Computational Neurodynamics

Topic 12
Plasticity

Murray Shanahan

Overview

- STDP characteristics
- STDP example
- Reward-modulated STDP
- A Braitenberg vehicle that learns

The Importance of Plasticity

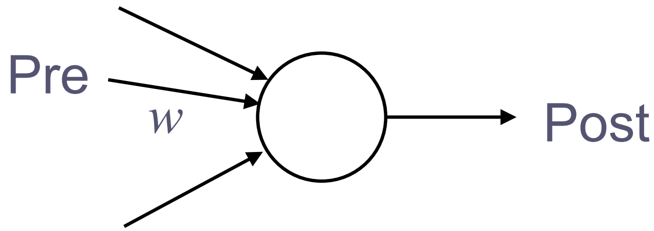
- This course has concentrated on dynamics in a fixed network with fixed weights
- But real brains are highly *plastic*, especially during development
- This enables an animal's brain to adapt to the specifics of the environment into which the animal is born, and to continue to adapt to changes in that environment throughout its lifetime
- There is plasticity both in the connections between neurons, which can grow and/or die, and in the synaptic weights of established connections
- We will consider only changes in synaptic weights

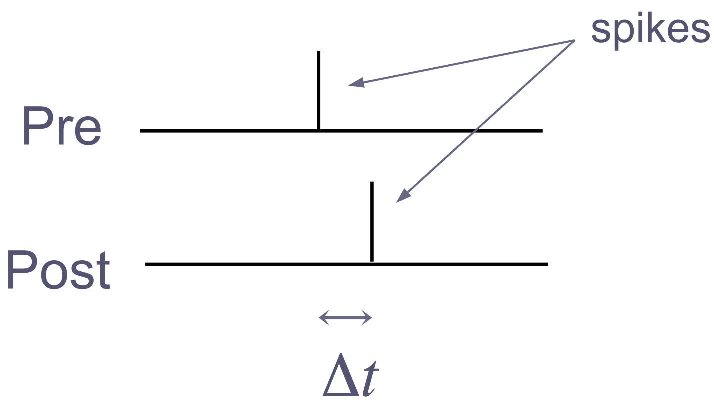
Learning Rules

- The function computed by a neural network is determined by the underlying neuron model, the network topology, by the training applied to the network, and by the *learning rule* used to train it
- There is a variety of learning rules for each type of artificial neuron, and for different network topologies
- For a feed-forward network of weighted sum (non-spiking) neurons, the best known learning rule is *back propagation*
- For a recurrent network of spiking neurons, the best known learning rule is *spike timing dependent plasticity* (STDP)
- STDP is a Hebbian rule – "neurons that fire together wire together"

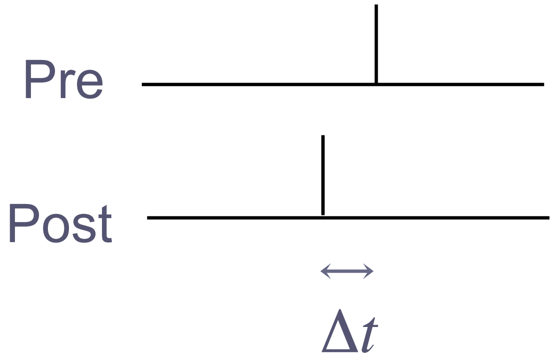
STDP Basics

- STDP adjusts the synaptic weightings between spiking neurons, to learn associations between firing patterns



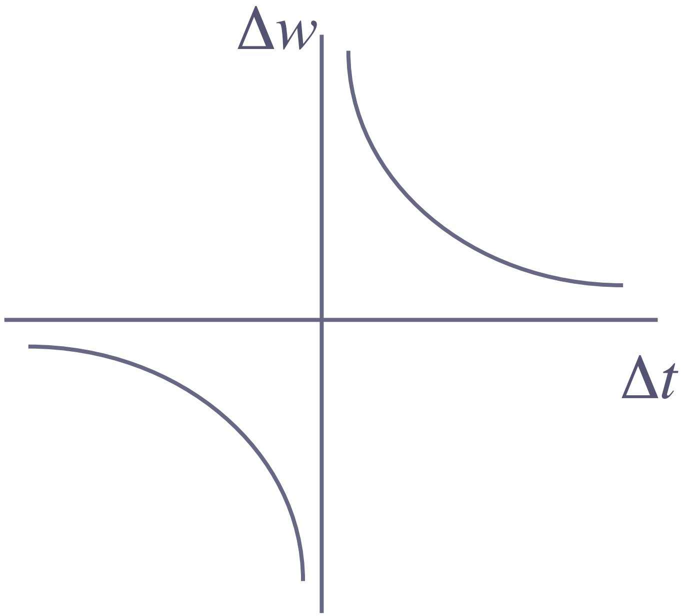


- If a pre-synaptic spike arrives just *before* a neuron fires, then the relevant synaptic weighting is *increased*



- If a pre-synaptic spike arrives just *after* a neuron fires, then the relevant synaptic weighting is *decreased*

The STDP Curve 1



- The change in weighting depends on the interval (Δt) between the pre-synaptic and post-synaptic spikes
- If t is positive, Δw is also positive, but if Δt is negative, Δw is negative
- If Δt is small, then the magnitude of Δw is large

The STDP Curve 2

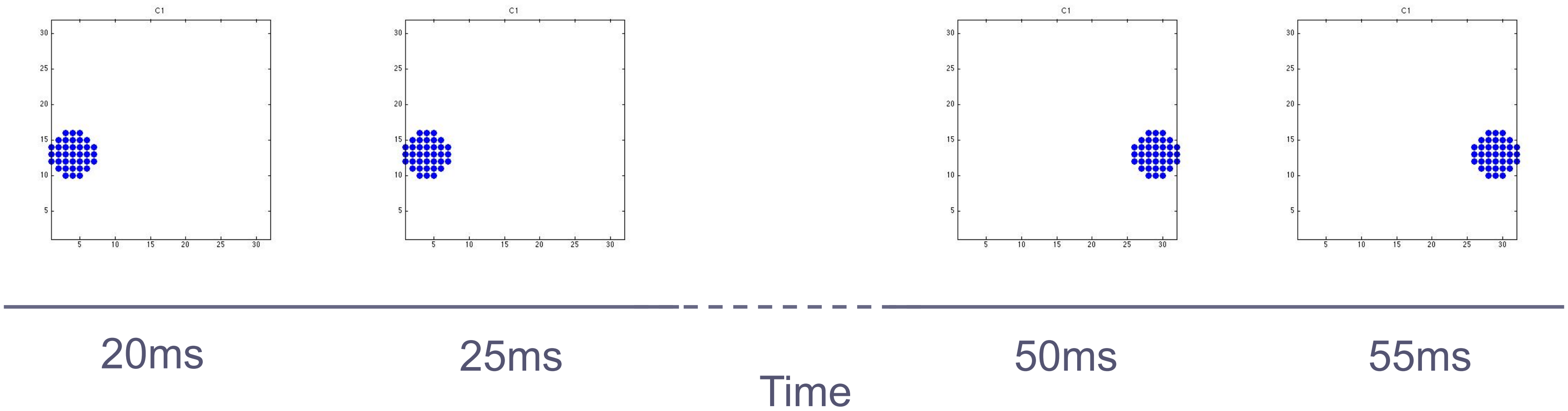
- More precisely, we have

$$\Delta w = \begin{cases} A^{+} e^{-\Delta t / \tau^{+}} & \text{if } \Delta t \geq 0 \\ -A^{-} e^{\Delta t / \tau^{-}} & \text{if } \Delta t < 0 \end{cases}$$

  where A^{+}, A^{-}, τ^{+}, and τ^{-} are constants
- Note that by varying these constants we can make the characteristic curve assymmetric
- Normally $(A^{-}\tau^{-})/(A^{+}\tau^{+}) > 1$, which prevents uncontrolled growth of w

STDP Example 1

- Here is an example of STDP in action on a 30x30 grid of fully interconnected spiking neurons with random initial weights and conduction delays
- In training, a spatiotemporal pattern of firing is presented
- A patch of neurons on the left fires several times at 5ms intervals
- Then after a delay, a different patch of neurons on the right fires several times at 5ms intervals



STDP Example 2

- After training, the network is presented with the first part of the spatiotemporal pattern only
- Then it reproduces (a statistical approximation of) the second part of the spatiotemporal pattern by itself

20ms 25ms 50ms 55ms

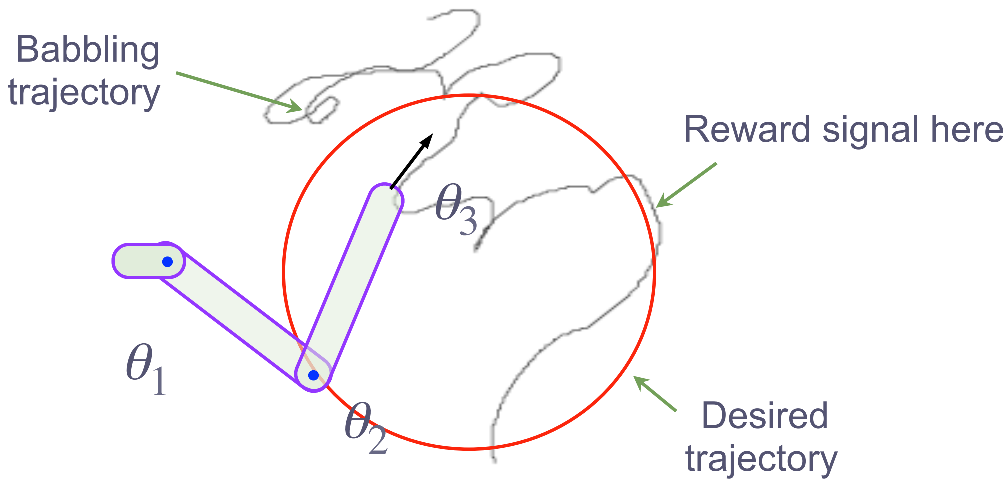
Time

Reward-Modulated STDP 1

- In an embodied setting, STDP can be used in conjunction with a reward signal to realise a form of *reinforcement learning*
- The animal or robot continuously executes actions, and sometimes receives a reward signal. In general, this will be because it has executed an action that has lead to something good (eg: food)
  - In the brain, reward signals are carried by a neurotransmitter called *dopamine* (which also indicates novelty)
- The challenge is to modify STDP so that it learns the correlations between certain actions and their positive outcomes

Reward-Modulated STDP 2

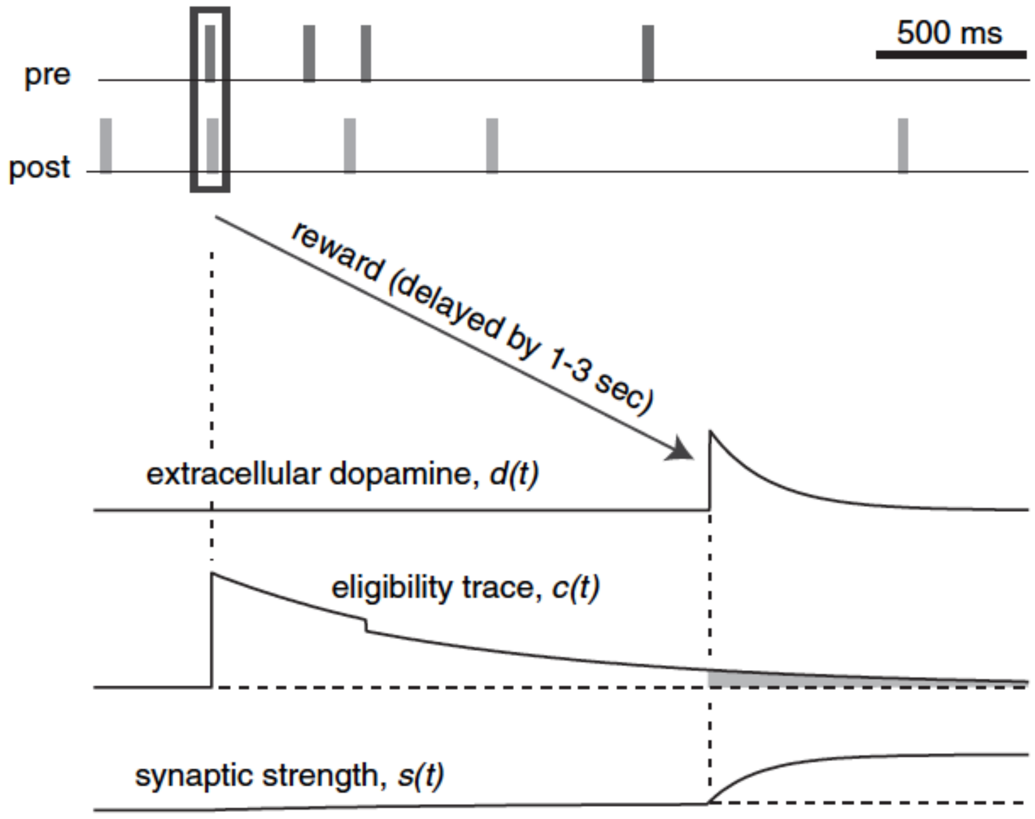
- Example: a 2-dof robot arm must learn a particular movement, such as motion in a circle
- It performs motor babbling (random movements), and whenever its actual trajectory (babbling) matches that of the desired trajectory (a circle), a reward signal is generated



Eligibility Traces 1

- Learning is carried out by a spiking neuron network with an input layer representing the arm configuration (θ_1 and θ_2), and an output layer representing the actual direction of movement of the end-effector (θ_3)
- If pre-synaptic and post-synaptic spikes occur just before or just after each other, then normally STDP would be applied
- But with reward-modulated STDP, this causes a jump in another (new) variable associated with the synapse which encodes the eligibility of a synapse for a weight change
- This *eligibility trace* decays over time
- When a reward is received, the synapse is updated according to the strength of the eligibility trace

Eligibility Traces 2



From Izhikevich, 2007

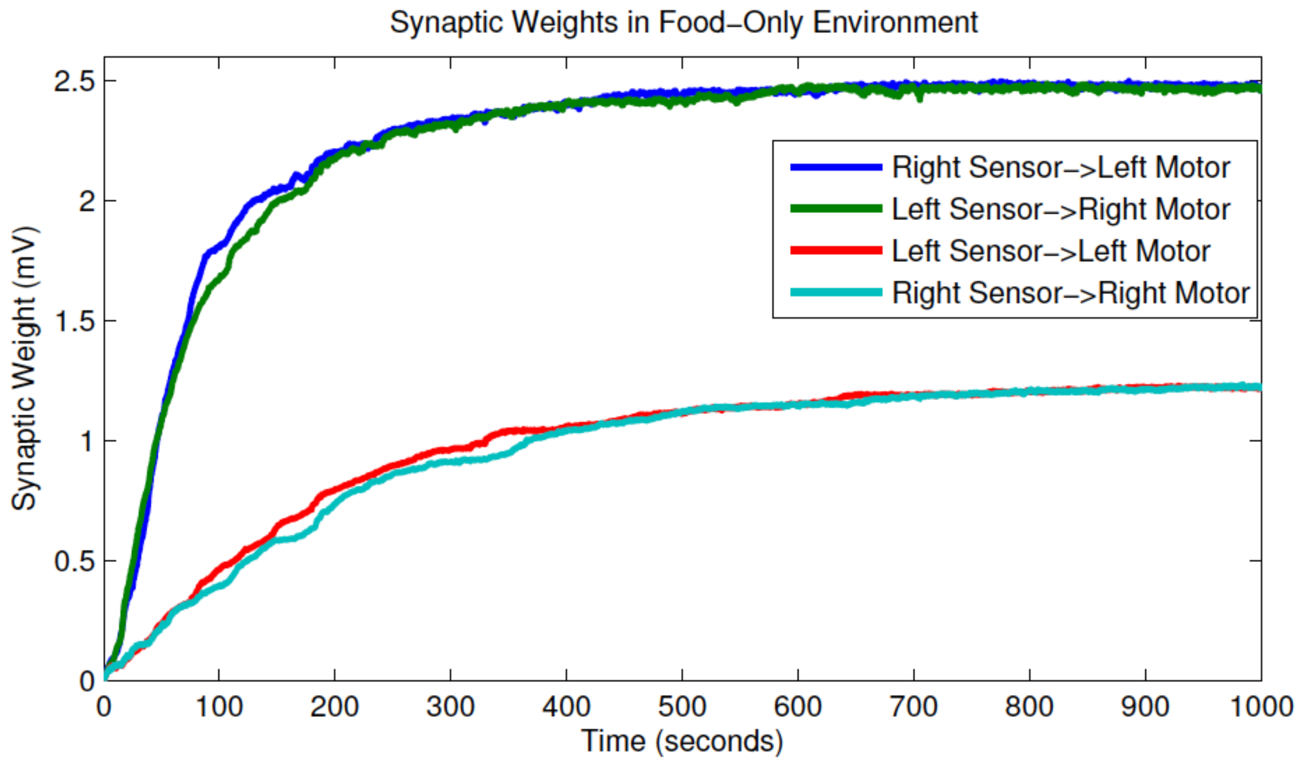
- This ensures that synaptic connections are strengthened between a given arm configuration and the corresponding required movement
- Using an eligibility trace allows *delayed rewards* to be dealt with effectively

Delayed Reward

- But what if the arm carries out both desired and undesired movements during a period, and STDP is applied?
- Because this is a delayed reward problem, there is a gap between the performance of an action and the reward signal it causes, rendering it ambiguous
- This will result in both desired and undesired connections being strengthened
- Using STDP, this problem is overcome with the injection of noisy firing. This ensures that there is a statistical tendency for synaptic weights to go down as much as up, and over time return to zero
- Over time, weights representing the occasional "false correlation" will go back to zero, but weights representing true correlations, which are rewarded more frequently, will migrate upwards

A Learning Braitenberg Vehicle 1

- Using reward-modulated STDP, we can build a neurally controlled Braitenberg vehicle that learns
- Here it gets rewarded when it touches “food”, and learns the synaptic connections that give rise to the familiar behaviour from Topic 5



MSc work by Richard Evans

A Learning Braitenberg Vehicle 2

- Here we see how the learning version of the robot is more successful over time, as the effects of learning kick in
- The robot without learning (which just wanders randomly) remains a poor forager

MSc work by Richard Evans

Related Reading

Cassenaer, S. & Laurent, G. (2012). Conditional Modulation of Spike-Timing Dependent Plasticity for Olfactory Learning. *Nature* 482, 47–53.

Izhikevich, E. (2007). Solving the Distal Reward Problem Through Linkage of STDP and Dopamine Signaling. *Cerebral Cortex* 17, 2443–2452.

Song, S. & Abbott, L.F. (2001). Cortical Development and Remapping Through Spike Timing-Dependent Plasticity. *Neuron* 32, 339–350.