Neuro-Inspired Processor Design for On-Chip Learning and Classification with CMOS and Resistive Synapses

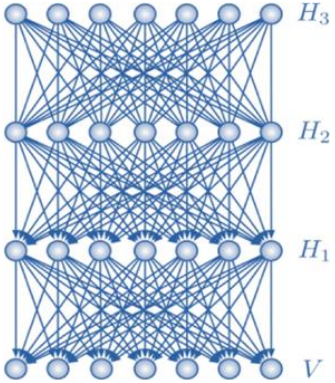
Jae-sun Seo

School of ECEE, Arizona State University

The 13th Korea-U.S. Forum on Nanotechnology

September 26, 2016

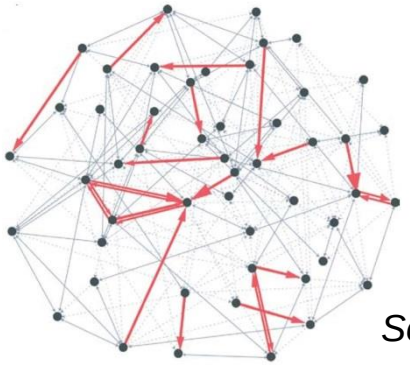
ML Literature (DNN)



Courtesy: Nuance

- **Dense connectivity**
- **Learning done offline**
  - **Back-propagation (requires labeled data)**
- **MNIST 99.79%, ImageNet 95%**
- **What about unlabeled data or customization?**
  - **Full computation on each layer → high power**

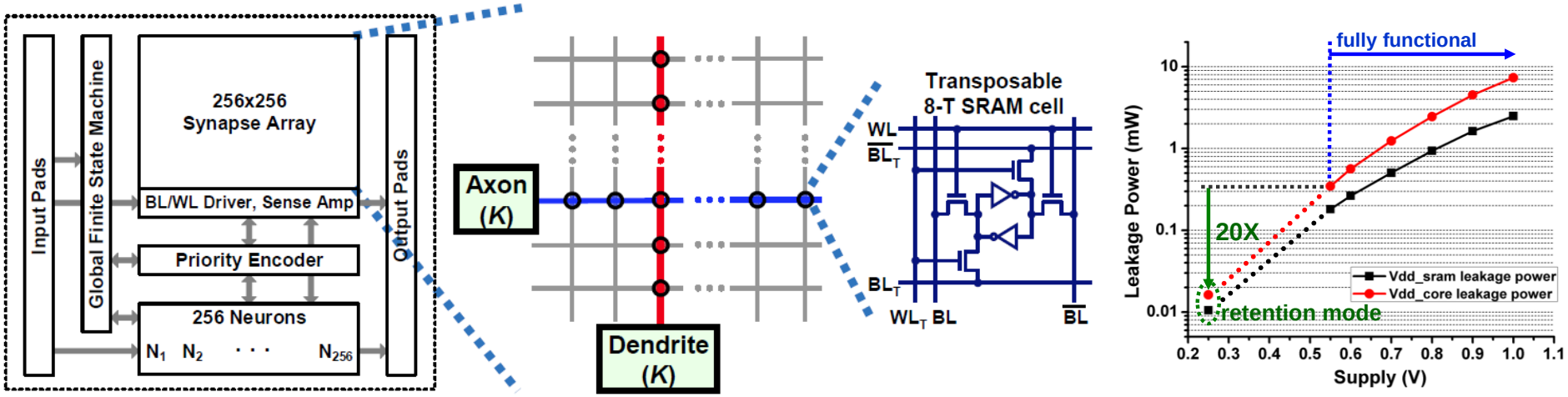
Neuromorphic (SNN)

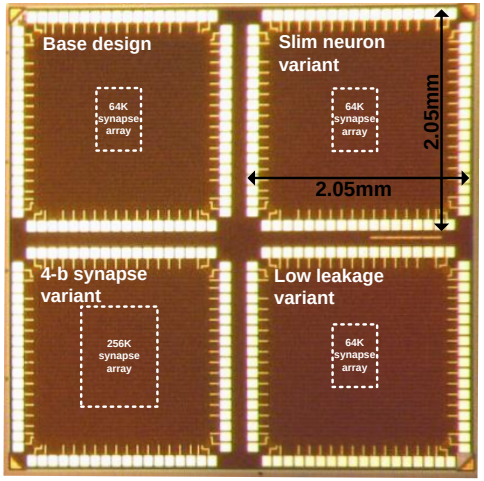


Song, PLoS Biol. 2005

- **Sparse connectivity**
- **Online learning**
  - **STDP, SRDP, Reward (biological evidence)**
- **MNIST 99.08%, ImageNet N/A**
- **Cont. learning & detection**
  - **Adaptable for input change**
  - **Sparse spiking, attention → low power**

Neuromorphic Core with On-Chip STDP



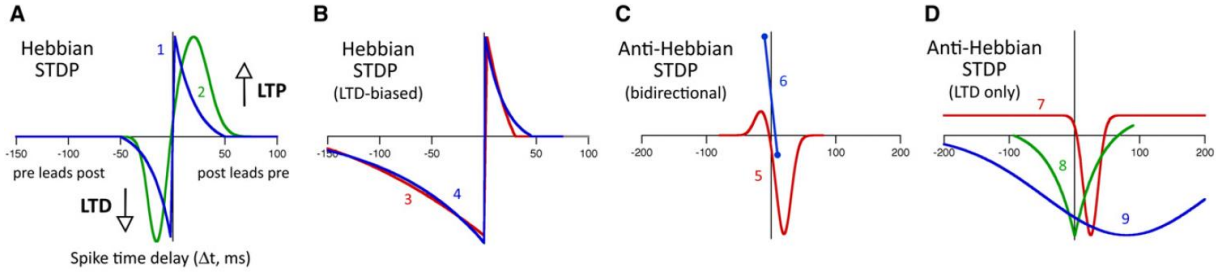


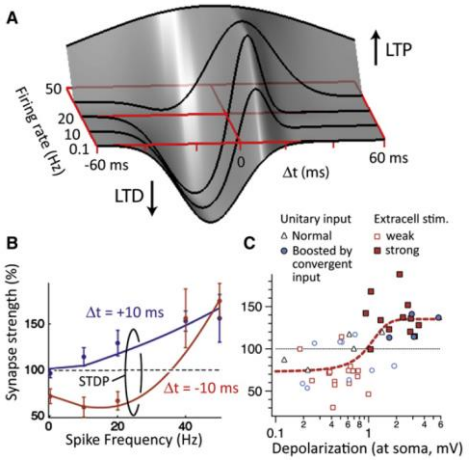
- Under STDP learning, when neuron K spikes, all synapses on row K and column K may update
- Transposable SRAM: single-cycle read & write in both row and col. directions
- Efficient pre- and post-synaptic update
- Near threshold operation
- Pattern recognition

Seo, CICC, 2011

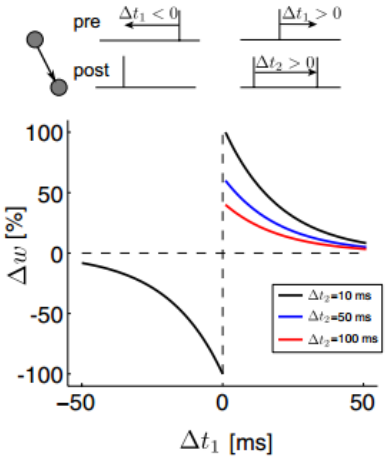
Versatile Learning in Neuromorphic Core

Various STDP Learning Rules
(Feldman, Neuron 2012)

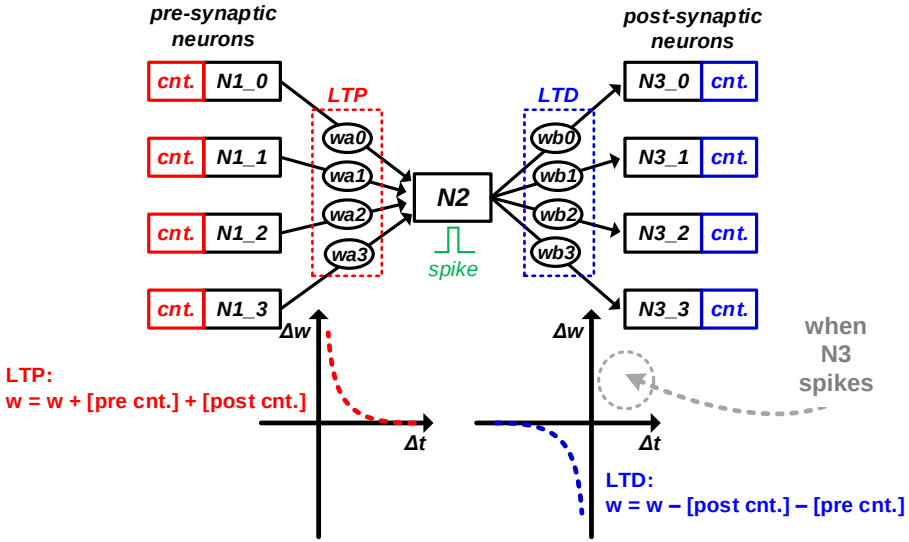




Multi-factor

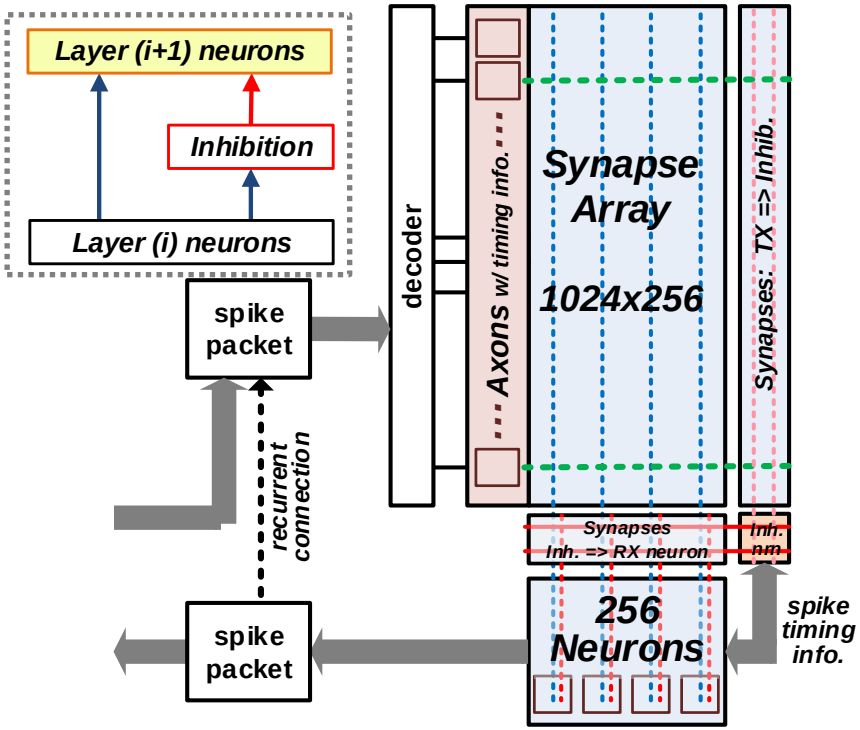


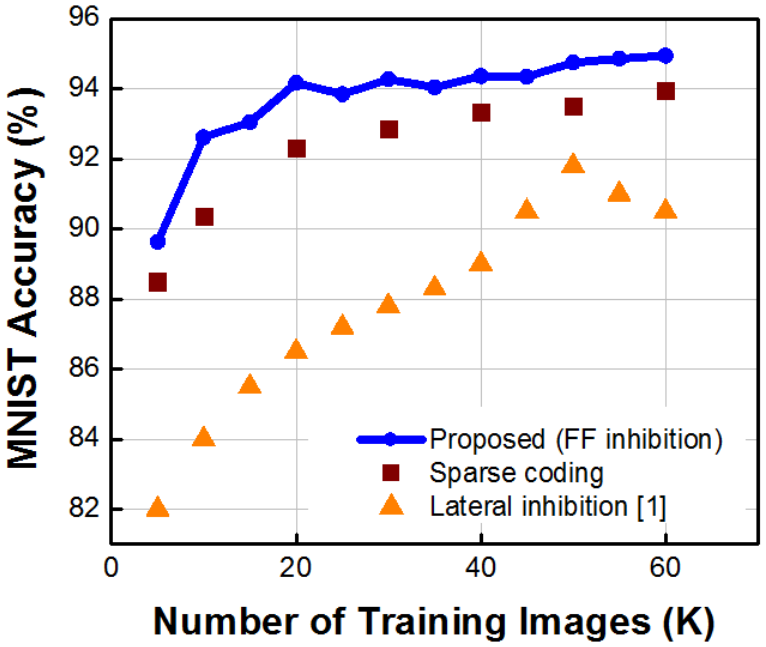
Triplet-STDP



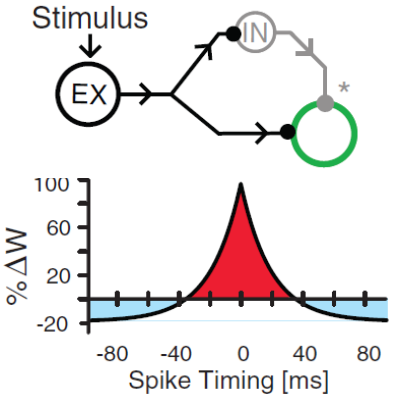
- **A versatile neurosynaptic core to support various learning rules, large fan-in/-out, sparse connectivity**
- **Triplet STDP (*Pfister, J. of Neuroscience, 2006, Gjorgjieva, PNAS 2011*)**
  - **post-pre-post: post nrn. spike & pre nrn. timing & post nrn. timing**
  - **pre-post-pre: pre nrn. spike & post nrn. timing & pre nrn. Timing**

Feedforward Excitation & Inhibition





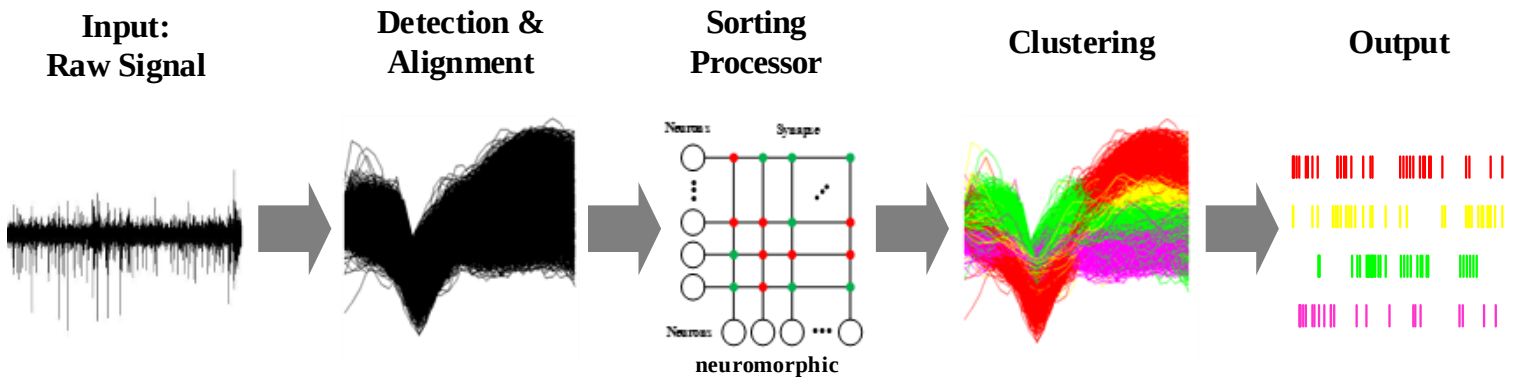
[1] Diehl, Front. of Neuroscience, 2015



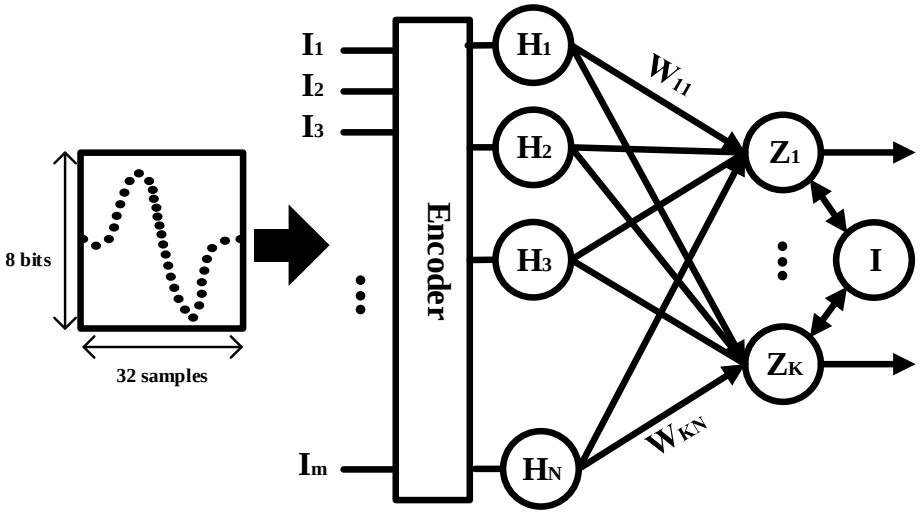
Vogels, Science, 2011

- **Joint feed-forward excitation and inhibition**
- **For a small number of inhibitory neurons, add pre=>inh, inh=>post synapses**
- **Balance excitatory & inhibitory synaptic inputs**

Neural Spike Sorting Processor (for deep brain sensing & stimulation)

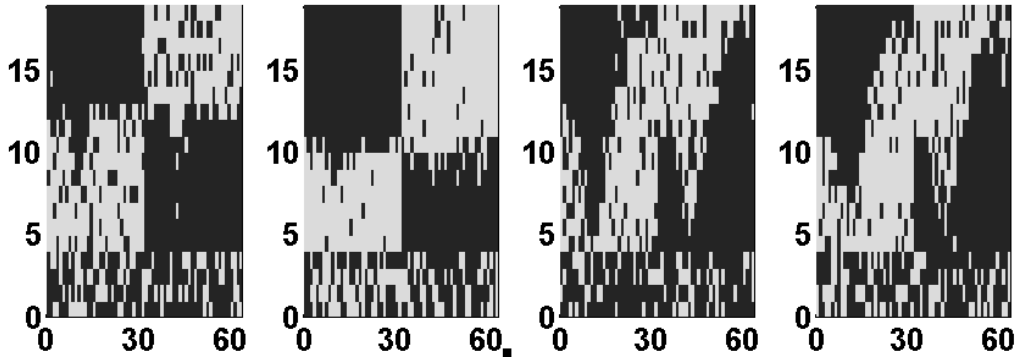


- Signals from invasive electrodes: spikes from multiple neurons
- Online, unsupervised neuromorphic spike-sorting processor
  Collaboration with Columbia University (ISLPED 2015)

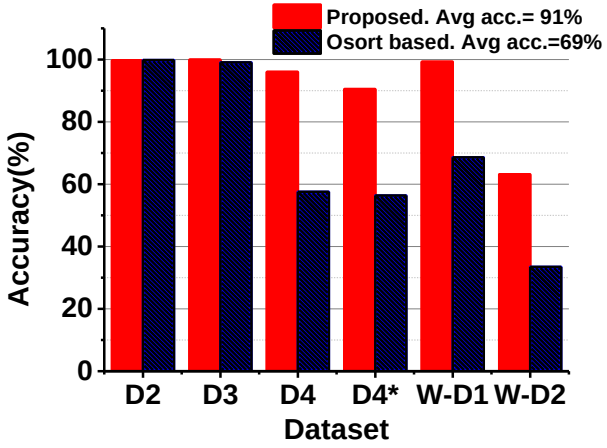


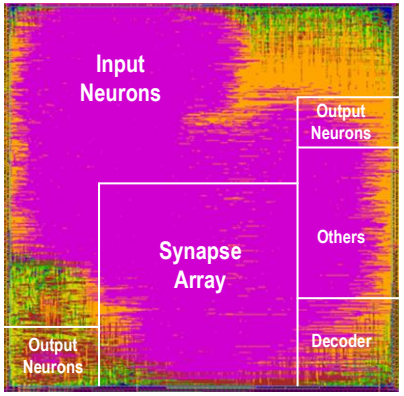
- Weight update through STDP
- Start with K=2, automatically increases # of output neurons if the spike difference is large enough (self-organized map)

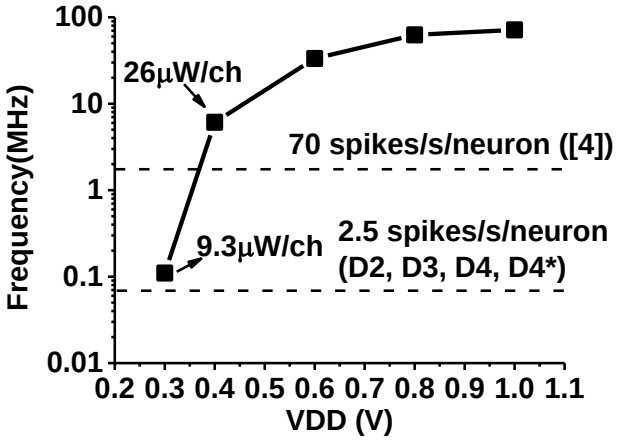
Exp. Results: Clustering Accuracy



Receptive field of dataset that contains 4 clusters in 3000 spikes





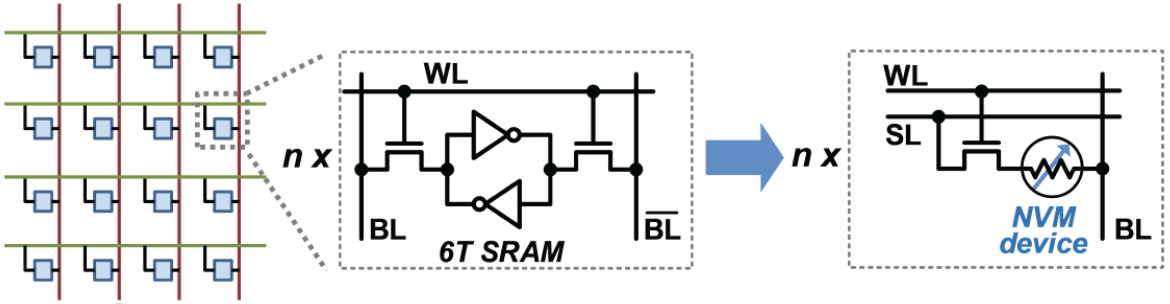


Spike sorting accuracy more reliable than other low-complexity algorithms such as O-sort

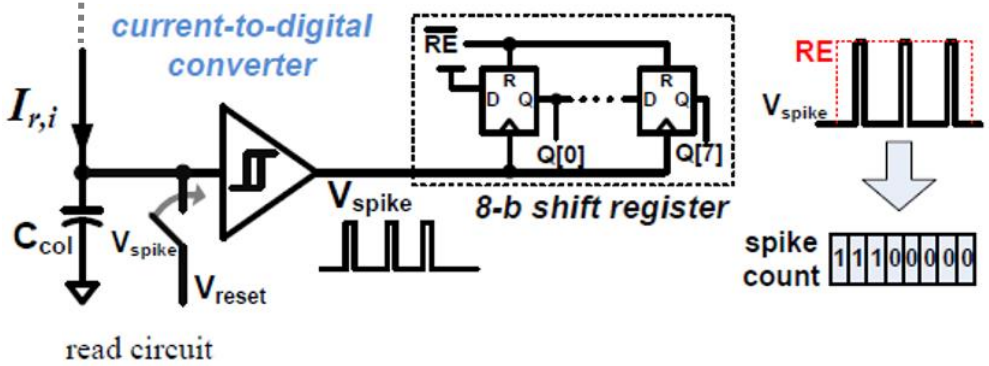
Avg. accuracy: 91% vs. 69%

- **65nm GP, high-Vth, 0.5x0.5mm^2**
- **9.3µW/ch at 0.3V**
- **Layout of the design is dominated by memory elements, as well as power.**

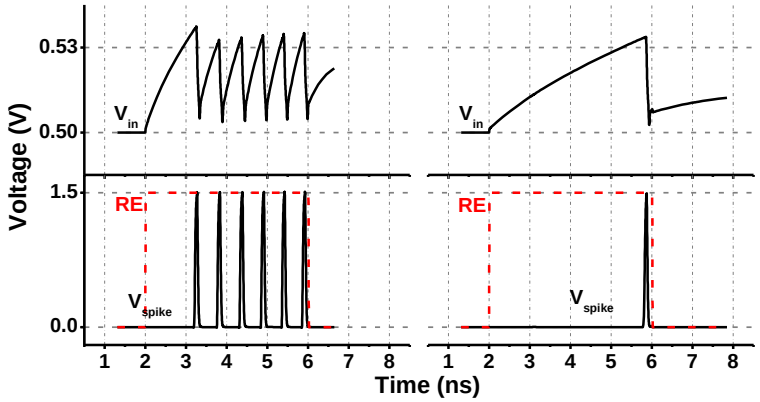
Neuromorphic Computing w/ NVMs



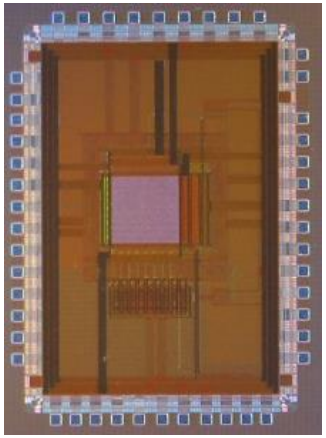
- Emerging NVMs (e.g. RRAM) could alleviate power/area bottleneck of conv. memories



- Read rows in parallel: weighted sum current
- Peripheral CMOS read: current-to-digital converter



Simulation results for 4ns read timing window



130nm
RRAM array +
CMOS read circuits
(under testing)

Summary

- **Neuromorphic computing hardware**
  - **45nm testchip with on-chip STDP learning**
  - **Versatile learning neuromorphic core & architecture**
  - **65nm spike clustering processor**
  - **Emerging NVM arrays + peripheral read/write circuits**

- **Future research with circuit-device-architecture co-design and optimization**

Collaborators

- **ASU**
  - **Faculty: Yu Cao, Shimeng Yu, Chaitali Chakrabarti, Sarma Vrudhula, Visar Berisha**
  - **Students: Minkyu Kim, Deepak Kadetotad, Shihui Yin, Abinash Mohanty, Yufei Ma**
- **Intel: Gregory Chen, Ram Krishnamurthy**
- **Columbia University: Mingoo Seok, Qi Wang**