# Architectural Exploration for On-chip, Online Learning in Spiking Neural Networks

**Subhrajit Roy, Sougata Kr. Kar and Arindam Basu**

# OUTLINE

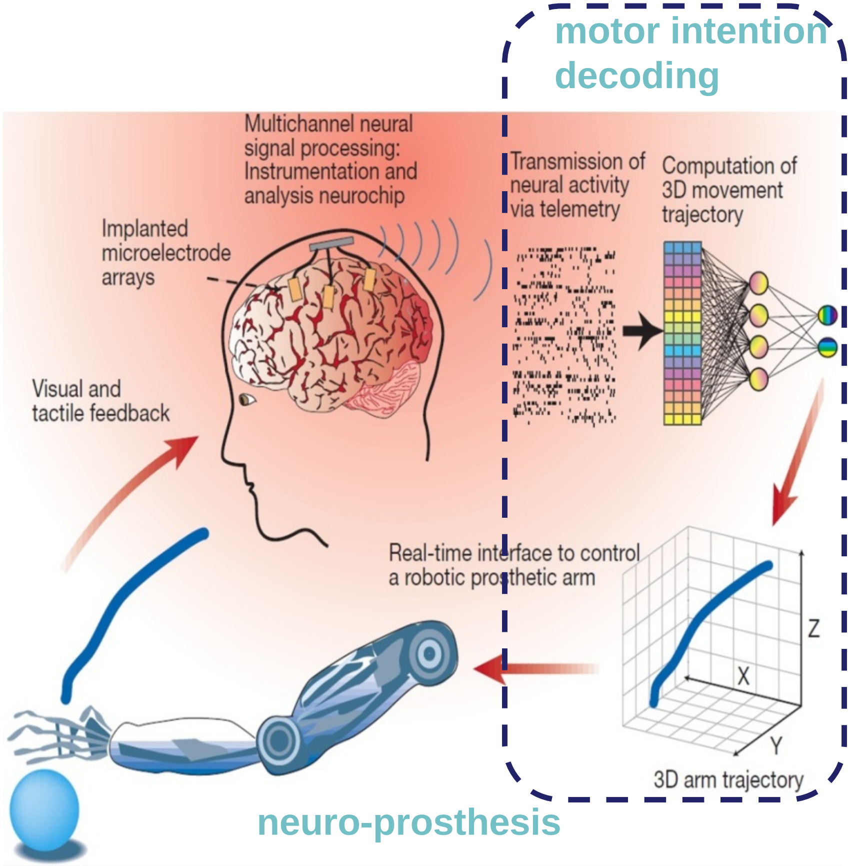
- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

# OUTLINE

- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

# Motivation for on-chip 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

# Motivation for online learning

- Storage space reduction by storing 'mini-batches'
- Accommodates time constant limitation for calculating 'true gradient'
- Robust to changes of the input data distribution

# OUTLINE

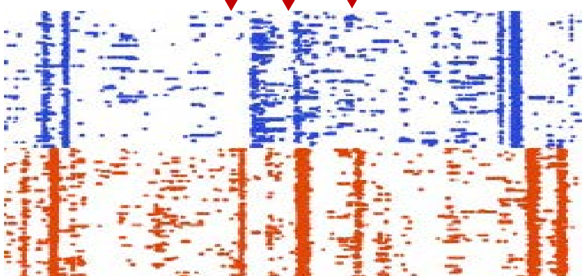
- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

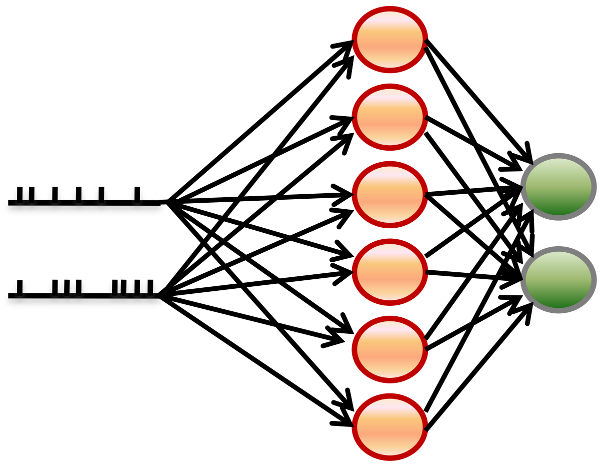
NANYANG TECHNOLOGICAL UNIVERSITY

# Motivation for active dendrites and structural plasticity

**Neuromorphic sensors – Retina-inspired Imager** (Lichtsteiner et al., 2008)

**Applications in face recognition, handwriting recognition**





**Spike inputs from biological neurons**

**Applications in brain-machine interfaces (BMI)**

**<u>Goals/Challenges Design classifiers:</u>**

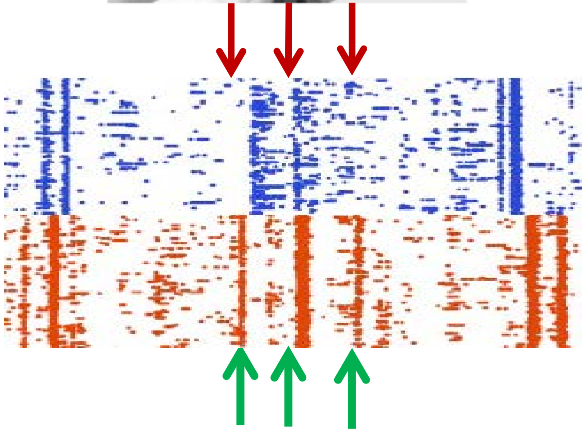
**1) Reduce number of synapses without hampering accuracy**
**2) Learning using synapses with low resolution weights**

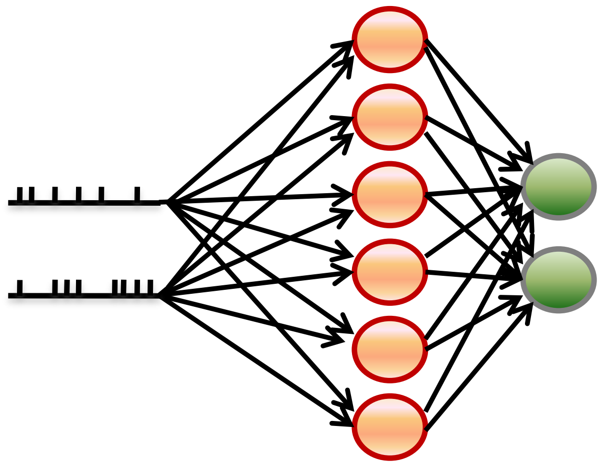
NANYANG TECHNOLOGICAL UNIVERSITY

# Motivation for active dendrites and structural plasticity

**Neuromorphic sensors – Retina-inspired Imager** (Lichtsteiner et al., 2008)

Applications in face recognition, handwriting recognition





**Spike inputs from biological neurons**

Applications in brain-machine interfaces (BMI)

## Solutions

1) **Reduce number of synapses without hampering accuracy**

**- Exploit dendritic properties**

2) **Learning using synapses with low resolution weights**

**- Exploit structural plasticity**

NANYANG TECHNOLOGICAL UNIVERSITY

Poirazi 2001, Roy 2014, Hussain 2013

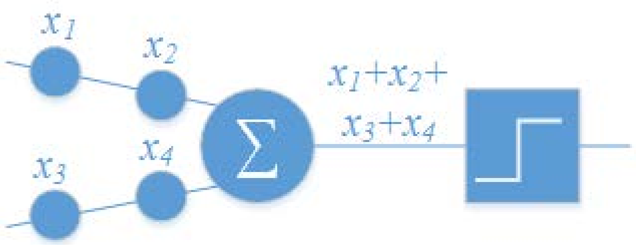
# OUTLINE

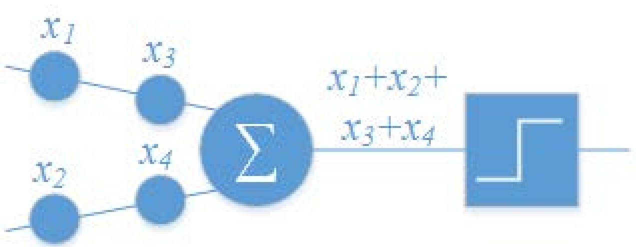
- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

# Neurons with Active Dendrites

Linear Cell w Binary Synapses



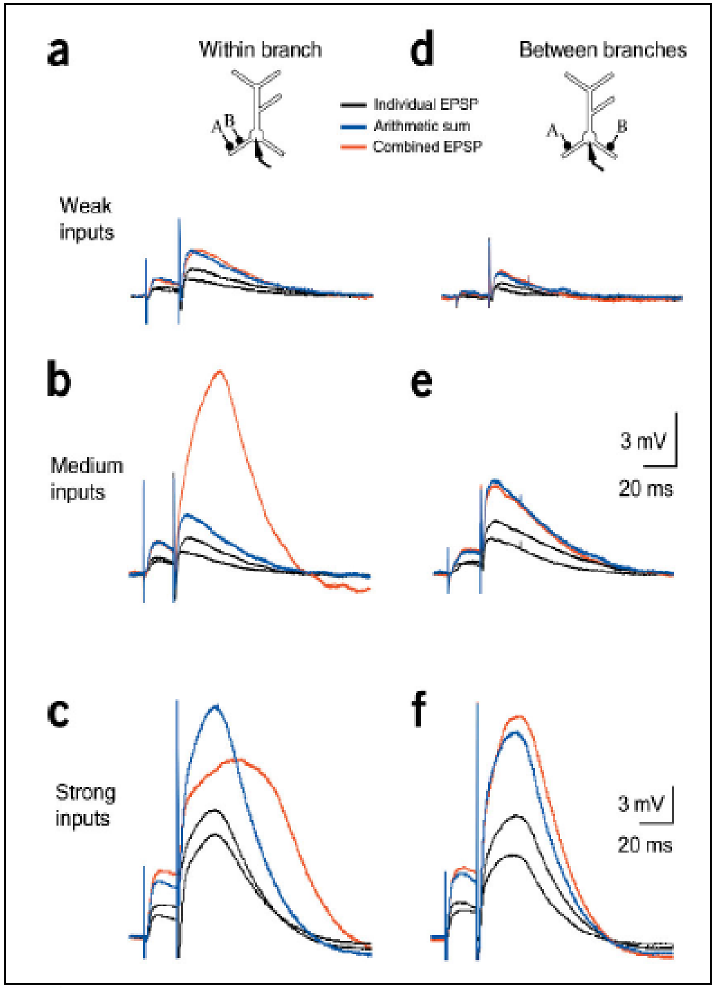




Unable to recognize the different combination of inputs

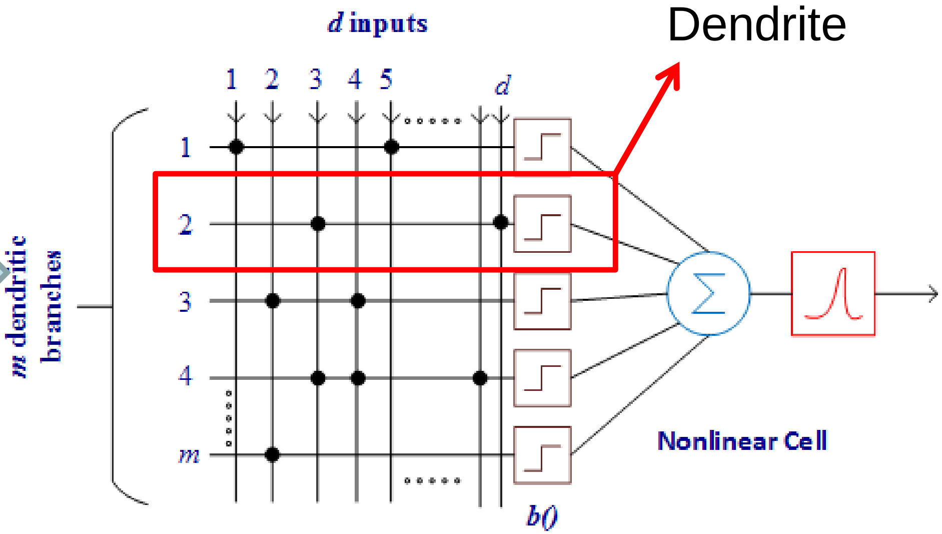
NANYANG TECHNOLOGICAL UNIVERSITY

# Neurons with Active Dendrites

Motivation




#synapses per branch : $k$ ($k << d$)




Source: Poirazi et al., 2001

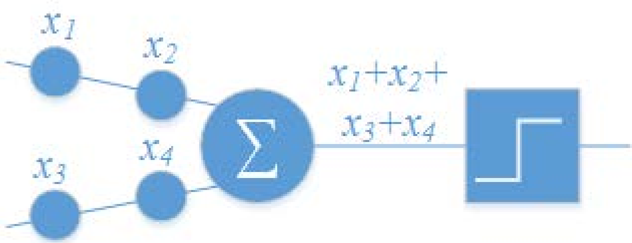
- Currently considering only lumped dendritic nonlinearity

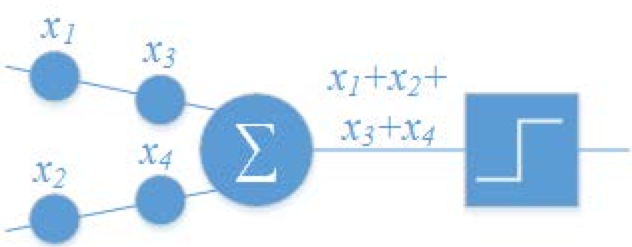
NANYANG TECHNOLOGICAL UNIVERSITY

Source: Polsky et al., 2004

# Neurons with Active Dendrites

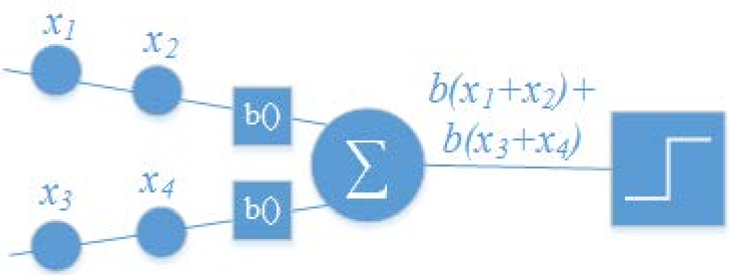
Linear Cell w Binary Synapses

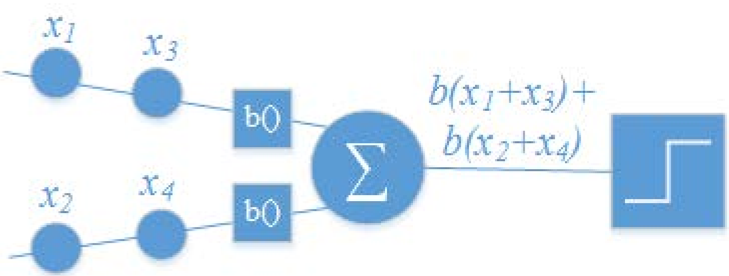



Unable to recognize the different combination of inputs

Non linear cell w Binary Synapses




Capable of recognizing the different combination of inputs

NANYANG TECHNOLOGICAL UNIVERSITY

# OUTLINE

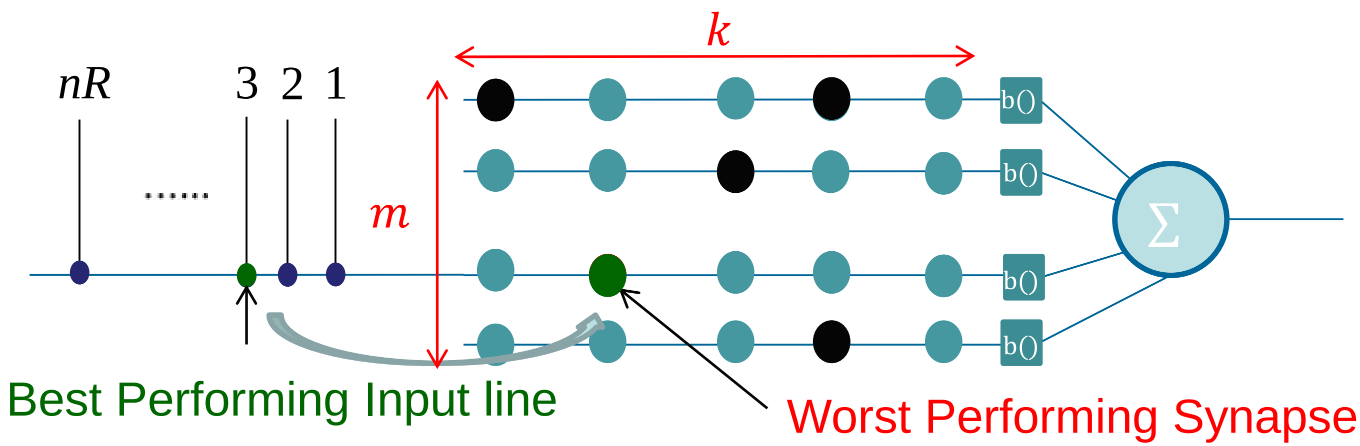
- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

# Network Re-Wiring Learning Rule

Employ two neuronal cells for binary classification

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



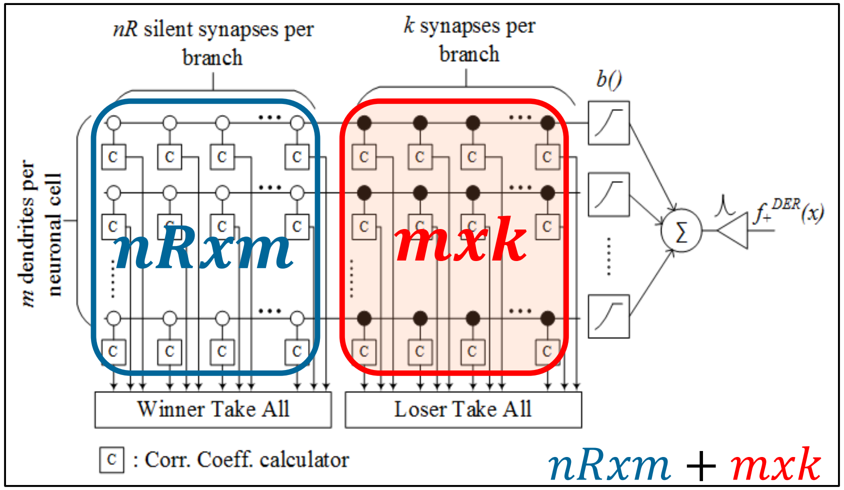

$n_T = 5$

- Search for the worst performing synapse i.e. lowest $c_{ij}$ synapse in the set $n_T$
- Choose a random set $n_R$ of the input lines → Replace by its best

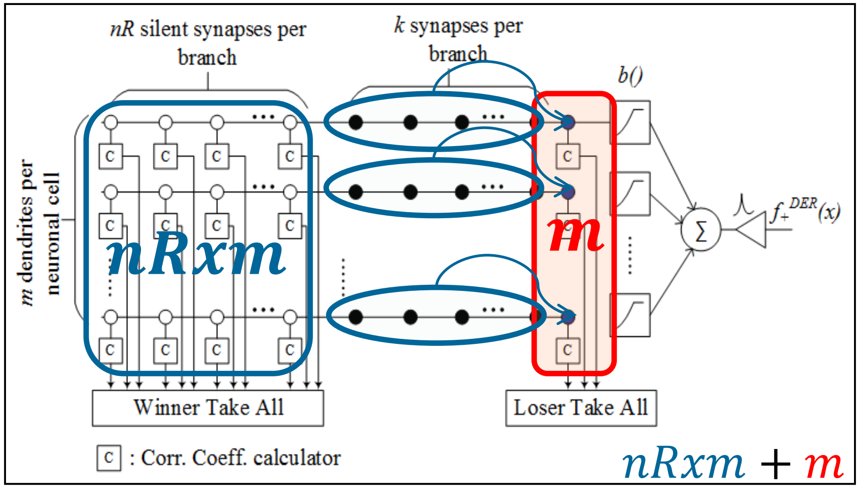
# OUTLINE

- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

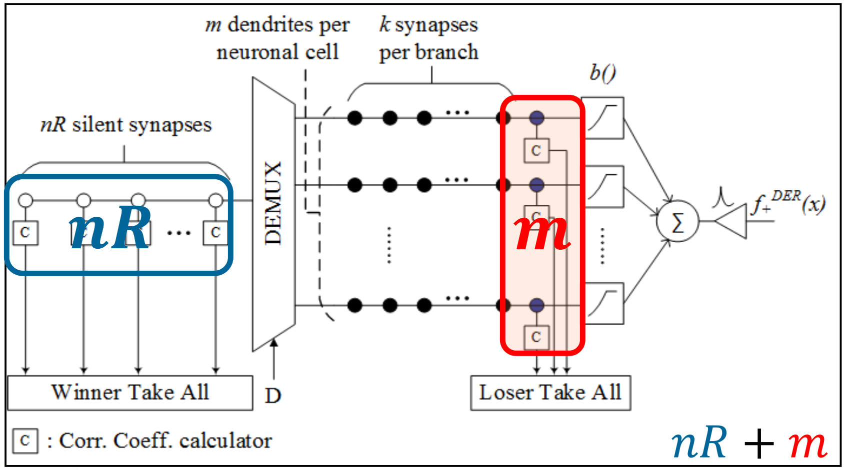
# Architectural Explorations



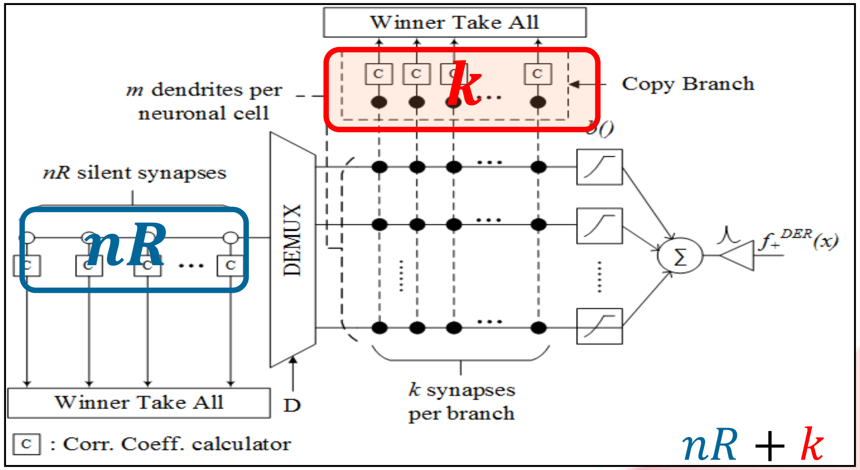

Arch. I




Arch. II




Arch. III




Arch. IV

NANYANG TECHNOLOGICAL UNIVERSITY

# Comparison of architectures

***$N_c$***: Number of $c_{ij}$ calculators per cell

Task considered : Classifying spatio-temporal Spike trains at the output of reservoir of Liquid State Machine

Table 1: Comparison of DER architectures*

| Arch. | $N_c$ | MAE | Remarks |
|---|---|---|---|
| I | $k \times m + nR \times m$ (119) | 0.0912 | Least $MAE = MAE_{min}$, Highest $N_c = N_{c,\max}$ |
| II | $m + nR \times m$ (56) | 0.0946 | $MAE \approx MAE_{min}$, $N_c < N_{c,\max}$ |
| III | $m + nR$ (14) | 0.1088 | $MAE = 1.1930 MAE_{min}$, $N_c \ll N_{c,\max}$ |
| IV | $k + nR$ (17) | 0.1094 | $MAE = 1.1996 MAE_{min}$, $N_c \ll N_{c,\max}$ |

*$k$=10, $m$=7 and $nR$=7

Only 1.76% MAE increase by using 8.5 times less $c_{ij}$ calculators

NANYANG TECHNOLOGICAL UNIVERSITY

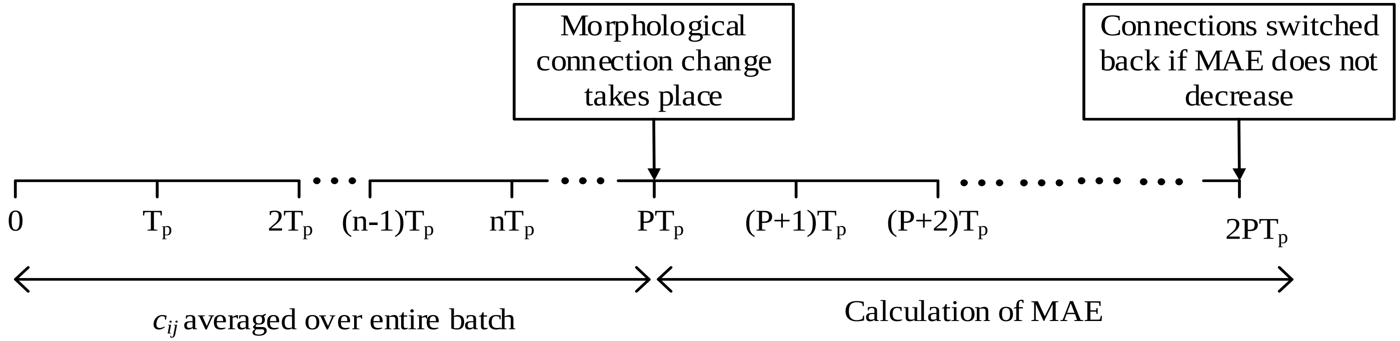
# OUTLINE

- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

NANYANG TECHNOLOGICAL UNIVERSITY

# Online NRW learning rule

Timeline of **batch** NRW rule:




$T_p$ : Pattern duration and $P$ = # patterns

**LIMITATIONS**

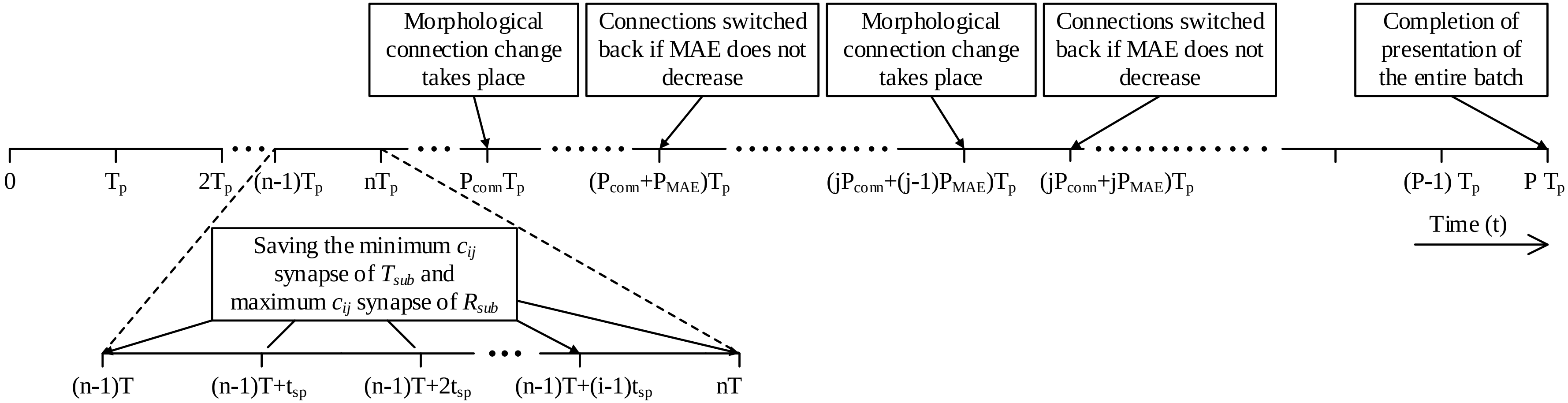
- Let $T_p = 0.5\,secs$ and $P$ = 100 $\rightarrow PT_p = 50\,secs$
- Analog averaging circuits operate in the order of milliseconds

**NOT SUITABLE** → Look for Online Learning

NANYANG TECHNOLOGICAL UNIVERSITY

# Online NRW learning rule

Timeline of **online** NRW rule:




$T_p$ : **Pattern duration**

$t_{sp}$ : **Averaging time**

$P_{conn}$ : **# patterns after which connection change takes place**

$P_{MAE}$ : **# patterns required for MAE calculation**

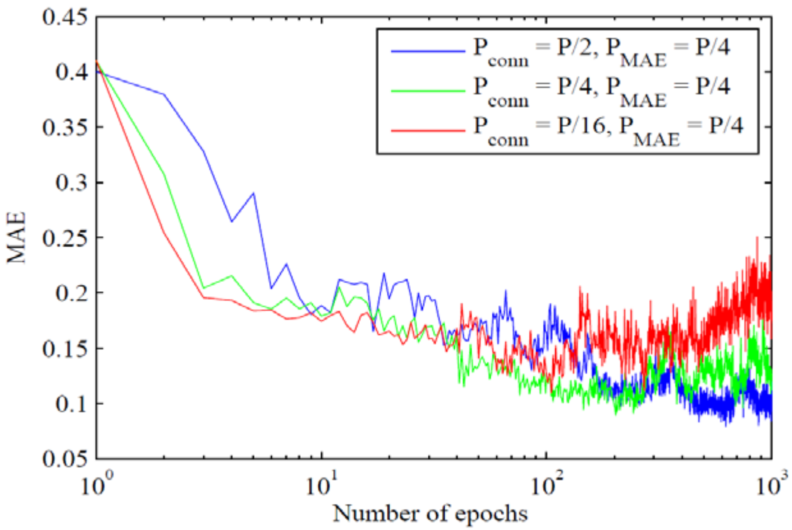
# OUTLINE

- Motivation for on-chip machine-learners and online learning
- Motivation for active dendrites and structural plasticity
- Neurons with active dendrites
- Network Re-Wiring (NRW) Learning rule for structural plasticity
- Architectural Explorations
- Online NRW learning rule
- Performance and Conclusion

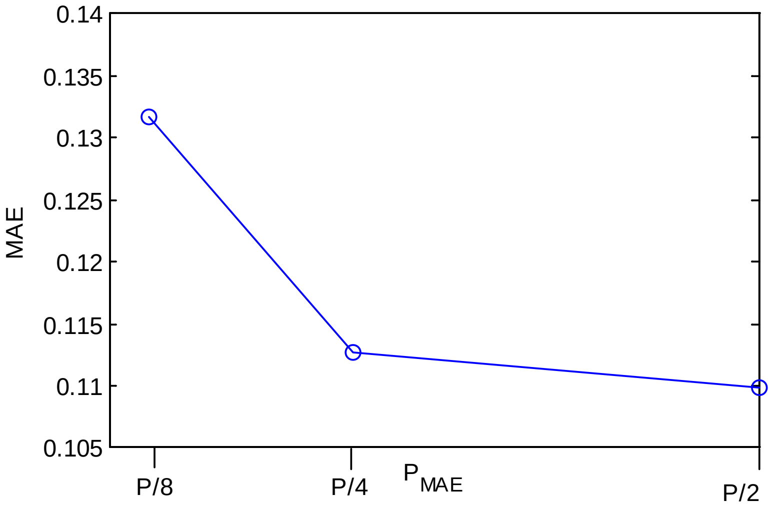
NANYANG TECHNOLOGICAL UNIVERSITY

# Performance of online NRW rule

$$P_{MAE} = \frac{P}{4}, P_{conn} \text{ varied}$$




$$P_{conn} = \frac{P}{4}, P_{MAE} \text{ varied}$$




When $P_{MAE} = P$ and $P_{conn} = P$ MAE = 0.1075 → Comparable performance by using 4200 times lesser time constant averaging circuit

NANYANG TECHNOLOGICAL UNIVERSITY

# Conclusion

- Various architectures for hardware implementation of DER and NRW
- Devised an optimized architecture maintaining similar performance.

  ⟹ Uses 8.5 times less resources while hurting performance by 1.76%

- Proposed an online version of the NRW learning rule

  ⟹ Uses an averaging circuit having 4200 times lesser time constant while providing similar performance.

THANK YOU

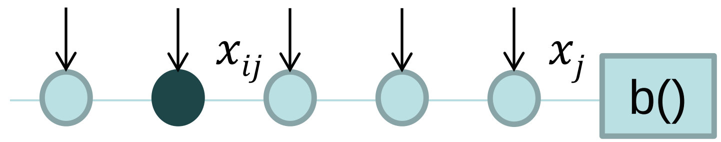
QUESTIONS?

# APPENDIX

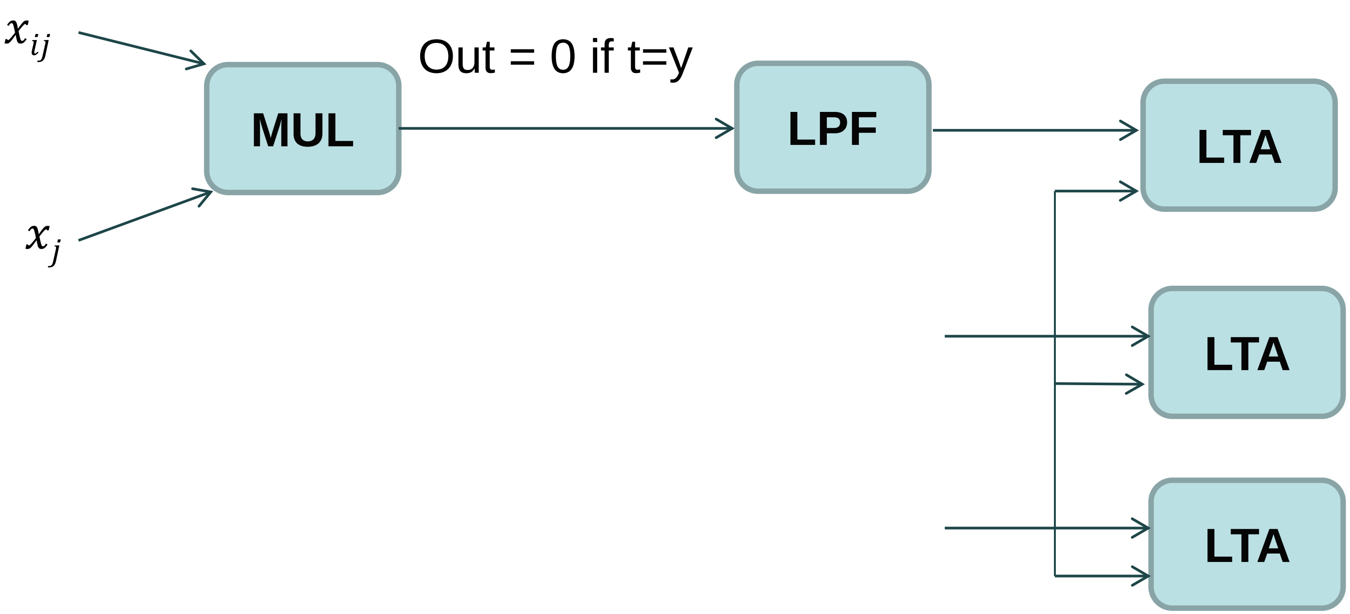
NANYANG TECHNOLOGICAL UNIVERSITY

# $c_{ij}$ calculator

$c_{ij} =< xijxj(t-y) >$




**If $t = 1$**

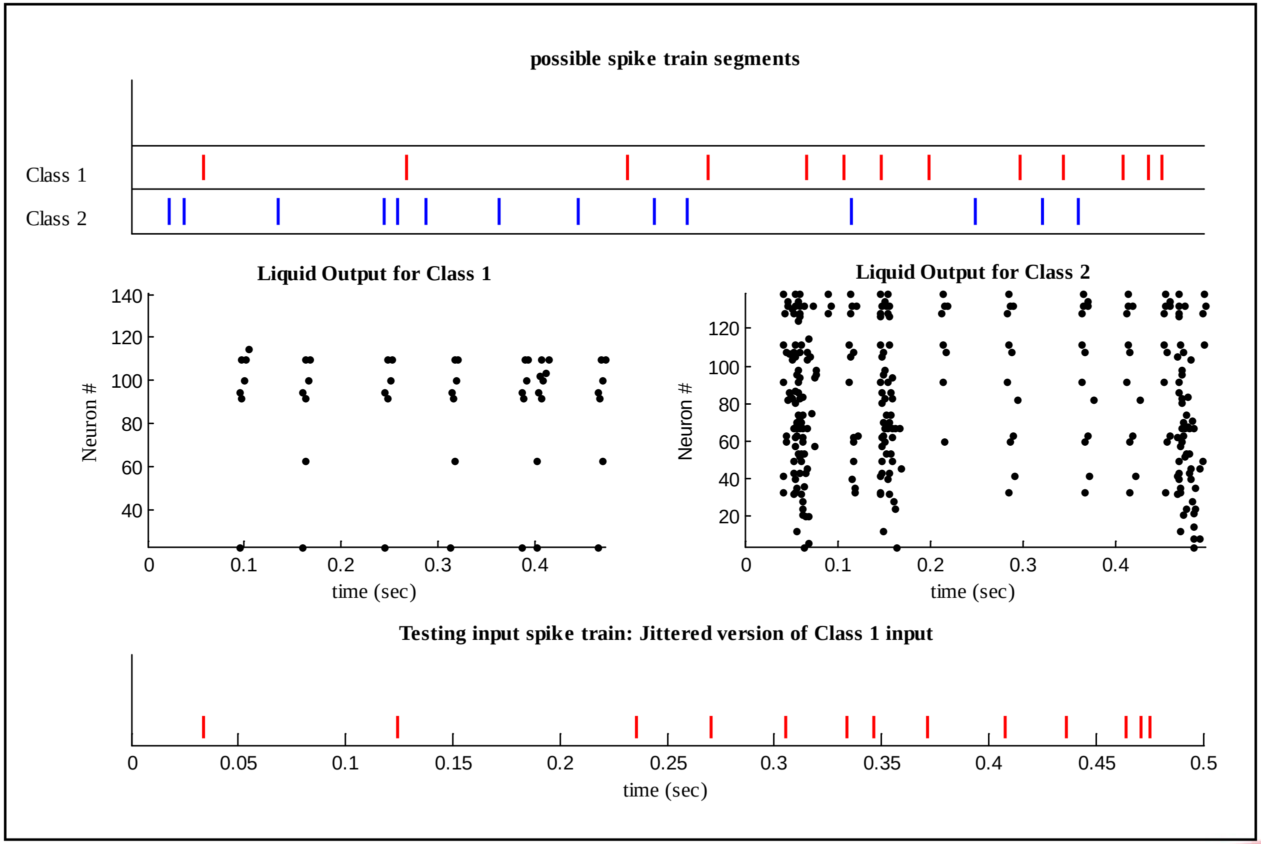



**If $t = 1$** LTA replaced by WTA

NANYANG TECHNOLOGICAL UNIVERSITY

# Experiments and Results

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




NANYANG TECHNOLOGICAL UNIVERSITY

# Liquid State Machine

**Liquid**

- Huge recurrent interconn-ection of spiking neurons

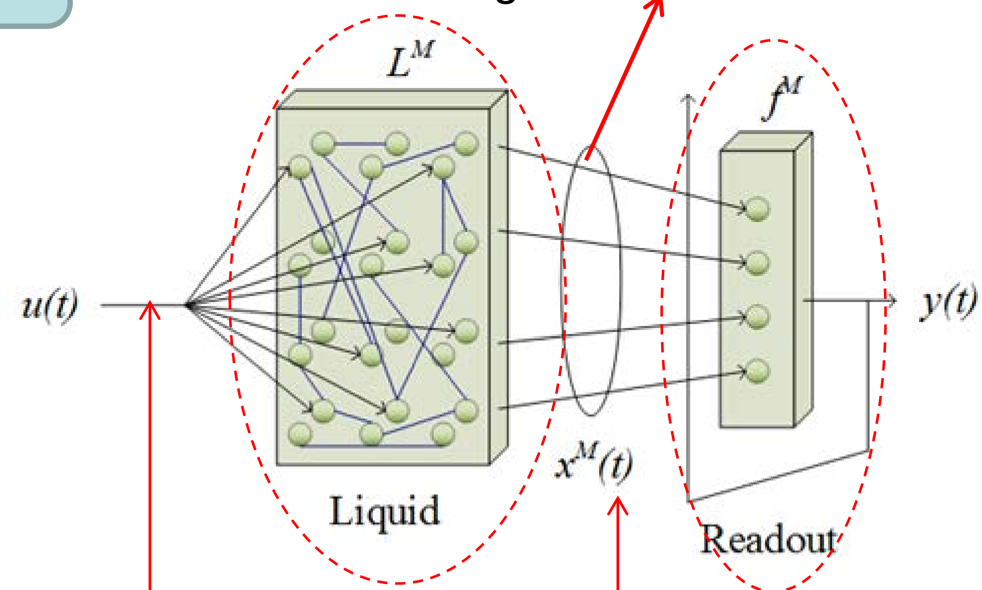
+

- Synapses with random weights

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

Learning of only these synapses

No learning



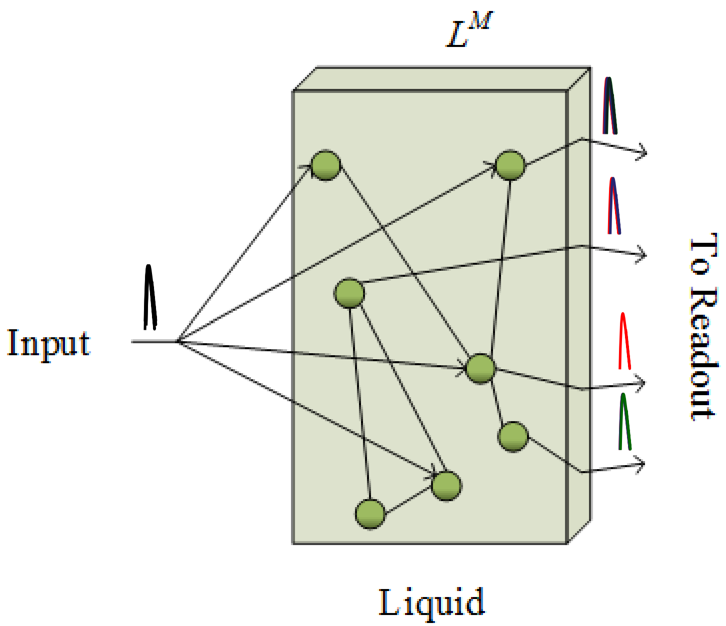

Inputs: Spike Trains or Analog current

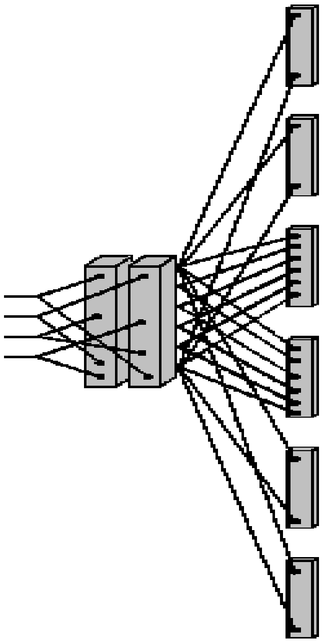
Task-specific training of readout

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**:

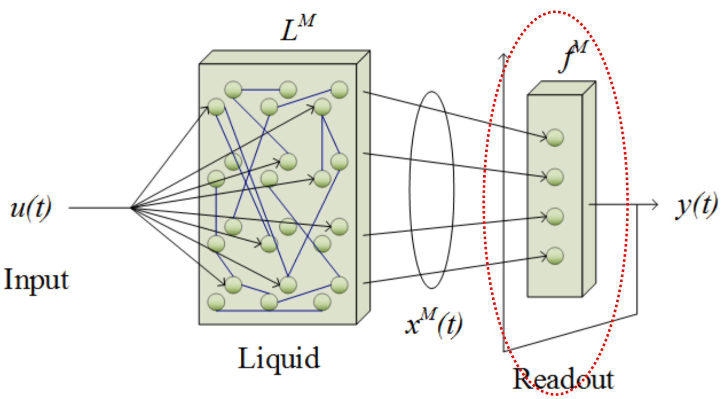
1. Encodes Spike Times
2. Low dimension to high dimension → Increases separability
3. Recurrence → Memory effect
4. General → Multiple features extracted

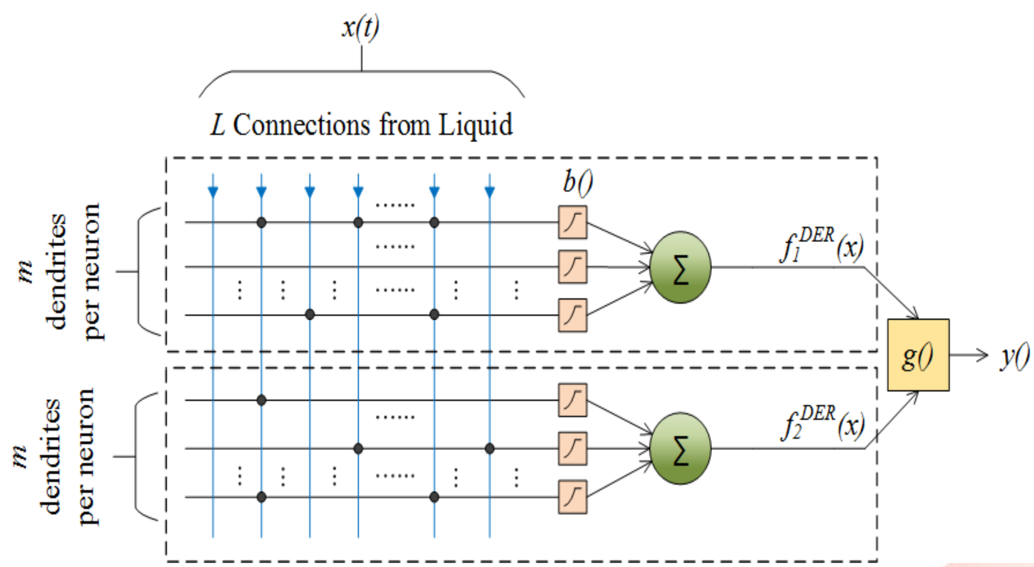
NANYANG TECHNOLOGICAL UNIVERSITY

# LSM w Dendritically Enhanced Readout (LSM-DER)

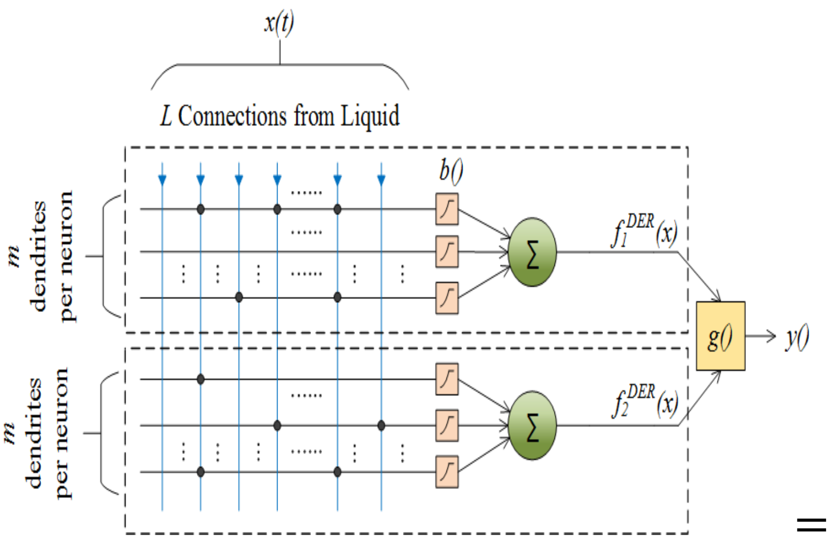
- The **proposed algorithm LSM-DER** constitutes the liquid of LSM followed by a two neuronal cell architecture
- The parallel perceptron stage of LSM-PPR has been replaced by a two neuronal cell architecture.

**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:

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

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

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

**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 $\Delta w_{ij}$ 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.