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
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
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
Liquid State Machine

Huge recurrent interconnection 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

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
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
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 (\textit{p}-delta learning rule) specifically designed for training of parallel perceptrons is used.

LSM-PPR limitation

1. If $L$ liquid neurons and $n$ readout neurons then total tunable synapses: $Lxn$
   - 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
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

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

Source: Polsky et al., 2004

Source: Poirazi et al., 2001
Neurons with Active Dendrites

Linear Cell w Binary Synapses

Non linear cell w Binary Synapses

Unable to recognize the different combination of inputs

Capable of recognizing the different combination of inputs
Neurons with Active Dendrites

\[ s: \text{Total number of synapses} \]
\[ m: \text{No. of dendritic branches} \]
\[ k: \text{No. of synapses per branch} \]
\[ d: \text{Dimension of input} \]

\[ B_N = \log_2 \left( \frac{s + d - 1}{s} \right) \]
\[ B_L = \log_2 \left( \frac{k + d - 1}{k} + m - 1 \right) \]

- Synaptic weight = 0/1
- Not weight update, but Connection change
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** : $L \times n$ synapses
**LSM-DER** : $L$ synapses
Network Re-Wiring (NRW) Learning Rule

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

\[
\Delta w_{ij} = - \frac{\partial e^2}{\partial w_{ij}} = 2 < (t - y) \frac{\partial y}{\partial w_{ij}} > \\
= 2 < (t - y) \frac{\partial g \left( f_1^{DER}(x) - f_2^{DER}(x) \right)}{\partial w_{ij}} > \\
= 2 < (t - y) \frac{\partial g \left( \sum_{j=1}^{m} b \sum_{i=1}^{k} w_{ij} \right)}{\partial w_{ij}} >
\]

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 \( c_{ij} \)

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

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

Remove the connected input line to the synapse by any random input line $\rightarrow$ No extra calculations but slow learning.

Place all the input lines in the dendrite of worst performing synapse $\rightarrow$ Replace with the best $\rightarrow$ Fast but exhaustive & requires lot of computations.

We take a middle path $\rightarrow$ Choose a random set $n_R$ of the input lines $\rightarrow$ Replace by its best.
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
Experiments and Results

- **Task 1**: Spike Train Classification Problem

![Graph showing possible spike train segments and liquid output for Class 1 and Class 2 neurons.](image)

- Testing input spike train: Jittered version of Class 1 input

![Graph showing liquid output for Class 1 and Class 2 neurons.](image)
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.
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
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

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

Task I : Classification

Task II : Approximation
Stability with respect to non-idealities

Hardware implementation of DER
[Amitava et al. 2015]

Hardware implementation of PPR
[Amitava et al. 2015]
Stability with respect to non-idealities

- Monte Carlo simulations of DPI synapse and Square Block Circuit

Variation in DPI Synapse

- DER: 2.33% increase
- PPR: 4.73% increase

Increased stability due to binary synapses

Maximum Variation
1. $I_0 \sim 13\%$
2. $\tau_s \sim 10.1\%$
3. $c_{ni} \sim 18\%$
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
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?