Low-Power Spiking Neural Network with Clock-gating technique

S1290033

Rui Shiota

Content

Content

SNNs and IF Neuron

Spiking Neural Networks (SNNs)

- Brain-inspired models
- Mimic biological neurons
- Communicate and compute using spikes [1]

Integrate-and-Fire (IF) neuron

$$\mathrm{V}_j(t) = V_j(t-1) + \sum_{i=1}^{n} w_{i,j} \times x_i(t-1) \quad \left(x_j(t) = \begin{cases} 1 \; if \; V_j(t) > V_{thres} \\ 0 \; otherwise \end{cases} \right)$$

- $V_j(t)$: membrane potential of neuron j at time step t
- $w_{i,j}$: synapse weight between neuron i and neuron j
- $x_i(t-1)$: output of the presynaptic neuron i
- V_{thres} : threshold value [4]

Power Consumption in VLSI

$$Total\ Power = Dynamic\ Power + Static\ Power$$

$$Dynamic\ Power = f_{sw} \times C_L \times V_{cc}^2 + T_{SC} \times I_{peak} \times V_{cc}$$

$$Static\ Power = V_{cc} \times I_{cc}$$

- f_{sw} : switching frequency
- C_L : dynamic effective capacitance
- V_{cc} : voltage applied to a logic IC
- T_{SC} : shortcut-circuit time period
- I_{peak} : peak current
- I_{CC} : static supply current of the IC [8]

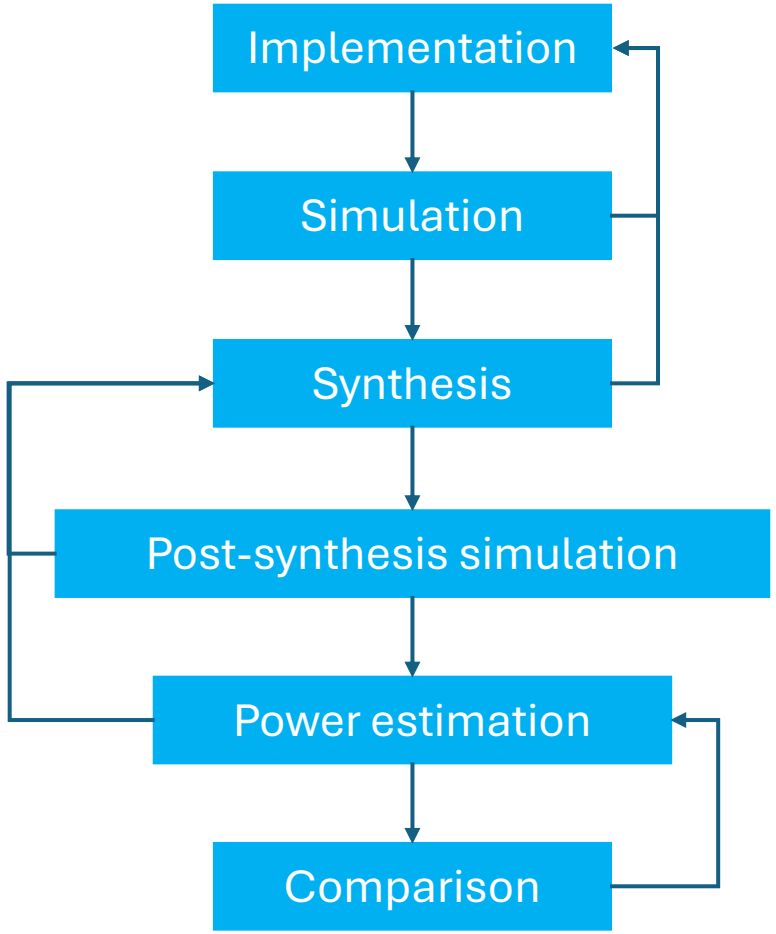
Why clock gating?

- Reduce operational costs and environmental impact [6][7]
- Clock gating: method to save power by turning off clock [9]
- Leading to significant power savings
- 100 to 1,000 times as energy-efficient as non-neuromorphic systems [5]

Content

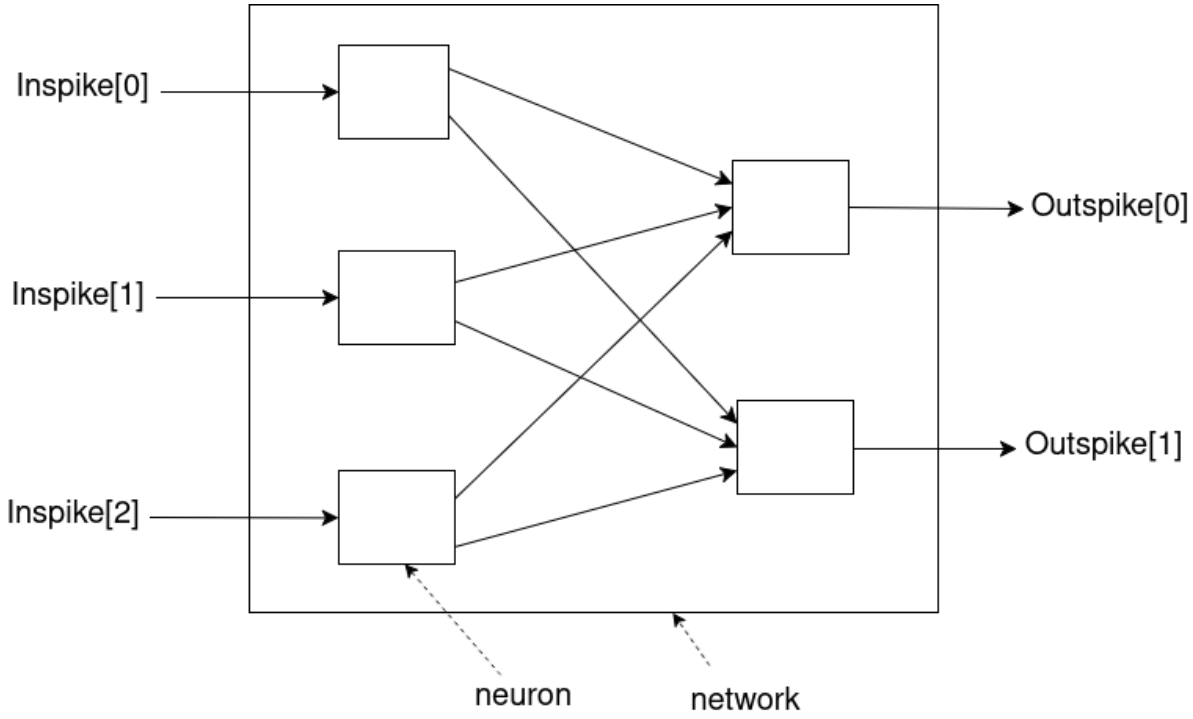
- Introduction
  - SNNs and IF Neuron
  - Power Consumption in VLSI
- Methodology
  - Procedure
  - Design of IF Neuron
  - Algorithm
  - Clock Gating Overview
- Results
  - Experimental Information
  - Simulation
  - Power Estimation
  - Discussion
- Conclusion

Procedure


Implementation
Simulation
Synthesis
Post-synthesis simulation
Power estimation
Comparison

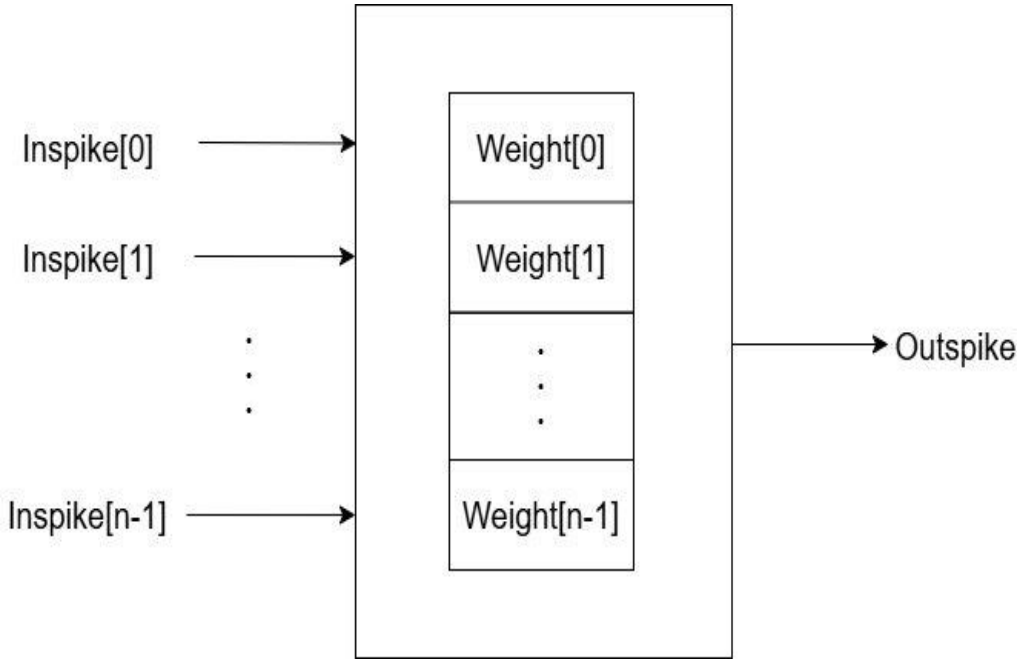
Design of IF Neural Network

Overall architecture of IF Neural Network



- Implement 3:2 network

Neuron architecture



- Each neurons have one memory
- The number of weight is determined by the number of inspike

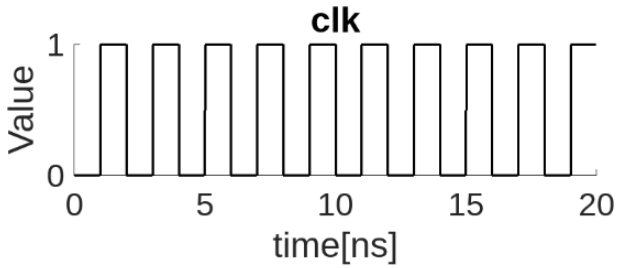
Outspike Computation

Algorithm 1 Calculate Outspike

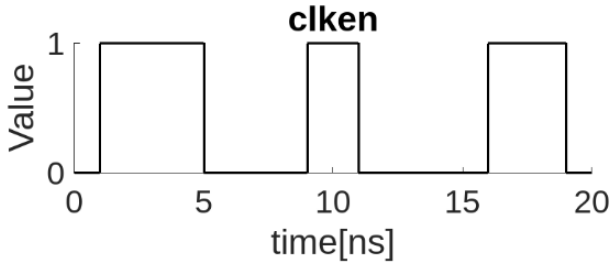
1: $V = V_{reset}$
2: $i = 0$
3: **for** $i < n$ **do**
4: $\quad V = V + inspike[i] * Weight[i]$
5: **end for**
6: **if** $V \geq Threshold$ **then**
7: $\quad outspike = 1$
8: $\quad V = V_{reset}$
9: **else**
10: $\quad ouspike = 0$
11: **end if**

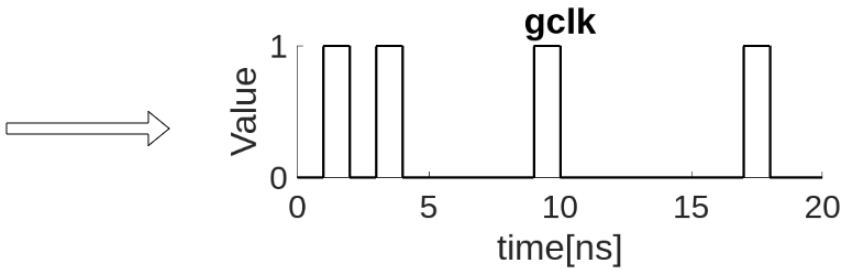
- V : membrane potential
- V_{reset} = 0
- i : index of inspike and Weight
- n : number of bit of inspike

Clock Gating Overview

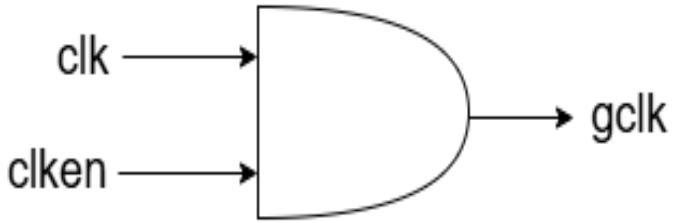


×





gclk = clk ⋀ clken



$$Dynamic\ Power = f_{sw} \times C_L \times V_{cc}^2 + T_{SC} \times I_{peak} \times V_{cc}$$

Content

- Introduction
  - SNNs and LIF Neuron
  - Power Consumption in VLSI
- Methodology
  - Procedure
  - Design of LIF Neuron
  - Algorithm
  - Clock Gating Overview
- Results
  - Experimental Information
  - Simulation
  - Power Estimation
  - Discussion
- Conclusion

Experimental Information

Coding: Verilog HDL

Simulation:

- Simulation time: 435 ns
- Clock cycle time: 10 ns

Tool:

- ModelSim (simulation)
- Synopsys Design Compiler (synthesis)
- Synopsys PrimeTime (power estimation)

Library: NangateOpenCellLibrary (45nm)

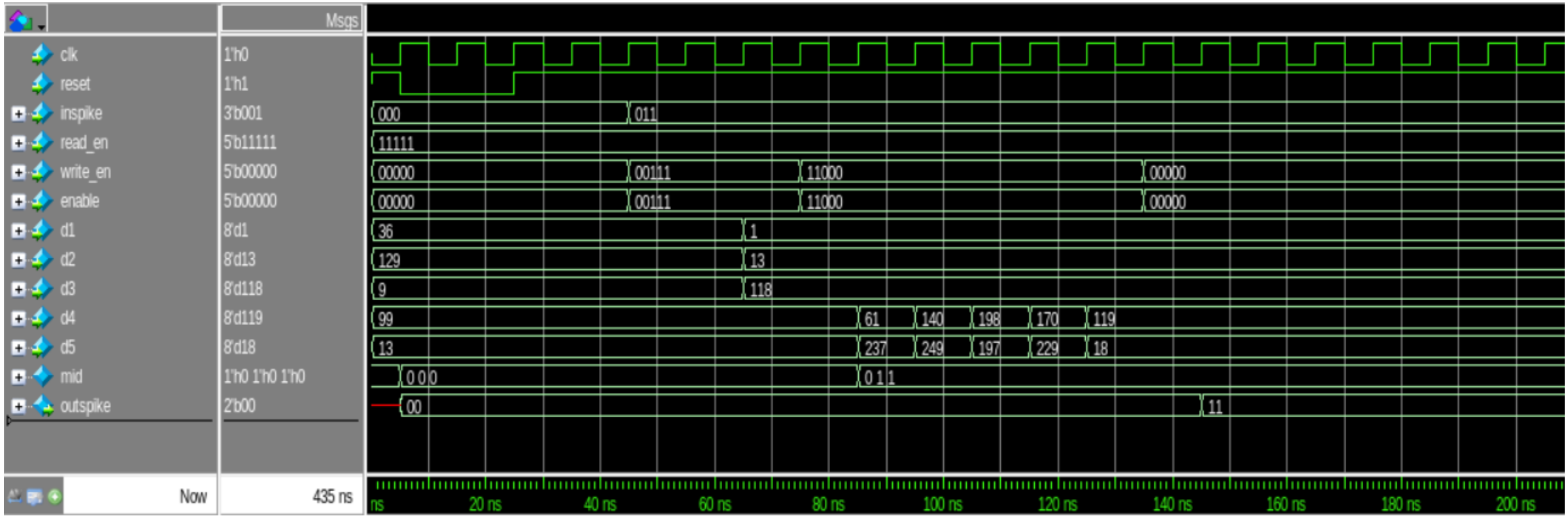
Experimental results

- Simulation
  - No clock gating network
  - Clock-gated network
  - Constantly clock-gated network
- Power estimation
  - No Clock Gating and Clock-gated Network
  - No Clock Gating and Constantly Clock-gated Network

Role of Signals

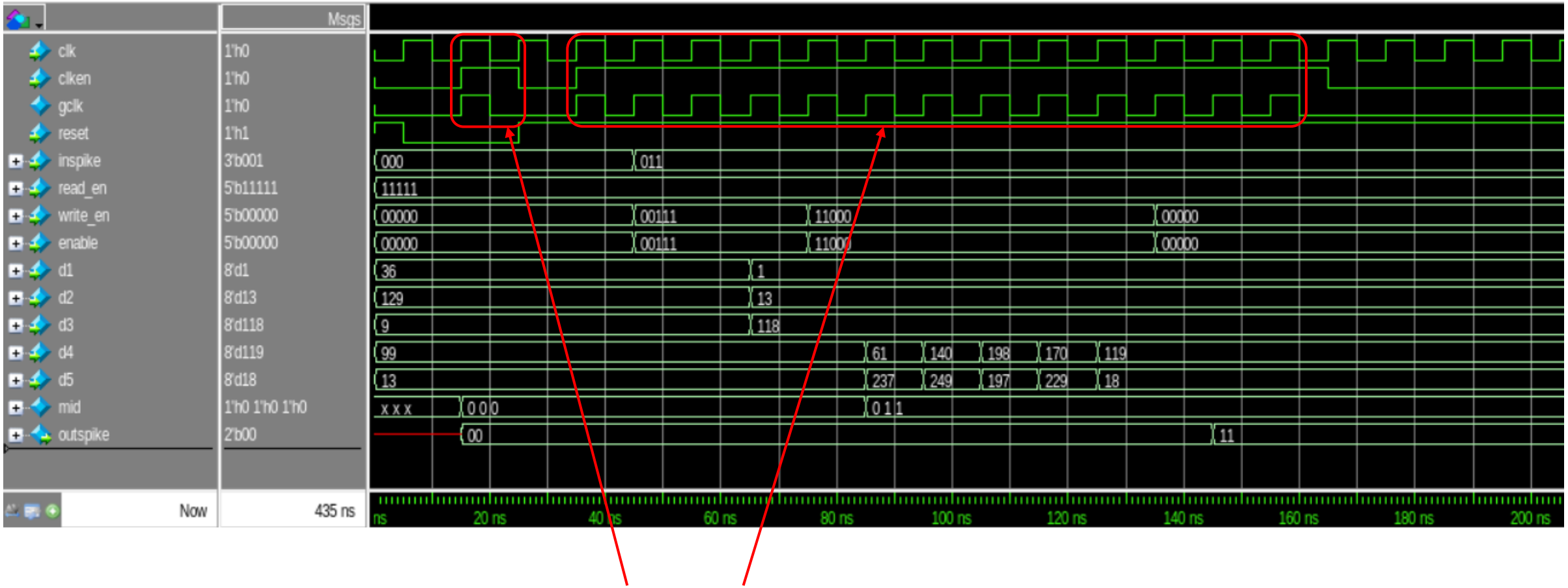
Signal	Role
clk	clock signal
clken	clock enable signal
gclk	clock-gated clock signal
reset	reset signal
inspike	input spike signal
read_en	signal to read weight from memory
write_en	signal to write weight to memory
enable	signal to work neurons
d1, d2, d3	weight of neurons in first layer
d4, d5	weight of neurons in Second layer
mid	signal to connect first and second layer
outspike	output signal

Simulation of No Clock Gating Network



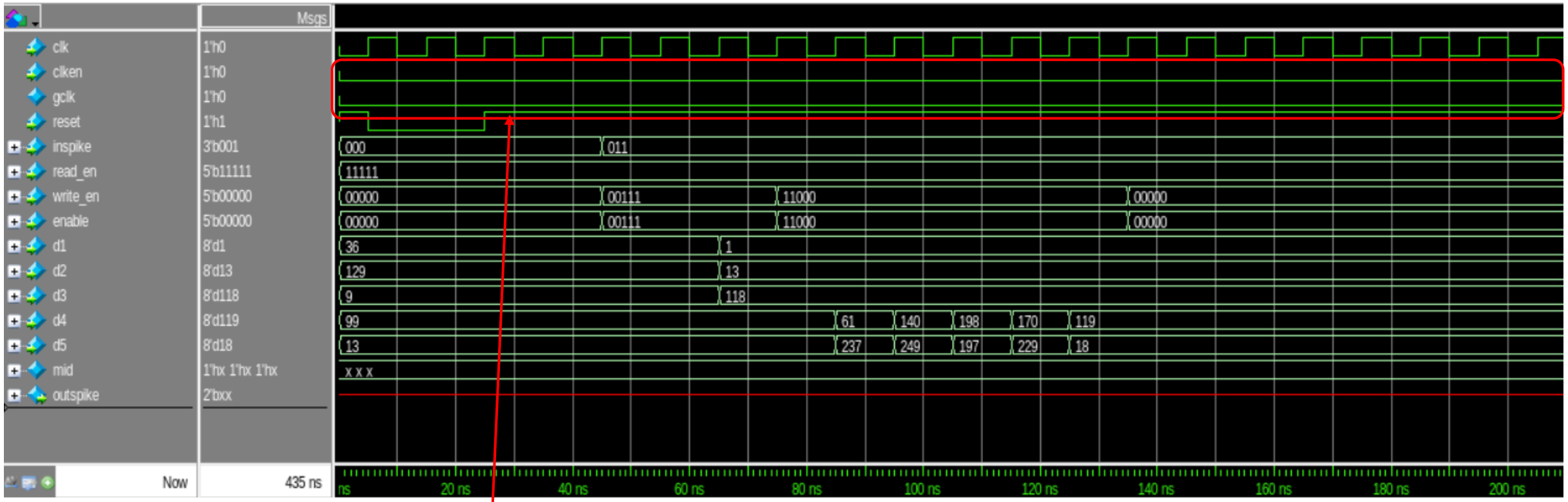
- Computation is done every time clock rises

Simulation of Clock-gated Network



Network works at this timing

Simulation of Constantly Clock-gated Network



- Clock enable is always "0"
- No computation

Power Comparison of No Clock Gating and Clock-gated Network

	No Clock Gating Network	Clock-gated Network	Constantly Clock-gated Network
Dynamic Power	$2.08 \times 10^{-4} W$	$1.45 \times 10^{-4} W$	$5.14 \times 10^{-6} W$
Static Power	$4.28 \times 10^{-5} W$	$4.29 \times 10^{-5} W$	$4.41 \times 10^{-5} W$
Total Power	$2.51 \times 10^{-4} W$	$1.89 \times 10^{-4} W$	$4.95 \times 10^{-5} W$

Comparison of No Clock Gating and Clock-gated Network

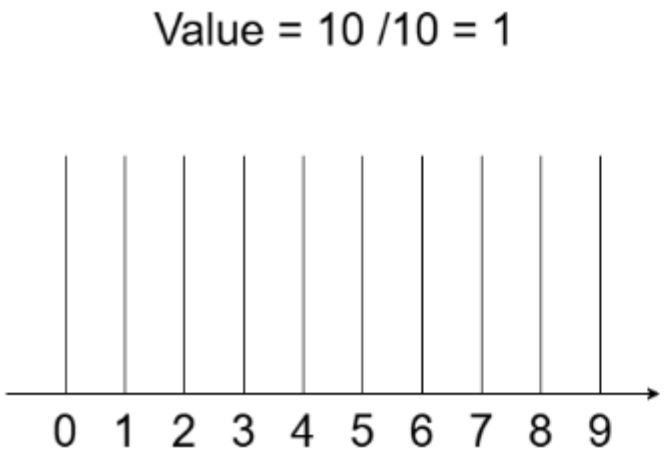
- Dynamic Power: 30.23% ↓
- Static Power: almost unchanged
- Total Power: 24.70% ↓

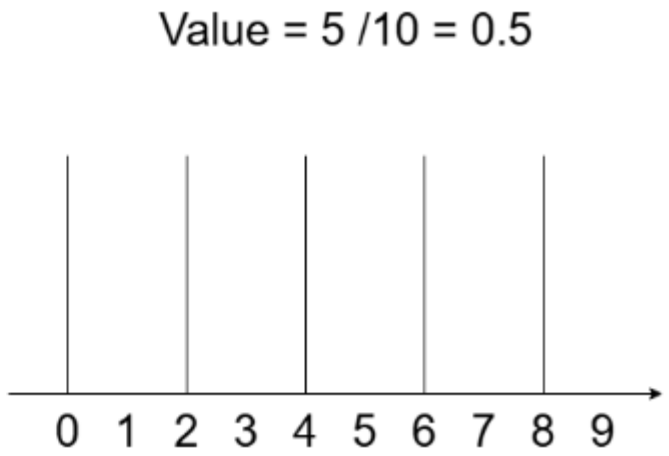
Comparison of No Clock Gating and Constantly Clock-gated Network

- Dynamic Power: 97.53% ↓
- Static Power: 3.04% ↑
- Total Power: 80.28% ↓

Discussion

- Successfully reduced power by clock gating
- Evaluate impact of clock gating
- Use “rate coding’’ to reduce power further in the future
- Presents information through spiking rates [10]





Content

Conclusion

- IF neuron with and without clock gating
- Simulation, synthesis and power estimation
- Reduce power consumption by 24.70%
- Reduce further by “rate coding’’

Reference

[1] Prithwineel Paul, Petr Sosik, Lucie Ciencialova, “A survey on learning models of spiking neural membrane systems and spiking neural net works“, arXiv, March 2024.

[2] Wei Fang, Zhaofei Yu, Yanqi Chen, Timothee Masquelier, Tiejun Huang, Yonghong Tian, “Incorporating Learnable Membrane Time Constant to Enhance Learning of Spiking Neural Networks“, arXiv, August 2021.

[3] Xiaoyan Fang, Derong Liu, Shukai Duan, Lidan Wang, “Memristive LIF Spiking Neuron Model and Its Application in Morse Code“, frontiers, April 2022.

[4] “Robust Cognitive Brain-inspired Computing System: Architectures and Algorithms“, https://u-aizu.ac.jp/~khanh/share/pubs/ETLTC-2022.pdf.

[5] Bennie Mols, “Making AI more energy efficient with neuromorphic computing“, CWI, March 2024.

[6] David Patterson, Joseph Gonzalez, Urs H¨olzle, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean, “The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink“, IEEE, July 2024.

[7] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga Behram, James Huang, Charles Bai, Michael Gschwind, Anurag Gupta, Myle Ott, Anastasia Melnikov, Salvatore Candido, David Brooks, Geeta Chauhan, Ben jamin Lee,Hsien-HsinS.Lee, BugraAkyildiz, MaximilianBalandat, Joe Spisak,Ravi Jain, MikeRabbat,KimHazelwood, “Sustainable AI: Environ mental Implications, Challenges and Opportunities“, arXiv, March 2024.

[8] “CMOS Power Calculation“, https://resources.pcb.cadence.com/blog/ 2023-cmos-power-calculation.

[9] Nandita Srinivasana, Navamitha.S.Prakasha, Shalakha.Da, Sivaran jani.Da, Swetha Sri Lakshmi.Ga, B.Bala Tripura Sundari, “Power Re duction by Clock Gating Technique“, ScienceDirect, November 2015.

[10] Wenzhe Guo, Mohammed E. Fouda, Ahmed M. Eltawil and Khaled Nabil Salama, “ Neural Coding in Spiking Neural Networks: A Comparative Study for Robust Neuromorphic Systems“, frontiers, March 2021

Thank you for your attention!