FPGA Implementation of the Huber-Braun Neuron Model

Marcel Beuler1, Alexander Krum2, Werner Bonath3 and Hartmut Hillmer1

1Department of Electrical Engineering and Computer Science, University of Kassel, Wilhelmshöher Allee 71-73, D-34121 Kassel, Germany
2School of Computer Science and Engineering, University of New South Wales, NSW 2052, Sydney, Australia
3Department of Electrical Engineering and Information Technology, University of Applied Sciences, Wiesenstr. 14, D-35390 Giessen, Germany

Keywords: Huber-Braun, Neuronal Network, FPGA, VHDL.

Abstract: The Hodgkin-Huxley model (HH) describes the initiation and propagation of action potentials in neurons closed to the biological conditions, but it is not well suited for large scale simulation of neuronal networks. In this paper, an implementation of the Huber-Braun model is presented. It is a simplified HH-type model and able to reproduce a wide variety of spiking patterns. An FPGA is selected as a reconfigurable hardware implementation platform to simulate the network functionality of the neurons. The 32-bit floating-point format and computation techniques (i.e. CORDIC) instead of LUTs are used to avoid loss of physiological information. We validated our design with a C++ program and report the synthesis result based on Xilinx Virtex 6 FPGA.

1 INTRODUCTION

In physiological research computer simulations play an important role in analyzing the neuronal information processing in the central nervous system. Basic elements are mathematical descriptions of neurons and synapses with different complexity. Hodgkin-Huxley-type neuron models generate action potentials by voltage-gated ion channels, therefore, they closely follow the biological concept. The original approach has four nonlinear differential equations and postulates three gating variables to model the dynamics of a sodium and potassium channel (Hodgkin and Huxley, 1952). Its solution requires a great deal of computing power and is thus not well suited for larger neuronal networks. That’s why several simplified models with reduced computing effort have been developed in the past, and their use depends on each specific problem (Izhikevich, 2004).

For example, the Hindmarsh-Rose model has three nonlinear differential equations and can be used to investigate the spiking-bursting behavior of two electrically coupled neurons (Xia and Qi-Shao, 2005). The FitzHugh-Nagumo model approximates the HH system by only two differential equations (FitzHugh, 1961), but it doesn’t generate bursts, which can be important for neuronal transmission (Izhikevich et al., 2003). One of the simplest models of a neuron’s electrical properties is the Integrate-and-Fire model. It is based on a threshold approach instead of generating action potentials and allows large network simulations, but the model is unrealistic from a physiological point of view (Gerstner et al., 2014). In this paper, we have chosen the Huber-Braun neuron model which is also a simplified HH-type model. With tonic, bursting, and chaotic spike patterns it can reproduce a large amount of neuronal activity. In contrast to the Hindmarsh-Rose or FitzHugh-Nagumo model, each parameter has a clearly defined biological correlate.

In general, analog and digital approaches can be used to implement neuron models. The latter are more popular as they have the advantage of higher accuracy, lower noise sensitivity, better testability, higher flexibility, etc. (Muthuramalingam et al., 2008). In this context, an analog circuit of a Huber-Braun neuron has problems with the first period-doubling bifurcation (Hermida et al., 2012).

Target hardware for digital implementations can be conventional CPUs, GPUs, ASICs, and FPGAs. Conventional CPU-based execution is much slower in comparison to specialized hardware because of its serial nature. GPUs can exploit the parallelism of neuronal networks better and can be a powerful alternative to general purpose processors. However, they may not be able to meet real-time requirements due to high rates of data exchange between the neurons (Du Nguyen, 2013). ASICs provide the best performance regarding chip area and clock frequency, but they are expensive and their functionality cannot be changed after chip manufacturing. Although FPGAs are not
as powerful as ASICs, they have the great benefit of reconfigurability during prototyping.

Several FPGA implementations of the HH neuron model have been implemented so far (Graas et al., 2004); (Zhang et al., 2009); (Bonabi et al., 2014). In this paper we present an FPGA-based processor architecture for the Huber-Braun neuron model. It is programmed in VHDL and can calculate up to 1600 neurons including electrical coupling in real-time on the Xilinx ML605 development board at 200 MHz clock frequency. To validate the design, two coupled neurons as well as a 20x20 network are tuned to different synchronization states, and these results are compared with a C++ simulation.

2 NEURON MODEL

The four-dimensional HH equations with transfer rates can be converted in a simplified HH model without the need for a power function and, finally, in a two-dimensional system as shown in (Postnova et al., 2012). These simplifications are acceptable, because the exact form of an action potential is less important than the modulation of the fire rate and impulse patterns. To consider subthreshold membrane potential oscillations which operate independently of the spike generation, the two-dimensional system is extended by two additional ion channels.

Within the model, spike generation is done by an HH-type component with a fast depolarizing sodium ion current (index d) and a fast repolarizing potassium ion current (index r) with voltage- and time-dependent conductances g_d and g_r. A slow depolarizing sodium ion current (index sd) and a slow repolarizing potassium ion current (index sr) with voltage- and time-dependent conductances g_sd and g_sr are responsible for subthreshold oscillations. Finally, a leakage current (index l) already known from Hodgkin-Huxley is included. Figure 1 (A) shows the equivalent circuit diagram of the neuron model with all ion currents as well as two current sources for an additional noise level J_noise and the synaptic process J_ext. The latter one consists of a current injected externally in the cell (J_m) and a gap-junction current (J_gap) which is explained further on.

A positive hyperpolarizing current J_ext > 0 decreases the fire rate, since the membrane potential gets more negative. With the current-voltage-relation of the membrane capacity and Kirchhoff’s Law we get Eq. (1). Leakage current and non-linear currents are modeled as Eq. (4) and Eq. (5), with g_i the conductances, a_i the voltage- and time-dependent activation variables, and V_i the related Nernst potentials.

\[
C_m \frac{dV}{dt} = - \sum J_m - J_{ext} + J_{noise} 
\]

\[
\sum J_m = J_d + J_r + J_{sd} + J_{sr} 
\]

\[
J_{ext} = J_m + J_{gap} 
\]

\[
J_i = g_i(V - V_i) i = \{d,r,sd,sr\} 
\]

Similar to the HH gating mechanism, the activation variables of the ion channels are described by differential equations first order. Steady-states are modeled by a sigmoid curve and time constants replace the voltage-dependent delay functions. The fast sodium channel J_d is instantaneous active, so a_d = a_d∞ can be estimated. Furthermore, J_d is directly coupled with a_{sr} in Eq. (9) and the related steady-state activation a_{sr∞} is omitted. Both factors \(\rho\) and \(\varphi\) are responsible for temperature scaling of conductances g_i and time constants \(\tau\), with \(T_0\) the reference temperature in °C.

\[
a_{sr∞} = \frac{1}{1 + \exp[-s_i(V - V_{th})]} i = \{d,r,sd\} 
\]

\[
a_d = a_d∞ 
\]

\[
d a_i \frac{dt}{\tau_i} = a_{i∞} - a_{i} \quad i = \{r,sd\} 
\]

\[
d a_{sr} \frac{dt}{\tau_{sr}} = -\eta \cdot J_{sd} - k \cdot a_{sr} 
\]

\[
g_i = g_0 \cdot \rho = g_0 \cdot 1.3(T - T_0)/10^°C 
\]

\[
1 - \frac{\varphi}{\varphi_0} = \frac{3.0(T - T_0)/10^°C}{\tau_0} 
\]

Network simulations are performed by bidirectional gap-junction coupling via J_gap with g_gap the coupling strength, V_{i,j} the membrane potential of an individual neuron at the position (i,j) in Figure 1 (B), and V_{i+n,j+m} the membrane potentials of neighboring neurons. The summation is taken over all pairs (m,n) with m,n \in \{-1,0,1\} (Postnova et al., 2009):

\[
J_{gap,i,j} = \sum_{m,n} g_{gap}(V_{i,j} - V_{i+n,j+m}) 
\]

Normally, border neurons are coupled with neurons from the opposite border to get a torus-like network:

\[
i + n = -1 \rightarrow i + n = 39 \quad j + m = -1 \rightarrow j + m = 39 
\]

\[
i + n = 40 \rightarrow i + n = 0 \quad j + m = 40 \rightarrow j + m = 0 
\]

All differential equations are solved numerically by using the Euler integration method with \(\Delta t = 0.1 \text{ ms}\):
In Eq. (13), \( g_W \) is a white Gaussian noise process given by the Box-Mueller transform (Fox et al., 1988), where \( a \) and \( b \) are two uniformly distributed random numbers in the interval \([0,1]\). The noise intensity can be adjusted with the parameter \( d \).

\[
\begin{align*}
V(t + \Delta t) &= V(t) + \frac{\Delta t}{C_m} f(V(t)) + g_W \\
\phi_i(t + \Delta t) &= \phi_i(t) + \frac{\Delta t}{\tau_i} (a_{\text{in}}^{(i)} - a_i^{(i)})
\end{align*}
\]  

(13)  

(14)

Analyzing the distribution of phase differences between spike times enables us to distinguish between different states of synchronization in a system of especially two coupled neurons (Postnova et al., 2007). A spike time is taken when the membrane potential exceeds a \(-35 \text{ mV}\) threshold. The phase difference is calculated by Eq. (16), where \( t_1 \) and \( t_2 \) are the times of subsequent spikes of the first neuron and \( \tau \) is the spike time of the second neuron (Pikovsky et al., 2001):

\[
\Delta \phi = 2\pi \frac{t_2 - t_1}{\tau_2 - \tau_1} \quad t_1 < \tau \leq t_2
\]

(16)

Typical values of the parameter in the above equations can be taken from (Postnova et al., 2009).

3 PROCESSOR ARCHITECTURE

Figure 2 shows the basic design of the implemented architecture with two separated modules for the Huber-Braun model equations and the electrical synapses. It contains all necessary hardware to compute 1600 neurons including gap-junction coupling in real-time at 200 MHz clock rate. Operations of the Huber-Braun model are allocated to 9 pipelined arithmetic units (AUs), where “ADD/SUB” is a combined 12-stage adder/subtractor and “MULT” an 11-stage multiplier. All the functions (\( \ln, \cos, \sqrt{\text{t}}, \text{and exp} \)) and, in contrast to (Bonabi et al., 2014; Zhang et al., 2009), the divisions are calculated by the iterative CORDIC algorithm (Walther, 1971). The remaining arithmetic is designed to ensure an optimum utilization of the 50-stage CORDIC pipeline. To solve the limited convergence domain problem of the algorithm we use a unified division-free argument reduction method proposed in (Hahn et al., 1994).

Each AU processes 32-bit floating-point values to avoid undefined effects due to significant loss of physiological information (Zhang et al., 2009). It has an own program memory for individually pro-
programming of the operations, up to four FIFOs to store the computed results and a ROM (a RAM used as ROM after initialization) to access model-specific parameters like \( g_l \) and \( V_{th} \). Two adders/subtractors also have a RAM for the temporary storage of model variables. These variables are the membrane potential \( V(i,j) \) (RAM of \( \text{ADD}_A \)) and three activation values \( a_i \) (RAM of \( \text{ADD}_B \)), see Figure 2. \( \text{OP}A \) and \( \text{OPB} \) represent both operands for each AU selectable by the machine instructions. To ensure an efficient programming, the RAM content of \( \text{ADD}_A \) is provided simultaneously to a second \( \text{ADD}_A \) module. FIFOs can store calculated results in cases where a subsequent operation on the same or another AU cannot be started immediately. First of all, each stored FIFO result is available as an operand for the own AU as well as all the others. After the arithmetic operations are allocated to the individual AUs, this wiring complexity is reduced to the required connections in order to meet the timing specification. “PRNG” generates pseudo random numbers via linear feedback shift registers. Here, a program memory determines how many random numbers are generated per time interval.

At 200 MHz clock rate, the step size of 0.1 ms is divided into 20 identical subneuron cycles with 1000 clocks in each case. Furthermore, all subneuron cycles are divided into 2 blocks with 500 clocks as the basic execution time, i.e. all program memories with a size of \( 2^5 = 512 \) cells can store up to 500 executable instructions. In such a block the operations of 40 neurons are allocated to the existing hardware components. During a neuron cycle each program memory is run through forty times with actualized index-registers to access variables, so that \( 40 \cdot 2 \cdot 20 = 1600 \) neurons per core are computable in real-time.

Before a simulation can start, a host software has to generate the machine instructions for each AU as well as the ROM contents and initial states for all variables by means of the user settings. This enables us to test different neuron models without changing the FPGA configuration. All generated data will then be transferred to the processor during the initialization sequence.

**Electrical Synapses.** A neuron at the position \((i,j)\) receives 8 partial gap-junction currents from neighboring neurons with \( J_{gap,i,j} \), the total current. In contrast to Eq. (12), not each potential difference is weighted by the coupling strength \( g_{gap} \), but only its summation. During a block calculation of 40 neurons all synaptic operations are allocated to two \( \text{ADD}_A \) modules and a multiplier. Because of direct further processing, both \( \text{ADD}_A \) modules store their results in registers. The synaptic process is independent of the neuron model and doesn’t change, so the related program memories have a fixed content and cannot be modified after the FPGA is configured. Membrane potentials are stored in two RAMs having the same content as the \( \text{ADD}_A \)-RAM of the model equations when the initialization sequence has finished. In each new neuron cycle, both RAMs operate alternately by a toggle-bit, that means while one RAM provides the membrane potentials for the gap-junctions, the other stores the updated potentials of the current neuron cycle parallel to \( \text{ADD}_A \)-RAM of the model equations and vice versa.

The module \( \text{SET}_{ADDR} \) determines the RAM addresses of the actual neuron (Addr0) as well as all neighboring neurons (AddrI to Addr8) which are identical to the neuron identification numbers (NIDs, 0···1,599). These addresses are then used in the \( \text{ADD}_A \) module to select the membrane potentials \( V_{i,j} \) and \( V_{i+n,j+m} \) with \( m,n \in \{-1,0,1\} \). \( V_{i,j} \) is stored in the register RegV as operand A during a separate load instruction and the other potentials represent operand \( B \), see Figure 2. Due to the additional load instruction, both RAMs can be implemented as single-port memories. The size of the neuronal network can easily be adjusted by the user settings from \( 1 \times 1 \) to \( 40 \times 40 \). Eight flags define whether the positions of the neighboring neurons are located inside the network borders and they’ve to be taken into account or not. All gap-junction currents are stored in FIFO \( \text{FGAP} \). Finally, these values can be accessed as operands by the \( \text{ADD}_B \) module of the model equations.

**Main-Controller.** The processor has a main-controller as higher level control which analyzes incoming initialization data from the host and relays them to the related memories and registers. The packet-based data transfer with the host is realized by a USB 3.0 connection. A FIFO-Management-Unit (FMU) with own buffers in both communication directions controls all packet transfers between main-controller and peripheral USB controller. Processor-generated packets always have a length of 512 bytes and are called \( \text{PPacket} \)s to distinguish them from USB packets (\( \text{UPacket} \)s) which can have a length of 512 or 1024 bytes depending on a high-speed and super-speed connection, respectively.

Because the processor is a 32-bit system, a 4-byte word is processed per clock cycle. A \( \text{PPacket} \) always consists of 4 header and 124 data words. The header contains information such as “Packet Type” and “Packet Number”. For example, the field “Packet Type” indicates if a certain program memory has to be initialized or the simulation has finished. User settings for core registers are summarized in a SETUP
packet. It mainly includes the mode, the selected neuron model, the gap-junction networking (torus / no torus), the network size, the threshold for spike detection, as well as initial values for the ramp function $g_{\text{gap}}$, $T$, and $J_{\text{inj}}$. During run-time, each ramp function enables an automatic incrementation of its parameter until maximum is reached. The parameter interval is divided into 1000 steps, this means $\Delta P = (P_{\text{max}} - P_{\text{init}}) / 1000$ with $P \in \{g_{\text{gap}}, T, J_{\text{inj}}\}$ and $P_{\text{max}} \geq P_{\text{init}}$. All ramp functions determine important tuning parameters for network analysis.

Core-Controller. During run-time, the core-
controller increments the program counters of all AUs, updates the ramp functions, alternates the RAM access for gap-junctions via a toggle-bit, and composes all P-Packets to be transferred to the host in a small buffer which can be read by the main-controller, see Figure 3. Here, both modules for the actual equations and the electrical synapses as well as the core-controller are summarized to a neuron core. Such a core includes all necessary hardware for 1600 neurons and supports several modes defining how the calculated results are further processed in the core-controller. In this paper, we mostly use an event-driven mode called Preselected-Neuron-Spike-Time (PNST). It contains the spike time and the related NIDs of up to 124 individually selectable neurons. With these spike times the host software can determine interspike intervals (ISI) and phase differences according to Eq. (16), since the necessary computing power is limited. The mode Preselected-Network-Integer (PNI) includes, among other things, the mean membrane potential of all interconnected neurons (mean field potential, MFP) as an indicator for synchronization. Finally, the fire state of all neurons at any time is composed in the mode Full-Network-Spike-Time (FNST). This mode is used to plot the dotted spike times of the entire network.

4 RESULTS

The presented architecture is written in VHDL and synthesized on the Xilinx ML605 development board. Buffer registers of memory outputs can be absorbed by the related RAM during the synthesis process with high logic delay of approximately 2 ns. In cases where the timing failed, this absorption can be avoided by the keep attribute in the VHDL code. Together with a minimum reset path we are able to meet the timing constraints with a minimum period of 4.996 ns. The synthesis results in Tab. 1 show that the processor with one implemented neuron core takes 16 % of all LUT resources including the FMU. Therefore, the used FPGA has enough resources for further development like chemical synapses and multi-core operation.

Due to the block-wise calculation of always 40 neurons, a neuron core is optimized for network sizes of 40x40 in real-time and 20x20 in quadruple speed. However, the simulation of two coupled neurons is not efficient and should be done software-based, since the results of the remaining neurons are rejected and the high parallelism is left unexploited. For such cases we have expanded our hardware to include a speedup option which only works in PNST mode and defines the number of calculated blocks per neuron cycle. This is still worse than a software-based solution, but it helps us in the evaluation period to reduce the real simulation time for interspike intervals (ISI) from several hours to less than 10 minutes.

To validate our design, the FPGA results are compared with a C++ simulation. Figure 4 summarizes the results of two electrically coupled neurons as well as a 20x20 network at T = 6 °C. All simulations use the coupling strength as tuning parameter from 0 to its maximum value. The two neuron system is well suited to study transitions between asynchronous and different types of synchronous states. Figure 4 (A) and (B) illustrate the interspike interval and the corresponding phase difference with \( J_{ij} = 0.045 \mu A/cm^2 \) and \( g_{gap,max} = 0.03 mS/cm^2 \) for the FPGA in PNST mode (left) and the C++ program (right). A comparison of these plots shows a good agreement between both implementations. The injection current is selected in a way that uncoupled neurons are closed to the first period-doubling. This is the region where the greatest variety of synchronous activity can be seen for coupled neurons (Postnova et al., 2007). At first, the neurons are in an out-of-phase synchronization state with a constant phase difference \( \Delta \phi \neq 0, 2\pi \)

---

**Table 1: Device utilization summary.**

<table>
<thead>
<tr>
<th>Slice Logic</th>
<th>Used</th>
<th>Available</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slice Registers</td>
<td>19,676</td>
<td>301,440</td>
<td>6 %</td>
</tr>
<tr>
<td>Slice LUTs</td>
<td>24,253</td>
<td>150,720</td>
<td>16 %</td>
</tr>
<tr>
<td>Occupied Slices</td>
<td>8,720</td>
<td>37,680</td>
<td>23 %</td>
</tr>
<tr>
<td>RAMB36E1</td>
<td>58</td>
<td>416</td>
<td>13 %</td>
</tr>
<tr>
<td>RAMB18E1</td>
<td>42</td>
<td>832</td>
<td>5 %</td>
</tr>
</tbody>
</table>
and a regular tonic activity. An increasing coupling strength leads to an abruptly asynchronous behavior with an irregular interspike interval distribution and then to an out-of-phase state of period 4. The latter means four lines in both diagrams. If the coupling strength in further increased, two additional regions with an asynchronous behavior can be seen until the neurons go into in-phase synchronization. In general, the simulated transitions are identical to (Postnova et al., 2007), where more information on physiological details can be found.

Figure 4 (C) shows, for the FPGA in PNI mode (left) and the C++ program (right), the mean field potential of a 20 × 20 network with $J_{inj} = 0.65 \mu A/cm^2$ and $g_{gap,max} = 0.15 mS/cm^2$. At this injection current, the uncoupled neurons are in the bursting regime. Here, an increasing coupling strength leads to a two step-like transition from the unsynchronized to the fully synchronized state with a MFP plateau at an almost synchronized state between 0.1 and 0.125 mS/cm². This transition was already published for a 10 × 10 network in (Postnova et al., 2009) and makes clear that the processor works accurately. At the end, the dotted spike times of the FPGA network taken in FNST mode with synchronized bursts at $t = 250s$ are illustrated in Figure 4 (D).

Finally, we have compared the real simulation time of our processor for the 20x20 network with an Intel Core i7 (6500U) with 3.0GHz clock rate, 2 CPU-cores, and full utilization. The CPU-based simulation takes about 2.9 times longer.

5 CONCLUSIONS

We have presented a new processor architecture for the Huber-Braun neuron model, a simplified HH-type model. It is optimized for network sizes of 40x40
in real-time and 20x20 in quadruple speed. Synaptic connections are realized by bidirectional gap-junction coupling. The synthesized design runs on the Xilinx ML605 development board at 200MHz clock frequency and takes 16% of all LUT resources. A comparison with C++ simulations under physiological conditions shows that the processor works accurately. Furthermore, it is about 2.9 times faster than an Intel core i7 with 2 CPU-cores. Our hierarchical structure with a main-controller as higher level control and a neuron core with all arithmetic components including a core-controller allows the implementation of physiologically more relevant chemical synapses in a special module and can easily be extended to a multi-core architecture. This will drastically increase the computing power and is one of the next targets of this research.

REFERENCES


