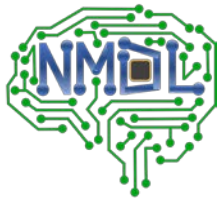
Neuromorphic Materials and Devices Lab.



IBM Research

Spiking Neural Network Neuromorphic Hardware Research for Novel Stochastic Algorithm

[Introduction to SNU class]

2021. 03. 30.

Uicheol Shin

Neuromorphic Device, Science & Technology, IBM Research - Tokyo

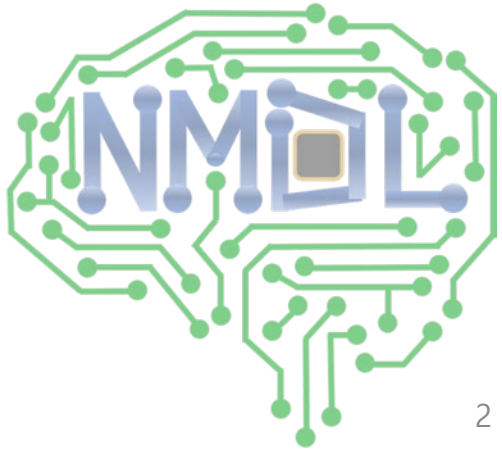
Department of Materials Science and Engineering, Seoul National University

Contents

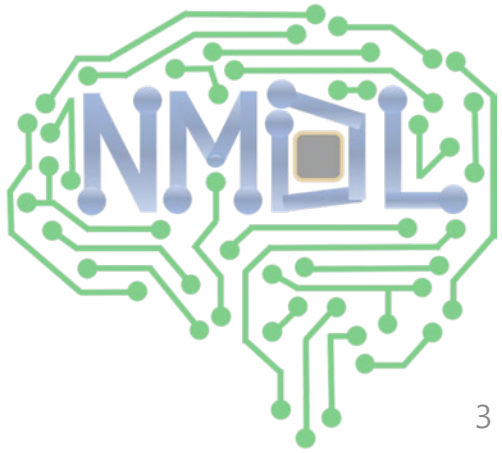
1. Introduction

2. My Previous works
   - SNN-RBM chip test results

3. Future Research Plans & Summary
   - Toward to novel algorithms



Contents

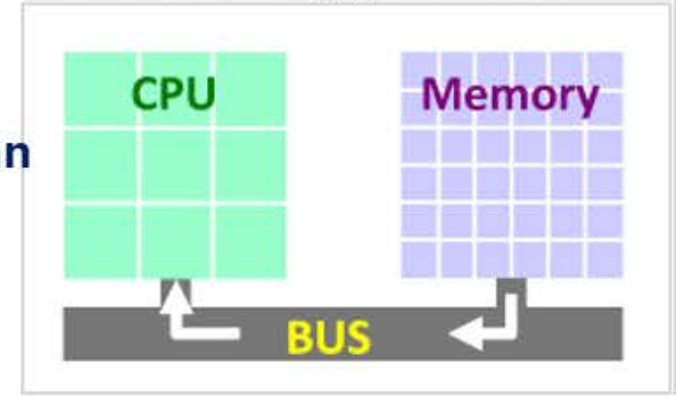


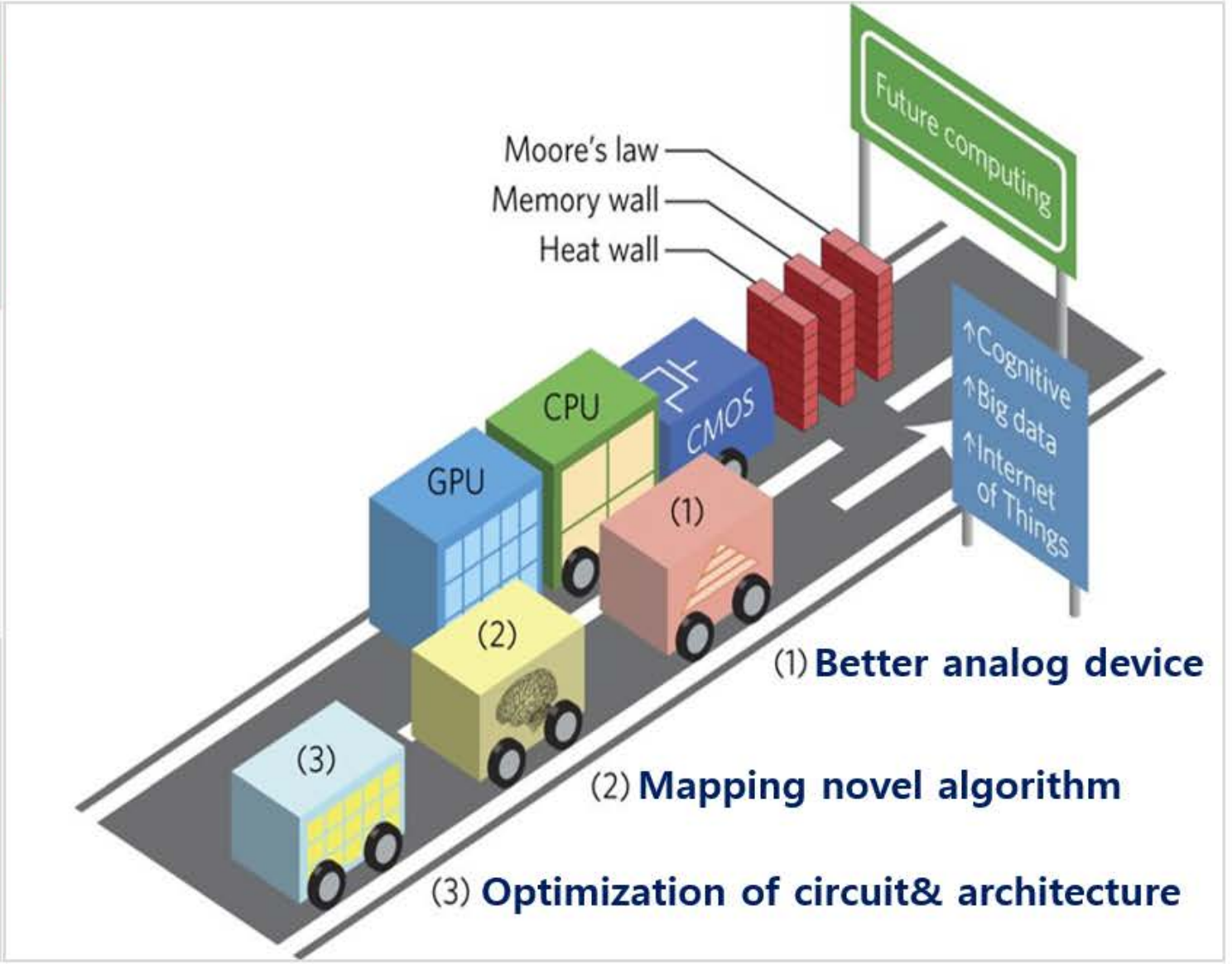
Why Neuromorphic?

End of Moore's Law

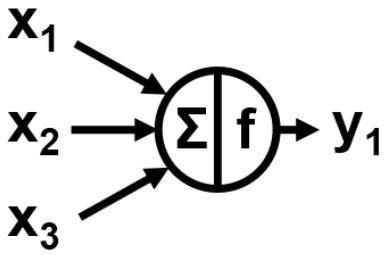
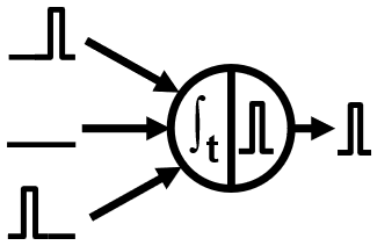
Explosion of data usage

Von Neumann bottleneck



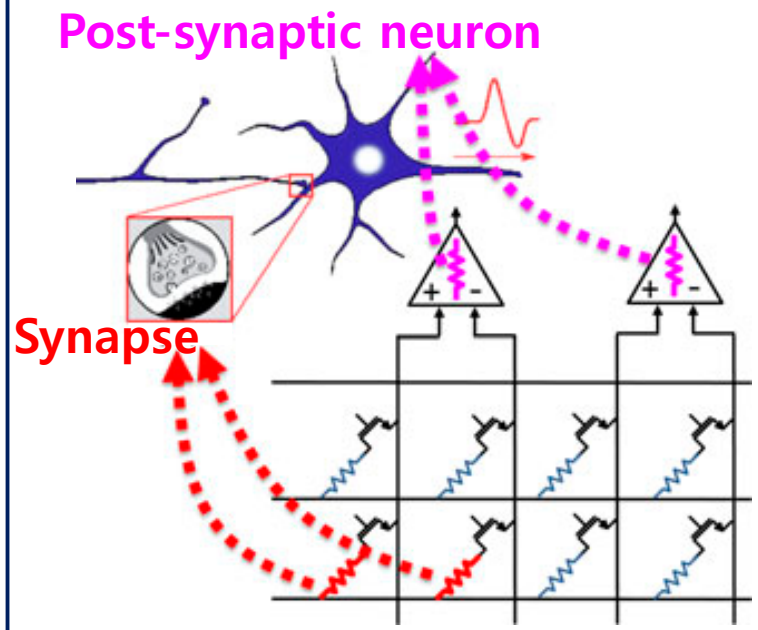


Why Spiking neural network neuromorphic?

	ANN : Non-spiking NN	Brain : Spiking NN
Inputs & Outputs	Real-valued numbers	Spikes
Neuron Operation		 S. Kim *et al.*, Springer International Publishing, **2017**, pp 153-164

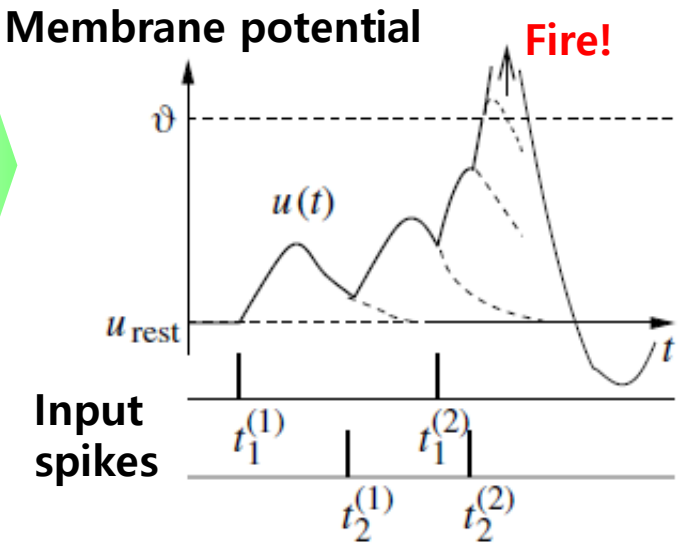
Asynchronous spikes bring us energy efficiency!

Concept of SNN hardware



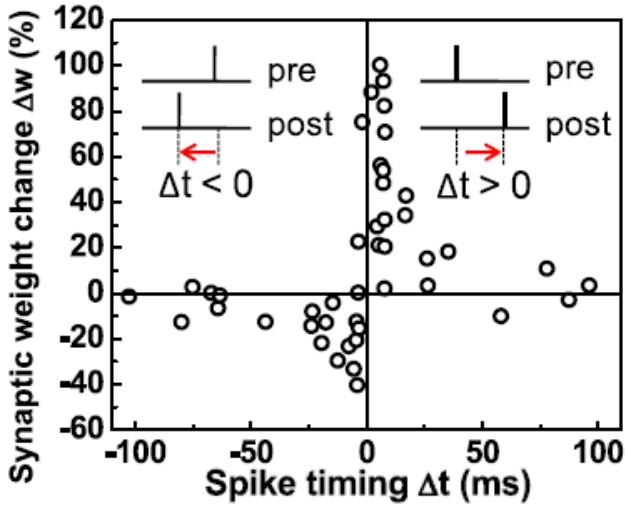
G. Burr *et al.*, Advanced in Physics: X, **2017**, *Vol.2 No.1*, 89-124

Leaky Integrate& Fire (LIF)



Gerstner W. *et al.*, Cambridge Univ. Press, **2017**

Spike-timing-dependent-plasticity (STDP)



Bi & Poo, J. Neurosci., **1988**, 18(*24*), 10464-10472

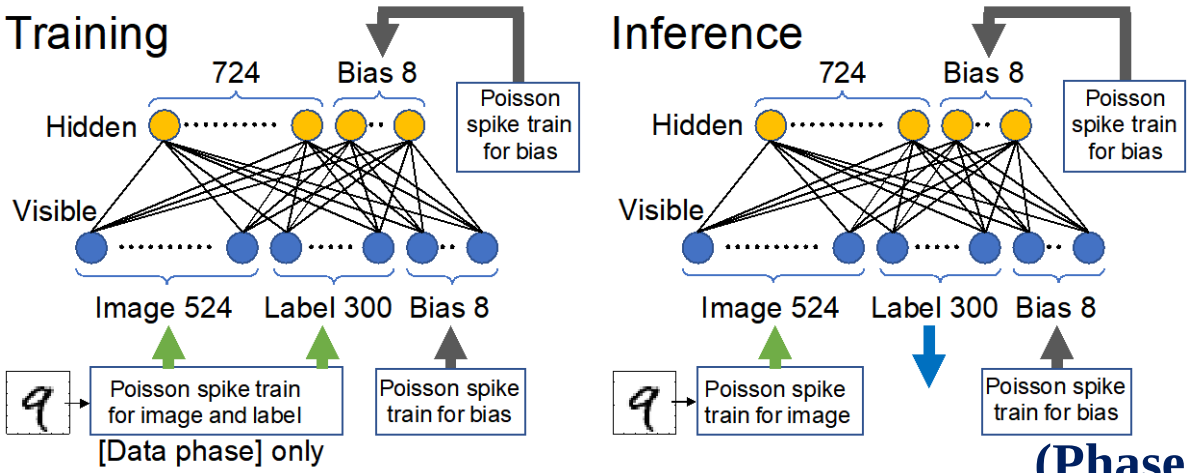
Hardware design of Spiking RBM chip

On-Chip Trainable 1.4M 6T2R PCM Synaptic Array with 1.6K Stochastic LIF Neurons for Spiking RBM

M. Ishii *et al., in IEDM*, **2019**, pp. 14.2.1-4.

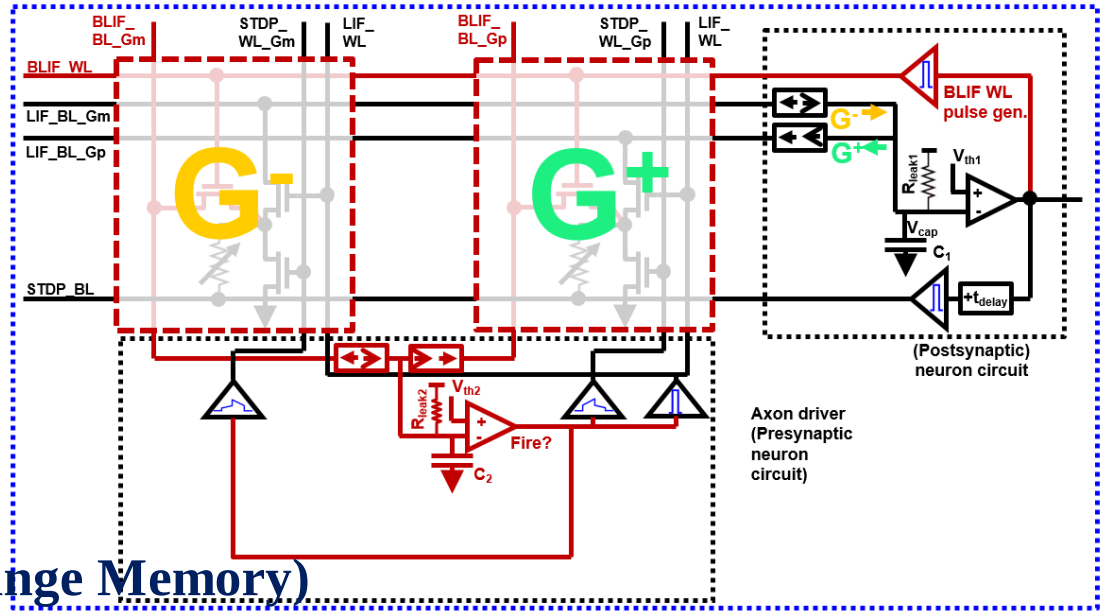
M. Ishii[1*], S. Kim[2*], S. Lewis[3], A. Okazaki[1], J. Okazawa[1], M. Ito[1], M. Rasch[3], W. Kim[3], A. Nomura[1], U. Shin[2], K. Hosokawa[1], M. BrightSky[3], and W. Haensch[3]
[1]IBM Research – Tokyo, Japan, [2]Seoul National University, South Korea, [3]IBM Research, T.J. Watson Research Center, USA
*These authors contributed equally to this work, email: ishiim@jp.ibm.com, sangbum.kim@snu.ac.kr

Restricted Boltzmann machine (RBM)



(eCD: event-driven Contrastive Divergence)

Neftci *et al.*, Frontiers in Neurosci., **2014**, *Vol.7*, 272



(Phase Change Memory)

"6T2R PCM-based unit cell & LIF neuron circuit"

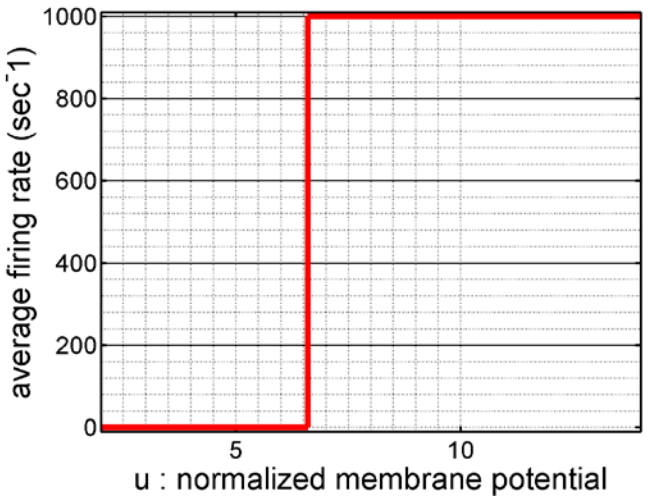
Algorithmic components:

i) Bidirectional activity
ii) Bipolar weight update
iii) Asynchronous and event-driven

i) Low-power consumption of PCM
ii) Asynchronous and parallel operation
iii) Undirected synaptic connection
iv) Bipolar synaptic weight (G^+ & G^-)

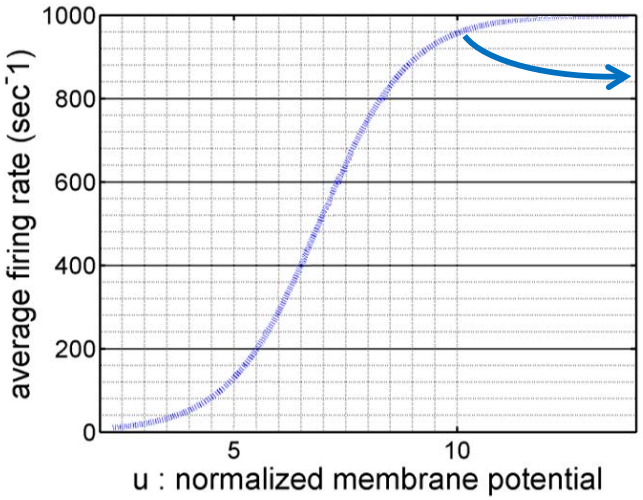
Implementation of stochasticity on Spiking RBM chip

["Deterministic" neuron]



Firing probability is either 0 or 1

["Stochastic" neuron]



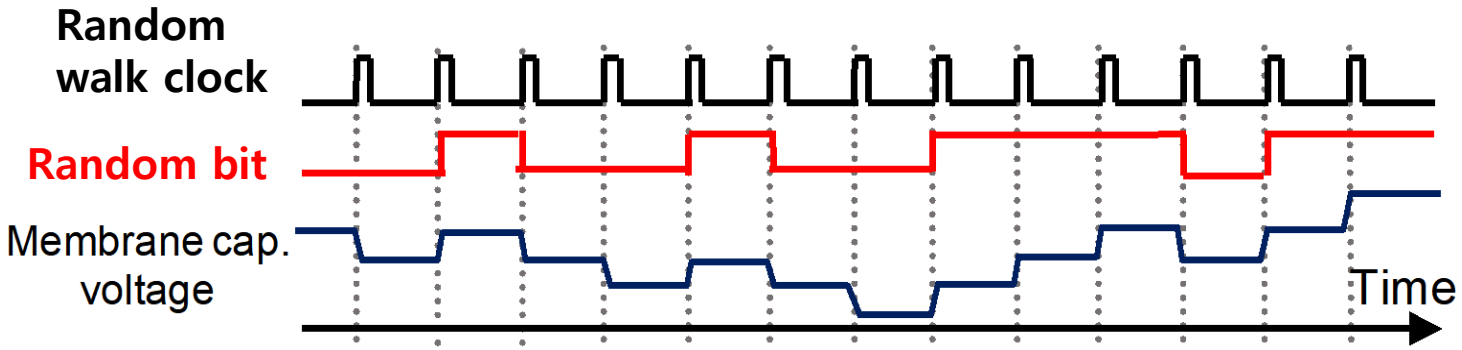
Ideal for spiking RBM

$$\rho(u) = (\tau_r + \exp(-u))^{-1}$$

Average firing rate for stationary $u(t)$,
($u(t)$: membrane potential,
τ_r: refractory period.)

Firing probability of a neuron is a function of membrane potential ≈ **NON-ZERO probability**

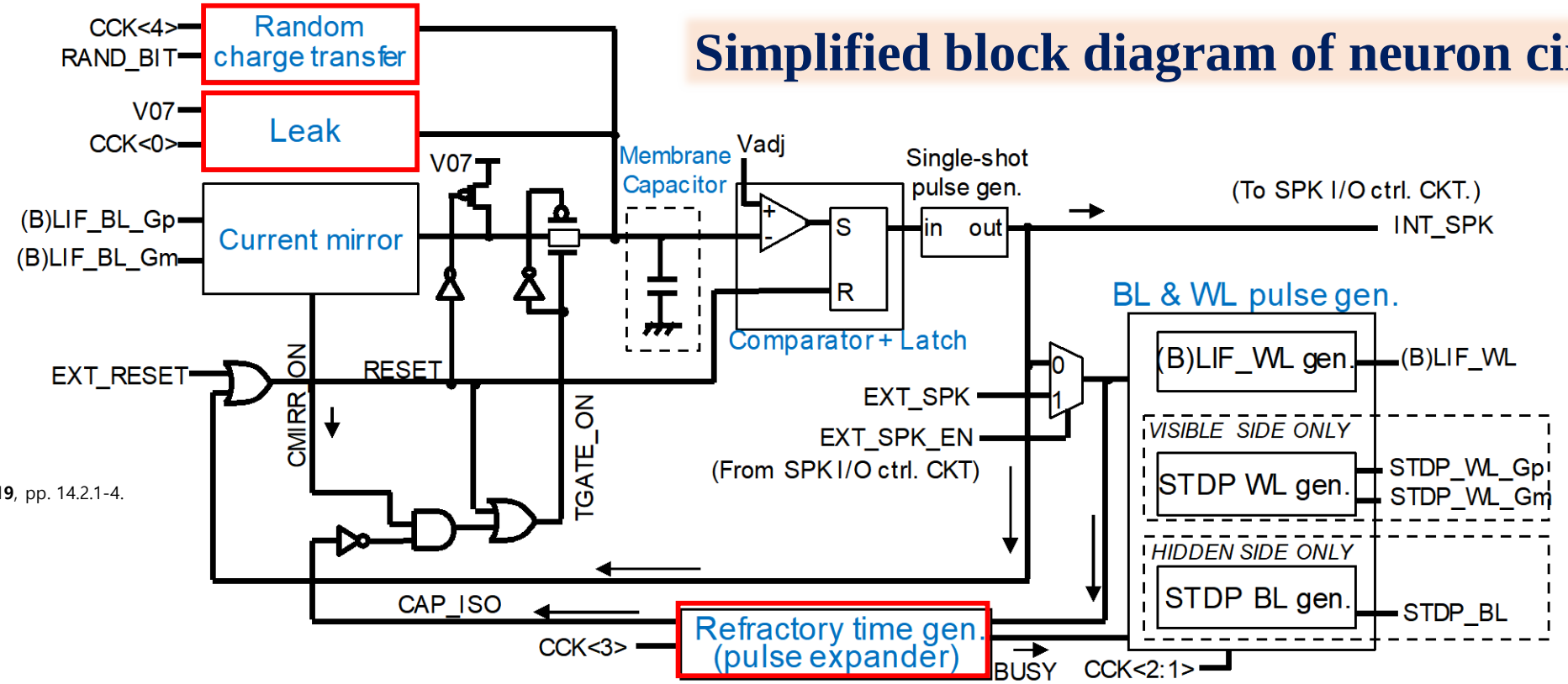
On-chip implementation: "Random walk function"



The firing probability is determined by the membrane potential.

M. Ishii *et al.*, *in IEDM*, **2019**, pp. 14.2.1-4.

Implementation of the other SNN components



Simplified block diagram of neuron circuits.

M. Ishii *et al.*, *in IEDM*, **2019**, pp. 14.2.1-4.

Implemented SNN primitives:

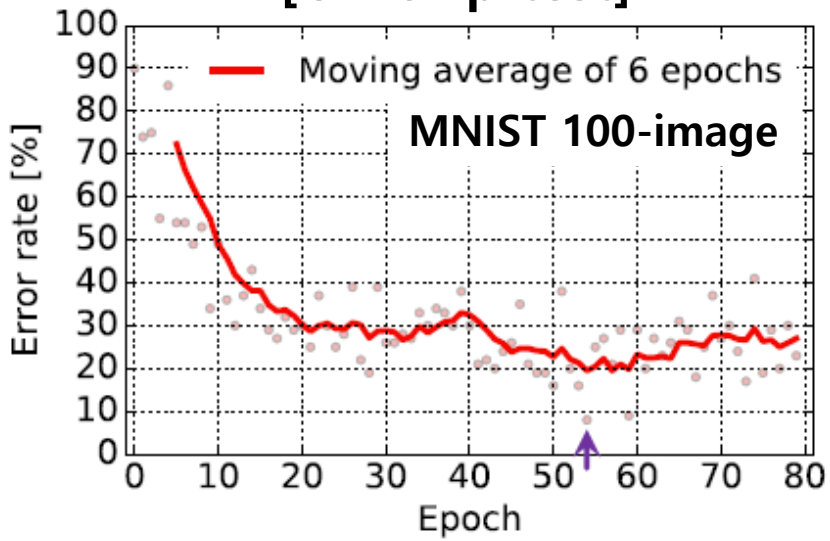
i) Random walk function – Ensure stochasticity of neurons

ii) Leak function:
The membrane potential gradually returns to resting potential.

iii) Refractory period:
Ignoring incoming spikes for a specific period after the fire.

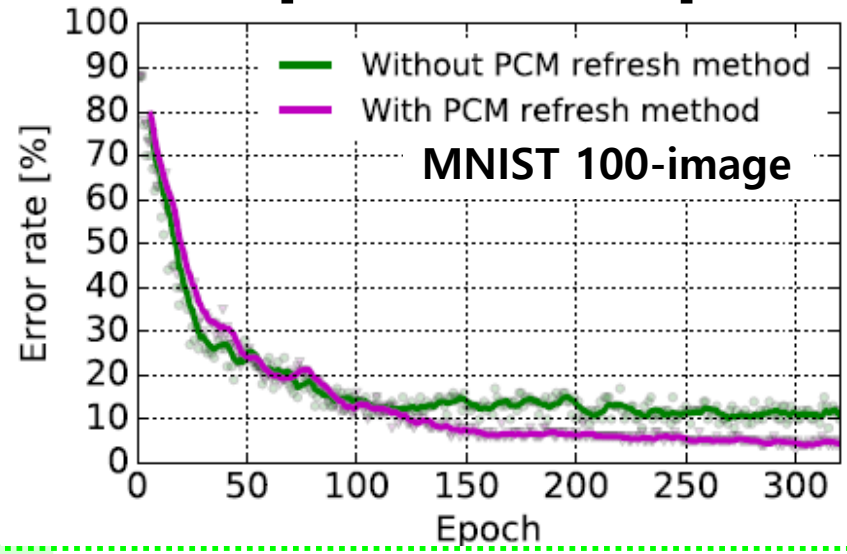
Further ‘Chip test’ works for improvements of chip performance

[On-chip test]

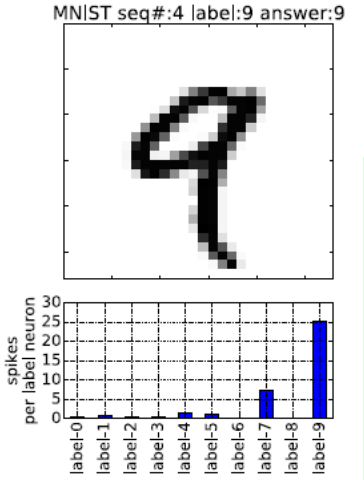


[Simulation test]

M. Ishii *et al.*, *in IEDM*, **2019**, pp. 14.2.1-4.



[Digit recognition]



Why is the error rate of real hardware higher than 20%?
Is there any issue that degrade the training accuracy?

Chip test can

i) Observe and analysis real phenomena on chip

ii) Optimize chip operations for algorithm efficiency

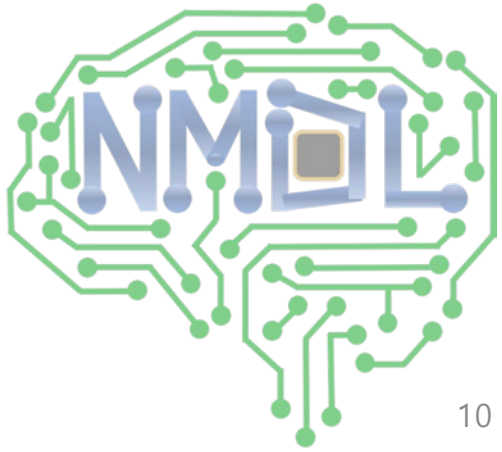
iii) Mapping novel algorithm using on-chip functions

Contents

1. Introduction

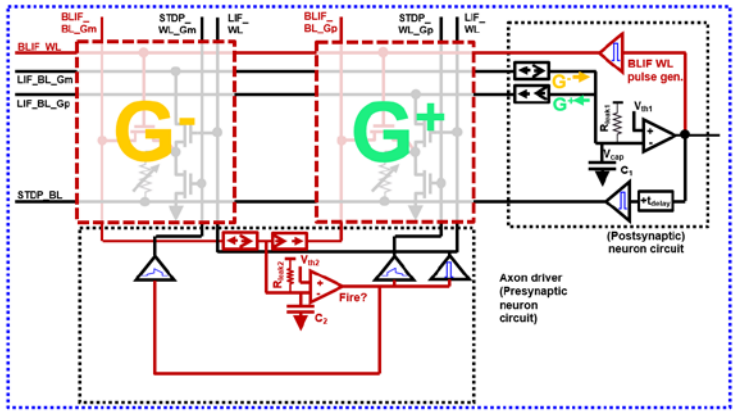
2. My Previous works
   - SNN-RBM chip test results

3. Future Research Plans & Summary
   - Toward to novel algorithms

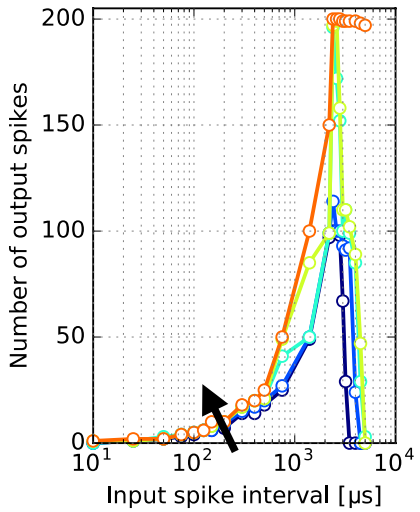


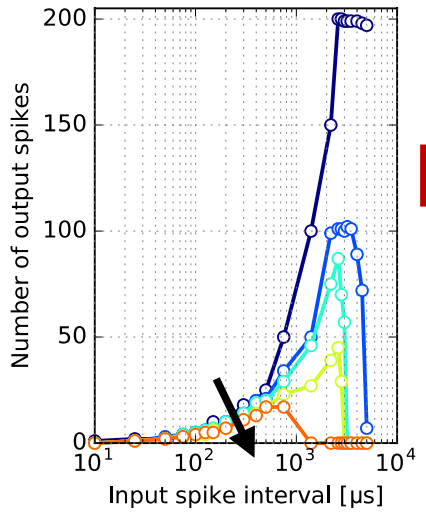
Experimental demonstration of spiking RBM chip (Restricted Boltzmann machine)

i) Bipolar synaptic weight



M. Ishii *et al.*, *in IEDM*, **2019**, pp. 14.2.1-4.

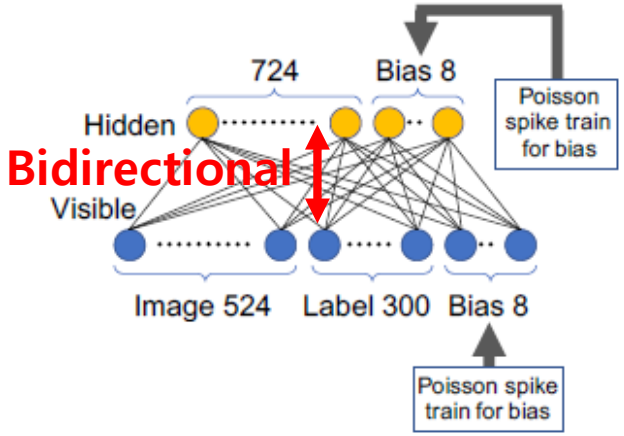




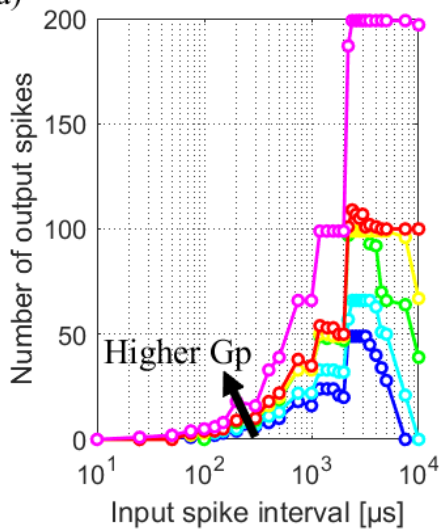
Programming Gp & Gm to implement potentiation & depression.

ii) Bidirectional synaptic connection

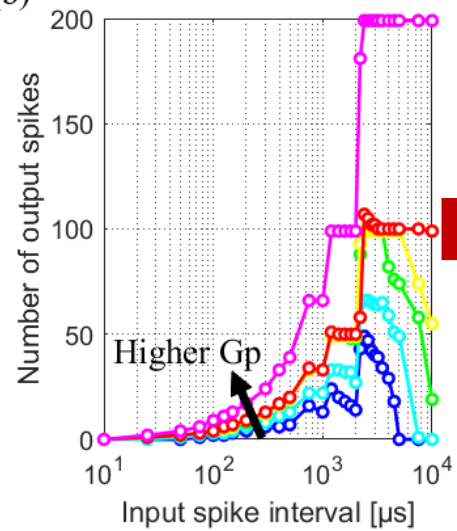
(Leaky Integrate-and-Fire)



[Forward LIF]



[Backward LIF]



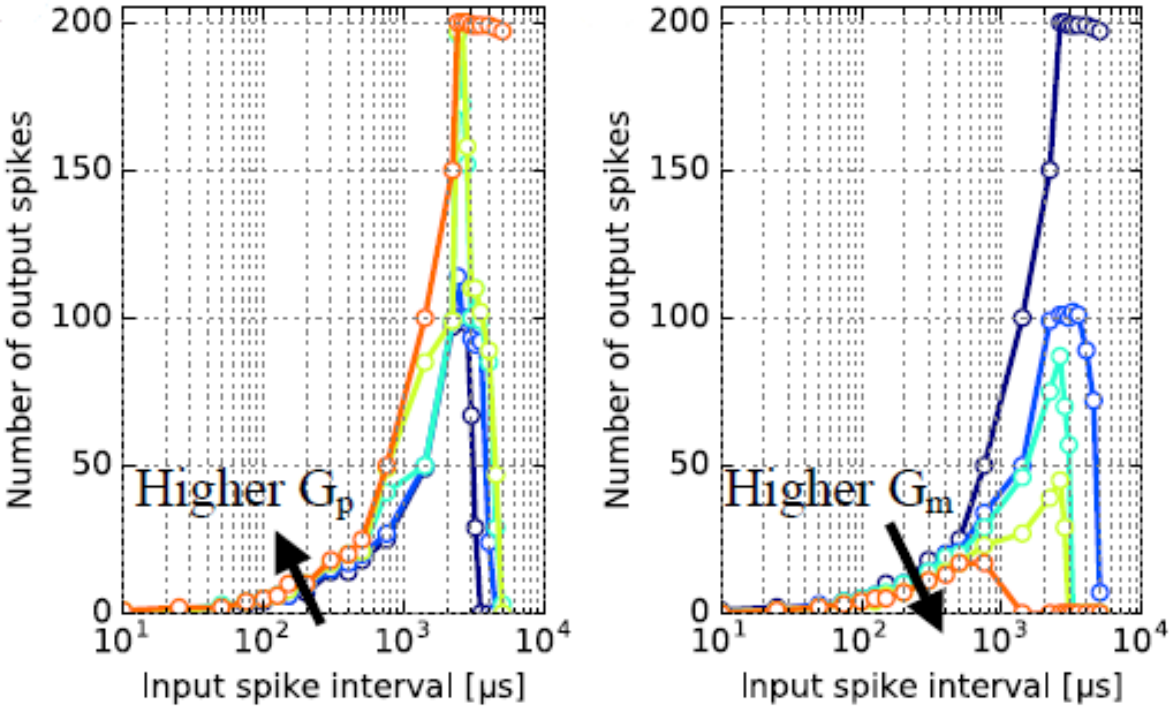
Demonstrating the symmetrical performance for the spiking RBM.

Experimental discussion of LIF functional test

Test condition

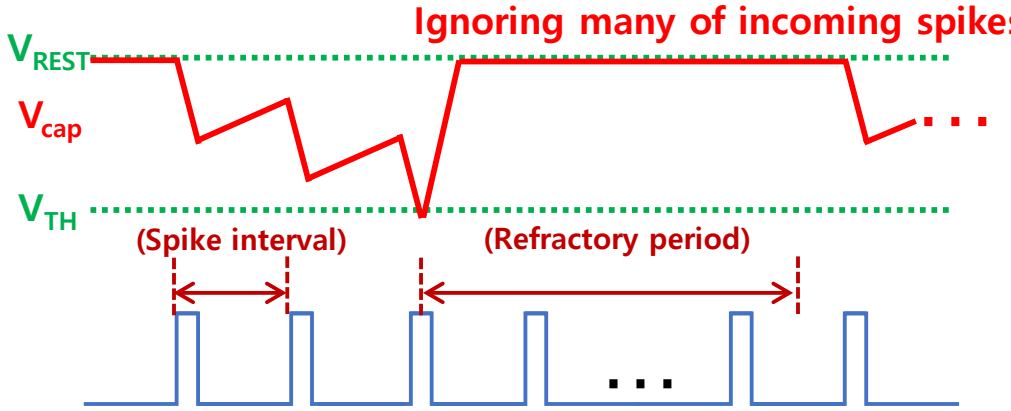
CCK_BASE 0, 1 = 10us, CCK_Leak = 80μs (→ fixed by PG)
SET voltage = 0.75V, RESET voltage = 1.50V
Test site = axon201-210 // neuron 211-220
Trigger period = 20ms, Gp_dly = 0.00130s, Gm_dly = 0.00020s
Spike period = 10-10000μs, Spike counter = 200

Measuring the LIF output by adjusting the interval of input spikes demonstrates that **leak function** and **refractory period** are properly working in our LIF circuitry.

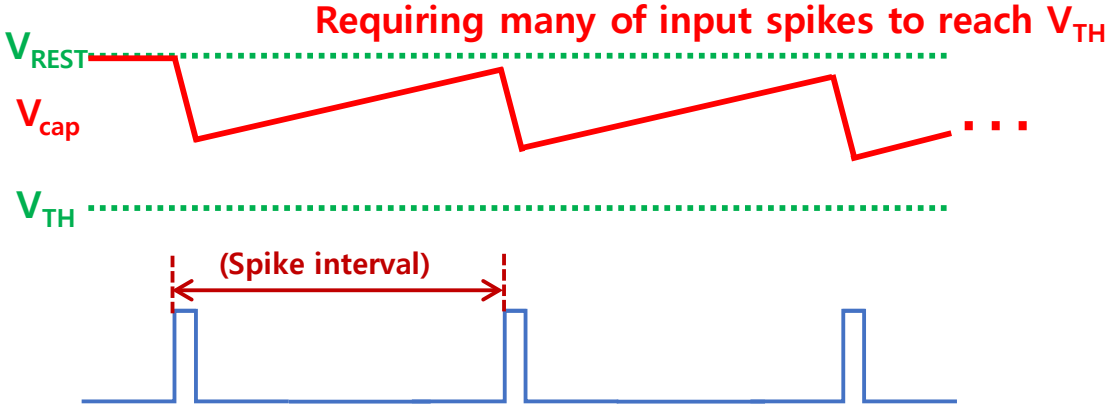


This work shows not only
bipolar synaptic weights are implemented well,
but also **i) the leaky function**
ii) the refractory period.

- Small input spike interval



- Large input spike interval



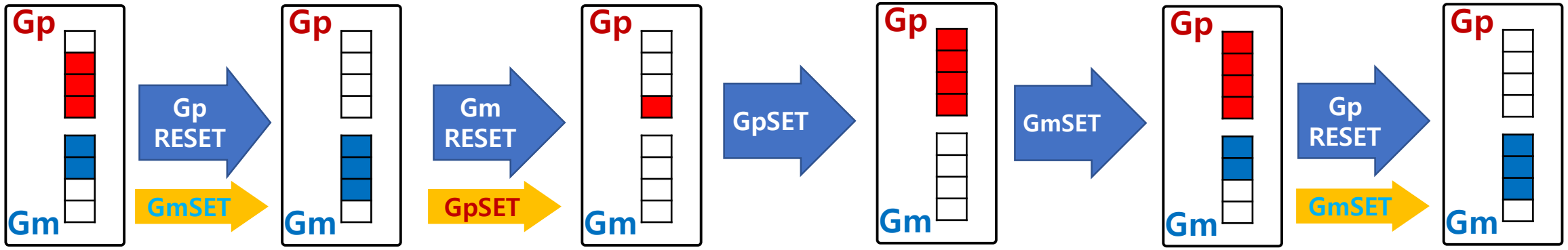
Gradual PGM & LIF results

[PGM & LIF test flow]

Virgin (Initial state?) → All RESET x3 (Clean up before the test) → Gp SET × 10 (Increased?) → Gm SET × 5 (Decreased?) → All RESET x3 (Clean up after the test)

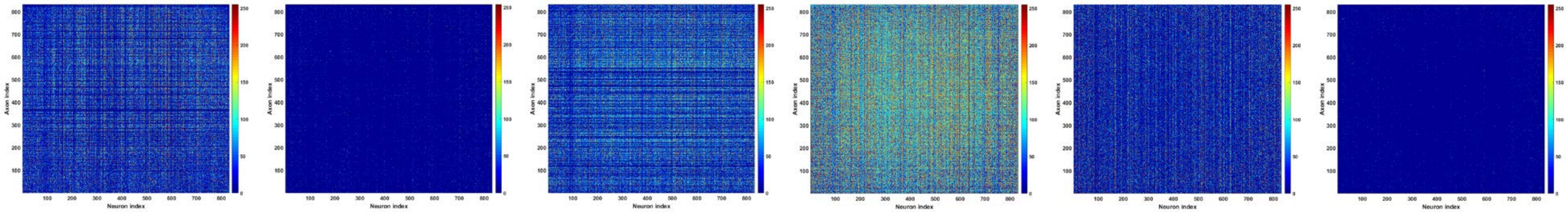
[Schematics of the programming] SET | RESET

SET : SET processes occur inevitably during the RESET process.



LIF output: - LIF output: ↓ LIF output: ↑ LIF output: ↑ ↑ LIF output: ↓ LIF output: ↓ ↓

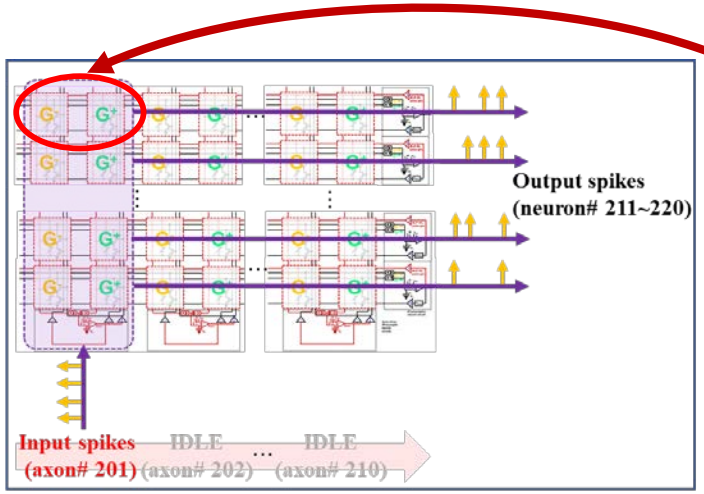
[LIF results of the programming, 832*832=692,224 synapses]



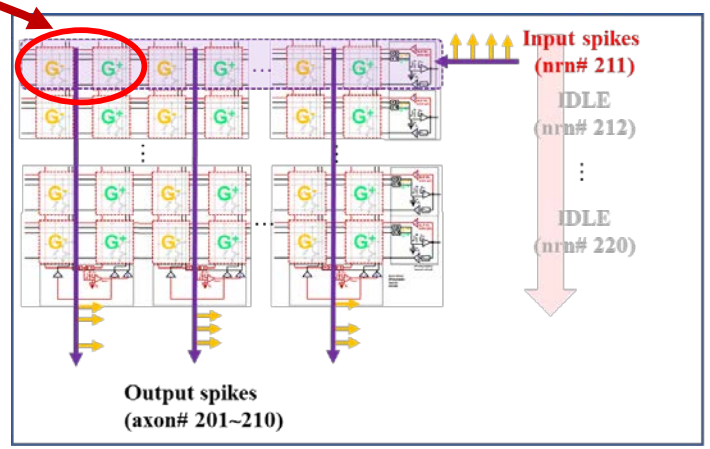
From the results, we can program all PCM cells gradually on both Gp & Gm

Why should we check ‘LIF vs BLIF’ = ‘Bidirectional Synaptic connection’

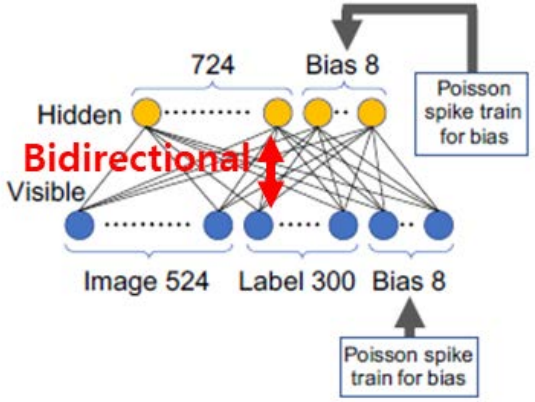
Bidirectional connection with the same synaptic weight.



[LIF test schematic diagram]



[BLIF test schematic diagram]



[Implemented RBM network]

Since the spiking RBM requires both forward- and backward-propagation,
We should demonstrate and optimize the symmetrical performance of LIF and Backward LIF (BLIF).

[LIF vs BLIF experiment]

All RESET every cell → Gp SET × 10 (Increased?) → LIF (Axon input) → BLIF (Nrn input) → All RESET x3 (Clean up after the test)

Programming condition: STDP BL PW : 5
STDP BL Times: 2
BL Interval : 500 (10μs)
AxNr Interval : 60μs

LIF & BLIF condition : # of input spikes: 200
Spike interval : 1,000μs

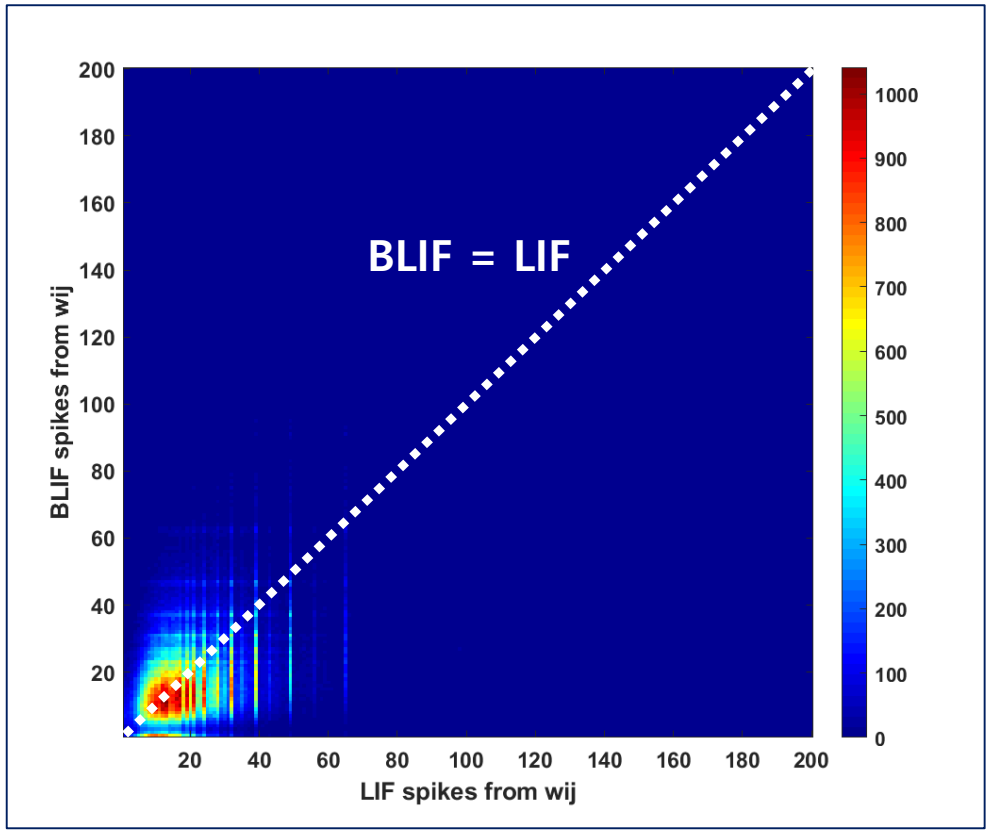
Finding the condition of **pulse widths** of LIF_WL & BLIF_WL,
which shows good balance in the output of LIF & BLIF.

Result of LIF vs BLIF

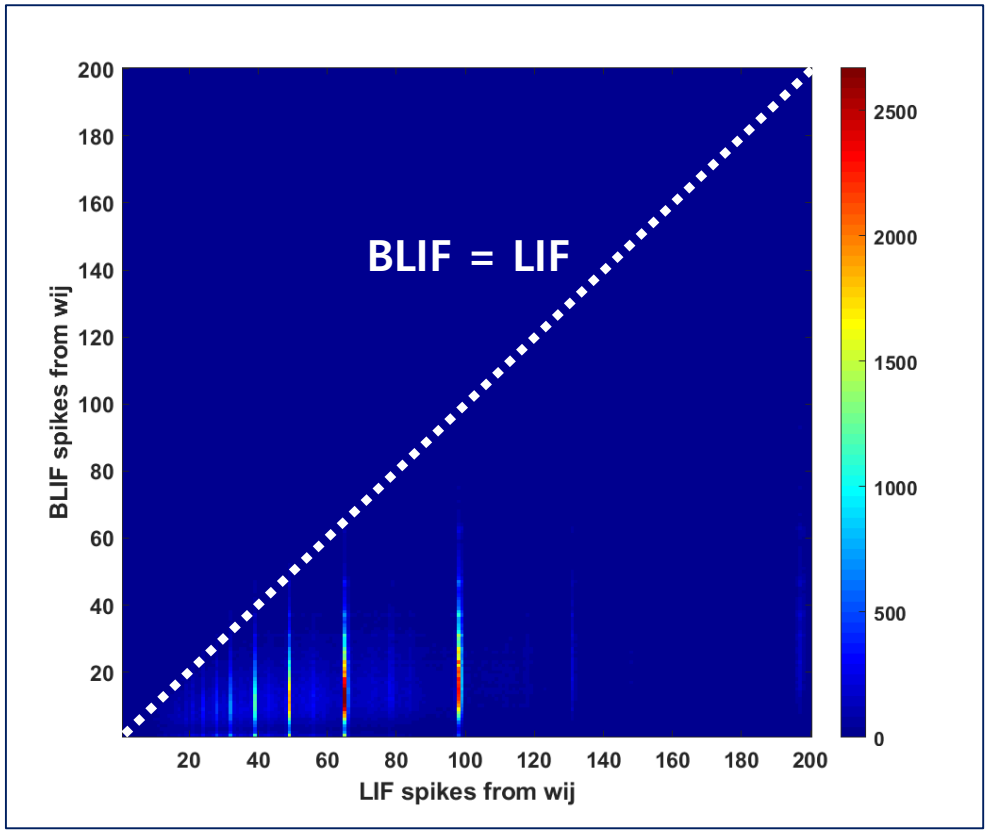
LIF conditions
Fire times : 200 (# of input spikes)
Fire interval : 1,000μs

[Comparing the best and the worst balance]

1. LIF PW0 (53.4 ns) vs BLIF PW7 (211.3 ns)



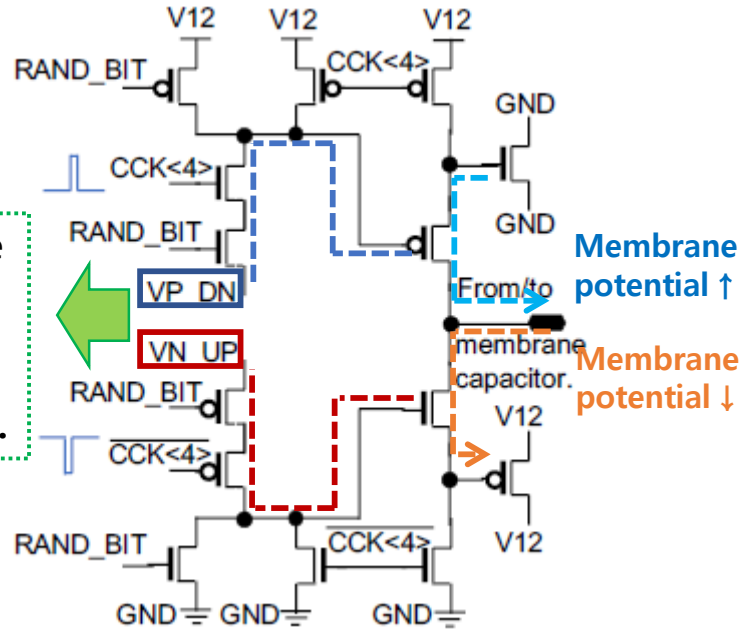
4. LIF PW7 (250.9 ns) vs BLIF PW7 (211.3 ns)



We finally fine-tuned the chip conditions for well-balanced between forward & backward.

Experimental optimization of random walk: Sigmoid firing probability

[Optimizing method]

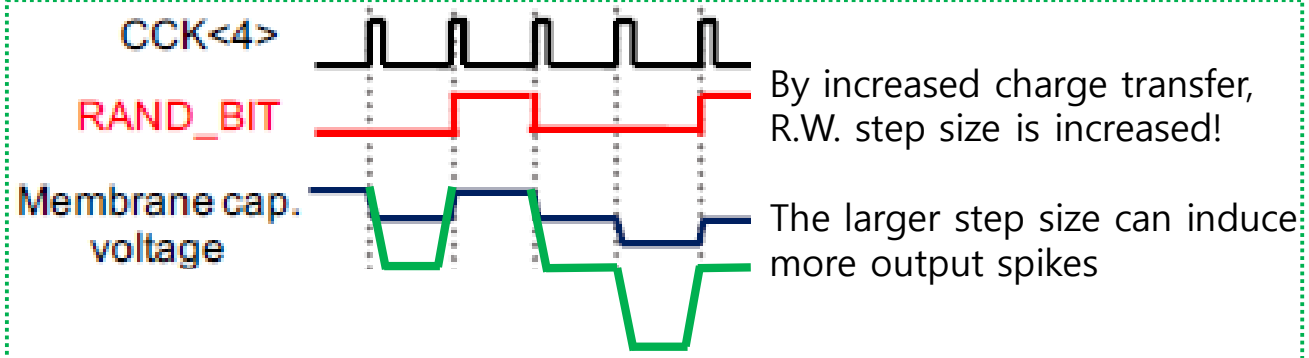


Membrane potential ↑

Membrane potential ↓

Those vol. sources can modulate the quantity of charge transfer.

[Vp_dn ↓] &
[Vn_up ↑] make larger change.



By increased charge transfer, R.W. step size is increased!

The larger step size can induce more output spikes

1. **Increasing the R.W. step size to obtain more output spikes.**

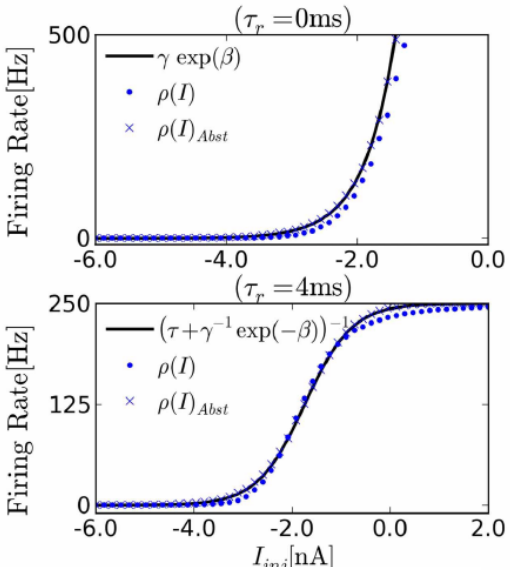
2. **We should maintain the step size of upper and lower. (A broken balance may affect the overall probability.)**

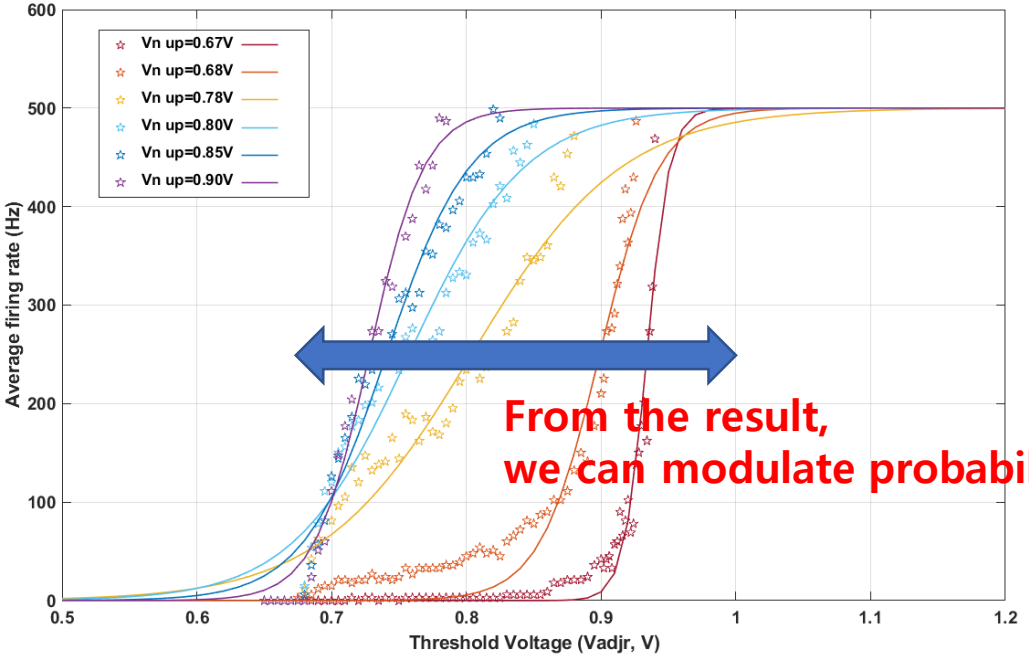
[Approximation with exponential fn.]

Neftci *et al.*, Frontiers in Neurosci., **2014**, *Vol.7*, 272

$$\rho = (\tau_r + \gamma^{-1}\exp(-\beta u))^{-1}$$

Fitted constants: $\gamma = 2.708 \times 10^{-4}$
$\beta = 17.95$
$(\tau_r \cong 2ms)$

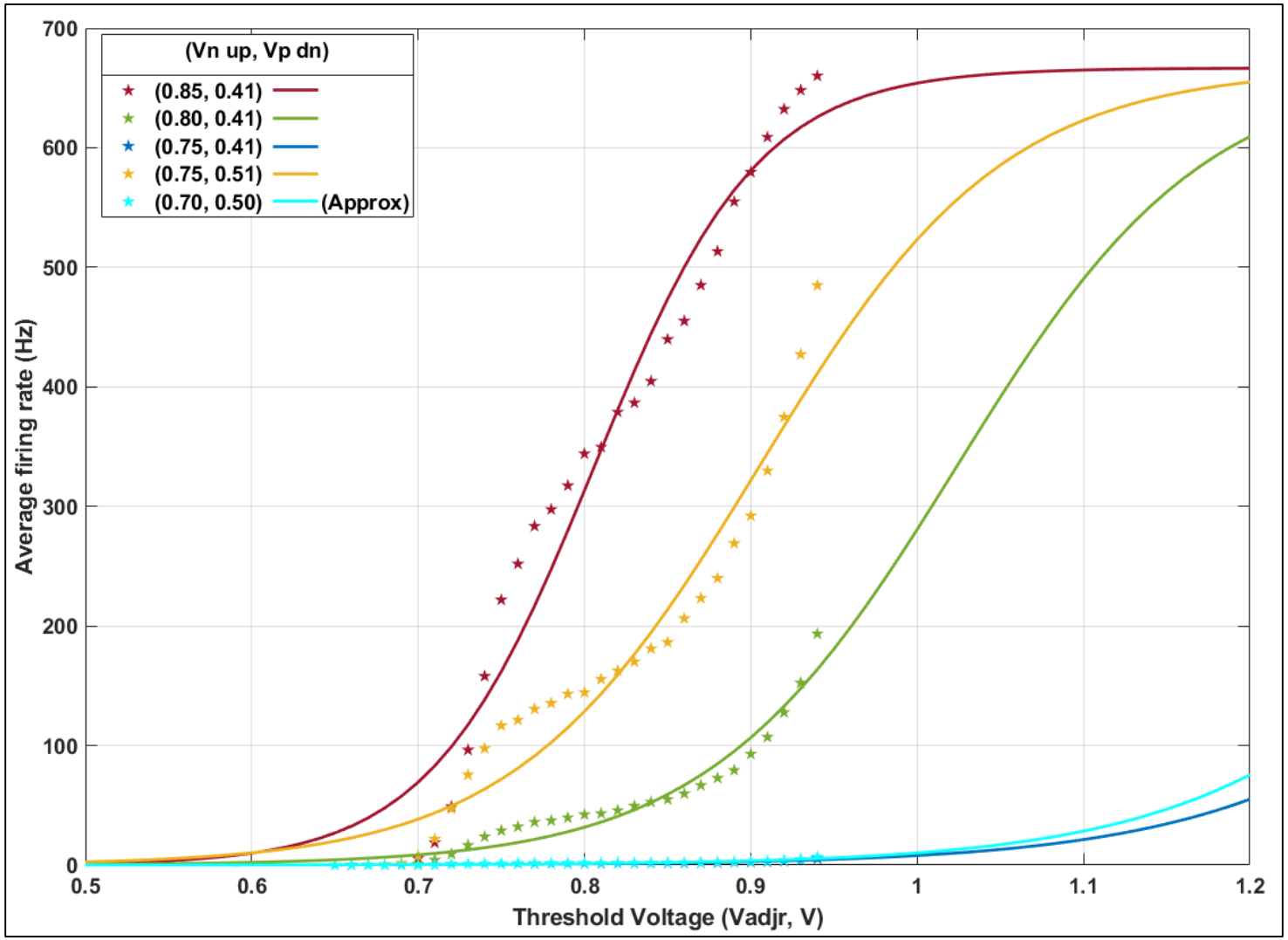




From the result, we can modulate probability!

Random walk results from sweeping Vadjr

[Vadjr sweep on 5 condition cases]



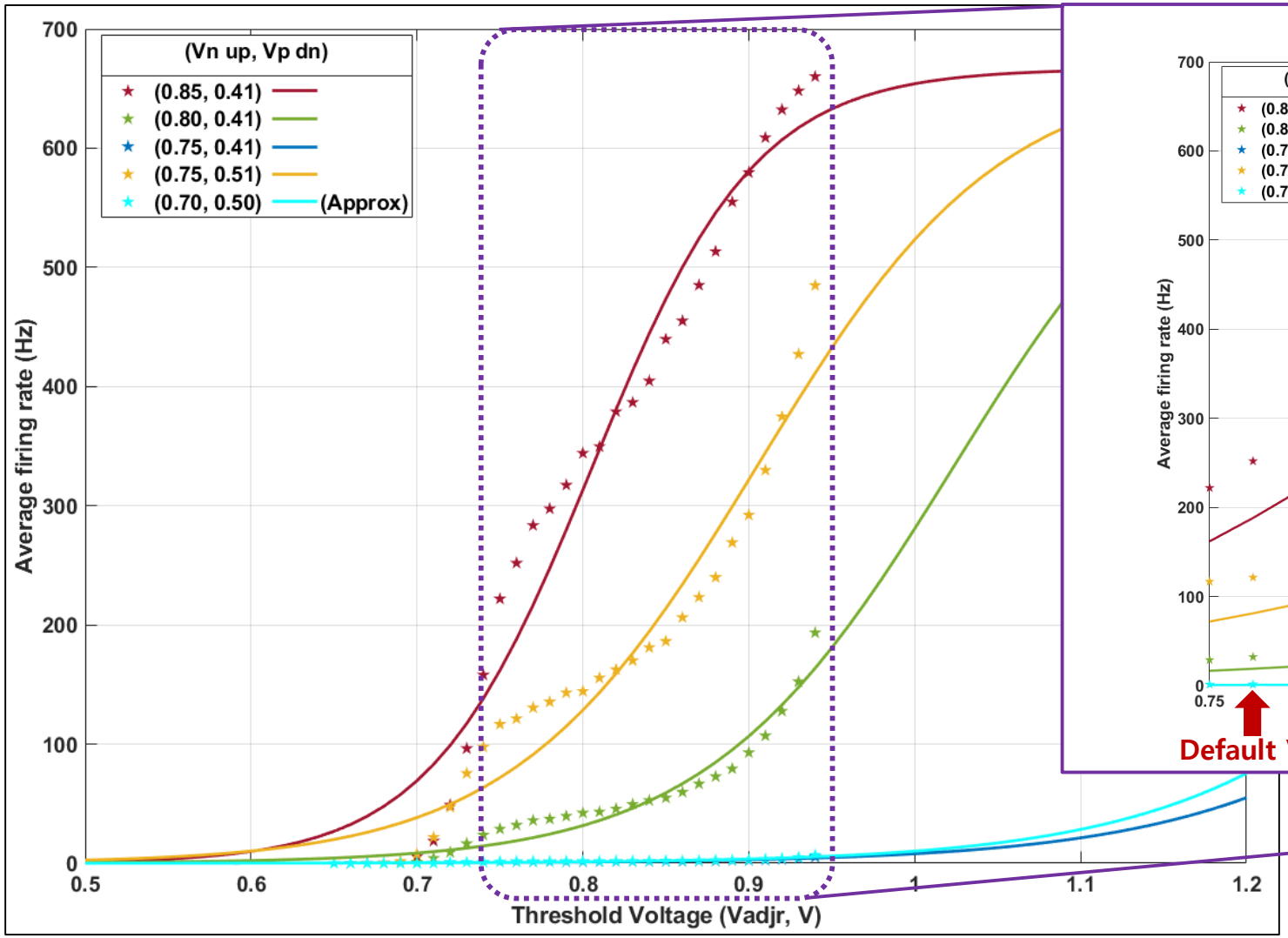
Rndwlk conditions

# of LFSR clks : 33280 clks (10μs/clk)

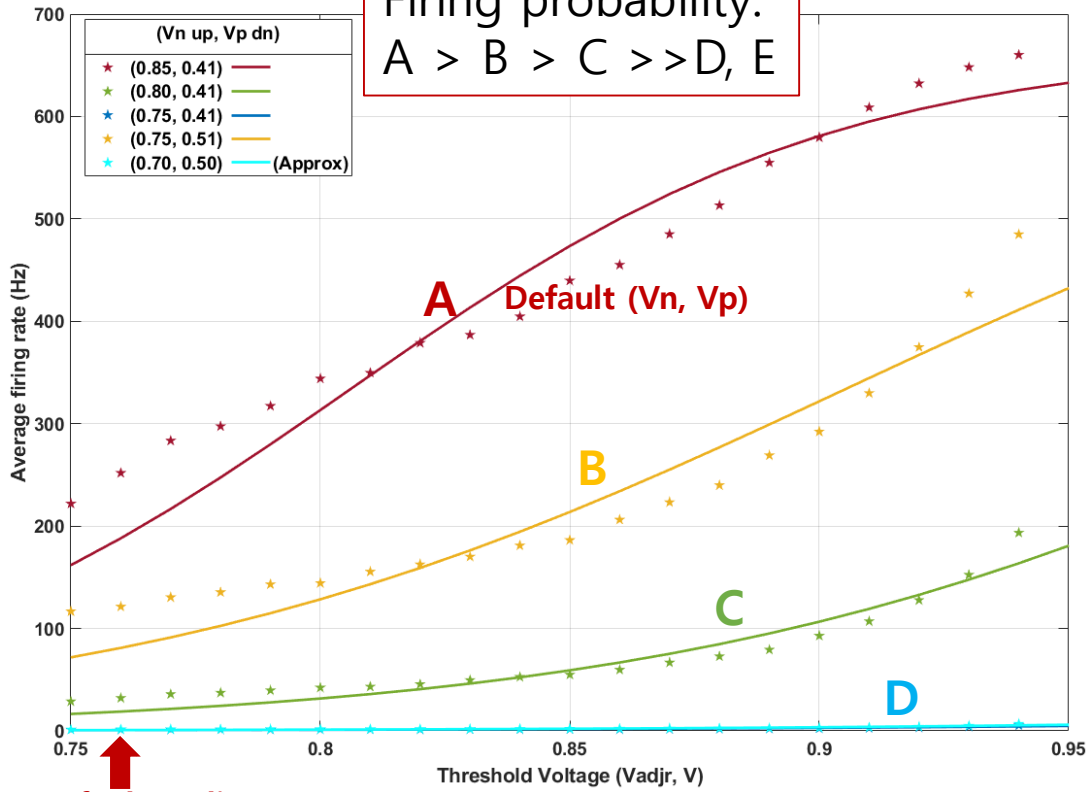
Vadjr sweep range : 0.65 ~ 0.94V

Random walk results from sweeping Vadjr

[Vadjr sweep on 5 condition cases]



Firing probability:
A > B > C >>D, E



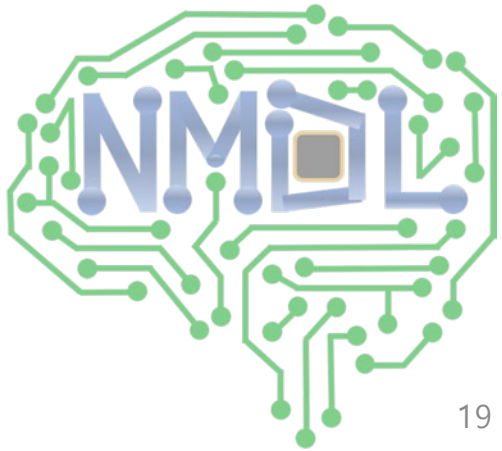
How do these voltage conditions bring differences in MNIST scores?

Contents

1. Introduction

2. My Previous works
   - SNN-RBM chip test results

3. Future Research Plans & Summary
   - Toward to novel algorithms

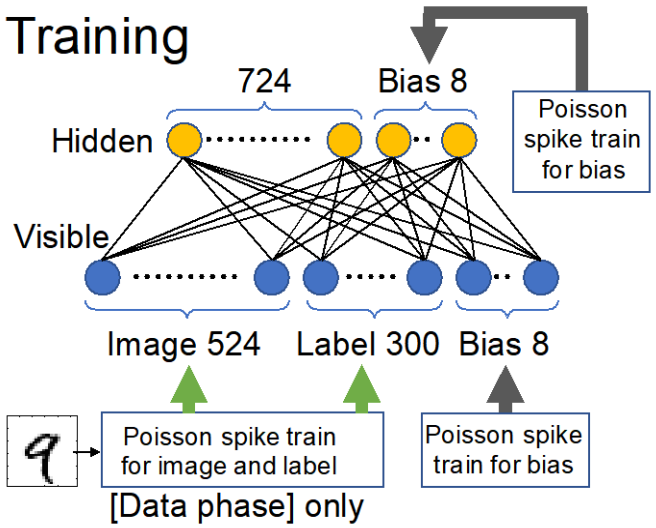


Algorithm implementation on the SNN hardware

[Mapping novel algorithms]

Combinatorial Optimization problem!

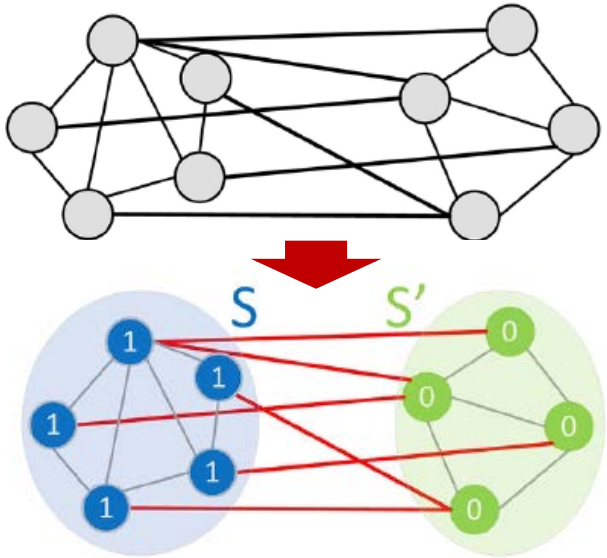
i. Spiking RBM (Restricted Boltzmann machine)



The goal:

Maximize training accuracy.

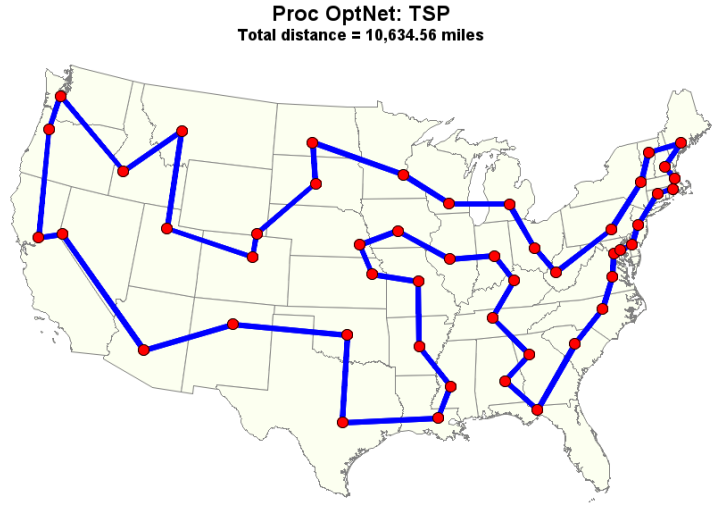
ii. Maximum cut problem (Max-Cut)



The goal:

Maximize the sum of edge weights between the sets. (red edges)

iii. Traveling Salesman Problem (TSP)



The goal:

Find out the minimal route visiting all cities.

Summary

- Neuromorphic hardware is promising technology for future edge devices.

: ① SNN is investigating as a 3rd generation of AI, with greatly reduced power consumption.
: ② RBM is one of the machine learning algorithms based on probability.
: ③ PCM is one of the emerging memory devices which is matured and commercialized(Intel).

- Demonstrations of SNN-RBM chip

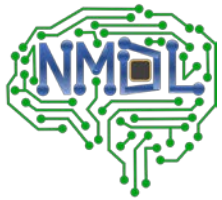
: ① Gradual change of bipolar synaptic weights.
: ② Symmetric operation on bi-directional synaptic connections.
: ③ Implementation of firing probability by random walk circuitry.

- Implementation of novel algorithms

: Not only MNIST-handwritten recognition, probabilistic problems such as 'Max-cut' and 'TSP'.

➡ We are still debugging test algorithms and searching proper conditions

Neuromorphic
Materials and
Devices
Lab.



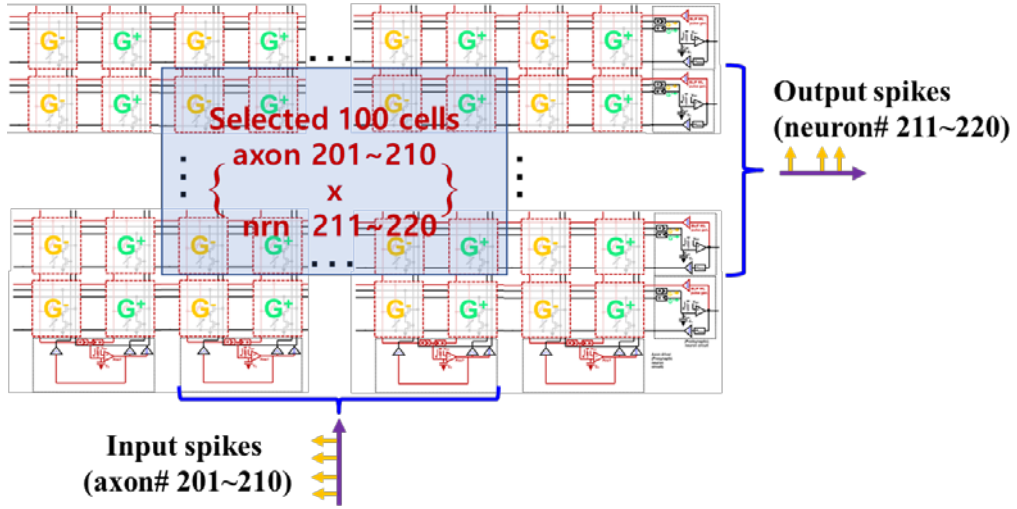
Thank You!

Neuromorphic
Materials and
Devices
Lab.

-Appendix-

LIF functional test

Leaky-Integrate and Fire (LIF) is core operation of our SNN chip!
So first, we should confirm that our circuitry executes LIF correctly.



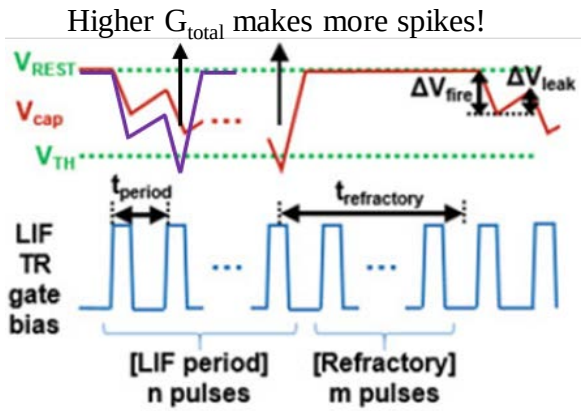
I tested on 10x10 cells sample group among the whole array.

LIF circuitry makes outputs according to G of synaptic cells.

{ SET programming on G_p → G_{total} increase → LIF outputs ↑
G_m → G_{total} decrease → LIF outputs ↓ }

($G_{total} = G_p - G_m$)

[Schematic firing process]



[Experimental results of LIF on IEDM 2019 paper]

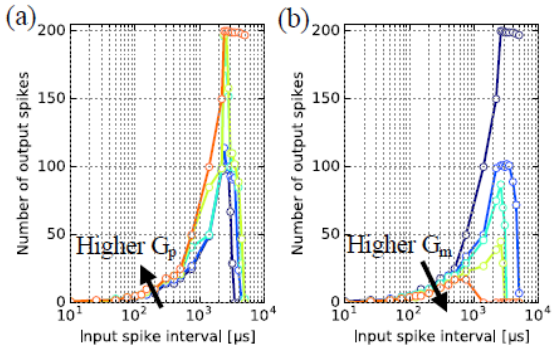


Fig. 14. Number of spikes from a LIF neuron (a) SET programing of G_p with fixed G_m results in more spikes. (b) SET programing of G_m with fixed G_p results in less spikes.

This work shows that
bipolar synaptic weights are implemented well,
and circuitry operate proper LIF function.

LIF vs BLIF experimental results

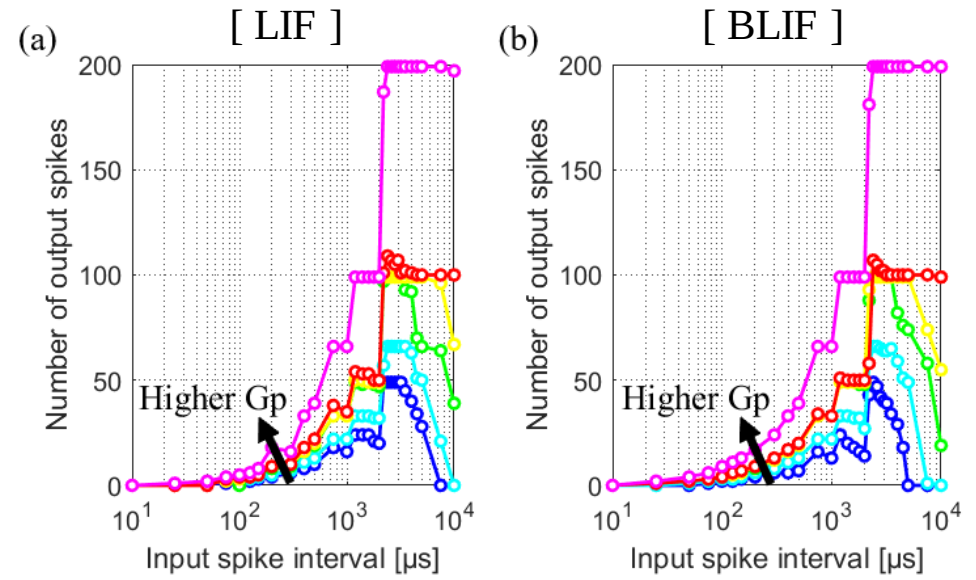
Demonstrating the symmetrical performance of LIF and Backward LIF (BLIF),
Since the spiking RBM requires both forward- and backward-propagation.

Test condition

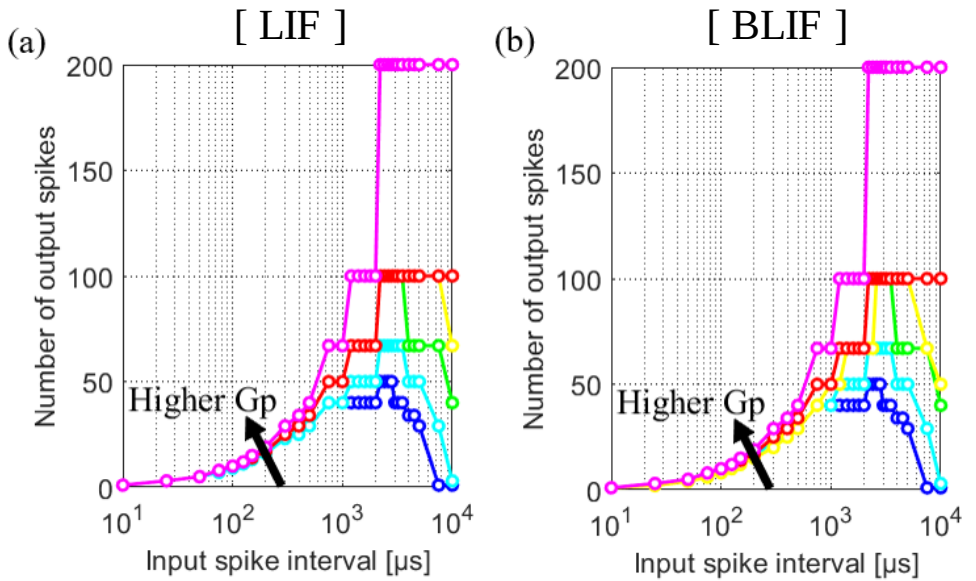
CCK_BASE 0, 1 = 10us, CCK_Leak = 800μs (→ fixed by PG)
SET voltage = 0.75V, RESET voltage = 1.50V
Test site = axon210 // neuron 214 **(One cell)**
Trigger period = 20ms, Gp_dly = 0.00130s, Gm_dly = 0.00020s
Spike period = 10-10000us, Spike counter = 200

Simulation figures are provided by *Wonseok Choi*, SNU

[Hardware in-situ results]

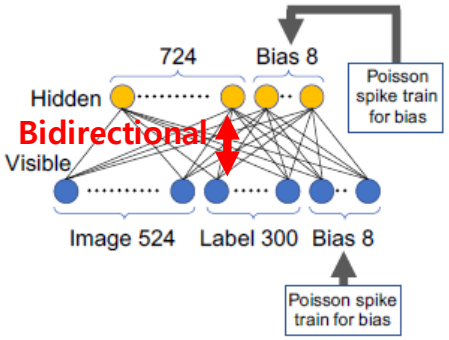


[Simulation results]



Performing the identical LIF functional experiment on the hidden side neuron circuit, we confirmed that bidirectional connections are implemented well.

In addition to hardware in-situ results,
the comparable tendency of simulation results indicates LIF circuitry is well-fabricated.



Conditions of pulse widths for ‘LIF vs BLIF’ test

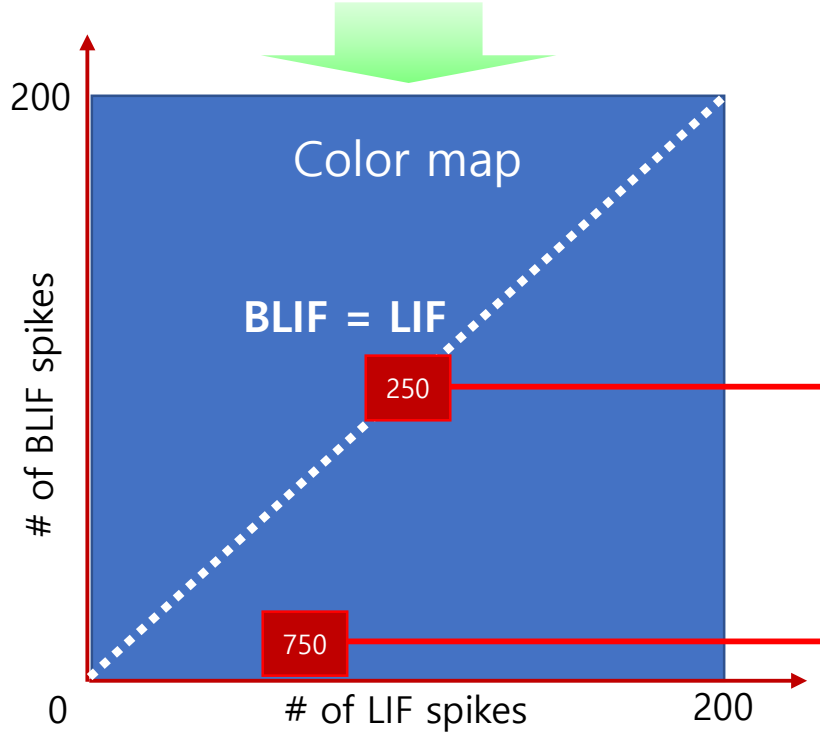
LIF_WL pulse width **(TMH_VPR_A2 <2:0>)**

1) PW0 (53.4 ns)
2) PW4 (116.2 ns)
3) PW6 (205.0 ns)
4) PW7 (250.9 ns)

BLIF_WL pulse width **(TMH_VRP_N1 <2:0>)**

1) PW7 (211.3 ns)

Since BLIF output is less than LIF, fixed to the longest width.



Example 1. (Good balance)
It means we have 250 synapses showing (LIF spikes, BLIF spikes) = (100, 100).

Example 2. (LIF dominates)
It means we have 750 synapses showing (LIF spikes, BLIF spikes) = (66, 0).