FACETS

FP6-2004-IST-FETPI 15879

Fast Analog Computing with Emergent Transient States

D15 –Equivalence report with the first generations of FACETS hardware from WP6 and WP7

Report Version: 1

Report Preparation: A. Destexhe, UNIC

Classification: PUBLIC

Contract Start Date: 01/09/2005 Duration: 4 Years

Project Coordinator: Karlheinz Meier (Heidelberg)

Partners: U Bordeaux, CNRS (Gif-sur-Yvette, Marseille), U Debrecen, TU Dresden, U Freiburg, TU Graz, U Heidelberg, EPFL Lausanne, Funetics S.a.r.l., U London, U Plymouth, INRIA, KTH Stockholm

Project funded by the European Community under the "Information Society Technologies" Programme

DELIVERABLE SUMMARY SHEET

Project Number: FP6-2004-IST-FETPI 15879

Project Acronym: FACETS

Title: Fast Analog Computing with Emergent Transient States

Deliverable N°: 15

Due date: September 2006

Delivery Date: October 12, 2006

Short description

We describe here the tests realized so far concerning the equivalence between specific integrated circuits to emulate neurons, and numerical simulations of the same models as those implemented on circuits. The two types of hardware developed in FACETS are described successively. First, we overview the analog circuits (ASICs) developed by the ENSEIRB partner, along with numerical simulations of the corresponding models. In the second part, we describe the circuits developed by the UHEI partner and their numerical counterpart.

It is important to note that both types of hardware are presently at the stage of prototype – a second generation is in progress and a second report about the equivalence between analog and numeric simulations is planned for the second project period (month 30).

Partners owning: UNIC, ENSEIRB, UHEI

Partners contributed: UNIC, ENSEIRB, UHEI

Made available to: PUBLIC

Part 1: ASIC circuits emulating Hodgkin-Huxley type neurons in real-time

1.1 Description of the ASICs and models

In this first part, we describe a platform for simulating networks of neurons based on the Hodgkin-Huxley (HH) formalism [1] and conductance-based synaptic interactions. The system is based on analog integrated circuits (ASICs) which solve the membrane equations of the neurons. Each ASIC neuron is equipped with a leak conductance, the I*Na* and I*Kd* voltage-dependent conductances for generating action potentials (HH model), a slow voltage-dependent K+ conductance for spike-frequency adaptation, and two conductance-based synaptic currents that implement kinetic models of glutamate and GABA receptors. The connectivity between the neurons is entirely managed digitally using a computer, which is interfaced in real-time (RT-LINUX) with the board containing the ASIC neurons. We demonstrate here the functionality of this system for small networks of excitatory and inhibitory neurons with excitatory synapses endowed with spike-timing dependent plasticity and fixed inhibitory synapses. We compare the behavior of the ASIC-based model with numerical simulations of the same models.

Models implemented. The neuron models implemented on circuits were Hodgkin- Huxley type representations of two main cell classes in cortex, the "regular spiking" (RS) neurons and "fast spiking" (FS) interneurons [2] (Fig 1). All neurons contain an adjustable leak conductance and voltage-dependent Na+ and K+ conductances necessary to generate action potentials. These currents define the FS phenotype for inhibitory neurons. Excitatory RS neurons possess in addition a slow voltage-dependent K+ current responsible for spike-frequency adaptation, which is adjustable. Synaptic interactions were described by conductance-based synaptic currents that implement kinetic models of glutamate and GABA receptors [3]. These models were conceived such as to handle any number and frequency of presynaptic signals, and therefore can represent multiple synapses. Plasticity rules were included at excitatory synapses, according to spike-timing dependent plasticity (STDP), in which the synaptic weight varies as a function of the relative timing of pre- and post-synaptic spikes [4]. The particular implementation of STDP included saturation terms which make the synaptic weight converge to a steady-state value (not shown). This type of convergence dynamics is particularly useful to compare the activity of analog networks (see description below), with the same networks simulated using traditional means (for this purpose we used the NEURON simulation environment [5]).

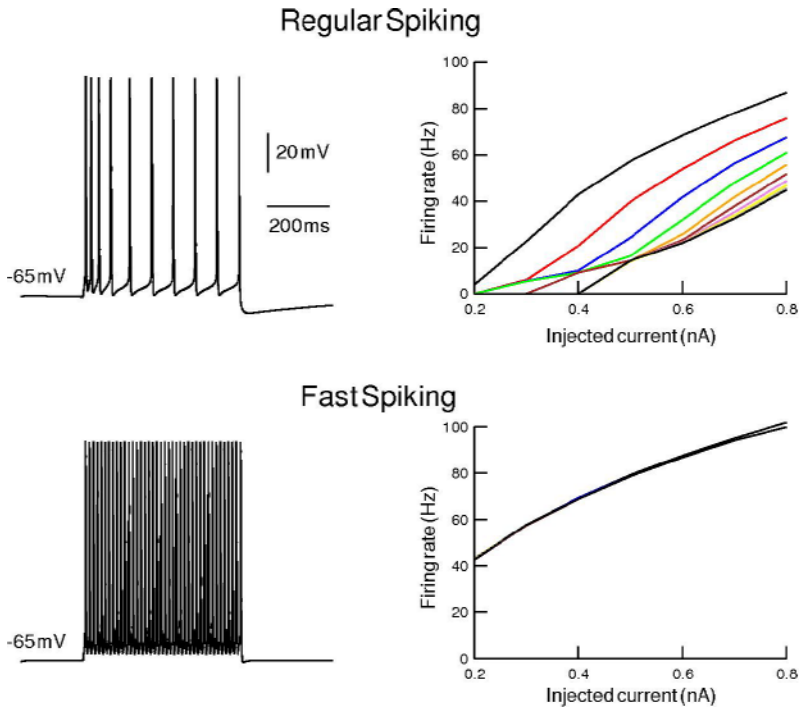

Figure 1: Hodgkin-Huxley type models of the two main types of cortical neurons implemented on ASIC circuits. Top panels: model of "regular spiking" excitatory neuron. The model contained the currents *INa* and *IK* responsible for spike generation, with an additional slow K+ current (*IM*) responsible for spike-frequency adaptation. These currents were simulated by Hodgkin–Huxley like models in NEURON. The model exhibited spike frequency adaptation following injection of depolarizing current pulses (left; 0.5 nA injected). The frequency–current (F/I) relation (right) represents the instantaneous firing rate (inverse of the interspike interval) as a function of spike number. Bottom panels: same paradigm for a model of "fast spiking" inhibitory neuron. Different colors indicate successive pairs of spikes (1st, 2nd, 3rd, . . . pairs are indicated in black, red, blue, green, orange, purple, brown, and yellow, respectively; note that colors are not visible in the bottom panel because the curves are nearly superimposed). This model contained only *INa* and *IK* (see details in text).

ASIC circuits. Specific analog VLSI circuits were designed and fabricated using a sub-micron BiCMOS (bipolar and CMOS) technology. The ASIC models for the first and second generation of ASICs are described in D18. In those devices, electronic currents modeling ionic currents of the HH formalism are produced by an analog modular circuit and summed on a capacitance. The voltage on this capacitance is equivalent to the neuron membrane voltage. Each module of the circuit, designed in current-mode, computes a mathematical function present in the RS and FS cells models description [6]; electronics modules are dimensioned at the transistor level and arranged to compute the ionic current models (4 types for the RS neurons, 3 for the FS neurons). An identical 2 arrangement is made to calculate on-chip the kinetic synapse currents, injected on the membrane capacitance. Kinetics and time constants of the original models are reproduced with no change in time scale, which imposes that the VLSI neurons run in biological real-time. This design principle and the electronics modules for neuromorphic devices have been validated in previous applications [7,8]. The devices we used here are specifically-designed prototypes, in which specific I/O have been added to allow the control of synaptic weights by STDP plasticity rules: the neuron activity is available in the form of a time-stamped event (1 bit). The kinetic synapse inputs (one excitatory and one inhibitory) are digitally coded using a simple width modulation for the weights.

Analog-digital interface. The connectivity is handled using a computer, via a PCI interface board supporting the ASICs and an FPGA to control digital inputs and outputs from and to the analog neurons (see scheme in Fig. 1A). The computer runs a LINUX operating system in soft real-time mode, and a specific software interface. The computer reads in real-time the

spike events from the board, computes synaptic plasticity algorithms, and dispatches the spike signals as inputs to the ASIC neurons, where they will trigger synaptic conductance changes. This way allows us to implement arbitrary schemes of connectivity and plasticity, since it is entirely managed by the computer, providing a great flexibility to the system.

1.2 Equivalence between analog and numeric simulations

We first characterized the individual ASIC neurons based on their frequency/current (f/i) curves obtained by constant current injection. The measurements were performed for both FS and RS neurons using different parameters, such as different levels of adaptation. The f/i curves were also computed from computer simulated neurons, and despite neuron-to-neuron variations in the ASIC circuits (the ASIC full-custom fabrication process results in circuits dispersion and mismatch at the transistor level), the agreement was excellent (Fig 2).

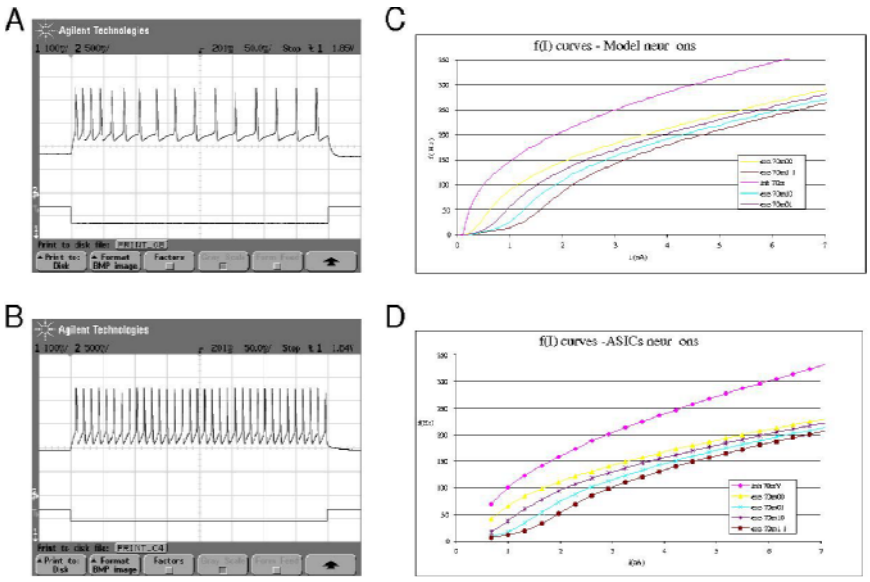

Figure 2. Single-neuron measurements on the ASIC analog circuits. A. Regular spiking (RS) cell: ASIC measurements showing the response to depolarizing current pulse with spike-frequency adaptation. The membrane potential was measured using an oscilloscope, and the current pulse was delivered using a current generator. B. Fast spiking (FS) cell: same as in A, showing the absence of spike-frequency adaptation in FS neurons. C. Model simulation of the steady-state frequency-current relation obtained for different maximal conductances of the slow potassium current responsible for spike-frequency adaptation. C. Same parameter settings simulated on the analog circuits.

A second comparison was to measure the frequency response of the ASIC neurons in response to synaptic noise (mixture of Poisson trains of excitatory and inhibitory events). In this case, ASIC and model neurons displayed similar responses, as shown by the raster-plots in Fig. 3. However, the exact timing of spikes of the model was in general not reproduced by the ASICs (due to dispersion of parameters). This difference is visible in the raster-plots in Fig. 3, but it was also quantified by calculating the crosscorrelation between model and ASICs spike trains (for the same injected noise). Correlations were broadly peaked (half-width of about 3 sec) and weak (peak included between 0.01-0.02). Despite these differences in spike timing, the frequency response of model and ASICs was remarkably similar (Fig. 3, right panels).

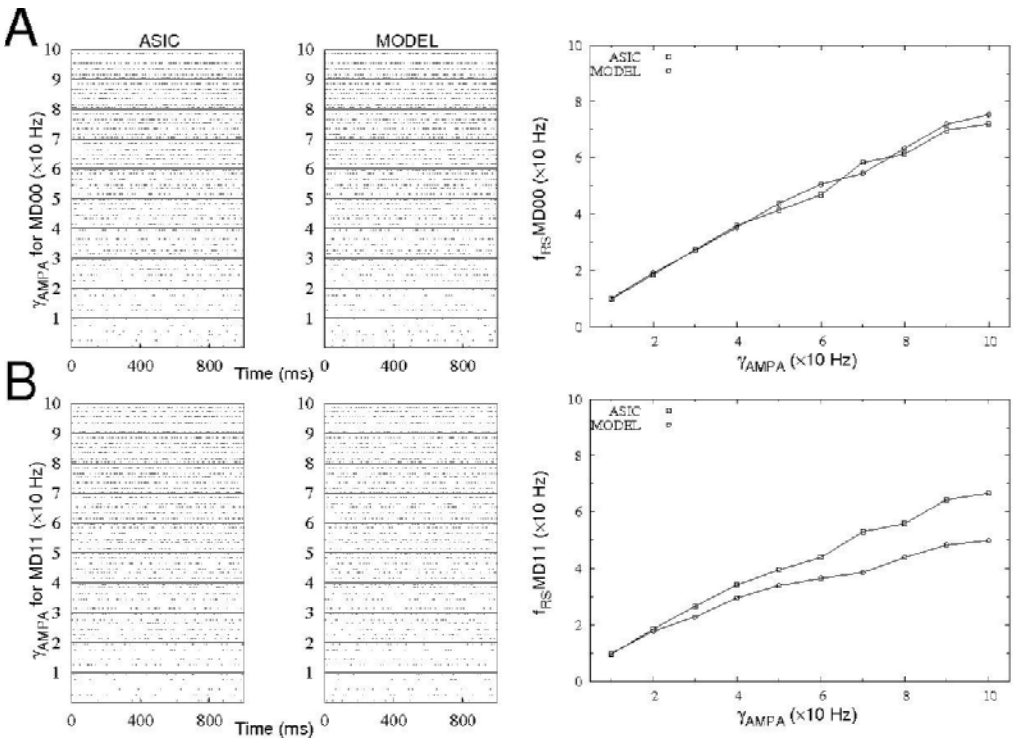

Figure 3: Firing properties of a single neuron bombarded by random synaptic inputs. A. Neuron 4 on board was bombarded by Poisson-distributed random excitatory (AMPA) synaptic inputs at different firing rates, with maximal conductance of 100 nS. Neuron 4 was setup as a "weakly adapting" RS cell (gM = 45.5 μS/cm2; MD00 on board). Both ASIC (left) and model (middle) simulation were run for 10 seconds under each bombardment, and split into 10 traces to form a raster plot against time. Right panel: firing frequency as a function of input rate. The firing frequency was averaged by spike counting during the simulations, and represented as a function of the mean input rate of the synaptic bombardment. The frequency response of ASIC (circles) was very close to that of the model (squares). B. Same simulation as A, with Neuron 4 set as a "strongly adapting" RS cell (gM = 181.8 μS/cm2; MD11 on board). In this case, the frequency response of the ASICs (right panel, circles) was higher than that of model (squares).

A more severe test of the ASIC circuits was to build a network of neurons in which the synaptic weights are subject to changes according to a well-defined plasticity rule. We used a network of 6 RS and 2 FS neurons, interconnected with excitatory (AMPA) and inhibitory (GABA) recurrent synapses, as shown in Fig. 1B. We used a paradigm in which neurons were subject to random synaptic inputs in addition to recurrent synapses from other neurons in the network. STDP [4] was applied to the recurrent excitatory synapses (from RS to FS cells), and their steady-state synaptic weight is represented as a function of the input rate for both ASIC (Fig. 1E) and model simulations (Fig. 1F). Although the synaptic weights in the ASIC network were generally more variable (presumably due to ASICs' fabrication – see above), the agreement between the ASIC and model neurons was excellent, for the whole range of inputs considered. In this case, the run time of the simulation was significantly faster on ASICs (real time) compared to PC-based workstations (about 5 times slower than real time).

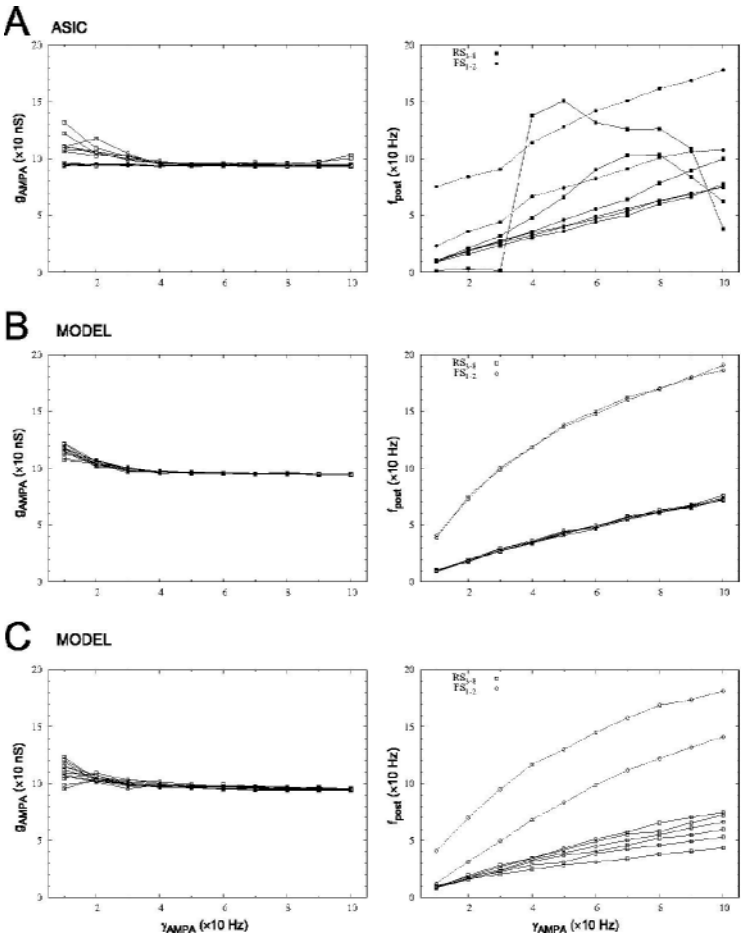

Figure 4. Steady-state synaptic weights and firing rates in ASIC and model 8-neuron circuits. A 8-neuron circuit was constructed from 6 weakly-adapting RS cells, 2 FS cells, all interconnected. STDP was present at excitatory synapses from RS to FS cells. Each cell was subject to random synaptic activity, with a mean excitatory input rate which was varied from 10 to 100 Hz, while the inhibitory input rate was fixed at 20 Hz. STDP was incorporated at 12 AMPA synapses during 50 seconds simulation time. After transients, the steady-state synaptic weight and firing rate was calculated at each synapse by averaging over the last 30 seconds. A. Steady-state synaptic weight (left) and rate (right) exhibited by ASIC circuits, for different values of the excitatory input rate *γAMPA*. B. Same description for a model circuit with the same values of parameters, and in which all RS neurons were identical. C. Model in which the properties of RS and FS neurons were dispersed in order to create cell-to-cell variability comparable to ASIC circuits.

1.3 Future simulations

The benchmark tests are currently limited by the small quantity of artificial neurons available on the ASICs. A new generation of ASICs has been designed and validated after fabrication (see D18). Each ASIC holds up to 5 neurons with 3 to 5 HH conductances. To optimally handle those ASICs, a new simulation setup is under development. Although its architecture is similar to the one presented in this deliverable, new features will be implemented so that the system will be able to simulate in real time up to 200 artificial neurons organized in a plastic network, with individually configurable adaptation algorithms on each synapse.

Part 2: Analog circuits emulating integrate-and-fire neurons with high speed

2.1 Description of the ASICs and models

In this second part we describe the VLSI model of a spiking neural network realizing an area-efficient mixed-signal implementation of synapse-based long term plasticity [14,15]. The artificial synapses are based on an implementation of STDP. In the biological specimen, STDP is a mechanism acting locally in each synapse. The presented electronic implementation succeeds in maintaining this high level of parallelism and simultaneously achieves a synapse density of more than 9k synapses per mm^2 in a 180 nm technology. This allows the construction of neural micro-circuits close to the biological specimen while maintaining a speed several orders of magnitude faster than biological real time. The large acceleration factor enhances the possibilities to investigate key aspects of plasticity, e.g. by performing extensive parameter searches.

Neuron model implemented. The ASIC membrane potential V is governed by the following differential equation:

$$c_m \frac{dV}{dt} = g_m(V - E_l) \quad + \sum_k p_k(t) g_k(V - E_x) + \sum_l p_l(t) g_l(V - E_i)$$

Each term on the right hand side contributes an individual current to the total membrane current, which by itself is equal to the derivative of the membrane potential multiplied by a constant c_m The first term models the contribution of the different ion channels that determine the potential E_l the membrane will eventually reach if no other currents are present. The synapses use different reversal potentials, E_i and E_x to model inhibitory and excitatory ion channels. The index k in the first sum runs over all excitatory synapses while the index l in the second covers the inhibitory ones. The individual activations of the synapses are controlled by the parameters $p_{k,l}(t)$. Plasticity is included in the model by varying g_k and g_l slowly with time. The synaptic weight $\omega_{k,l}(t)$ denotes the relative synaptic strength at a given time t:

$$g_{k,l}(t) = \omega_{k,l}(t) \cdot g_{\max k,l}$$

ASIC circuits. Fig. 5 shows the operation principle of the synapse and neuron circuits. The synapses form an array of 256 rows x 384 columns below which 384 neurons are located. Each neuron contains a capacitance C_m that represents the membrane capacitance. Three different conductances model the different ion channel currents. The membrane leakage current flows through g_{leak}. It can be individually controlled for each neuron. The leakage reversal potential E_l, the excitatory and inhibitory reversal potentials of the synapse conductances g_x and g_i as well as the threshold and reset voltages V_{th} and V_{reset} can be set for groups of 64 neurons each.

Each neuron receives its input signals from one column of the synapse array. Two separate conductances are connected to the membrane capacitance inside the neuron circuit, one representing the excitatory and one the inhibitory synapses' ion channels. Each is controlled by the sum of the currents generated by the active synapses located in the respective synapse column. A third conductance models all ion channels contributing by their respective leakage currents to the neuron's resting potential E_l.

In contrast to a biological neuron the axon of its VLSI counterpart is electrically isolated from its input. It carries a digital signal that encodes the exact time of each spike's occurrence by its

rising edge. This signal is also routed back along the same column of synapses that comprises the neuron's input. This allows the STDP circuit located inside each synapse to measure the time between a pre-synaptic pulse and a post-synaptic spike.

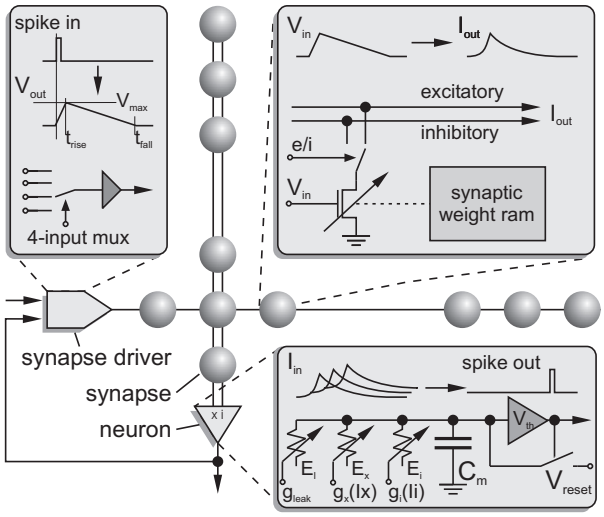

Figure 5: Operating principle of the spiking neural network. The three boxes show the signal processing done by synapse drivers, synapses and neurons.

STDP model. The correlation measurement for STDP is part of every synapse. It is based on the biological mechanism as described in [4,13]. For each occurrence of a pre- or post-synaptic action potential the synapse circuit must change the synaptic strength by a factor of $1+F(\Delta t)$. F is called the STDP modification function [13] and is defined as follows:

$$F(\Delta t) = \begin{cases} A_+ \exp(\frac{\Delta t}{\tau_+}) & \text{if } \Delta t < 0 \quad \text{(causal)} \\ -A_- \exp(-\frac{\Delta t}{\tau_-}) & \text{if } \Delta t > 0 \quad \text{(acausal)} \end{cases}$$

Experimental data suggest a value of about 20 ms for the time constants for causal, τ_+, and acausal, τ_-, events. The parameters A_+ and A_- have also been determined experimentally by dividing the total modification of the synaptic strength measured for multiple spike pairs by the number of pairs. Fig. 6 shows the causal modification function for $A = 0.005$.
The synapse strength ω changes with each pre- or postsynaptic action potential according to:

$$\omega_{\text{new}} = \omega_{\text{old}}(1 + F(\Delta t))$$

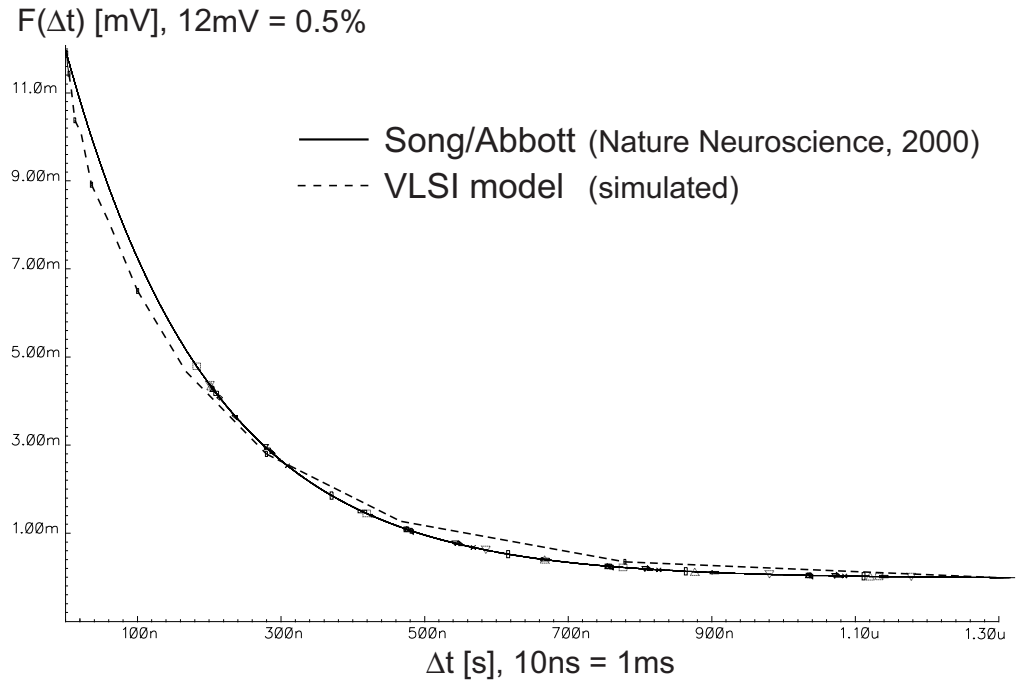

Figure 6: Comparison between the measured modification function from [13] and the presented VLSI model.

2.2 Methods for efficient comparison between hardware and software simulations

Within the FACETS project, a common unified procedural language for the description of neuron-scientific experiments on both software and hardware platforms is under development. This language is based on the programming language *Python* and will provide the possibility to set up a neural network simulation experiment in only one way and then execute it on multiple platforms, i.e. different software simulators and different hardware setups available. For both the simulator NEST and the hardware system described here, a Python interface already has been developed and is under further development. Both are ready to be integrated into a high-level Python language which also already is in work.
For the equivalence check method presented below it was the first time members of the FACETS consortium were able to make use of this approach by easily sharing the externally generated input data and by using the same analysis and plotting tools for both software and hardware simulations.

2.3 Equivalence between analog and numeric simulations

Since the production of a first prototype of this hardware was finished just recently, it still is in a testing phase and not ready to be operated completely. Calibration routines that will balance out the chip-inherent differences between the behavior of single neurons and synapses due to hardware process variations and parasitic effects still have to be established for the given concrete system. Specifications of neurons, synapses, of the communication system, of the embedding hardware and of chip specific noise sources have to be retrieved by extensive measurements which are in progress. Thus, for this report we can only show the establishment of *methods* for a future serious equivalence check between the ASIC and numerical software simulations. It is exemplarily shown in terms of membrane potential traces, but it is also already set up for output firing rate comparison and can be easily applied and extended to all kind of spike-based analysis (see next subsection). We also show a first fit of the measured hardware STDP functionality to data given in literature.

Membrane potential trace. For the given system, frequency/current curves as described for the first hardware system (see section 1.2) cannot be measured due to the fact that only externally generated spikes can be fed into the ASIC, but no currents. Thus, for a single arbitrarily selected neuron (neuron number 4 in the low level hardware enumeration scheme) the following configuration was set up:
The neuron receives 128 excitatory and 128 inhibitory inputs. The synaptic time constants are selected to be equal for both input types since a differentiated configuration of these parameters is not supported yet (but will be after more specification work). To get a similar effective excitatory and inhibitory input, the weights are also set to equal values for both types of synapses. For each input a randomly generated Poisson distributed spike train is applied. For all inputs, a fixed rate of 3 Hz (simulated biological time) is selected.

In a first experiment the neuron's analog membrane potential trace is recorded for a period of 10 seconds of simulated real time, which is achieved by a hardware operation real time of approximately 1 msec. The simulation speedup factor was estimated by determining the length of an excitatory postsynaptic potential measured on the chip and claiming it to correspond to a biological real time of 5 msec. The value retrieved by this method is a speedup factor from hardware to biology of 10^4. It has to be pointed out that this is only a very coarse estimation which will be specified further in the near future.

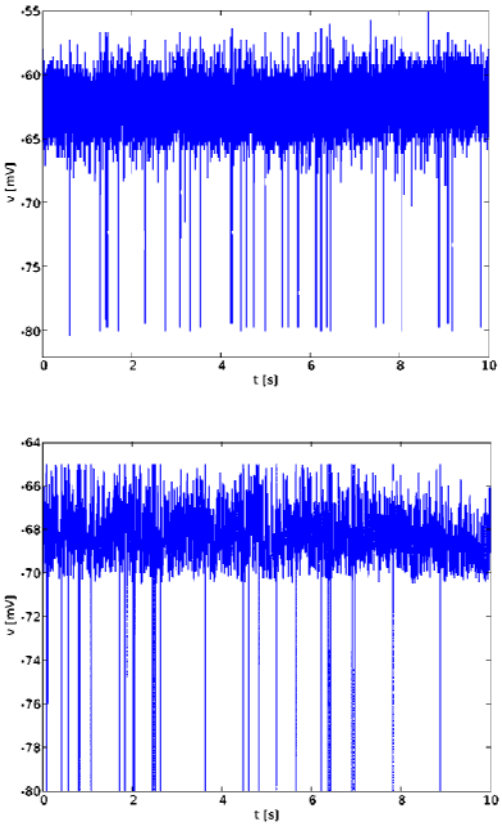

Figure 7: Direct comparison of a single neuron's membrane potential trace under Poisson bombardment. Left: FACETS Stage 1 prototype hardware. Right: NEST software simulation. For further parameter values see text.

The membrane potential trace is then compared to a pure software simulation using the NEST simulator, implementing the integrate-and-fire model described above with basically the same parameter setup as for the hardware (see Fig. 7). The input for the synapses is generated using numerical extensions for the language Python and thus is interpretable by both the Python based interface to the hardware and by the NEST simulator. The procedural descriptions for the experiments on both platforms are written in Python and need only a small step to be

integrated into the unified common experimental description language currently developed within FACETS.

Note again that the purpose of this comparison is just to illustrate the established method and that no effort was taken to fit the graphs since the system is not calibrated yet. Note also that for both plots the action potential (spike) itself is not modeled, because its shape is stereotypic and thus only its exact time of occurrence has to be recorded. Each firing of an action potential can be detected on the graphs by a strong subsequent negative peak due to the immediate jump of the membrane potential to some reset voltage. By regarding this, a qualitative similarity between hardware and software trace can be detected in terms of the neuron's output spiking frequency.

STDP. Fig. 8 shows first measurements of the STDP modification function from the fabricated chip. Due to the early state of the test setup these measurements are limited to a single synapse, located at row 252 and column 66 in the left half of the chip.
Pairs of pre- and post-synaptic spikes with a given time difference Δt were sent into the chip. After each pair the voltage on C_2 was compared against a threshold voltage of 50 mV. The number of pairs necessary to cross this threshold was recorded. This process was repeated 10 times for each time difference Δt. After each data point Δt was incremented by 5 ns.

The step-like behavior of the activation function for small values of Δt, as it is visible in Fig. 8, is an artefact of this measurement procedure, since for small time differences only a few pulse-pairs are necessary to cross the 50 mV threshold. This results in a rather large quantization error inversely proportional to Δt. Since the threshold voltage will be increased in the final setup, this effect can be greatly reduced.

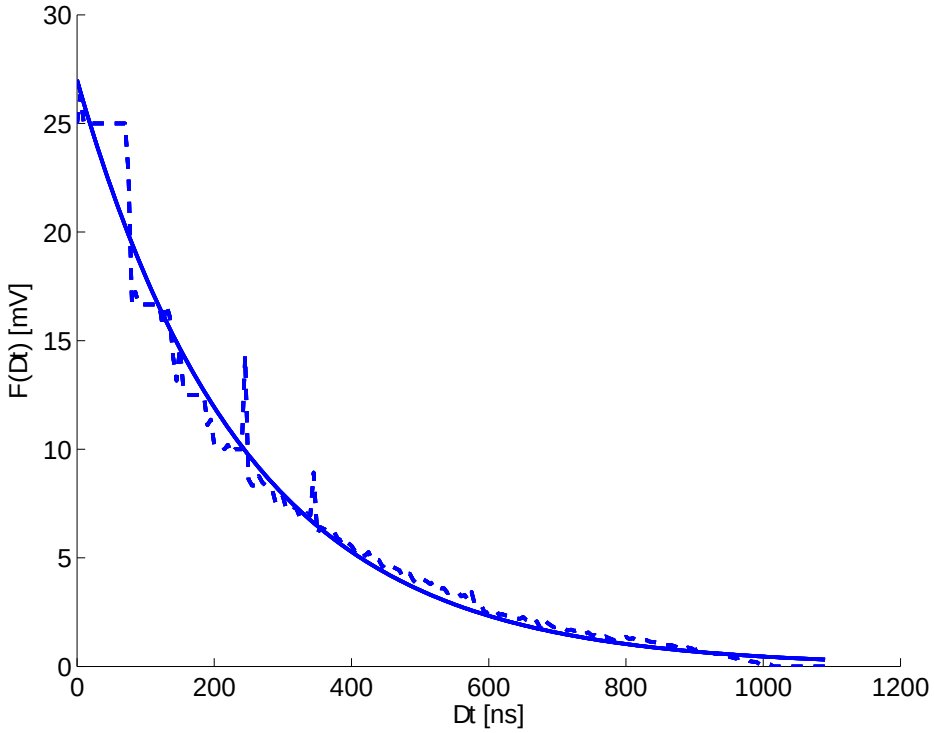

Figure 8: Measurement of the STDP modification function for a single, arbitrary synapse (dashed line). The theoretical model is shown as a reference (solid line). The parameters were set as follows: A=27 mV, Δ t=245 ns.

References

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

[2] B.W.Connors and M.J.Gutnick, Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* **13** (1990) 99-104.

[3] A.Destexhe, Z.F.Mainen and T.J.Sejnowski, An efficient method for computing synaptic conductances based on a kinetic model of receptor binding. *Neural Comput.* **6** (1994) 14-18.

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

[5] M.L.Hines and N.T.Carnevale, The Neuron simulation environment. *Neural Comput.* **9** (1997) 1179-1209.

[6] S.Saighi, J.Tomas, L.Alvado, Y.Bornat and S.Renaud, Integration of Biological Neurons Models, XVIII Conference on Design of Circuits and Integrated Systems – DCIS2003, Ciudad real (Spain), (2003).

[7] S.Saighi, J.Tomas, Y.Bornat and S.Renaud, A Conductance-based silicon neuron with dynamically tunable model parameters, 2nd International IEEE EMBS Conference on Neural Engineering, Arlington (VA), USA, March 16-19 2005, ISBN 0-7803-8710-4, (2005) pp. 285-288.

[8] S.Renaud-Le Masson, G.Le Masson, L.Alvado, S.Saighi and J.Tomas, A neural simulation system based on biologically-realistic electronic neurons. *Information Science.* **161** (2004) 57-69.

[9] L.Carota, G.Indiveri and V.E.Dante, A software-hardware selective attention system, *Neurocomputing* **58-60** (2004) 647-653.

[10] M.Simoni, G.Cymbalyuk, M.Sorensen, R.Calabrese and S.DeWeerth, A Multiconductance silicon neuron with biologically matched dynamics. *IEEE Transactions on Biomedical Engineering* **51** (2004) 342-354.

[11] G.LeMasson, S.Renaud-LeMasson, D.Debay and T.Bal, Feedback inhibition controls spike transfer in hybrid thalamic circuits. *Nature* **417** (2002) 854-858.

[12] Q.Zou, Y. Bornat, S.Saighi, J.Thomas, S.Renaud and A.Destexhe. Analog-digital simulations of full conductance-based networks of spiking neurons with spike timing dependent plasticity. *Network,* **in press** (2006)

[13] S. Song, K. D. Miller, and L. F. Abbott. Competitive hebbian learning though spike-timing-dependent synaptic plasticity. *Nature Neuroscience* **3** (2000), 919–926

[14] Schemmel, J. and Gruebl,A. and Meier, K. and Mueller, E., Implementing Synaptic Plasticity in a VLSI Spiking Neural Network Model. *Proceedings of the 2006 International Joint Conference on Neural*
Networks (IJCNN'06) (2006)

[15] Schemmel, J. and Meier, K. and Mueller, E., A New VLSI Model of Neural Microcircuits Including Spike Time Dependent Plasticity. *Proceedings of the 2004 International Joint Conference on Neural*
Networks (IJCNN'04) (2004),), IEEE Press, 1711-1716