Interface Engineering of PCM for Improved Synaptic Performance in Neuromorphic Systems

M. Suri[1], O. Bichler[2], Q. Hubert[1] ,L. Perniola[1], V. Sousa[1], C. Jahan[1], D. Vuillaume[3], C. Gamrat[2], B. De Salvo[1]

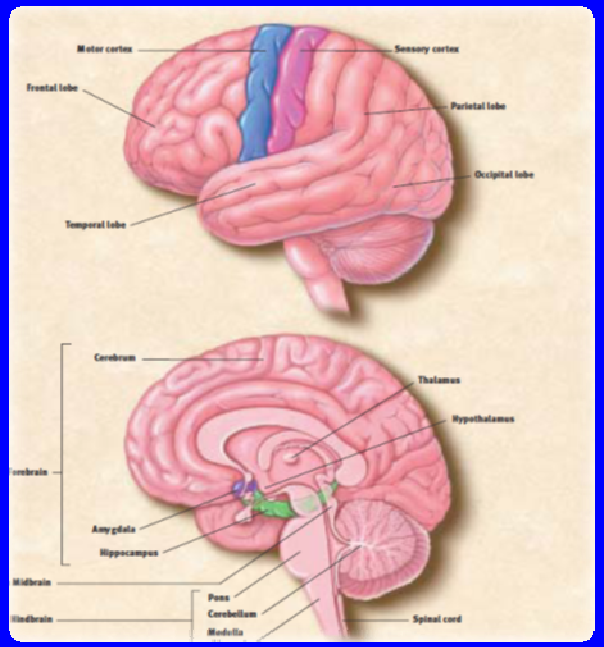
[1]CEA-LETI, [2]CEA-LIST, [3]CNRS-IEMN (France)

What… ?

Taking inspiration form computational neuroscience and learning mechanisms inside the brain

Why?

- **Massive Parallelism**
- **Scalability (< 2L : the brain)**
- **Low Power (~ 20W: the brain)**
- **Immunity to Variability**



Information Storage and Processing are not entirely different tasks

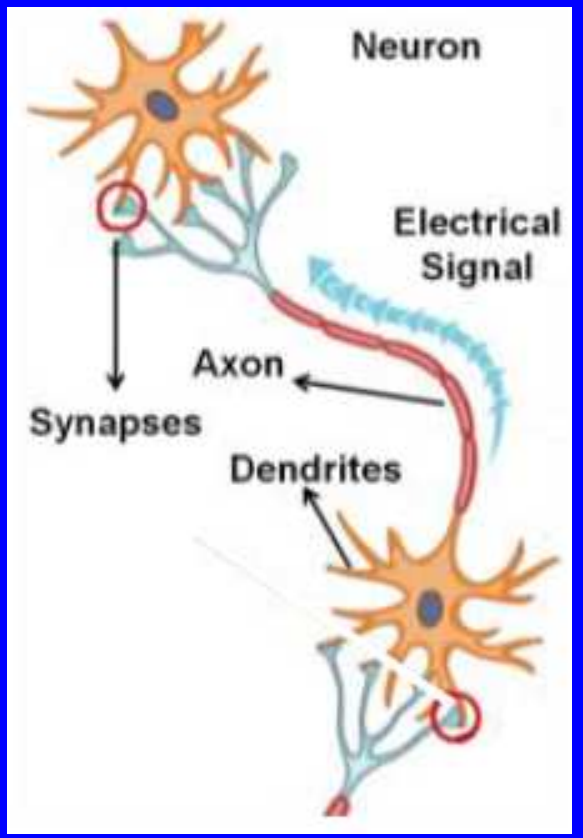
"Memory is Intelligent"

How? Neuron + Synapse + Learning Rule

Neuron

Source of Spikes or Action Potentials (electrical pulses)

Integrates all the incoming spikes
and fires when a threshold is reached



Synapse

Communication channel or the medium between neurons.

Plasticity – can either Potentiate (LTP) or Depress (LTD)

Learning Rule

The protocol to change the synaptic weights (STDP)

Source of Inspiration…?

Region of the brain responsible for vision
Cerebral cortex, hippocampus, hypothalamus…

Main Issues?

(Very Large Numbers)

Neurons: 1.1 x 10^{10}

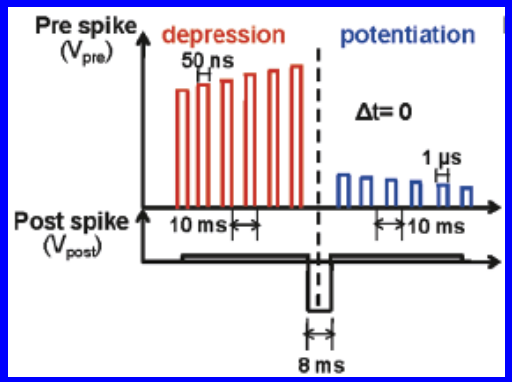
Synapses: 10^{14} – 10^{15}

Need a Device which can change and remember simultaneously!! High transistor count to implement synapse circuit capable of learning !!

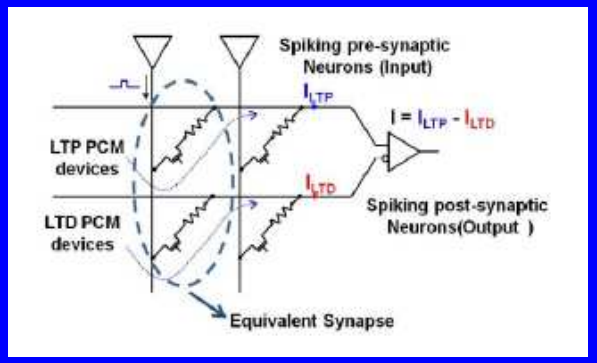
Resistive memories for synapses…

State-of-the-art

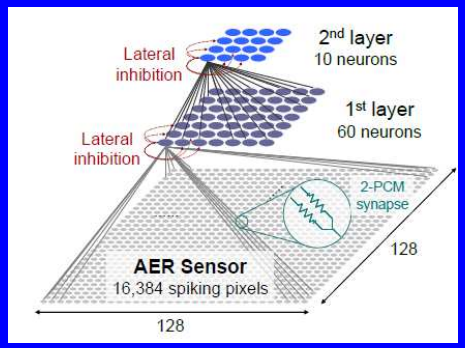
Different Strategies of Implementing STDP on PCM devices



D. Kuzum et. al . Nano Letters, 2011



The *'2-PCM Synapse'*



Spiking Neural Network

Advantages

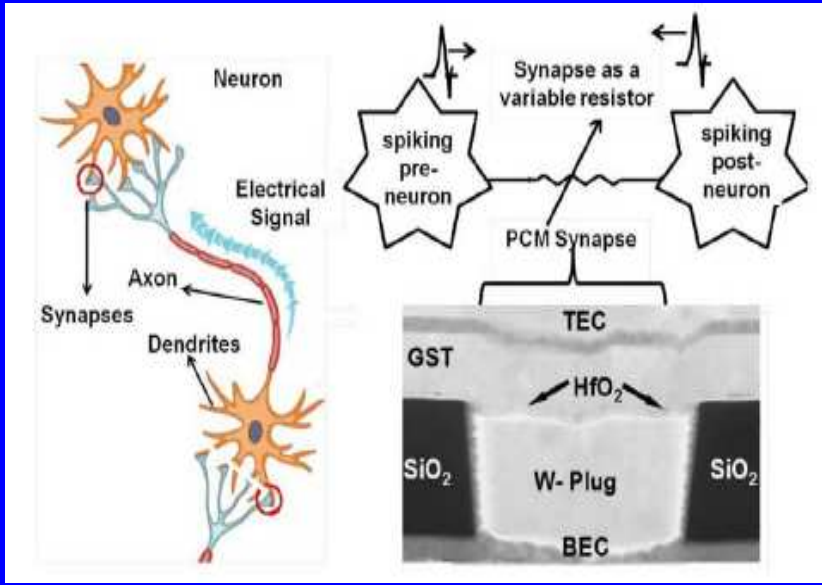
1. Simplified Pulse Schemes
2. Very Low-Power
3. Based on Crystallization

M. Suri et. al . IEDM, 2011

In this Work

Improved System Efficiency compared to GST synapses

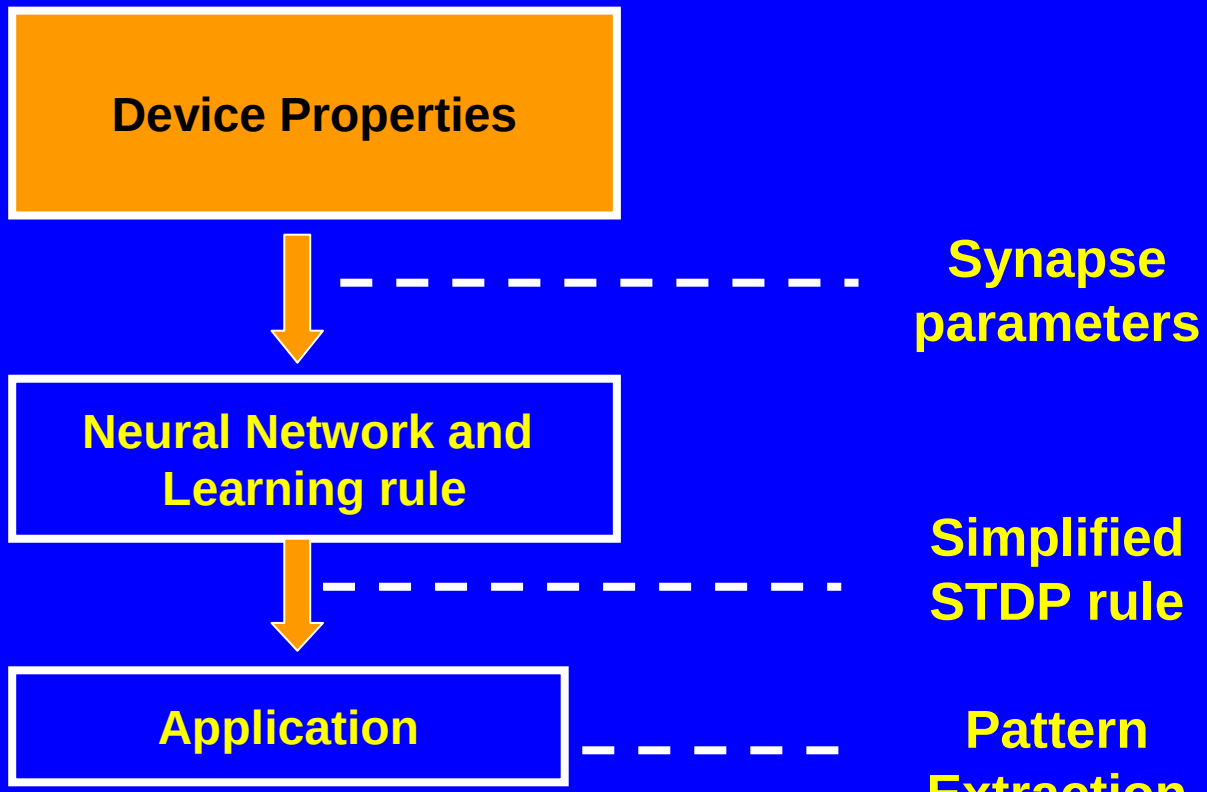
- Low Power at device Level
- Low Power at System Level



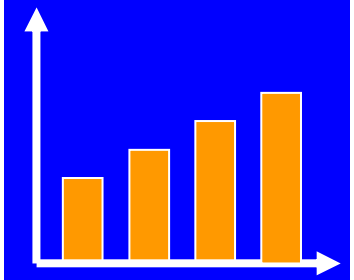
- Large Scale Demonstration (2 million synapses)
- Real World Learning Application (Pattern Extraction)

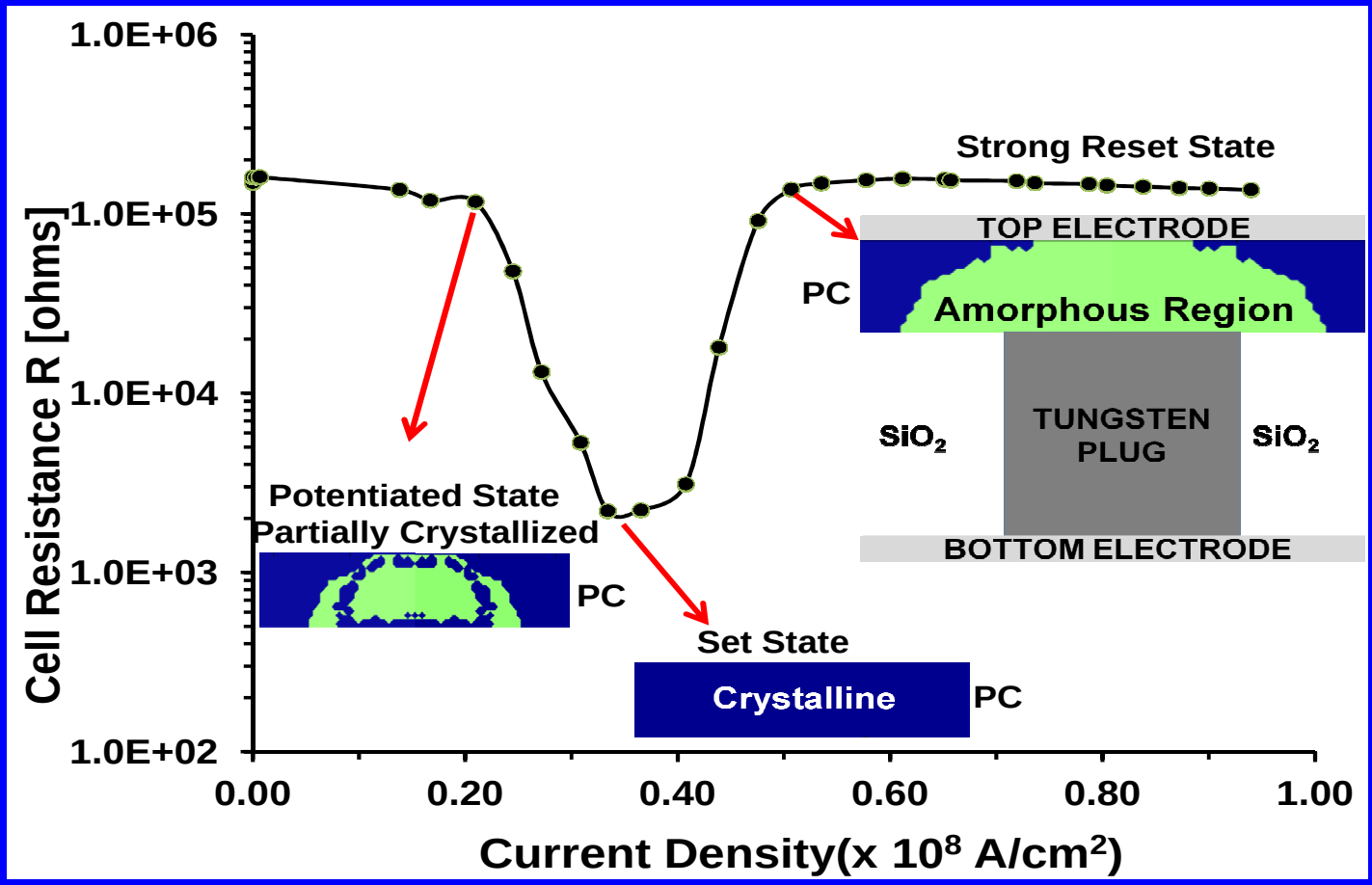
CMOS Neuron + PCM Synapse

Our Approach

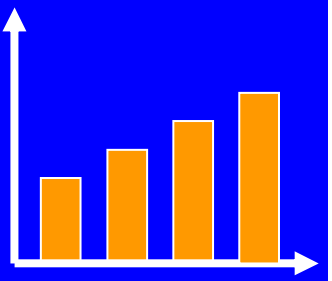


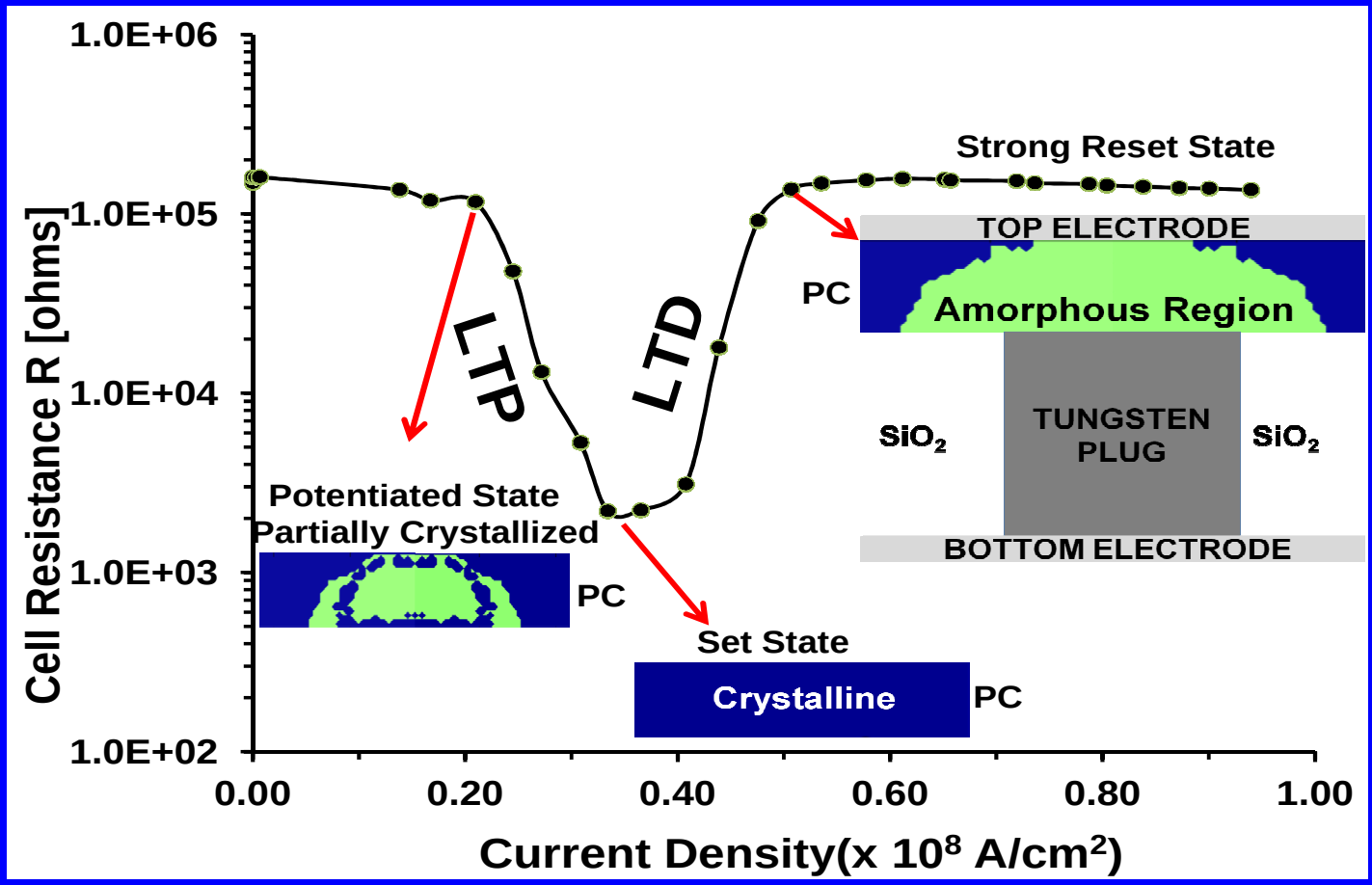
Resistance – Current (LTP/LTD)





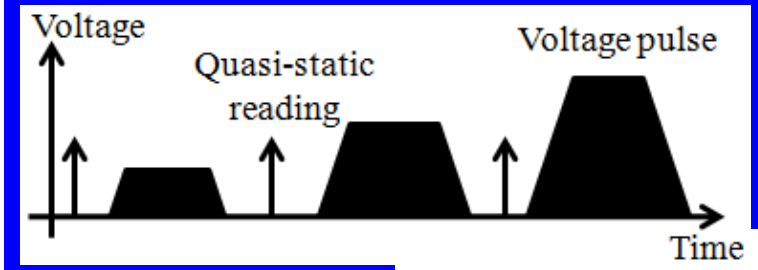
Resistance – Current (LTP/LTD)



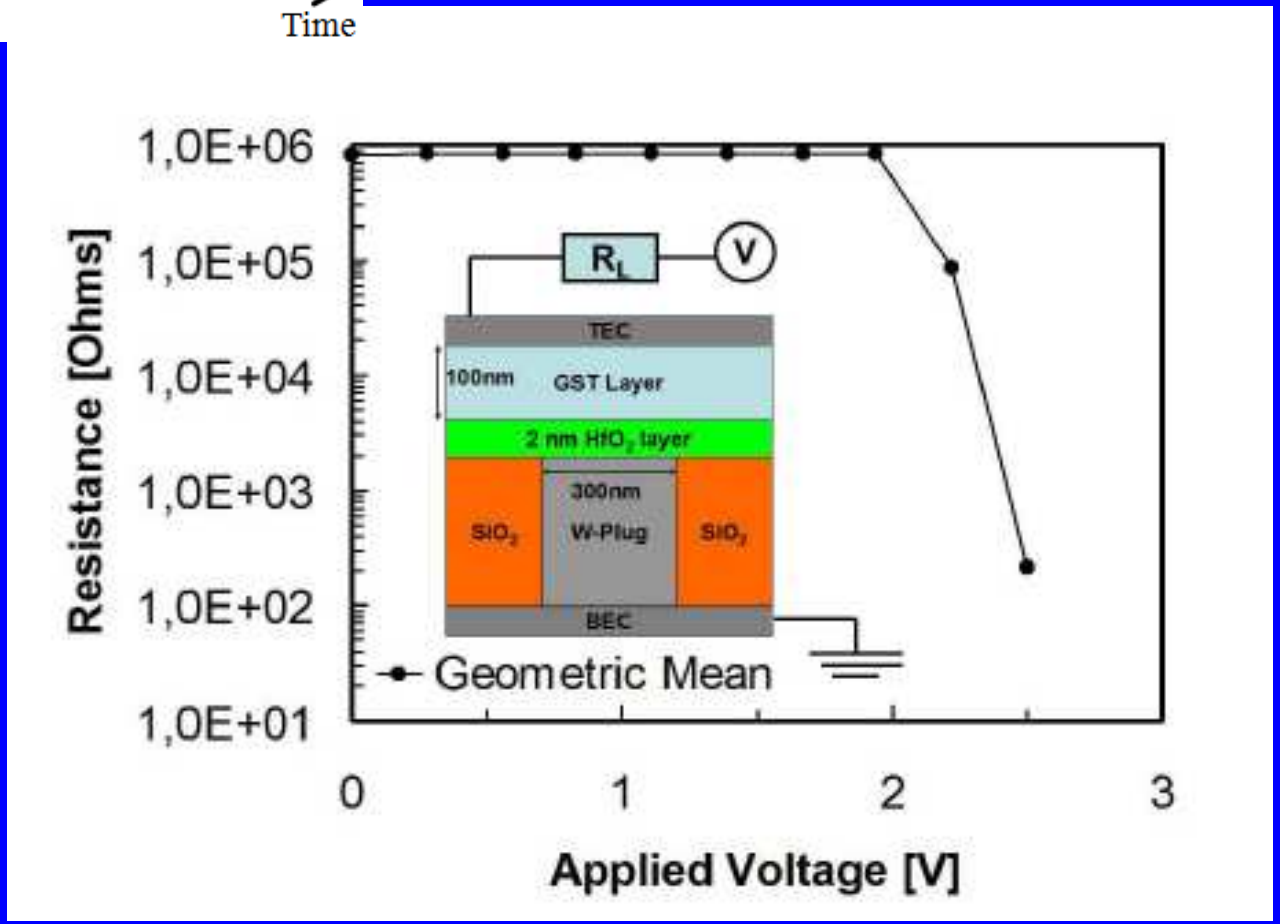


SET to RESET transition = Synaptic Depression (LTD)
RESET to SET transition = Synaptic Potentiation (LTP)

Forming Step – to break the HfO_2 Layer



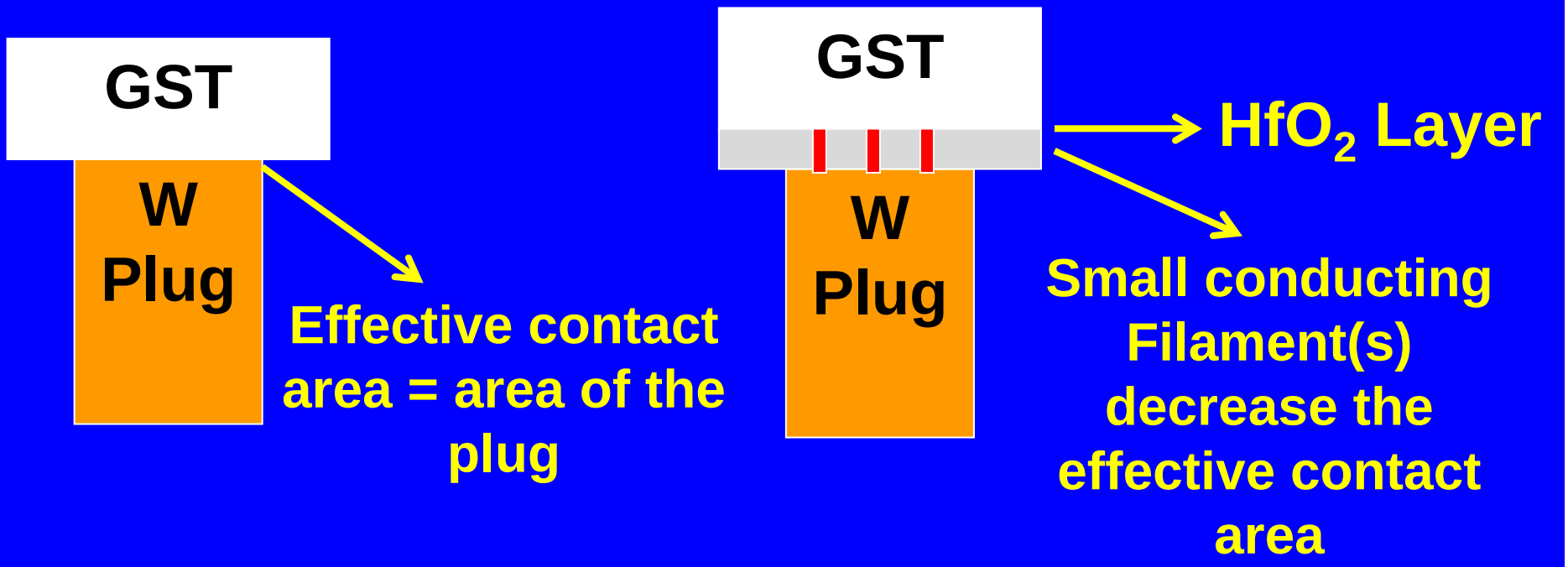
V_{bd} = 2.2V for 2nm thick



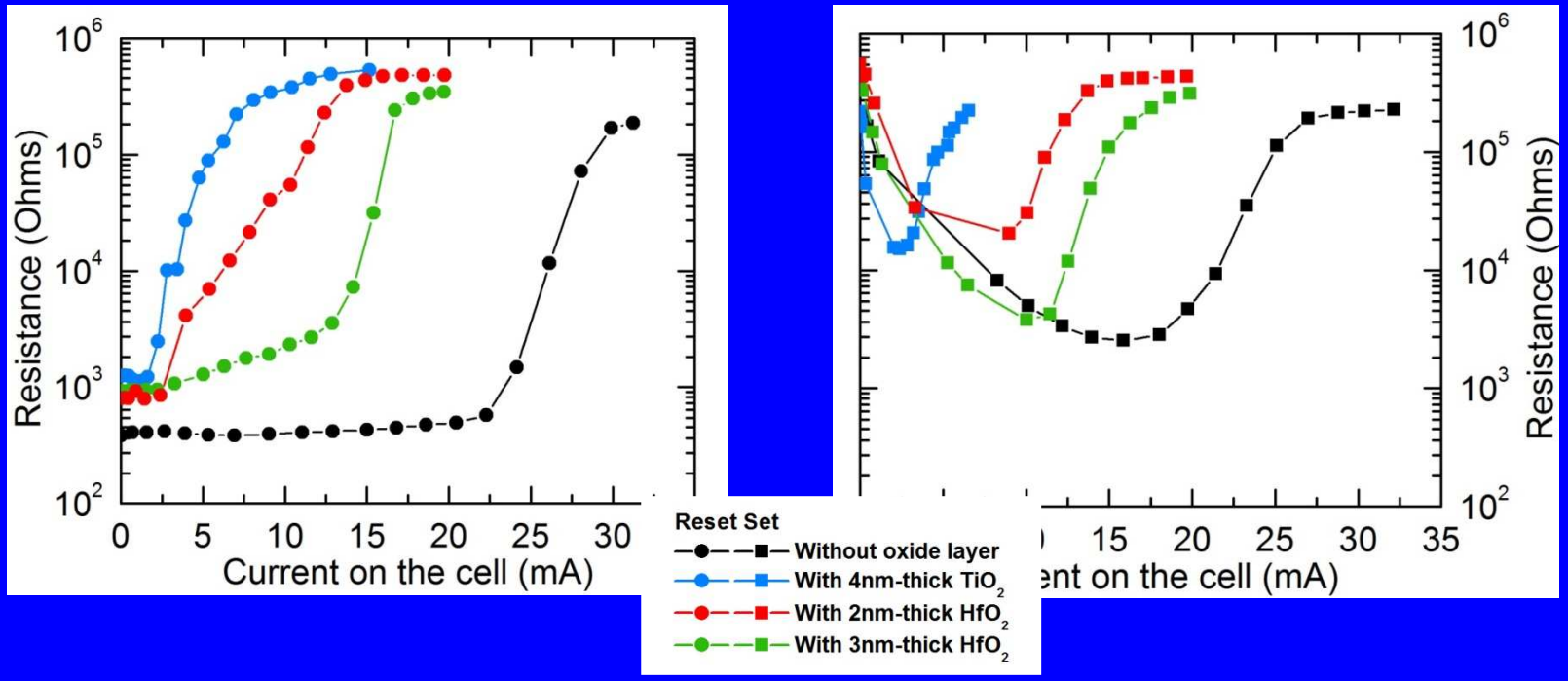
Improvement compared to GST devices

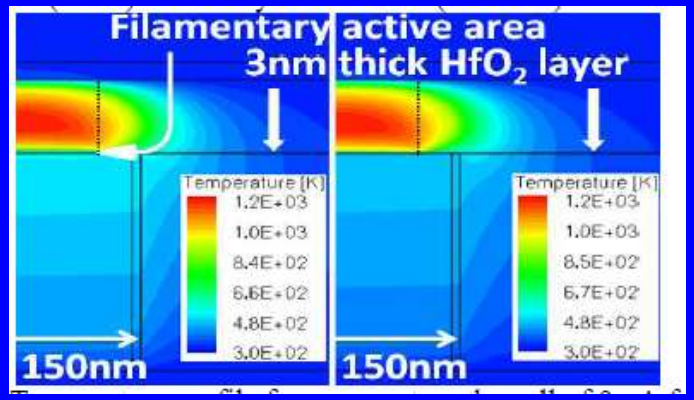
Reduction in Set and Reset Current (> 60%)

- **reduction in the effective contact area between the plug and the GST layer**
- **Less heat-loss (better thermal insulation)**



Programming Current reduction due to Interface layers (HfO_2 and TiO_2)

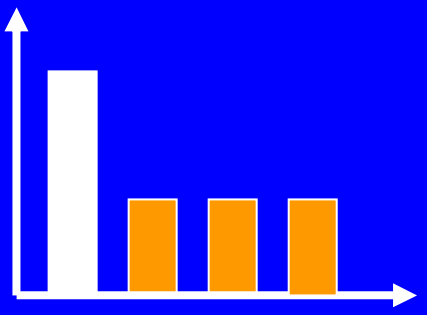




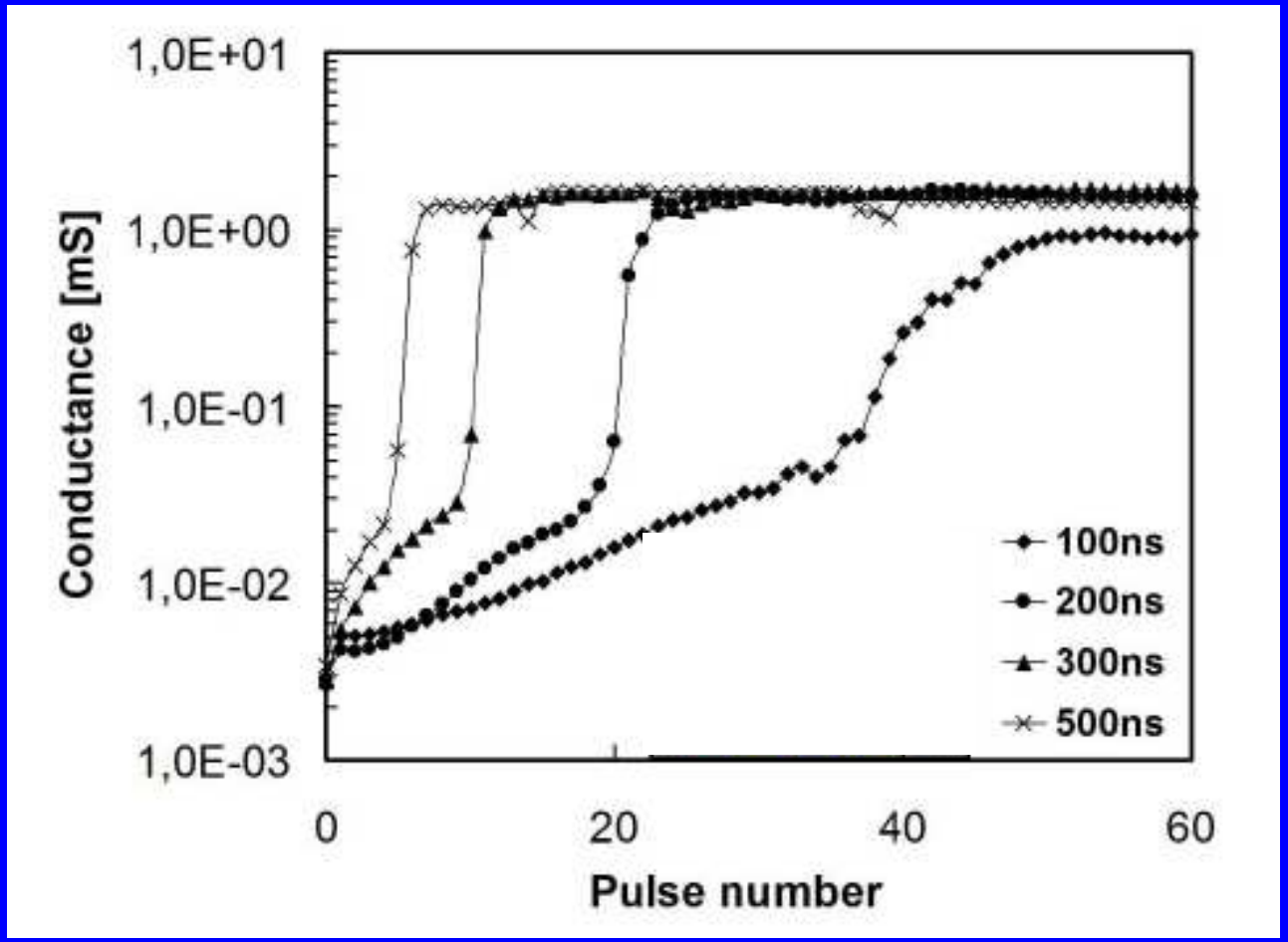
Q. Hubert et. al . ESSDERC, 2011

Less heat loss

Potentiation (LTP) – GST + HfO_2



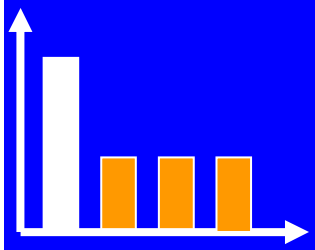
Gradual Crystallization

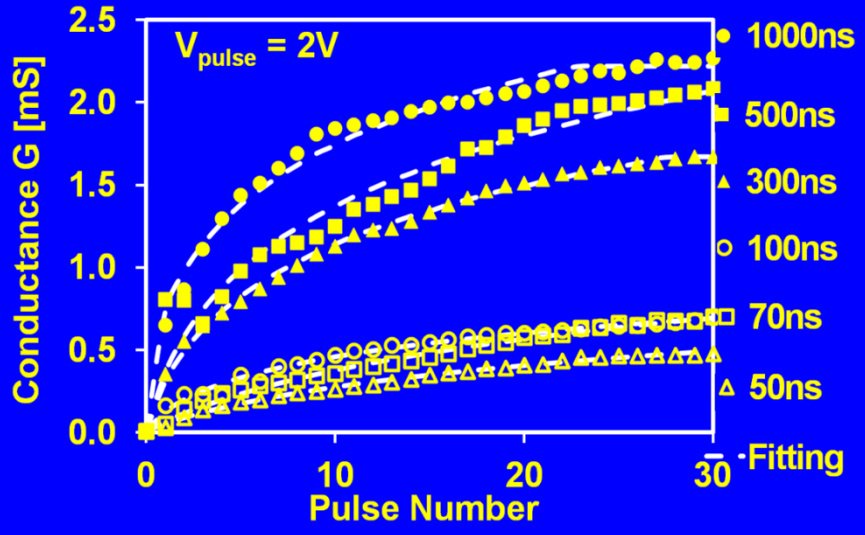


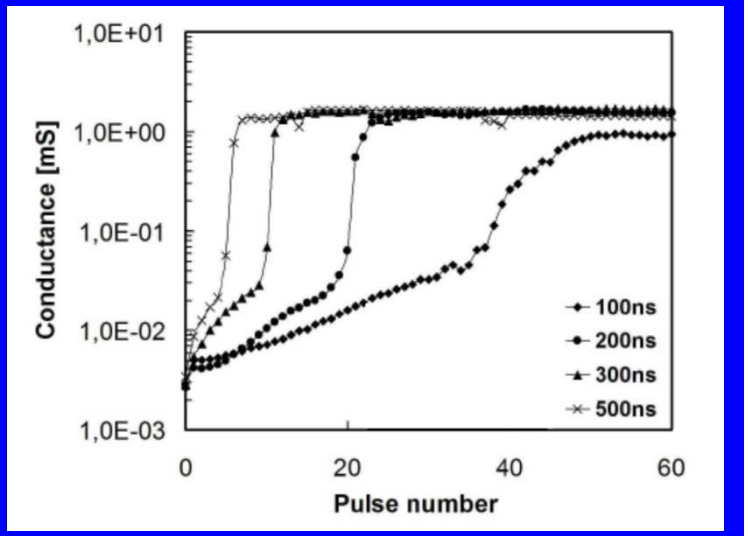
Conductance (Crystallization) increases with increase in pulse number and width

Improvement compared to GST devices

Increased number of intermediate resistance points in the LTP curve (more than double of GST)







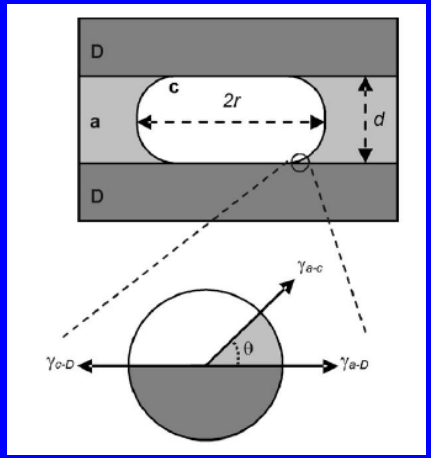
GST

GST + HfO_2

Possible impact of interface layer on crystallization

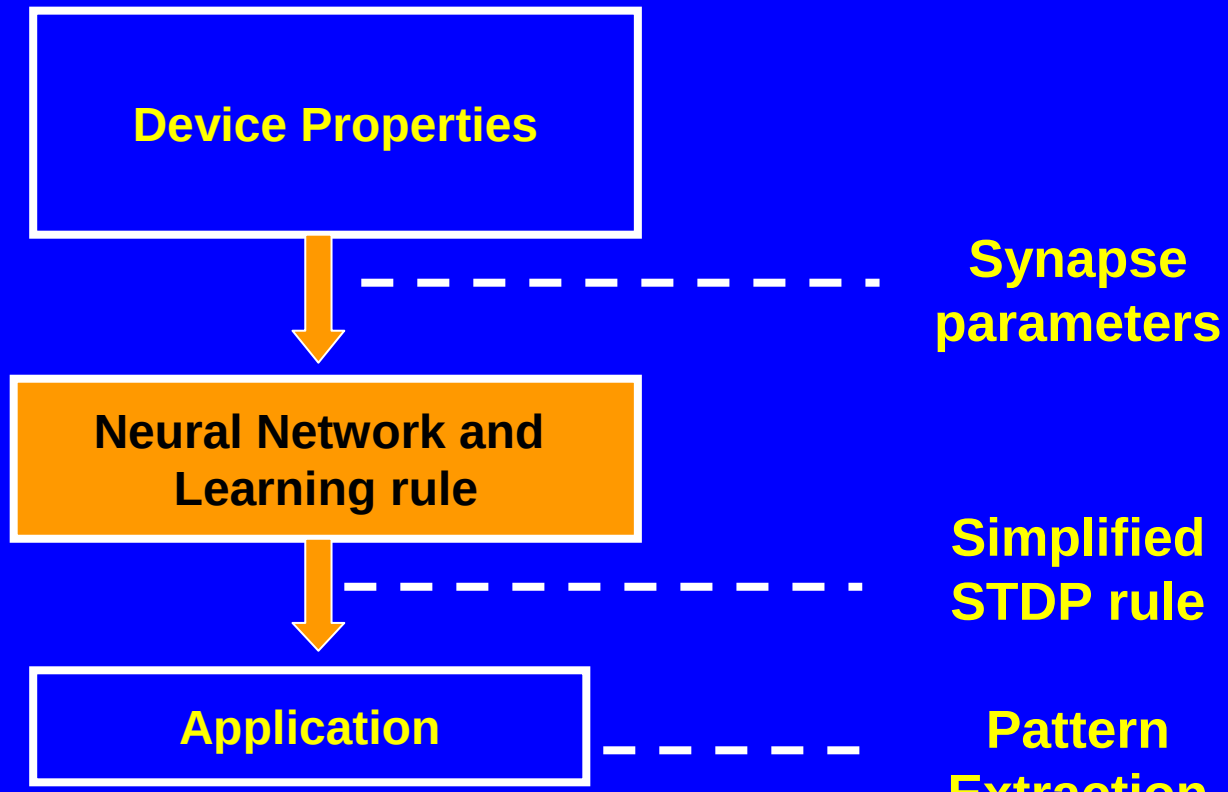
Capping layer can affects the Growth and the Nucleation rate

Based on the type of capping layer and the film thickness, the activation energy for crystal growth can either increase of decrease

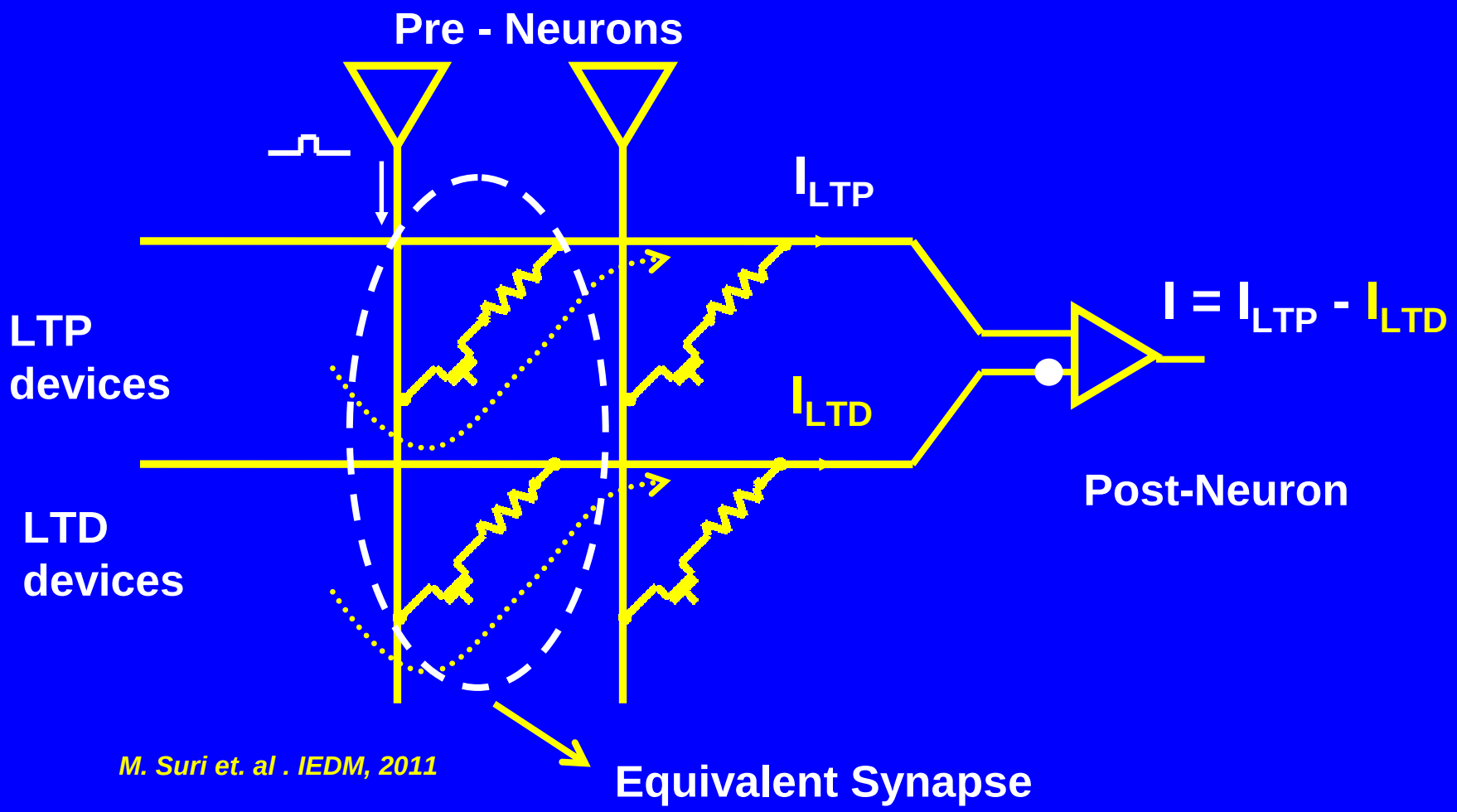


R. Pandian, **et al.,JAP 100, 123511 (2006)**

Our Approach

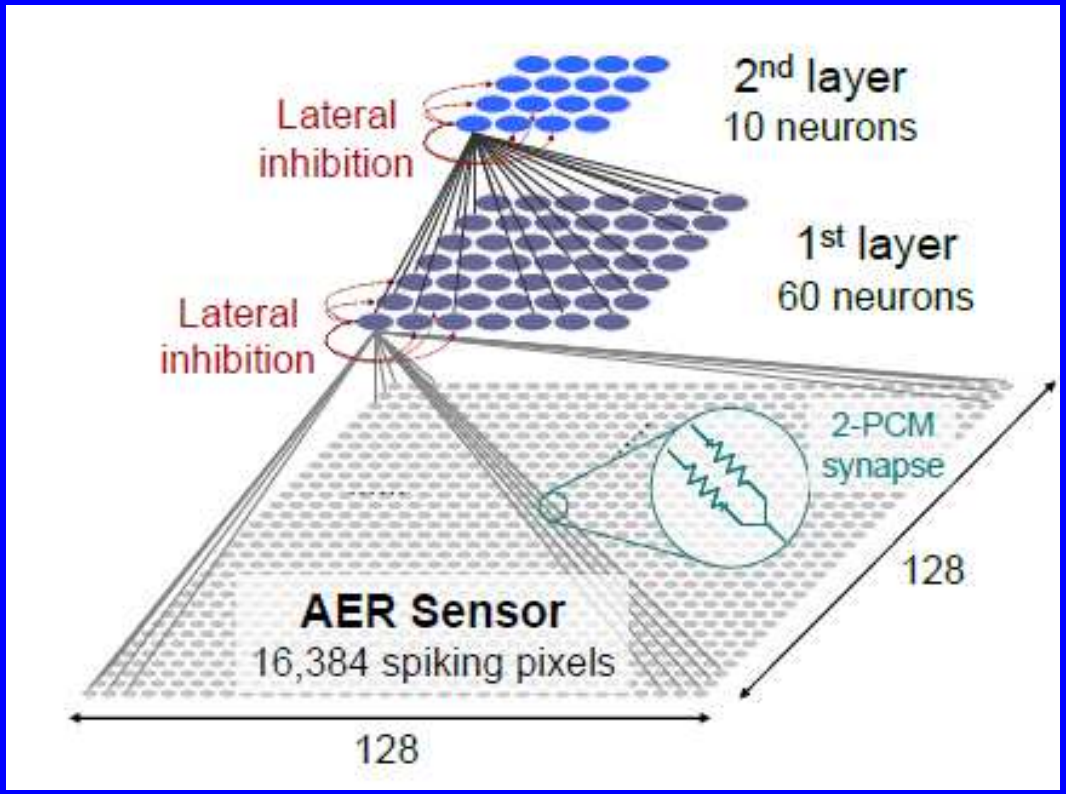


The '2-PCM Synapse'



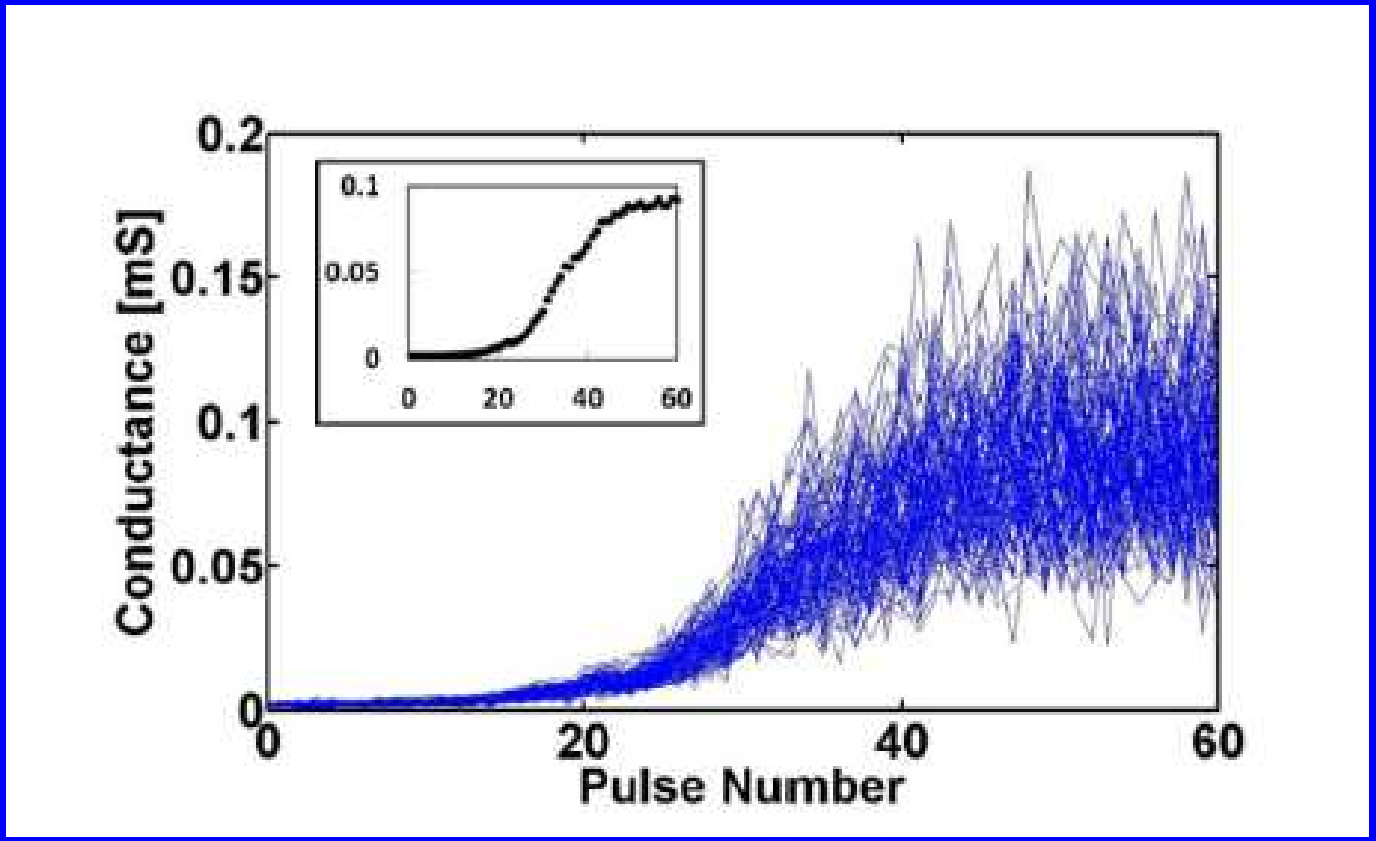
LTP device: Current is Added, LTD device: Current is Subtracted
We use two devices to avoid the problem of abrupt LTD in PCM

Feed Forward Neural Network: Simulation



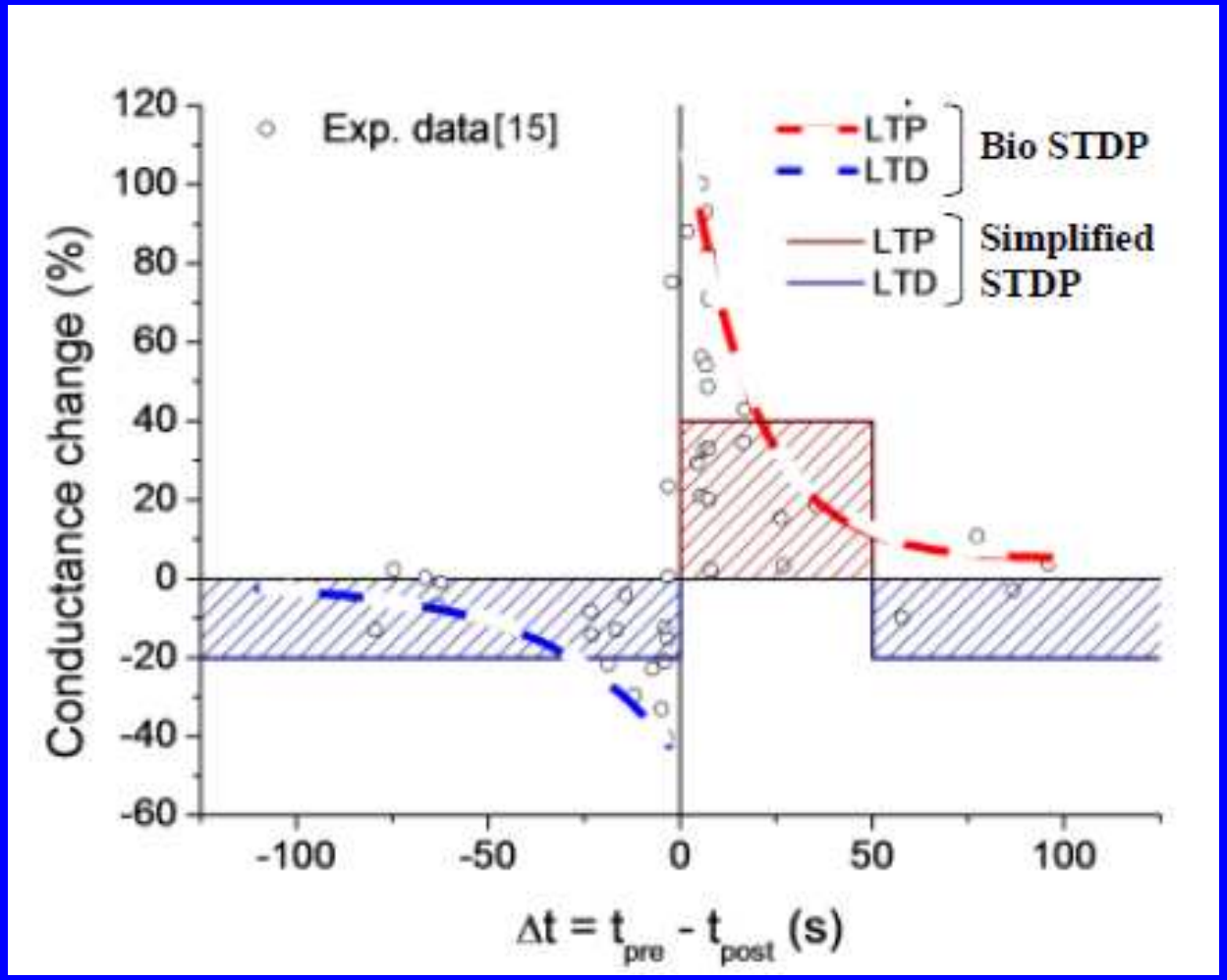
2- Layer Spiking Neural Network
Total Neurons: 70 (60 + 10)
2 Synapses per neuron (Inhibiting/Exhibiting)
Total Synapse ~ 2 million => 4 million PCM Devices

LTP characteristics used in simulations



20% Standard Dev. Dispersion added to model robustness to variability of the neural network

Biological and simplified STDP learning rule

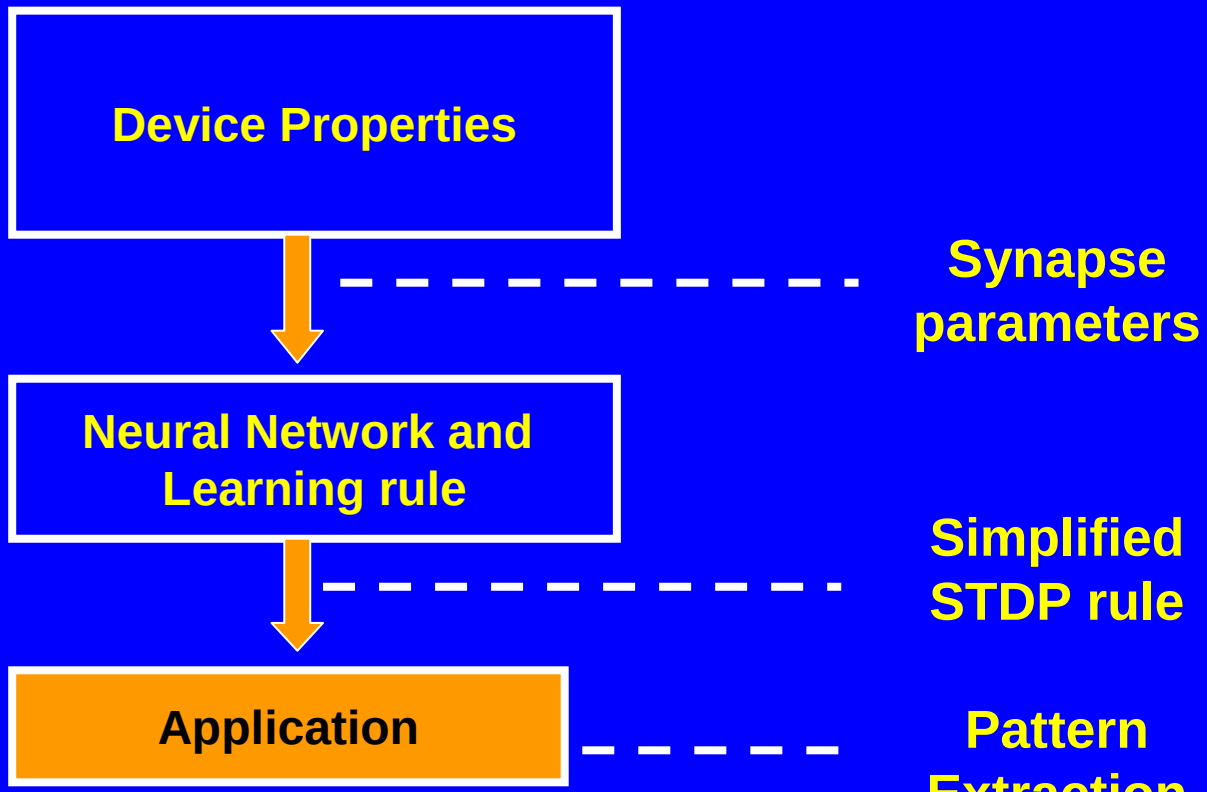


Pre- Neuron spikes before Post- Neuron: Potentiation (LTP)

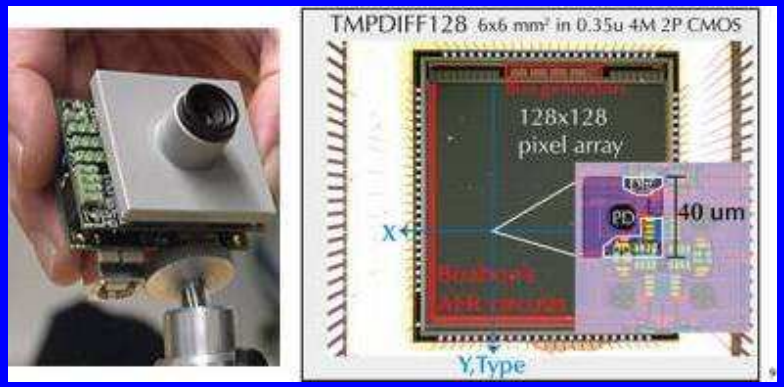
Post- Neuron spikes before Pre- Neuron: Depression (LTD)

% Change in Synaptic Weight is constant
Extended LTD window

Our Approach



Data for our Neural Network

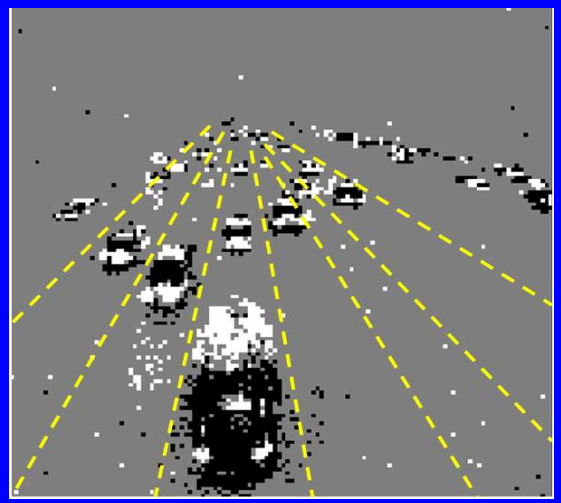


DVS Silicon Retina- Generates AER data

128 x 128 pixels
(Address Event Representation)

P. Lichtsteiner, et al , IEEE J. Solid-State Circuits, Vol. 43, 2008.

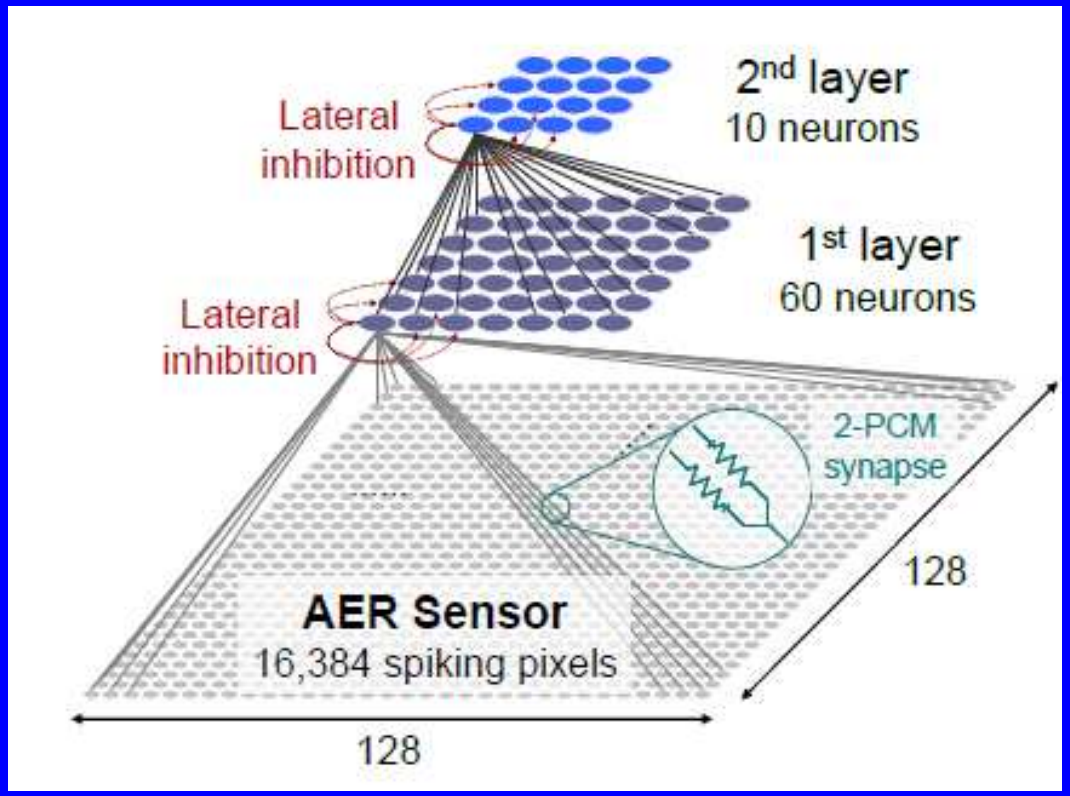
Recorded Data – Cars passing on a Freeway

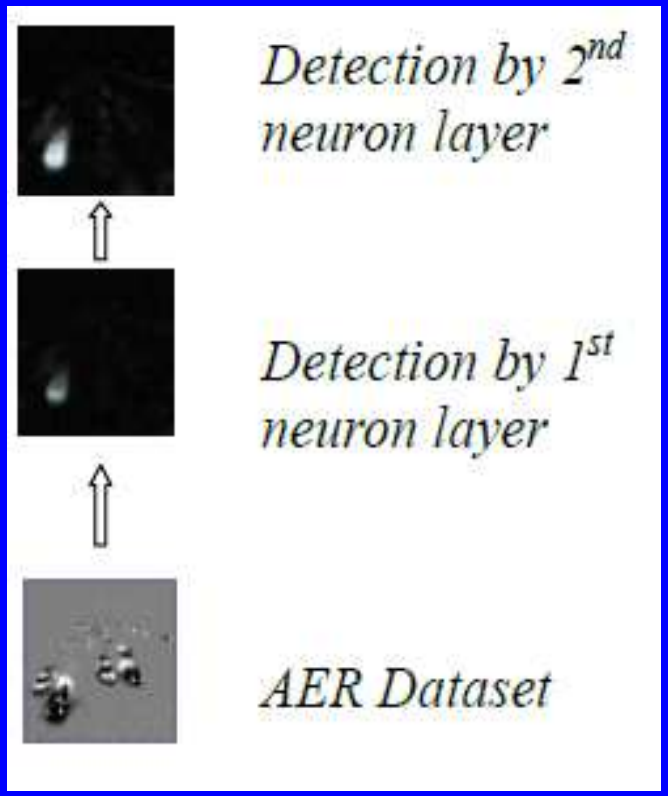


Relative Change in the luminous intensity at each pixel is an event

AER = Pixel Address + Time Stamp + Event

Feed Forward Neural Network: Simulation



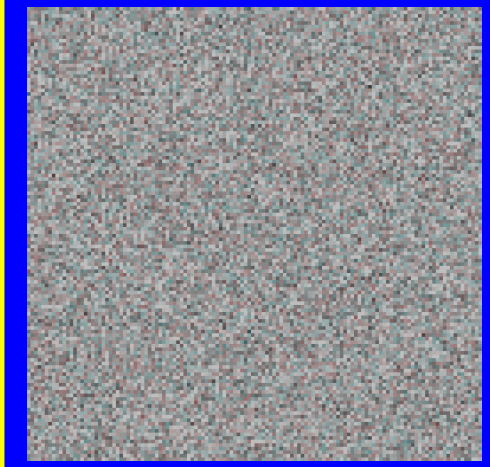


2- Layer Spiking Neural Network
Total Neurons: 70 (60 + 10)
2 Synapses per neuron (Inhibiting/Exhibiting)
Total Synapse ~ 2 million => 4 million PCM Devices

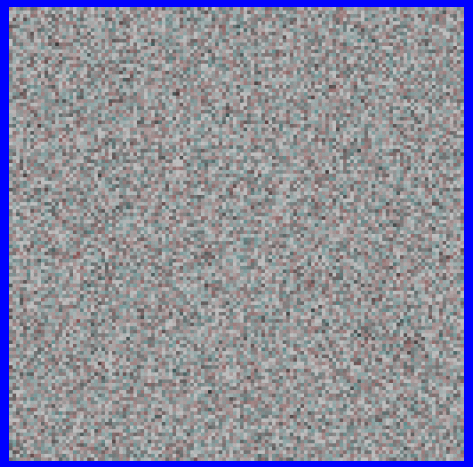
Evolution of Neuron Sensitivity during Learning

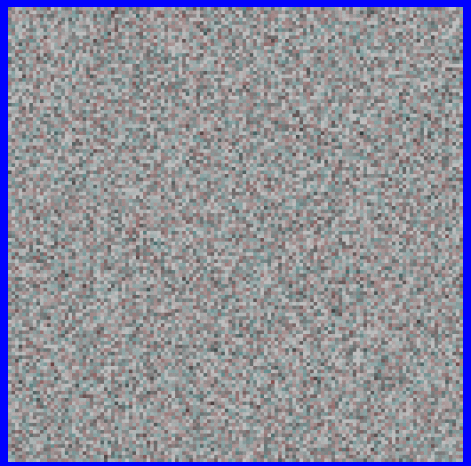
Recorded Stimuli

Neuron - 1

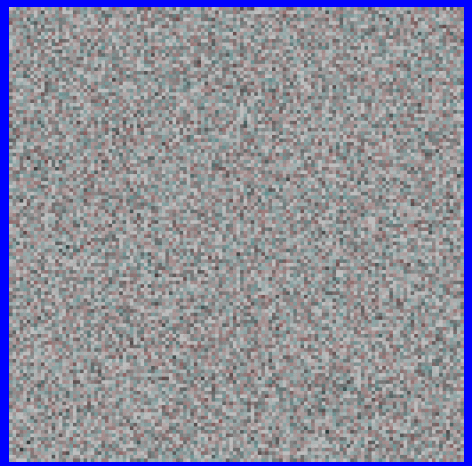


Neuron - 2

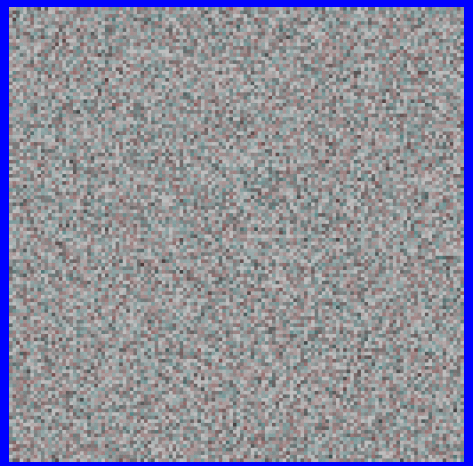




Neuron-3



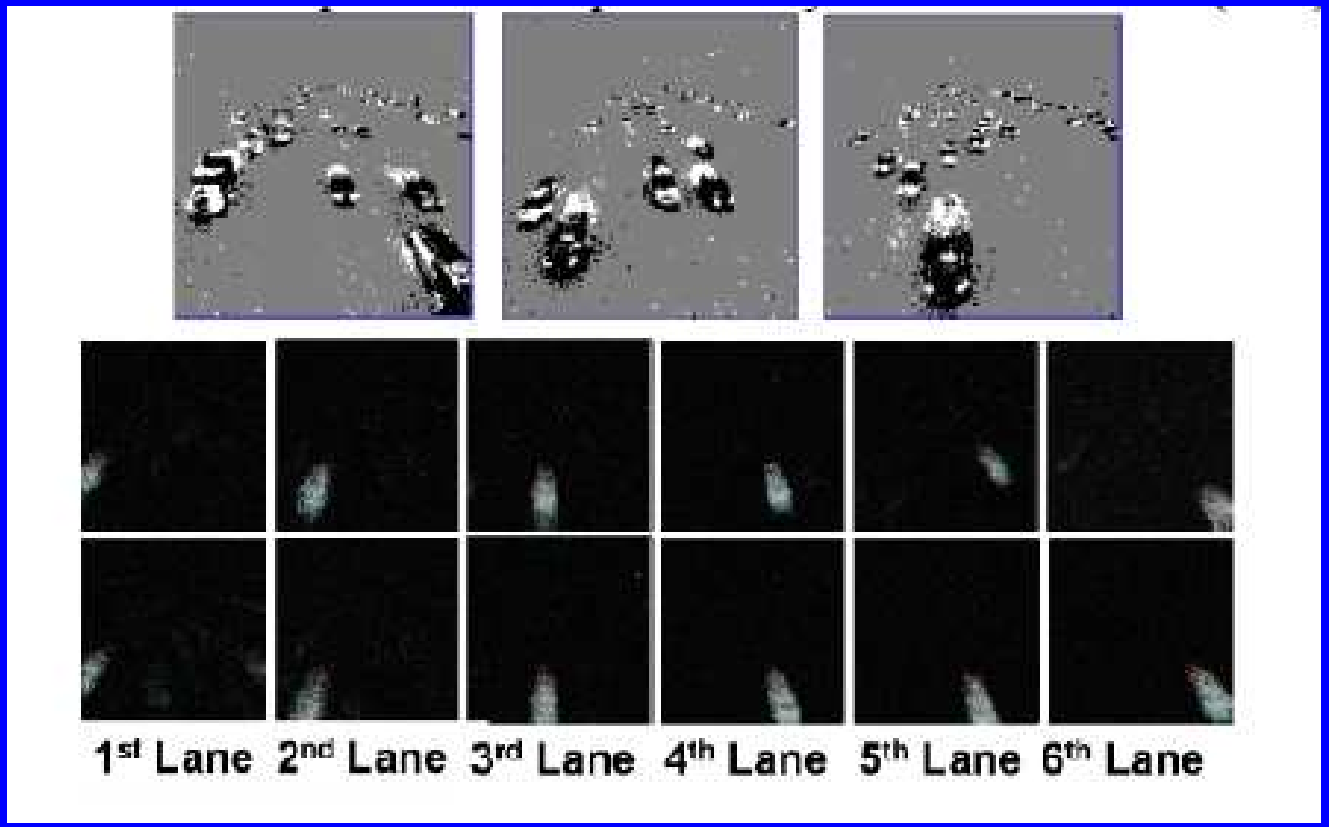
Neuron - 4



Neuron - 5

Learning Results
(Detection of Cars in individual lanes)

Total Learning duration ~ 10 minutes
Avg. Lane Detection Rate > 90 %



Final Sensitivity Map for 10 Neurons

Performance and comparison to GST devices

Pulse Type	/device	/device/sec	Total
GST Devices			
Read	1265	1.9	4.97×10^9
Set	106	0.16	4.16×10^8
Reset	4.2	0.0062	1.65×10^7
Total System Learning Power = 112µW			
GST + HfO2 Devices			
Read	1265	1.9	4.97×10^9
Set	144	0.21	5.64×10^8
Reset	2.6	0.0038	1.00×10^7
Total System Learning Power = 60µW			

Performance and comparison to GST devices

Pulse Type	/device	/device/sec	Total
GST Devices			
Read	1265	1.9	4.97×10^9
Set	106	0.16	4.16×10^8
Reset	4.2	0.0062	1.65×10^7
Total System Learning Power = 112µW			
GST + HfO2 Devices			
Read	1265	1.9	4.97×10^9
Set	144	0.21	5.64×10^8
Reset	2.6	0.0038	1.00×10^7
Total System Learning Power = 60µW			

Conclusions

- We propose the '*2-PCM Synapse*' , a unique programming methodology and a novel feed forward Neural Network based on PCM synapses.

-We show a real world application using our neuromorphic system based on PCM synapses – complex visual pattern extraction

- Addition of HfO_2 interface layer to GST devices increases the system power/energy efficiency by-

a)Decreasing the Iset and Ireset
b)Increasing the number of intermediate resistance states in the LTP curve

-Total synaptic programming power is as low as 60μW with interface layer PCM synapses.

Questions??

Thank you for your Attention!!

(manan.suri@cea.fr)