Liquid State Machine with Dendritically Enhanced Readout for Low-power, Neuromorphic implementations

Subhrajit Roy
Supervisor : Dr. Arindam Basu

OUTLINE

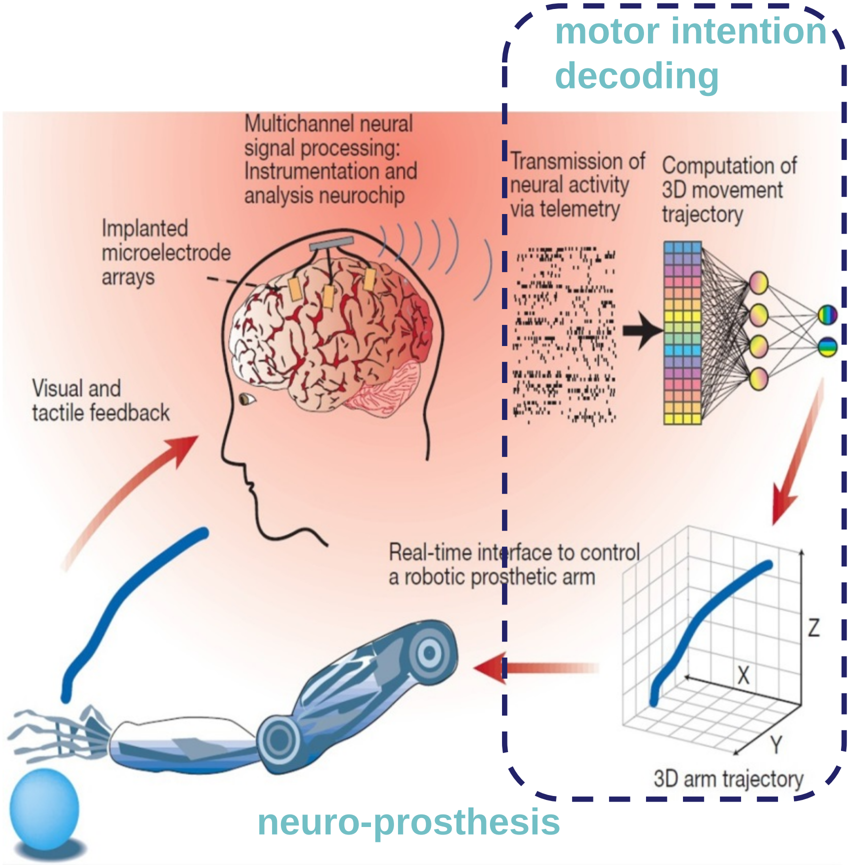
- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion

OUTLINE

- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

Motivation for developing low-power, machine-learners



PROBLEMS

- Data rate / channel ~ 200 Kbps
- 1000 channels → 200 Mbps
- Huge power dissipation

UNSUSTAINABLE

SOLUTIONS

- On chip neural processing unit
- Requirement for low-power hardware implementations of supervised classifiers

NANYANG TECHNOLOGICAL UNIVERSITY

OUTLINE

- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

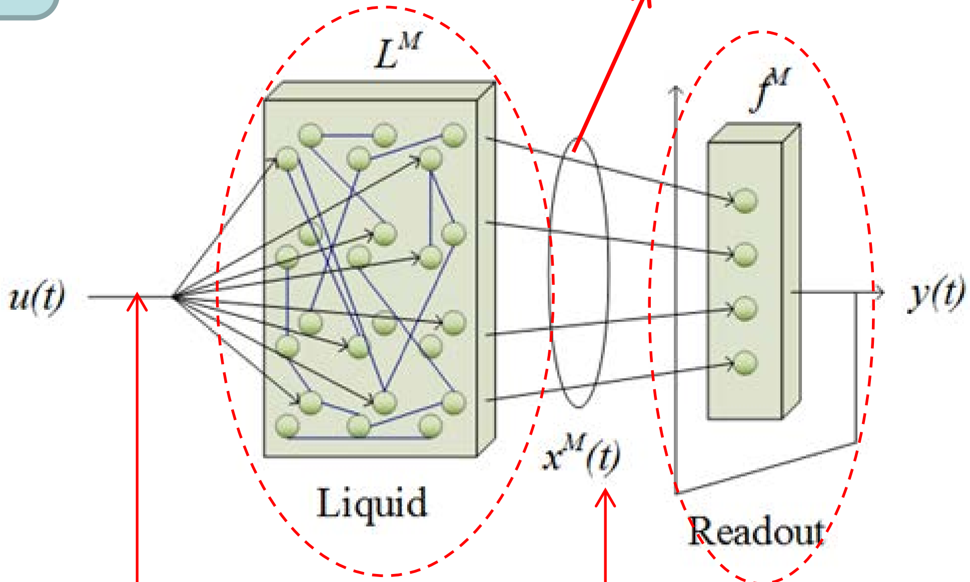
Liquid State Machine

Liquid

- Huge recurrent interconnection of spiking neurons
- +
- Synapses with random weights

No learning

Learning of only these synapses



Inputs: Spike Trains or Analog current

Task-specific training of readout

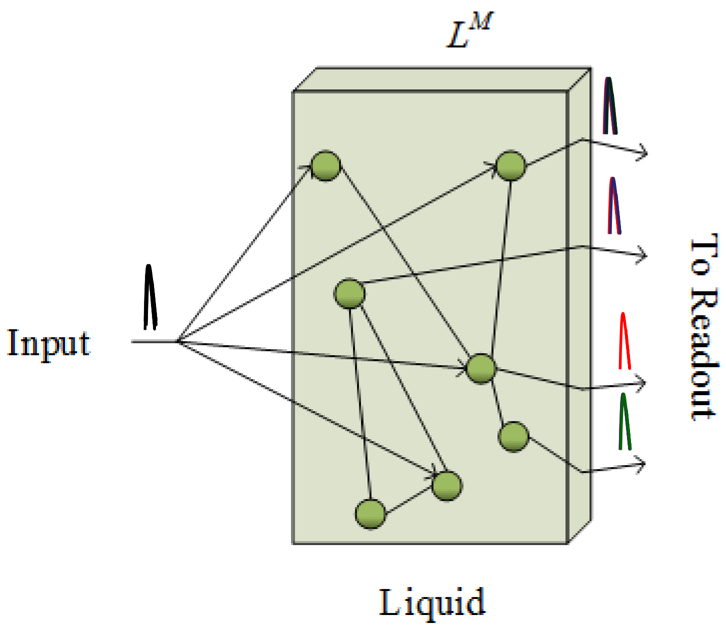
Readout can be:
- Single Perceptron
- **Parallel layer of perceptrons**
- Multilayer perceptrons

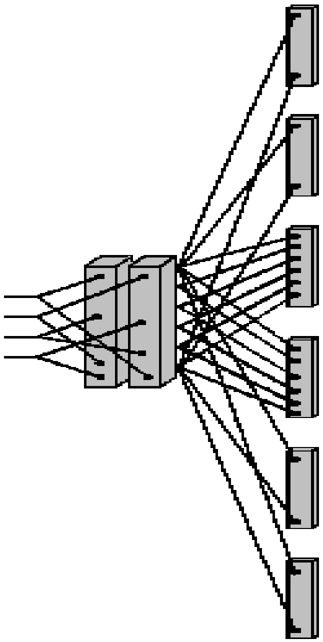
NANYANG TECHNOLOGICAL UNIVERSITY

W. Maass, T. Natschlager, and H. Markram, “Real-time computing without stable states: a new framework for neural computation based on perturbations,” *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, Nov. 2002.

Properties of LSM

- **Why the term Liquid?** Short term memory effect leading to **Temporal Integration**.





Parallel Processing :
•Multiple readouts with the same liquid
•Each trained to perform different tasks on same inputs

Liquid Advantages:
1. Encodes Spike Times
2. Low dimension to high dimension → Increases separability
3. Recurrence → Memory effect
4. General → Multiple features extracted

NANYANG TECHNOLOGICAL UNIVERSITY

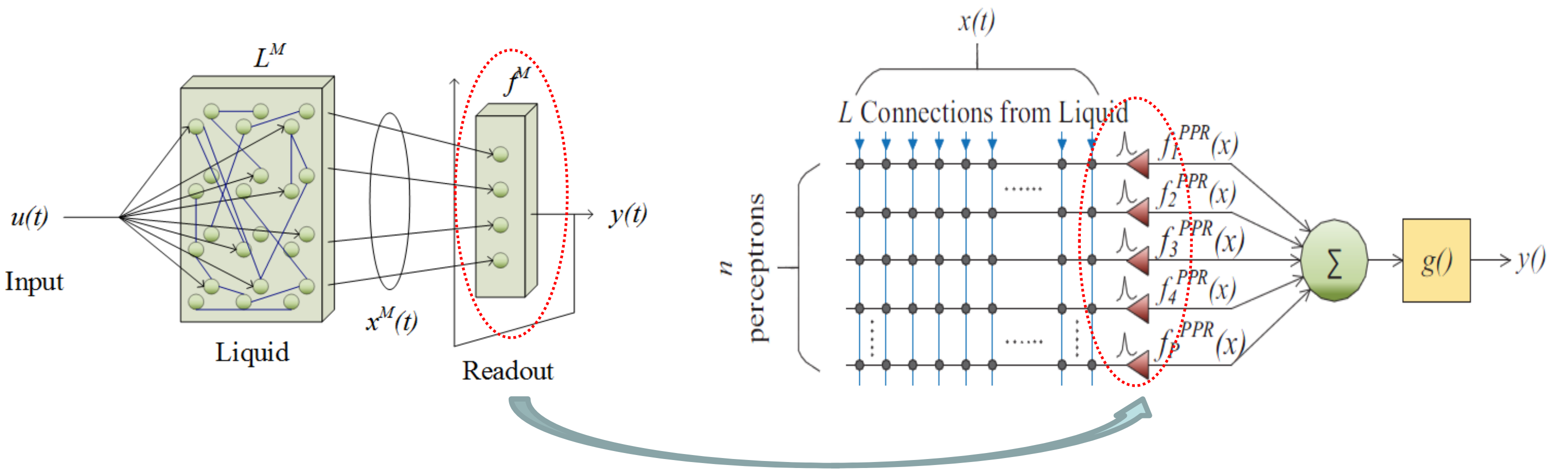
OUTLINE

- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

LSM w Parallel Perceptron Readout (LSM-PPR)

- **MLP** : Works in both classification and approximation → H/W constraints
- **Parallel Perceptron**: A single layer composed of finite number of perceptrons each receiving the same input is called a parallel perceptron.



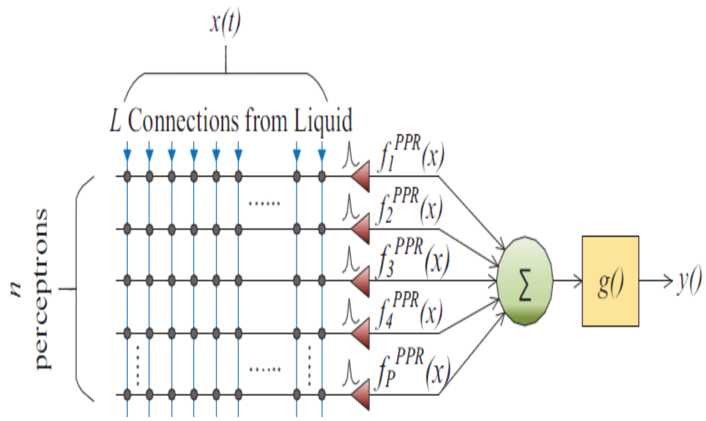
- An efficient algorithm (***p*-delta learning rule**) specifically designed for training of parallel perceptrons is used.

State-of-the-art for single layer readout

NANYANG TECHNOLOGICAL UNIVERSITY

W. Maass, T. Natschlager, and H. Markram, “Real-time computing without stable states: a new framework for neural computation based on perturbations,” *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, Nov. 2002.

LSM-PPR limitation



1. If L liquid neurons and n readout neurons then total tunable synapses : $\boldsymbol{Lxn}$

⟹ **Difficult for VLSI implementation**

2. Tunable synapses require high-resolution and non-volatile weights

⟹ **Difficult for VLSI implementation**

USE LOWER RESOURCES FOR GIVING SAME PERFORMANCE

+

USE BINARY SYNAPSES INSTEAD OF HIGH RES SYNAPSES

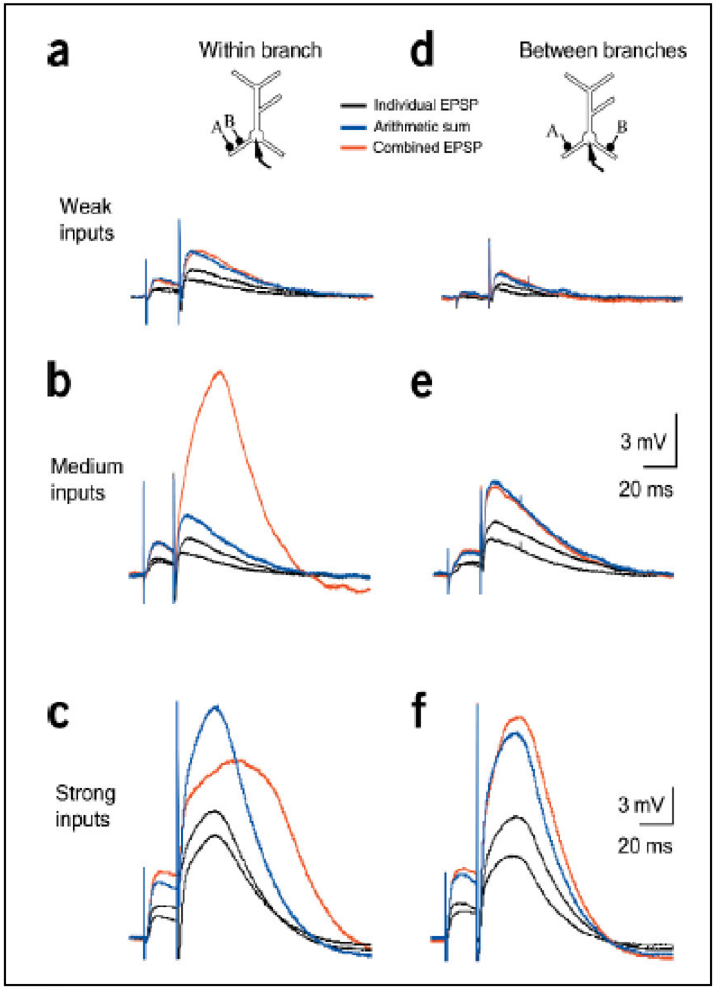
NANYANG TECHNOLOGICAL UNIVERSITY

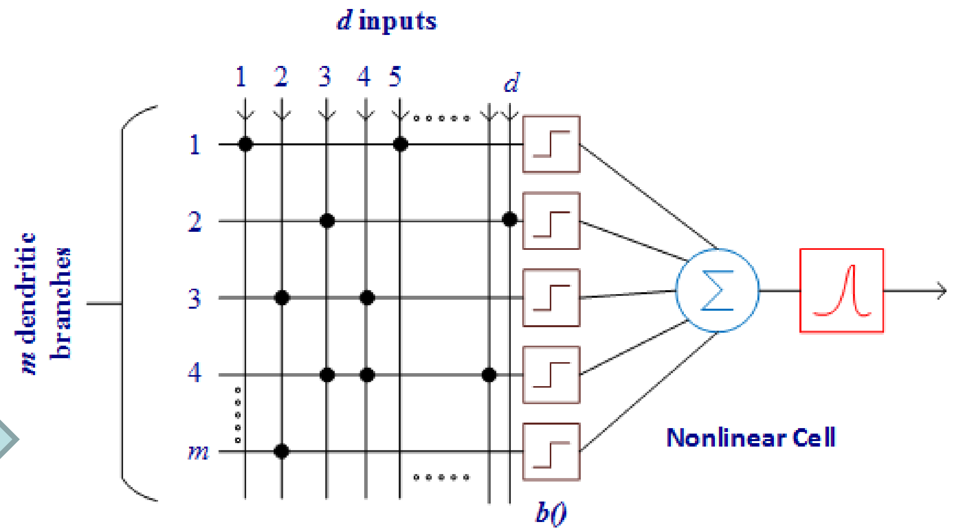
OUTLINE

- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion

Neurons with Active Dendrites

Motivation





Source: Poirazi et al., 2001

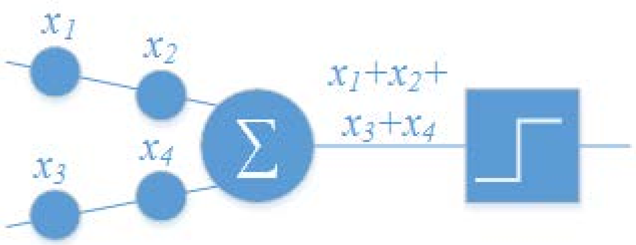
- Currently considering only lumped dendritic nonlinearity
- Neurons with active dendrites have higher storage capacity

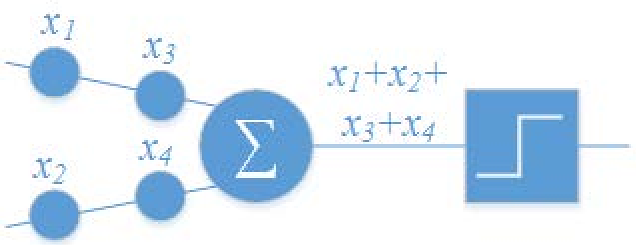
NANYANG TECHNOLOGICAL UNIVERSITY

Source: Polsky et al., 2004

Neurons with Active Dendrites

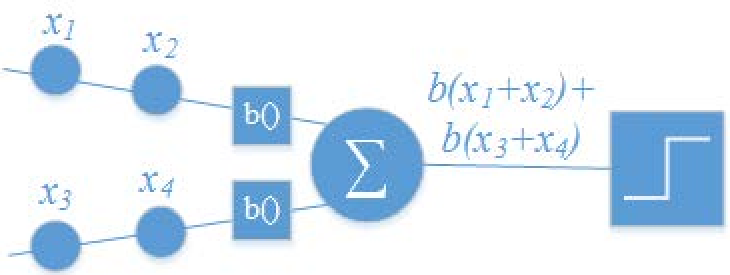
Linear Cell w Binary Synapses

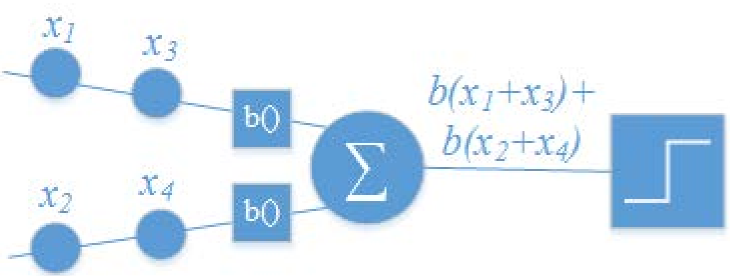




Unable to recognize the different combination of inputs

Non linear cell w Binary Synapses



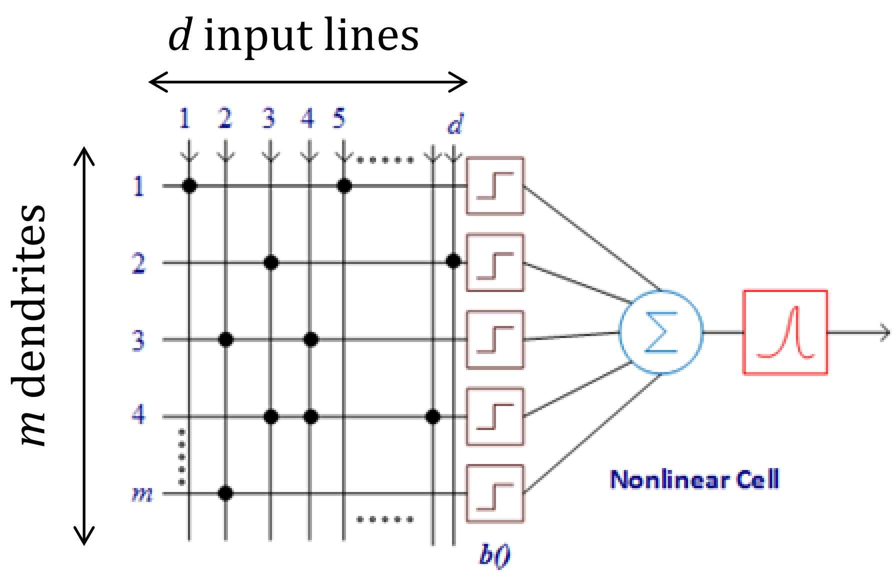


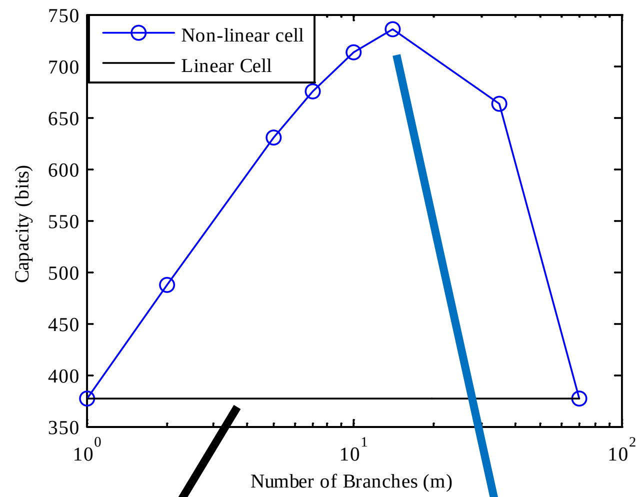
Capable of recognizing the different combination of inputs

NANYANG TECHNOLOGICAL UNIVERSITY

Neurons with Active Dendrites

s: Total number of synapses
m: No. of dendritic branches
k: No. of synapses per branch
d: Dimension of input





$$B_L = log_2 \binom{s+d-1}{s} \qquad B_N = log2 \binom{\binom{k+d-1}{k}+m-1}{m}$$

$k << d$

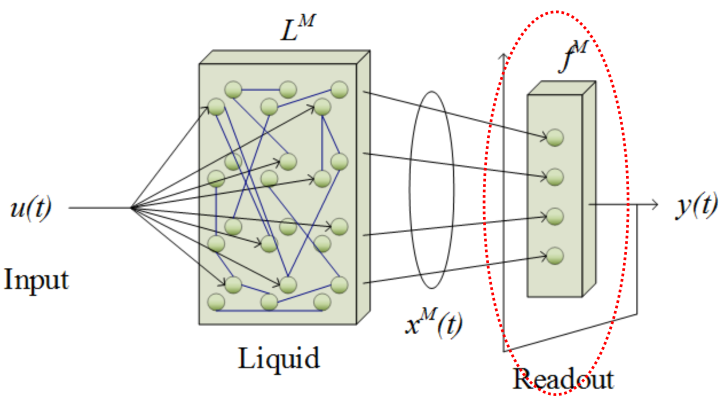
- Synaptic weight = 0/1
- Not weight update, but Connection change

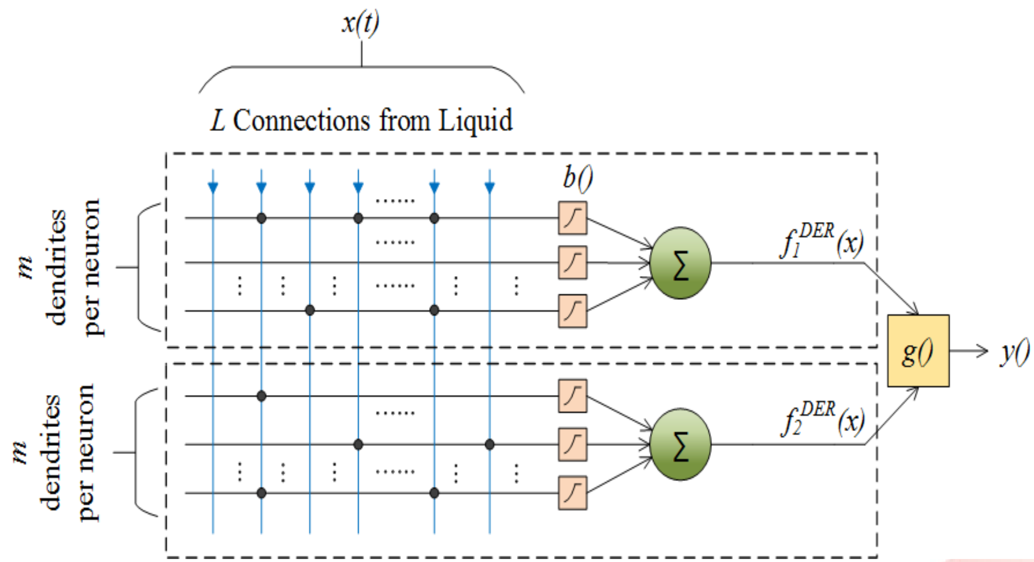
NANYANG TECHNOLOGICAL UNIVERSITY

LSM w Dendritically Enhanced Readout (LSM-DER)

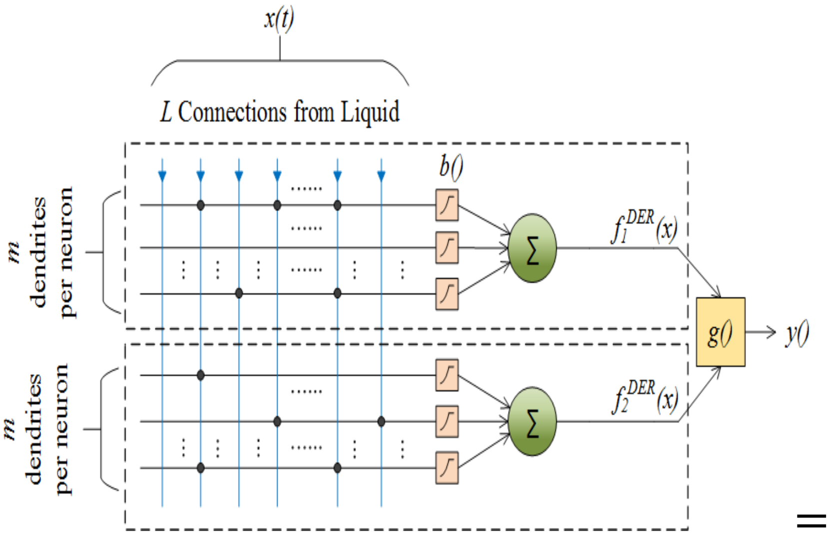
- The **proposed algorithm LSM-DER** constitutes the liquid of LSM followed by a two neuronal cell architecture
- The parallel perceptron stage of LSM-PPR has been replaced by a two neuronal cell architecture.

LSM-PPR : Lxn synapses
LSM-DER : L synapses





Network Re-Wiring(NRW) Learning Rule



- t = Teacher signal,y = LSM-DER output
- Applying Gradient Descent algorithm:

$$\Delta w_{ij} = -\frac{\partial e^2}{\partial w_{ij}} = 2 < (t-y)\frac{\partial y}{\partial w_{ij}} >$$

$$= 2 < (t-y)\frac{\partial g\left(f_1^{DER}(x) - f_2^{DER}(x)\right)}{\partial w_{ij}} >$$

$$= 2 < (t-y)\frac{\partial g\left(\left(\sum_{j=1}^{m} b\left(\sum_{i=1}^{k} w_{ij}\right)\right)_1 - \left(\sum_{j=1}^{m} b\left(\sum_{i=1}^{k} w_{ij}\right)\right)_2\right)}{\partial w_{ij}}$$

For positive cell: $\Delta w_{ij} = 2 < (t-y)g' b_j' x_{ij} >$

For negative cell: $\Delta w_{ij} = -2 < (t-y)g' b_j' x_{ij} >$

Binary synapses in our case so Δw_{ij} can be considered as a fitness parameter / correlation calculator c_{ij}

For positive cell: $c_{ij} =< (t-y)x_j x_{ij} >$

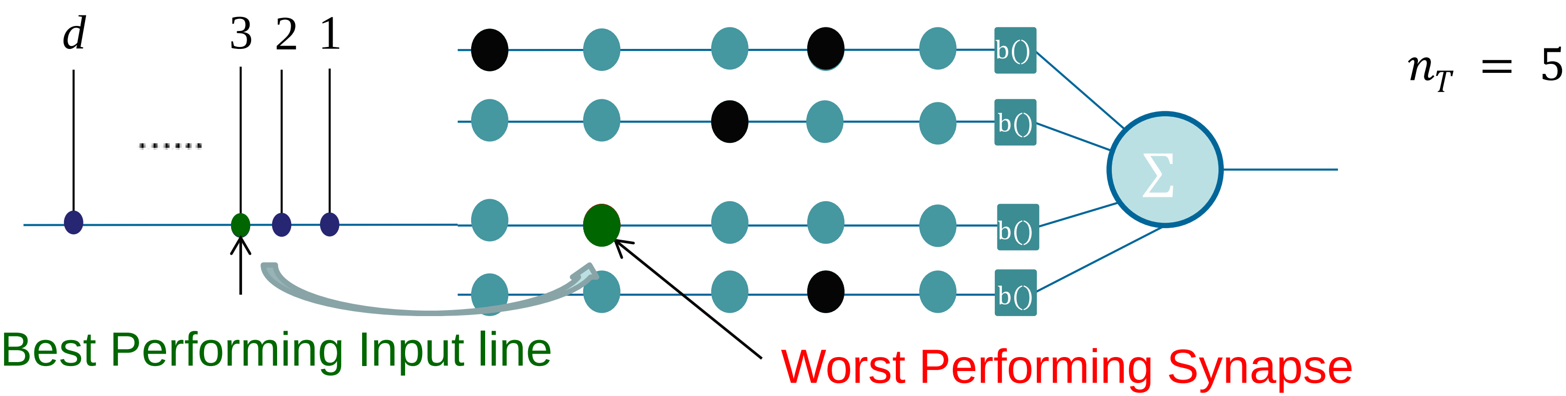
For negative cell: $c_{ij} =-< (t-y)x_j x_{ij} >$

- g' dropped for ease in h/w implementation
- $b()$ is a saturating squared non linearity

NANYANG TECHNOLOGICAL UNIVERSITY

Widrow et al., " 30 Years of Adaptive Neural…..," Proceedings of the IEEE, 1990.

Network Re-Wiring Learning Rule



We search for the worst performing synapse i.e. lowest c_{ij} synapse in the set n_T

Replacement Strategies

- Remove the connected input line to the synapse by any random input line → No extra calculations but slow learning
- Place all the input lines in the dendrite of worst performing synapse → Replace with the best → Fast but exhaustive & requires lot of computations
- We take a middle path → Choose a random set n_R of the input lines → Replace by its best

NANYANG TECHNOLOGICAL UNIVERSITY

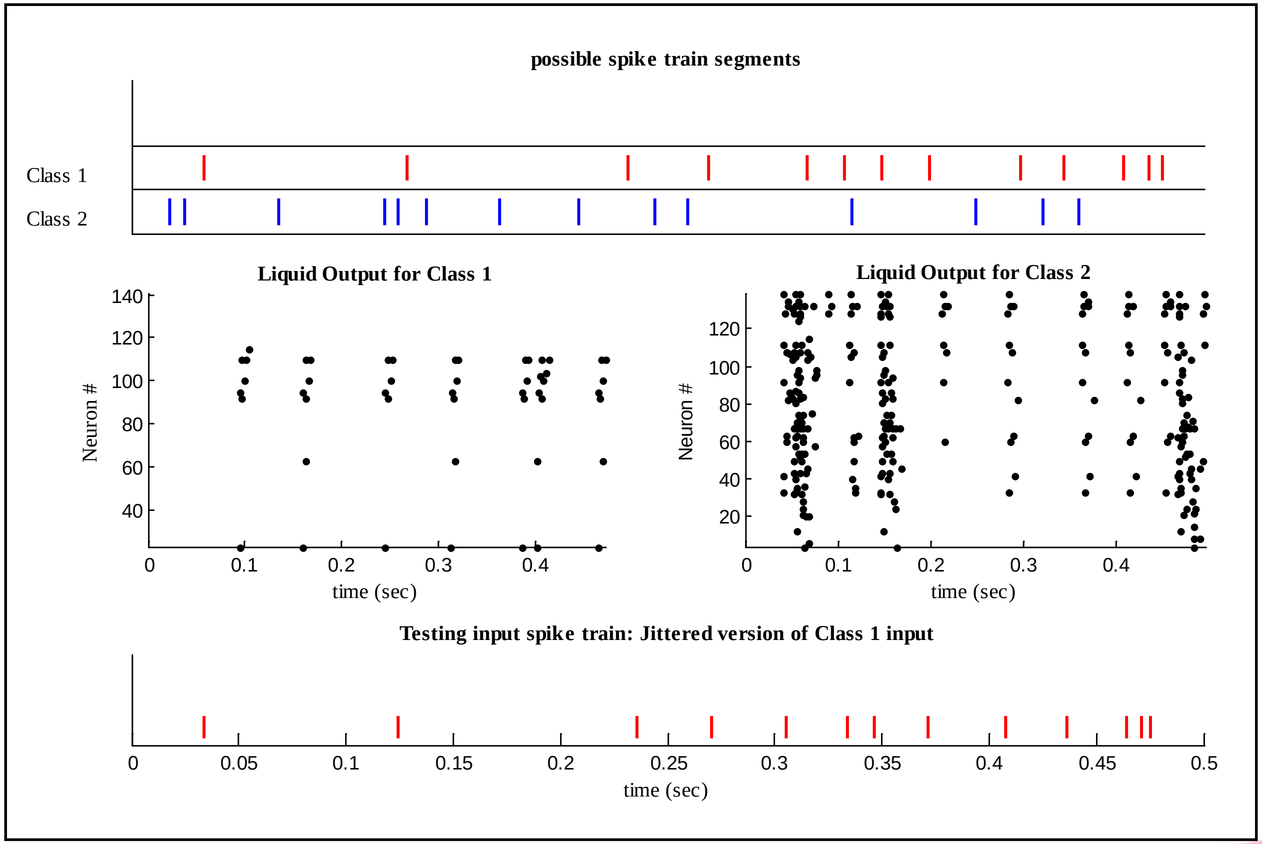
OUTLINE

- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion, Publications and Future Work

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments and Results

- **Task I** : Spike Train Classification Problem

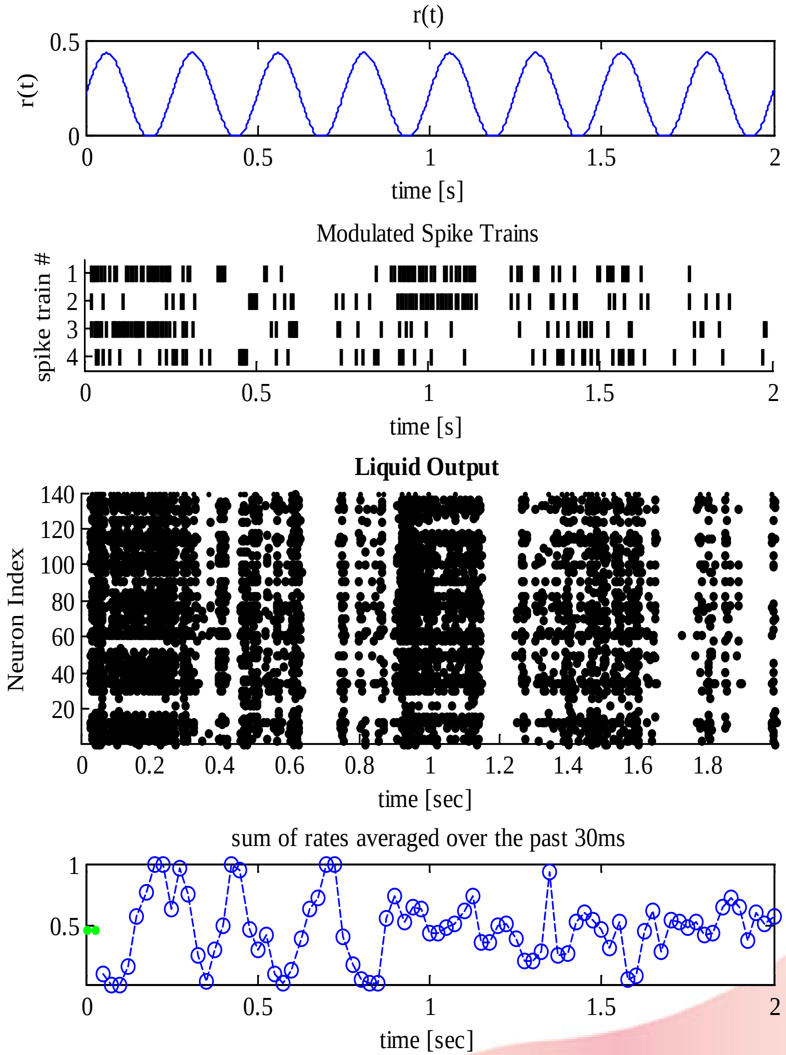


Experiments and Results

- **Task II**:

  Retrieval of sum of rates:

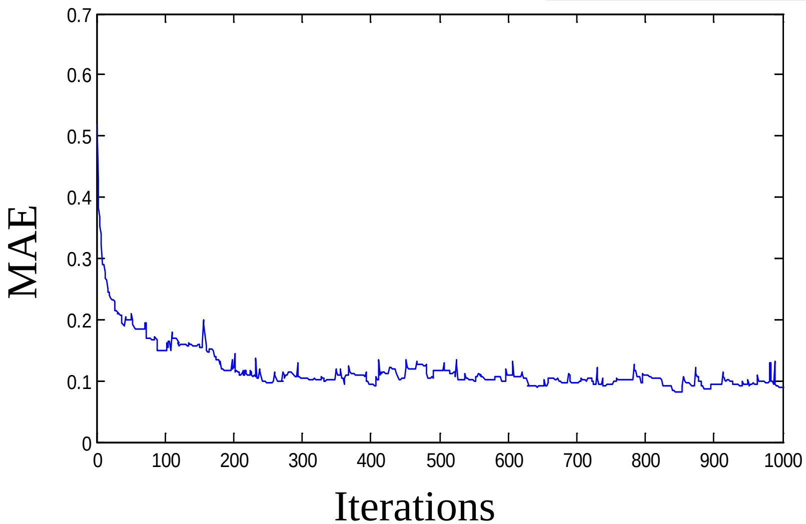
  - 4 Poisson spike trains with randomly modulated firing rates are injected into the liquid.
  - At any point of time *t*, the job of the network is then to give as output the normalized sum of input rates averaged over the last 30 ms.



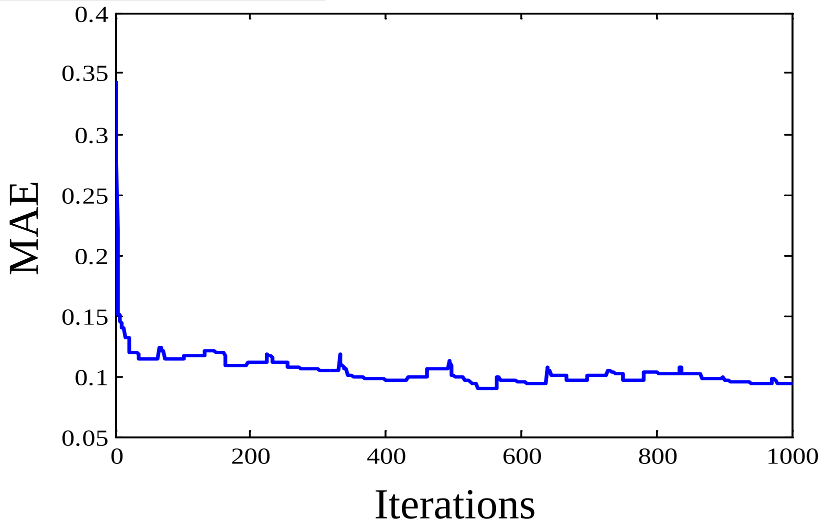
NANYANG TECHNOLOGICAL UNIVERSITY

Results: Performance of LSM-DER and NRW algorithm

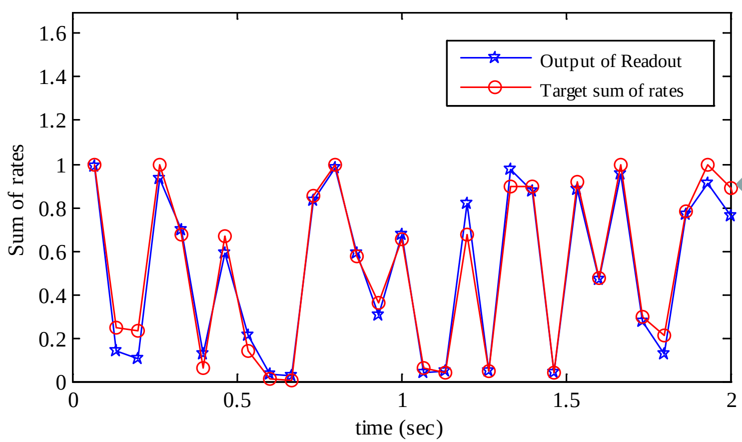
Training Error vs Iterations



Task I : Classification



Task II : Approximation



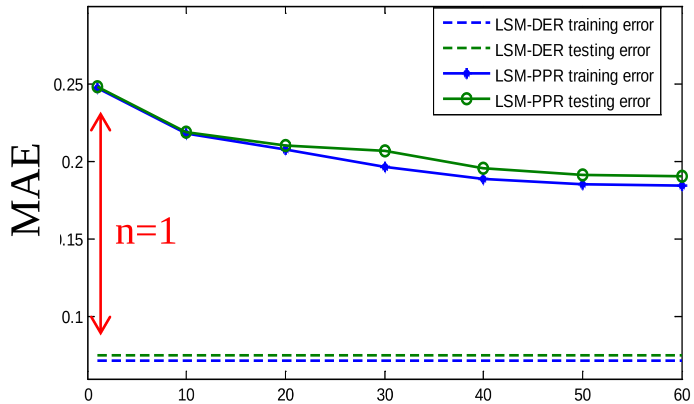
Successful Approximation of target function by LSM-DER

NANYANG TECHNOLOGICAL UNIVERSITY

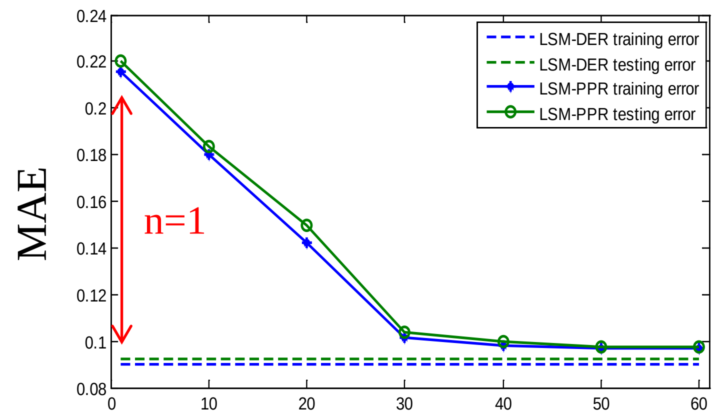
Results: Comparison between LSM-DER and LSM-PPR

- Performance comparison of LSM-DER and LSM-PPR with varying n
- LSM-PPR with $n = 1$ (i.e. single perceptron readout) has same number of tunable synapses as LSM-DER
- For $n = 1$, LSM-DER gives 3.3 and 2.4 times less error for Task I and II respectively.

With the requirement of 1 perceptron we are getting better performance than n perceptrons



Task I : Classification

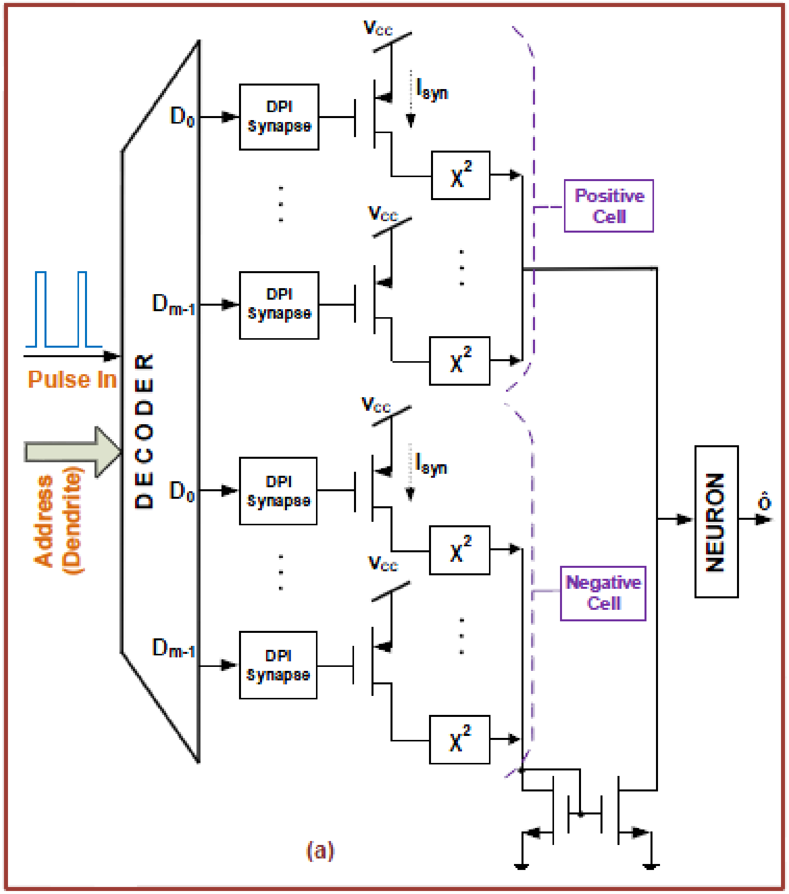


Task II : Approximation

NANYANG TECHNOLOGICAL UNIVERSITY

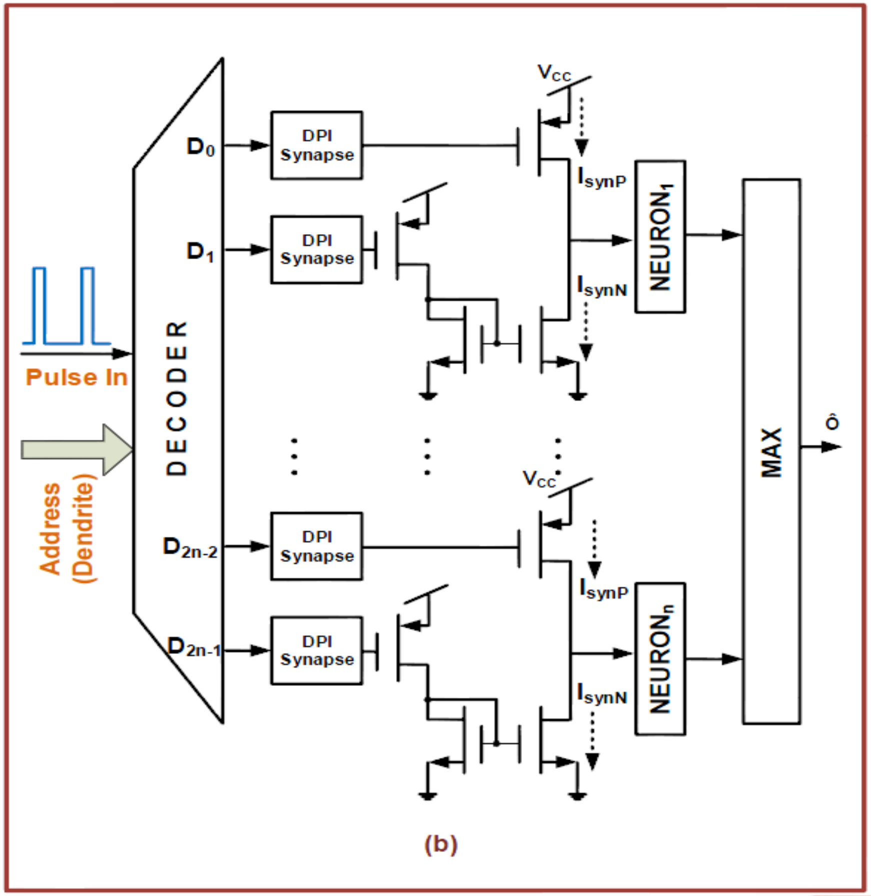
LSM-PPR can never achieve the performance of LSM-DER

Stability with respect to non-idealities



Hardware implementation of DER

[Amitava et al. 2015]



Hardware implementation of PPR

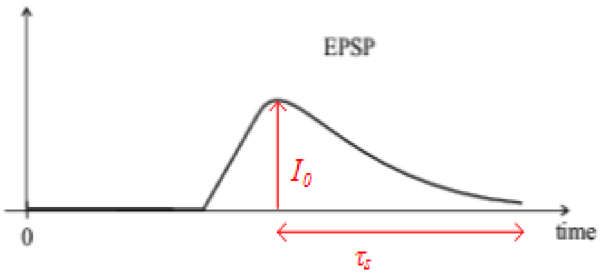
[Amitava et al. 2015]

NANYANG TECHNOLOGICAL UNIVERSITY

Stability with respect to non-idealities

- Monte Carlo simulations of DPI synapse and Square Block Circuit

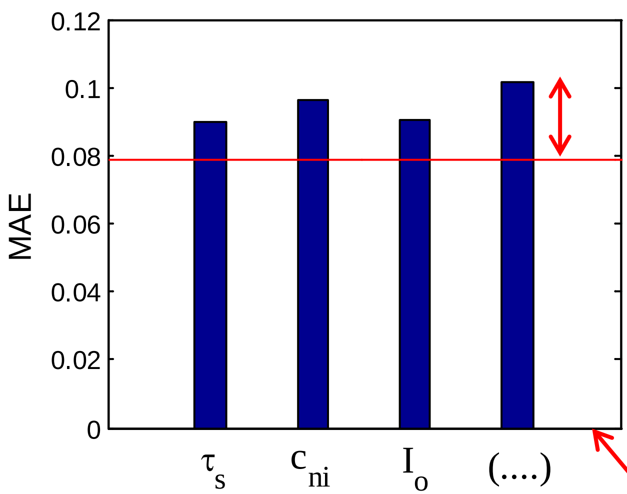
Variation in DPI Synapse



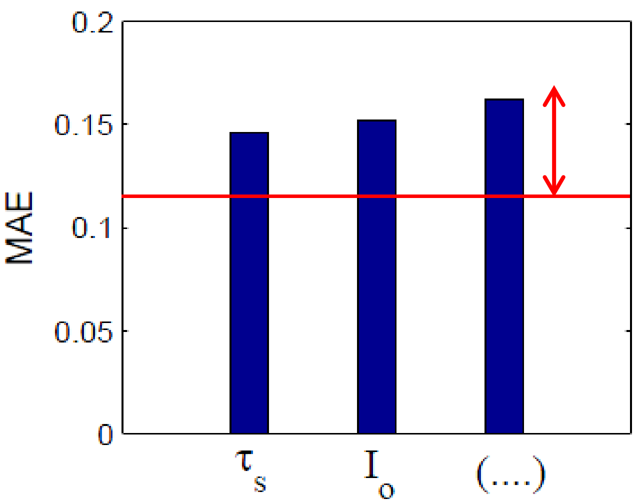
Maximum Variation

1. I_0 ~ 13%
2. τ_s ~ 10.1 %
3. c_{ni} ~ 18%

DER : 2.33 % increase



PPR : 4.73 % increase



Increased stability due to binary synapses

NANYANG TECHNOLOGICAL UNIVERSITY

OUTLINE

- Motivation for low-power machine learners and why Liquid State Machine?
- Liquid State Machine (LSM)
- LSM with parallel perceptron readout (LSM-PPR)
- LSM with dendritically enhanced readout (LSM-DER)
  - Concept of neuron with active dendrites
  - Architecture of LSM-DER
  - Network Re-wiring(NRW) rule
- Experiments and Results
  - Problem description
  - LSM-DER and NRW rule performance
  - Comparison with LSM-PPR
  - Stability with respect to non-idealities
- Conclusion

NANYANG
TECHNOLOGICAL
UNIVERSITY

Conclusion

- Proposed new hardware-friendly readout stage for Liquid State Machine
- LSM-DER and NRW rule achieves better results using less resources
- LSM-DER uses binary synapses
- Resilient to VLSI mismatch

Thank You

Questions?

NANYANG TECHNOLOGICAL UNIVERSITY