Qualitative-Modeling-Based Silicon Neuronal Networks

Takashi Kohno

Institutue of Industrial Science
University of Tokyo
kohno@sat.t.u-tokyo.ac.jp
http://www.sat.t.u-tokyo.ac.jp/~kohno

at Université de Picardie Jules Verne
31/10/2016
Topics

- Background and overview of silicon neuronal networks
- An ultralow-power *analog* silicon neuron circuit
- A *digital* silicon neuronal network platform
- A *silicon neuronal network* on an FPGA chip
The nervous system

- **low-power consumption** (human brain: ~20W)
- **intelligent, flexible, and high-speed** processing of a huge amount of multimodal information flow
- Its information processing ability is **gained autonomously** and adapted to the environment.
- high durability (~ 100 years)
  (autonomous restoration, fault tolerance)
- a **massively parallel connection** of low-speed devices

**Neuromorphic Information Processing Systems**

- information processing systems with massively parallel architecture that inherits the excellent properties in the nervous system
- **neuro-mimetic** or **neuro-inspired** approaches

**background**

- rapid growth of information network (ex.) IoT
- reduction in the labor force
- energy challenge
Silicon neuronal networks

Electronic-circuit version of nervous system composed of silicon neuron and synapse circuits which reproduce the electrophysiological activities of neurons and synapses in real-time or faster.

- a most fine-granular bottom-up approach to neuromorphic systems
- neuro-mimetic: intends to catch all characteristics of the nervous system

Different from artificial neural networks (neuro-inspired)
Silicon neuronal networks

Electronic-circuit version of nervous system composed of silicon neuron and synapse circuits which reproduce the electrophysiological activities of neurons and synapses in real-time or faster.

- a most fine-granular bottom-up approach to neuromorphic systems
- neuro-mimetic: intends to catch all characteristics of the nervous system

Different from artificial neural networks (neuro-inspired)
Application fields of SNNs

Neuronal cells and Synapses

Nervous system

Silicon neurons + Silicon synapses

Hybrid systems

Model validation
Biomedical device
BMI device

Large-scale real-time simulators

Neuromorphic Hardwares

Autonomous information processing systems that mimic nervous information processing

Silicon retina, cochlea
Selective attention system
Associative memories
Pattern recognition

Le Masson et al., Nature, 2002
Indiveri, ISCAS, 2003
[Schemmel et al., Proc. IJCNN, 2008]
Difficult in silicon neuronal networks

Precisely reproducing various complex neuronal activities.

V.S.

Power consumption and integration degree.

It is **UNKNOWN** what properties in the neuronal activites contribute what information processing.
Ionic mechanisms of membrane potential
Membrane capacitance is (dis)charged by ionic currents.

Extracellular fluids

Intracellular fluids

$\textbf{I}_{\text{stim}}$: stimulus current

$\textbf{v}$: membrane potential

spikes' timing is modulated

extracellular fluid

intracellular fluid

axon

synapse
Ionic mechanisms of membrane potential

Membrane capacitance is (dis)charged by ionic currents.

- **membrane capacitance**: capacitance of the cell membrane
- **ionic current**: generated by the flow of ionic particles through ionic channels
Ionic mechanisms of membrane potential
Membrane capacitance is (dis)charged by ionic currents.

- **membrane capacitance**: capacitance of the cell membrane
- **ionic current**: generated by the flow of ionic particles through ionic channels
Ionic mechanisms of membrane potential

Membrane capacitance is (dis)charged by ionic currents.

- **Voltage-dependent channels**: conductance is dependent on \( v \)
  - Fast channels: follows quickly to \( v \)
  - Slow channels: takes time to respond to change of \( v \)
Conventional silicon neuron models

Ionic conductance models

Leech Heart Interneuron

\[ C \frac{dv}{dt} = I_{CaF} + I_{CaS} + I_P + I_{Na} + I_{K_1} + I_{K_2} + I_A + I_h + I_l + I_{SynG} + I_{SynS} \]

I\&F-based models

- Focused on the timing of spike generation.
- Neuronal activities are reproduced partially.
- Simple low-dimensional ODE.
- Simple circuitry & low power

Ionic current: \( I_j = \bar{g}_j m_j^M_j h_j^H_j (v - E_j) \)

- Mechanisms of neuronal activities are described.
- Neuronal activities are reproduced precisely.
- High-dimensional nonlinear ODE
- complex circuitry & high power

[Nadim et al. 1995]
Limitations of I&F-based models

Variability of spike intensity

- Peak of excitatory post-synaptic potential is dependent on the stimulus intensity.  
  [Alle et al., 2006]

Phase response curve (PRC)

- PRC of the Izhikevich model in the Class 2 mode is not typical Type 2 PRC.
- Low error recovery performance of auto-associative memory in Class 2 mode.  
  [Osawa and Kohno, 2015]
Qualitative neuronal modeling

Ionic conductance models
Describe dynamics of ionic charges
High-dimensional complex eqs.

\[ C \frac{dv}{dt} = \sum_j I_j (v, w), \]
\[ \frac{dw_j}{dt} = f_j (v, w_j). \]

Dimension reduction, bifurcation analysis, ...

Qualitative models
Describe mathematical structures in ionic conductance models
Low-dimensional polynomial eqs.

\[ \frac{dv}{dt} = \sum_j g_j (v, w), \]
\[ \frac{dw_j}{dt} = f_j (v, w_j). \]

e.g. Hodgkin-Huxley model

\[ C \frac{dv}{dt} = g_{Na} m^3 h (E_{Na} - v) + g_K n^4 (E_K - v) + g_L (E_L - v), \]
\[ \frac{dm}{dt} = (f_m (v) - m) / \tau_m (v), \]
\[ \frac{dh}{dt} = (f_h (v) - h) / \tau_h (v), \]
\[ \frac{dn}{dt} = (f_n (v) - n) / \tau_n (v), \ldots \]

FitzHugh-Nagumo model

\[ \frac{dx}{dt} = c (y + x - \frac{x^3}{3} + z), \]
\[ \frac{dy}{dt} = - (x - a + by) / c. \]
Qualitative SN modeling

Ionic conductance models

Describe dynamics of ionic charges
High-dimensional complex eqs.

\[ C \frac{dv}{dt} = \sum_j I_j (v, w), \]

\[ \frac{dw_j}{dt} = f_j (v, w_j). \]

dimension reduction, bifurcation analysis, ...

Qualitative models
compatible with target devices

[ Kohno & Aihara, 2005 ]
[ Takemoto, Kohno, & Aihara, 2007 ]
[ Kohno & Aihara, 2008 ]

Qualitative models

Describe mathematical structures in ionic conductance models
Low-dimensional polynomial eqs.

\[ \frac{dv}{dt} = \sum_j g_j (v, w), \]

\[ \frac{dw_j}{dt} = f_j (v, w_j). \]
A qualitative-modeling-based SN circuit

- a 3-variable model designed for
  subthreshold MOSFET circuits
- power consumption: ~ 70nW (TSMC 0.35μm)
- configurability: Class I/II (FS), RS, square-wave bursting, elliptic bursting

Parameter tuning (nullcline-drawing) mode:

Neuron mode:
An ultralow-power QMB silicon neuron circuit based on a three-variable silicon neuron model

\[ C \frac{dv}{dt} = I_{in} + I_{\text{leak}} + I_{Na} + I_{Kd} + I_{KM} + I_{CaT} + I_{CaL} \]

- **fast dynamics (spikes)**
- **slow dynamics**

- supports Class I and II spike generation, RS, FS, LTS, SWB, and EB neuron classes. (partial support for IB class)
- equipped with parameter tuning helper (nullcline-drawing) circuit
QMB design approach

- Design the fast subsystem using phase-plane and bifurcation analyses. Bifurcation parameter: \( q \) (the variable for slow dynamics)
- Bifurcation diagram is the phase portrait projected to the \( v-q \) plane.
Design the fast subsystem using phase-plane and bifurcation analyses. Bifurcation parameter: $q$ (the variable for slow dynamics)

Bifurcation diagram is the phase portrait projected to the $v$-$q$ plane.
An ultralow-power QMB silicon neuron model

Equations

\[ C \frac{dv}{dt} = f_v(v) - g_v(v) + I_{av} - r_n(n) - r_q(q) + I_{\text{stim}}, \]

\[ \frac{dn}{dt} = f_n(v) - g_n(v) + I_{an} - r_n(n), \]

\[ \frac{dq}{dt} = f_n(v) + I_{aq} - r_q(q). \]

\( v \): membrane potential,

\( n \): variable for fast dynamics,

\( q \): variable for slow dynamics.

\[ f_x(v) \text{ circuits: steeper sigmoidal curve} \]

\[ f_x(v) = \frac{M_x}{1 + \exp \left( -\frac{\kappa}{U_T} (v - \delta_x) \right)} \]

- current “booster” circuit (M7, M8, M10, M11):
  reduces width of M1, M2, M3 and total power consumption
An ultralow-power QMB silicon neuron model

Equations

\[
C \frac{dv}{dt} = f_v(v) - g_v(v) + I_{av} - r_n(n) - r_q(q) + I_{stim},
\]

\[
\frac{dn}{dt} = f_n(v) - g_n(v) + I_{an} - r_n(n),
\]

\[
\frac{dq}{dt} = f_n(v) + I_{aq} - r_q(q).
\]

\(v\): membrane potential,

\(n\): variable for fast dynamics,

\(q\): variable for slow dynamics.

g\(_x(v)\) and r\(_x(x)\) circuits: shallower sigmoidal curve

- M2 and M3: cascode circuit
- M1: responsible for square root operation \((\rightarrow\) shallowness)\n- M6, M7, M8, M9: displaces the curve horizontally \((\text{only in } g_n(v) \text{ circuit})\).

\[
g_x(v) = I_0 \sqrt{\frac{R_{x20} \exp\left(\frac{\kappa}{U_T} \theta_x\right)}{1 + R_{x21} \exp\left(-\frac{\kappa}{U_T}(v - \theta_x)\right)}}
\]
Overall circuit

16 parameter voltages to be tuned for an intended neuronal activity.

- The voltage clamp amp. assists the parameter tuning procedure.
Overall circuit

16 parameter voltages to be tuned for an intended neuronal activity.

- **Nullcline mode**

- the $v$- and $n$-nullclines:
  - indicators for the dynamical structure in the spike generating system.
  - the $q$-nullcline: indicates the dynamical property of $q$. 

T. Kohno Qualitative-modeling-based SNN 23/50
Overall circuit

16 parameter voltages to be tuned for an intended neuronal activity.

Bifurcation diagram mode

- indicates the overall dynamical structure in the system.
Spike generating subsystem

Nullcline mode  Pulse stim. response  Sustained stim. response

Class I

Class II

- Spectre simulation results (circuits experiments are not shown)
- Both Class I and II spike generation mechanisms are supported.
Regular spiking (RS) mode

Spectre simulation

Biological examples

- The spike frequency adapts to a lower value in response to sustained stimulus.
- The spike frequency is higher in response to stronger stimulus.
The rebound bursting (burst spiking just at the offset of inhibitory stimulus) The number of spikes is dependent on cells and stimulus intensity.
Autonomous bursting

Spectre simulation

Elliptic bursting mode

Square-wave bursting mode

Two types of rhythmic bursting.

Tonic firing in a square-wave bursting cells.
### Summary

<table>
<thead>
<tr>
<th></th>
<th><strong>our circuit</strong></th>
<th>[ Brink et al. 2013 ]</th>
<th>[ Qiao et al. 2015 ]</th>
<th>[ Rangan et al. 2010 ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>QM-based</td>
<td>QM-based</td>
<td>I&amp;F-based</td>
<td>I&amp;F-based</td>
</tr>
<tr>
<td>power consumption</td>
<td>&lt; 5 nW</td>
<td>2 – 3 nW</td>
<td>2 – 3 nW</td>
<td>0.2 – 7 nW</td>
</tr>
<tr>
<td>Class I</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Class II</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓ (partially)</td>
</tr>
<tr>
<td><strong>spike variability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>LTS</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>SWB</strong></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>SNN chip available</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

- Test circuit has been fabricated. Preparing for circuit experiments.
- **Automatic tuning algorithm** for parameter voltages is to be developed.
Ultrapower analog SNN (to be realized)

- **(A) Ultra-low-power silicon neurons**
  \(~ 5 \text{ nW} \) with a 3-variable model, 10–25 neurons in a 25mm\(^2\) chip

- **(B) Kinetic-model-based silicon synapses**
  \(~ 2 \text{ nW} / 1000 \text{ synapses} \) , 1000 synapses / neuron

- **(C) inter-neuron routing** and **(D) inter-chip connection**

- **(E) Memory circuit for synaptic efficiency**
An analog SNN chip by INI

- AMS 0.35 μm CMOS, 2.1mm x 2.5mm
- 32 I&F-based silicon neurons in a chip
- 4 silicon synapses for a neuron (3 excitatory and 1 inhibitory)
- 1024 Synaptic inputs /w programmable connection a silicon synapse receives multiple inputs
- 5-bit synaptic weight (STDP learning rule is to be implemented)
Soft Winner-Take-All (sWTA) network

- Blue: Excitatory neuron
  - output to the first and second nearest neighbors
  - output to the inhibitory neuron (red neuron)

- Red: Inhibitory neuron
  - output to all the excitatory neurons (blue neurons)
Firing rate of the group with slightly stronger input is extremely higher.

- Blue: random stimulus inputs (poisson process)
- Black: spike outputs

[Indiveri et al. 2014]
Orientation detector

TMPDIFF chip
- A neuromorphic image sensor
- $32 \times 32$ pixels
- Generates spikes proportional to log of the “intensity” of a pixel.

AER bus
- A off-chip bus that transmits the timing of spikes.

IFWTA chip
- The I&F-based SNN chip.
- Each of 31 silicon neurons receives spikes from its own “receptive field” bar with different orientation.

[Chicca, et al., IEEE CAS, 2007]
Output is coded by **firing rate** of each silicon neurons.

- **Selectivity is enhanced by local connection**
  - Neurons for **similar** orientation **facilitate** each other via the neighborhood connection.
  - Neurons for **different** orientation are **depressed** via the global inhibitory neuron.
Digital Spiking Silicon Neuronal Network Platform

A high-speed neuronal network simulation platform which supports various neuronal activities (optimized for digital circuits)

Elements (models suitable for fixed-point operation and Euler’s method)

- **Neurons**: Digital Spiking Silicon Neuron (DSSN) model
  - NO reset of state variables, supports: variability of spikes in Class II, Class I*
- **Synapses**: Digital Silicon Synapse model
  - Transmits information of spike variation

**Learning rule**: Hebbian and STDP rules

All-to-all network of 1000 neurons in a Virtex-7 VX690T chip

Improvement of FPGAs

→ lower price, higher speed

Direct application of results
Digital Spiking Silicon Neuron Model

[ Nanami & Kohno 2016 ]

\[
\frac{dv}{dt} = \frac{\phi(u)}{\tau}(f_v(v) - n - q + u + I_0 + I_{\text{stim}}),
\]

\[
\frac{dn}{dt} = \frac{1}{\tau}(f_n(v) - n),
\]

\[
\frac{dq}{dt} = \frac{1}{\tau}(f_q(v) - q),
\]

\[
\frac{du}{dt} = \frac{\epsilon_u}{\tau}(f_u(v) - u),
\]

\[f_x(v) \equiv \begin{cases} 
  a_{xn}(v - b_{xn})^2 + c_{xn} & (v < r_x) \\
  a_{xp}(v - b_{xn})^2 + c_{xp} & (v \geq r_x)
\end{cases}
\]

or

\[f_x(v) \equiv \frac{v - x_0 - x}{\alpha_x}
\]

- Solved by \textbf{Euler's method} with \( \Delta t = 0.1 \text{ ms} \), 18-bit fixed point operations.
- 1 multiplication / step
- Supports: RS, FS (Class I/II), LTS, IB, SWB, EB, PB
Activities of DSSN Model (1)

Pospischils’ model  
DSSN model

Regular Spiking

Fast Spiking (Class I)
Activities of DSSN Model (1)

Pospischils’ model

Low Threshold Spiking

Intrinsically Bursting
Activities of DSSN Model (1)

- Fitting two important statistic measures for many (40–50) values of $I_{\text{stim}}$
Activities of DSSN Model (2)

- Elliptic bursting
- Parabolic bursting
- Square-wave bursting

- Supports three types of well-known autonomous bursting activities.
Digital Silicon Synapse Model

- Simplified version of the kinetic models of chemical synapses
  $I_{syn}$ reflects the spike width of pre-synaptic neurons

![Graphical representation of pre-synaptic, silicon synapse, and post-synaptic neuron circuits]
Digital Silicon Synapse Model

- Simplified version of the kinetic models of chemical synapses
  \( I_{\text{syn}} \) reflects the spike width of pre-synaptic neurons

\[
T = \begin{cases} 
1 & \text{when } v \geq 0 \\
0 & \text{when } v < 0 
\end{cases}
\]

\[
\frac{dI_{\text{syn}}}{dt} = \begin{cases} 
\alpha (1 - I_{\text{syn}}) & \text{when } T = 1 \\
\beta I_{\text{syn}} & \text{when } T = 0 
\end{cases}
\]

\( v \) : the membrane potential of a pre-synaptic neuron

\( T \) : represents the release of chemical transmitters

\( I_{\text{syn}} \) : synaptic current
Auto-associative memory by all-to-all connected DSSNN

[ Li, Katori, & Kohno, 2012 ]
[ Li, Katori, & Kohno, 2013 ]
Memorized patterns and input patterns all-to-all connected DSSN network

(A) 1 2 3 4

(B) 5% 10% 15% 20% 25%

30% 35% 40% 45% 50%

- 256-neuron network on Spartan-6 LX45
- 4 patterns were memorized by correlation learning
- A part of pixels are inverted in the input patterns (error insertion)
- Recall performance (ability to recall a correct pattern with more errors in input patterns)

Silicon neuron models are in: Class II > Class I
Auto-associative memory by all-to-all connected DSSNN

[Li, Katori, & Kohno, 2012][Li, Katori, & Kohno, 2013]

Memorized patterns and input patterns

Recall performance

(A) 1 2 3 4

(B) 5% 10% 15% 20% 25%

30% 35% 40% 45% 50%

- 256-neuron network on Spartan-6 LX45
- 4 patterns were memorized by correlation learning
- A part of pixels are inverted in the input patterns (error insertion)
- Recall performance (ability to recall a correct pattern with more errors in input patterns)

Silicon neuron models are in: Class II > Class I
Auto-associative memory by all-to-all connected DSSNN
[ Li, Katori, & Kohno, 2012 ][ Li, Katori, & Kohno, 2013 ]

Raster plots in recall process

Synchronization index (PSI)

SNs fire **in synchrony** when a correct pattern is recalled.
**NO synchronous firing** in case of unsuccessful recall
(synchronization reflects the correctness of the results)

Both of the patterns and synchronization are meaningful
an advantage to ANNs
A compact representation of multi-neuron’s spike trains.

A dot is plotted at the time when a neuron generated a spike.
Online learning in all-to-all connected DSSNN

[ Li, Katori, & Kohno, to be submitted ]

Hebbian learning

Asymmetric STDP learning

- **Hebbian learning** boosts the recall performance of auto-associative memory. Performance boost in both Class I and II modes

- **Asymmetric STDP learning** realizes a sequential memory.
  - strong stimulus induces a pattern transition
Auto-associative memory with Izhikevich model

[Osawa & Kohno, 2015]

PRC in Class II mode

Recall performance of auto-associative memory

- **Low performance with the Class II mode**
  - NOT typical type 2 PRC
  - discontinuity at $\theta = 0$
- Increase in parameter $d$ improves performance but with **distorted spike shape**
- **PRC is important in auto-associative memory**
  PRC can be distorted by I&F approximation
Summary

- **Biomimetic approach** is important for reaching intelligent computing systems comparable to the brain as is in the bio-silico hybrid systems.
  
  “Analysis by synthesis”

- Qualitative-modeling approach can realize simple and low-power biomimetic circuits.
  
  l&F-based circuits are simpler and lower-power, but can have **critical limitations**
  
  - Spiking neuronal networks can process information not only by spike rate but also by spike timing.
  
  - l&F-based circuits have poor ability in a spike-timing-based processing.