Neuromorphic Computing

2. Fundamentals

Ben Abdallah Abderazek, Khanh N. Dang
E-mail: {benab, khanh}@u-aizu.ac.jp

This lecture is based on the book ''Neuromorphic Computing Principles and Organization,'' Publisher: Springer; 2022 edition, ISBN-10 : 3030925242, ISBN-13 : 978-3030925246, by Abderazek Ben Abdallah, Khanh N. Dang

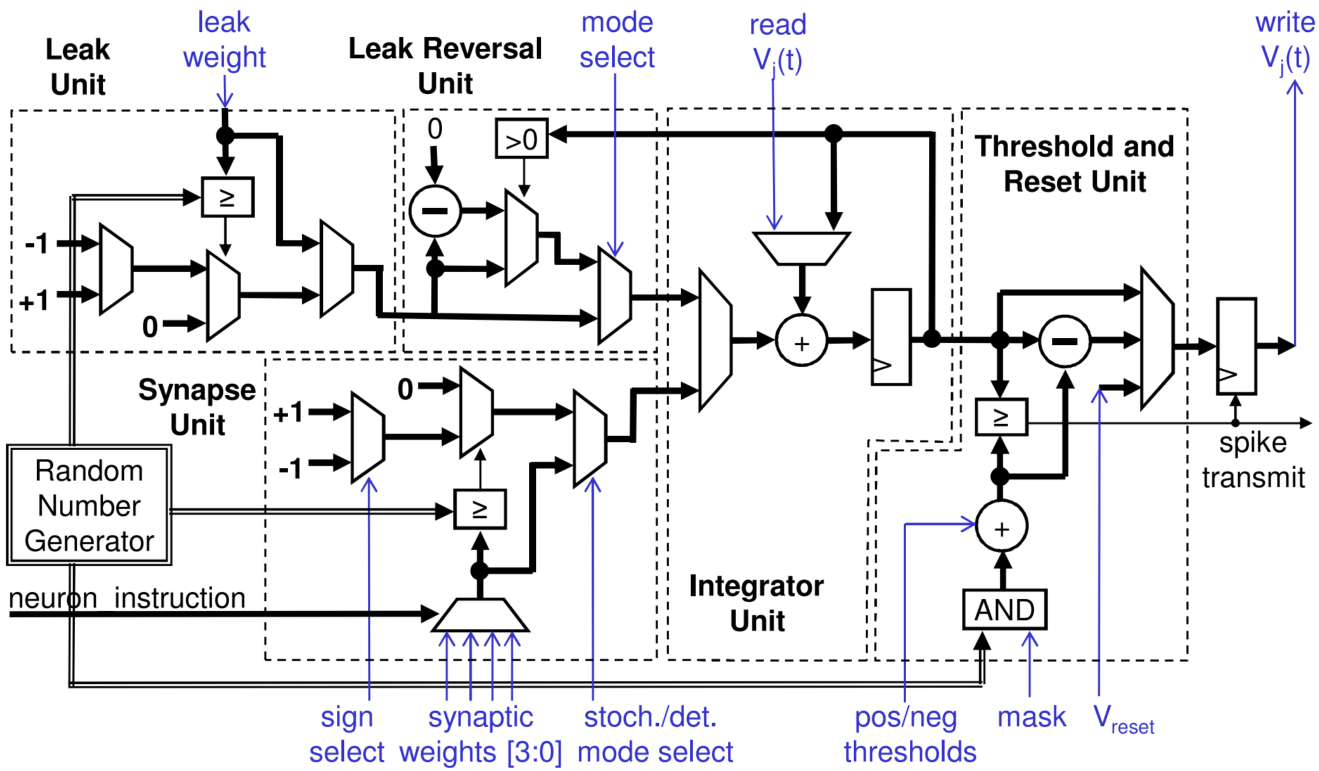
Lecture Contents

1. Spiking Neural Networks
2. Neural Coding Schemes
3. Spiking Neuron Models
4. Learning Algorithms
5. Synapses & Inter-neuron communication
6. Conclusions

1. Spiking Neural Network Introduction

- The term "neuromorphic engineering" was coined by Carverd Mead in the later 1980s
  - The usage of VLSI systems containing electronics analog circuits to mimic neuro-biological architectures.
- Recent neuromorphic systems can be:
  - Digital VLSI systems
  - Analog mixed signal VLSI systems
  - Software based system by CPU/GPU

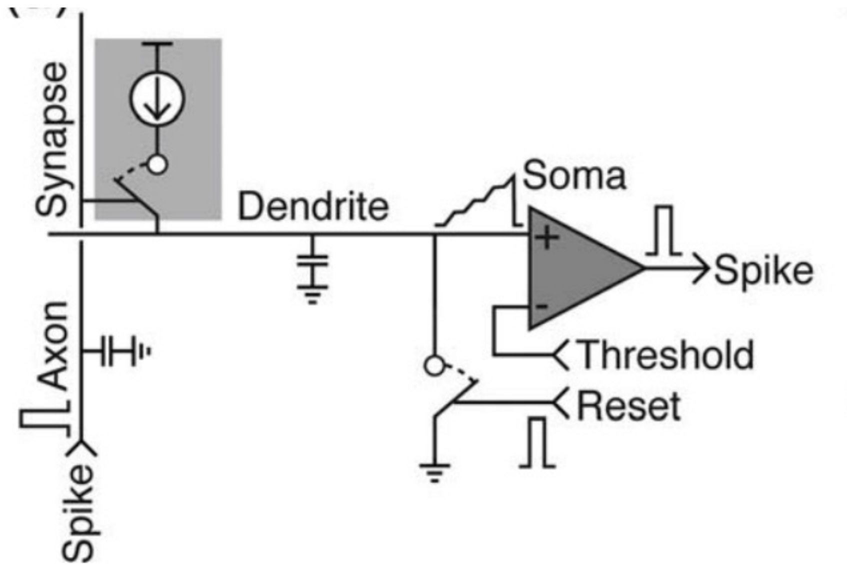
1. Spiking Neural Network
Digital VLSI system



TrueNorth's neuron block

F. Akopyan et al., "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 34, no. 10, pp. 1537-1557, Oct. 2015, doi: 10.1109/TCAD.2015.2474396.

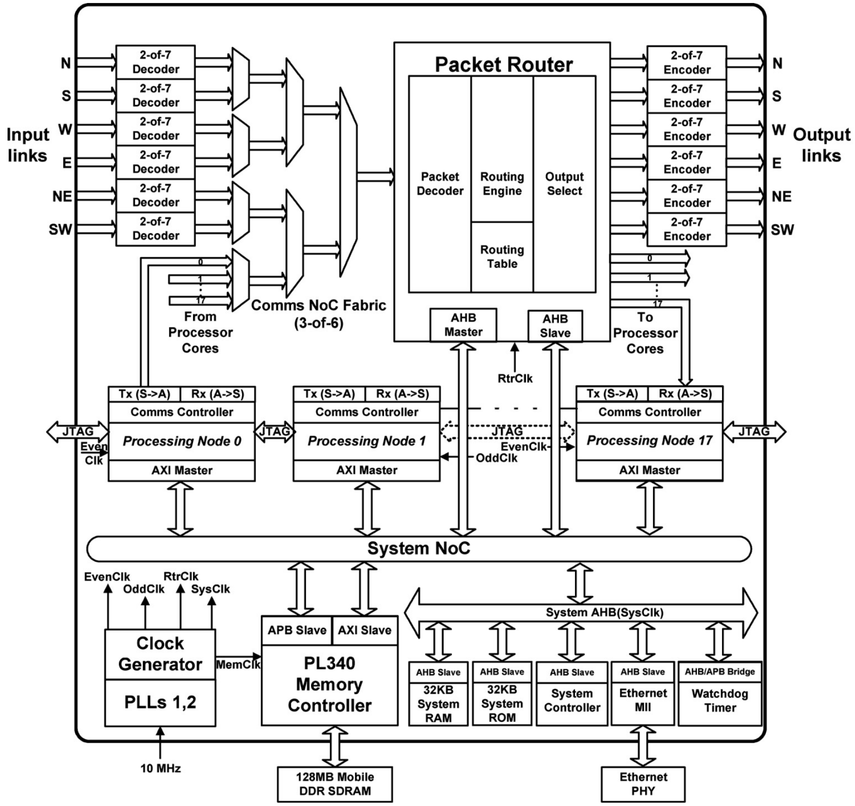
1. Spiking Neural Network
Analog Mixed Signal VLSI system



NeuroGrid's simplified neuron block

B. V. Benjamin et al., "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations," in Proceedings of the IEEE, vol. 102, no. 5, pp. 699-716, May 2014, doi: 10.1109/JPROC.2014.2313565.

1. Spiking Neural Network
Software Based system



SpiNNaker’s simplified chip architecture

E. Painkras et al., "SpiNNaker: A 1-W 18-Core System-on-Chip for Massively-Parallel Neural Network Simulation," in IEEE Journal of Solid-State Circuits, vol. 48, no. 8, pp. 1943-1953, Aug. 2013, doi: 10.1109/JSSC.2013.2259038.

1. Spiking Neural Network
Biological structure

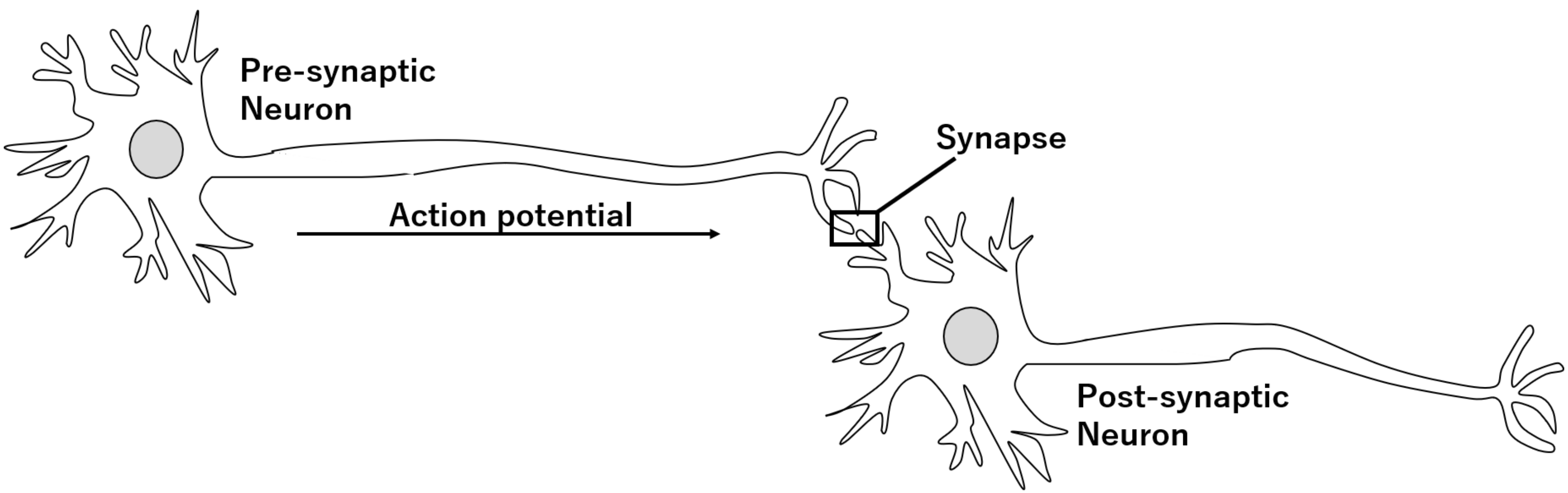


Fig. 2.1: Two neurons communicating via a synapse.

Lecture Contents

1. Spiking Neural Networks
2. Neural Coding Schemes
3. Spiking Neuron Models
4. Learning Algorithms
5. Synapses & Inter-neuron communication
6. Conclusions

2. Neural Coding Schemes
Introduction

- There are several type of neural coding schemes
  - Rate coding
    - Spike count rate coding
    - Spike density rate coding
    - Population activity rate coding
  - Temporal Coding
    - Time-to-first spike
    - Inter-spike-interval
    - Phase coding
    - Rank order coding
    - Correlation and synchrony
    - Threshold coding

2. Neural Coding Schemes
Rate Coding

Spike Count Rate

$$v_{sc} = \frac{n_{spike}}{\Delta_t}$$

n_{spike}: number of spikes

Δ_t is the period of repetition

2. Neural Coding Schemes
Rate Coding (cnt.)

Spike Density Rate

$$v_{sd} = \frac{n_K}{\mathrm{K}\Delta_t}$$

the same stimulation sequence is repeated K times,

n_K is summed over all repetitions

2. Neural Coding Schemes
Rate Coding (cnt.)

Population Activity Rate

$$v_{pa} = \frac{n_p}{\mathrm{N}\Delta_t}$$

N: number of neurons

n_p is the total number of spikes generated by N neurons,

2. Neural Coding Schemes
Temporal Coding

- Time-to-first spike
- Inter-spike-interval
- Phase coding
- Rank order coding
- Correletion and synchrony
- Threshold coding

2. Neural Coding Schemes
Time to First Spike

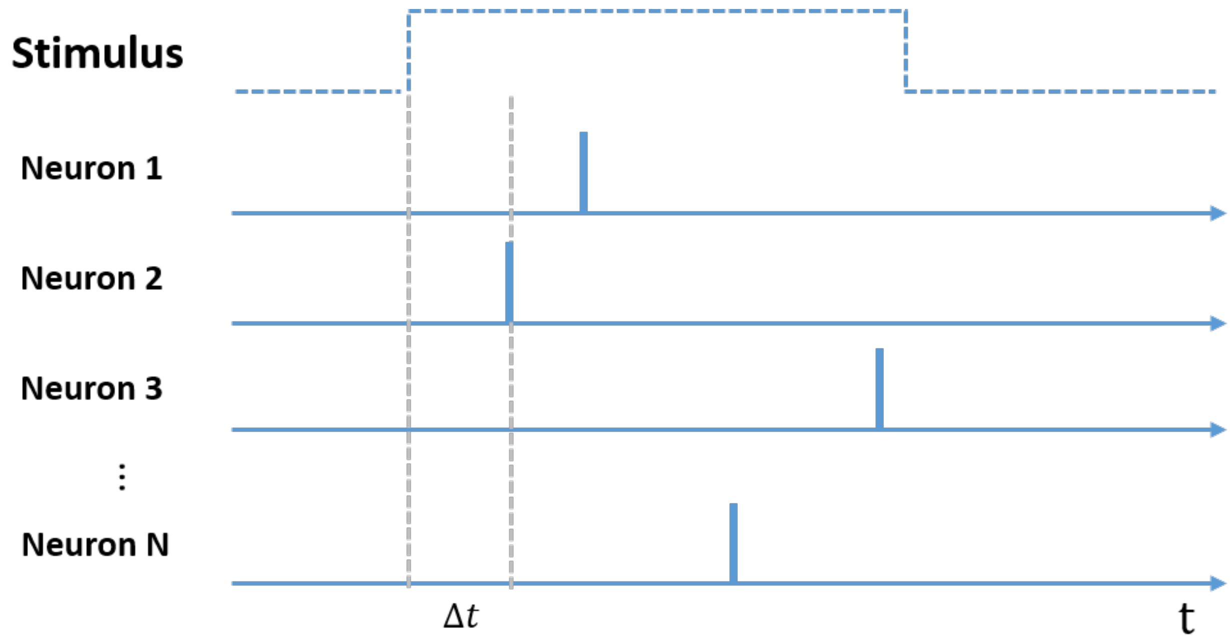


Fig. 2.2: Time to first spike

The idea is to encode the latency information of the first spike event given a stream of spikes, where the precise timing of the spikes indicates the strength of the stimulation

2. Neural Coding Schemes
Phase Coding

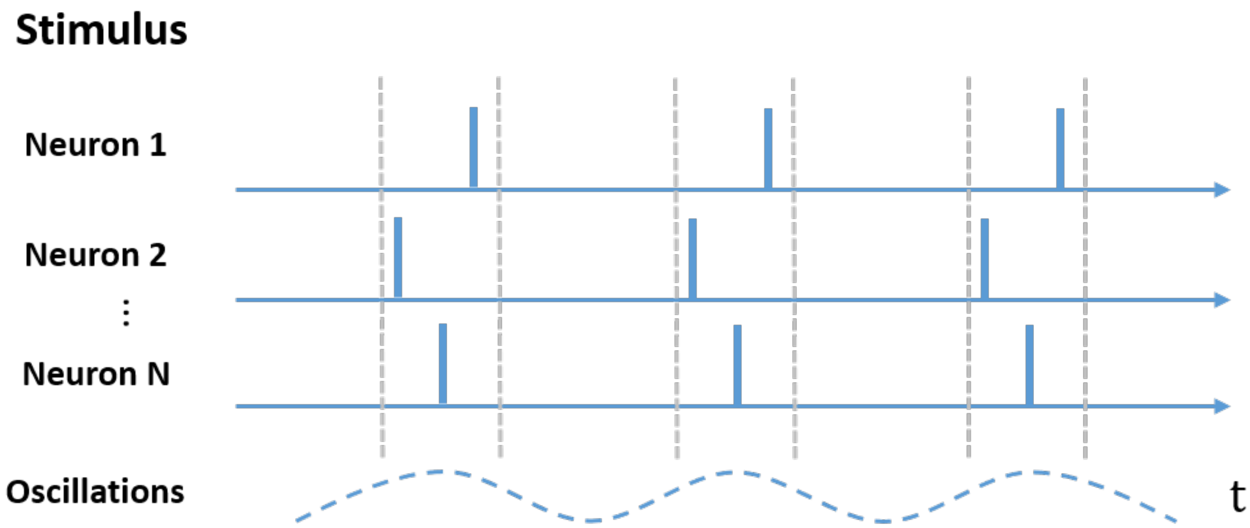


Fig. 2.4: Phase coding

The reference signal is not a single event, but a periodic signal (oscillations)
Coding depends on the “phase” of the signal

2. Neural Coding Schemes
Inter-spike-interval

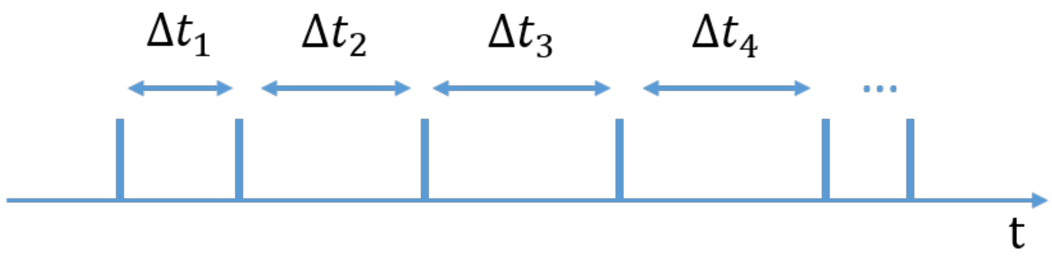


Fig. 2.3: Inter-spike-interval

Coding is based on the distance (“interval”) between consecutive spikes.
The differences between inter-spike-interval is the reference point is the previous spike, not at stimulus

2. Neural Coding Schemes
Rank order

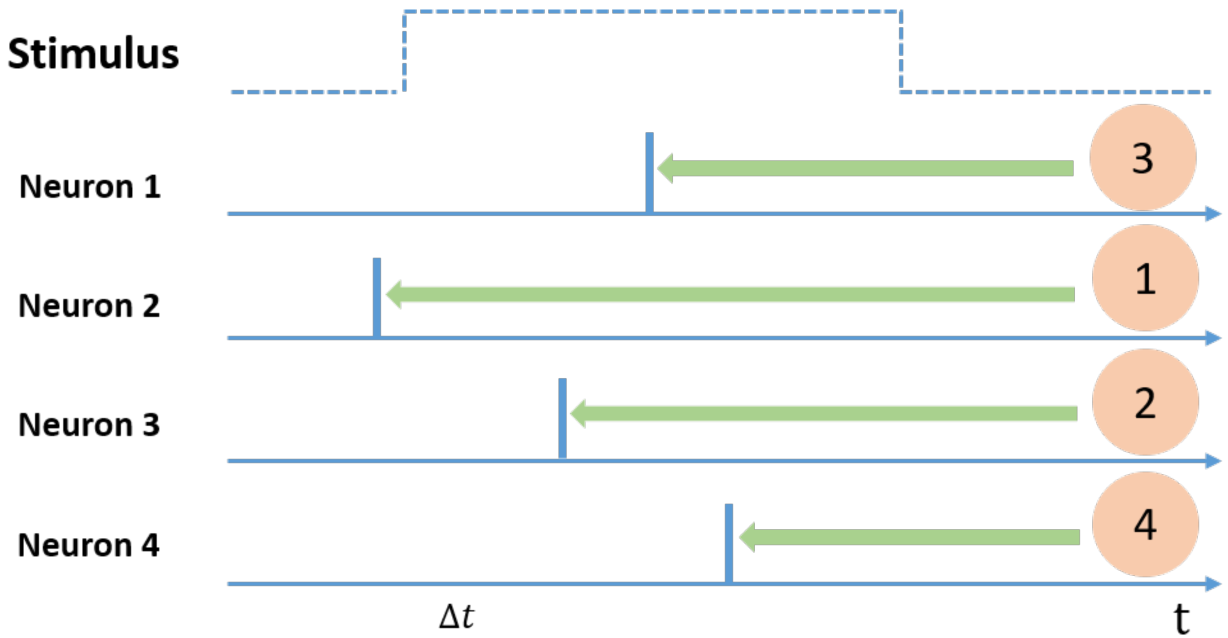


Fig. 2.5: Rank order

This rank order coding gets rid of the timing; here, we only consider the order of the spikes

2. Neural Coding Schemes
Correlation and synchrony

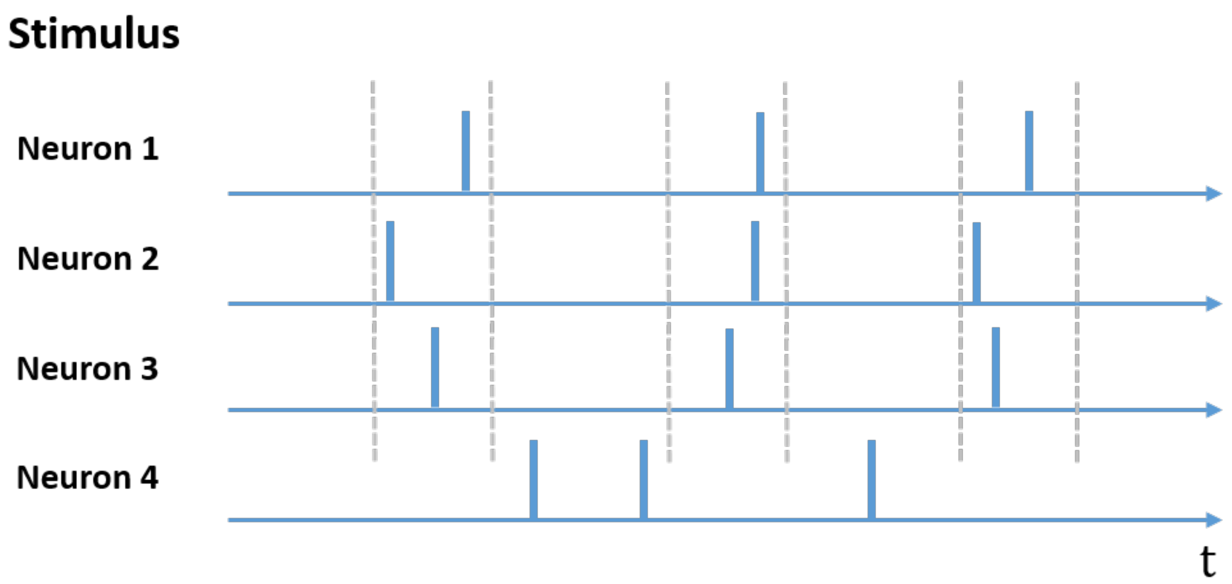


Fig. 2.6: Correlation and synchrony

Synchrony of a pair or of many neurons could signify special events and convey information which is not contained in the firing rate of the neurons

Fig 2.6, top three neurons are synchronized which means they “belong together” (i.e. same group of neuron)

Lecture Contents

1. Spiking Neural Networks
2. Neural Coding Schemes
3. Spiking Neuron Models
4. Learning Algorithms
5. Synapses & Inter-neuron communication
6. Conclusions

3. Spiking Neuron Models
Overview

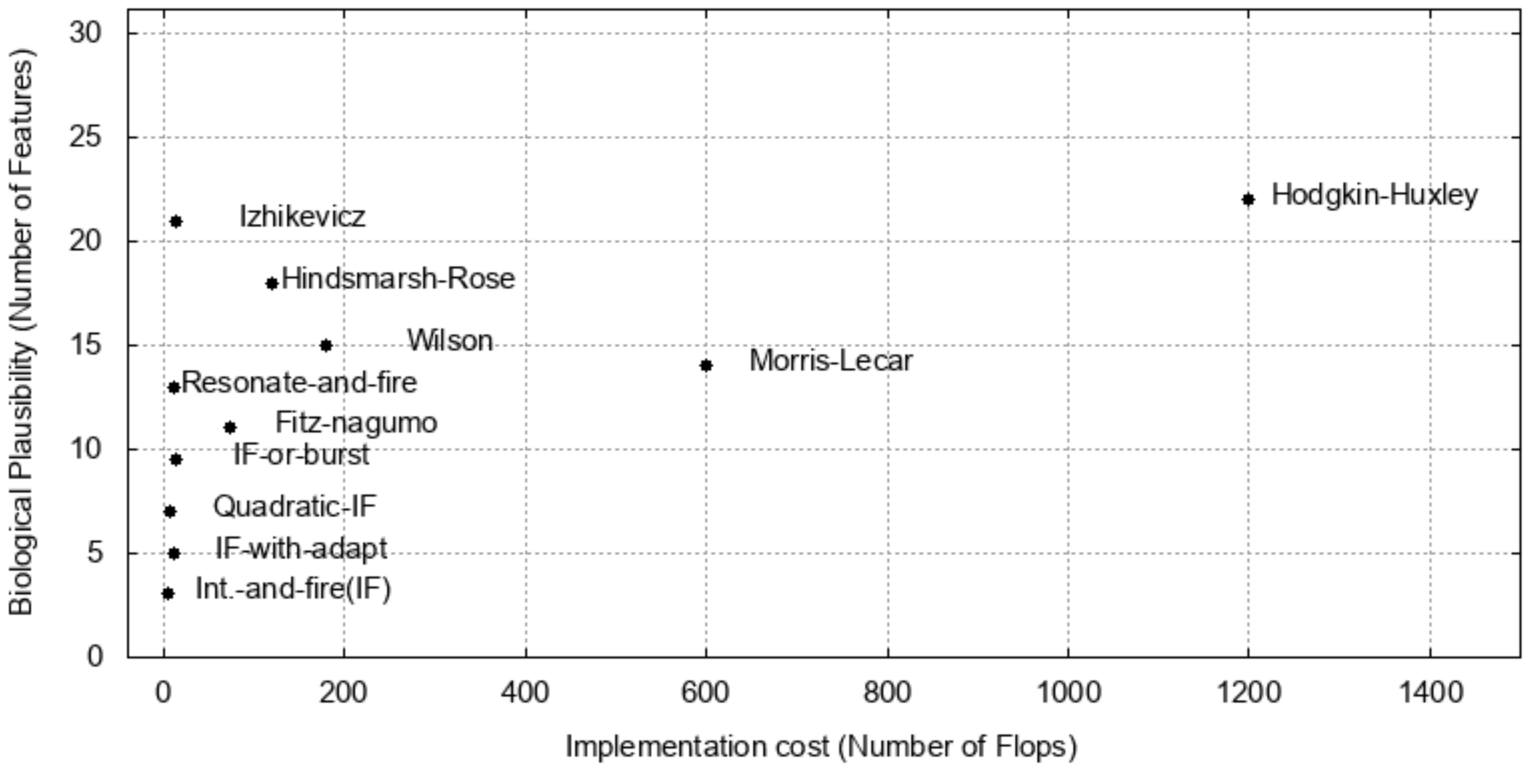


Fig. 2.9: A comparison of spiking neuron models in terms of implementation cost and biological plausibility

3. Spiking Neuron Models
Hodgkin-Huxley

- The Hodgkin-Huxley neural model was proposed in the 1950s.
- Mathematical model:

$$\frac{dv}{dt} = \left(\frac{1}{C}\right) I \; - g_k n^4 (V - V_k) - g_{N_a} m^3 \left(V - V_{N_a}\right) - g_L (V - V_L)$$

- where C is the capacitance of the circuit, I is the external current, conductances are potassium g_k, sodium g_{N_a}, and leakage g_L.

3. Spiking Neuron Models

Hodgkin-Huxley

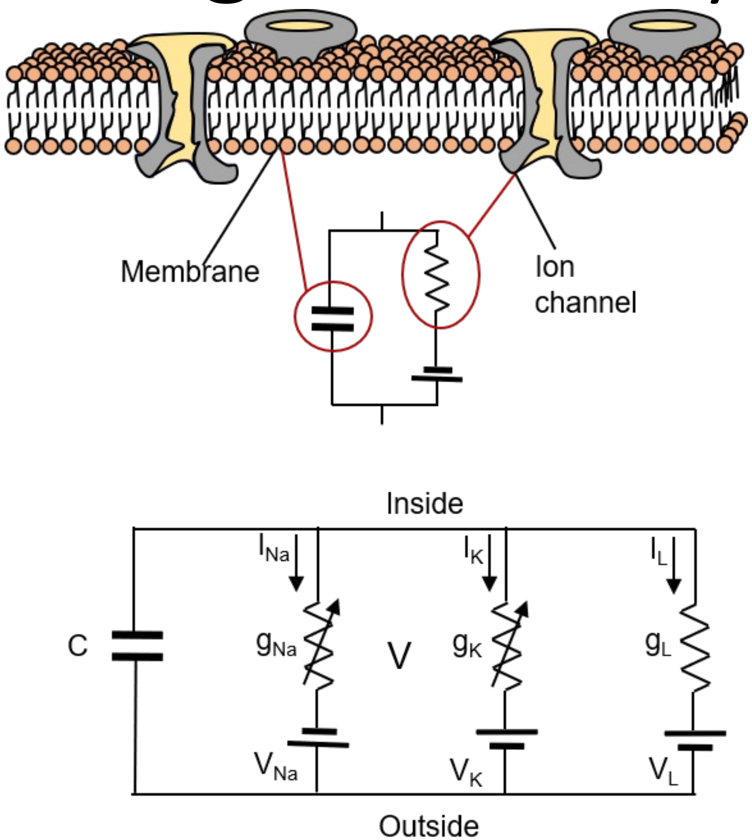


Fig. 2.10: The Hodgkin-Huxley model: (a) the schematic diagram presents the membrane potential, in which current injection starts at t = 5 ms as (b), while (c) and (d) show the dependency of the gating variables n, m and h on the membrane potential.

3. Spiking Neuron Models
Izhikevich Model

- Mathematical model

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I$$

$$\frac{du}{dt} = a(bv - u)$$

$$\begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} if\ v \geq 30mV$$

- where v is membrane potential, u is membrane recovery variable, a, b and c are model parameters.

3. Spiking Neuron Models
Leaky Integrate and Fire

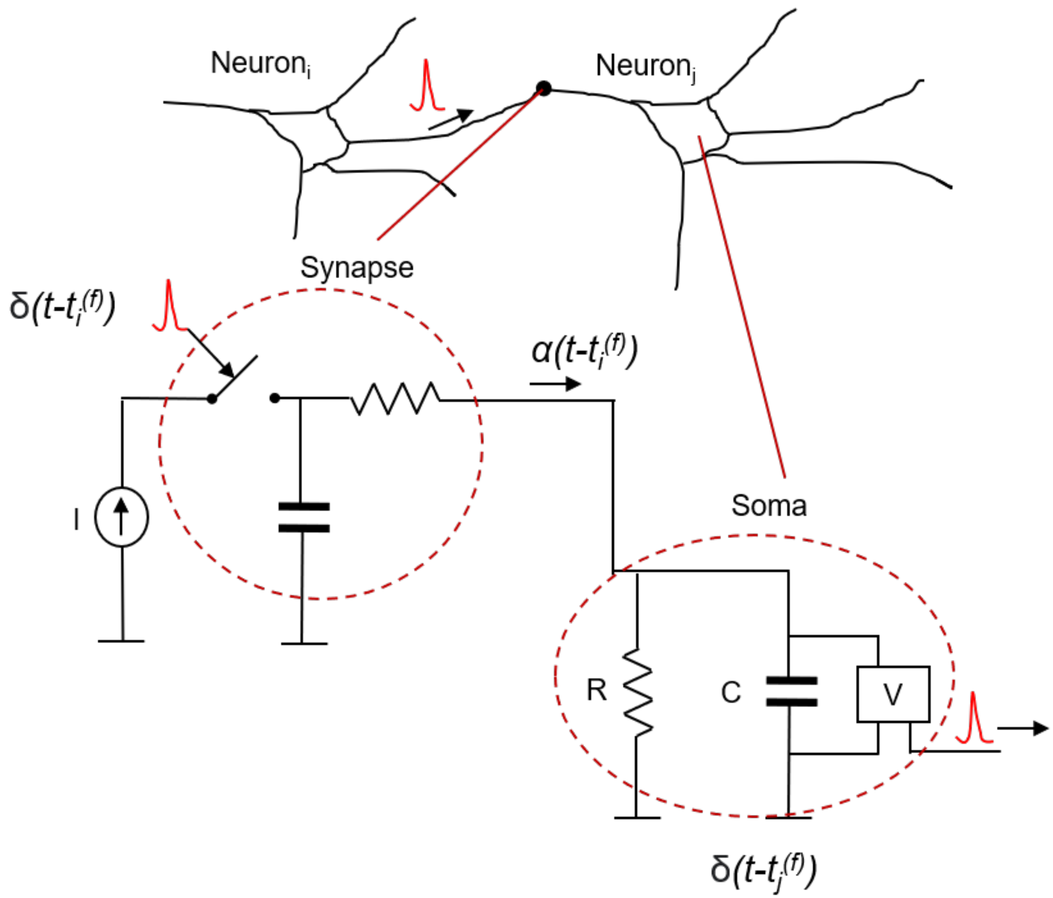


Fig. 2.11: Schematic diagram of the LIF model.

3. Spiking Neuron Models
Leaky Integrate and Fire

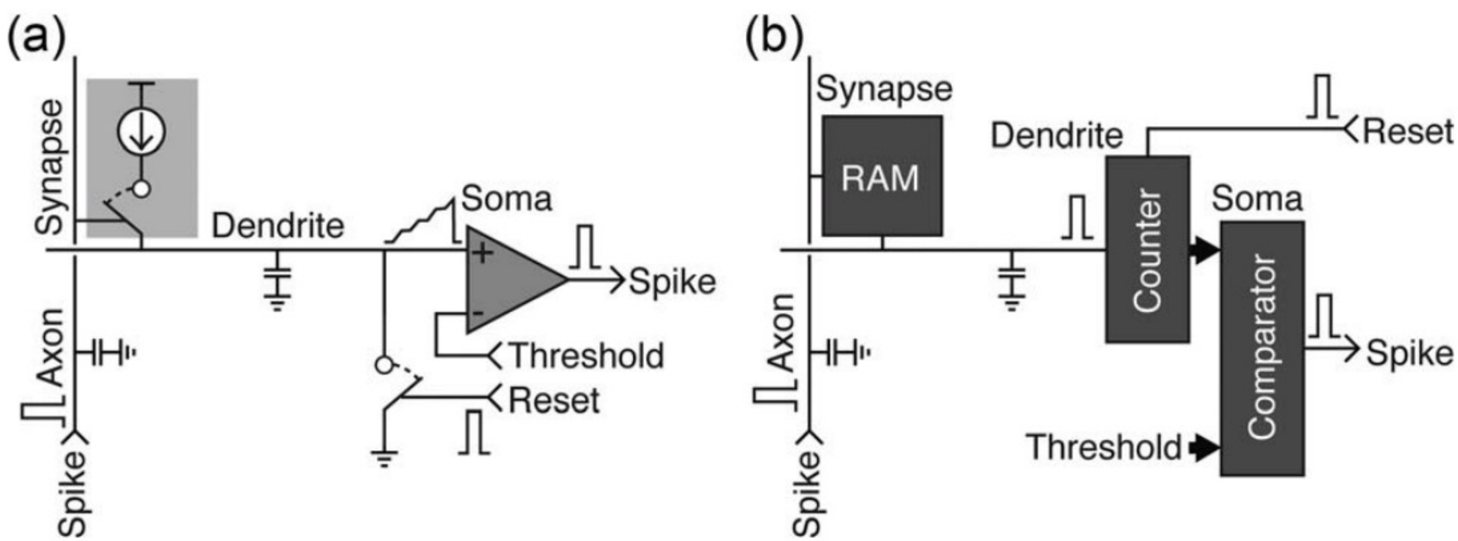
- Mathematical model

$$\frac{dv}{dt} = I + a - bv$$

$$v \leftarrow c, if\ v \geq v_{th}$$

where v is membrane potential, I is neuron current, a ,b and c are the neuron parameters, v_{th} is the threshold voltage.

3. Spiking Neuron Models
Leaky Integrate and Fire Design



Neurogrid: (a) Analog mixed signal neuron; (b) Digital Neuron

Lecture Contents

1. Spiking Neural Networks
2. Neural Coding Schemes
3. Spiking Neuron Models
4. Learning Algorithms
5. Synapses & Inter-neuron communication
6. Conclusions

4. Learning Algorithm
Introduction

- Learning in SNN is the process to modify the strength of the synaptic connections between neurons
- Two major types:
  - Supervised Learning
  - Unsupervised Learning

4. Learning Algorithm
Supervised Learning

- The most notable method is “backpropagation”.
- Usually an off-chip method:
  - Train on a host PC
  - Synaptic weights are download to the neuromorphic chip later
- Two approaches:
  - Training directly the spiking model
  - Training the similar ANN model and convert it to SNN

4. Learning Algorithm
Unsupervised Learning

- Base on self-organizing map and learning rules.
- Notable methods:
  - The spike-timing-dependent plasticity (STDP), which is a Hebbian-based learning
  - The Spike Driven Synaptic Plasticity (SDSP) which adapts synaptic weights at the arrival of presynaptic spikes

Lecture Contents

1. Spiking Neural Networks
2. Neural Coding Schemes
3. Spiking Neuron Models
4. Learning Algorithms
5. Synapses & Inter-neuron communication
6. Conclusions

5. Synapses & Inter-neuron communication
Introduction

- In hardware neuromorphic system, memories usually embedded into the system to store the synaptic weight.
- Primary operation of SNN are storing and reading synaptic weight.
- Several technologies are avaible for SNNs:
  - SRAM, e-DRAM, non-volatile (i.e., memristor)

5. Synapses & Inter-neuron communication
SRAM

- A typical SRAM contains six transistors (6-T) offering read and write at the same time.
- Compatible with CMOS technology
- Transpose access can be obtained using 8-T design

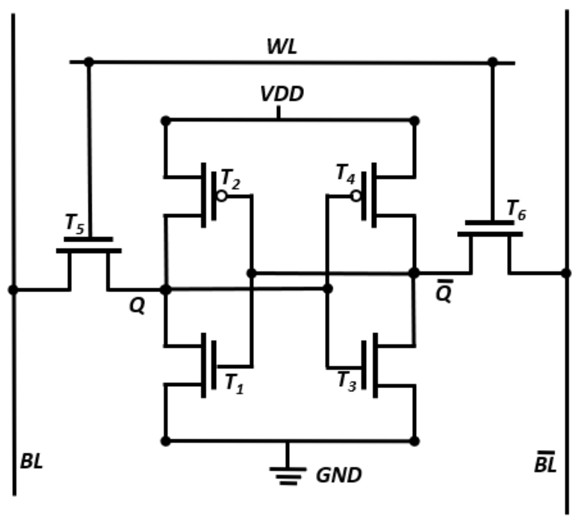


Fig. 2.12: SRAM.

5. Synapses & Inter-neuron communication
e-DRAM

- e-DRAM (embedded dynamic random access memory) is a technology to embed DRAM cells into CMOS technology.
- Since DRAM cell (1T1C) is much smaller the SRAM cell (6T), it offers higher density

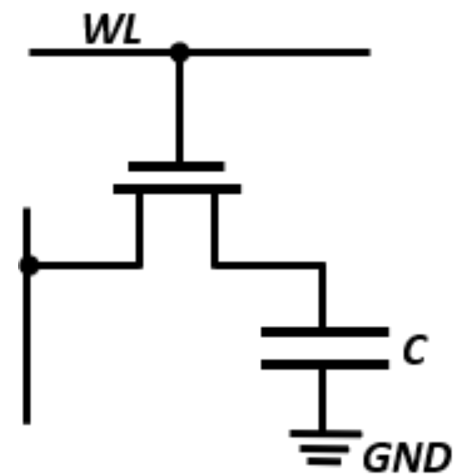


Fig. 2.13: EDRAM.

5. Synapses & Inter-neuron communication
Memristor

- Memristor offer high density as 1T1M (similar to e-DRAM)
- Unlike e-DRAM which loss it value once turning off the power, memristor retains it value
- Memristor usually offer two type of resistive state:
  - High resistance
  - Low resistance

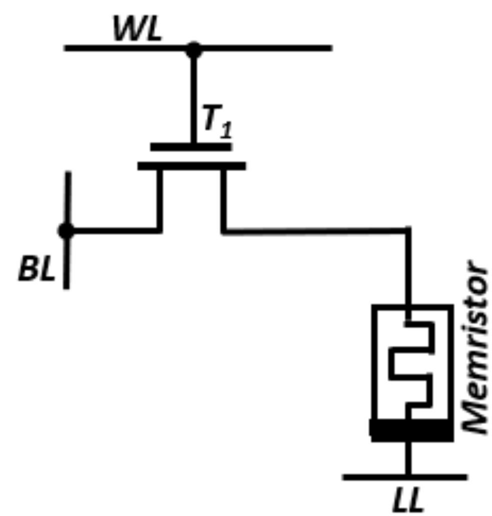


Fig. 2.14: MEMRISTOR.

5. Synapses & Inter-neuron communication
AER-protocol

- Communication architectures for multicore spiking neuromorphic systems are responsible for delivering spikes between neuro-cores/tiles
- Since spikes are usually sparse (majority is non-spike), sending signals to indicate non-spiking neurons are not efficient.
- Instead of that, Address Event Representation is used.
    - Once a neuron fires, the system encapsulate its spike as its address.
    - If there is no spike, no information is sent.

5. Synapses & Inter-neuron communication

AER-protocol

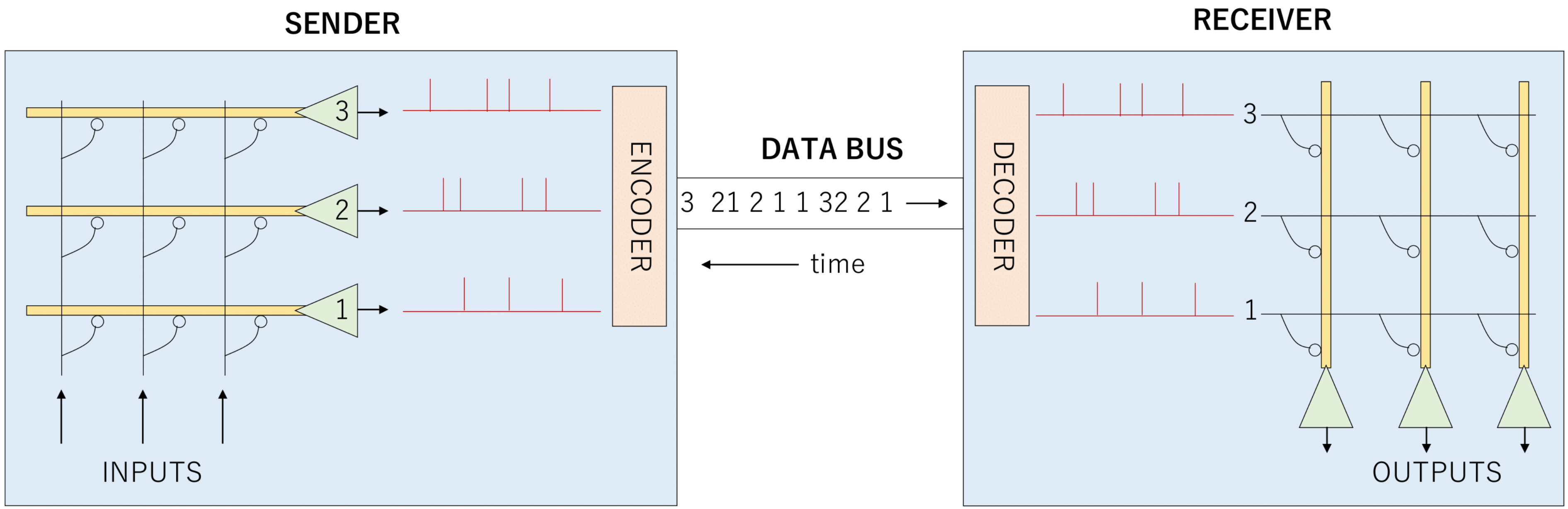


Fig. 2.15: AER protocol

5. Synapses & Inter-neuron communication
AER-protocol

- AER has low encoding complexity and exploit the sparsity of spike trains (spatially and temporally).
  - Instead of sending the spike, AER signal is shorter.
    - Assuming the system with 1024 neurons, spike vector need 1024 bits while AER requires 10 bits.
    - If less than 102 neurons fired, the AER uses less bit in the single time-step
  - AER is more scalable:
    - With 1 million neurons, transmitting 1 Mbit vs 20 bit per spikes is scalable

5. Synapses & Inter-neuron communication
Communication Fabrics

- Conventionally, CPU/GPU based system use bus (multi-level) as the communication fabrics
- For accelerators/standalone neuromorphic system, we classify as two type:
  - On-chip
  - Off-chip
- Communication could be:
  - Unicast-based
  - Multicast based

Lecture Contents

1. Spiking Neural Networks
2. Neural Coding Schemes
3. Spiking Neuron Models
4. Learning Algorithms
5. Synapses & Inter-neuron communication
6. Conclusions

6. Conclusion

- In this lecture, we reviewed the neuron structure and the notable models such as Hodgkin-Huxley, Izhikevich and LIF.
- We also discussed about how neurons inform each other via spikes:
  - Rate Coding
  - Temporal Coding
- The training method for neuromorphic systems is also reviewed.
- Synaptic weight realization and inter-neuron communication in neuromorphic system is summarized.