A Neurally controlled Robot that learns

MSc Thesis by: Benjamin Walther Büel

22. Sept. 2011

Problem statement

”Understand if and how learning processes driven by dopamine modulated STDP can enable specific behaviours through controlled motor movements in a Braitenberg inspired mobile robot”

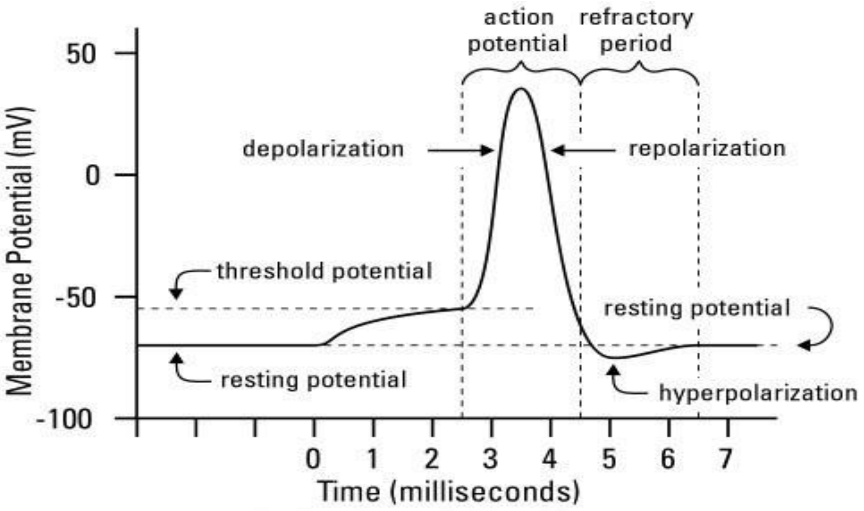
Outline

- Background
- Model
- Results & Demo

Spiking neural networks

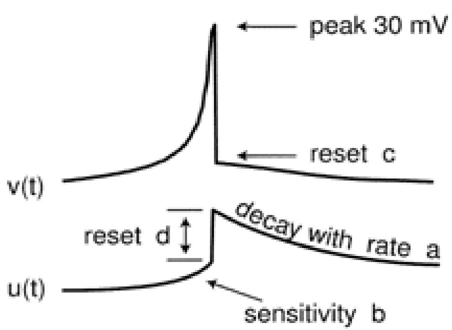
- Biological inspired neural networks : neurons + connectivity
- Focus on understanding the biological foundations, vs. solving a concrete AI-problem
- Incorporate spatial-temporal dynamics

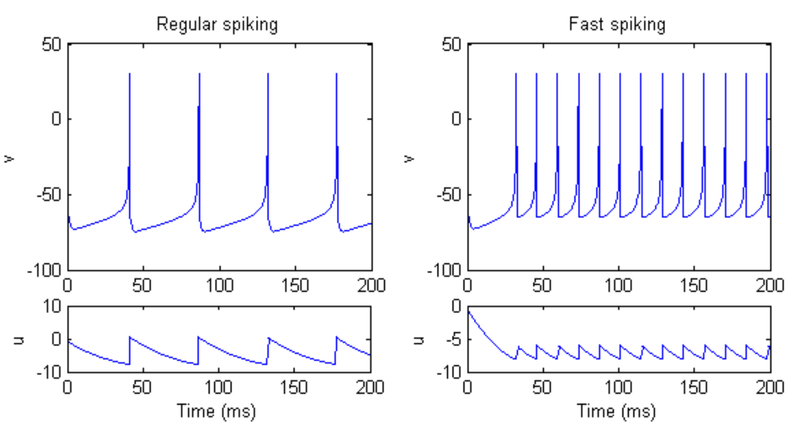
Action potential of a neuron



Neuron Model

- Izhikevich model (2003)
- $\dot{v} = 0.004v^2 + 5v + 140 - u + I$
- $\dot{u} = a(bv - u)$
- $if\ v \geq 30\ then \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases}$





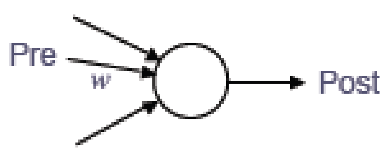
Plasticity

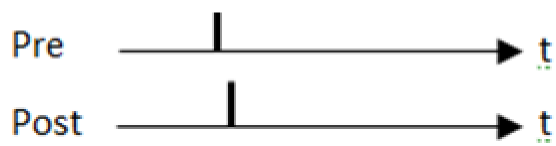
How the connections between neurons change

- Hebbian/correlation based: STDP: fosters competition, tendency to destabilize.
- Homeostatic plasticity: compensate for chronic changes, stabilizing form of plasticity.
  - Synaptic scaling: scale all weights per synapse according to target firing rate
  - Intrinsic plasticity: change the excitability of a synapse depending on postsynaptic firing (change the transfer function)
- Short term plasticity
- Structural plasticity
- Meta plasticity

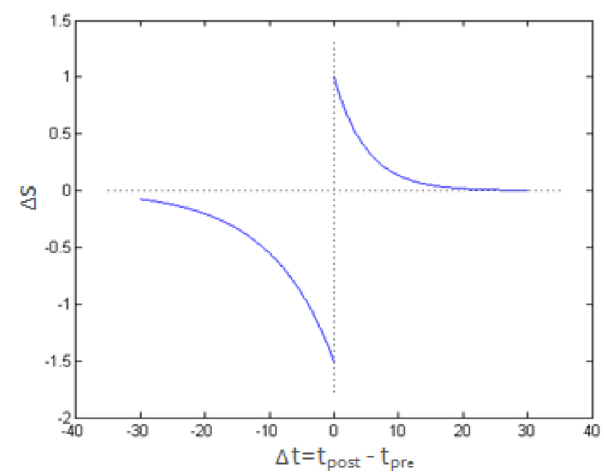
Spike time dependent plasticity (1)

- Temporal correlation of spikes determine the strengthening or weakening
- Hebbian learning rule: “neurons that fire together, wire together”
- At the molecular level: receptor channel is blocked (by Mg^{2+}) until postsynaptic neuron is activated





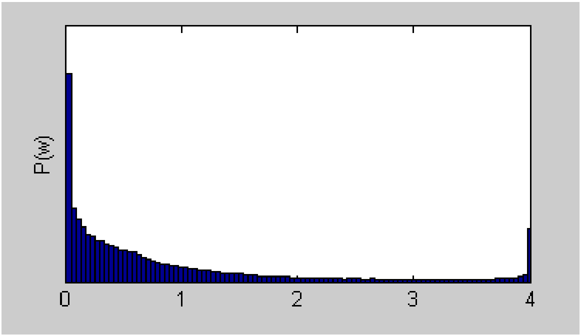
$$\Delta w = \begin{cases} A_+ \exp(-\Delta t/\tau^+) \;\; if \; \Delta t > 0 \\ A_- \exp(\Delta t/\tau^-) \;\; if \;\; \Delta t < 0 \end{cases}$$



Spike time dependent plasticity (2)

- linear relationship between the strength of a synapse and the probability of a postsynaptic spike
- Need to limit synaptic strength growth
- Hard bounding strength limit leads to bimodal soft bounding leads to unimodal bell-shaped distribution

Bimodal weight distribution



Dopamine modulated STDP (1)

- Dopamine (DA) regulates reward and learning among other things
- Unexpected rewards trigger large amounts of DA
- Temporal difference prediction error in reinforcement learning
- DA modifies the synapses to modulate STDP

Dopamine modulated STDP (2)

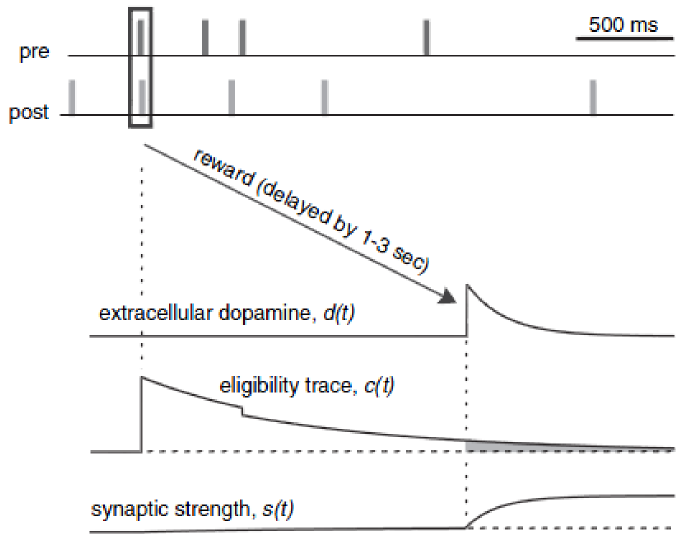
- Eligibility traces: solve the credit assignment problem

$$\Delta w_{ji}(t) = c_{ji}(t)d(t)$$

$$\dot{c_{ji}} = -c_{ji}/\tau_c + STDP(t_{post} - t_{pre})$$

$$\dot{d} = -d/\tau_d + DA(t)$$

$$DA(t) = \begin{cases} 0.01\,\mu M/s + 0.5\mu M & if\ reward \\ 0.01\,\mu M/s & else \end{cases}$$



3 decay parameters:

- STDP window decay: tau=20ms
- Eligibility trace decay: 1‰/ms
- Dopamine decay: 5‰/ms

Dopamine modulated STDP (3)

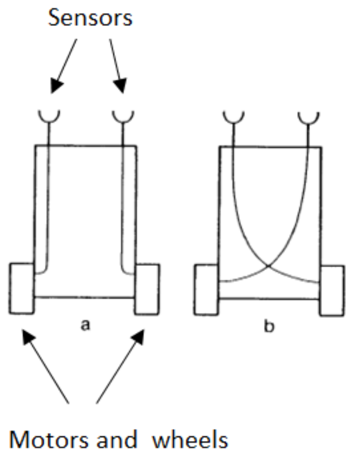
- Many pairs of neurons fire nearly coincident
- But firing is random, cancels out over time
- Important functional role of noise: maintain suitable level of spontaneous firing
  - Too low: Neurons cannot find out if it would be rewarded when they never fire
  - Too high: Probability high that sequential spike pairs fall into STDP window by chance; neurons only fire because of noise

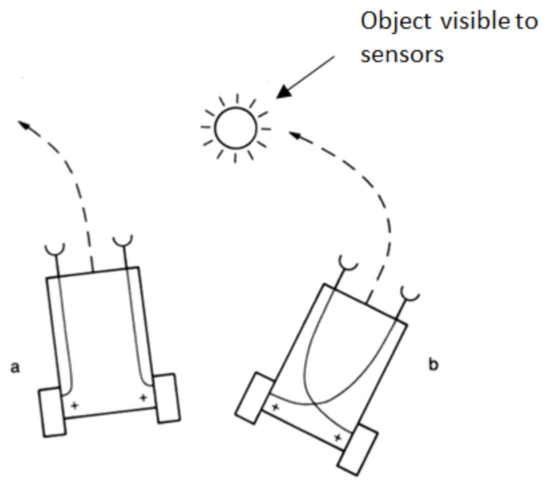
Neural coding

- How to encode stimulus from environment into spike trains, and decode spike trains into motor actions or other output?
- Possible schemes: rate coding, temporal coding, rank order coding, population coding
- 2 key problems:
  - Noisy neural responses
  - Noisy stimulus

Braitenberg vehicles

- Vehicles: two wheeled differential drive robots
- Controller is simple neural network
- Seemingly simple internal structures can exhibit surprisingly complex behaviour

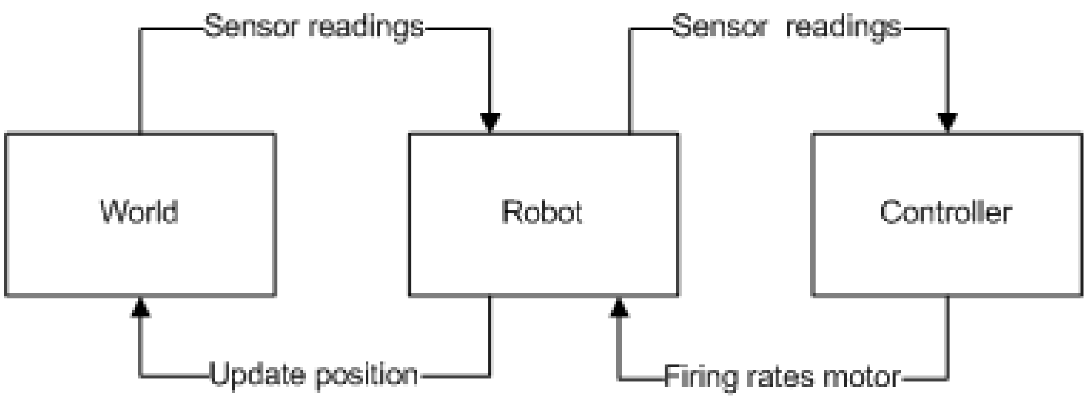




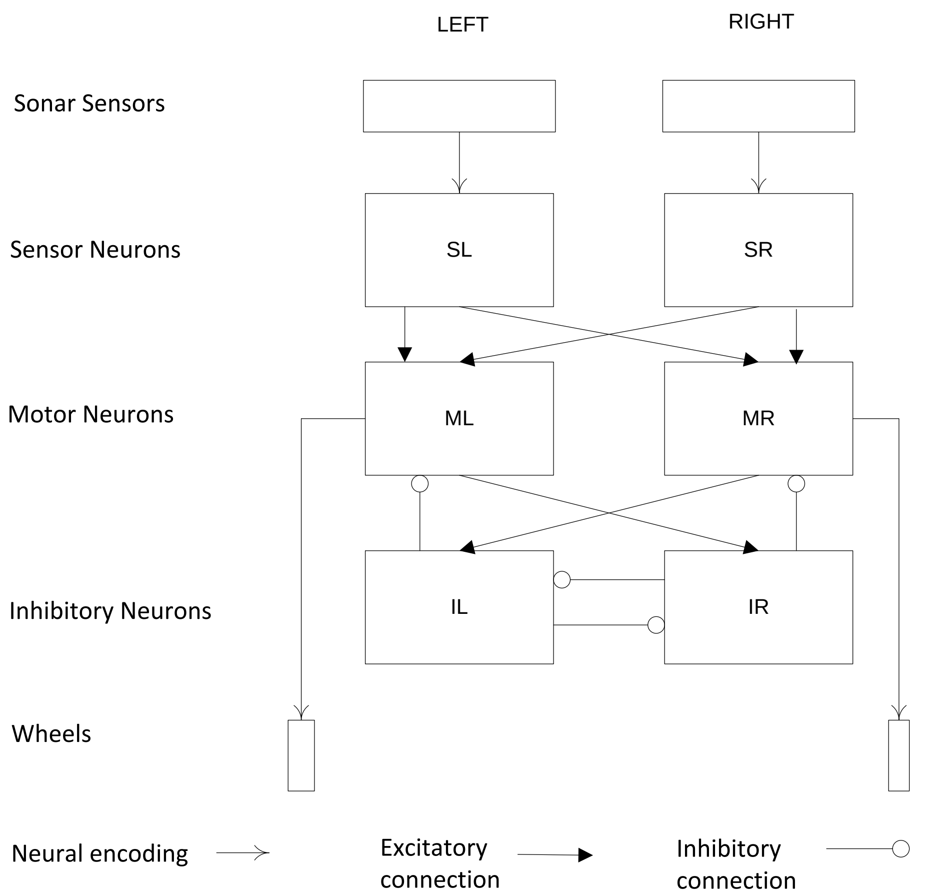
Model

Model

- Braitenberg-inspired mobile robot with SNN controller
- SNN controller with DA-modulated STDP

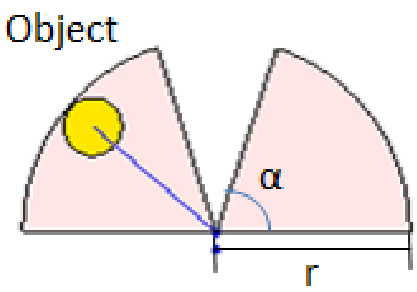


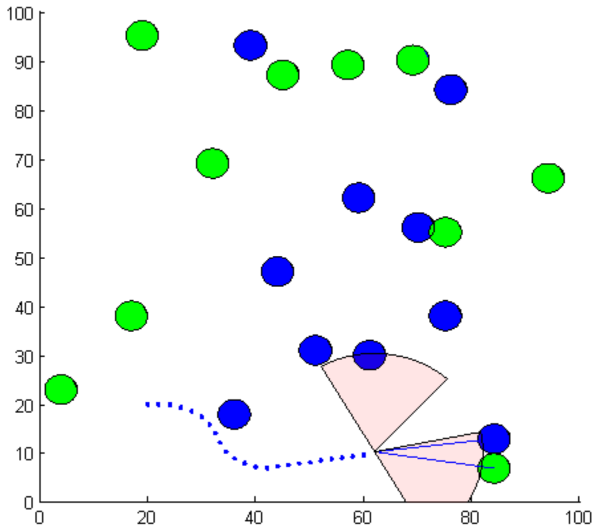
Controller Architecture



Robot and Environment

- 1 sonar left, 1 sonar right
- Environment is contionous torous (without walls) containing resources of 2 types: food(green) or/and obstacles (blue)





Encoding

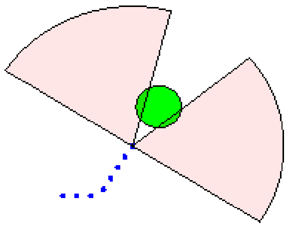
- Sensor readings to spike patterns
  - Poisson input stream with λ inversely proportional to distance measured; constant when very close (<20% of sonar distance)
- Spike patterns to motor commands
  - Rate coding, mean firing rate

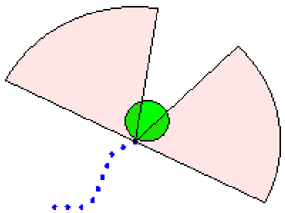
$$v = U_{min} + r(U_{min} - U_{max}) \qquad r = {}^{MFR}/_{MFR_{max}}$$

$$MFR_{t,max} = \max(MFR_{t-1}, MFR_{t-2}, \dots, MFR_{t-50})$$

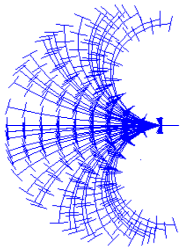
Significance of Encoding Parameters

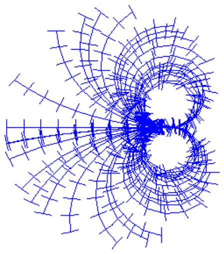
- Sonar angle





- Relation U_{min}/U_{max}: steering angle
- MFR_{max}





Training

- Instrumental conditioning
- Directed: Inducing external training current
- Dynamic: Adapt training current to learning progress
- Isolated 1 obstacle vs. Random walk vs. Random walk with «curiosity» training

Convergence/when to stop

- Linear relationship between orientation change (per Δt) and wheel velocity difference:

$$\Delta\theta = \frac{(v_R - v_L)\Delta t}{W}$$

- Linear relationship between wheel velocity and mean firing rate:

$$v = U_{min} + r(U_{min} - U_{max})$$

-> Linear relation between orientation change and mean firing rate difference

- Stop condition using mean firing rate difference:

$$mean(s(SLMR)) > \lambda * mean(s(SLML)$$

$$mean(s(SRML)) > \lambda * mean(s(SRMR))$$

Results and Demo

Parameter effects

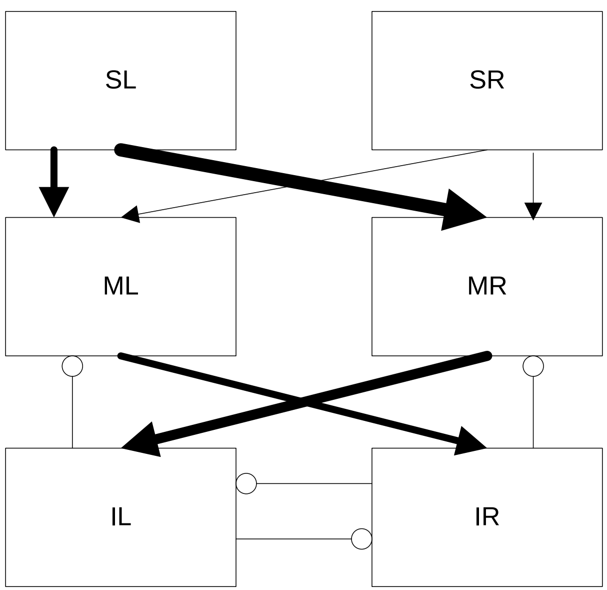
- Large sonar angle causes oscillations
- Correct amount of background noise relative to training current
- Inhibitory neurons affect sensor to motor connections in random ways at times
- Miniumum network ~1000 neurons

<u>Finding 1:</u> Plasticity between motor neurons and inhibitory neurons is significant

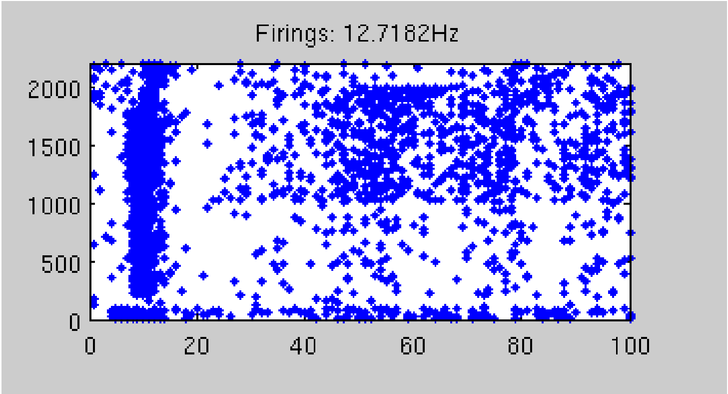
<u>Finding 2:</u> Activity dependent scaling induces stability and prevents early failure

<u>Finding 3:</u> Level of STDP learning rate has dramatic effects on dynamics

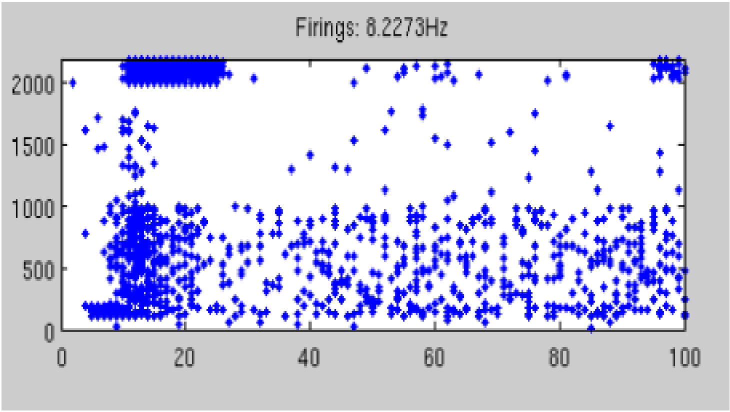
Significance of plasticity between motor neurons and inhibitory layer



Without plasticity

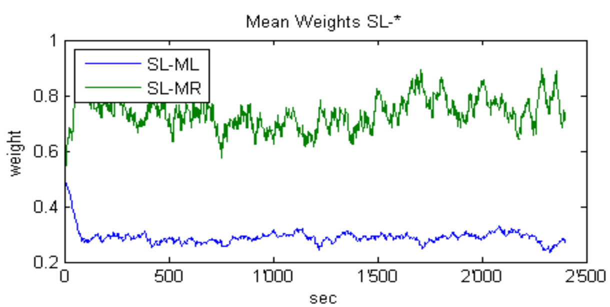


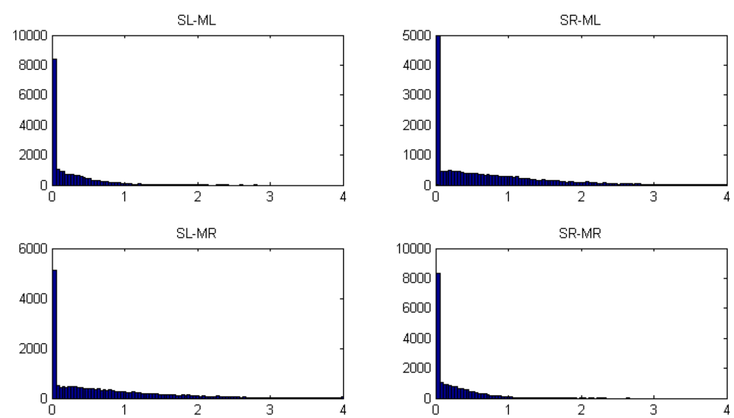
With plasticity



Activity dependent scaling effects

- Modulate the excitability of a neuron towards a target firing rate
- Or a group of neurons: Normalise synaptic strengths with a fixed weight sum



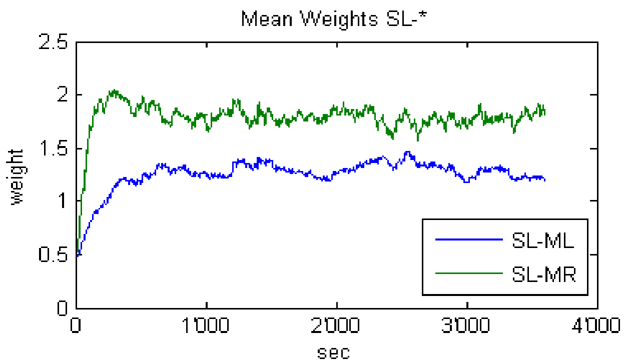


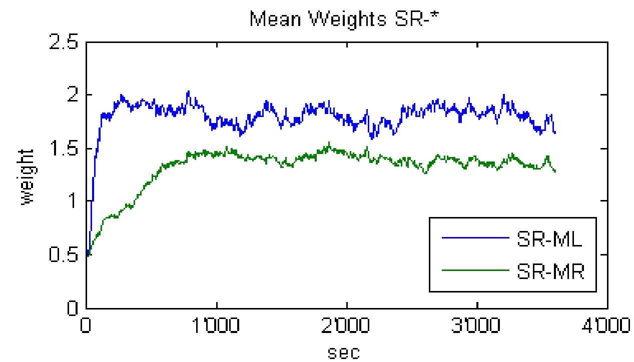
Learning rate effects

$$\Delta w = \begin{cases} A_+ \exp(-\Delta t/\tau^+) \ \ if\ \Delta t > 0 \\ A_- \exp(\Delta t/\tau^-) \ \ if\ \ \Delta t < 0 \end{cases}$$

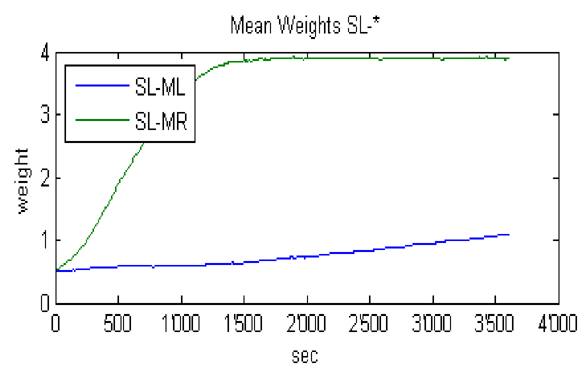
$\tau^+ = \tau^- = 20$ms

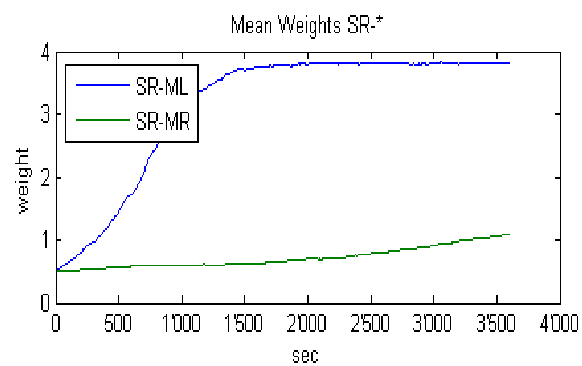
$A_+=0.1$; $A_-=-0.15$





$A_+=0.01$; $A_-=-0.011$





Behaviours

Directed training

- ✓ Attraction
- ✓ Avoidance
- ✓ Attraction and avoidance

Random walk:

- ✓ Attraction
- ✓ Avoidance
- ○ Attraction and avoidance

Stats

Attraction behaviour

	mean	std.dev
Random walk	99.5	10.3
Benchmark	371.2	7.63
Training directed	369.1	8.87
Training rdwk	301.8	6.87

10 trials, 15 objects each type, 10 min simulation time

Repulsive behaviour

	mean	std.dev
Random walk	99.5	10.3
Benchmark	36.2	3.63
Training directed	37.2	3.8
Training rdwk	54.9	4.87

Attraction and repulsive behaviour

	Rewards		Punishments	
	mean	std.dev	mean	std.dev
Random walk	145.5	20.3	152.4	17.2
Benchmark	243.2	31.89	40.8	5.6
Training rdwk	164.75	30.8	131.5	12.2

10 trials, 10 objects each type, 30 min simulation time

Directed training:

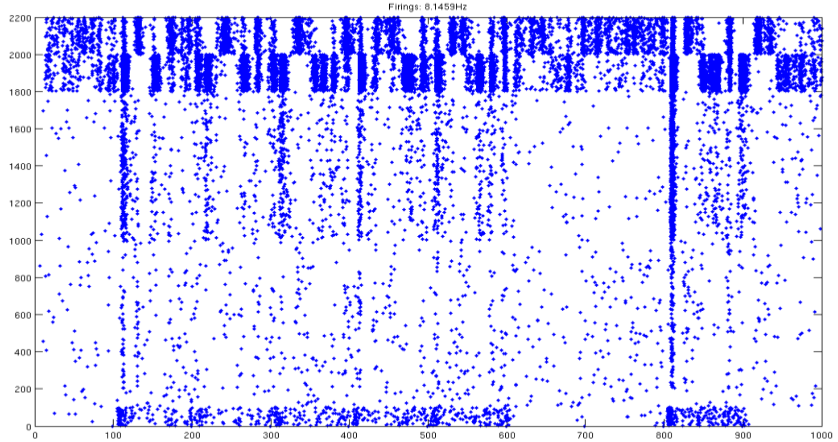
Behaviour	λ factors
Attractor	$SLMR/_{SLML} = 7.15$
Attractor	$SRML/_{SRMR} = 7.08$
Repulsive	$SLML/_{SLMR} = 2.15$
Repulsive	$SRMR/_{SRML} = 2.44$

Summary

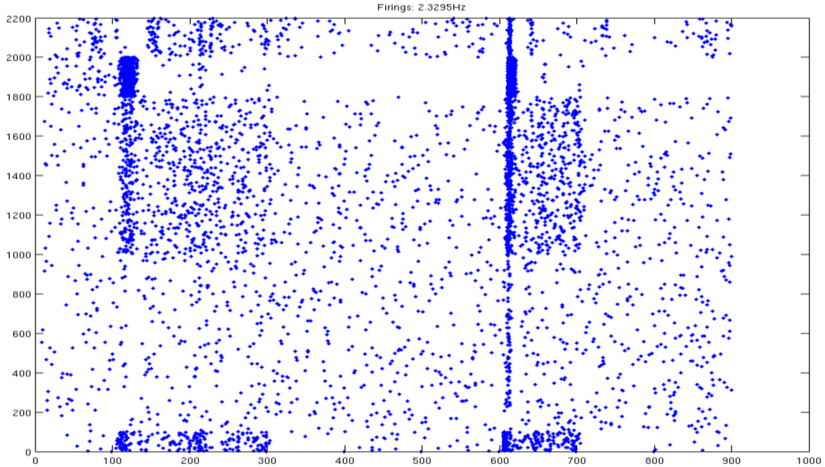
- Robot controlled by Spiking neural network subject to DA-modulated STDP is able to successfully learn autonomously in a previously unknown environment
- Such learning without imposing neuro-anatomical constraints has been a problem addressed in the literature, see Chorley (2008)
- Dedicated training phase quickly enables robot to adapt connections
- Significant effects of tuning parameters, especially learning rate

Thank you!

Appendix: Experiments with decay rate a



Inhibitory layer with a=0. 1



Inhibitory layer with a=0.001