiation: outstanding questions and attempted synthesis. Phil Trans R Soc Lond B: Biol Sci 358: 829–842.
50. Fonseca R, Nägerl U, Morris R, Bonhoeffer T (2003) Neuronal activity determines the protein synthesis dependence of long-term potentiation. Nat Neurosci 9: 478–480.
51. Sajikumar S, Frey J (2004) Resetting of 'synaptic tags' is time- and activity-dependent in rat hippocampal CA1 in vitro. Neuroscience 129: 503–507.
52. Sutton R, Barto A (1998) Reinforcement Learning. Cambridge, MA: MIT Press.
53. Arleo A, Gerstner W (2000) Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. Biol Cybern 83: 287–299.
54. Pfister JP, Toyoizumi T, Barber D, Gerstner W (2006) Optimal spike-timing dependent plasticity for precise action potential firing in supervised learning. Neural Comput 18: 1309–1339.
55. Izhikevich E (2007) Solving the distal reward problem through linkage of STDP and dopamine signaling. Cereb Cortex 17: 2443–2452.
56. Legenstein R, Pecevski D, Maass W (2008) A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. PLoS Comput Biol 4: e1000180. doi:10.1371/journal.pcbi.1000180.
57. Schultz W, Dayan P, Montague R (1997) A neural substrate for prediction and reward. Science 275: 1593–1599.
58. Navakkode S, Sajikumar S, Frey J (1007) Synergistic requirements for the induction of dopaminergic D1/D5-receptor-mediated LTP in hippocampal slices of rat CA1 in vitro. Neuropharmacology 52: 1547–1554.
59. Badel L, Lefort S, Brette R, Petersen C, Gerstner W, et al. (2008) Dynamic I-V curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. J Neurophysiol 99: 656–666.

A TOP DOWN APPROACH TO A RULE PERFORMING INDEPENDENT COMPONENT ANALYSIS

INDEPENDENT Component Analysis (ICA) is a technique able to find original source signals out of mixtures of those signals, the so-called cocktail party problem. How do we recover a voice in a cocktail party when mixed with music? There are powerful ICA algorithms that take into account either the spatial distribution of the signal ("spatial ICA") (Hyvärinen, Karhunen, and Oja 2001) or the temporal correlations of the signal ("temporal ICA") (Tong, Liu, Soon, and Huang 1991; Molgedey and Schuster 1994; Belouchrani, Abed-Meraim, Cardoso, and Moulines 1997; Ziehe and Müller 1998). Those algorithms are however difficult to interpret biologically since they are either not online or require difficult preprocessing of the data such as whitening (Hyvärinen, Karhunen, and Oja 2001). Thus the way the brain is able to solve the cocktail party problem is still unclear. In Chapter 3 however, the induction model was shown to perform spatial ICA. We were wondering if a model could solve temporal ICA and what would be the link to the induction model. In the following paper (Clopath, Longtin, and Gerstner 2008) we present a top down approach leading to a biologically plausible model performing ICA. It takes into account the temporal correlations of the signals. By decorrelating the signal mixtures at 2 different time lags, it is possible to recover the original sources. The rule is a standard rate-based Hebbian rule where the rate is taken at those two different time lags. This rule does not require any preprocessing of the data. It allows to de-mix the different sound signals of the standard ICA benchmarks (Hyvärinen, Karhunen, and Oja 2001). The link between this model and the early phase of plasticity model is explained in the "future work" section.

An online Hebbian learning rule that performs Independent Component Analysis

Claudia Clopath
School of Computer Science and Brain Mind Institute
Ecole polytechnique federale de Lausanne
1015 Lausanne EPFL
claudia.clopath@epfl.ch

Andre Longtin
Center for Neural Dynamics
University of Ottawa
150 Louis Pasteur, Ottawa
alongtin@uottawa.ca

Wulfram Gerstner
School of Computer Science and Brain Mind Institute
Ecole polytechnique federale de Lausanne
1015 Lausanne EPFL
wulfram.gerstner@epfl.ch

Abstract

Independent component analysis (ICA) is a powerful method to decouple signals. Most of the algorithms performing ICA do not consider the *temporal* correlations of the signal, but only higher moments of its *amplitude* distribution. Moreover, they require some preprocessing of the data (whitening) so as to remove second order correlations. In this paper, we are interested in understanding the neural mechanism responsible for solving ICA. We present an online learning rule that exploits delayed correlations in the input. This rule performs ICA by detecting joint variations in the firing rates of pre- and postsynaptic neurons, similar to a local rate-based Hebbian learning rule.

1 Introduction

The so-called cocktail party problem refers to a situation where several sound sources are simultaneously active, e.g. persons talking at the same time. The goal is to recover the initial sound sources from the measurement of the mixed signals. A standard method of solving the cocktail party problem is independent component analysis (ICA), which can be performed by a class of powerful algorithms. However, classical algorithms based on higher moments of the signal distribution [1] do not consider temporal correlations, i.e. data points corresponding to different time slices could be shuffled without a change in the results. But time order is important since most natural signal sources have intrinsic temporal correlations that could potentially be exploited. Therefore, some algorithms have been developed to take into account those temporal correlations, e.g. algorithms based on delayed correlations [2, 3, 4, 5] potentially combined with higher-order statistics [6], based on innovation processes [7], or complexity pursuit [8]. However, those methods are rather algorithmic and most of them are difficult to interpret biologically, e.g. they are not online or not local or require a preprocessing of the data.

Biological learning algorithms are usually implemented as an online Hebbian learning rule that triggers changes of synaptic efficacy based on the correlations between pre- and postsynaptic neurons. A Hebbian learning rule, like Oja's learning rule [9], combined with a linear neuron model, has been shown to perform principal component analysis (PCA). Simply using a nonlinear neuron combined with Oja's learning rule allows one to compute higher moments of the distributions which yields ICA if the signals have been preprocessed (whitening) at an earlier stage [1]. In this paper, we are
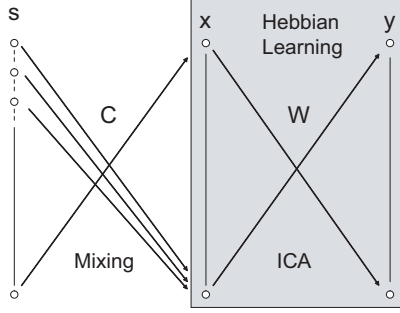
1

Figure 1: The sources **s** are mixed with a matrix C, $\mathbf{x} = C\mathbf{s}$, **x** are the presynaptic signals. Using a linear neuron $\mathbf{y} = W\mathbf{x}$, we want to find the matrix W which allows the postsynaptic signals **y** to recover the sources, $\mathbf{y} = P\mathbf{s}$, where P is a permutation matrix with different multiplicative constants.

interested in exploiting the correlation of the signals at different time delays, i.e. a generalization of the theory of Molgedey and Schuster [4]. We will show that a linear neuron model combined with a Hebbian learning rule based on the joint firing rates of the pre- and postsynaptic neurons of different time delays performs ICA by exploiting the temporal correlations of the presynaptic inputs.

2 Mathematical derivation of the learning rule

2.1 The problem

We assume statistically independent autocorrelated source signals s_i with mean $< s_i > = 0$ ($<>$ means averaging over time) and correlations $< s_i(t)s_j(t') > = K_i(|t - t'|)\delta_{ij}$. The sources **s** are mixed by a matrix C

$$\mathbf{x} = C\mathbf{s}, \tag{1}$$

where **x** are the mixed signals recorded by a finite number of receptors (bold notation refers to a vector). We think of the receptors as presynaptic neurons that are connected via a weight matrix W to postsynaptic neurons. We consider linear neurons [9], so that the postsynaptic signals **y** can be written

$$\mathbf{y} = W\mathbf{x}. \tag{2}$$

The aim is to find a learning rule that adjusts the appropriate weight matrix W to W^* (* denotes the value at the solution) so that the postsynaptic signals **y** recover the independent sources **s** (Fig 1), i.e. $\mathbf{y} = P\mathbf{s}$ where P is a permutation matrix with different multiplicative constants (the sources are recovered in a different order up to a multiplicative constant), which means that, neglecting P,

$$W^* = C^{-1}. \tag{3}$$

To solve this problem we extend the theory of Molgedey and Schuster [4] in order to derive an online biological hebbian rule.

2.2 Theory of Molgedey and Schuster and generalization

The paper of Molgedey and Schuster [4] focuses on the instantaneous correlation matrix but also the time delayed correlations $M_{ij} = < x_i(t)x_j(t + \tau) >$ of the incoming signals. Since the correlation matrix M_{ij} is symmetric, it has up to $n(n + 1)/2$ independent elements. However, the unknown mixing matrix C has potentially n^2 elements (for n sources and n detectors). Therefore, we need to evaluate two delayed correlation matrices M and $\bar{M}$ with two different time delays defined as

$$M_{ij} = < x_i(t)x_j(t + \tau_2) > \qquad \bar{M}_{ij} = < x_i(t)x_j(t + \tau_1) > \tag{4}$$

to get enough information about the mixing process [10].

From equation 1, we obtain the relation $M_{ij} = \sum_l C_{il}C_{jl}\Lambda_{ll}$ and similarly $\bar{M}_{ij} = \sum_l C_{il}C_{jl}\bar{\Lambda}_{ll}$ where $\Lambda_{ij} = \delta_{ij}K_i(\tau_2)$ and $\bar{\Lambda}_{ij} = \delta_{ij}K_i(\tau_1)$ are diagonal matrices. Since $M = C\Lambda C^T$ and $\bar{M} = C\bar{\Lambda}C^T$, we have

$$(M\bar{M}^{-1})C = C(\Lambda\bar{\Lambda}^{-1}). \tag{5}$$

It follows that C can be found from an eigenvalue problem. Since C is the mixing matrix, a simple algorithmic inversion allows Molgedey and Schuster to recover the original sources [4].

2.3 Our learning rule

In order to understand the putative neural mechanism performing ICA derived from the formalism developed above, we need to find an online learning rule describing changes of the synapses as a function of pre- and postsynaptic activity. Taking the inverse of (5), we have $C^{-1}\bar{M}M^{-1} = \bar{\Lambda}\Lambda^{-1}C^{-1}$. Therefore, for weights that solve the ICA problem we expect because of (3) that

$$W^*\bar{M} = \bar{\Lambda}\Lambda^{-1}W^*M, \tag{6}$$

which defines the weight matrix W^* at the solution.

For the sake of simplicity, consider only one linear postsynaptic neuron. The generalization to many postsynaptic neurons is straightforward (see section 4). The output signal y of the neuron can be written as $y = \mathbf{w}^{*\mathbf{T}}\mathbf{x}$, where $\mathbf{w}^{*\mathbf{T}}$ is a row of the matrix W^*. Then equation 6 can be written as

$$\mathbf{w}^{*T}\bar{M} = \lambda\mathbf{w}^{*\mathbf{T}}M, \tag{7}$$

where λ is one element of the diagonal matrix $\bar{\Lambda}\Lambda^{-1}$.

In order to solve this equation, we can use the following iterative update rule with update parameter γ.

$$\dot{\mathbf{w}} = \gamma[\mathbf{w}^{\mathbf{T}}\bar{M} - \lambda\mathbf{w}^{\mathbf{T}}M]. \tag{8}$$

The fixed point of this update rule is giving by (7), i.e. $\mathbf{w} = \mathbf{w}^*$. Furthermore, multiplication of (7) with $\mathbf{w}$ yields $\lambda = \frac{\mathbf{w}^{\mathbf{T}}\bar{M}\mathbf{w}}{\mathbf{w}^{\mathbf{T}}M\mathbf{w}}$.
If we insert the definition of M from (2), we obtain the following rule

$$\dot{\mathbf{w}} = \gamma[< y(t)\mathbf{x}(\mathbf{t} + \tau_{\mathbf{1}}) > -\lambda < y(t)\mathbf{x}(\mathbf{t} + \tau_{\mathbf{2}}) >], \tag{9}$$

with a parameter λ given by

$$\lambda = \frac{< y(t)y(t + \tau_1) >}{< y(t)y(t + \tau_2) >}.$$

It is possible to show that $\dot{\mathbf{w}}$ is orthogonal to $\mathbf{w}$. This implies that to first order (in $|\dot{\mathbf{w}}/\mathbf{w}|$), $\mathbf{w}$ will keep the same norm during iterations of (9).

The rule 9 we derived is a batch-rule, i.e. it averages over all sample signals. We convert this rule into an online learning rule by taking a small learning rate γ and using an online estimate of λ.

$$\dot{\mathbf{w}} = \gamma[y(t)\mathbf{x}(\mathbf{t} + \tau_{\mathbf{1}}) - \frac{\lambda_1}{\lambda_2}y(t)\mathbf{x}(\mathbf{t} + \tau_{\mathbf{2}})] \tag{10}$$

$$\tau_\lambda\dot{\lambda_1} = -\lambda_1 + y(t)y(t + \tau_1)$$

$$\tau_\lambda\dot{\lambda_2} = -\lambda_2 + y(t)y(t + \tau_2).$$

Note that the rule defined in (10) uses information on the correlated activity $\mathbf{x}y$ of pre- and postsynaptic neurons as well as an estimate of the autocorrelation $< yy >$ of the postsynaptic neuron. τ_λ is taken sufficiently long so as to average over a representative sample of the signals and $|\gamma| \ll 1$ is a small learning rate. Stability properties of updates under rule (10) are discussed in section 4.

3 Performances of the learning rule

A simple example of a cocktail party problem is shown in Fig 2 where two signals, a sinus and a ramp (saw-tooth signal), have been mixed. The learning rule converges to a correct set of synaptic
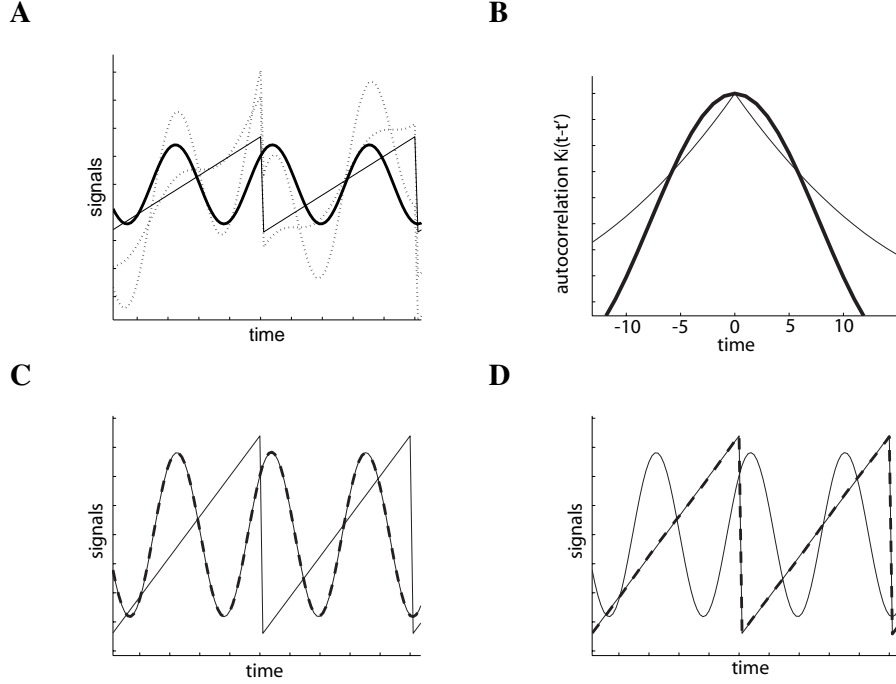
Figure 2: A. Two periodic source signals, a sinus (thick solid line) and a ramp (thin solid line), are mixed into the presynaptic signals (dotted lines). B. The autocorrelation functions of the two source signals are shown (the sinus in thick solid line and the ramp in thin solid line). The sources are normalized so that $\Lambda(0) = 1$ for both. C. The learning rule with $\tau_1 = 3$ and $\tau_2 = 0$ extracts the sinusoidal output signal (dashed) composed to the two input signals. In agreement with the calculation of stability, $\gamma > 0$, the output is recovering the sinus source because $\Lambda_{sin}(3) > \Lambda_{ramp}(3)$. D. The learning rule with $\tau_1 = 10$, $\tau_2 = 0$, converges to the other signal (dashed line), i.e. the ramp, because $\Lambda_{ramp}(10) > \Lambda_{sin}(10)$. Note that the signals have been rescaled since the learning rule recovers the signals up to a multiplicative factor.

weights so that the postsynaptic signal recovers correctly one of the sources. Postsynaptic neurons with different combinations of τ_1 and τ_2 are able to recover different signals (see the section 4 on Stability). In the simulations, we find that the convergence is fast and the performance is very accurate and stable. Here we show only a two-sources problem for the sake of visual clarity. However, the rule can easily recover several mixed sources that have different temporal characteristics.

Fig 3 shows an ICA problem with sources $s(t)$ generated by an Ornstein-Uhlenbeck process of the form $\tau_{s_i}\dot{s}_i = -s_i + \xi$, where ξ is some gaussian noise. The different sources are characterized by different time constants. The learning rule is able to decouple these colored noise signals with gaussian amplitude distribution since they have different temporal correlations.

Finally, Fig 4 shows an application with nine different sounds. We used 60 postsynaptic neurons with time delays τ_1 chosen uniformly in an interval [1,30ms] and $\tau_2 = 0$. Globally 52 of the 60 neurons recovered exactly 1 source (A, B) and the remaining 8 recovered mixtures of 2 sources (E). One postsynaptic neuron is recovering one of the sources depending on the source's autocorrelation at time τ_1 and τ_2 (.i.e. the source with the biggest autocorrelation at time τ_1 since $\tau_2 = 0$ for all neurons, see section Stability). A histogram (C) shows how many postsynaptic neurons recover each source. However, as it will become clear from the stability analysis below, a few specific postsynaptic neurons tuned to time delays, where the autocorrelation functions intersect (D, at time $\tau_1 = 3$ms and $\tau_2 = 0$), cannot recover one of the sources precisely (E).
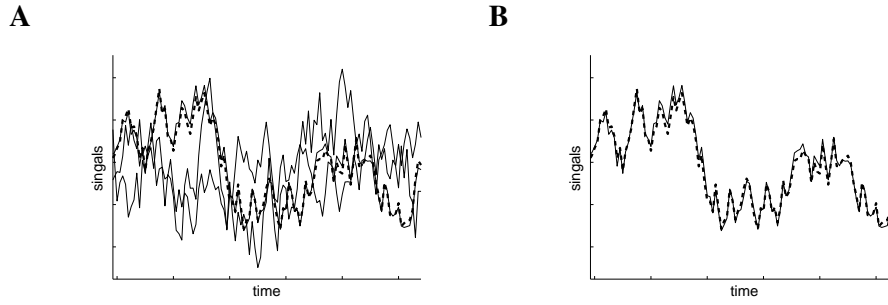
4

Figure 3: A. The 3 source signals (solid lines generated with the equation $\tau_{s_i}\dot{s}_i = -s_i + \xi$ with different time constants, where ξ is some gaussian noise) are plotted together with the output signal (dashed). The learning rule is converging to one of the sources. B. Same as before, but only the one signal (solid) that was recovered is shown together with the neuronal output (dashed).
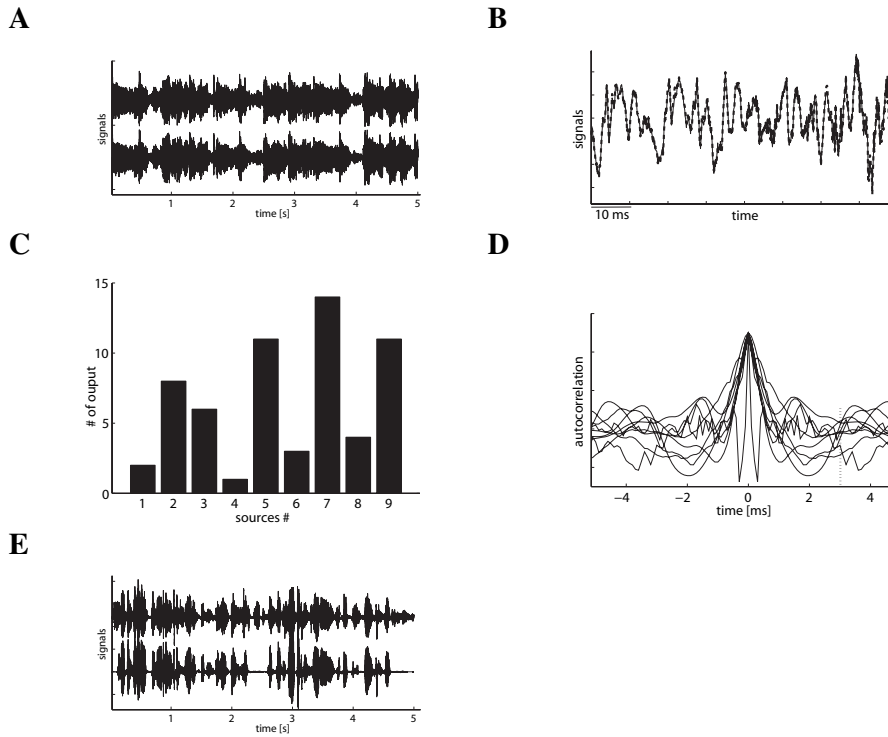

Figure 4: Nine different sound sources from [11] were mixed with a random matrix. 60 postsynaptic neurons tuned to different τ_1 and τ_2 were used in order to recover the sources, i.e. τ_1 varies from 1ms to 30ms by steps of 0.5ms and $\tau_2 = 0$ for all neurons. A. One source signal (below) is recovered by one of the postsynaptic neurons (above, for clarity reason, the output is shifted upward). B. Zoom on one source (solid line) and one output (dashed line). C. Histogram of the number of postsynaptic neurons recovering each sources. D. Autocorrelation of the different sources. There are several sources with the biggest autocorrelation at time 3ms. E. The postsynaptic neuron tuned to a $\tau_1 = 3$ms and $\tau_2 = 0$ (above) is not able to recover properly one of the sources even though it still performs well except for the low amplitude parts of the signal (below).

5

4 Stability of the learning rule

In principle our online learning rule (10) could lead to several solutions corresponding to different fixed points of the dynamics. Fixed points will be denoted by $\mathbf{w}^* = \mathbf{e}_k$, which are by construction the row vectors of the decoupling matrix W^* (see (5) and (7)). The rule 10 has two parameters, i.e. the delays τ_1 and τ_2 (the τ_λ is considered fixed). We assume that in our architecture, these delays characterize different properties of the postsynaptic neuron. Neurons with different choices of τ_1 and τ_2 will potentially recover different signals from the same mixture. The stability analysis will show which fixed point is stable depending on the autocorrelation functions of the signals and the delays τ_1 and τ_2.

We analyze the stability, assuming small perturbation of the weights, i.e. $\mathbf{w} = \mathbf{e}_i + \epsilon \mathbf{e}_j$ where $\{\mathbf{e}^k\}$, the basis of the matrix C^{-1}, are the fixed points. We obtain the expression (see Appendix for calculation details)

$$\dot{\epsilon} = \gamma \epsilon \frac{\Lambda_{jj}(\tau_1)\Lambda_{ii}(\tau_2) - \Lambda_{ii}(\tau_1)\Lambda_{jj}(\tau_2)}{\Lambda_{ii}(\tau_2)}, \qquad (11)$$

where $\Lambda(\tau)_{ij} = \langle s_i(t)s_j(t+\tau)\rangle$ is the diagonal correlation matrix.

To illustrate the stability equation (11), let us take $\tau_1 = 0$ and assume that $\Lambda_{ii}(0) = \Lambda_{jj}(0)$, i.e. all signals have the same zero-time-lag autocorrelation. In this case (11) reduces to $\dot{\epsilon} = \gamma\epsilon[\Lambda_{jj}(\tau_1) - \Lambda_{ii}(\tau_1)]$. That is the solution e_i is stable if $\Lambda_{jj}(\tau_1) < \Lambda_{ii}(\tau_1)$ for all directions e_j (with biggest autocorrelation at time τ_1) for $\gamma > 0$. If $\gamma < 0$, the solution e_i is stable for $\Lambda_{jj}(\tau_1) > \Lambda_{ii}(\tau_1)$.

This stability relation is verified in the simulations. Fig 2 shows two signals with different autocorrelation functions. In this example, we chose $\tau_1 = 0$ and $\Lambda(0) = \mathbb{I}$, i.e. the signals are normalized. The learning rule is recovering the signal with the biggest autocorrelation at time τ_1, $\Lambda_{kk}(\tau_1)$, for a positive learning rate.

5 Comparison between Spatial ICA and Temporal ICA

One of the algorithms most used to solve ICA is FastICA [1]. It is based on an approximation of negentropy and is purely spatial, i.e. it takes into account only the amplitude distribution of the signal, but not it's temporal structure. Therefore we show an example (Fig. 5), where three signals generated by Ornstein-Uhlenbeck processes have the same spatial distribution but different time constants of the autocorrelation. With a spatial algorithm data points corresponding to different time slices can be shuffled without any change in the results. Therefore, it cannot solve this example. We tested our example with FastICA downloaded from [11] and it failed to recover the original sources (Fig. 5). However, to our surprise, FastICA could for very few trial solve this problem even though the convergence was not stable. Indeed, since FastICA algorithm is an iterative online algorithm, it takes the signals in the temporal order in which they arrive. Therefore temporal correlations can in some cases be taken into account even though this is not part of the theory of FastICA.

6 Discussions and conclusions

We presented a powerful online learning rule that performs ICA by computing joint variations in the firing rates of pre- and postsynaptic neurons at different time delays. This is very similar to a standard Hebbian rule with exception of an additional factor λ which is an online estimate of the output correlations at different time delays. The different delay times τ_1, τ_2 are necessary to recover different sources. Therefore properties varying between one postsynaptic neuron and the next could lead to different time delays used in the learning rule. We could assume that the time delays are intrinsic properties of each postsynaptic neuron due to for example the distance on the dendrites where the synapse is formed [12], i.e. due to different signal propagation time. The calculation of stability shows that a postsynaptic neuron will recover the signal with the biggest autocorrelation at the considered delay time or the smallest depending of the sign of the learning rates. We assume that for biological signals autocorrelation functions cross so that it's possible with different postsynaptic neurons to recover all the signals.
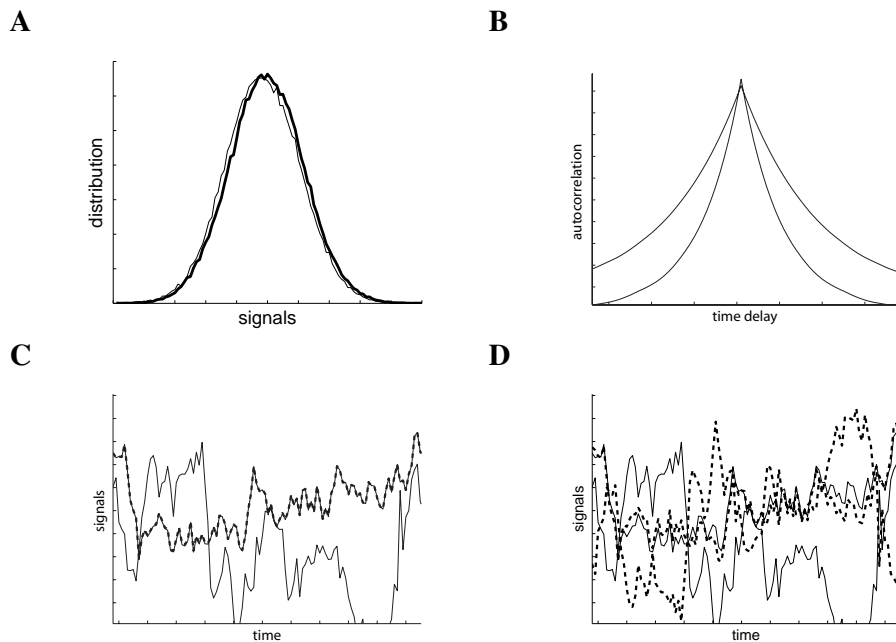
Figure 5: Two signals generated by an Ornstein-Uhlenbeck process are mixed. A. The signals have the same spatial distributions. B. The time constants of the autocorrelations are different. C. Our learning rule converges to an output (dashed line) recovering one of the signals source (solid line). D. FastICA (dashed line) doesn't succeed to recover the sources (solid line).

The algorithm assumes centered signals. However for a complete mapping of those signals to neural rates, we have to consider positive signals. Nevertheless we can easily compute an online estimate of the mean firing rate and remove this mean from the original rates. This way the algorithm still holds taking neural rates as input.

Hyvaerinen proposed an ICA algorithm [8] based on complexity pursuit. It uses the non-gaussianity of the residuals once the part of the signals that is predictable from the temporal correlations has been removed. The update step of this algorithm has some similarities with our learning rule even though the approach is completely different since we want to exploit temporal correlations directly rather than formally removing them by a "predictor". We also do not assume pre-whitened data and are not considering nongaussianity.

Our learning rule considers smooth signals that are assumed to be rates. However, it is commonly accepted that synaptic plasticity takes into account spike trains of pre- and postsynaptic neurons looking at the precise timing of the spikes, i.e. Spike Timing Dependent Plasticity (STDP) [13, 14, 15]. Therefore a spike-based description of our algorithm is currently under study.

Appendix: Stability calculation

By construction, the row vectors $\{\mathbf{e}_k, k = 1,..,n\}$ of $W^* = C^{-1}$, the inverse of the mixing matrix, are solutions of the batch learning rule 9 (n is the number of sources). Assume one of these row vectors $\mathbf{e}_i^T$, (i.e. a fixed point of the dynamic), and consider $\mathbf{w} = \mathbf{e}_i + \epsilon \mathbf{e}_j$ a small perturbation in direction $\mathbf{e}_j^T$. Note that $\{\mathbf{e}_k\}$ is a basis because $det(C) \neq 0$ (the matrix must be invertible). The rule (9) becomes:

7

$$\dot{\epsilon}\mathbf{e}_i = \gamma[<x(t+\tau_1)(\mathbf{e}_i+\epsilon\mathbf{e}_j)^Tx(t)> \tag{12}$$

$$-\frac{<(\mathbf{e}_i+\epsilon\mathbf{e}_j)^Tx(t)(\mathbf{e}_i+\epsilon\mathbf{e}_j)^Tx(t+\tau_1)>}{<(\mathbf{e}_i+\epsilon\mathbf{e}_j)^Tx(t)(\mathbf{e}_i+\epsilon\mathbf{e}_j)^Tx(t+\tau_2)>}<x(t+\tau_2)>(\mathbf{e}_i+\epsilon\mathbf{e}_j)^Tx(t)>].$$

We can expand the terms on the righthand side to first order in ϵ. Multiplying the stability expression by $\mathbf{e}_j^T$ (here we can assume that $\mathbf{e}_j^T\mathbf{e}_j = 1$ since the recovering of the sources are up to a multiplicative constant), we find:

$$\dot{\epsilon} = \gamma\epsilon\frac{[\mathbf{e}_j^TC\Lambda(\tau_1)C^T\mathbf{e}_j][\mathbf{e}_i^TC\Lambda(\tau_2)C^T\mathbf{e}_i]-[\mathbf{e}_i^TC\Lambda(\tau_1)C^T\mathbf{e}_i][\mathbf{e}_j^TC\Lambda(\tau_2)C^T\mathbf{e}_j]}{\mathbf{e}_i^TC\Lambda(\tau_2)C^T\mathbf{e}_i} \tag{13}$$

$$-\epsilon\frac{4[\mathbf{e}_i^TC\Lambda(\tau_1)C^T\mathbf{e}_j][\mathbf{e}_j^TC\Lambda(\tau_2)C^T\mathbf{e}_i]}{\mathbf{e}_i^TC\Lambda(\tau_2)C^T\mathbf{e}_i}.$$

where $\Lambda(\tau)_{ij} = <s_i(t)s_j(t+\tau)>$ is the diagonal matrix.

This expression can be simplified because $\mathbf{e}_i^T$ is a row of $W^* = C^{-1}$, so that $\mathbf{e}_i^TC$ is the unit vector of the form $(0,0,...,1,0,...)$ where the position of the "1" indicates the solution number 0. Therefore, we have $e_i^TC\Lambda(\tau)C^Te_k = \Lambda(\tau)_{ik}$.

The expression of stability becomes

$$\dot{\epsilon} = \gamma\epsilon\frac{\Lambda_{jj}(\tau_1)\Lambda_{ii}(\tau_2)-\Lambda_{ii}(\tau_1)\Lambda_{jj}(\tau_2)}{\Lambda_{ii}(\tau_2)} \tag{14}$$

References

[1] A. Hyvaerinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001.

[2] L Tong, R Liu, VC Soon, and YF Huang. Indeterminacy and identifiability of blind identification. *IEEE Trans. on Circuits and Systems*, 1991.

[3] A. Belouchrani, KA. Meraim, JF. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Sig. Proc.*, 1997.

[4] L. Molgedey and H.G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72:3634–37, 1994.

[5] A. Ziehe and K. Muller. Tdsep – an efficient algorithm for blind separation using time structure.

[6] KR. Mueller, P. Philips, and A. Ziehe. Jade td : Combining higher-order statistics and temporal information for blind source separation (with noise). *Proc. Int. Workshop on ICA*, 1999.

[7] A. Hyvaerinen. Independent component analysis for time-dependent stochastic processes. *Proc. Int. Conf. on Art. Neur. Net.*, 1998.

[8] A. Hyvaerinen. Complexity pursuit: Separating interesting components from time-series. *Neural Computation*, 13:883–898, 2001.

[9] E. Oja. A simplified neuron model as principal component analyzer. *J. Math. Biol.*, 15:267 –273, 1982.

[10] J.J. Hopfield. Olfactory computation and object perception. *PNAS*, 88:6462–6466, 1991.

[11] H. Gavert, J. Hurri, J. Sarela, and A. Hyvarinen. Fastica and cocktail party demo. *http://www.cis.hut.fi/projects/ica/*.

[12] R. C. Froemke, M. Poo, and Y. Dan. Spike-timing dependent synaptic plasticity depends on dentritic location. *Nature*, 434:221–225, 2005.

[13] G. Bi and M. Poo. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 2001.

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

[15] W. Gerstner, R. Kempter, JL. van Hemmen, and H. Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383:76–78, 1996.

CONCLUSIONS AND FUTURE WORK

6.1 SUMMARY OF THE RESULTS

THE four papers presented in this thesis aim for a better understanding of synaptic plasticity. The first paper (Clopath, Jolivet, Rauch, Luescher, and Gerstner 2007) is dedicated to choose an appropriate neuron model that is compact and reproduces well experimental voltage traces of layer V pyramidal cells. The same neuron model, the adaptive Exponential Integrate and Fire model (AdEx), was shown later on to operate in different dynamical regimes and exhibit various firing patterns (Naud, Marcille, Clopath, and Gerstner 2008). The voltage time course of the neuron turns out to be important for the description of the early phase of plasticity (Clopath, Vasilaki, Buesing, and Gerstner xxxx), i.e. not only is the postsynaptic spike timing important but so is the level of depolarization of the postsynaptic neuron (Artola, Bröcher, and Singer 1990; Ngezahayo, Schachner, and Artola 2000; Sjöström, Turrigiano, and Nelson 2001). Thus the model for the early phase of plasticity presented in the second paper (Clopath, Vasilaki, Buesing, and Gerstner xxxx) combines the presynaptic spike time arrival and the postsynaptic membrane potential filtered with different time constants. It is very robust in reproducing experimental data like voltage clamp experiments (Artola, Bröcher, and Singer 1990; Ngezahayo, Schachner, and Artola 2000), frequency dependence (Sjöström, Turrigiano, and Nelson 2001), burst-timing dependent plasticity (Nevian and Sakmann 2006) and a several subtle combinations of spike-timing and voltage depolarization or hyperpolarization. Moreover, this model exhibits some important functional implications like selectivity in the input, for example receptive field development, ICA-like computation and exhibits a tight relation between coding and connectivity. Rate coding leads to few strong bidirectional connections in a sea of weak connections as measured in visual cortex (Song, Sjöström, Reigl, Nelson, and Chklovskii 2005) in contrast to temporal coding which expresses few unidirectional connections as measured in the barrel cortex (Lefort, Tomm, Sarria, and Petersen 2009). This is an interesting interpretation suggesting that different coding schemes exist in different cortical areas. Interestingly, standard STDP models (Gerstner, Kempter, van Hemmen, and Wagner 1996; Gerstner and van Hemmen 1993; Roberts and Bell 2000; Kistler and van Hemmen 2000;

Mehta, Quirk, and Wilson 2000; Song, Miller, and Abbott 2000; Legenstein, Naeger, and Maass 2005; Guyonneau, VanRullen, and Thorpe 2005) cannot sustain stable bidirectional connections (Song, Miller, and Abbott 2000). This early phase model is combined with a late-phase model of long term synaptic plasticity, the TagTriC model (Clopath, Ziegler, Vasilaki, Buesing, and Gerstner 2008) presented in the third paper. It takes into account discrete synapses for the early phase where the probabilities of transition are given by the model described in the previous paper (Clopath, Vasilaki, Buesing, and Gerstner xxxx). These discrete states also set the tags at the synapses. It is combined with a triggering process of plasticity related proteins and a consolidated phase which includes bistable states. This model reproduces a large number of synaptic tagging experiments and cross tagging. Finally, a top down approach of computing temporal ICA in a biologically plausible manner was derived in the last paper (Clopath, Longtin, and Gerstner 2008). This rate-based rule decorrelates the signal mixtures at different time points (lags) so that the original signal recovered is the one with the biggest autocorrelation at these time points. It allows different neurons, with different properties (i.e. lag) to recover different sources. The model performed greatly on standard ICA benchmarks with sound sources.

6.2 Open questions and future work

This thesis work has revealed a range of open questions and therefore offers a list of possible future work:

Spike-based rule performing temporal ICA

The last paper of the thesis presents a rate-based rule performing ICA (Clopath, Longtin, and Gerstner 2008). The next step is to develop a spiking version of this model. The type of encoding has to be decided but the most straightforward is to use Poisson neurons that fire with probability corresponding to the signal. Thus the presynaptic spike trains would encode the signal mixtures. Postsynaptic neurons can also be considered as Poisson neurons that fires with the probability reflecting the weighted sum of the inputs. Synaptic plasticity should converge to a set of weights so that the postsynaptic neuron encodes one of the original sources. The plasticity model can take the form of a standard STDP learning window (Markram, Lübke, Frotscher, and Sakmann 1997; Bi and Poo 1998) reflecting the correlation at two different lags in the rate-based framework. The amplitudes of this window have to vary depending on the output correlation, like the λ term in the rate-based rule. This variable amplitudes can be interpreted for instance as homeostasis. This STDP model with homeostasis could be sufficient to perform temporal ICA. The encoding with Poisson neurons is however not exploring precise timing of the spikes. A temporal code could be considered, like the time to first spike.

Biological ICA

Humans have two ears to solve the cocktail party problem! At one given point in time they have access to two different mixtures of the original sources. However, classical ICA benchmarks contain the same number of sources as mixtures to prevent loss of information. The question remains how humans do to decouple more than two different sources? Can they recover only two sources at the same time and attention is then shifted to recover the other ones? Do they have to move slightly their head to decode different mixtures? Does the auditory preprocessing help for this problem? In fact auditory preprocessing is known to decompose the sources in different frequency bands due to the hair cells in the cochlea. Should Fourier decomposition be taken into account in this case?

Relation between spike-based rule performing temporal ICA and the model for early phase plasticity

The second paper (Clopath, Vasilaki, Buesing, and Gerstner xxxx) shows that a nonlinear rule in the postsynaptic term combined with homeostasis, i.e. BCM form, is sufficient to perform spatial ICA. It would be interesting to find out whether the nonlinearity could be shifted from the postsynaptic term to a simple non linear neuron or not. Moreover the link between the spiking spatial ICA model (Clopath, Vasilaki, Buesing, and Gerstner xxxx) (i.e. taking into account spatial distribution of the signal) and the spiking temporal ICA model (follow up of Clopath, Longtin, and Gerstner 2008, taking into account temporal correlations of the signal) is not yet clear. It remains an open question whether different models exist in the brain and operate with different inputs. For example visual inputs would be processed by spatial algorithms and audio inputs by temporal algorithms. On the contrary, a combined model performing temporal and spatial ICA (Müller, Philips, and Ziehe 1999; Hyvärinen 1998; Hyvärinen 2001) would be more robust to any kind of sources, like for example a movie. If that is the case, the way those two models (Clopath, Vasilaki, Buesing, and Gerstner xxxx and follow up of Clopath, Longtin, and Gerstner 2008) should be combined has to be clarified.

Extension of the model for early phase of plasticity

This model takes into account the voltage of the postsynaptic neuron which turns out the be critical for synaptic plasticity (Artola, Bröcher, and Singer 1990; Sjöström, Turrigiano, and Nelson 2001). However the proposed plasticity rule uses a neuron model with a single compartment, i.e. equal voltage at the soma and at the synapses. This is definitely not the case and a more detailed neuron model giving the voltage at the synapse would be helpful, especially describing the back propagating action potential. It would then be possible to describe quantitatively plasticity depending on dendritic location, e.g. the differ-

ence between apical plasticity when the back propagating action potential fails and basal dendrite plasticity (Sjöström and Häusser 2006). Moreover some additional non-linearities should be taken into account. For example, there seems to be a non-linearity in the number of pairing, where a small number of pairing only leads to LTP (Wittenberg and Wang 2006) and a saturation effects may occur. Finally different weight dependencies should be explored. Note that optimal models clearly predict that nonlinearities are crucial (Toyoizumi, Pfister, Aihara, and Gerstner 2005).

Interestingly, pre-post pairing at low frequency when the neuron is stimulated intracellularly does not change the weights (Sjöström, Turrigiano, and Nelson 2001) whereas extracellular stimulation leads to potentiation (Froemke and Dan 2002). This results suggest that pre-post pairing would lead to potentiation only if previously there was some depolarization (Sjöström, Turrigiano, and Nelson 2001), high activity of the slices via extracellular stimulation (Froemke and Dan 2002), or neuromodulation (Schultz and Dickinson 2000) (the last two could be the same since extracellular stimulation can excite dopaminergic fibers). The model should then be adjusted so that the potentiation term contains a correlation between pre-post with an additional term which could be either voltage (Clopath, Vasilaki, Buesing, and Gerstner xxxx) or neuromodulation (Schultz and Dickinson 2000), unifying triplet rule and reward-modulated learning rules. Moreover, the neuromodulation term seems to be complex and has to be tuned with additional experiments.

Finally it would be interesting to validate the model with the predictions explained in the paper (Clopath, Vasilaki, Buesing, and Gerstner xxxx).

Combing short term plasticity with long term plasticity

In the second paper (Clopath, Vasilaki, Buesing, and Gerstner xxxx) the model describes long term plasticity lasting 2 to 3 hours. However, synapses are also plastic at a much shorter time scale lasting hundreds of milliseconds, the so-called short-term plasticity. It would be worthwhile to combine short term (Markram and Tsodyks 1996; Abbott, Varela, Sen, and Nelson 1997) and long term plasticity models especially in protocols using high frequency spiking. The impact of short term plasticity would then be more clear and the long term model may have to be readjusted.

Extension of the TagTriC model

The TagTriC model has some limitations at least concerning one set of experimental data set which resets the tags (Sajikumar and Frey 2004b). This model is thus oversimplified and needs to differentiate between setting the tags and the early phase of plasticity. A more elaborate model with additional states should allow for tag resetting, possibly a hybrid model between our model and another model for synaptic tagging (Barrett, Billings, Morris, and van Rossum 2009).

84

Many open questions still exist concerning the late-phase of long-term plasticity. For example it is not clear what are the states of the synaptic weights 10 hours after a STDP protocol, since weights are recorded usually for 30 minutes. Do they undergo consolidation? If yes, does reducing the number of pairing prohibits maintenance? If not, is a longer STDP protocol sufficient to reach the late phase or is it necessary to pair a strong extracellular protocol in a different pathway, stimulating dopaminergic fibers? Maintenance depends on neuromodulation but how? More experiments are required using pharmacological tools.

Link between the TagTriC model and reinforcement learning models

The TagTriC model presented in the third paper (Clopath, Ziegler, Vasilaki, Buesing, and Gerstner 2008) exhibits some structural similarities to the reinforcement learning framework: (i) The selection of "definitive" memories depends in both cases on neuromodulators, such as dopamine (Schultz and Dickinson 2000). In reinforcement learning, the weights are updated in the presence of dopamine encoding prediction of reward and in the tagging experiment maintenance requires stimulation of dopaminergic fibers. (ii) The presence of a memory trace. In reinforcement learning the eligibility trace keeps a memory of the pre-post correlation and in the tagging experiment, the early phase of plasticity keeps a memory of the induction. However, the time scales for the early phase and the eligibility trace seem to be different. Thus the TagTriC model should be tested against standard reinforcement learning tasks like learning a location in a maze.
Moreover, recent experiments provide evidence for behavioral tagging (Moncada and Viola 2007), memory reconsolidation and extinction (Eisenhardt and Menzel 2007) where the time scales seem to be more consistent with the ones measured in tagging experiments (Frey and Morris 1997). The TagTriC model should be further validated with these recent experimental findings.

Functional implications of the early phase of plasticity

It is necessary to check that the functional implications of the early phase of our plasticity model (Clopath, Vasilaki, Buesing, and Gerstner xxxx) still hold when combined with the late phase model, i.e. selectivity, ICA computation, relation between connectivity and coding. Would it open up more computational possibilities? What are the consequences when applied to bigger networks?
Analytical study of a plastic recurrent network under this nonlinear model should be investigated, similarly to the case with standard STDP rule (Kempter, Gerstner, and van Hemmen 1999; Burkitt, Gilson, and van Hemmen: 2007; Gilson, Burkitt, Grayden, Thomas, and van Hemmen 2009a; Gilson, Burkitt, Grayden, Thomas, and van Hemmen 2009b).

Functional implications of the voltage dependency

The functional consequences of the early phase plasticity model (Clopath,

Vasilaki, Buesing, and Gerstner xxxx) were essentially taking advantage of the nonlinearity of the model and the homeostatic properties. However, the consequence of the voltage dependence was not explored. It would be interesting to study the implications in a network with background activity that probably leads to a much higher average voltage. What are the consequences in vivo, where the voltage is closer to threshold?

Functional implications of the TagTriC model in a large network

What are the functional consequences of the TagTriC model in a large network? Does the network learn over a long period? In contrary, a network with multiplicative STDP shows no stable strong synapses over time (Morrison, Aertsen, and Diesmann 2007). What is the final weight distribution? Does memory consolidation increase memory capacity in a network? In fact the TagTriC model can be seen as a shallow cascade model (Fusi, Drew, and Abbott 2005) with only two levels. How good is the trade-off between a plastic network and a long lasting memory, i.e. palimpsest property (Nadal, Toulouse, Changeux, and Dehaene 1986; Amit and Fusi 1994)? However the exact construction of such a network is not straightforward: (i) Should the inhibitory plasticity in such a network be taken into account? For this, more experimental data (Gaiarsa, Caillard, and Ben-Ari 2002; Haas, Nowotny, and Abarbanel 2006) and an appropriate model is needed. (ii) Plasticity is very diverse across synapses (Caporale and Dan 2008; Sjöström, Rancz, Roth, and Hausser 2008). Should this inhomogeneity be taken into account? (iii) What would be the inputs to the network and the learning benchmarks?

6.3 Conclusion

The present thesis explores different types of long-term synaptic plasticity by developing models based upon simple mechanistic principles. Such models can be applied to plastic artificial networks mimicking brain areas, offering a direct interpretation of the obtained results. They reveal induction mechanisms of synaptic plasticity as well as mechanisms of its maintenance and consolidation.

The proposed models offer a good trade-off between the brain plasticity and maintenance of previous memories (known as stability-plasticity dilemma). Under different coding scheme scenarios they produce distinctive connectivity patterns that can be related to different brain areas, suggesting that the underlying reason of the experimentally observed differences may be the different encoding schemes used in the various areas, for instance rate coding in visual cortex and temporal coding in barrel cortex.

Finally these models explain how experience modifies structural properties like receptive field development and how the brain computes Independent Component Analysis when in a noisy environment such as the cocktail party problem.

Appendix A

ADDITIONAL WORK

I<small>N</small> this appendix, a few extra experiments with the induction model from (Clopath, Vasilaki, Buesing, and Gerstner xxxx) is shown. These are unpublished results.

E<small>XTENDED INDUCTION MODEL FITS TRIPLET AND QUADRUPLET DATA</small>

We simulate three sets of experiments on hippocampal cultures from (Wang, Gerkin, Nauen, and Bi 2005). In the first experiment, we apply a standard pairing protocol. In the second experiment we apply the triplet protocol, i.e. a set of pre-post-pre and post-pre-post triplets with variable time distances between the spikes. In the third experiment we apply the quadruplet protocol, i.e. either a post-pre pair of spikes is followed after time T by a pre-post pair of spikes (if T positive) or a pre-post pair of spikes is followed after time -T by a post-pre pair of spikes (if T negative). For a detailed description of the protocols see (Wang, Gerkin, Nauen, and Bi 2005).

We fit the model parameters such that the data of the pairing, triplet and the quadruplet protocol are reproduced altogether with the same parameters (Fig. A.1). It turns out that for the hippocampal cultures, an additional term is necessary, i.e. a pre-post pair-term for potentiation, similar to the triplet rule from (Pfister and Gerstner 2006).

F<small>UNCTIONAL CONSEQUENCES OF THE INDUCTION MODEL</small>

The induction model is used in different scenarios of rather simple feedforward networks. Selectivity is observed when a group of inputs undergoes rate modulation (Fig. A.2A) or shares spike-spike correlation (Fig. A.2B). Additionally the weights of set of common inputs exhibit spontaneous selectivity after a very long time, showing the unstable dynamics (Fig. A.2C).
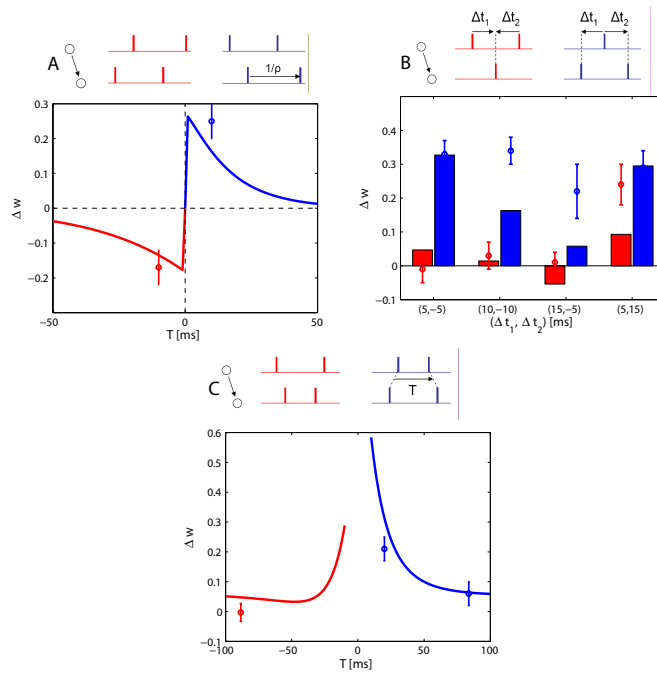
Figure A.1: Weight change according to pair (A), triplet (B) and quadruplet protocols (C) following the induction model with an additional pre-post term for potentiation. A. 60 pairing at 1Hz where T is the time between the pairs (red: post-pre, blue: pre-post). B. 60 triplets at 1Hz (red:pre-post-pre, blue: post-pre-post) C. 60 quadruplet at 1Hz (red:pre-post-post-pre, blue:post-pre-pre-post). Dots are experimental data taken from (Wang, Gerkin, Nauen, and Bi 2005).
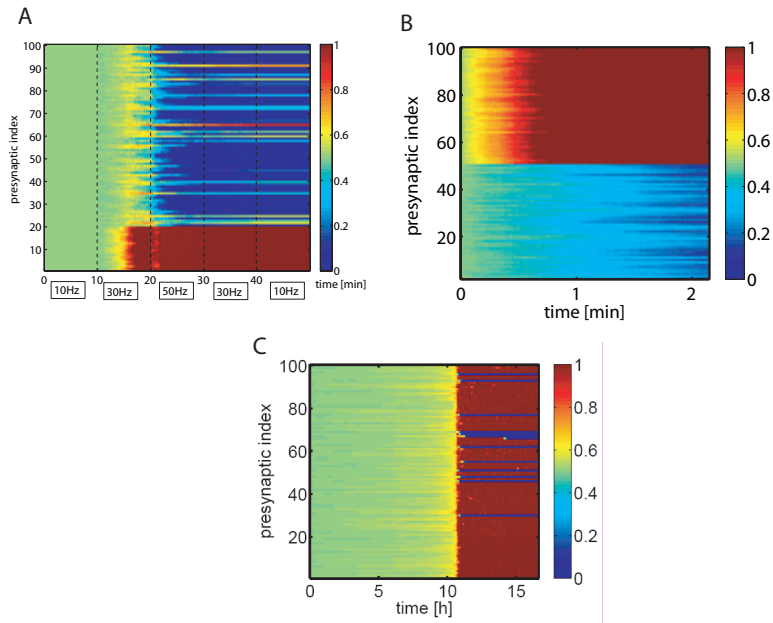
Figure A.2: A. Rate modulation. 100 presynaptic Poisson inputs are connected to one postsynaptic neuron that receives as a current the weighted sum of the inputs. The 80 last inputs fire at 10Hz. The 20 first inputs undergo a rate-modulation: a random number is chosen every 200ms between 0-10, 0-30, 0-50, 0-30, 0-10Hz for 10min each. Selectivity is observed as well as a hysteresis meaning that the selection is stable within those 10min. B. Spike-spike correlation. The 100 inputs fire at 10Hz but the 50 lasts share spike-spike correlation of c=2. Selectivity is observed. C. Spontaneous selection. The 100 inputs fire at 10Hz. Spontaneous selectivity is observed after 10hours indicating two stable points of the dynamics.

REFERENCES

Abarbanel, H., L. Gibb, R. Huerta, and M. Rabinovich (2003). Biophysical model of synaptic plasticity dynamics. *Biological cybernetics 89*, 214–226.

Abarbanel, H., R. Huerta, and M. Rabinovich (2002). Dynamical model of long-term synaptic plasticity. *Proc. Natl. Academy of Sci. USA 59* (10137-10143).

Abbott, L. F., J. A. Varela, K. Sen, and S. B. Nelson (1997). Synaptic depression and cortical gain control. *Science 275*, 220–224.

Amit, D. and S. Fusi (1994). Learning in neural networks with material synapses. *Neural Computation 6*, 957–982.

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

Badel, L., S. Lefort, R. Brette, C. Petersen, W. Gerstner, and M. Richardson (2008). Dynamic i-v curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J Neurophysiol 99*, 656 – 666.

Bagal, A., J. Kao, C. Tang, and S. Thompson (2005). Long-term potentiation of exogenous glutamate responses at single dendritic spines. *Proc. Natl. Acad. Sci. USA 102(40)*, 14434–14439.

Barrett, A., G. Billings, R. Morris, and M. van Rossum (2009). State based model of long-term potentiation and synaptic tagging and capture. *Plos Comp Biol 5(1)*, e1000259. doi:10.1371/journal.pcbi.1000259.

Bednar, J. and R. Miikkulainen (2000). Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neural Computation 12*(7), 1721–1740.

Bell, A. and T. Sejnowski (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation 7*, 1129–1159.

Bell, C., V. Han, Y. Sugawara, and K. Grant (1997). Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature 387*, 278–281.

Belouchrani, A., K. Abed-Meraim, J.-F. Cardoso, and E. Moulines (1997). A blind source separation technique using second order statistics. *IEEE Transactions on Signal Processing 45*, 434–444.

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

Bienenstock, E., L. Cooper, and P. Munroe (1982). Theory of the development of neuron selectivity: orientation specificity and binocular interaction in

visual cortex. *Journal of Neuroscience 2*, 32–48. reprinted in Anderson and Rosenfeld, 1990.

Bienenstock, E. L., L. N. Cooper, and P. W. Munro (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience 2*(1), 32–48.

Bliss, T. and A. Gardner-Medwin (1973). Long-lasting potentation of synaptic transmission in the dendate area of unanaesthetized rabbit following stimulation of the perforant path. *J. Physiol. 232*, 357–374.

Brette, R. and W. Gerstner (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol. 94*, 3637 − 3642.

Burkitt, A., M. Gilson, and J. van Hemmen: (2007). Spike-timing-dependent plasticity for neurons with recurrent connections. *Biological Cybernetics 96(5)*, 533–546.

Cai, Y., J. Gavornik, L. Cooper, L. Yeung, and H. Shouval (2007). Effect of stochastic synaptic and dendritic dynamics on synaptic plasticity in visual cortex and hippocampus. *J Neurophysiol 97*, 375–386.

Caporale, N. and Y. Dan (2008). Spike timingŨdependent plasticity: A hebbian learning rule. *Annual Reviews Neuroscience 31*, 25–46.

Clopath, C., R. Jolivet, A. Rauch, H.-R. Luescher, and W. Gerstner (2007). Predicting neuronal activity with simple models of the threshold type: Adaptive exponential integrate-and-fire model with two compartments. *Neurocomputing 70*, 1668 − 1673.

Clopath, C., A. Longtin, and W. Gerstner (2008). An online hebbian learning rule that performs independent component analysis. *Advances in Neural Information Processing Systems 20*, 312–328.

Clopath, C., E. Vasilaki, L. Buesing, and W. Gerstner (xxxx). Connectivity reflects coding: A model of voltage-based spike-timing-dependent-plasticity with homeostasis. *Under review in Nature Neuroscience x*, xx–xx.

Clopath, C., L. Ziegler, E. Vasilaki, L. Buesing, and W. Gerstner (2008, Dec). Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression. *PLoS Comput Biol 4*(12).

Dudek, S. M. and M. F. Bear (1993). Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. *J. Neuroscience 13*, 2910–2918.

Egger, V., D. Feldmeyer, and B. Sakmann (1999). Coincidence detection and changes of synaptic efficacy in spiny stellated neurons ins barrel cortex. *Nature Neuroscience 2*, 1098–1105.

Eisenhardt, D. and R. Menzel (2007). Extinction learning, reconsolidation and the internal reinforcer hypothesis. *Neurobiol. Learn. & Mem. 87*, 167–173.

Florian, R. V. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation 19*, 1468–1502.

Foster, D., R. Morris, and P. Dayan (2000). Models of hippocampally dependent navigation using the temporal difference learning rule. *Hippocampus 10*, 1–16.

Fourcaud-Trocme, N., D. Hansel, C. van Vreeswijk, and N. Brunel (2003). How spike generation mechanisms determine the neuronal response to fluctuating input. *J. Neuroscience 23*, 11628–11640.

Frey, U. and R. Morris (1997). Synaptic tagging and long-term potentiation. *Nature 385*, 533 − 536.

Froemke, R. and Y. Dan (2002). Spike-timing dependent plasticity induced by natural spike trains. *Nature 416*, 433–438.

Fusi, S., P. Drew, and L. Abbott (2005). Cascade models of synaptically stored memories. *Neuron 45*, 599–611.

Gaiarsa, J., O. Caillard, and Y. Ben-Ari (2002). Long-term plasticity at gabaergic and glycinergic synapses: mechanisms and functional significance. *Trends in Neuroscience 25(11)*, 564–570.

Gerstner, W., R. Kempter, J. van Hemmen, and H. Wagner (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature 383*(6595), 76–78.

Gerstner, W. and J. L. van Hemmen (1993). Spikes or rates? − stationary, oscillatory, and spatio-temporal states in an associative network of spiking neurons. In S. Gielen and B. Kappen (Eds.), *ICANN'93, Proceedings of the International Conference on Artificial Neural Networks, Amsterdam, 13-16 September 1993*, pp. 633–638. London: Springer-Verlag.

Gilson, M., A. Burkitt, D. Grayden, D. Thomas, and J. van Hemmen (2009a). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. i. input selectivity strengthening correlated input pathways. *Biological Cybernetics*, 10.1007/s00422−009−0319−4.

Gilson, M., A. Burkitt, D. Grayden, D. Thomas, and J. van Hemmen (2009b). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. ii. input selectivity symmetry breaking. *Biological Cybernetics*, 10.1007/s00422−009−0320−y.

Graupner, M. (2008). *Induction and Maintenance of Synaptic Plasticity*. Ph. D. thesis, Uni. Pierre et Marie Curie Paris V AND TU Dresden.

Graupner, M. and N. Brunel (2007). Stdp in a bistable synapse model based on CaMKII and associate signaling pathways. *PLOS Comput. Biol. 3*, e 221 doi:10.1371/journal.pcbi.0030221.

Gupta, A., Y. Wang, and H. Markram (2000). Organizing principles for a diversity of gabaergic interneurons and synapses in the neocortex. *Science 287*, 273–278.

Gustafsson, B., H. Wigstrom, W. C. Abraham, and Y.-Y. Huang (1987). Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus. *J. Neurosci. 7*, 774–780.

Gütig, R., R. Aharonov, S. Rotter, and H. Sompolinsky (2003a). Learning input correlations through non-linear temporally asymmetric hebbian plasticity. *J. Neuroscience 23*, 3697–3714.

Gütig, R., S. Aharonov, S. Rotter, and H. Sompolinsky (2003b). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *Journal of Neuroscience 23*(9), 3697–3714.

Gütig, R. and H. Sompolinsky (2006, March). The tempotron: a neuron that learns spike timing-based decisions. *Nat Neurosci 9*(3), 420–428.

Guyonneau, R., R. VanRullen, and S. Thorpe (2005). Neurons tune to the earliest spikes through stdp. *Neural Computation 17*(4), 859–879.

Haas, J., T. Nowotny, and H. Abarbanel (2006). Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *Journal of neurophysiology doi:10.1152/jn.00551.2006.*

Hebb, D. O. (1949). *The Organization of Behavior.* New York: Wiley.

Hinton, G. E. and T. J. Sejnowski (1983). Analyzing cooperative computation. *In Proceedings of the 5th Annual Congress of the Cognitive Science Society, Rochester, NY.*

Hodgkin, A. L. and A. F. Huxley (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol 117*(4), 500–544.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA 79*, 2554–2558.

Hyvärinen, A. (1998). Independent component analysis for time-dependent stochastic processes. *Proc. Int. Conf. on Art. Neur. Net..*

Hyvärinen, A. (2001). Complexity pursuit: Separating interesting components from time-series. *Neural Computation 13*, 883–898.

Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent Component Analysis.* Wiley.

Iglesiasa, J., J. Erikssonb, F. Grize, M. Tomassini, and A. Villa (2005). Dynamics of pruning in simulated large-scale spiking neural networks. *Biosystems 79(1-3)*, 11–20.

Izhikevich, E. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks 15*, 1063–1070.

Izhikevich, E. (2007). Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral Cortex 17*, 2443–2452.

Izhikevich, E. M. and G. M. Edelman (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences 105*, 3593–3598.

Karmarkar, U. and D. Buonomano (2002). A model of spike-timing dependent plasticity: one or two coincidence detectors. *J. Neurophysiology 88*, 507–513.

Karmarkar, U., M. Najarian, and D. Buonomano (2002). Mechanisms and significance of spike-timing dependent plasticity. *Biol. Cybernetics 87*, 373–382.

Kelso, S. R., A. H. Ganong, and T. H. Brown (1986). Hebbian synapses in hippocampus. *Proc. Natl. Acad. Sci. USA 83*, 5326–5330.

Kempter, R., W. Gerstner, and J. L. van Hemmen (1999). Hebbian learning and spiking neurons. *Phys. Rev. E 59*, 4498–4514.

Kistler, W. M. and J. L. van Hemmen (2000). Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic potentials. *Neural Comput. 12*, 385–405.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE 78*, 1464–1480.

Kozloski, J. and G. A. Cecchi (2008). Topological effects of spike timing-dependent plasticity. *arxiv.org abs*, 0810.0029.

Lefort, S., C. Tomm, J. Sarria, and C. Petersen (2009). The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron 61*, 301–316.

Legenstein, R., C. Naeger, and W. Maass (2005). What can a neuron learn with spike-timing dependent plasticity. *Neural Computation 17*, 2337–2382.

Ling, D., L. Benardo, P. Serrano, N. Blace, M. Kelly, J. Crary, and T. Sacktor (2002). Protein kinase Mζ is necessary and sufficient for ltp maintenance. *Nature Neuroscience 5*, 295–296.

Linsker, R. (1986). From basic network principles to neural architecture: emergence of orientation selective cells. *Proc. Natl. Acad. Sci. USA 83*, 8390–8394.

Lisman, J. (1985). A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proc. Natl. Acad. Sci. USA 82*, 3055–3057.

Lisman, J. (1989). A mechanism for hebb and anti-hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. USA 86*, 9574–9578.

Lisman, J., H. Schulman, and H. Cline (2002). The molecular basis of camkii function in synaptic and behavioural memory. *Nat Rev Neurosci 3*, 175–190.

Lisman, J. and N. Spruston (2005). Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. *Nature Neuroscience 8*(7), 839–841.

Lisman, J. and A. Zhabotinsky (2001). A model of synaptic memory: A CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron 31*, 191–201.

Lu, J., C. Li, J.-P. Zhao, M. ming Poo, and X. Zhang (2007). Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type. *J. Neuroscience 27*, 9711–9720.

Lynch, G., T. Dunwiddie, and V. Gribkoff (1977). Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature 266*, 737 – 739.

Markram, H., J. Lübke, M. Frotscher, and B. Sakmann (1997). Regulation of synaptic efficacy by coincidence of postysnaptic AP and EPSP. *Science 275*, 213–215.

Markram, H. and M. Tsodyks (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature 382*, 807–810.

Markram, H., Y. Wu, and M. Tosdyks (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. USA 95*, 5323–5328.

McCormick, D. A., Z. Wang, and J. Huguenard (1993). Neurotransmitter control of neocortical neuronal activity and excitability. *Cereb Cortex 3*(5), 387–398.

Mehta, M. R., M. Quirk, and M. Wilson (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron 25*, 707–715.

Miller, K. D. and D. J. C. MacKay (1994). The role of constraints in hebbian learning. *Neural Computation 6*, 100–126.

Miller, P., A. Zhabotinsky, J. Lisman, and X. Wang (2005). The stability of s stochastic camkii switch: dependenc on the number of enzyme molecules and protein turnover. *PLOS Biology 3*, e107.

Molgedey, L. and H. G. Schuster (1994). Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters 72*(23), 3634–3637.

Moncada, D. and H. Viola (2007). Induction of long-term memory by exposure to novelty requires protein synthesis: Evidence for a behavioral tagging. *Journal of Neuroscience 27(28)*, 7476–7481.

Morrison, A., A. Aertsen, and M. Diesmann (2007). Spike-timing dependent plasticity in balanced random networks. *Neural Computation 19*, 1437–1467.

Müller, K., P. Philips, and A. Ziehe (1999). Jade td : Combining higher-order statistics and temporal information for blind source separation (with noise). *Proc. Int. Workshop on ICA*.

Nadal, J.-P., G. Toulouse, J.-P. Changeux, and S. Dehaene (1986). Networks of formal neurons and memory palimpsests. *Europhysics Letters 1*, 349–381.

Naud, R., N. Marcille, C. Clopath, and W. Gerstner (2008). Firing patterns in the adaptive exponential integrate-and-fire model. *Biological Cybernetics 99*, 335–347.

Nevian, T. and B. Sakmann (2006). Spine ca2+ signaling in spike-timing-dependent plasticity. *J. Neurosci. 26*(43), 11001–11013.

Ngezahayo, A., M. Schachner, and A. Artola (2000). Synaptic activation modulates the induction of bidirectional synaptic changes in adult mouse hippocamus. *J. Neuroscience 20*, 2451–2458.

O'Connor, D., G. Wittenberg, and S.-H. Wang. (2005). Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. USA 102*, 9679–9684.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Mathematical Biology 15*, 267–273.

Okamoto, H. and K. Ichikawa (2000). Switching characteristics of a model for biochemical-reaction networks describing autophosphorylation versus dephosphorylation of ca2+/calmodulin-dependent protein kinase ii. *Biological Cybernetics 82*, 35–47.

Olshausen, B. A. and D. J. Field (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature 381*, 607–609.

Pawlak, V. and J. Kerr (2008). Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *Journal of Neuroscience 28(10)*, 2435–2446.

Petersen, C., R. Malenka, R. Nicoll, and J. Hopfield (1998). All-or-none potentiation of ca3-ca1 synapses. *Proc. Natl. Acad. Sci. USA 95*, 4732–4737.

Pfister, J.-P. and W. Gerstner (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neuroscience 26*, 9673–9682.

Pfister, J.-P., T. Toyoizumi, D. Barber, and W. Gerstner (2006). Optimal spike-timing dependent plasticity for precise action potential firing in supervised learning. *Neural Computation 18*, 1309–1339.

Piomelli, D. (2003). The molecular logic of endocannabinoid signalling. *Nature Reviews Neuroscience 4*, 873–884.

Reymann, K. and J. Frey (2007). The late maintenance of hippocampal ltp: requirements, phases,synaptic tagging, late associativity and implications. *Neuropharmacology 52*, 24–40.

Roberts, P. and C. Bell (2000). Computational consequences of temporally asymmetric learning rules: II. Sensory image cancellation. *Computational Neuroscience 9*, 67–83.

Rubin, J., R. Gerkin, G.-Q. Bi, and C. Chow (2005). Calcium time course as a signal for spike-timing-dependent plasticity. *J. Neurophysiology 93*, 2600–2613.

Rubin, J., D. D. Lee, and H. Sompolinsky (2001). Equilibrium properties of temporally asymmetric Hebbian plasticity. *Physical Review Letters 86*, 364–367.

Sajikumar, S. and J. Frey (2004a). Late-associativity, synaptic tagging, and the role of dopamine during ltp and ltd. *Neurobiology of Learning and Memory 82*, 12–25.

Sajikumar, S. and J. Frey (2004b). Resetting of synaptic tags is time- and activity dependent in rat hippocampal ca1 in vitro. *Neuroscience 129*, 503–507.

Saudargiene, A., B. Porr, and F. Wörgötter (2003). How the shape of pre- and postsynaptic signals can influence stdp: A biophysical model. *Neural Computation 16*, 595–626.

Schultz, W., P. Dayan, and R. Montague (1997). A neural substrate for prediction and reward. *Science 275*, 1593–1599.

Schultz, W. and A. Dickinson (2000). Neuronal coding of prediction errors. *Annual Reviews of Neuroscience 23*, 472–500.

Sejnowski, T. J. and G. Tesauro (1989). The Hebb rule for synaptic plasticity: algorithms and implementations. In J. H. Byrne and W. O. Berry (Eds.), *Neural Models of Plasticity*, Chapter 6, pp. 94–103. Academic Press.

Senn, W., M. Tsodyks, and H. Markram (2001). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Computation 13*, 35–67.

Seung, H. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron 40*, 1063–1073.

Sheynikhovich, D., R. Chavarriaga, T. Strosslin, A. Arleo, and W. Gerstner (2009). Is there a geometric module for spatial orientation? insights from a rodent navigation model. *Psychological Review to appear*.

Shouval, H. Z., M. F. Bear, and L. N. Cooper (2002). A unified model of nmda receptor dependent bidirectional synaptic plasticity. *Proc. Natl. Acad. Sci. USA 99*, 10831–10836.

Sjöström, P. and M. Häusser (2006). A Cooperative Switch Determines the Sign of Synaptic Plasticity in Distal Dendrites of Neocortical Pyramidal Neurons. *Neuron 51*(2), 227–238.

Sjöström, P., E. Rancz, A. Roth, and M. Hausser (2008). Dendritic excitability and synaptic plasticity. *Physiological Review 88*, 769–840.

Sjöström, P., G. Turrigiano, and S. Nelson (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron 32*, 1149–1164.

Sjöström, P., G. Turrigiano, and S. Nelson (2003). Neocortical ltd via coincident activation of presynaptic nmda and cannabinoid receptors. *Neuron 39*, 641–654.

Sjöström, P., G. Turrigiano, and S. Nelson (2004). Endocannabinoid-dependent neocortical layer-5 ltd in the absence of postsynaptic spiking. *J. Neurophysiol. 92*, 3338–3343.

Song, S., K. Miller, and L. Abbott (2000). Competitive Hebbian learning through spike-time-dependent synaptic plasticity. *Nature Neuroscience 3*, 919–926.

Song, S., P. Sjöström, M. Reigl, S. Nelson, and D. Chklovskii (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology 3*, 507–519.

Sprekeler, H., C. Michaelis, and L. Wiskott (2007). Slowness: An objective for spike-timing-plasticity? *PLoS Computational Biology 3*(6), e112.

Sutton, R. and A. Barto (1998). *Reinforcement learning.* MIT Press, Cambridge.

Tong, L., R. Liu, V. Soon, and Y. Huang (1991). Indeterminacy and identifiability of blind identification. *IEEE Trans. on Circuits and Systems.*

Toyoizumi, T., J.-P. Pfister, K. Aihara, and W. Gerstner (2005). Generalized bienenstock-cooper-munro rule for spiking neurons that maximizes information transmission. *Proc. National Academy Sciences (USA) 102*, 5239–5244.

Toyoizumi, T., J.-P. Pfister, K. Aihara, and W. Gerstner (2005). Spike-timing dependent plasticity and mutual information maximization for a spiking neuron model. In L. K. Saul, Y. Weiss, and L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17*, pp. 1409–1416. Cambridge, MA: MIT Press.

Turrigiano, G. and S. Nelson (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience 5*, 97–107.

Tzounopoulos, T., Y. Kim, D. Oertel, and L. O. Trussell (2004). Cell-specific, spike timingâĽŠdependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience 7*, 719–125.

van Rossum, M. C. W., G. Q. Bi, and G. G. Turrigiano (2000). Stable Hebbian learning from spike timing-dependent plasticity. *J. Neuroscience 20*, 8812–8821.

von der Malsburg, C. (1973). Self-organization of orientation selective cells in the striate cortex. *Kybernetik 14*, 85–100.

Wang, H., R. Gerkin, D. Nauen, and G. Bi (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience 8*, 187–193.

Williams, R. (1992). Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Reinforcement Learning 8*, 229–256.

Wiskott, L. and T. Sejnowski (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation 14*(4), 715–770.

Wittenberg, G. M. and S. S.-H. Wang (2006). Malleability of spike-timing-dependent plasticity at the ca3Ŭca1 synapse. *Journal of Neuroscience 26(24)*, 6610–6617.

Xie, X. and H. Seung (2004). Learning in neural networks by reinforcement of irregular spiking. *Physical Review E 69*(4), 41909.

Zhabotinsky, A. (2000). Bistability in the ca2+/calmodulin-dependent protein kinase-phosphatase system. *Biophysical Journal 79(5)*, 2211–2221.

Ziehe, A. and K. Müller (1998). TDSEP–an efficient algorithm for blind separation using time structure. *Proc. Int. Conf. on Artificial Neural Networks (ICANN '98)*, 675–680.

Claudia CLOPATH

Contact:
Address: EPFL-LCN
             Station 15
             CH-1015 Lausanne
Telephone: 0041.21.693.18.37
E-mail   : claudia.clopath@epfl.ch

CURRICULUM VITAE

EDUCATION

Swiss Federal Institute of Technology (Switzerland): PhD	from 2005
Advanced Course in Computational Neuroscience Freiburg	2008
Massachusetts Institute of Technology (USA): Master project	2004-2005
University of Waterloo (Canada): Exchange year in Physics	2002-2003
Swiss Federal Institute of Technology (Switzerland): BA/MA in Physics	2000-2005

EXPERIENCE

PhD thesis, EPFL (supervised by W. Gerstner)   from 2005
- Models of early and late long-term plasticity and their functional implications
Teaching assistant, EPFL   2005-2008
- Neural networks and biological modeling
Master Project, MIT (supervised by M. L. Gray)   2004-2005
- Study of cartilage biosynthesis using magnetic resonance spectroscopy

PUBLICATIONS

Connectivity reflects coding: A model of voltage-based spike-timing-dependent-plasticity with homeostasis
Clopath C, Vasilaki E, Büsing L and Gerstner W. Under review in Nature Neuroscience

Tag-Trigger-Consolidation: A model of early and late long-term potentation and depression
Clopath C, Ziegler L, Vasilaki E, Büsing L and Gerstner W. PLoS Computational Biology, 2008, 4, e1000248

An online Hebbian learning rule that performs Independent Component Analysis
Clopath C, Longtin A and Gerstner W. Advances in Neural Information Processing Systems 20, MIT Press, 2008, 312-328

Firing Patterns in the Adaptive Exponential Integrate-and-Fire
Naud R, Marcille N, Clopath C, Gerstner W. Biological Cybernetics, 2008, 98:459-478

Predicting Neuronal Activity with Simple Models of the Threshold Type: Adaptive Exponential Integrate-and-Fire Model with Two Compartments
Clopath C, Jolivet R, Rauch A, Lüscher H-R and Gerstner W. Neurocomputing, 2007, 70, 1668 – 1673

TALKS

Modeling synaptic plasticity across different time scales: LTP induction and consolidation.
- BMI-EPFL retreat, Villars. Feb 09

Modeling induction of LTP/LTD: the role of voltage and spike timing.
- Fast Analog Computing with Emergent Transient State (Facets, European project), Lausanne. Oct 08

Effect of lateral connections in a plastic mainly feed-forward network.
- Advanced Course in Computational Neuroscience, Freiburg. Aug 08

A unified voltage-based model for timing and voltage dependent synaptic plasticity.
- Society for Industrial and Applied Mathematics (SIAM), Montreal. Aug 08

An online Hebbian learning rule that performs Independent Component Analysis.
- Annual Computational Neuroscience Meeting (CNS), Portland. Jul 08

Voltage model of STDP leads to BCM and ABS.
- Computational and Systems Neuroscience (Cosyne), Salt Lake City. Ma 08

Modeling Synaptic Plasticity.
- Fast Analog Computing with Emergent Transient State (Facets, European project), Debrecen. Fev 08

An online Hebbian learning rule that performs Independent Component Analysis.
- Invited talk in André Longtin's lab, Ottawa. Dec 07

REVIEWING ACTIVITY
Annual Computational Neuroscience Meeting (CNS)

GRADES AND AWARDS

Grade for the Master project: 6/6	2005
Final grade of the Master: 5.62/6	2005
Physics and Mathematics award for the scientific maturity, Nyon	2000

LANGUAGES
French: Mother tongue
English: Fluent
German: Good knowledge

COMPUTER KNOWLEDGE
Programming languages: Matlab, HTML, LATEX
Graphics: Illustrator

MISCELLANEOUS
-Sports trainer for disabled people, Plussport CH
-Blood donation assistant for the Red Cross, Nyon, Switzerland
Hobbies:
Tango, traveling, swimming, mountain climbing, skiing, squash