

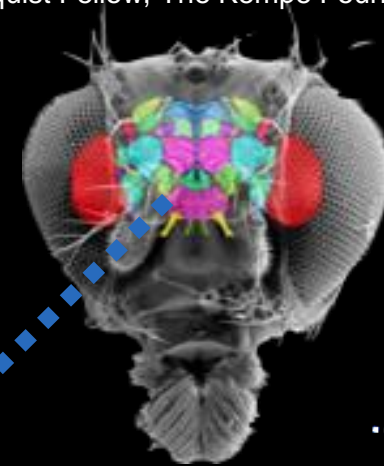
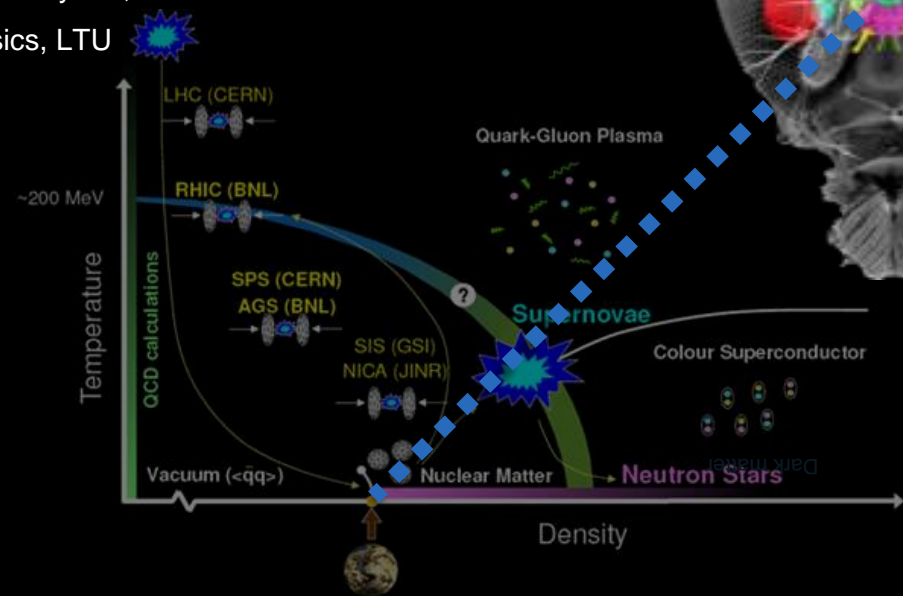
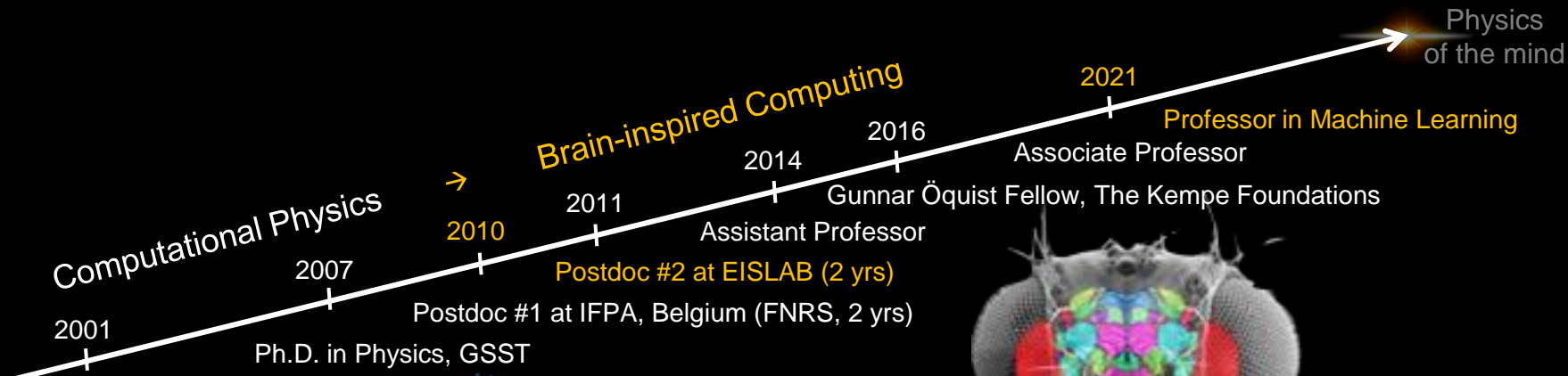
Temporal Dynamics in Neuromorphic Computing

A New Frontier for Co-Design and Optimization in Physics and ML

Fredrik Sandin

LULEÅ
TEKNISKA
UNIVERSITETEN

LULEÅ
UNIVERSITY
OF TECHNOLOGY



Machine Learning Group



Marcus



Pedro



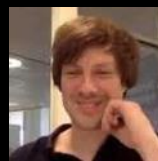
Konstantina



Gustav



Fotini



Christian



Kanjar



Vibha



Fredrik



Priyamvada



György



Saleha



Rajkumar



Oluwatosin



Homam



Mattias



Daniel



Nosheen



Sana



Ali



András



Richa



Karl



Carl



Prakash



Lama



Elisa



Notice something?
Almost 40%
woman



WASP

Machine Learning for the welfare of society

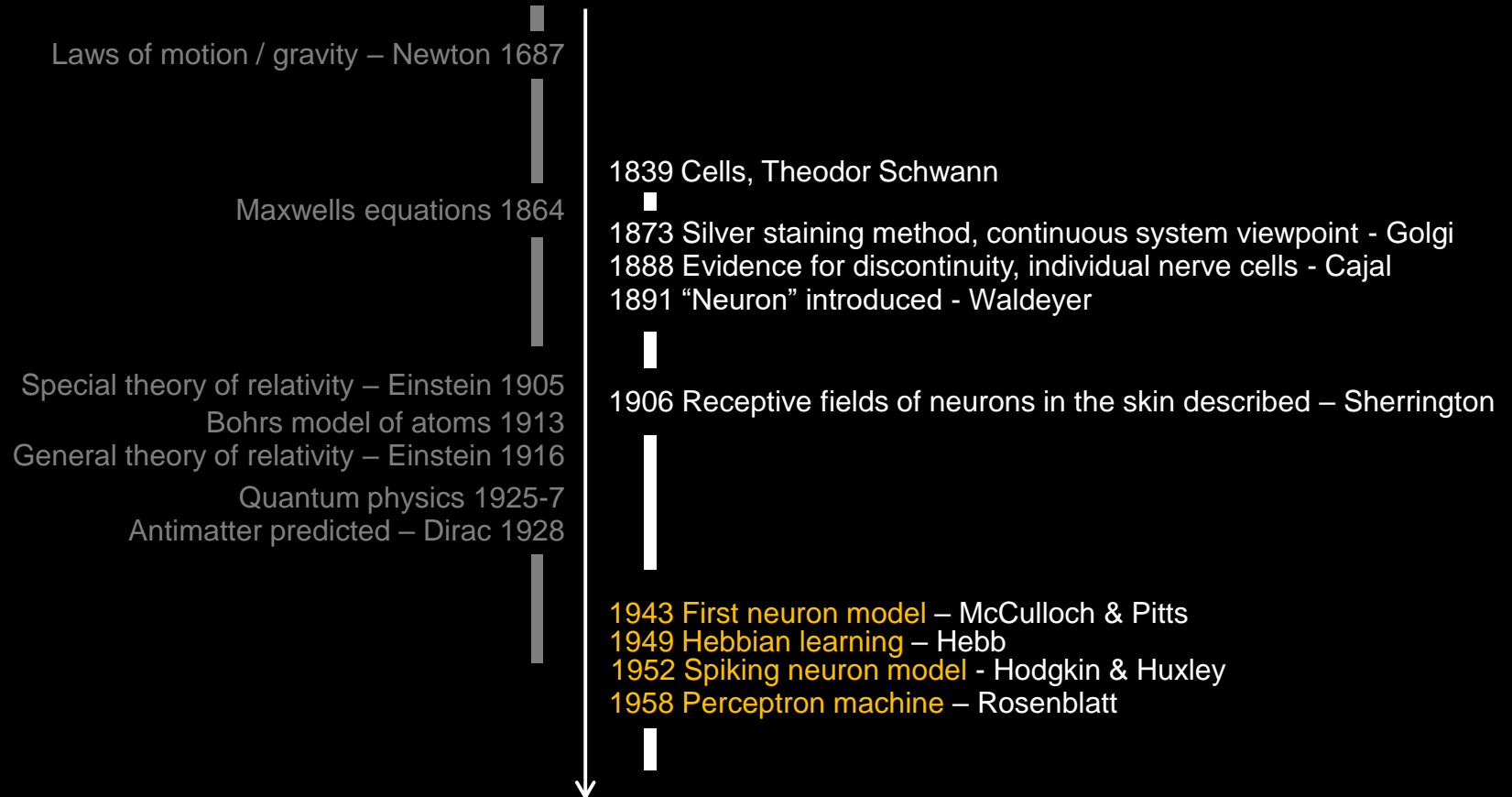
Neuromorphic Computing

Spiking Neural Networks (SNN)

Examples, work in progress

History of Physics

... and Neuroscience



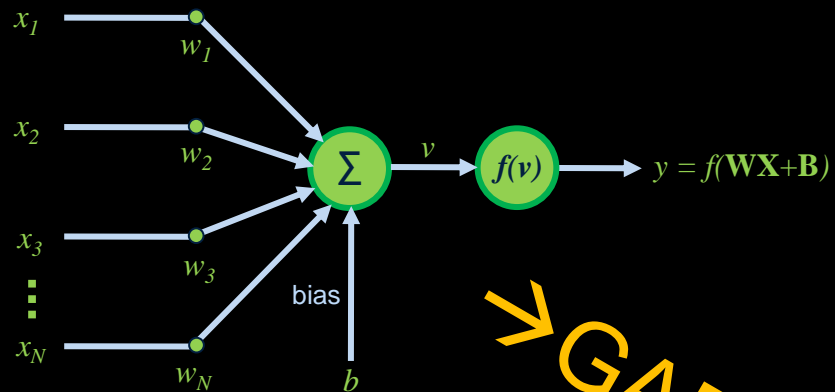
Neuromorphic engineering aims to create computing hardware that mimics biological nervous systems



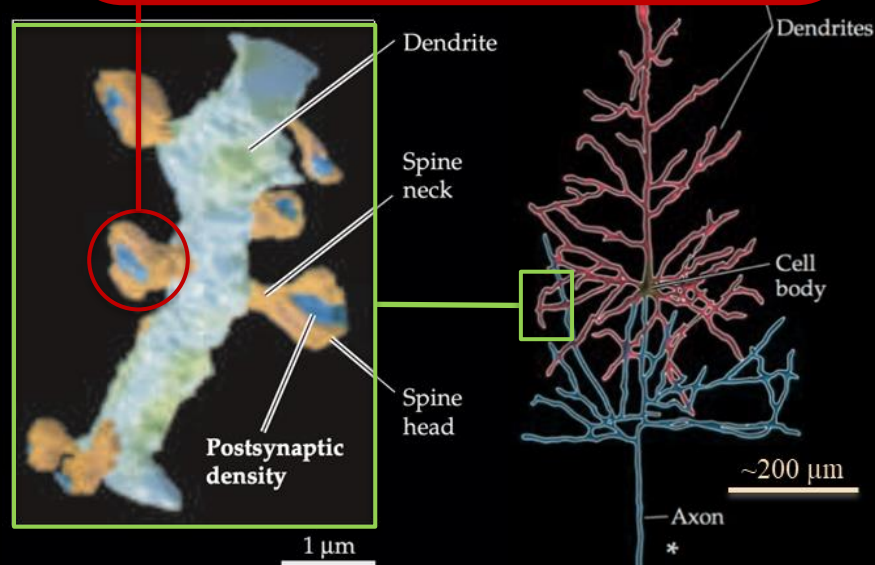
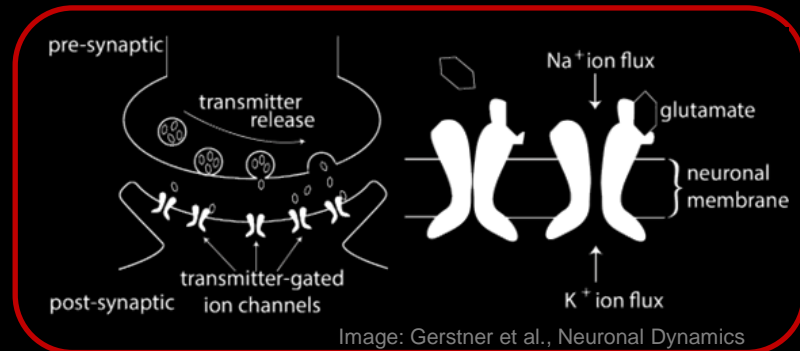
Image credit: Rodney Douglas

It had become clear to me that industrial practice was on a problematic path: in the race to release new product generations, it was faster to scale old designs to smaller feature sizes than to innovate at the architecture level.

Carver Mead, *How we created neuromorphic engineering*, Nature Electronics (2020)



→ GAP ←



Images: Purves et al., Neuroscience

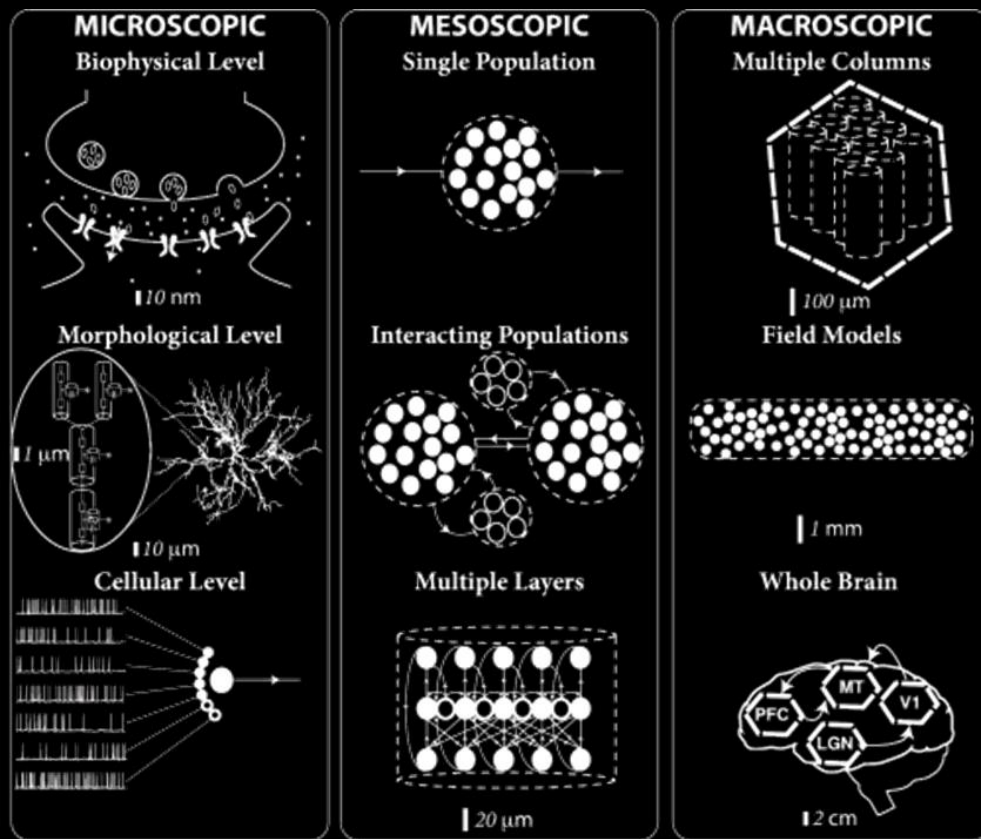


Image: Neuronal Dynamics, Chapter 12

10^4 cells and several km
of nerves (axons) per mm^3

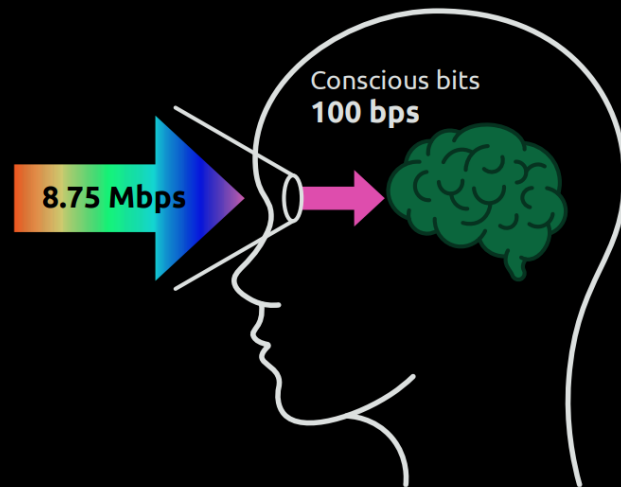
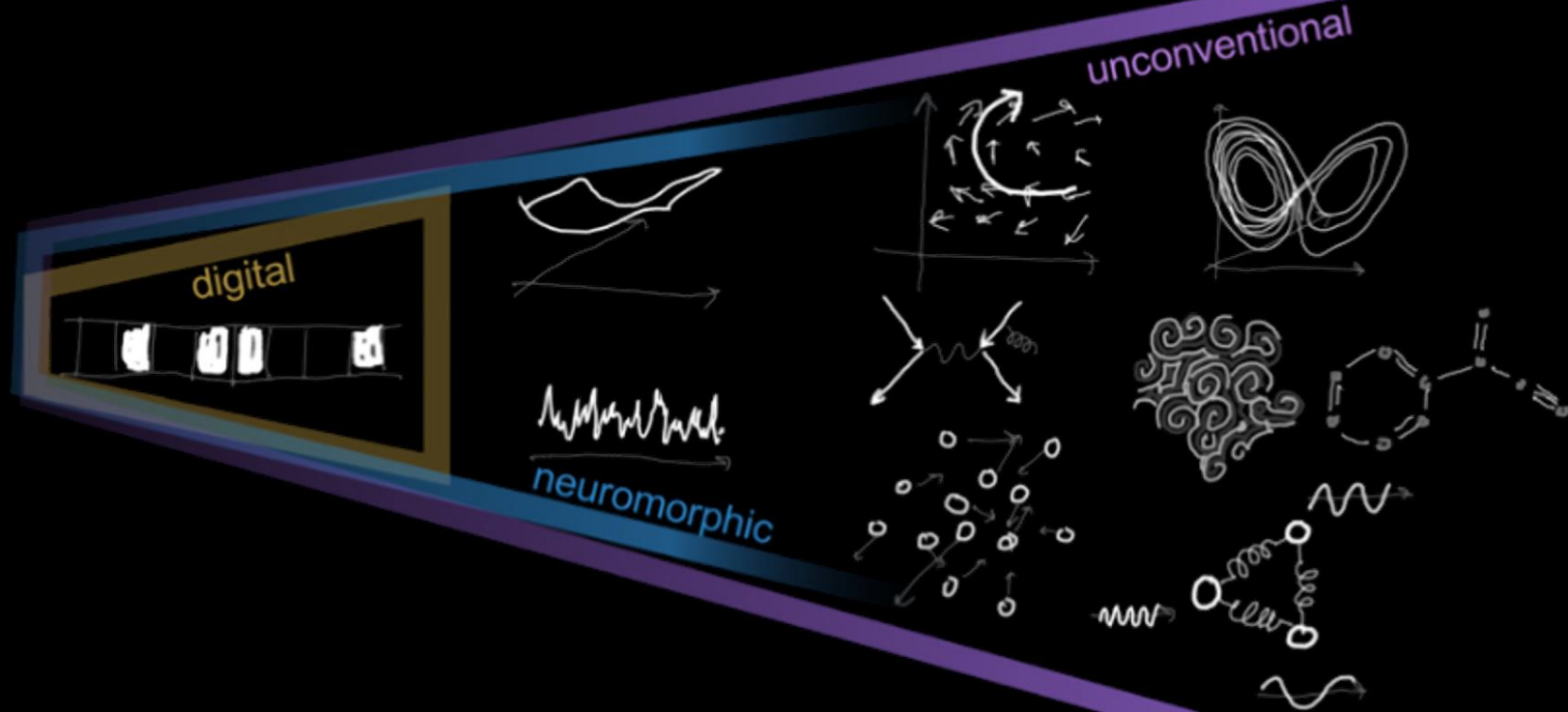


Image: Decadal Plan for Semiconductors

See: Frenkel, Bol and Indiveri, *Bottom-Up and Top-Down Approaches for the Design of Neuromorphic Processing Systems: Tradeoffs and Synergies Between Natural and Artificial Intelligence*, Proceedings of the IEEE (2023); DOI: 10.1109/JPROC.2023.3273520.

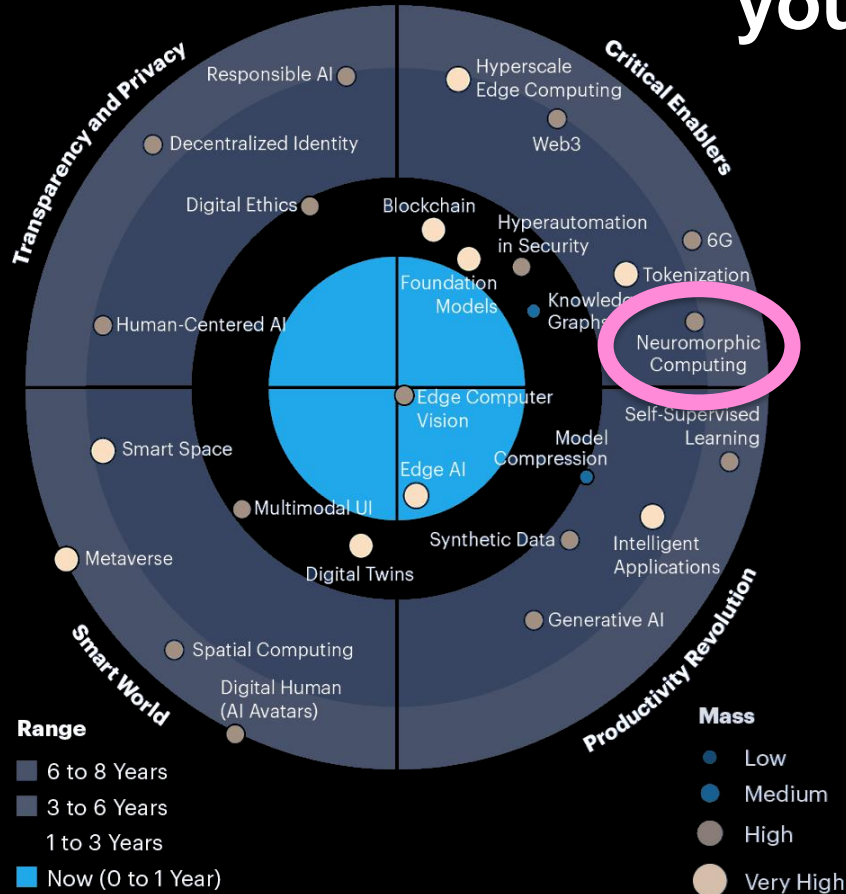
Towards a generalized theory of computing



Herbert Jaeger, Neuromorph. Comput. Eng.1 (2021); <https://doi.org/10.1088/2634-4386/abf151>

Lecture: <https://www.youtube.com/playlist?app=desktop&list=PL2Mh0Lr7X-WWEk8-NRB28rUqrLXitFaZ4>

Gartner #1 emerging technology that you need to know about

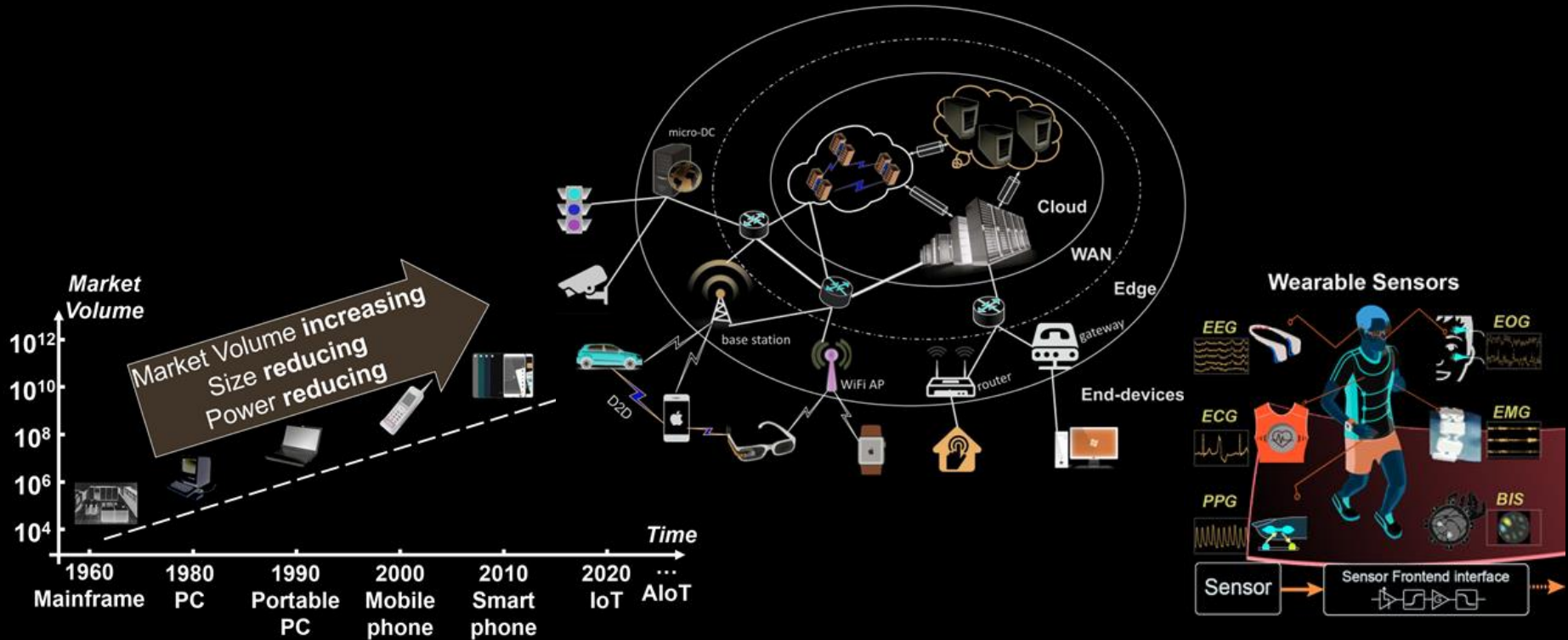


expected to disrupt many of the current AI technology developments

substantial impact on existing products and markets

3–6 yrs to cross over from early-adopter status to early majority adoption

Data at network edge increasingly valuable

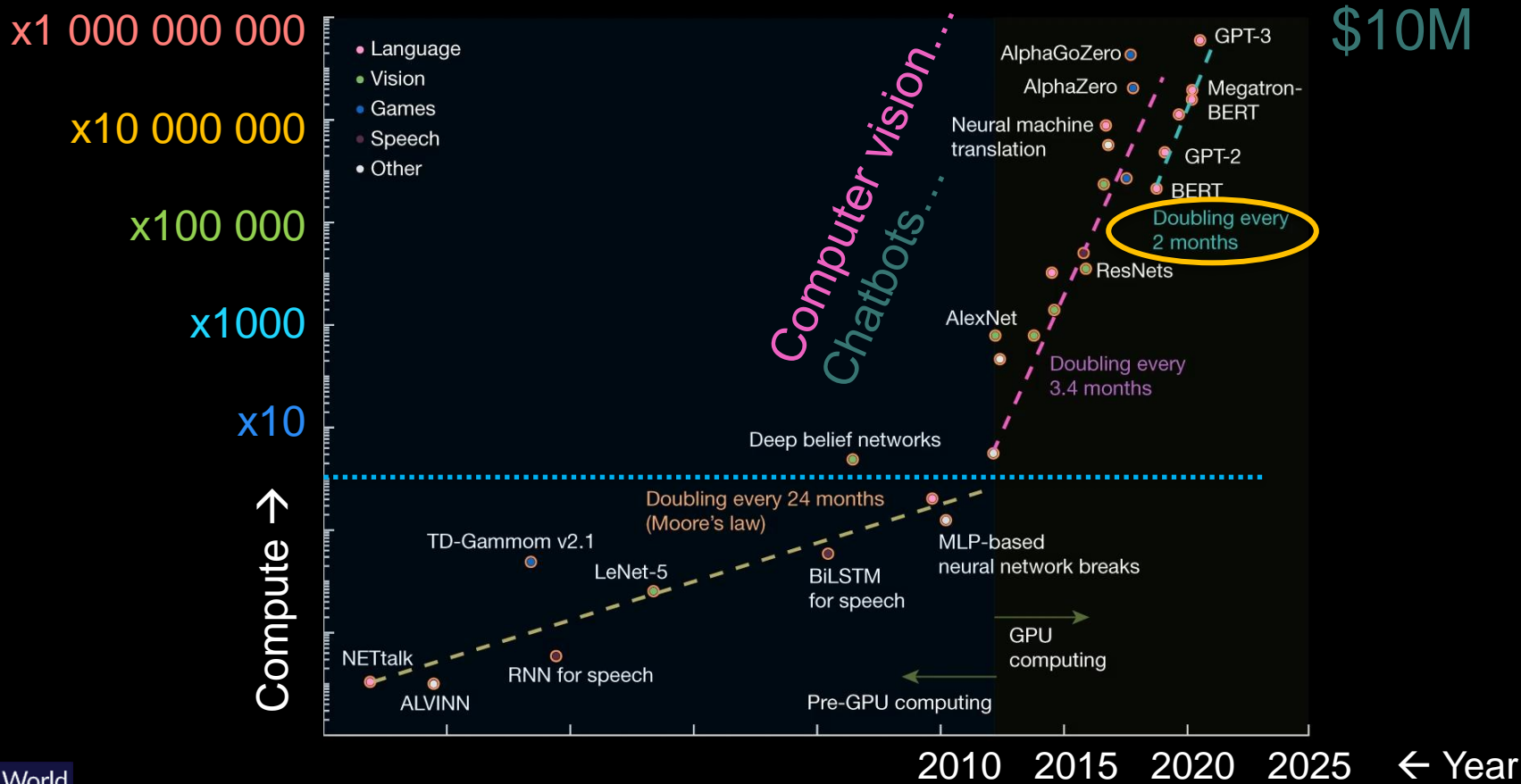


Ye, et al. (2021) "Challenges and Emerging Technologies for Low-Power Artificial Intelligence IoT Systems"

Zhou, et al. (2019) "Edge Intelligence"

Covi, et al. (2021) "Adaptive Extreme Edge Computing for Wearable Devices"

Deep learning is too resource intensive



Energy-efficiency gap

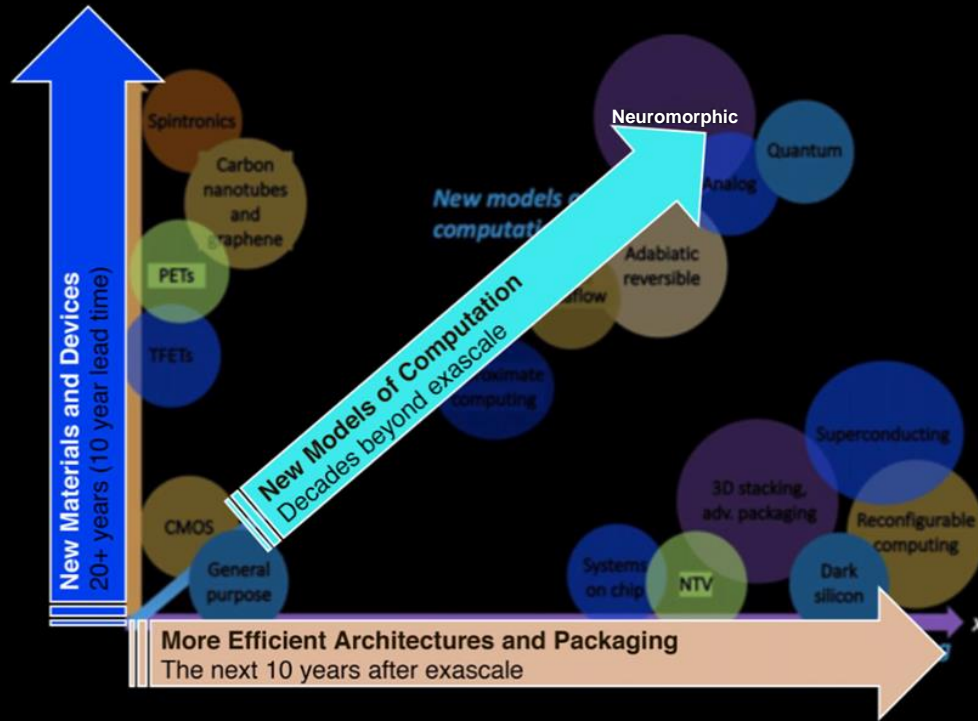


~20 W



21 MW

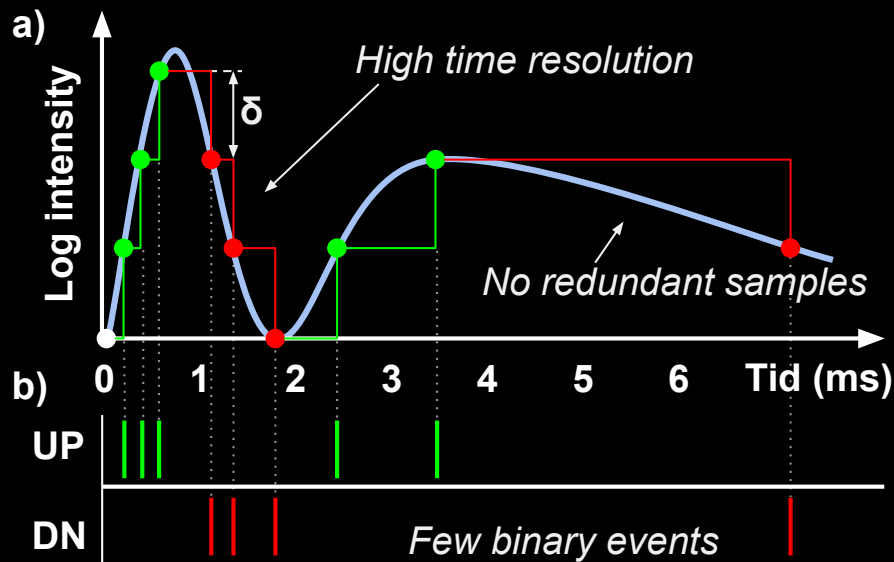
“Beyond Moore”, “Time Moore”



*Rethink the constraint that
“all cycles must have the
same duration in time”*

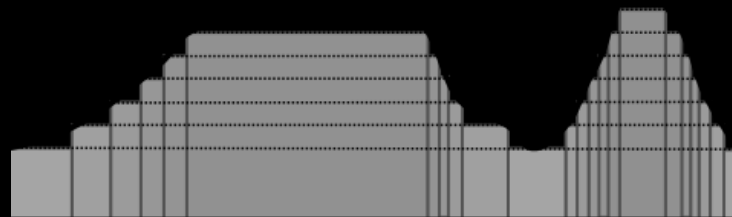
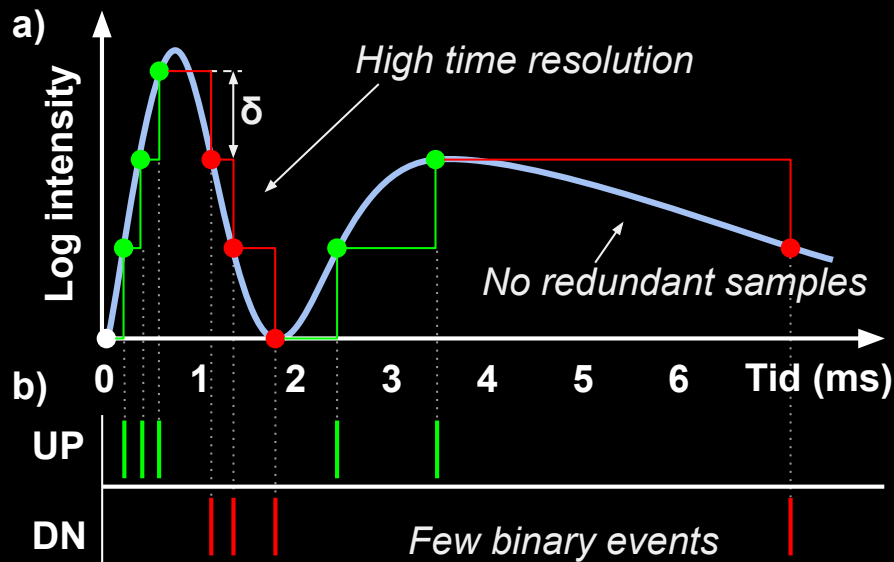
Event-triggered sampling,
processing and control

Level-crossing ADC



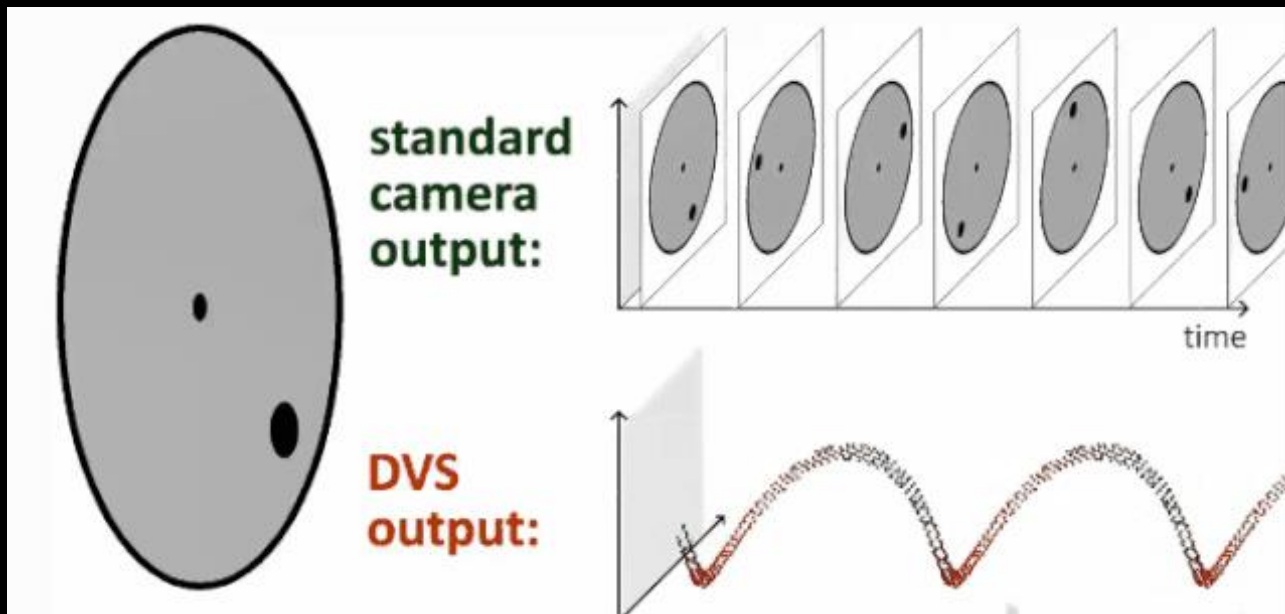
Signal changes by some delta \rightarrow new information

Lebesgue sampling paradigm

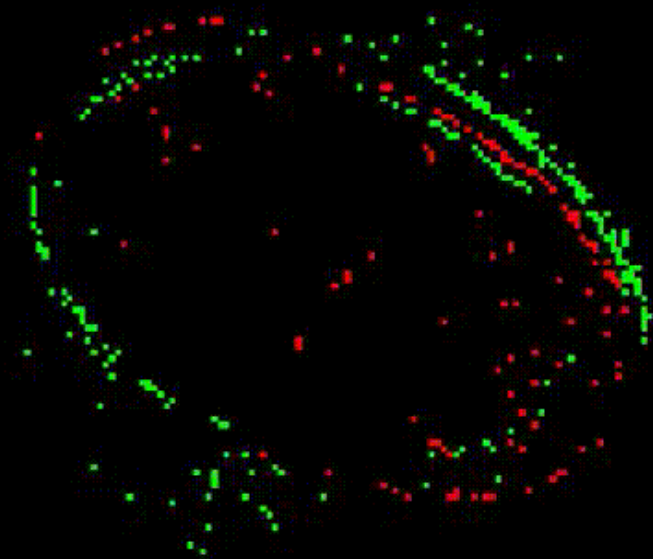


Lebesgue integration instead of Riemann integration with constant dt (clock cycle)

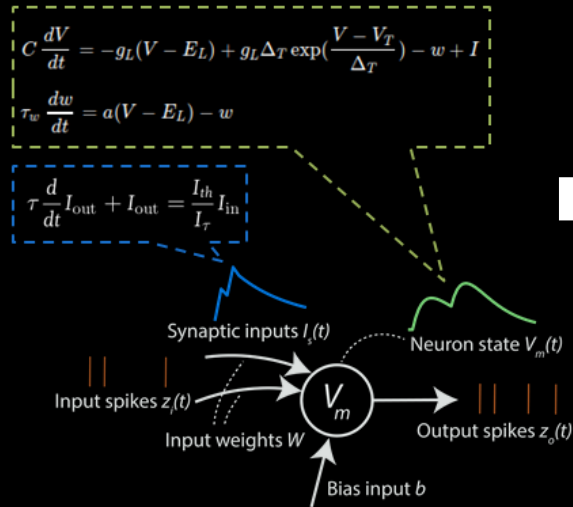
Dynamic vision sensors (DVS)



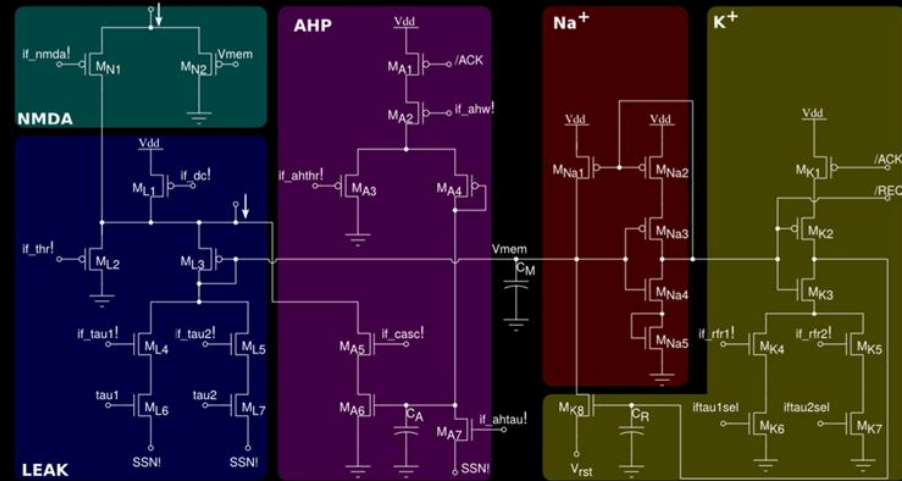
Source: INI & RPG Zurich



Dynamic models of neurons and synapses



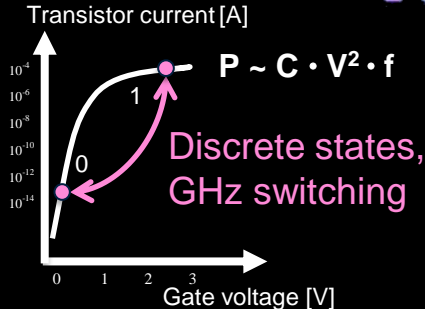
Analog “silicon-neuron” circuit Conventional CMOS-transistor technology



Information Processing Concepts

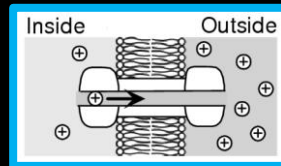
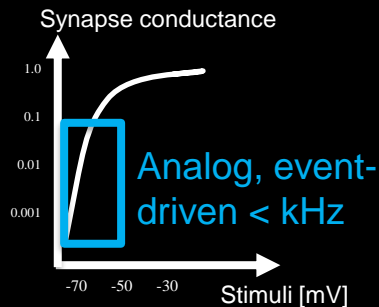
Conventional computers

*mimic
logical and
analytical
thinking*

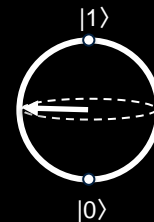


Neuromorphic processors

*mimic
the senses,
learning and
perception*

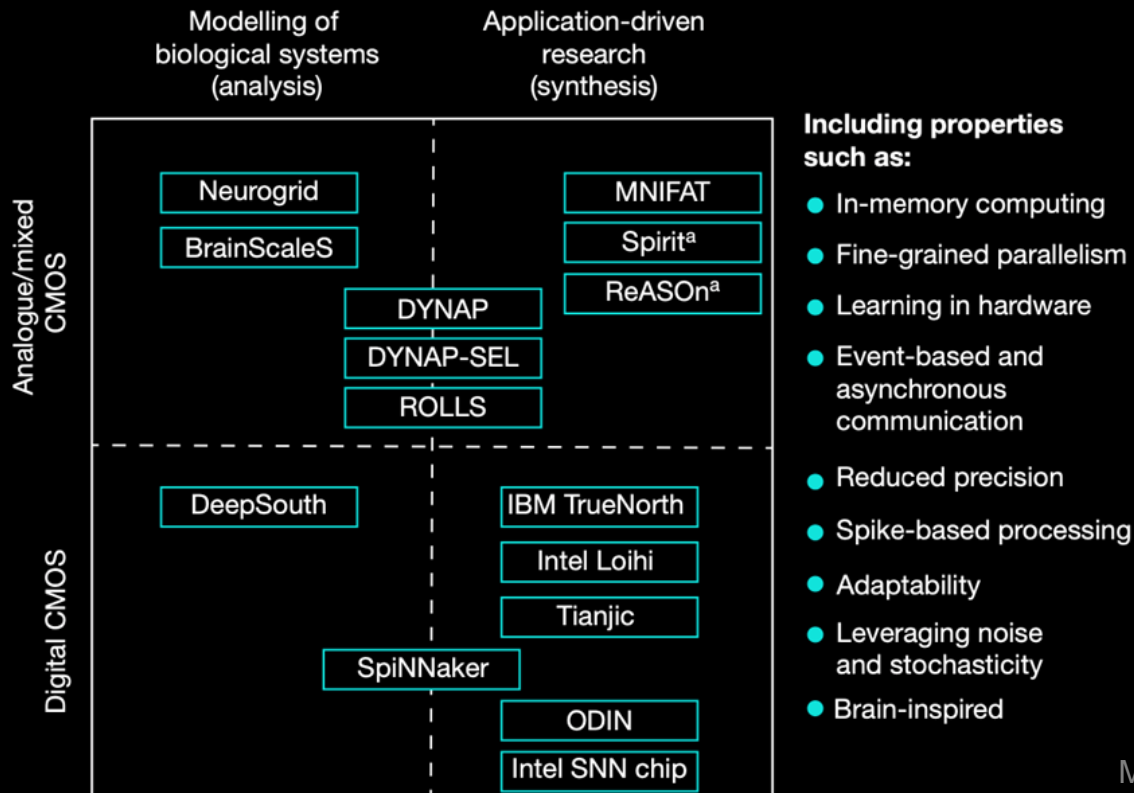


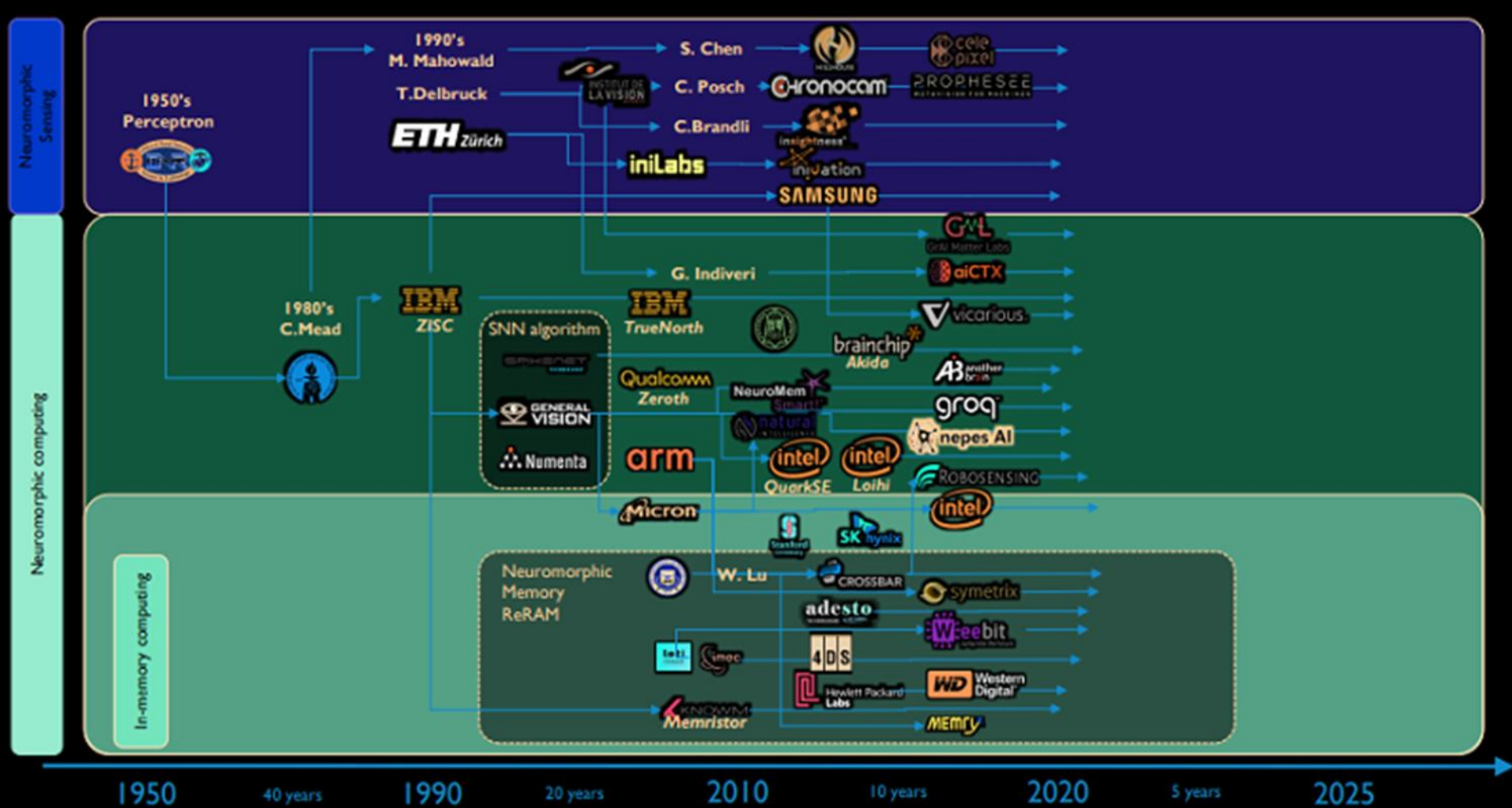
Quantum
processors
*exploit quantum
superpositions
for probabilistic
inference*



Cell membrane
ion diffusion

Neuromorphic Computing Systems





Energy-efficiency comparison

	Human brain	Technology	Chip	Worse by
Housekeeping	4.8E-11	9.8E-07	RTX2070	2.0E+04
Resting potential	5.8E-11	3.8E-08	Spikey	6.6E+02
Action potential	2.0E-11	4.4E-10	Spikey	2.2E+01
Spike transmission	8.2E-15	1.1E-11	Spikey	1.3E+03
Single neuron	2.5E-10	1.5E-06	Spikey	1.3E+03
Full brain	2.1E+01	1.3E+05	Spikey	6.2E+03

Neurons

Neuron = Processor

Produces an output in response to the various incoming signals based on its own dynamics or algorithm

Dendrites

Axon = Wire

Relays the neural output to the next layer of processing

Synapse = Memory

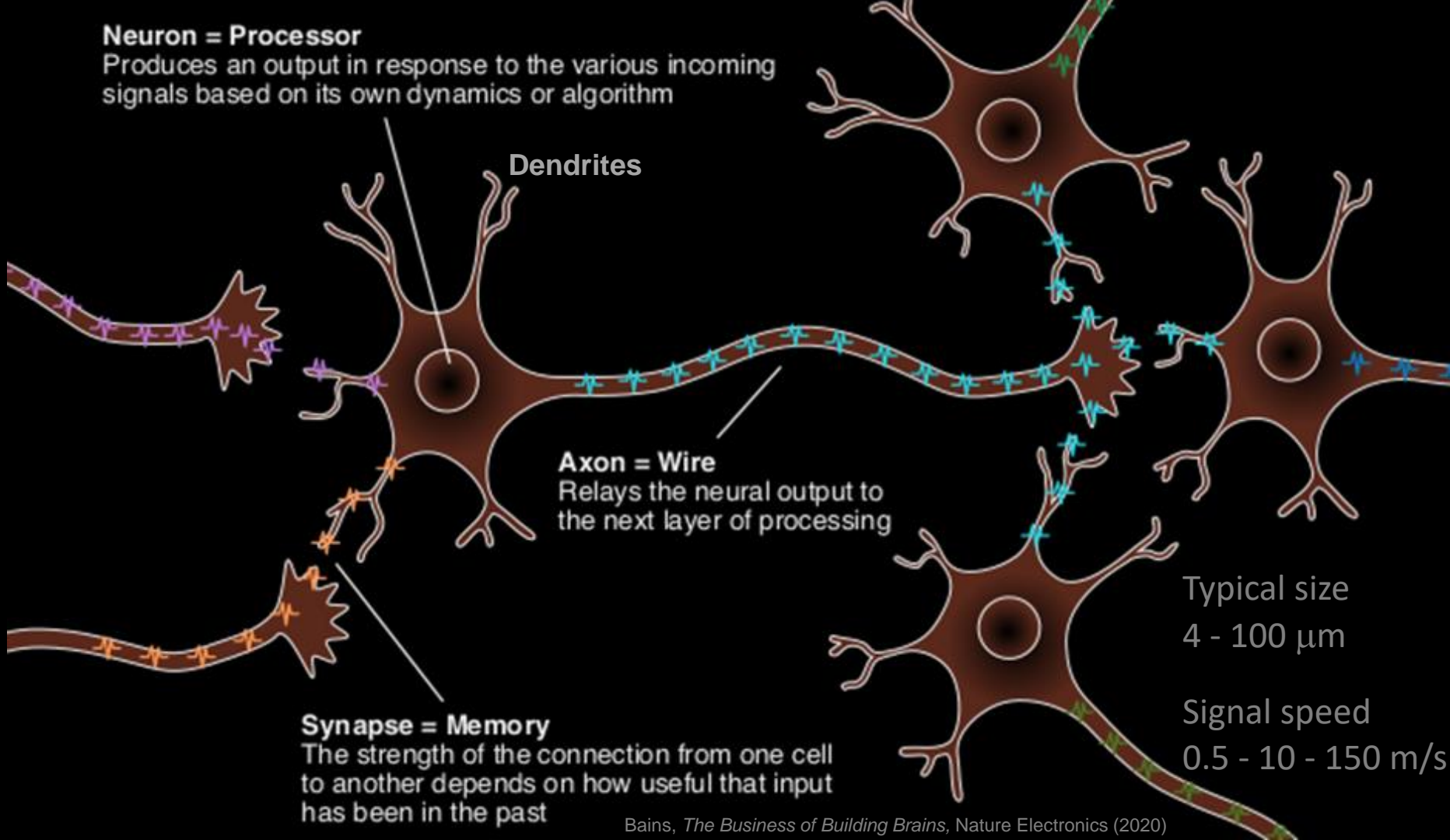
The strength of the connection from one cell to another depends on how useful that input has been in the past

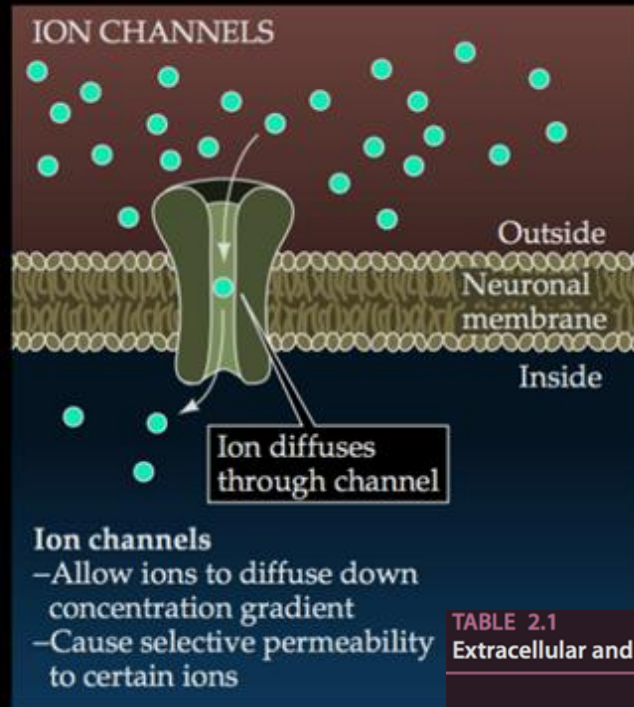
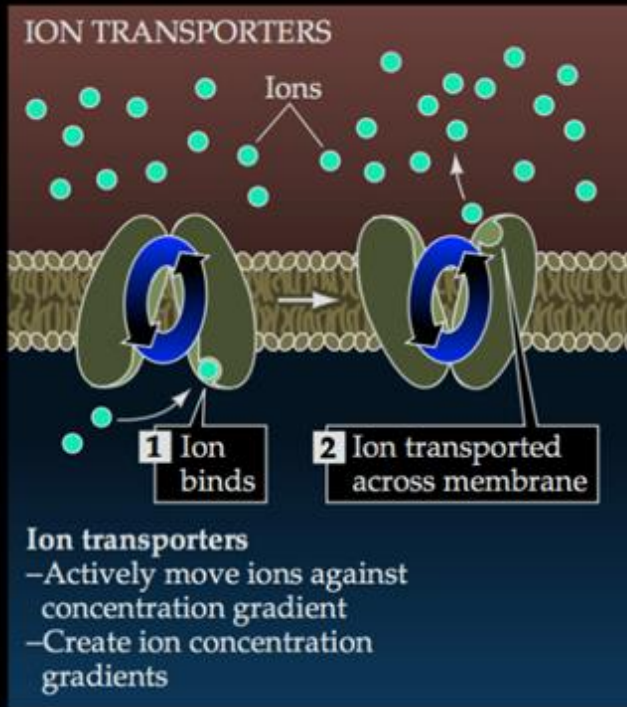
Typical size

4 - 100 μm

Signal speed

0.5 - 10 - 150 m/s





Nernst potential

$$\Delta u = \frac{kT}{q} \ln \frac{n_2}{n_1}$$

Image credit: Neuroscience, D. Purves et al.

TABLE 2.1

Extracellular and Intracellular Ion Concentrations

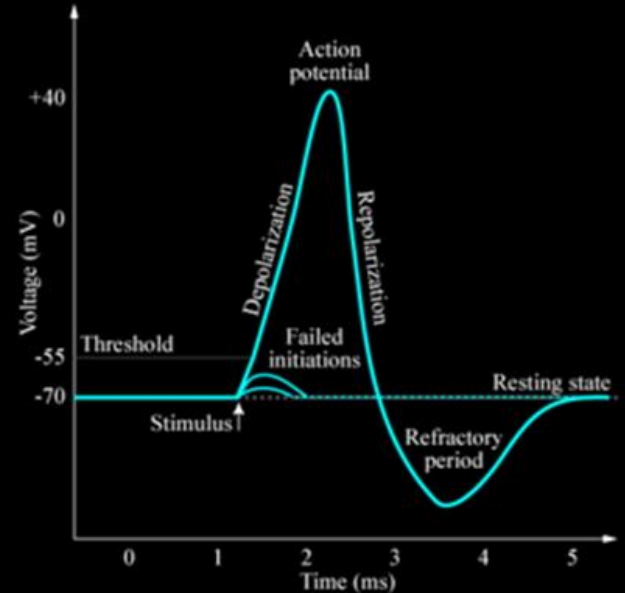
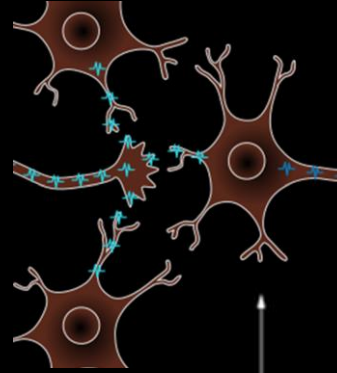
Ion	Concentration (mM)	
	Intracellular	Extracellular
Squid neuron		
Potassium (K ⁺)	400	20
Sodium (Na ⁺)	50	440
Chloride (Cl ⁻)	40–150	560
Calcium (Ca ²⁺)	0.0001	10
Mammalian neuron		
Potassium (K ⁺)	140	5
Sodium (Na ⁺)	5–15	145
Chloride (Cl ⁻)	4–30	110
Calcium (Ca ²⁺)	0.0001	1–2

Action potentials

About -65 mV resting potential

Surrounding bath \equiv 0 mV

Positive Na^+ feedback process
if threshold voltage reached



Action potentials and “spike” approximation

About -65 mV resting potential

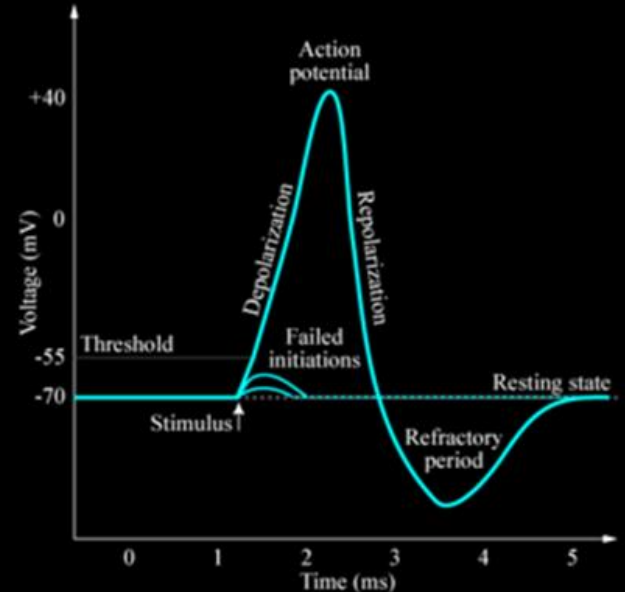
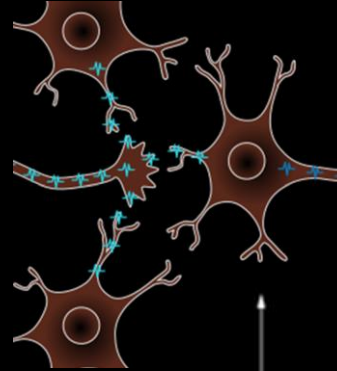
Surrounding bath \equiv 0 mV

Positive Na^+ feedback process
if threshold voltage reached

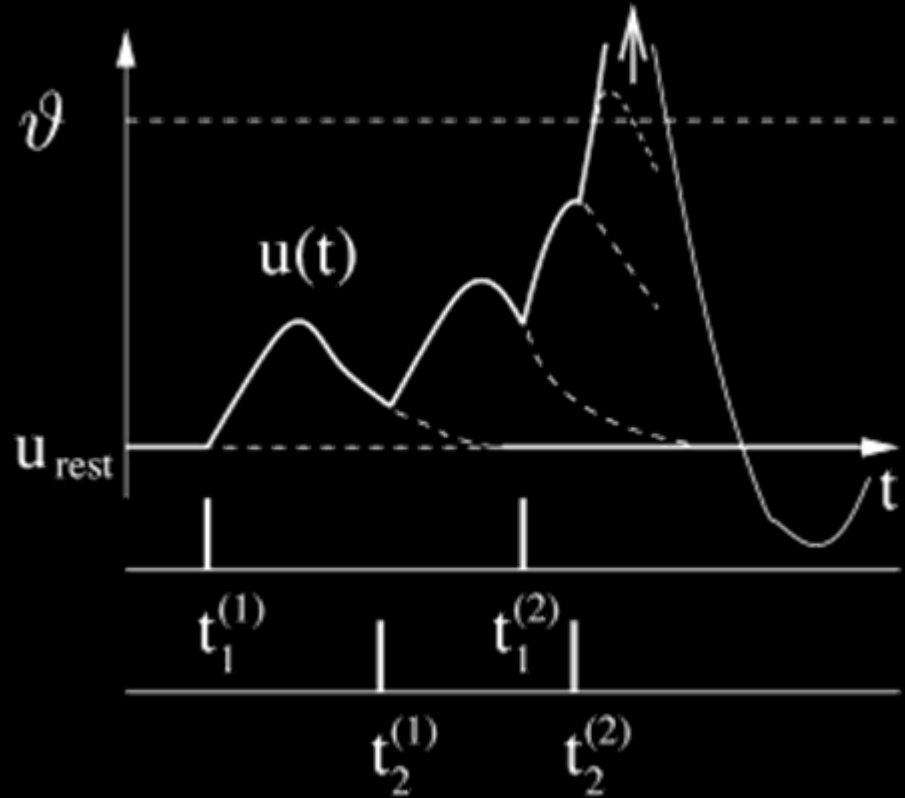
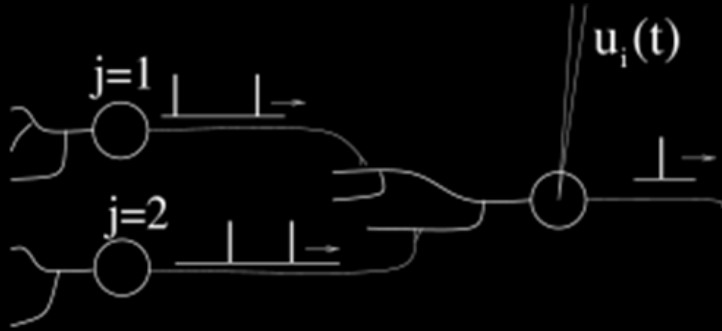


~100 mV fluctuation for ~1 ms

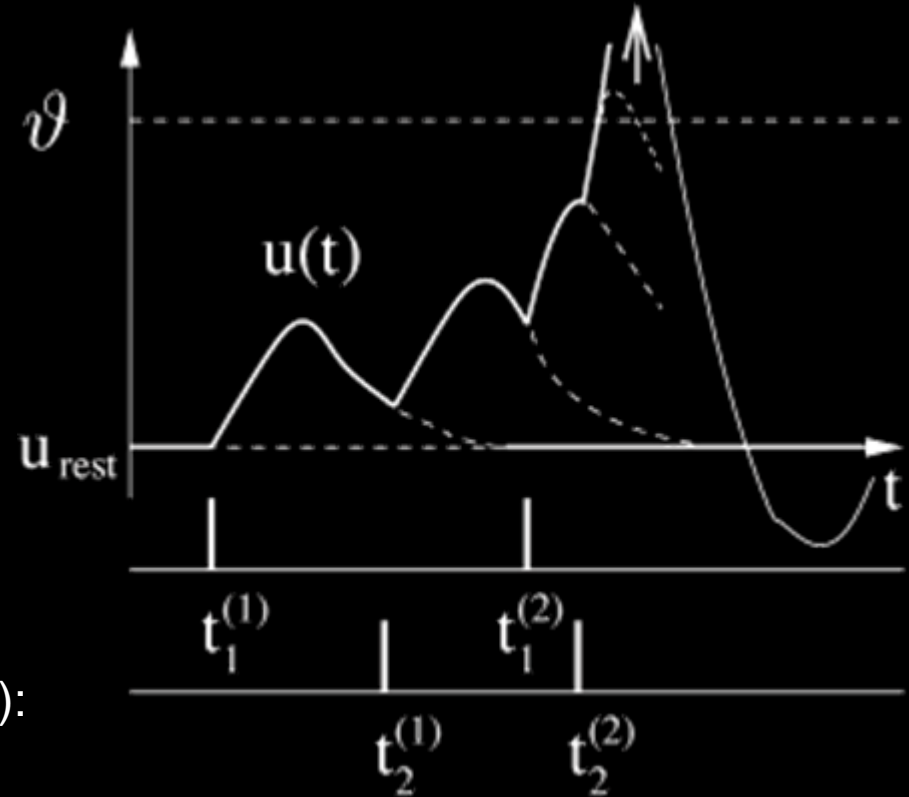
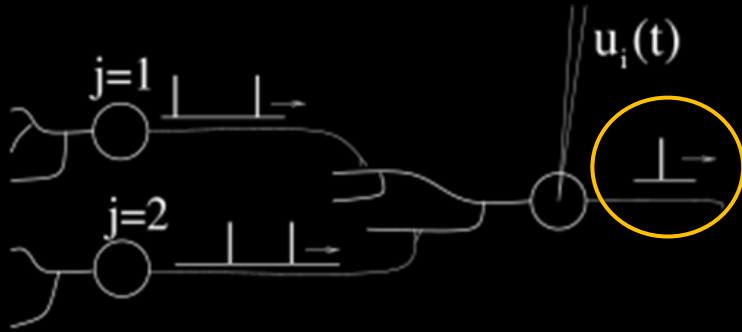
- Action potential
- Mostly stereotype events, **spikes**, characterized by the **spike time**



Spiking Neural Networks (SNN)



Spiking Neural Networks (SNN)



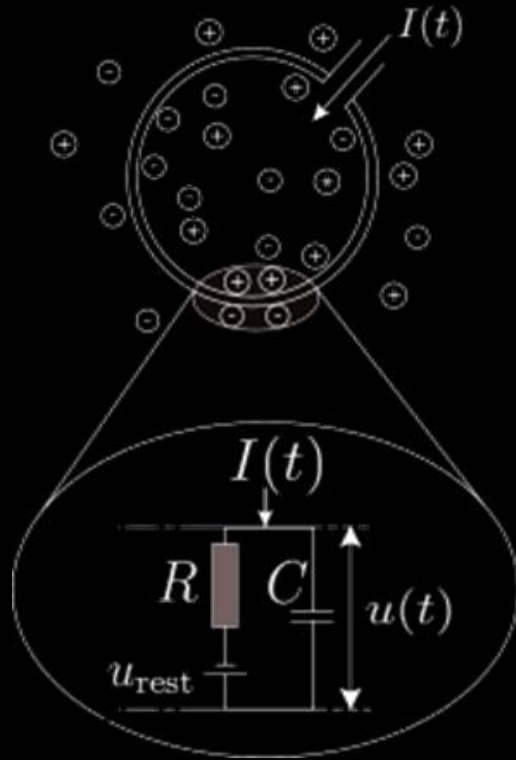
Spike

Address Event Representation (AER):

Address of source neuron

Spike time \equiv Physical time

Leaky Integrate and Fire (LIF) neuron model



$$\tau_m \frac{du}{dt} = u_{rest} - u(t) + RI(t)$$

Threshold condition: $u(t) = \vartheta$

Then generate spike at t and set $u = u_r < \vartheta$

```
#!/usr/bin/env python

from brian2 import *

n = 100                # No. of neurons
duration = 2*second    # Simulation time

R = 90e6                # Membrane resistance, Ohm
tau = 30*ms            # Leakage time constant
E_L = -65*mV           # Leakage resting potential
v_thres = -50*mV       # Threshold potential

# LIF equations, define RI = R*I(t) as a voltage for simplicity
eqs = '''
    dv/dt = (E_L - v + RI) / tau : volt (unless refractory)
    RI : volt
    '''

# Define a population of LIF neurons
group = NeuronGroup(n, eqs, threshold='v > v_thres',
    reset='v = E_L', refractory=5*ms, method='linear')

# Define the initial value of the membrane potential
group.v = E_L

# Each neuron is fed by a different current via the RI term
group.RI = '100*mV * i / (n-1)'

# Create monitors to enable plotting of variables
monitor1 = SpikeMonitor(group)
monitor2 = StateMonitor(group, 'v', record=True)

# Run the simulation
run(duration)

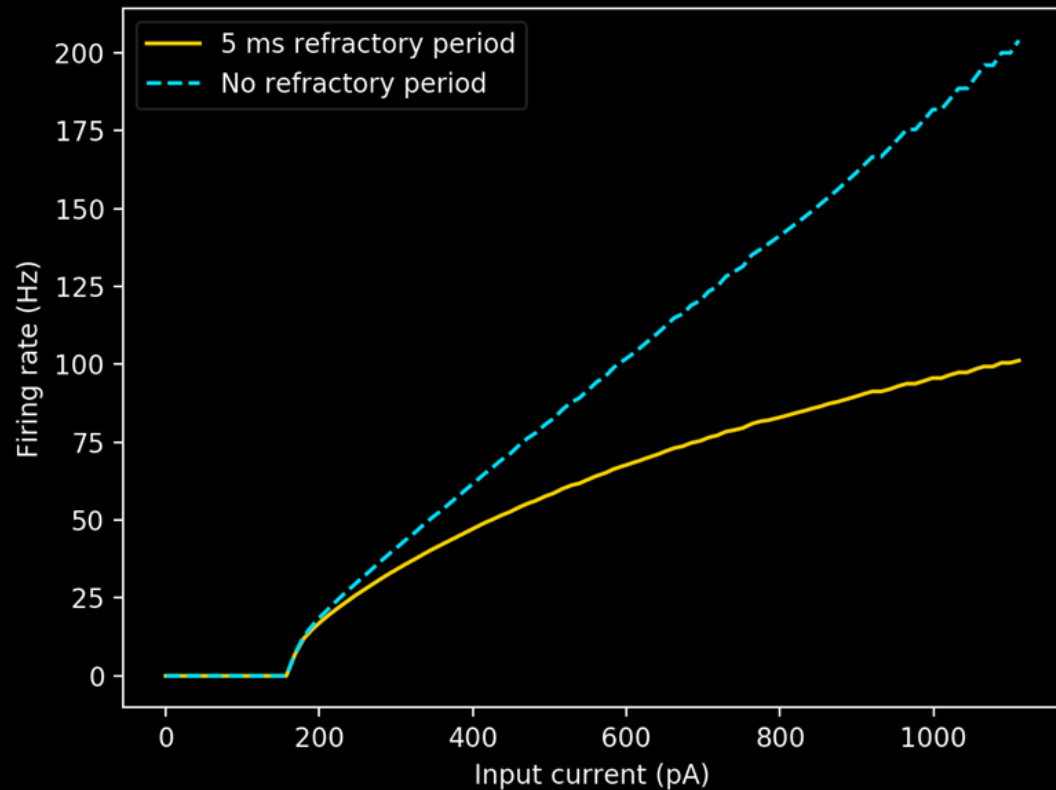
# Plot results
figure(figsize=(12,4))
subplot(121)
plot(group.RI*(1e-3/mV)/R/1e-12, monitor1.count / duration, '-b')
xlabel('Input current (pA)')
ylabel('Firing rate (Hz)')
subplot(122)
plot(monitor2.t, monitor2.v[15]*1e-3/mV, '-r', label='Neuron 15')
plot(monitor2.t, monitor2.v[20]*1e-3/mV, '--g', label='Neuron 20')
legend()
xlabel('Time (s)')
ylabel('Membrane potential (mV)')
show()
```



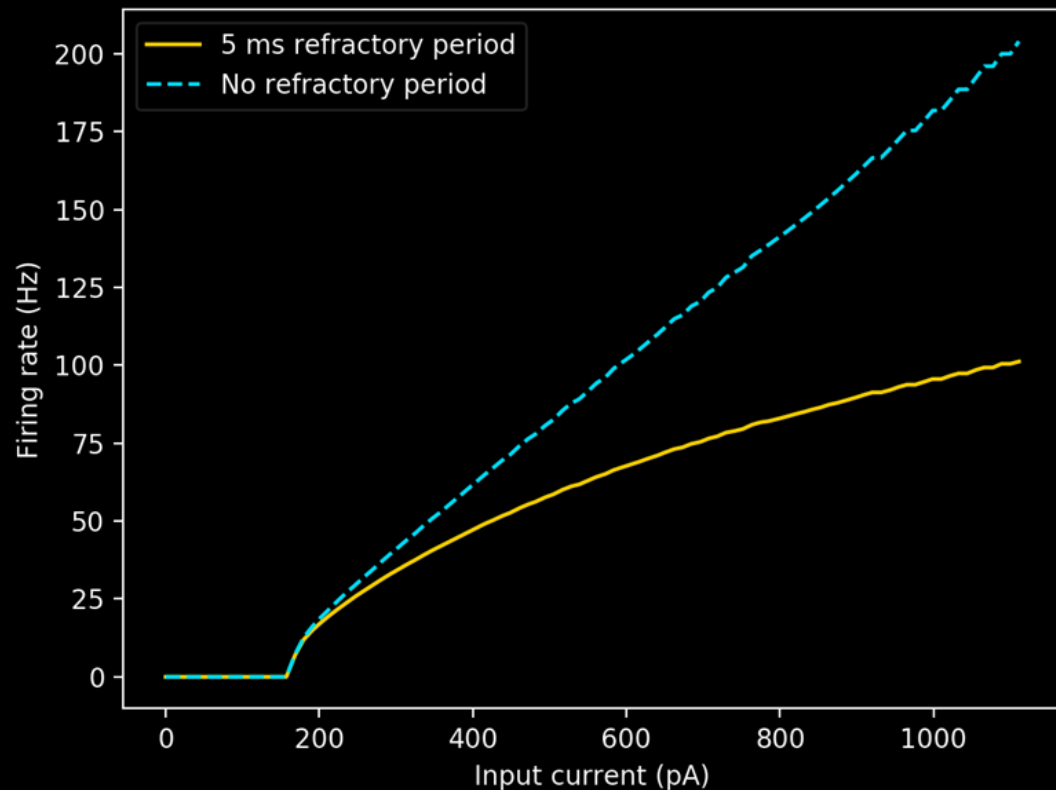
<http://brian2.readthedocs.io>

Brette, Goodman, Stimberg, 2016

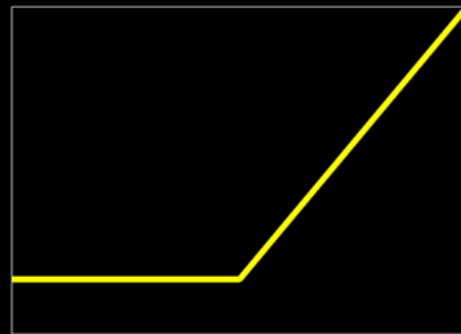
LIF model spike rate versus input current



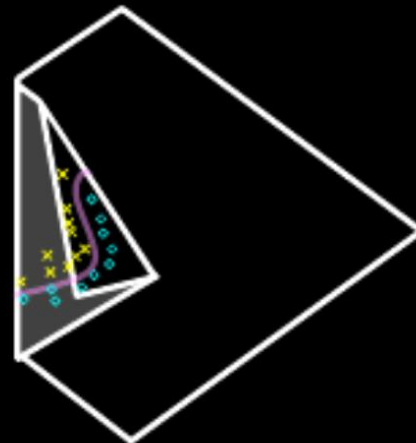
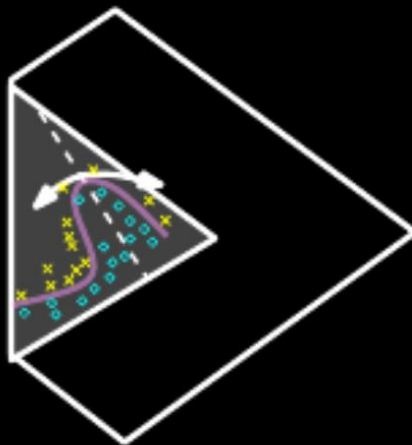
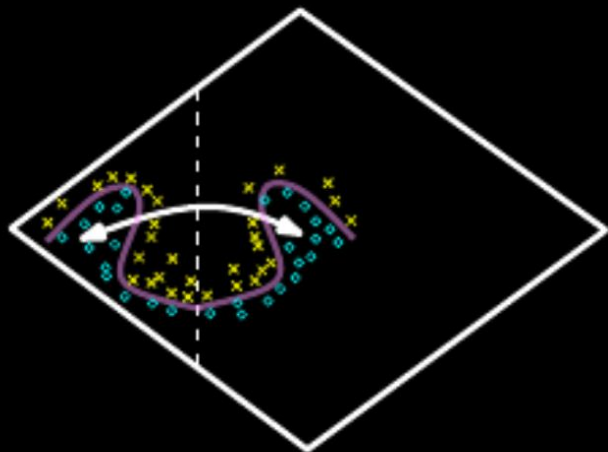
Similar to ReLU in quasistatic regime



ReLU activation in ANN

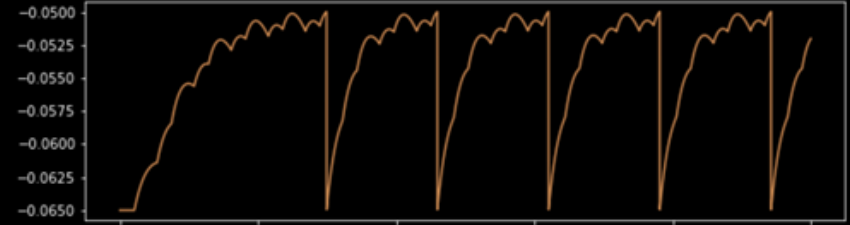
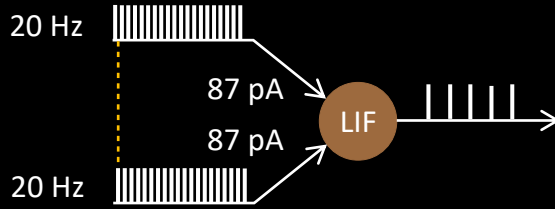


$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$



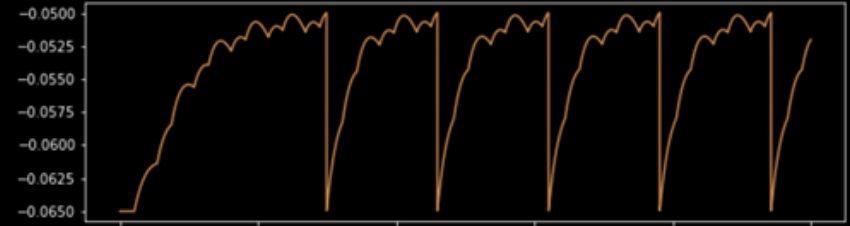
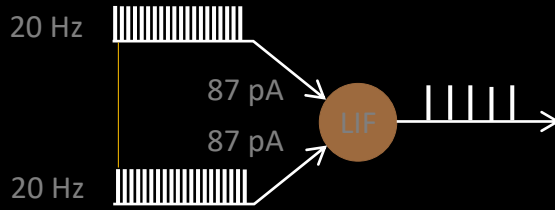
LIF-model example

1) Neuron responds to 20Hz stimuli that is **out of phase**

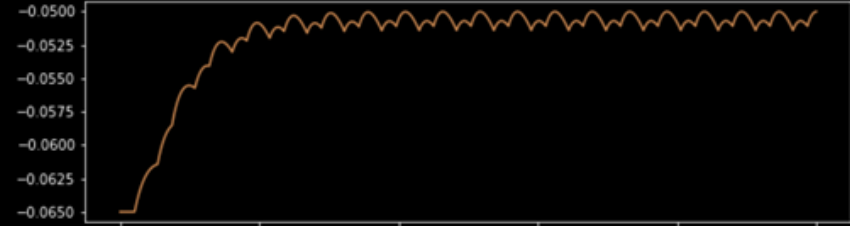
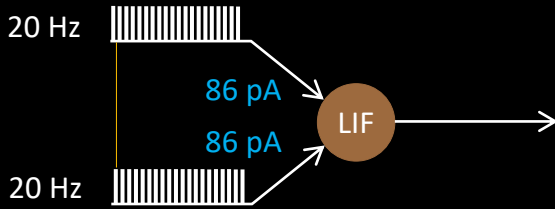


LIF-model example

1) Neuron responds to 20Hz stimuli that is out of phase

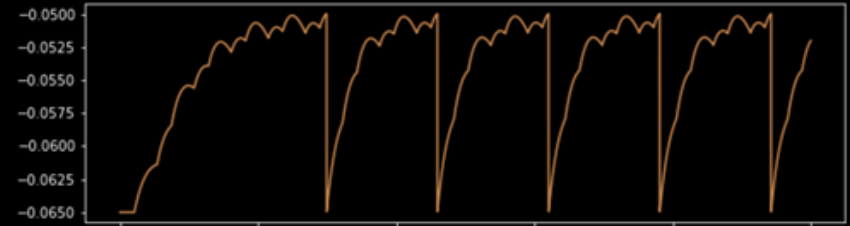
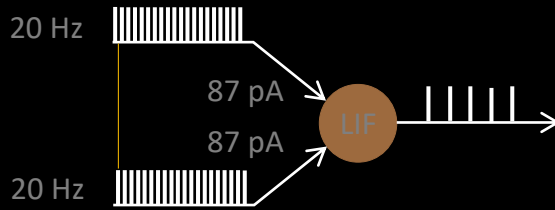


2) **Weights** are lower. Neuron does not respond anymore

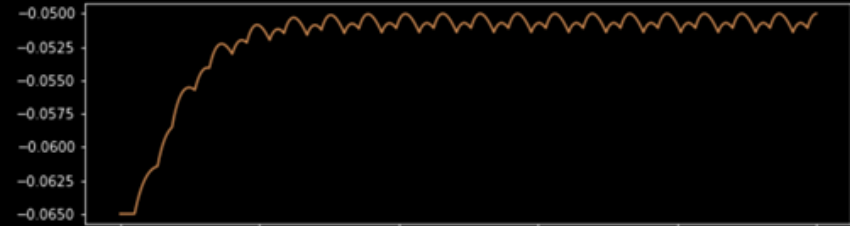
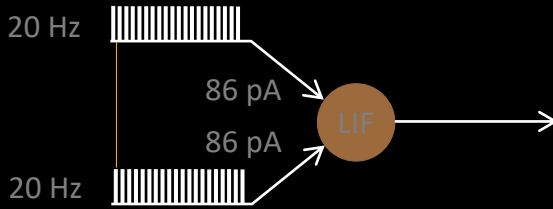


LIF-model example

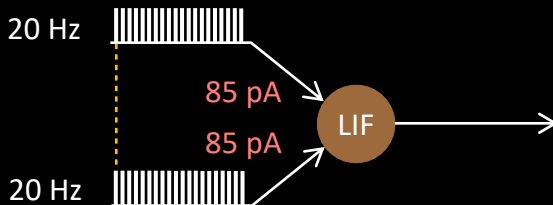
1) Neuron responds to 20Hz stimuli that is out of phase



2) Weights are lower. Neuron does not respond anymore

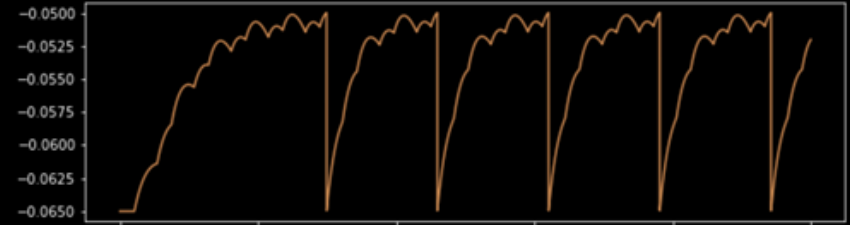
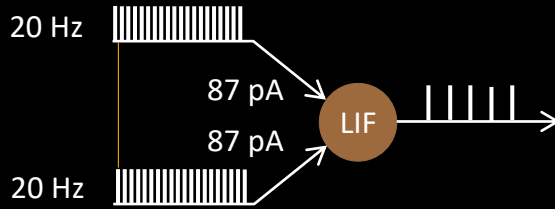


3) Lower weights more. Does it respond to 20Hz in-phase stimuli?

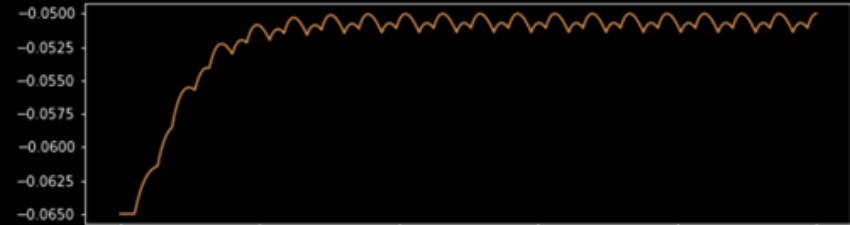
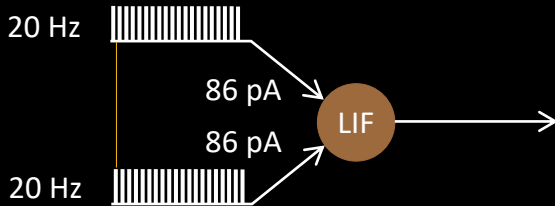


LIF-model example

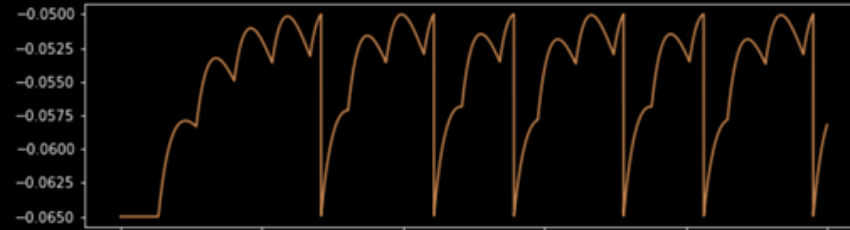
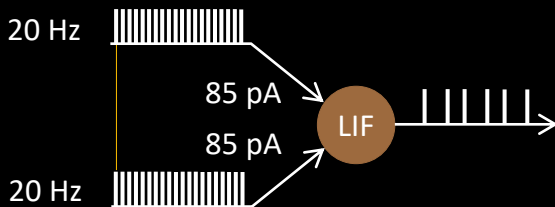
1) Neuron responds to 20Hz stimuli that is out of phase



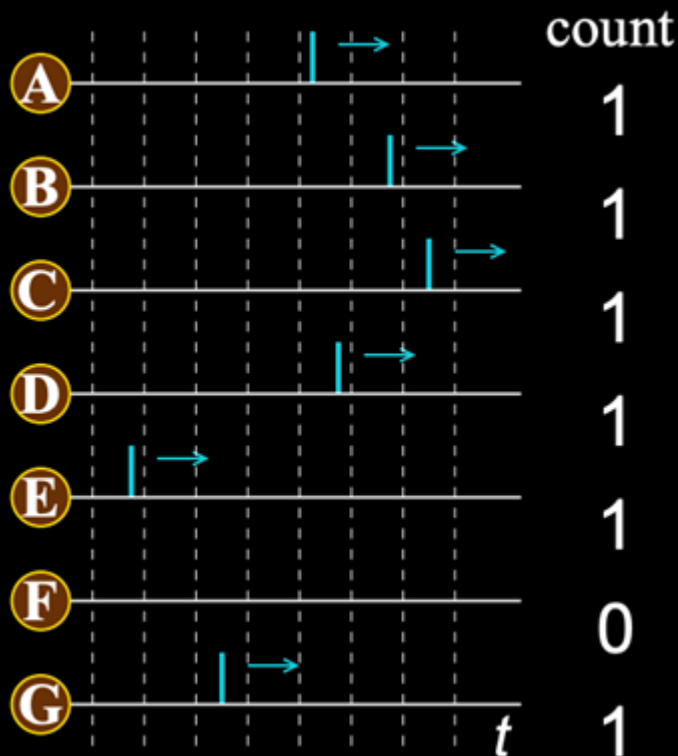
2) Weights are lower. Neuron does not respond anymore



3) Lower weights more. Does it respond to 20Hz in-phase stimuli? -YES



Representational power of spike coding schemes?



Population rate code $\sim \log_2(n+1)$ bits
(inefficient)

Binary code $\sim n$ bits

Timing code $\sim n \log_2(\text{time precision})$

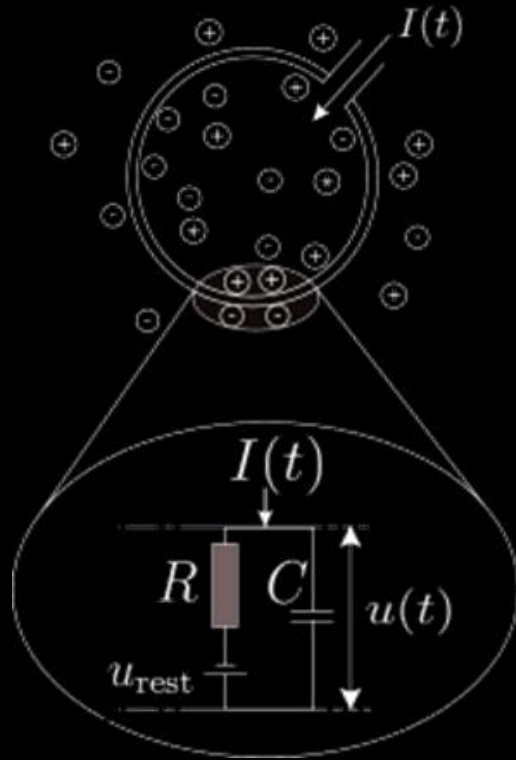
Ordering code $\sim n \log_2(n)$

...

Neural coding in biology ?

Comment on end-to-end
optimization of SNNs

Discretize the LIF model



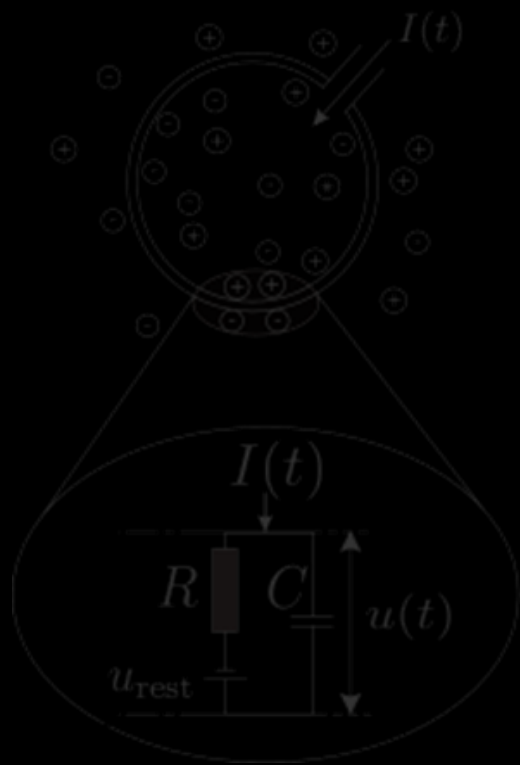
$$\tau_m \frac{du}{dt} = u_{rest} - u(t) + RI(t)$$

Threshold condition: $u(t) = \vartheta$

Then generate spike at t and set $u = u_r < \vartheta$

$$\frac{du}{dt} \approx \frac{u[t+1] - u[t]}{\Delta t}$$

$$u[t+1] = \alpha u[t] + \beta I[t] + \mu$$



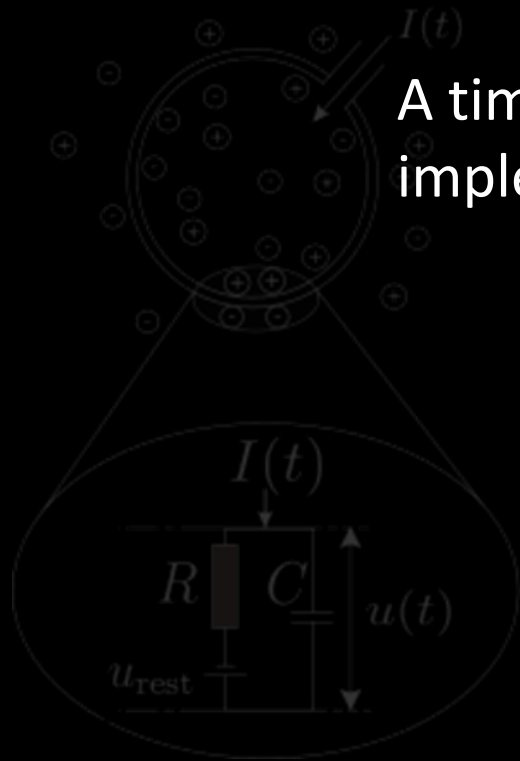
Define spiking output, S

$$\frac{du}{dt} \approx \frac{u[t+1] - u[t]}{\Delta t}$$

$$u[t+1] = \alpha u[t] + \beta I[t] + \mu$$

$$S[t+1] = H(u[t+1] - u_{thr})$$

$$u \rightarrow u_{reset}$$



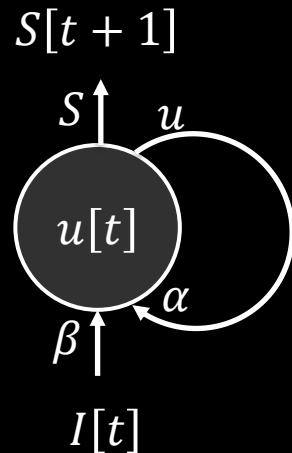
A time-discretized feed-forward SNN can be implemented and optimized as a recurrent ANN

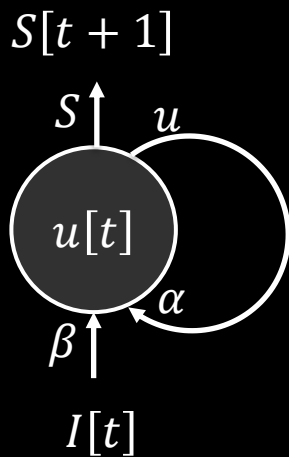
$$\frac{du}{dt} \approx \frac{u[t+1] - u[t]}{\Delta t}$$

$$u[t+1] = \alpha u[t] + \beta I[t] + \mu$$

$$S[t+1] = H(u[t+1] - u_{thr})$$

$u \rightarrow u_{reset}$

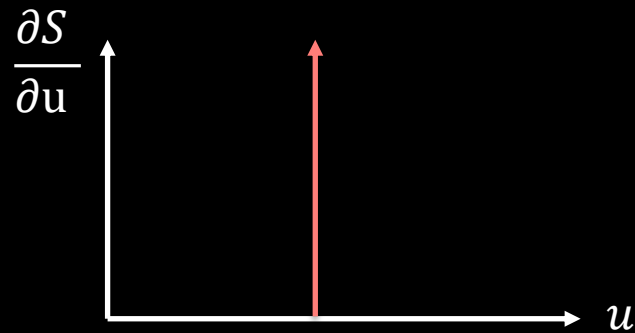
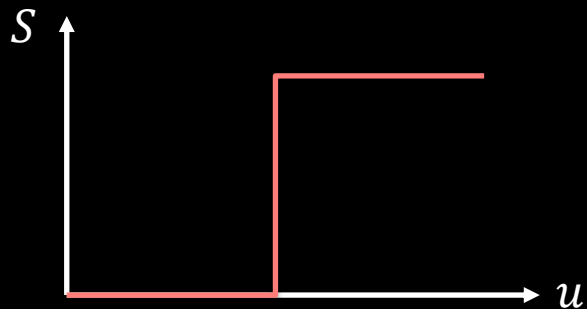


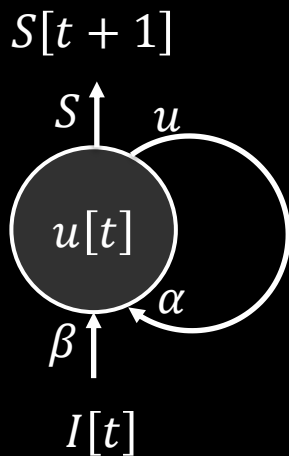


$$u[t+1] = \alpha u[t] + \beta I[t] + \mu$$

$$S[t+1] = H(u[t+1] - u_{thr})$$

$u \rightarrow u_{reset}$

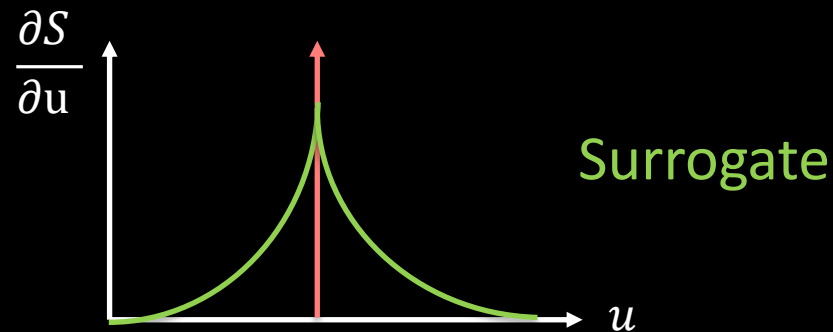
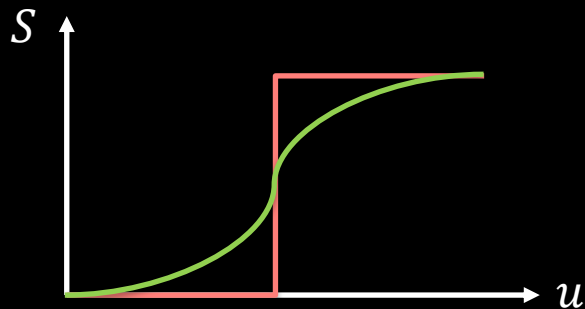




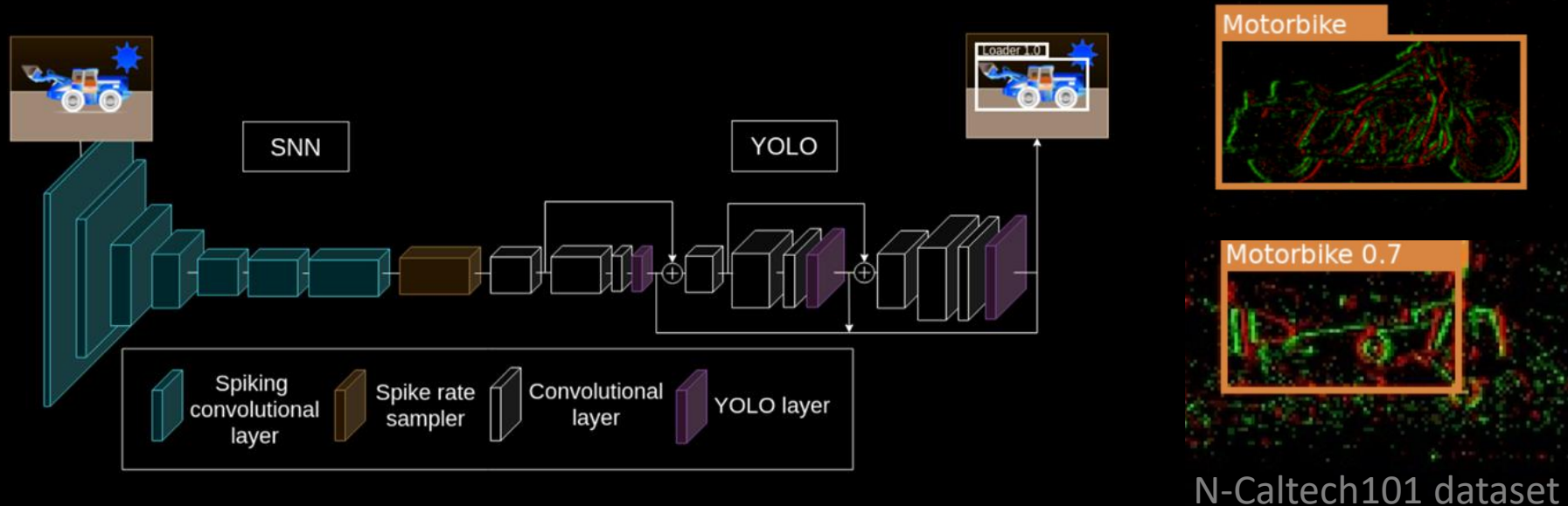
$$u[t+1] = \alpha u[t] + \beta I[t] + \mu$$

$$S[t+1] = H(u[t+1] - u_{thr})$$

$u \rightarrow u_{reset}$



Example: Optimization of Hybrid SNN-ANN

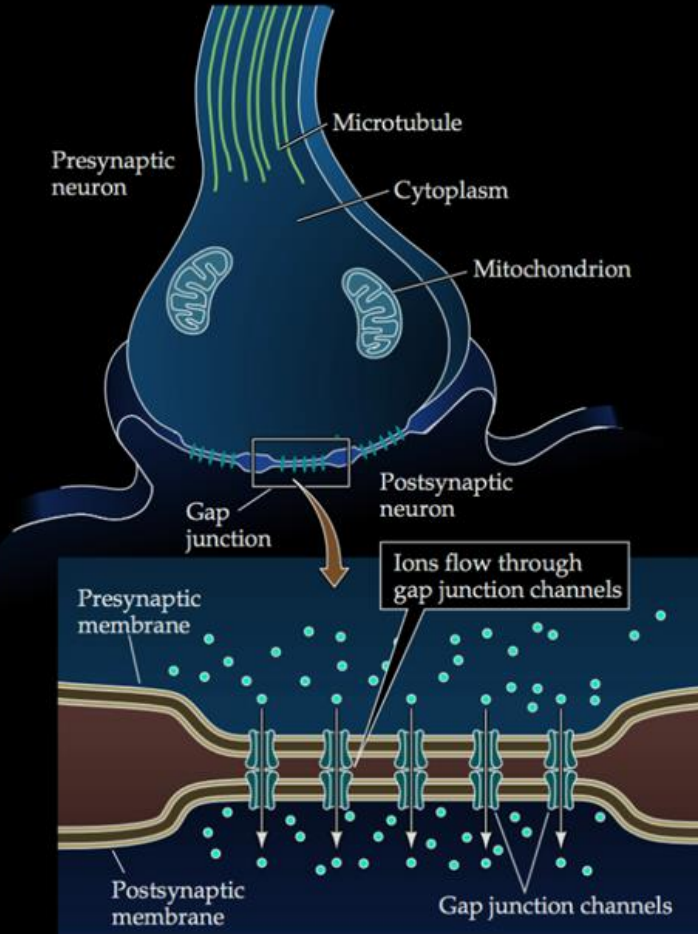


Olof Johansson, Training of Object Detection Spiking Neural Networks for Event-Based Vision, MSc thesis, LTU (2021)

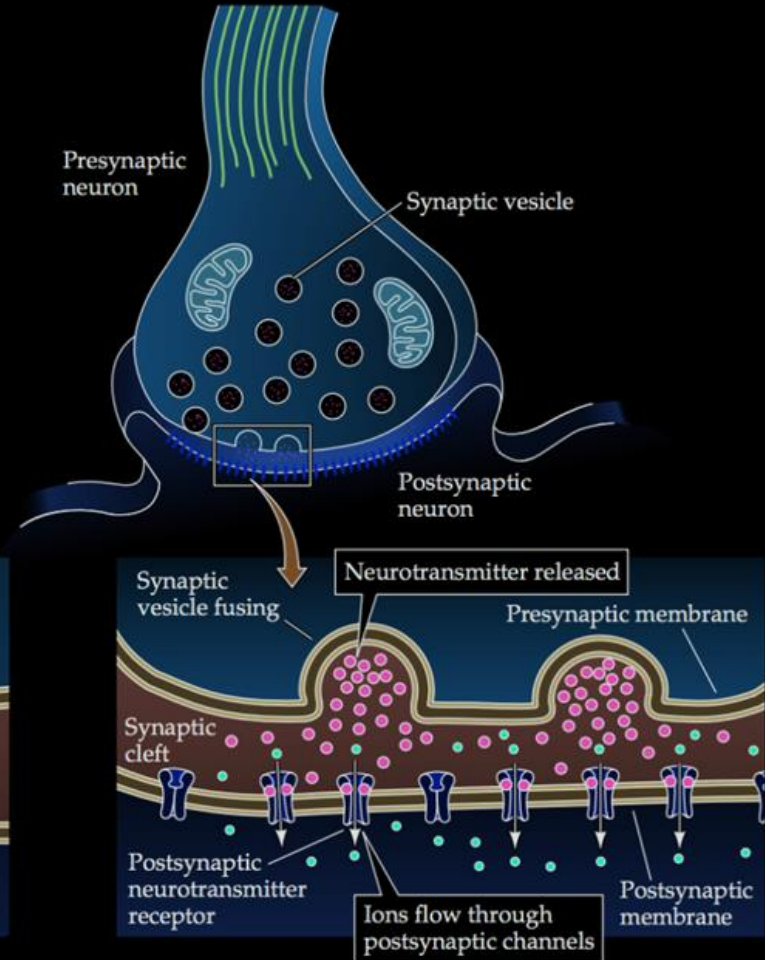
See also Kugele et al, Hybrid SNN-ANN: Energy-Efficient Classification and Object Detection for Event-Based Vision (January 2022)

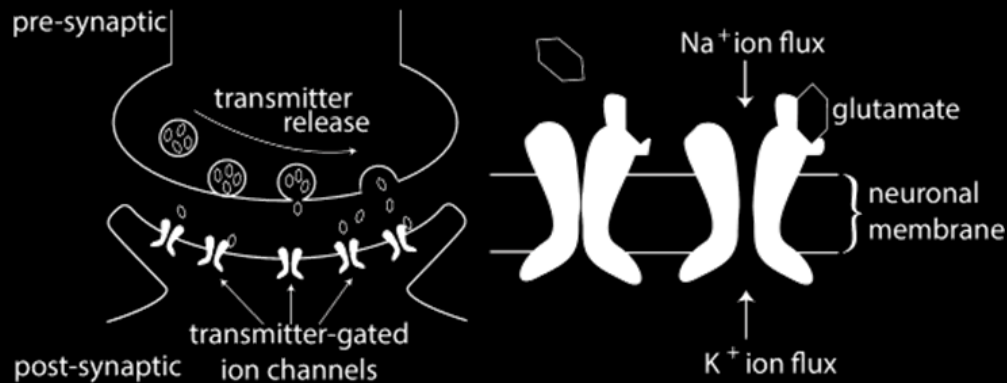
Synapses and learning

(A) ELECTRONIC SYNAPSE

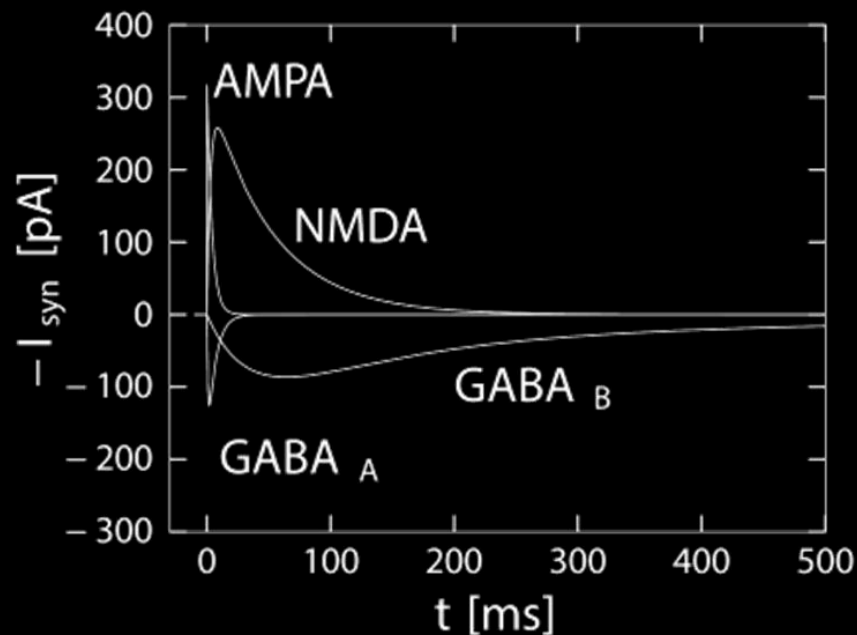


(B) CHEMICAL SYNAPSE





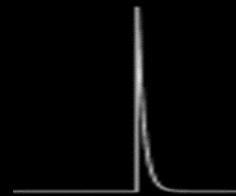
$$g_{\text{syn}}(t) = \sum_f \bar{g}_{\text{syn}} e^{-(t-t^{(f)})/\tau} \Theta(t - t^{(f)})$$



Current-based synapse approximation

$$\tau_{syn} \frac{dI_{syn}}{dt} = -I_{syn}$$

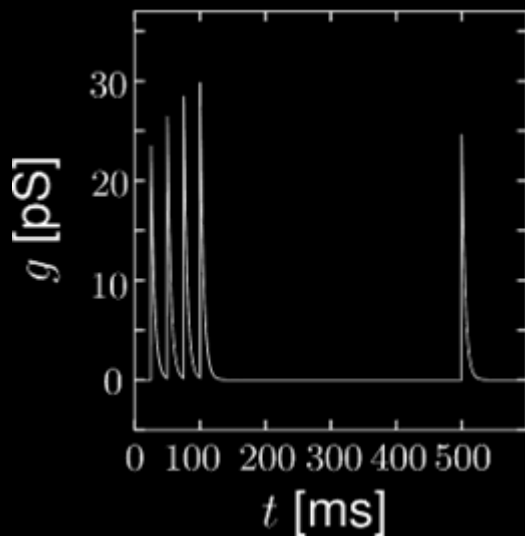
On presynaptic spike: $I_{syn} \leftarrow I_{syn} + w$



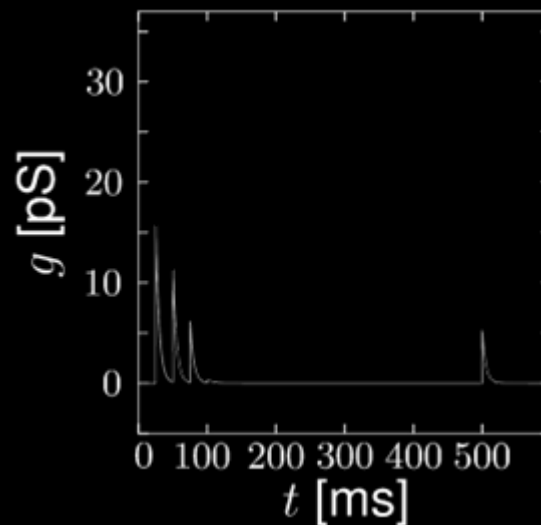
Where the **weight**, w , is positive for excitatory synapses and negative for inhibitory synapses.

Short-term plasticity (STP) of synapses

Synaptic facilitation



Synaptic depression



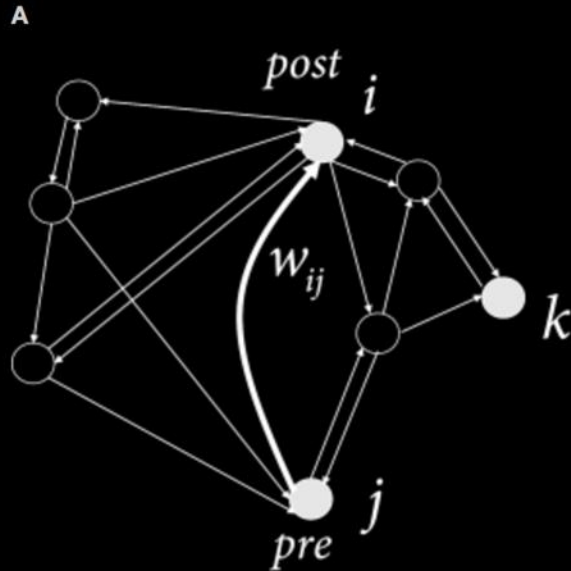
Long-term plasticity of synapses

Long-lasting changes of synaptic efficacies are neural correlates of **learning and memory** formation

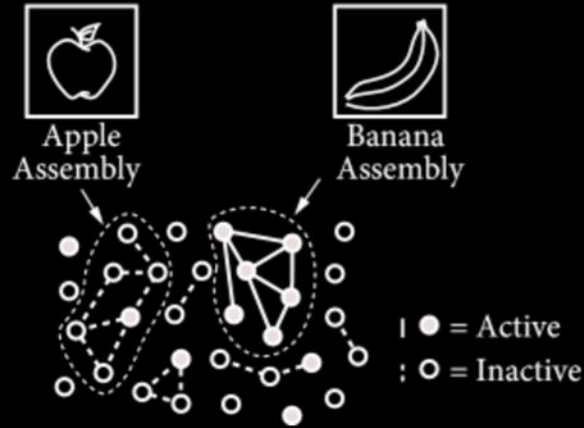
Long-term potentiation (LTP)

Long-term depression (LTD)

Hebbian Learning



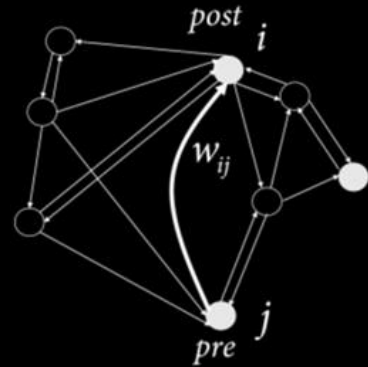
B “Neurons that fire together, wire together”



Locality
Joint activity

$$\frac{d}{dt} w_{ij} = F(w_{ij}; v_i, v_j) . \quad (19.1)$$

Stability of Hebbian learning



- Hebb's postulate

$$\frac{d}{dt} w_{ij} = F(w_{ij}; v_i, v_j) . \quad (19.1)$$

- Taylor expansion of F

$$\begin{aligned} \frac{d}{dt} w_{ij} = & c_0(w_{ij}) + c_1^{\text{pre}}(w_{ij}) v_j + c_1^{\text{post}}(w_{ij}) v_i + c_2^{\text{pre}}(w_{ij}) v_j^2 \\ & + c_2^{\text{post}}(w_{ij}) v_i^2 + \underline{c_{11}^{\text{corr}}(w_{ij}) v_i v_j} + \mathcal{O}(v^3) . \end{aligned} \quad (19.2)$$

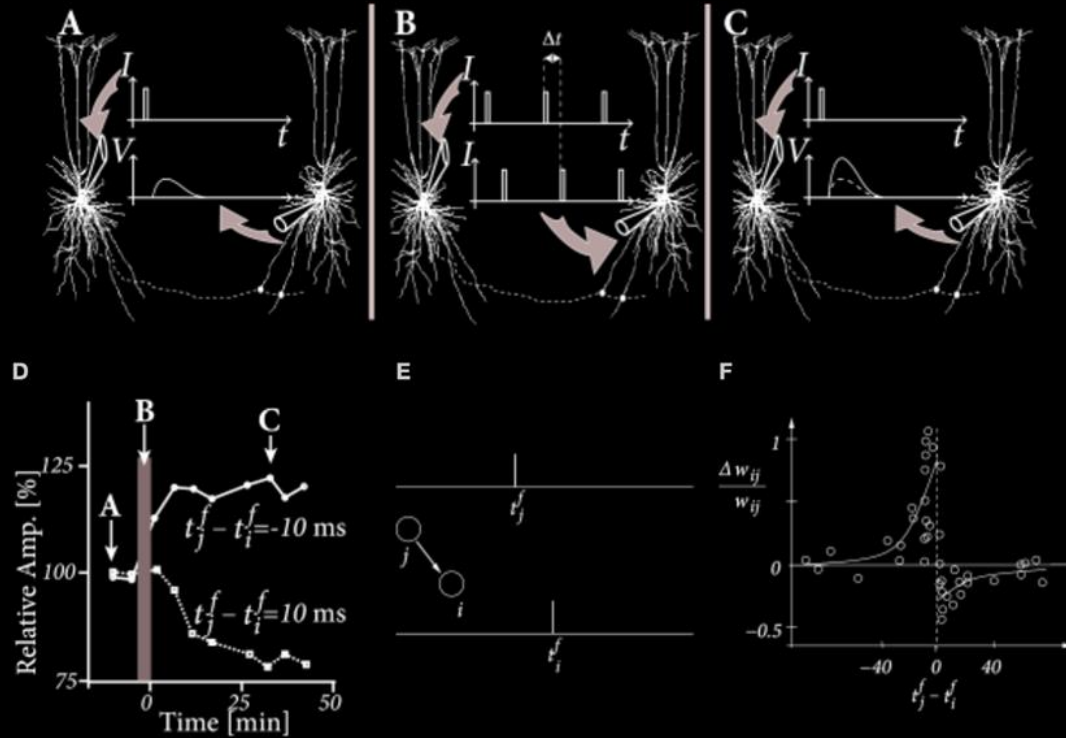
- The naive learning rule, problematic since w_{ij} cannot decrease

$$\frac{d}{dt} w_{ij} = c_{11}^{\text{corr}} v_i v_j . \quad (19.3)$$

- Stable learning rules by keeping additional terms, e.g., Oja's rule

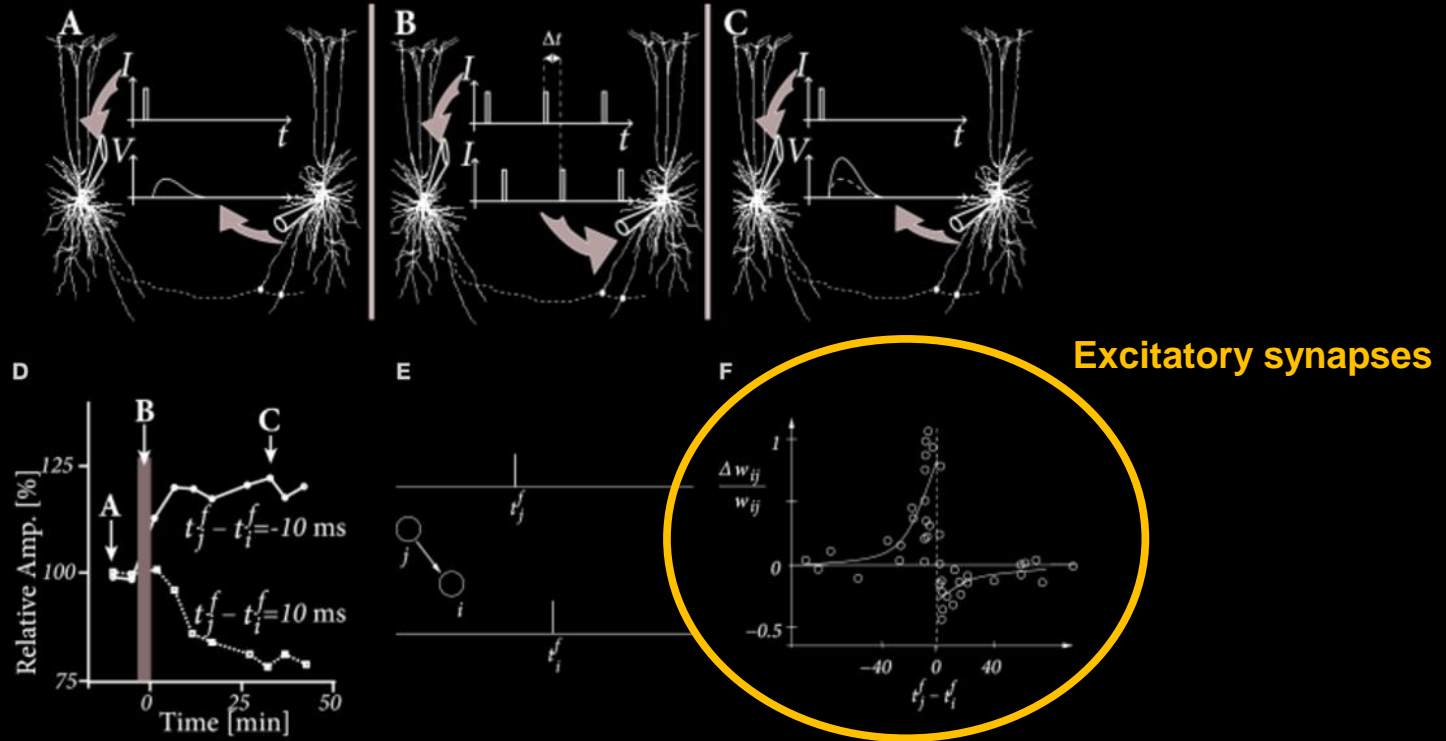
$$\frac{d}{dt} w_{ij} = \gamma [v_i v_j - w_{ij} v_i^2] \quad (19.7)$$

Spike-timing-dependent plasticity (STDP)



$$\begin{aligned} \Delta w_+ &= A_+(w) \cdot \exp(-|\Delta t|/\tau_+) & \text{at } t_{\text{post}} & \quad \text{for } t_{\text{pre}} < t_{\text{post}} \\ \Delta w_- &= A_-(w) \cdot \exp(-|\Delta t|/\tau_-) & \text{at } t_{\text{pre}} & \quad \text{for } t_{\text{pre}} > t_{\text{post}} \end{aligned} \quad (19.10)$$

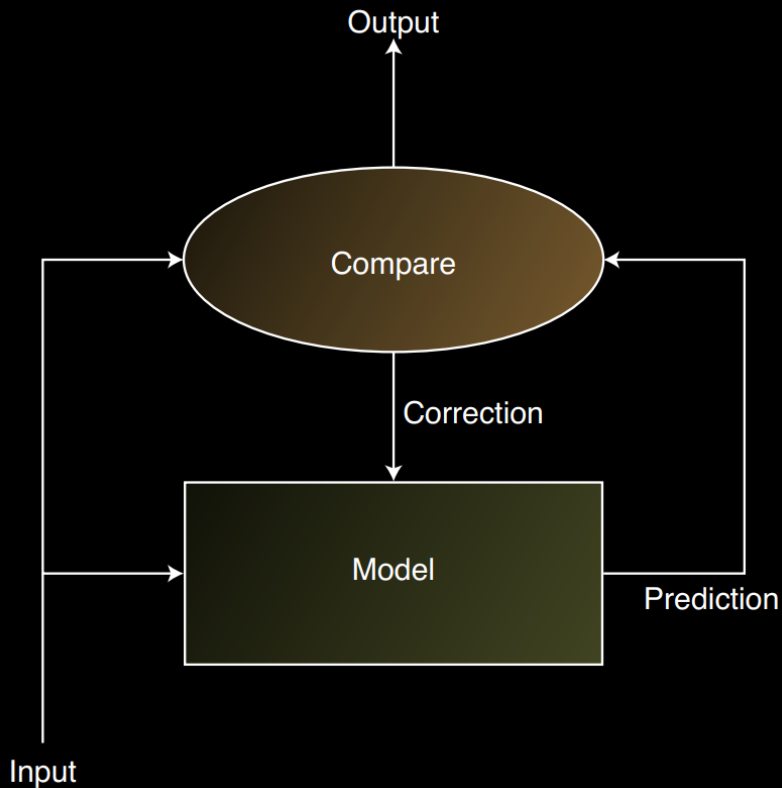
Spike-timing-dependent plasticity (STDP)



Excitatory synapses

$$\begin{aligned} \Delta w_+ &= A_+(w) \cdot \exp(-|\Delta t|/\tau_+) & \text{at } t_{\text{post}} & \quad \text{for } t_{\text{pre}} < t_{\text{post}} \\ \Delta w_- &= A_-(w) \cdot \exp(-|\Delta t|/\tau_-) & \text{at } t_{\text{pre}} & \quad \text{for } t_{\text{pre}} > t_{\text{post}} \end{aligned} \quad (19.10)$$

Next steps?



Online learning (backprop on stored data is too resource intensive)

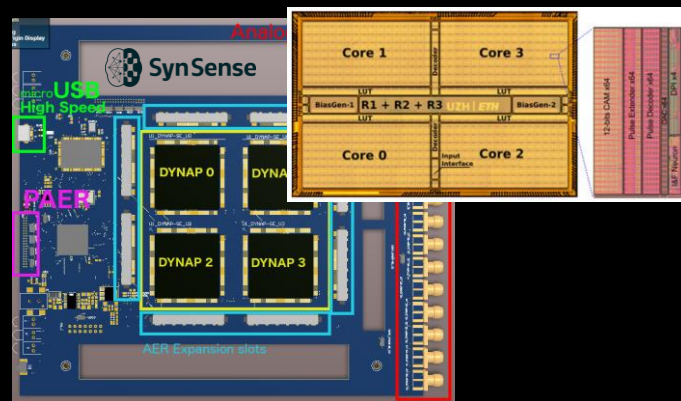
Dendrocentric learning (point neuron approximation is too simplistic)

Inhibitory and excitatory based learning

Examples, work on DYNAP-SE

DYNAP-SE mixed-signal processor

- 4 chips each having
 - 1k neurons (AdEx)
 - 64k dynamic synapses (4 types, DPI)
 - Programmable connectivity via CAM
 - 25 bias-parameters/core (4 cores/chip)



Moradi, Qiao, Stefanini, Indiveri (2018). doi: 10.1109/TBCAS.2017.2759700

Chicca, Stefanini, Bartolozzi, Indiveri (2014). doi: 10.1109/JPROC.2014.2313954

Comparison of chips: Basu, Deng, Frenkel, Zhang (2022). doi: 10.1109/CICC53496.2022.9772783

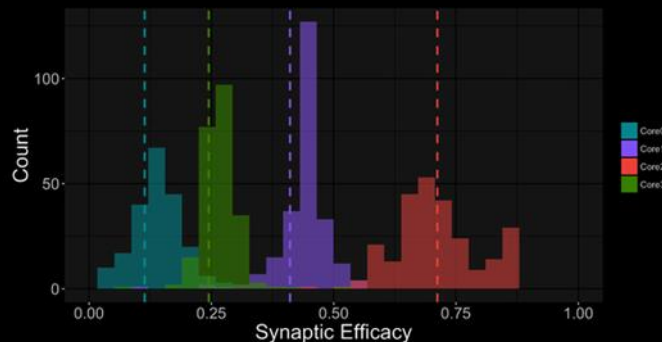
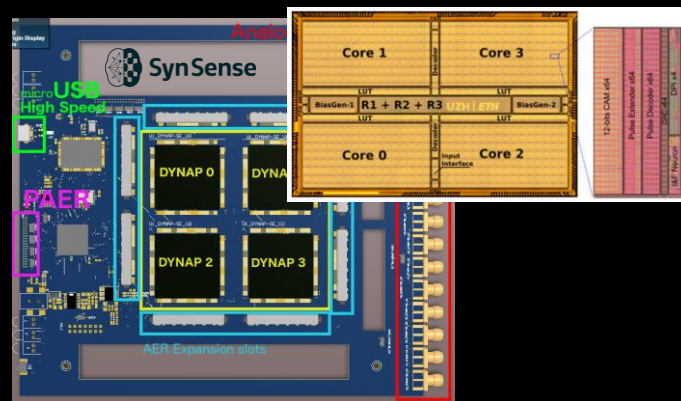
DYNAP-SE mixed-signal processor

4 chips each having

- 1k neurons (AdEx)
- 64k dynamic synapses (4 types, DPI)
- Programmable connectivity via CAM
- 25 bias-parameters/core (4 cores/chip)

Analog neuron & synapse circuits
→ distributions of parameter values

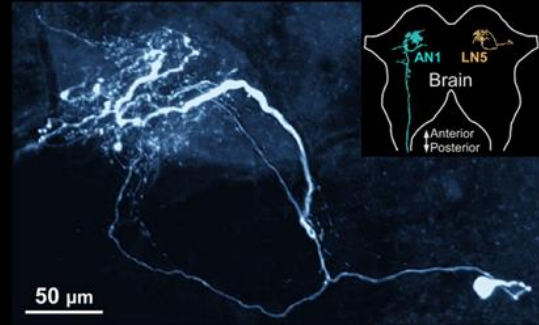
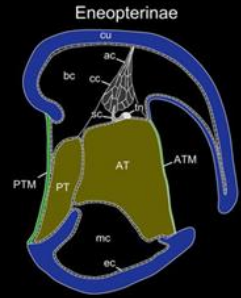
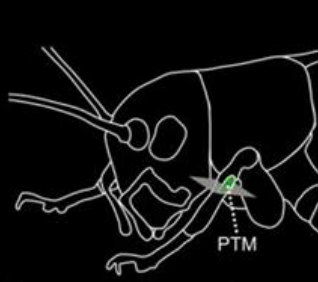
Digital spike-event communication



Moradi, Qiao, Stefanini, Indiveri (2018). doi: 10.1109/TBCAS.2017.2759700
Chicca, Stefanini, Bartolozzi, Indiveri (2014). doi: 10.1109/JPROC.2014.2313954

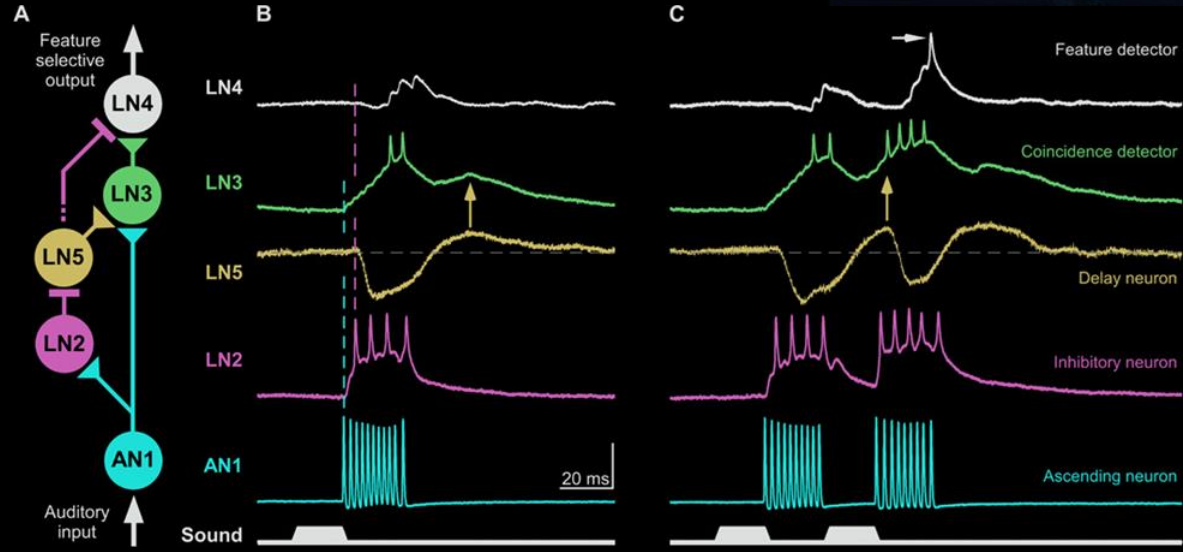
Comparison of chips: Basu, Deng, Frenkel, Zhang (2022). doi: 10.1109/CICC53496.2022.9772783

Cricket auditory feature detector

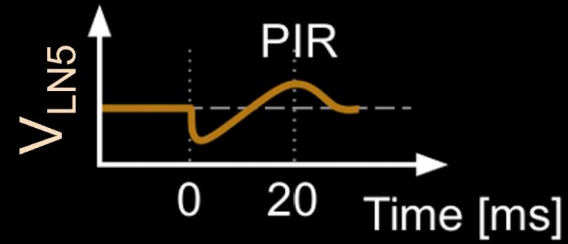
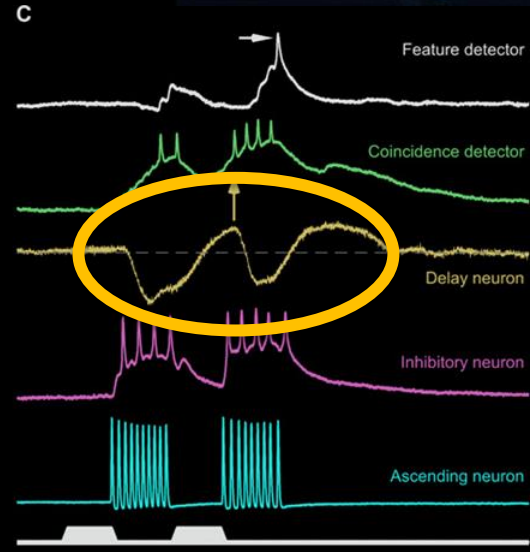
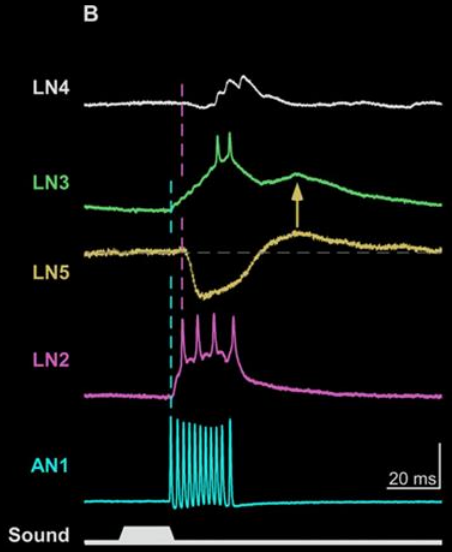
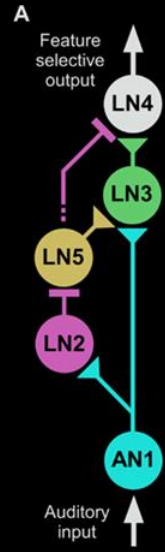
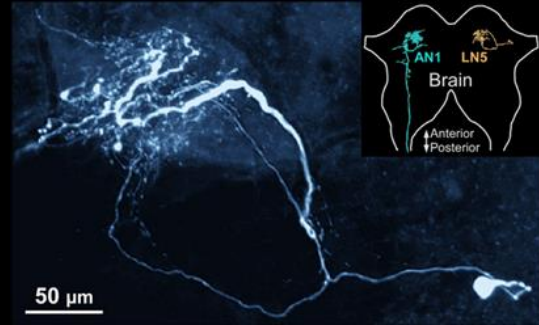
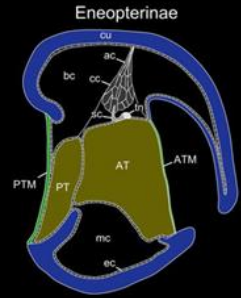
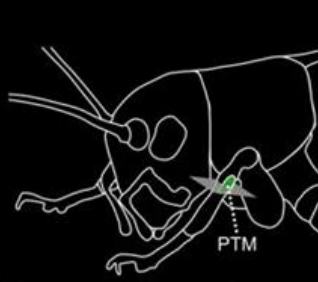


www.youtube.com/watch?v=Pb8vhhbLwBM

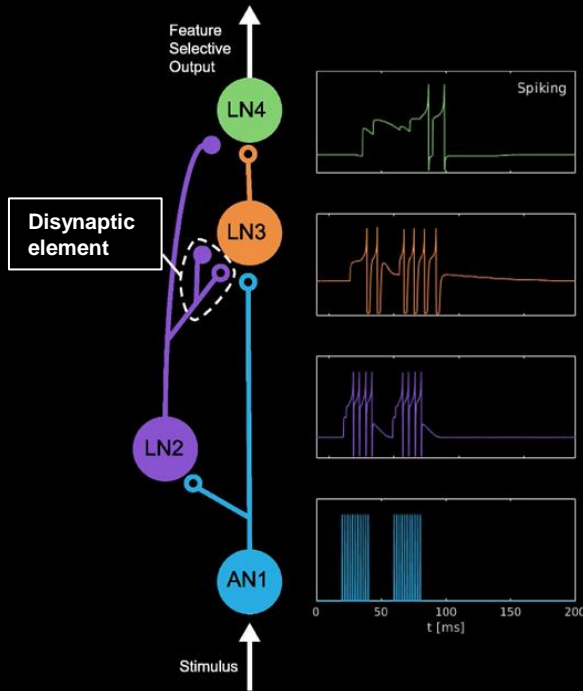
<https://doi.org/10.1038/s41598-017-15282-z>



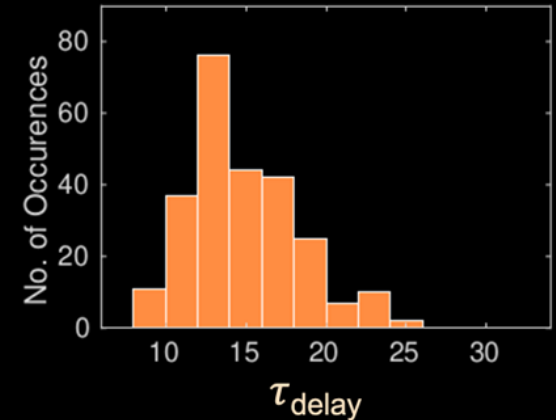
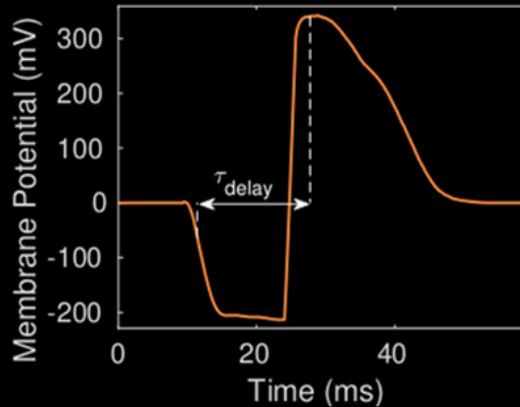
Cricket auditory feature detector



Configuration of cricket circuit in DYNAP-SE

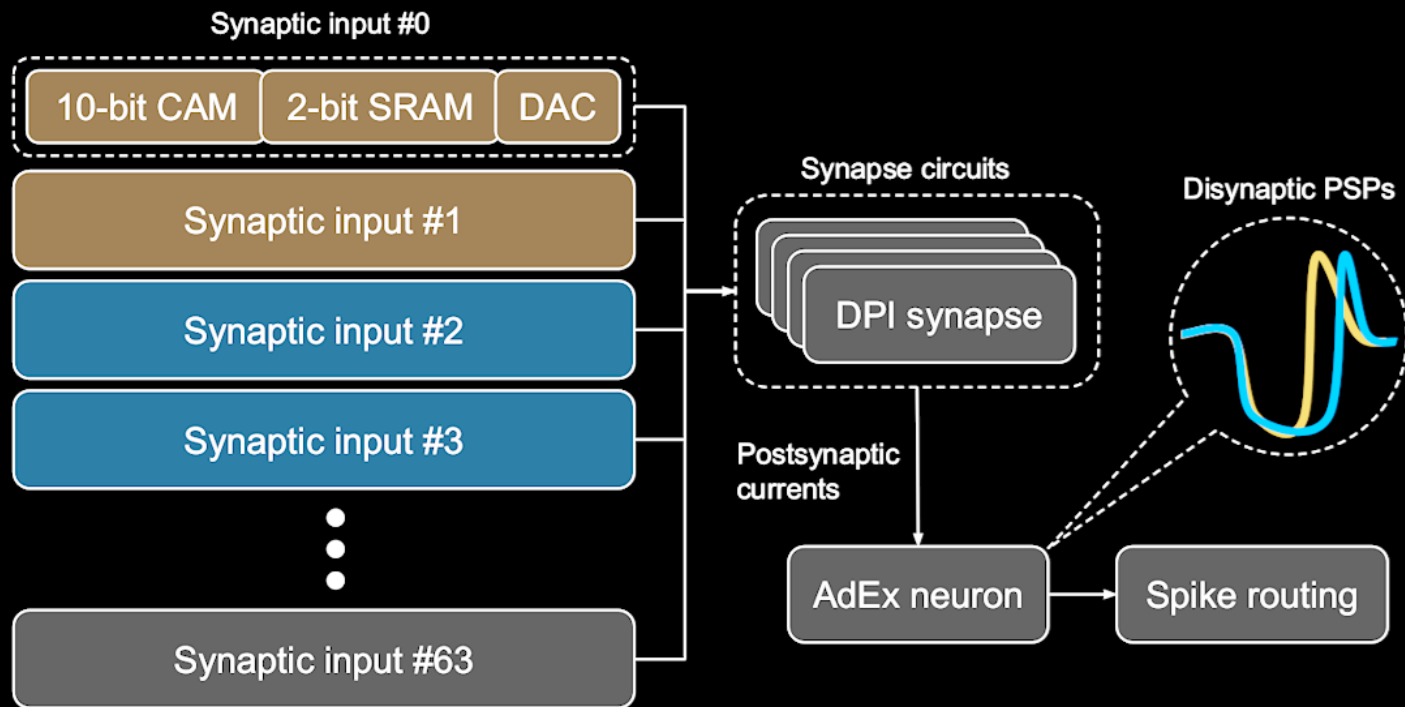


Dynamics of non-spiking LN5 approximated with two synapses (disynaptic element)



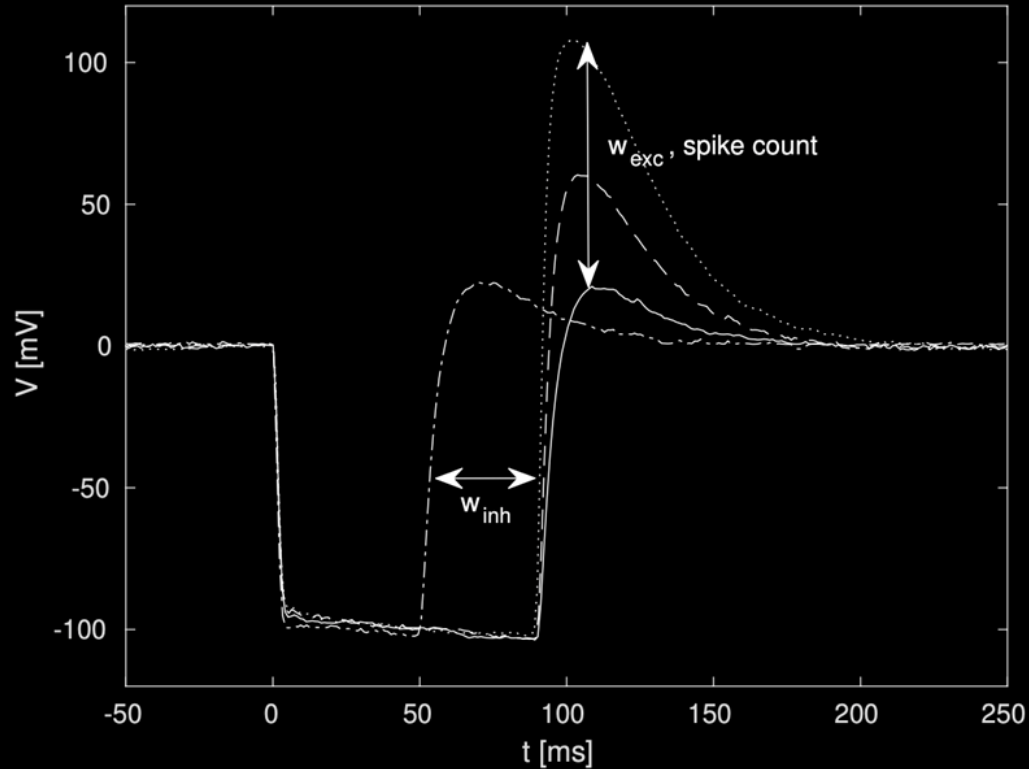
Balance of excitatory and inhibitory postsynaptic currents (via bias tuning & Hebbian learning protocol)

Disynaptic configuration in DYNAP-SE



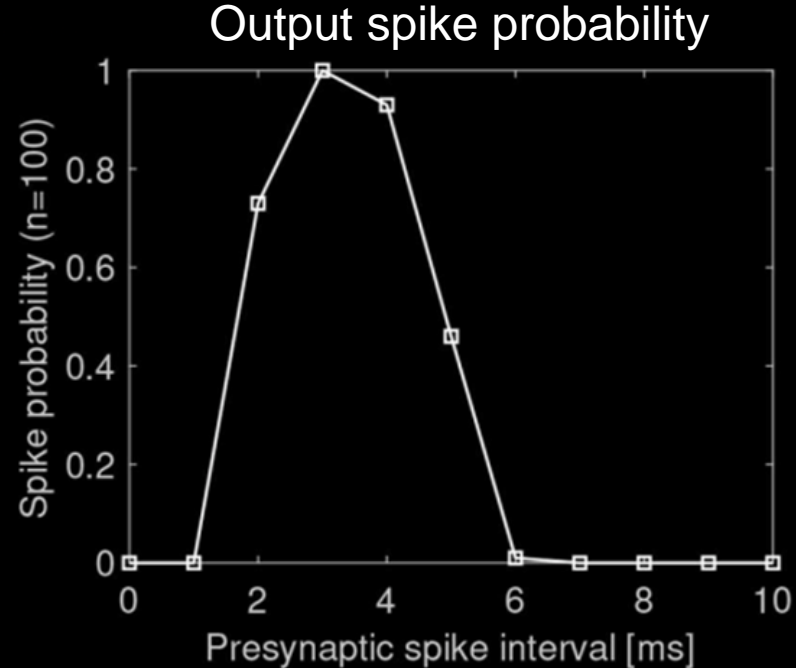
Different PSPs due to **DAC** device-to-device mismatch

Control of PSP via disynaptic parameters



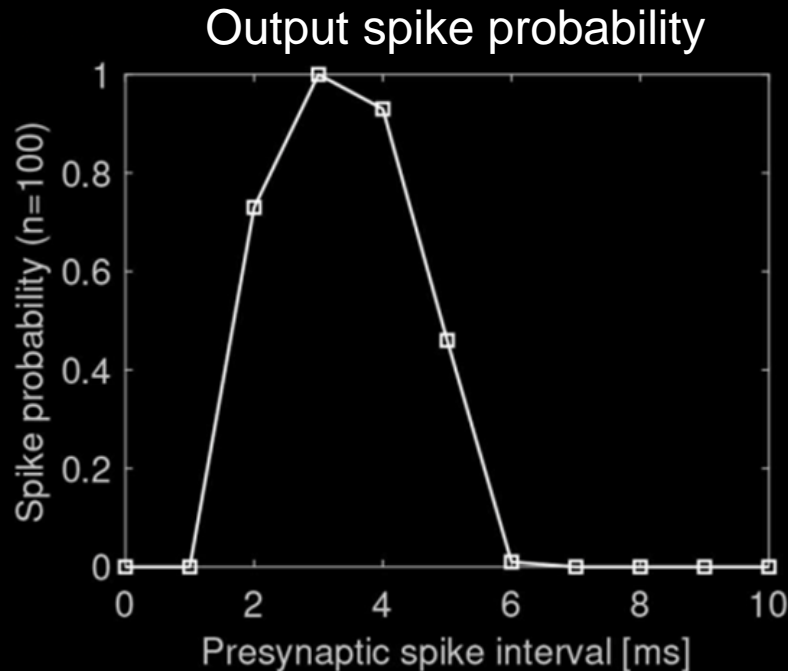
Spike-pair selectivity with one neuron

Single neuron with two disynaptic inputs



Spike-pair selectivity with one neuron

Single neuron with two disynaptic inputs

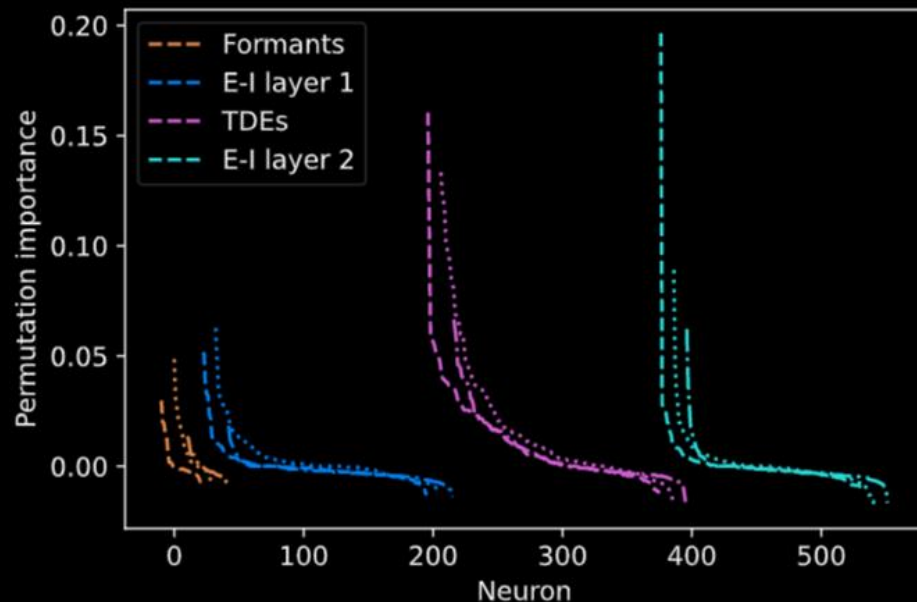
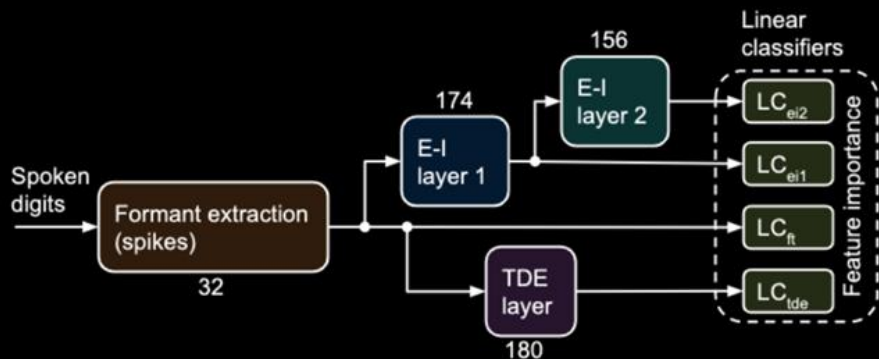
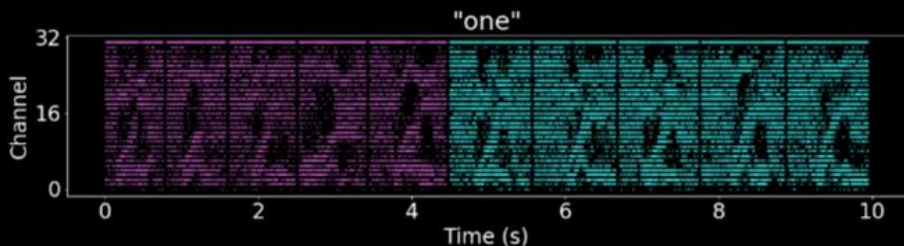


Sandin & Nilsson, Frontiers in Neuroscience (2020). doi: 10.3389/fnins.2020.00150

For spike triplets: Nilsson, Liwicki and Sandin, IJCNN (2020). doi: 10.1109/IJCNN48605.2020.9207210

More general: Nilsson, Liwicki and Sandin, ICONS (2022)

Keyword Spotting with Few (~10) Neurons



Work in progress

Optimize vibration monitoring system design

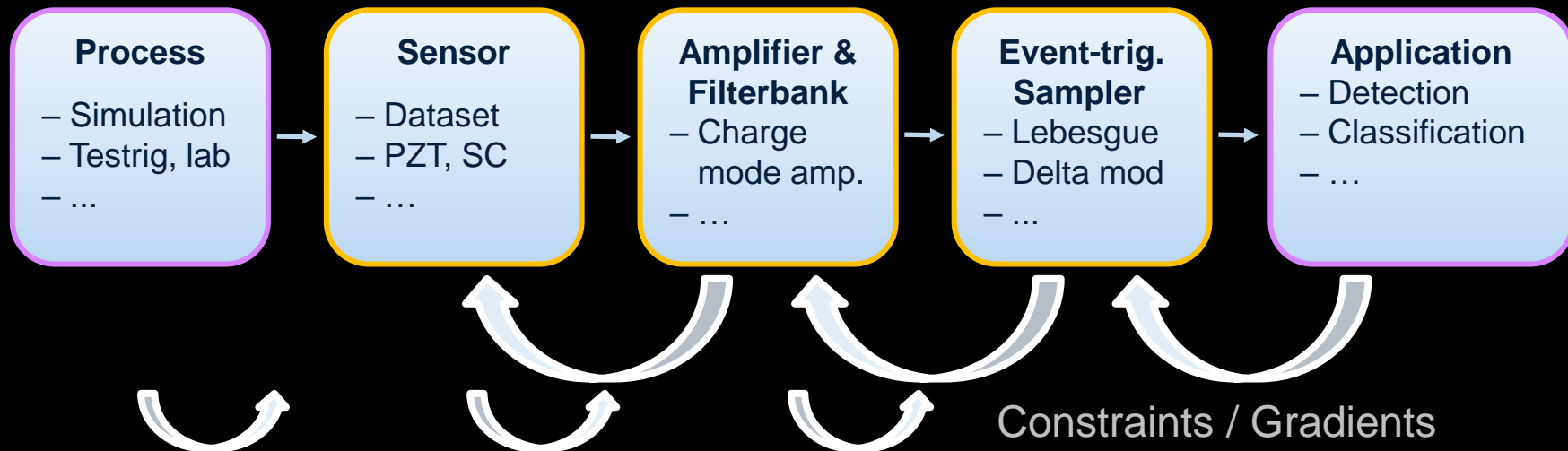
Modular software for optimization of the sensor, filterbank, amplifier and event-triggered sampler



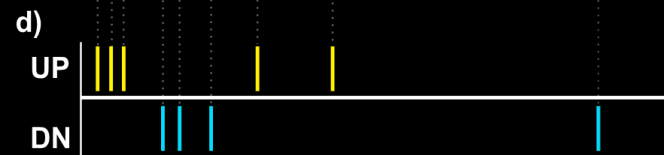
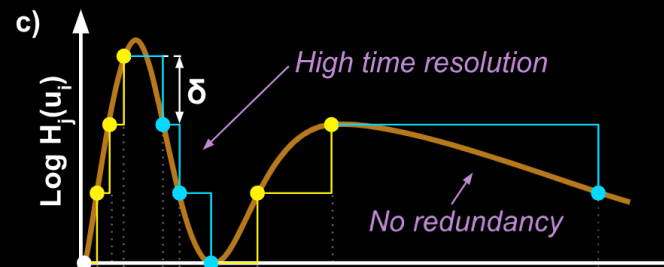
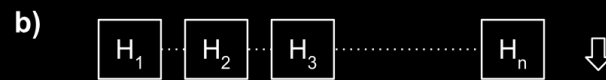
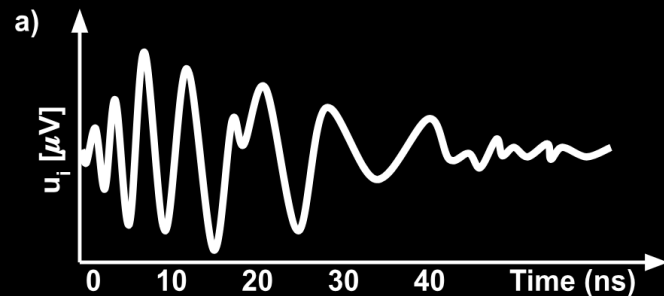
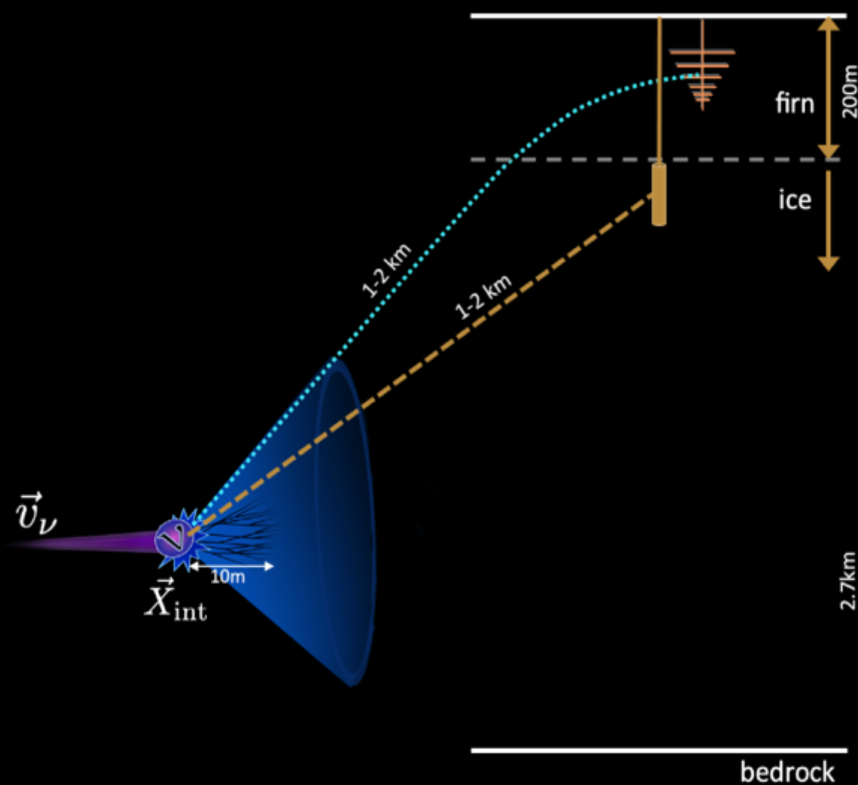
Daniel



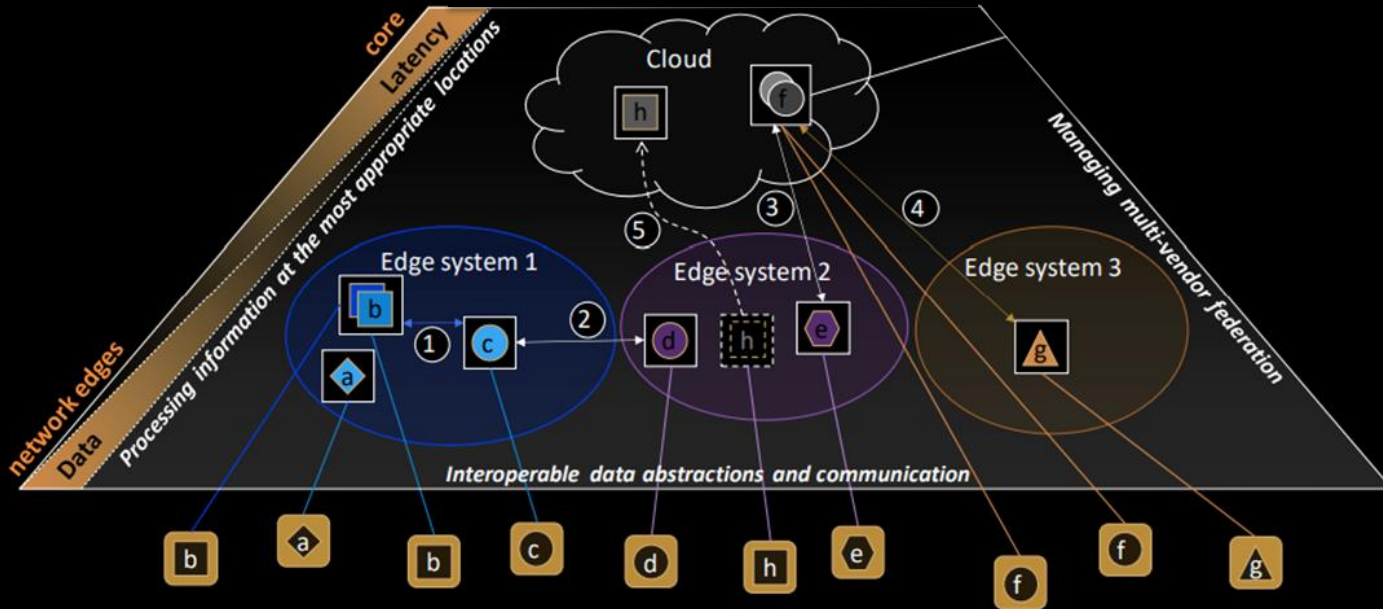
Ashwani



Proposal with Christian Glaser (Uppsala) & Tommaso: Improving radio-detection methods for neutrino astronomy



Cloud-to-Edge Computing Continuum Optimization



Proposal under review

Related paper: <https://doi.org/10.3389/fnins.2023.1074439>



Acknowledgements

STINT (IG2011-2025)

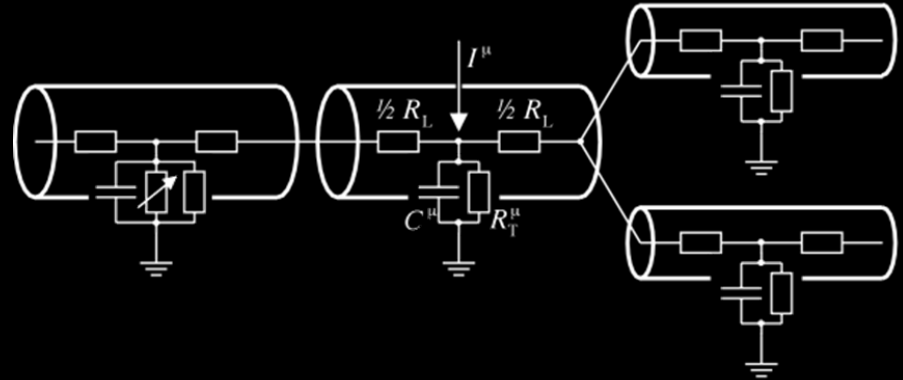
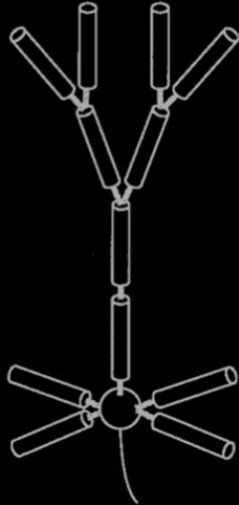
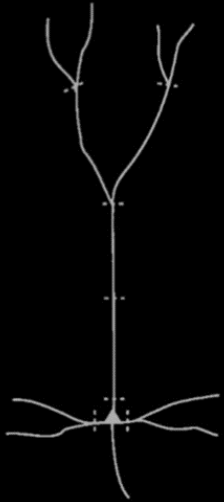
The Kempe Foundations (SMK1429, JCK-1809, SMK21-0046)

ECSEL JU (Arrowhead Tools, 737459)

LTU Jubilee Fund & Creaternity

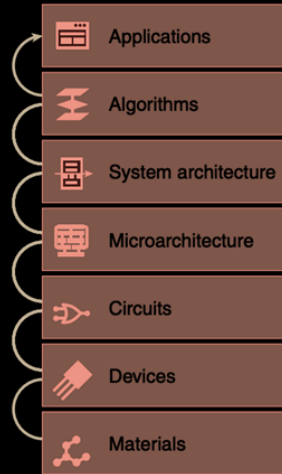
CapoCaccia Neuromorphic Engineering Workshop

Dendrites (compartment models), AdEx, ...

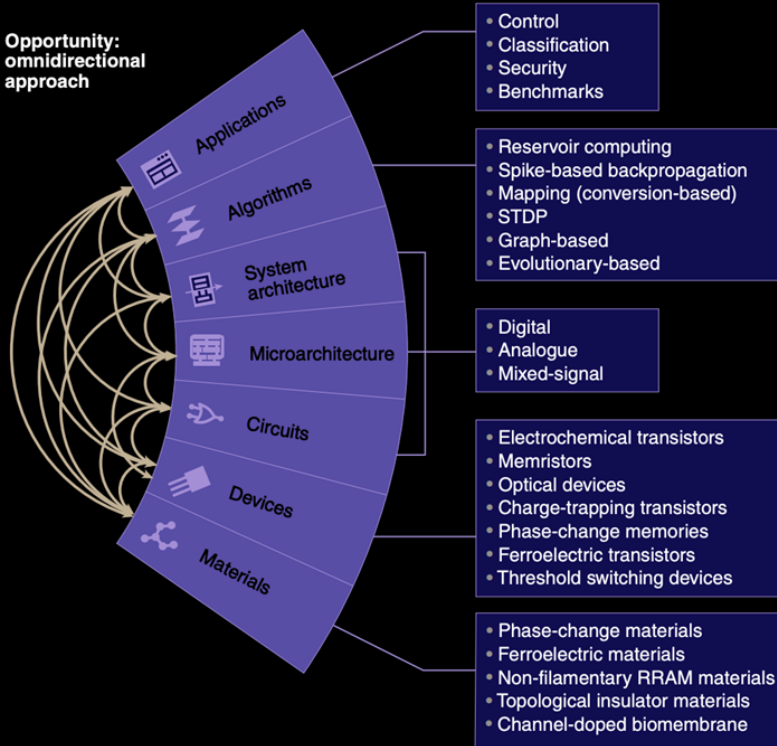


Neuromorphic Compute Stack Co-design

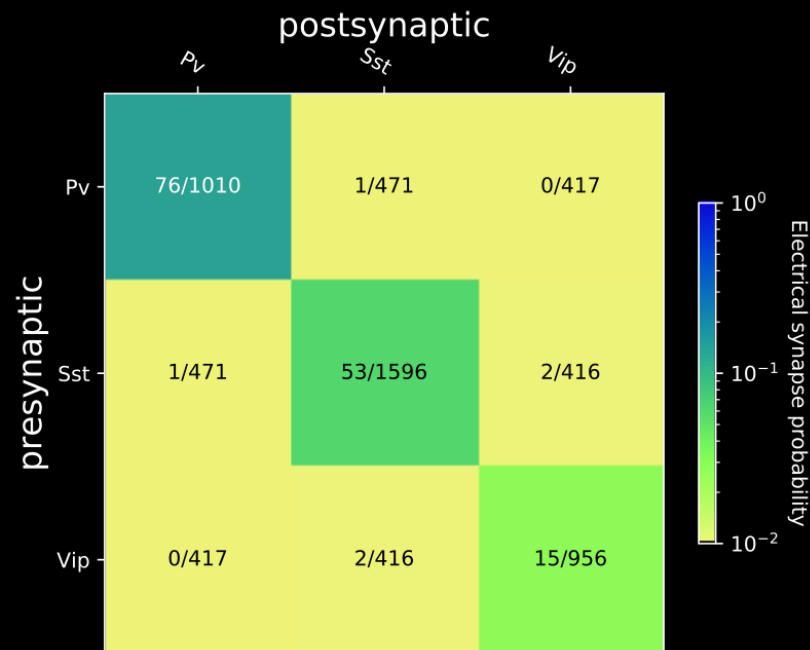
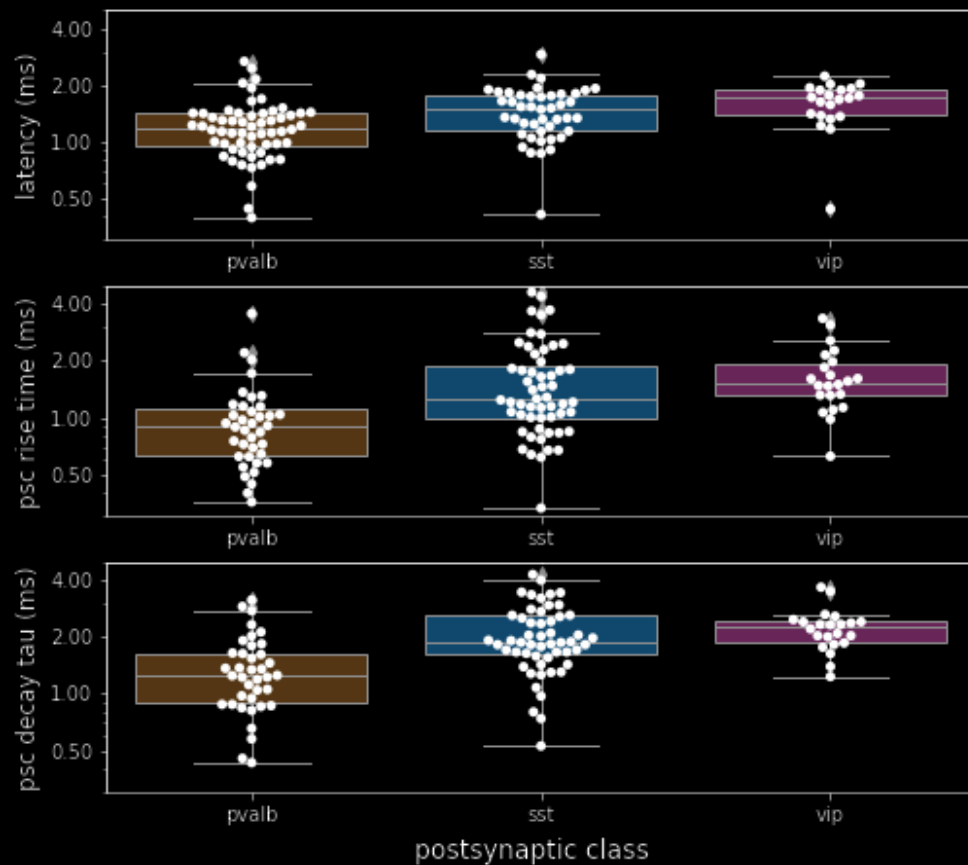
**State of the art:
bottom-up approach**



**Opportunity:
omnidirectional
approach**



Synaptic Physiology Dataset, three inhibitory interneuron subclasses (Pv, Sst, and Vip)



Idea of Tommaso Dorigo: High-Granularity Hadron Calorimeters with Embedded Neuromorphic Computing

Charged pions, kaons, and protons constitute the bulk of the hadrons flowing into a hadron calorimeter

Being able to distinguish them would bring in **very large gains**:

- to flavour tagging (killer app: $H \rightarrow ss$ at a future collider, where you need to tag the fast kaon from s hadronization)
- to energy reconstruction (improved through particle flow techniques)
- to boosted-jet tagging (from improved inner structure reconstruction of jet cores)

