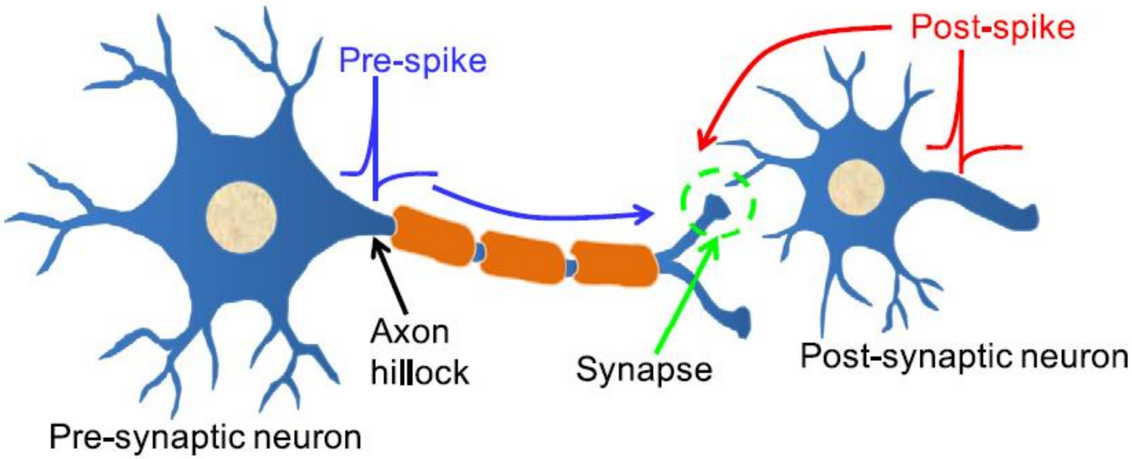
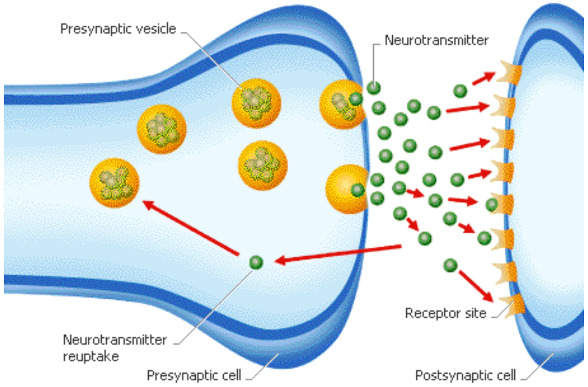
L8: Neuromorphic Learning

Instructor: Prof. Feng Xiong

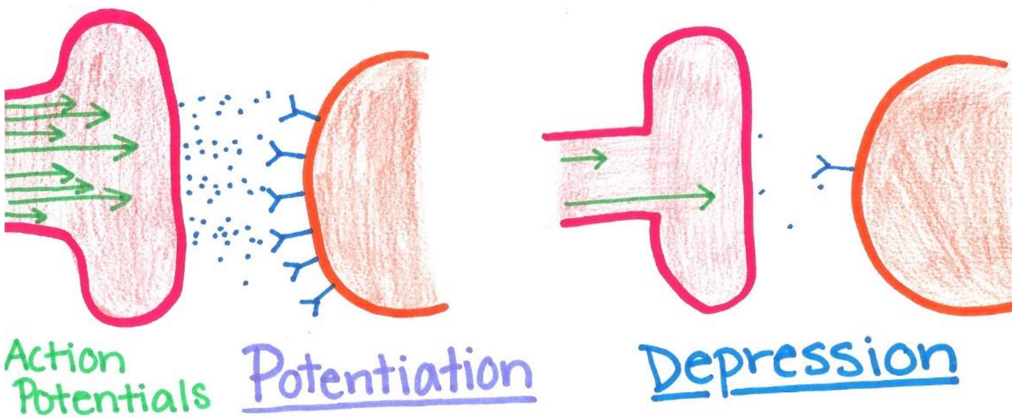
Recap



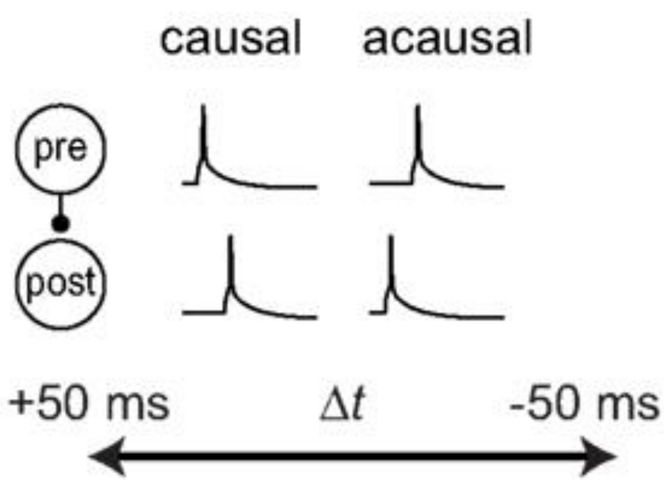
Pre-spike
Post-spike
Axon
hillock
Synapse
Post-synaptic neuron
Pre-synaptic neuron



Presynaptic vesicle
Neurotransmitter
Receptor site
Neurotransmitter
reuptake
Presynaptic cell
Postsynaptic cell



Action
Potentials
Potentiation
Depression



causal
acausal
pre
post
+50 ms
Δt
-50 ms

Outline

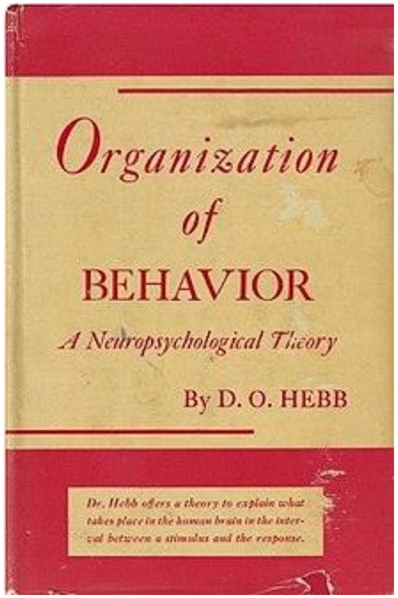
- Spike-timing dependent plasticity (STDP)
- Additive and Multiplicative learning
- Associative learning
- Synaptic electronics
  - requirements
  - current approach and limitations
  - promising candidates

Hebbian Learning

- Neurons that fire together, wire together!
- First introduced by Donald Hebb in 1949 in his book “The Organization of Behavior”

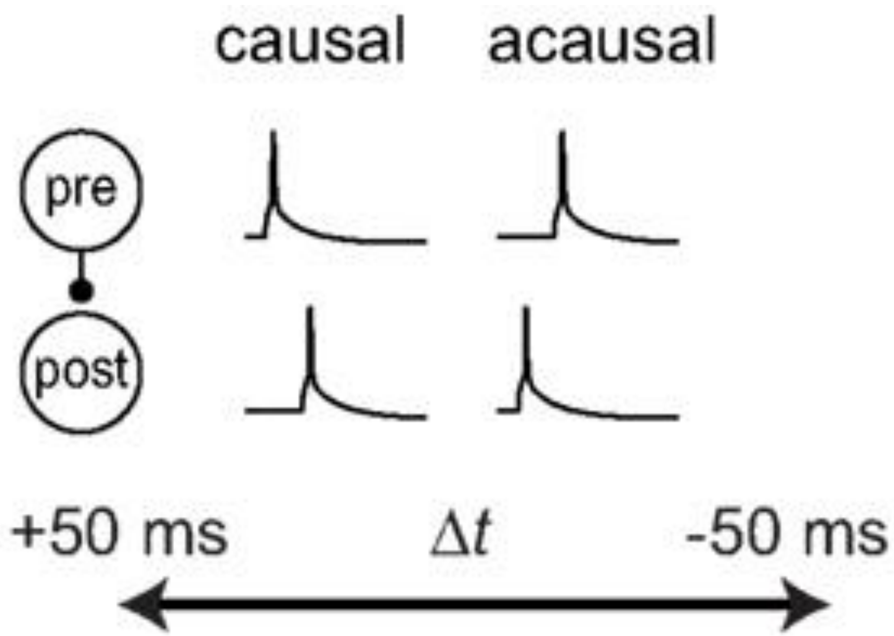
- Principle: any two cells or systems of cells that are repeatedly active around the same time tend to become associated, so that activity in one facilitates activity in the other

Donald Hebb
Canadian Psychologist



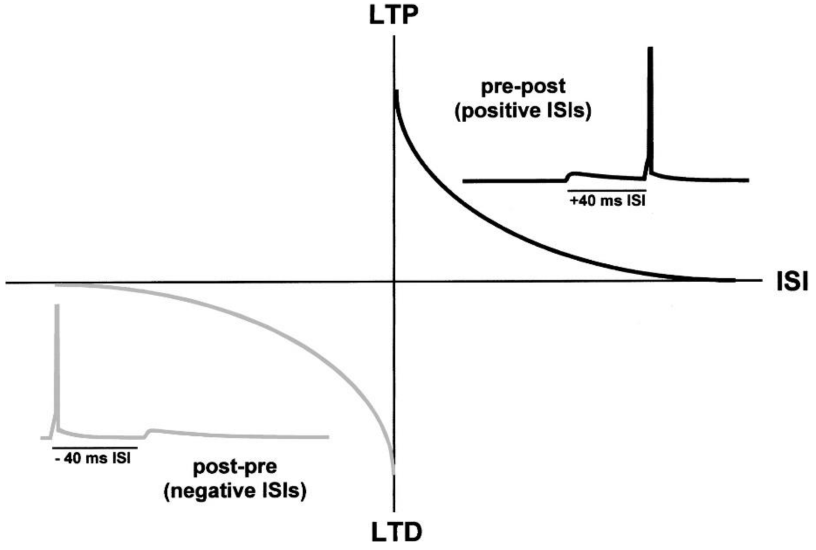
Spike Timing Dependent Plasticity (STDP)

- Hebbian learning: connection strength between neurons are modified based on neural activities during learning
- STDP
  - – spike timing dependent plasticity
  - – focus on **temporal order** of spikes in cellular learning
  - – plasticity depends on relative timing of pre- and post-synaptic spikes

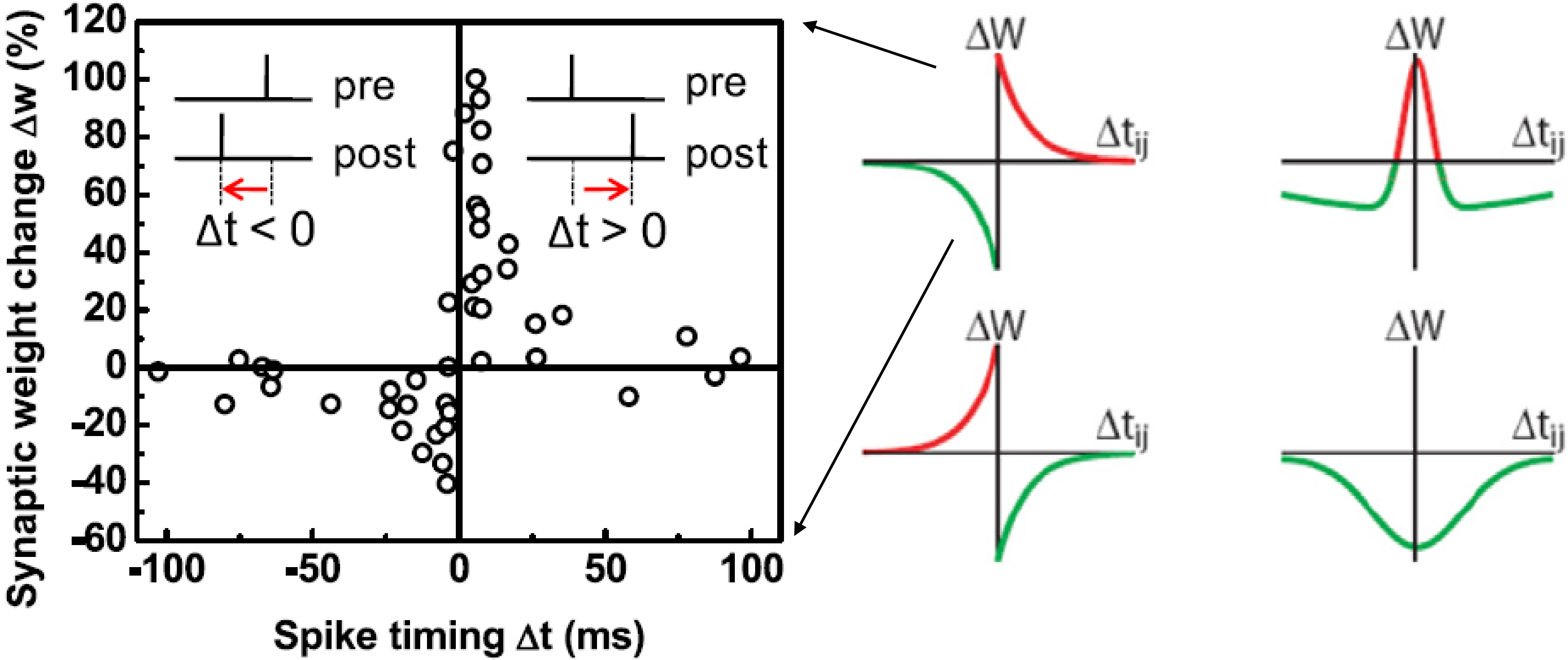


STDP Timing Window

- Hebbian rule: *those who fire together, wire together*
- Firing needs to be related, therefore close in window →
- Causal: pre-synaptic neuron fire just before the post-synaptic neuron; suggesting prediction and thus potentiation
- Acausal: pre-synaptic neuron firing after the post-synaptic neuron; suggesting unrelated and thus depression



Different Forms of STDP



- Change in ΔW is bigger if Δt is smaller → implies urgency

Other Plasticity

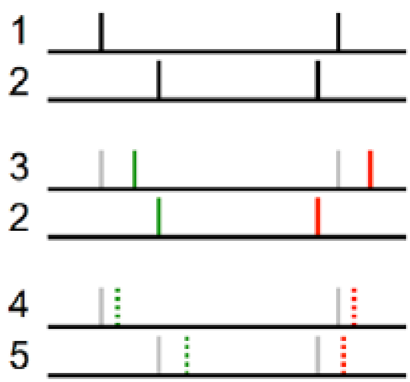
- STDP:
  - simple
  - biologically plausible
  - computationally powerful
  - +100 to -100 ms window
  - weight change +100% to -50% in biological STDP

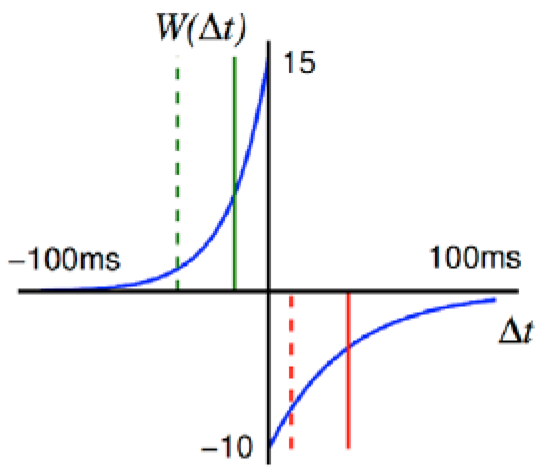
- STDP is NOT the only form of synaptic plasticity

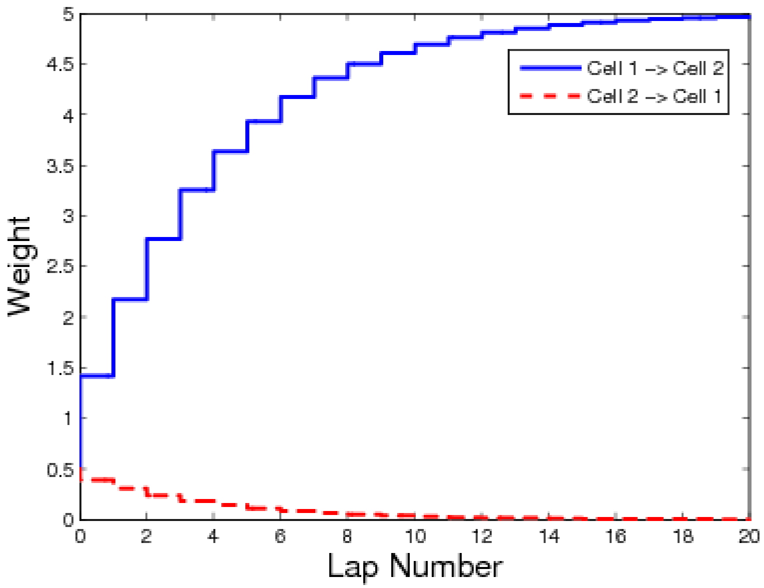
- Other forms include:
  - firing rate
  - spiking orders
  - dendritic locations and etc

Additive and Multiplicative Learning

- STDP is an asymmetric form of Hebbian learning
- Two types of STDP: additive and multiplicative
- Additive: learning only depends on ΔT; not on the actual weight
- Multiplicative: learning is a function of both ΔT and weight







Short Term and Long Term Learning

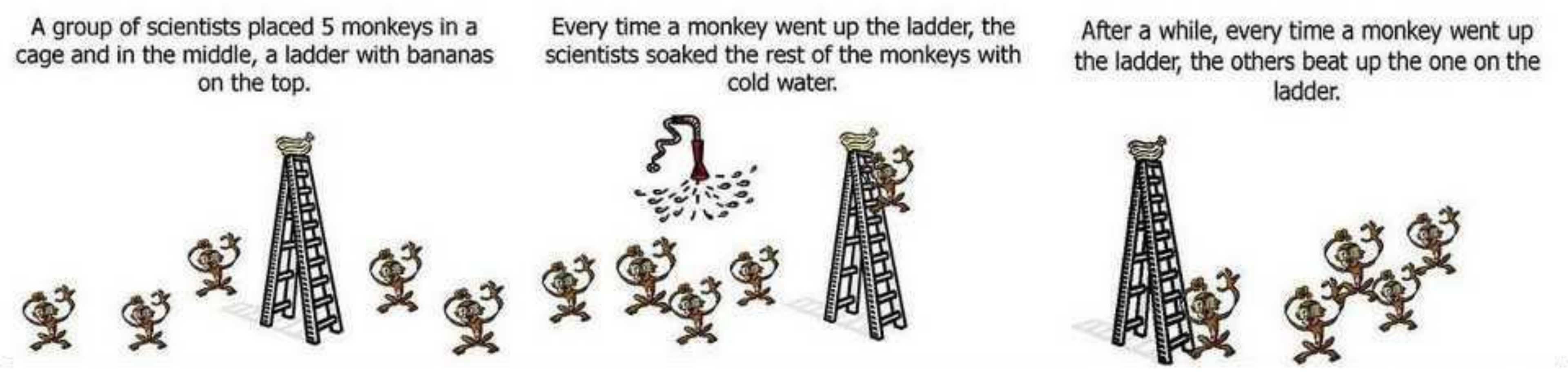
- Short term learning
  - temporary: synaptic strength returns to initial level when pre-synaptic activity stops
- Long term plasticity
  - durable and persistent; activity dependent
- Additive learning, more short term
- Multiplicative learning, more long term learning

Associative Learning

- Associative learning goal: complete an incomplete representation from a previously learnt pattern
- Network learns a pattern by strengthening certain synapses between neurons
- With incomplete pattern, synapses can still recruit the missing neurons to recall the original pattern

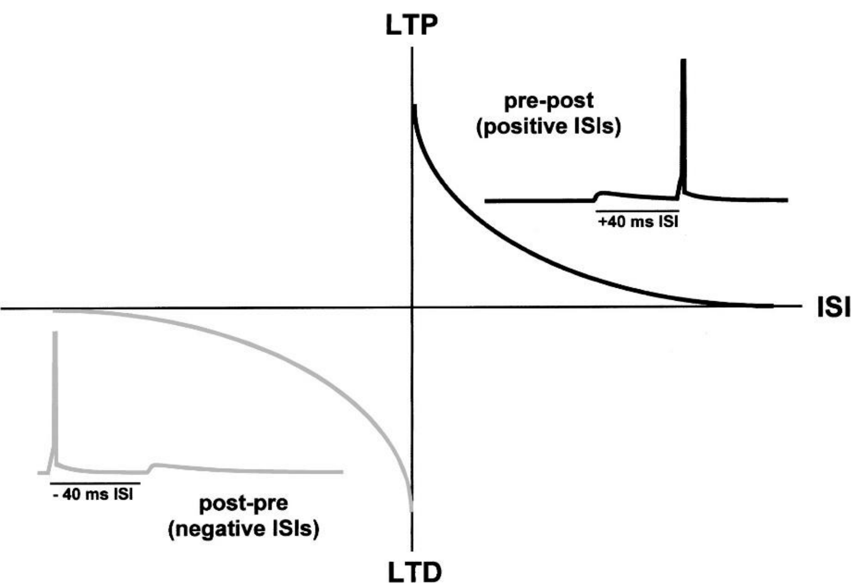
Sequential Learning

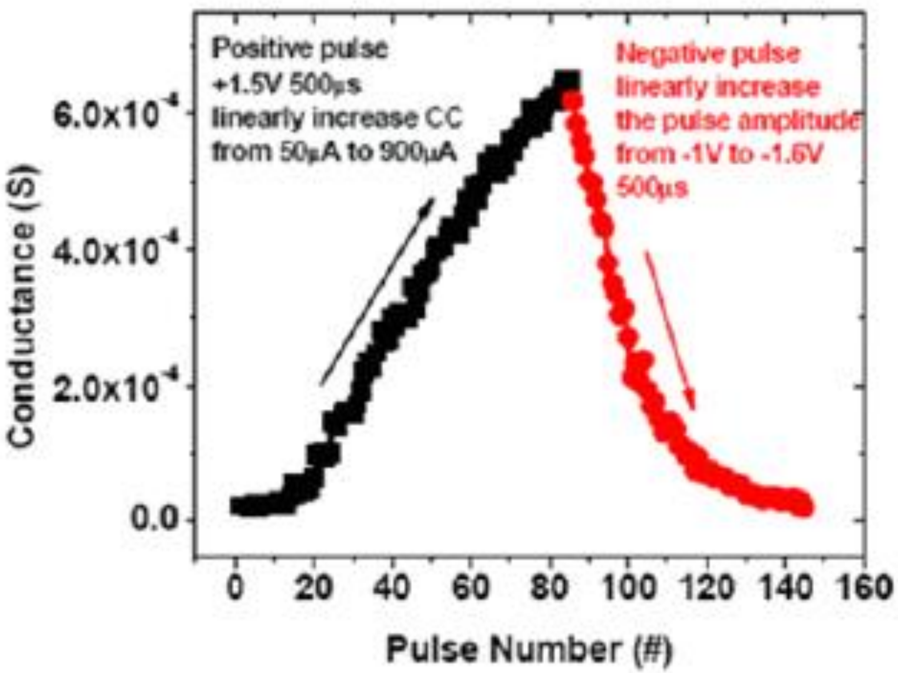
- Sequential learning goal: retrieve the next member of a sequence of patterns
- Asymmetric STDP → temporal sequences of the events are encoded in synaptic weight
- Predicting future events based on previous experience



Synaptic Electronics

- Synapse-like electronics to mimic neural network
- non-volatile to store information
- Analog, not digital





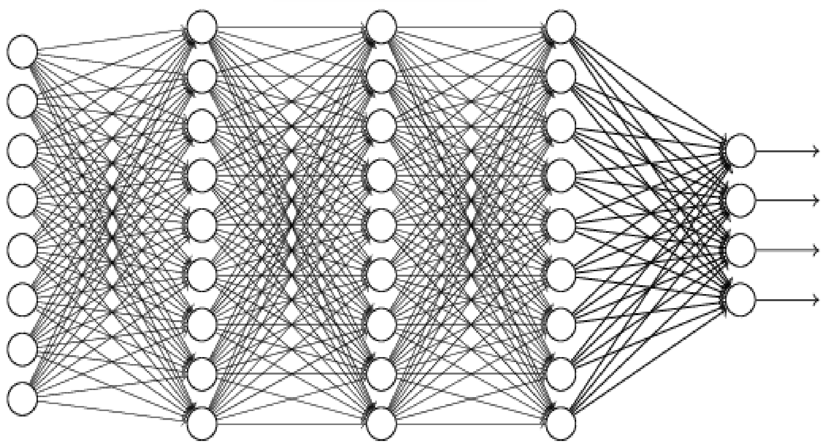
Deep vs. Spiking Neural Network

Deep Neural Network (DNN)

Input

hidden layer

Output



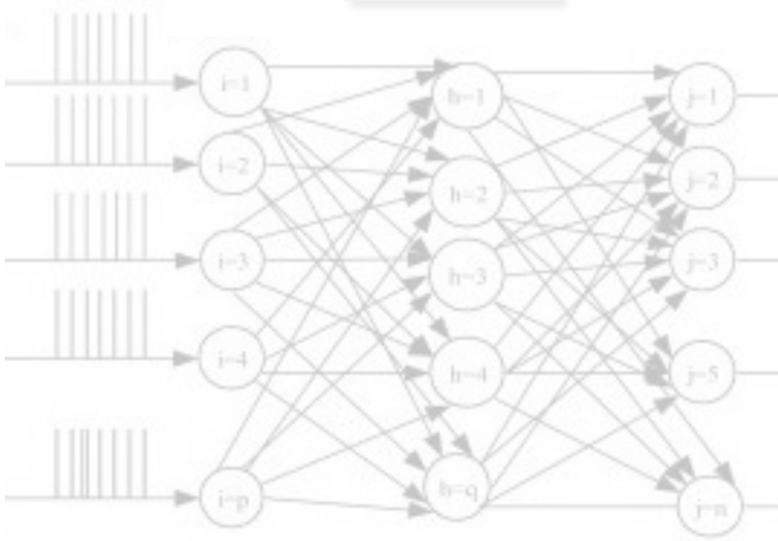
- Synaptic weight only
- Data-driven
- “Black box”

Spiking Neural Network (SNN)

Input

hidden layer

Output



- Weight + Temporal dynamics
- Event-based
- Potential for cognitive computing

M. Nielsen, “Neural Networks and Deep Learning”, 2015

R.V. Florian, “Reinforcement Learning in SNNs”, 2012

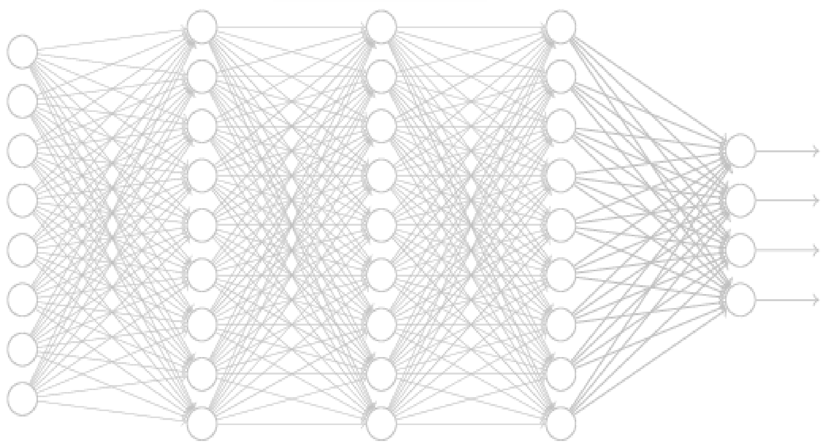
Deep vs. Spiking Neural Network

Deep Neural Network (DNN)

Input

hidden layer

Output



- Synaptic weight
- Data-driven
- "Black box"

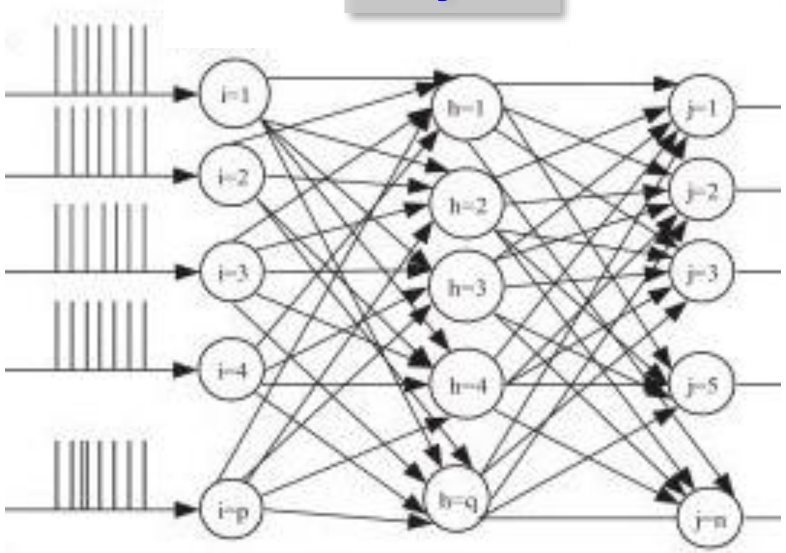
M. Nielsen, "Neural Networks and Deep Learning", 2015

Spiking Neural Network (SNN)

Input

hidden layer

Output



- Weight + Temporal dynamics
- Potential for cognitive computing
- Event-based

R.V. Florian, "Reinforcement Learning in SNNs", 2012

Synaptic Device Requirements

- Synaptic device = Memory?

- Similarities
  - data retention
  - Energy and speed
  - Reliability: variations, endurance, stability
  - scalability

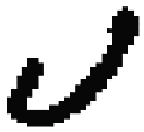
- Differences
  - High-precision
  - High linearity and symmetry for DNNs
  - Temporal dynamics (both short-term and long-term) for SNNs

Fault and Variation Tolerant

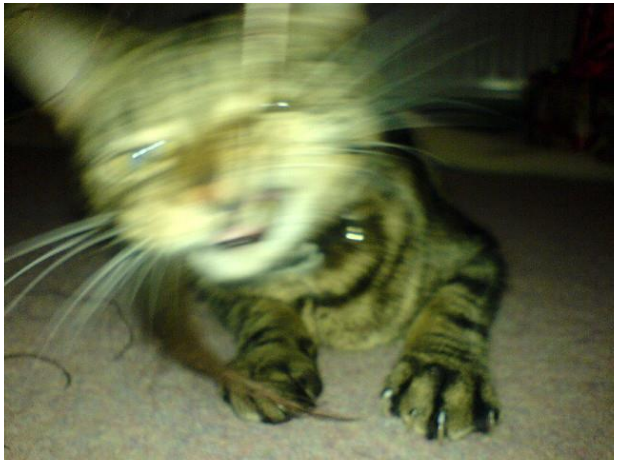
- More fault tolerant compared to memory

- Variation resistant
  - e.g. fuzzy pattern recognition
  - Able to recognize with partial input

"Ha" or "tta"

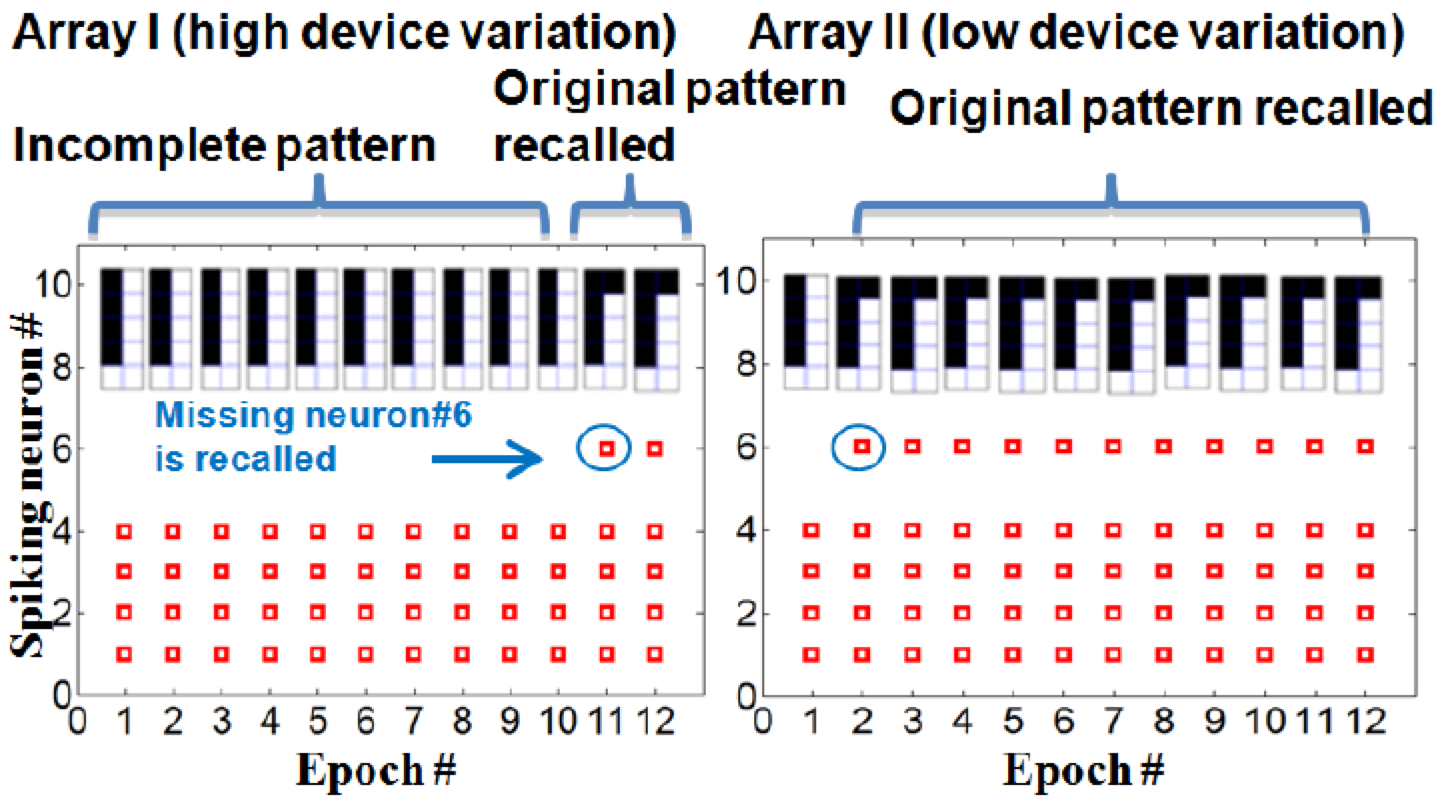


"J" or "U"



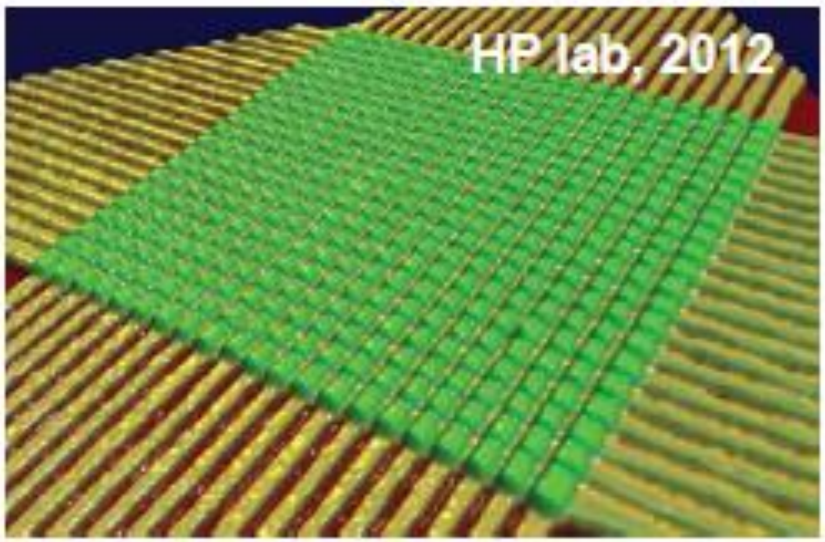
Variation Tolerance

- High variation array
  - – Requiring a higher threshold current to reach acceptable accuracy
  - – More training epochs required
  - – Consuming more energy

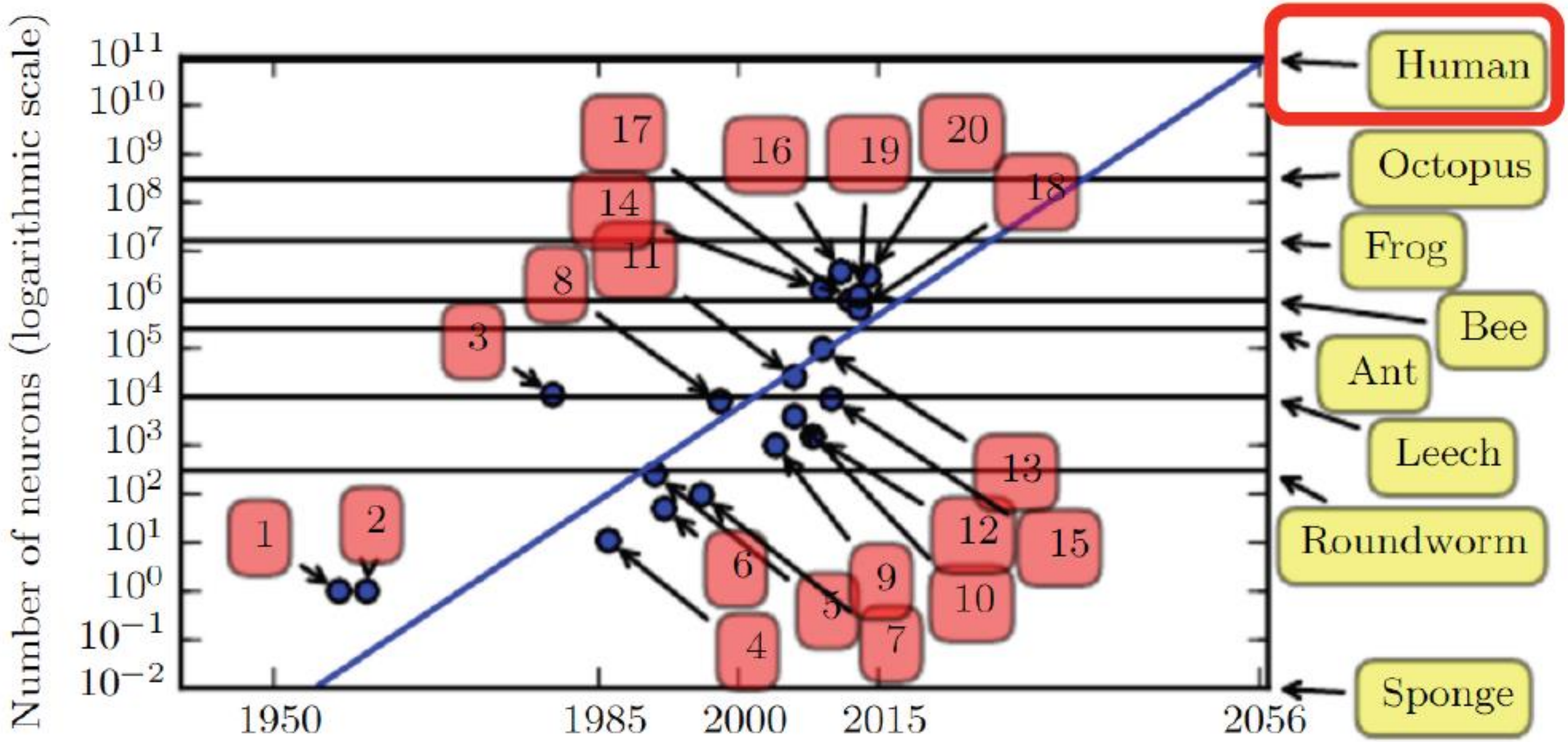


Scaling up

- Compact: preferably 2-terminal
- Massive parallelism in arrays
- Crossbar bar vs. neural network



Growth Trend in Neural Network



Source: I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, 2016

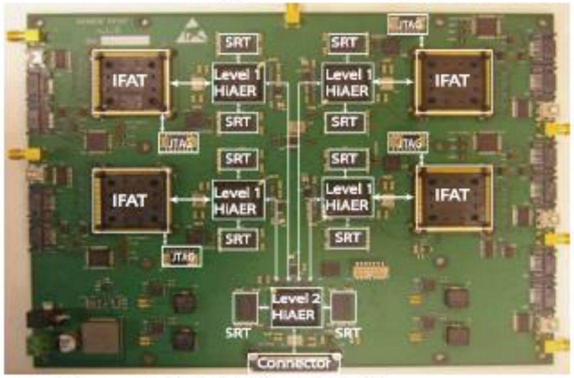
Energy Efficiency

	Application	Hardware used	Estimated power consumption
Large scale	Emulating 4.5% of human brain: 10^{13} **synapses**, 10^9 neurons [7]	Blue Gene/P: 36,864 nodes, 147,456 cores	**2.9 MW** (LINPACK)[45]
	Deep sparse autoencoder: 10^9 **synapses**, 10M images [8]	1,000 CPUs (16,000 cores)	**~100 kW** (cores only)

- Human Brain
- 10^{11} neuron and 10^{15} synapse
- Power ~10-20 W

Current Large Scale Architecture

DRAM (off-chip)



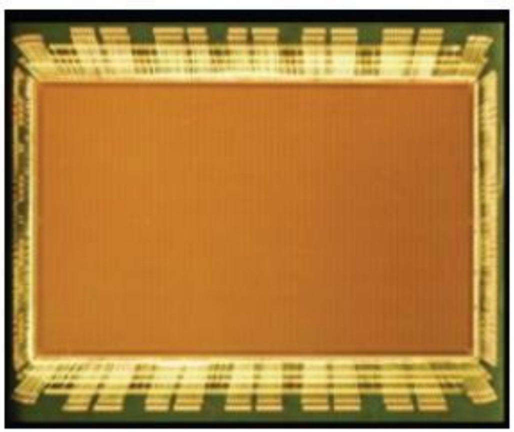
IFAT+HiAER (UCSD)

DRAM (off-chip)

Neurogrid (Stanford)

- Energetically expensive
  - Refresh
  - Off-chip access

SRAM (on-chip)



TrueNorth (IBM)

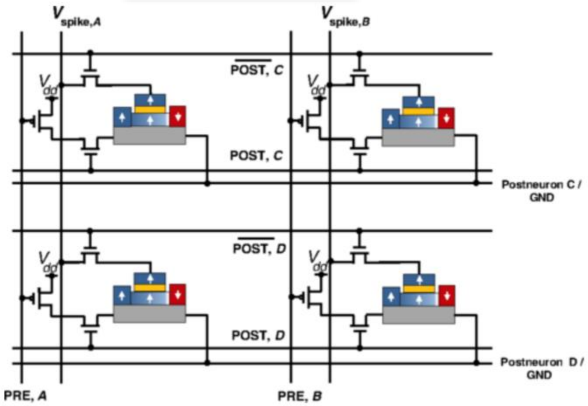
- Area inefficient

SRAM (on-chip)+ **DRAM**(off-chip)

SpiNNaker (U of Manchester)

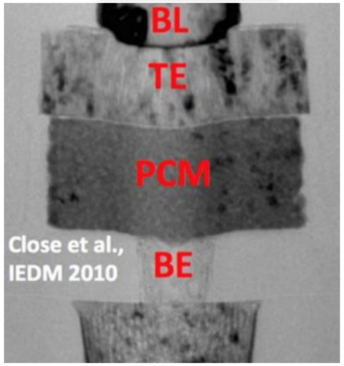
Synaptic Device – Current Approaches

CMOS



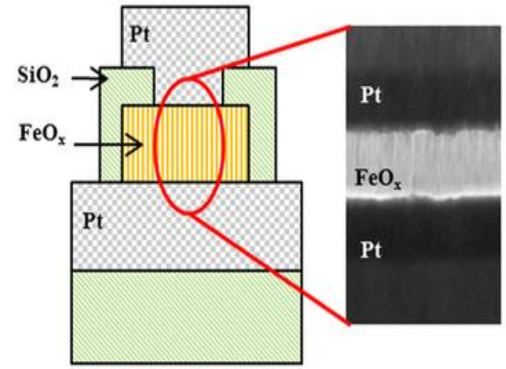
A. Sengupta, Phys. Rev. Applied 6, 2016

PCM



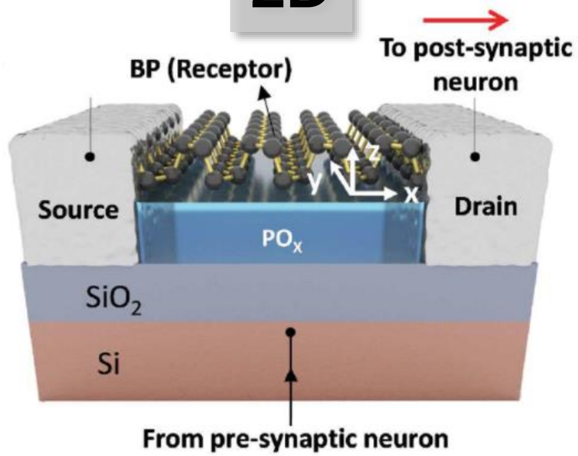
H.-S. Philip Wong et al., Front. Neurosci., 2014

RRAM



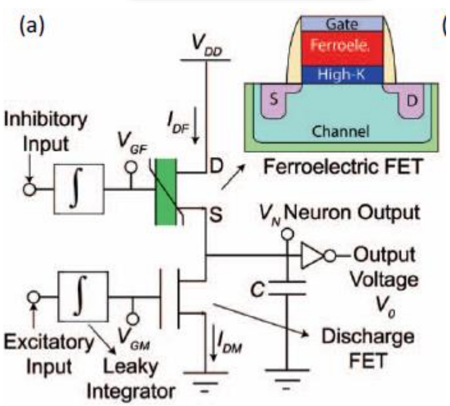
W. He et al., Sci. Rep., 4, 2014

2D



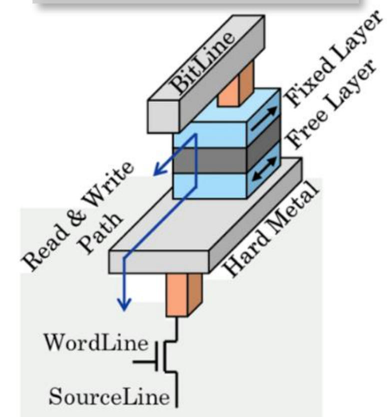
H. Wang et al., Adv. Mater., 28, 2016

FEFET



Z. Wang et al., IEDM, 300, 2018

STT-RAM



S. Mittal et al., J. Low Power Elec.2017