Stimulus-Driven Unsupervised Synaptic Pruning in Large Neural Networks

Javier Iglesias[1,3,4], Jan Eriksson[3], Beatriz Pardo[2], Marco Tomassini[1], and Alessandro E.P. Villa[1,3,4]

[1] Information Systems Department, University of Lausanne, Switzerland
{Javier.Iglesias, Marco.Tomassini}@unil.ch
http://inforge.unil.ch/
[2] Centro de Biologia Molecular Severo Ochoa, Universidad Autonoma, Madrid, Spain
bpardo@cbm.uam.es
[3] Laboratory of Neuroheuristics, University of Lausanne, Switzerland
jan@lnh.unil.ch
http://www.nhrg.org/
[4] Inserm U318, Laboratory of Neurobiophysics, University Joseph Fourier, Grenoble, France
Alessandro.Villa@ujf-grenoble.fr

Abstract. We studied the emergence of cell assemblies out of locally connected random networks of integrate-and-fire units distributed on a 2D lattice stimulated with a spatiotemporal pattern in presence of independent random background noise. Networks were composed of 80% excitatory and 20% inhibitory units with initially balanced synaptic weights. Excitatory–excitatory synapses were modified according to a spike-timing-dependent synaptic plasticity (STDP) rule associated with synaptic pruning. We show that the application, in presence of background noise, of a recurrent pattern of stimulation let appear cell assemblies characterized by an internal pattern of converging projections and a feed-forward topology not observed with an equivalent random stimulation.

1 Introduction

Massive synaptic pruning following over-growth is a general feature of mammalian brain maturation [11]. Pruning starts near time of birth and is completed by time of sexual maturation. Trigger signals able to induce synaptic pruning could be related to dynamic functions that depend on the timing of action potentials. Spike-timing-dependent synaptic plasticity (STDP) is a change in the synaptic strength based on the ordering of pre- and postsynaptic spikes. This mechanism has been proposed to explain the origin of long-term potentiation (LTP), i.e. a mechanism for reinforcement of synapses repeatedly activated shortly before the occurrence of a postsynaptic spike [8,2]. STDP has also been proposed to explain long-term depression (LTD), which corresponds to the weakening of synapses strength whenever the presynaptic cell is repeatedly activated shortly after the occurrence of a postsynaptic spike [7]. The relation

M. De Gregorio et al. (Eds.): BVAI 2005, LNCS 3704, pp. 59–68, 2005.

between synaptic efficacy and synaptic pruning [3,9], suggests that the weak synapses may be modified and removed through competitive "learning" rules. Competitive synaptic modification rules maintain the average neuronal input to a postsynaptic neuron, but provoke selective synaptic pruning in the sense that converging synapses are competing for control of the timing of postsynaptic action potentials [12,13].

This article studies the emergence of cell assemblies out of a locally connected random network of integrate-and-fire units distributed on a 2D lattice. The originality of our study stands on the size of the network, between 8,100 and 12,100 units, the duration of the experiment, 500,000 time units (one time unit corresponding to the duration of a spike), and the application of an original bio-inspired STDP modification rule compatible with hardware implementation [4]. In this study the synaptic modification rule was applied only to the exc–exc connections. This plasticity rule might produce the strengthening of the connections among neurons that belong to cell assemblies characterized by recurrent patterns of firing. Conversely, those connections that are not recurrently activated might decrease in efficacy and eventually be eliminated. The main goal of our study is to determine whether or not, and under which conditions, such cell assemblies may emerge from a large neural network receiving background noise and content-related input organized in both temporal and spatial dimensions.

2 Model

The complete neural network model is described in details in [5]. Some aspects that were not discussed in that reference are presented here, along with a sketch description of the model. Integrate-and-fire units (80% excitatory and 20% inhibitory) were laid down on a squared 2D lattice according to a space-filling quasi-random Sobol distribution. Network sizes of $[90 \times 90]$, $[100 \times 100]$, and $[110 \times 110]$ were simulated. Sparse connections between the two populations of units were randomly generated according to a two-dimensional Gaussian density function such that excitatory projections were dense in a local neighbourhood, but low probability long-range excitatory projections were allowed [5]. Edge effects induced by the borders were limited by folding the network as a torus.

All units of the network were simulated by leaky integrate-and-fire neuromimes. The state of the unit (spiking/not spiking) was a function of the membrane potential and a threshold. After spiking, the membrane potential was reset, and the unit entered an absolute refractory period set to 3 ms for excitatory units, and 2 ms for inhibitory units. Each unit received a background excitatory input (corresponding to a depolarization of 60 mV) that followed an independent and uncorrelated Poisson process of mean $\lambda = 5$ spikes/s.

It is assumed *a priori* that modifiable synapses are characterized by discrete activation levels that could be interpreted as a combination of two factors: the number of synaptic *boutons* between the pre- and postsynaptic units and the changes in synaptic conductance as a result of Ca^{2+} influx through the NMDA receptors. In the current study we attributed a fixed activation level (meaning

no synaptic modification) $A_{ji}(t) = 1$, to exc–inh, inh–exc, and inh–inh synapses while activation levels were allowed to take one of $A_{ji}(t) = \{0, 1, 2, 4\}$ for exc–exc synapses, $A_{ji}(t) = 0$ meaning that the projection was permanently pruned out (see [5] for more details).

3 Stimulus Protocol

Each simulation was running for $5 \cdot 10^5$ discrete time steps (1 ms per time step), corresponding to about 8.5 minutes. After a stabilization period of 1 s without any external input, a stimulus was presented every 2 seconds. Overall this represented 250 presentations of the stimulus along one simulation run. Three stimulus durations were used: 50 ms followed by 1,950 ms without any external input, 100 ms followed by 1,900 ms without any external input, 200 ms followed by 1,800 ms without any external input. The stimulus was composed of vertical bars uniformly distributed over the 2D lattice surface, each bar being 1 column wide. The number of bars composing the stimulus was a function of the simulated network sizes: 9 bars for $[90 \times 90]$ networks, 10 bars for $[100 \times 100]$ networks, and 11 bars for $[110 \times 110]$ networks, such that the bars were always distant of 10 columns one from another and spanning all over the available surface. At each time step during stimulus presentation, the bars were simultaneously moved one column to the right, such that each bar slipped over the entire surface of the network.

The stimulus was applied only to a fraction of the population formed by excitatory units; these units are called *input units*. The number of input units used for the simulations was a ratio (i.e. 3, 5, 7, or 10%) of the initial number of excitatory units. For a $[100 \times 100]$ network, 10% of input units corresponds to 800 input units, i.e. 10% of the 80% excitatory units of the 10,000 units. The stimulus applied on a particular input unit provoked a depolarization on its membrane with amplitudes equal to 0 (i.e. no stimulation), 30, 40, 50, and 60 mV, depending on the protocol. Notice that the stimulus amplitude was selected in the beginning and did not vary during the simulations.

The three following presentation protocols were applied: (i) *'No stimulation'*: this condition corresponds to a stimulation of zero amplitude which is necessary to check computing artefacts that might be associated to the programming routines used to "stimulate" the units; (ii) *'Random stimulation'*: at each stimulus presentation, the input units were randomly chosen, such that the input units changed at any new stimulus presentation; (iii) *'Fixed stimulation'*: the input units were selected in the beginning of the simulation run and remained the same at any new stimulus presentation. The total amount of applied stimulation is equal in both random and fixed protocols.

To summarize the stimulation procedure, let us consider the following example. For each of the input units, randomly selected among the 10% of excitatory units, of a $[100 \times 100]$ network stimulated with a 100 ms stimulus, one stimulus presentation resulted in a sequence of 10 external inputs equally distributed in time every 10 ms. At the network level, each stimulus presentation corresponded

to a spatiotemporal sequence characterized by 10 groups of 80 synchronously excited units stimulated 10 times during 10 ms.

4 Computer Implementation

The simulator was a custom C program that relies on the GNU Scientific Library (GSL) for random number generation and quasi-random Sobol distribution implementations. With our current implementation and setup at the University of Lausanne, a 10,000 units network simulation for a duration of 500 seconds lasted approximatively 3 hours, depending on the network global activity. We performed simulations with both fixed and random input stimulations, using the same model parameters and pseudo-random number generator seed, and compared the cell assemblies that emerged. Network activity was recorded as a multivariate time series akin of multisite multiple spike train recordings at a resolution of 1 ms. The firing pattern of each unit could be characterized by first- and second-order time domain analyses using the programs and tools accessible from the OpenAdap.Net[1] project.

The complete status of the network was dumped when the simulations were stopped, providing information on the strength of the connections after the STDP-driven synaptic plasticity and pruning. A set of custom scripts were used to extract emerged cell assemblies from the dumped status. The extracted weighted and oriented graphs were further analyzed by means of a tool built on top of the Java Universal Network/Graph Framework (JUNG[2]). Some typical graph measurements were computed, including the number of incoming projections (k_{in}, *in-degree*) and outgoing projections (k_{out}, *out-degree*) for each vertex of the graph.

5 Results

The pool of excitatory units whose incoming and/or outgoing excitatory projections were not entirely pruned and that were *not* directly depolarized by the external stimulus was identified at time t= 500 seconds (from the beginning of the simulation). Among the units of this pool a subset of units is selected on the basis of their connection pattern to– and from the pool itself. The units with at least three strong incoming ($k_{in} \geq 3$) and three strong outgoing projections ($k_{out} \geq 3$) within the pool are dubbed *strongly interconnected units* (*SI*-units).

The activity of all the *SI*-units was affected by the *fixed stimulation* presentation. Fig. 1 shows the response of two *SI*-units to an external stimulus lasting 200 ms, during the fixed stimulus presentation. About 22% of the *SI*-units were strongly inhibited during the stimulus presentation (e.g. Fig. 1a), despite the fact that the stimulus was delivered only to excitatory units. This effect is due to the activity of the local inhibitory units that receive excitatory projections from

[1] `http://www.openadap.net/`
[2] `http://jung.sourceforge.net/`

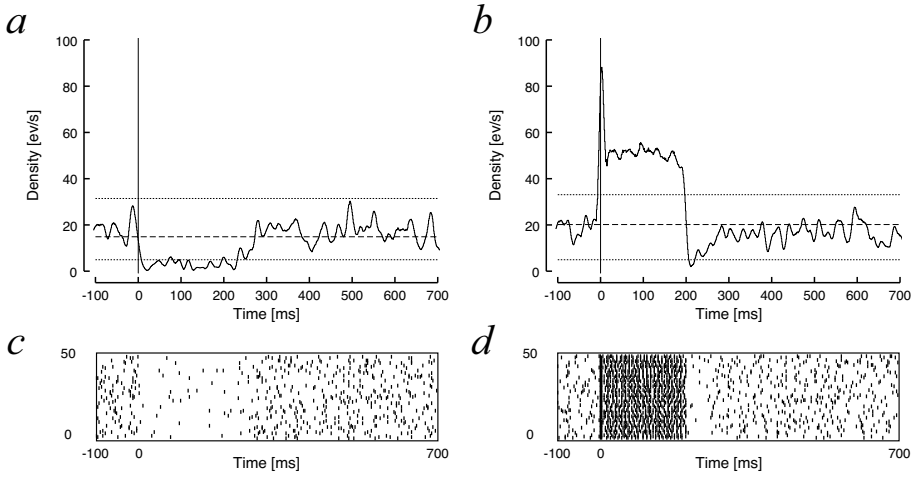

Fig. 1. Response of two strongly interconnected sample units to 50 presentations of the *fixed stimulation* between time $t = 450$ and $t = 500$ seconds from the simulation start. Network size: [100 × 100]; background activity: 5 spikes/s; stimulus duration: 200 ms; stimulus intensity: 60 mV; ratio of input units: 10%; *fixed stimulation* protocol. ***(a, b):*** peri-event densities (PSTH) for the last 50 presentations of the stimulus; smoothed with a Gaussian kernel, bin=5ms. *Dashed line* corresponds to the mean firing rate; *dotted lines* represent the 99% confidence limits assuming a Poissonian distribution. Time zero corresponds to the stimulus onset; ***(c, d):*** corresponding raster plots.

the input units. The peristimulus histogram of the other *SI*-units showed that the firing rate strongly increased during the stimulus presentation (e.g. Fig. 1b) with a "primary-like" response pattern, despite the fact that none of the units belonging to this pool was directly stimulated.

The effect of the different stimulation protocol was complex. The overall number of *SI*-units found in absence of stimulation was similar to the number of *SI*-units found with *random stimulation* ($n \approx 400$, 6% of excitatory units not directly stimulated). In the *fixed stimulation* protocol, the number of *SI*-units was much smaller (about 2%), but depended on the stimulus-induced depolarization amplitude (Fig. 2). Conversely, in the *random stimulation* protocol condition, we did not observe a significant change of the number of *SI*-units in response to stimulus intensity.

During the process of pruning only the modifiable connections that kept a sufficient level of activity driven by the STDP rule could "survive". Then, the first step for searching an oriented topology after 500 seconds consisted to detect the *excitatory neighbourhood* of the *SI*-units. This neighbourhood corresponds to the set of those excitatory units that send a projection to the *SI*-units, receive a projection from the *SI*-units, or both send and receive projections. Thus, this neighbourhood may also include input units, i.e. the units that are directly receiving the stimulus. The ratio between the number of input units belonging

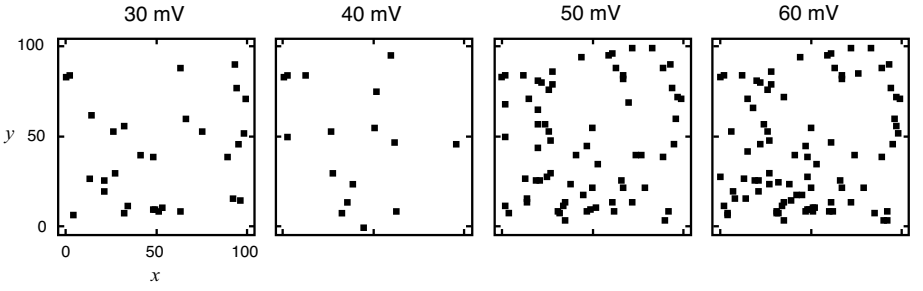

Fig. 2. Example of the location of strongly interconnected units as a function of the amplitude of the stimulus-induced depolarization. Network size: [100 × 100]; stimulus duration: 200 ms; input units: 10% of the excitatory units; *fixed stimulation* protocol.

to the neighbourhood and the number of *SI*-units defines the *index of connected units* (ICU). The larger the ICU, the larger the influence of the input units on the *SI*-units.

The value for the ICU was computed for different network dimensions, stimulus durations and ratio of input units during the *fixed stimulation* protocol (Fig. 3). With a ratio of input units equal to 3%, we observed that the value of ICU was almost zero and independent of the other parameters, because the amount of stimulus delivered to the network was not sufficiently large to let appear a noticeable stimulus-driven pruning. Such pruning appeared with 5% of input units and became clearly visible with 7 and 10% of input units. It is worth to note that a stimulus lasting 200 ms provoked an effect similar to a stimulus lasting 50 ms. The "network size" effect is not so interesting by itself, as it is consistent with the fact that the smaller the network, the larger is the impact of a certain ratio of the input units. Besides, the application of a parameter scaling factor introduced in [5] almost suppressed the size effect (compare Fig. 3a and b).

The evolution of k_{in} and k_{out} for the *SI*-units and their neighbourhood was studied as a function of the simulation duration for a [100 × 100] network. The state of the network was analysed at $t = 50$, $t = 200$ and $t = 500$ seconds (Fig. 4). In the beginning of a simulation, an average excitatory unit receives and sends projections to about 190 other excitatory units, i.e. $k_{in} = k_{out} \approx 190$ (see Fig. 4a). The variability comes from the projection two-dimensional Gaussian density function (see Model description). As no new connections are established during the simulation, k_{in} and k_{out} can only decrease under the pressure of the pruning process. Some units tend to loose their incoming connections first, others tend to loose their outgoing connections first. The existence of other processes combining different speeds for the loss of input and output connections results in the smear of points visible in Fig. 4b-d.

We observed that as soon as $t = 50$ seconds, corresponding to 25 stimulus presentations with the *fixed stimulation* protocol (Fig. 4d), the evolution of the *SI*-neighbour units k_{in} and k_{out} was different from the other two protocols.

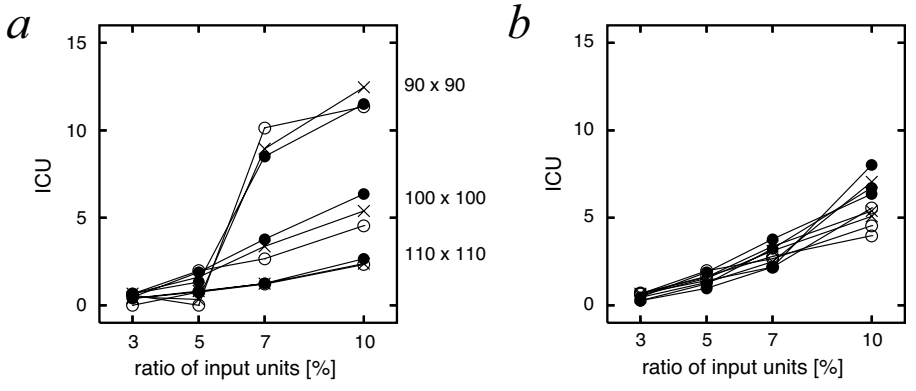

Fig. 3. The index of connected units (ICU), i.e. the ratio between the number of input units and the number of *SI*-units, as a function of the ratio of input units, stimulus duration, and network dimensions. Labels of stimulus duration: ○: 50 ms; ×: 100 ms; •: 200 ms. ***(a):*** simulations performed with unscaled parameter values for all network sizes; ***(b):*** like *(a)* except for the size-specific scaled variables defined in [5]. Stimulus intensity: 60 mV; *fixed stimulation* protocol.

Plots for $t = 200$ and $t = 500$ seconds show that most units have $k_{out} \ll k_{in}$, which indicates that the pruning modified the topology of the connections and favored the emergence of a converging pattern. The comparison of these degrees between $t = 200$ and $t = 500$ s (Fig. 4e-g *vs.* Fig. 4h-j) shows that the tendency to loose outward projections continued during the last part of the simulation. In particular, notice that a large part of the neighbourhood population lost all its input connections ($k_{in} = 0$); these units 'survived' only because the background noise maintained some of their outward connections timely tuned with the discharges of their targets.

Figure 4 shows that the distribution patterns, for the *random stimulation* protocol (Fig. 4c,f,i) and in absence of stimulation (Fig. 4b,e,h) are very similar. A random stimulus could not drive any significant effect, which was somehow expected, but it was necessary as a control experiment to detect any bias introduced in the simulation program. In the *fixed stimulation* protocol (Fig. 4g,j), we observed $n = 415$ units with $30 \leq k_{in} \leq 130$ at $t = 200$ s that are maintained at $t = 500$ s. There are only 26 units with these properties in the other two conditions. This population is composed of 407 input units belonging to the neighbourhood. These input units maintained a large k_{in}, because of the synchronization of their activity during the stimuli presentations. The vast majority of the input units ($> 85\%$) were presynaptic with respect to the *SI*-units, thus confirming that the topology organized towards a feed-forward converging pattern of connections.

In the *fixed stimulation* protocol, the number of incoming and outgoing projections of the *SI*-units was $k_{in} \approx 180$ and $3 \leq k_{out} \leq 20$. It is important to notice that the distribution of the k_{in} of the *SI*-units did not change in time.

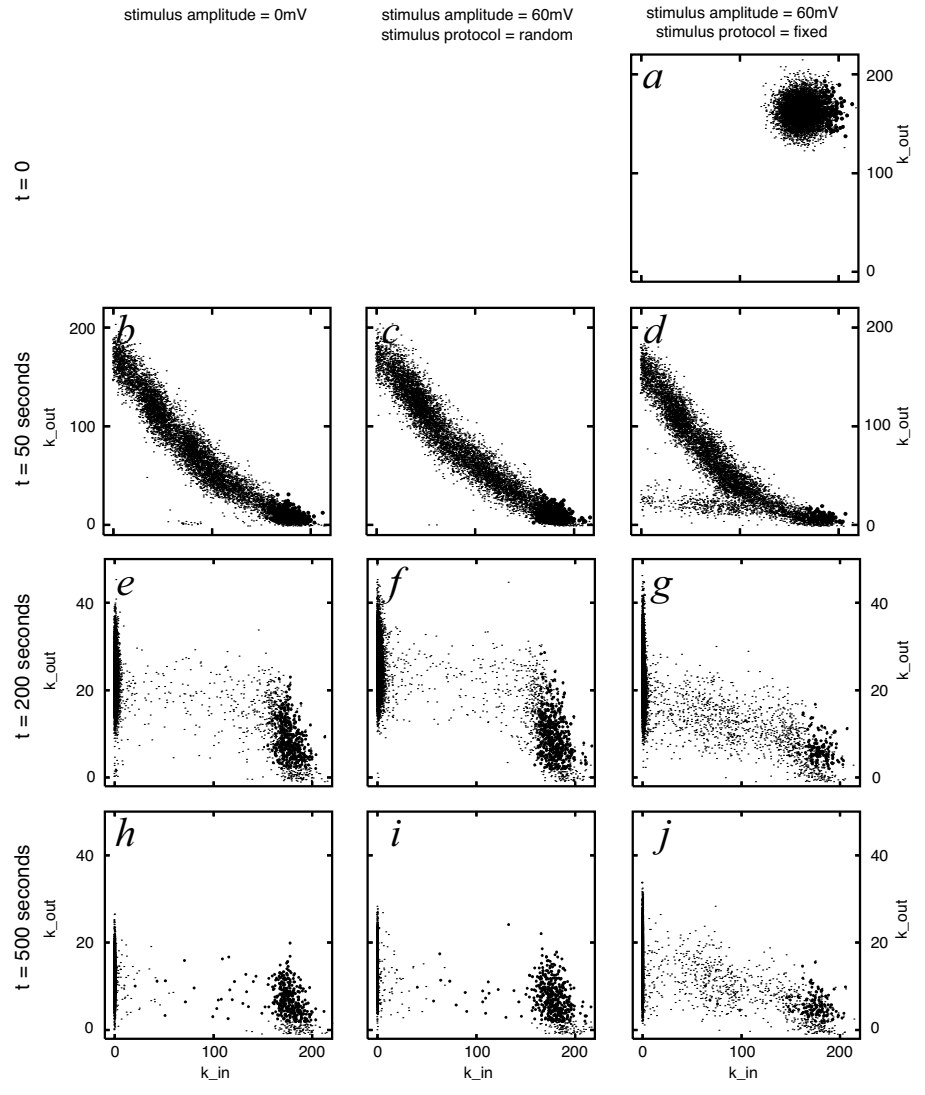

Fig. 4. Evolution of the out-degrees (k_{out}) vs. in-degrees (k_{in}): (•) *SI*-units; (·) *SI*-neighbour units. ***(b, e, h):*** in absence of any input (362 *SI*-units, 6,954 neighbours); ***(c, f, i):*** *random stimulation* protocol (425 *SI*-units, 6,996 neighbours); ***(a, d, g, j):*** *fixed stimulation* protocol (123 *SI*-units, 6,762 neighbours). ***(a):*** initial situation at $t = 0$ is identical for all three protocols; ***(b, c, d):*** at $t = 50$ seconds; ***(e, f, g):*** at $t = 200$ seconds; ***(h, i, j):*** at $t = 500$ seconds; Note the scale of vertical axis k_{out} is 200 in panels *(a-d)*, and 40 for panels *(e-j)*. Network size: [100 × 100]; stimulus duration: 100 ms; input units ratio: 10%.

In fact, the *SI*-units were characterized by an input pattern very close to that they had since the very beginning of the simulation. Different random seeds generated different populations of *SI*-units but the number of these units did not vary much as a function of the random seed.

6 Discussion

The main result has been to show that the application, in presence of background noise, of a recurrent pattern of stimulation let appear cell assemblies of excitatory units when associated to STDP-driven pruning. The vast majority of the connections that are modifiable by the spike-timing dependent plasticity rule were eliminated during the first thousands time steps of the simulation run [5]. Among the remaining active synapses, almost all were characterized by the highest possible activation level, in accordance with previous results [12].

We observed that the unsupervised pruning mechanisms tended to organize a feed-forward cell assembly of strongly interconnected units on top of the input units selected by the pruning process. Inhibitory responses observed in the pool of the *SI*-units are due to a balanced network reaction to the overall increased firing rate by increasing the activity within the pool of inhibitory units. The connectivity pattern of *SI*-units, initially set at random, appeared to match some requirements for maintaining almost all the input connections. The interpretation is that the cell assembly formed by the *SI*-units was initially determined by chance and when the pruning process started to select the active connections, these were maintained because of their connectivity pattern, thus letting emerge a particular circuit that was embedded in the network at time $t = 0$. However, the emergence of the diverging projections was much more difficult to observe than the convergence.

The self-organization of spiking neurons into cell assemblies was recently described in a study featuring large simulated networks connected by STDP-driven projections [6]. They studied the spatiotemporal structure of emerging firing patterns, finding that if axonal conduction delays and STDP were incorporated in the model, neurons in the network spontaneously self-organized into neuronal groups, even in absence of correlated input. The study [6] emphasizes the importance of axonal conduction delays that we did not initially consider in our model.

The choice of our neuromimetic model was justified by its compatibility with a novel hardware architecture [14]. Instead of leaky integrate-and-fire neuromimes, the use of biophysical models of neuromimes based on the Hodgkin-Huxley framework with multistate neurons and the associated multidimensional synapses [10] could bring better insight into the biological rationale of the emergence of cell assemblies by synaptic pruning.

Acknowledgments. The authors thank Dr. Yoshiyuki Asai for discussions and comments on the results and the manuscript. This work was partially funded by the European Community Future and Emerging Technologies pro-

gram, grant #IST-2000-28027 (POETIC), and Swiss grant OFES #00.0529-2 by the Swiss government.

References

1. Abeles, M.: Corticonics: Neural Circuits of the Cerebral Cortex. Cambridge University Press (1991)
2. Bi, G. Q., Poo, M. M.: Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci. **18** (1998) 10464-72
3. Chechik, G., Meilijson, I., Ruppin, E.: Neuronal Regulation: A Mechanism for Synaptic Pruning During Brain Maturation. Neural Computation **11** (1999) 2061–80
4. Eriksson, J., Torres, O., Mitchell, A., Tucker, G., Lindsay, K., Rosenberg, J., Moreno, J.-M., Villa, A.E.P.: Spiking Neural Networks for Reconfigurable POEtic Tissue. Lecture Notes in Computer Science **2606** (2003) 165–73
5. Iglesias, J., Eriksson, J., Grize, F., T., Marco, Villa, A.E.P.: Dynamics of Pruning in Simulated Large-Scale Spiking Neural Networks. Biosystems **79** (2005) 11–20
6. Izhikevich, E. M., Gally, J. A., Edelman, G. M.: Spike-timing Dynamics of Neuronal Groups. Cerebral Cortex **14** (2004) 933–44
7. Karmarkar, U. R., Buonomano, D. V.: A model of spike-timing dependent plasticity: one or two coincidence detectors? J Neurophysiol. **88** (2002) 507–13
8. Kelso, S. R., Ganong, A. H., Brown, T. H.: Hebbian synapses in hippocampus. Proc. Natl. Acad. Sci. USA **83** (1986) 5326–30
9. Mimura, K., Kimoto, T., Okada, M.: Synapse efficiency diverge due to synaptic pruning following over-growth. Phys Rev E Stat Nonlin Soft Matter Phys **68** (2003) 031910
10. Quenet, B., Horcholle-Bossavit, G., Wohrer, A., Dreyfus, G.: Formal modeling with multistate neurones and multidimensional synapses. Biosystems **79** (2005) 21–32
11. Rakic, P., Bourgeois, J. P., Eckenhoff, M. F., Zecevic, N., Goldman-Rakic, P. S.: Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. Science **232** (1986) 232–5
12. Song, S., Miller, K. D., Abbott, Larry F.: Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nature Neuroscience **3** (2000) 919–26
13. Song, S., Abbott, Larry F.: Cortical Development and Remapping through Spike Timing-Dependent Plasticity. Neuron **32** (2001) 339–50
14. Tyrrell, A., Sanchez, E., Floreano, D., Tempesti, G., Mange, D., Moreno, J.M., Rosenberg, J., Villa, A.E.P.: POEtic: An Integrated Architecture for Bio-Inspired Hardware. Lecture Notes in Computer Science **2606** (2003) 129–40