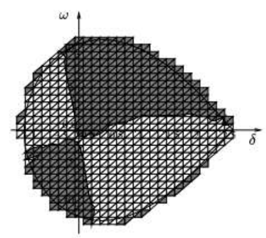
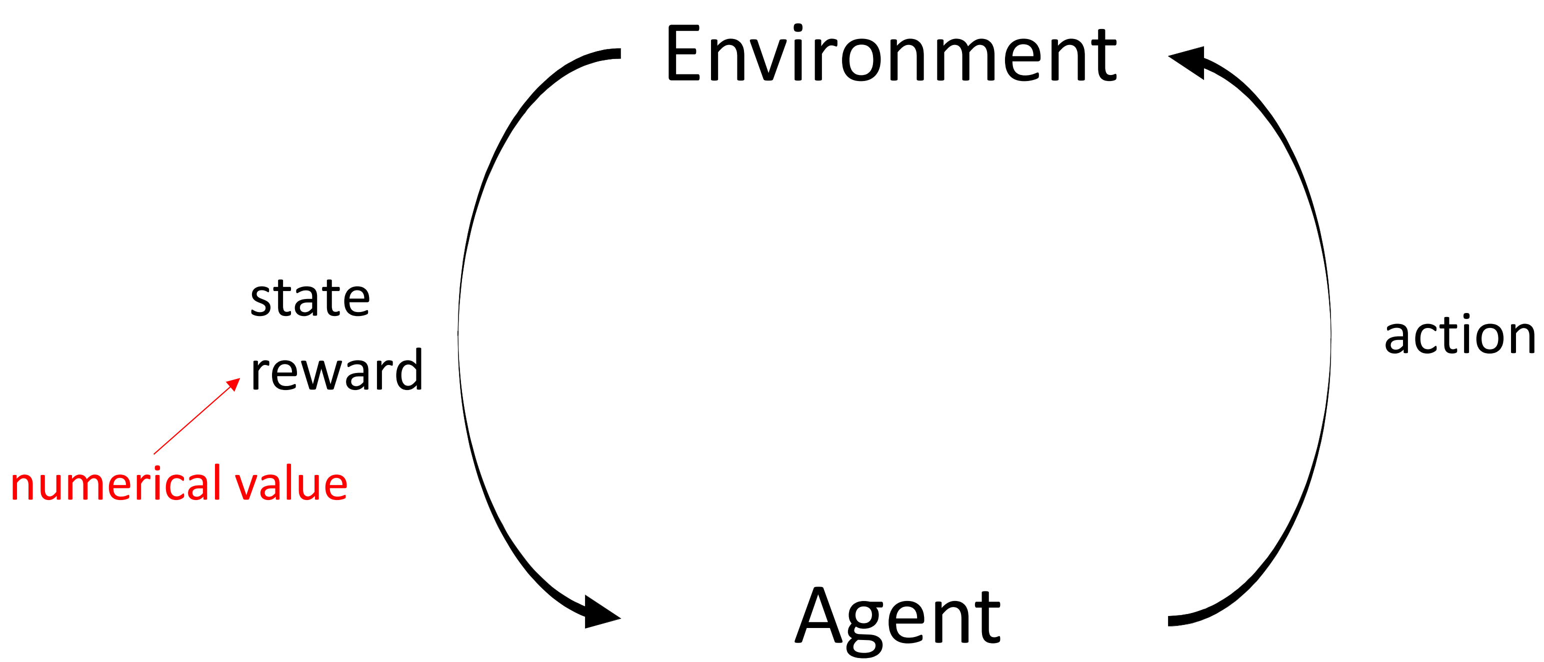
Reinforcement learning, energy systems and deep neural nets



Prof. Damien ERNST

LIÈGE université

Reinforcement learning agent



+ the energy market

State: (i) the battery level (ii) *Everything you know about the market*
Reward: The money you make during the market period.

The battery setting for the next market period.

The battery controller

Table 1. Summary of RL considerations for electric power system control/decision

Problem	Type of control	RL method	Reference(s)
Electricity market simulation	Market decision	Q-learning	Harp et al. (2000) Rahimiyan et al. (2010) Nanduri and Das (2007) Lincoln et al. (2012) Kim et al. (2016) Krause et al. (2006)
Transient angle instability	Emergency	Q-learning	Ernst et al. (2004) Glavic (2005) Glavic et al. (2005a) Glavic et al. (2005b) Li and Wu (1999)
		Fitted Q iteration	Ernst et al. (2009)
		Policy search	Mohagheghi et al. (2006)
Oscillatory angle instability	Emergency	Q-learning	Ernst et al. (2004) Wang et al. (2014) Glavic et al. (2005a) Ademoye and Feliachi (2012) Karimi et al. (2009)
Voltage control	Normal	Q-learning	Xu et al. (2012) Vlachogiannis et al. (2004)
AGC (Automatic) generation control)	Normal	$Q(\lambda)$ with elig. traces	Yu et al. (2011)
		Q-learning	Daneshfar and Bevrani (2010) Ahamed et al. (2002)
		$R(\lambda)$	Yu et al. (2012b)
Economic dispatch	Normal	Q-learning	Jasmin et al. (2011) Yu et al. (2016)
Wide-area control	Emergency	TD	Yousefian et al. (2016)
		Q-learning	Yan et al. (2016) Hadidi and Jeyasurya (2013)
Households control	Normal	Q-learning	Wang et al. (2016) Yan et al. (2016)
Wind generation control	Normal	Q-learning	Wei et al. (2015) Tang et al. (2015)
		$Q(\lambda)$	Yu et al. (2012a)
Demand control	Normal	Fitted Q iteration	Ruelens et al. (2016) Vandael et al. (2015)
System restoration	Restorative	Q-learning	Ye et al. (2011)
Congestion management	Emergency	Q-learning	Zarabbian et al. (2016)
Microgrids control	Normal	Q-learning	Khorramabady et al. (2015) Li et al. (2012)
		Policy search	Venayagamorthy et al. (2016)

Table taken from: "Reinforcement Learning for Electric Power System Decision and Control: Past Considerations and Perspectives". M. Glavic, R. Fonteneau and D. Ernst. Proceedings of the 20th IFAC World Congress.

Learning:

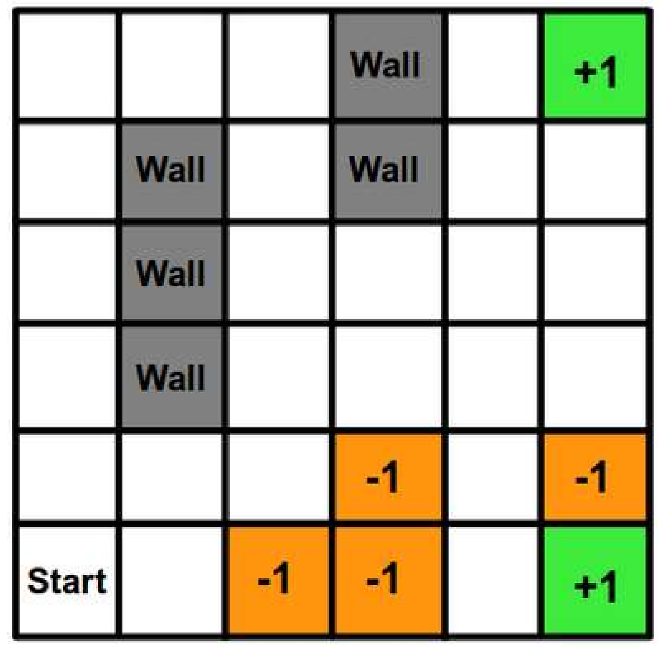
Input : x_t, u_t, r_t and x_{t+1}

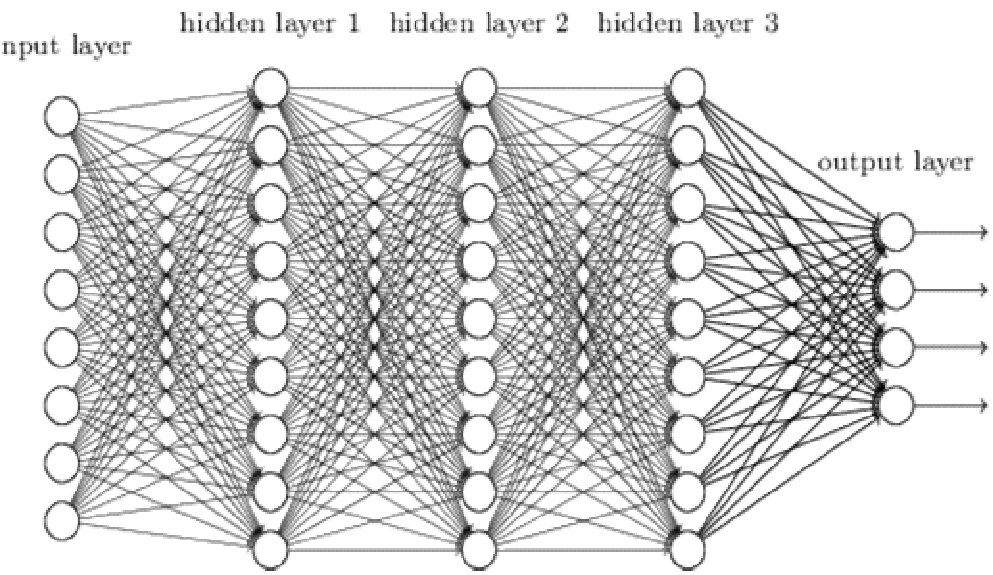
$$\delta \leftarrow (r_t + \gamma \max_{u \in U(x_{t+1})} Q(x_{t+1}, u)) - Q(x_t, u_t)$$

$$Q(x_t, u_t) \leftarrow Q(x_t, u_t) + \alpha\delta$$

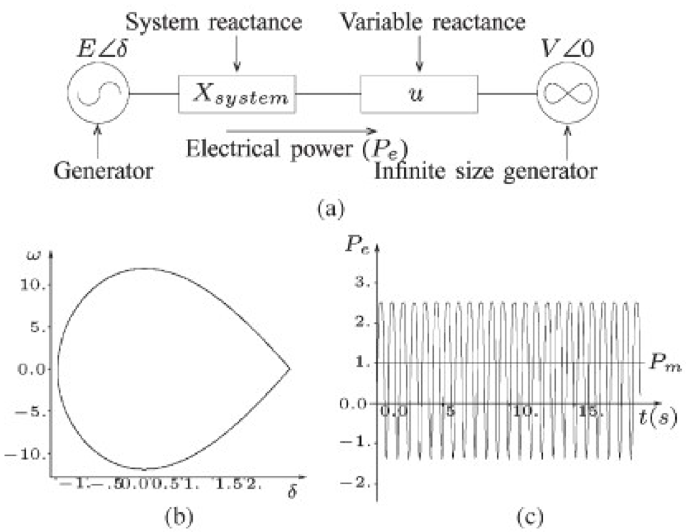
Exploration/exploitation: Not always take the action that is believed to be optimal to allow exploration.

Generalization: Generalize the experience gained in some states to other states.

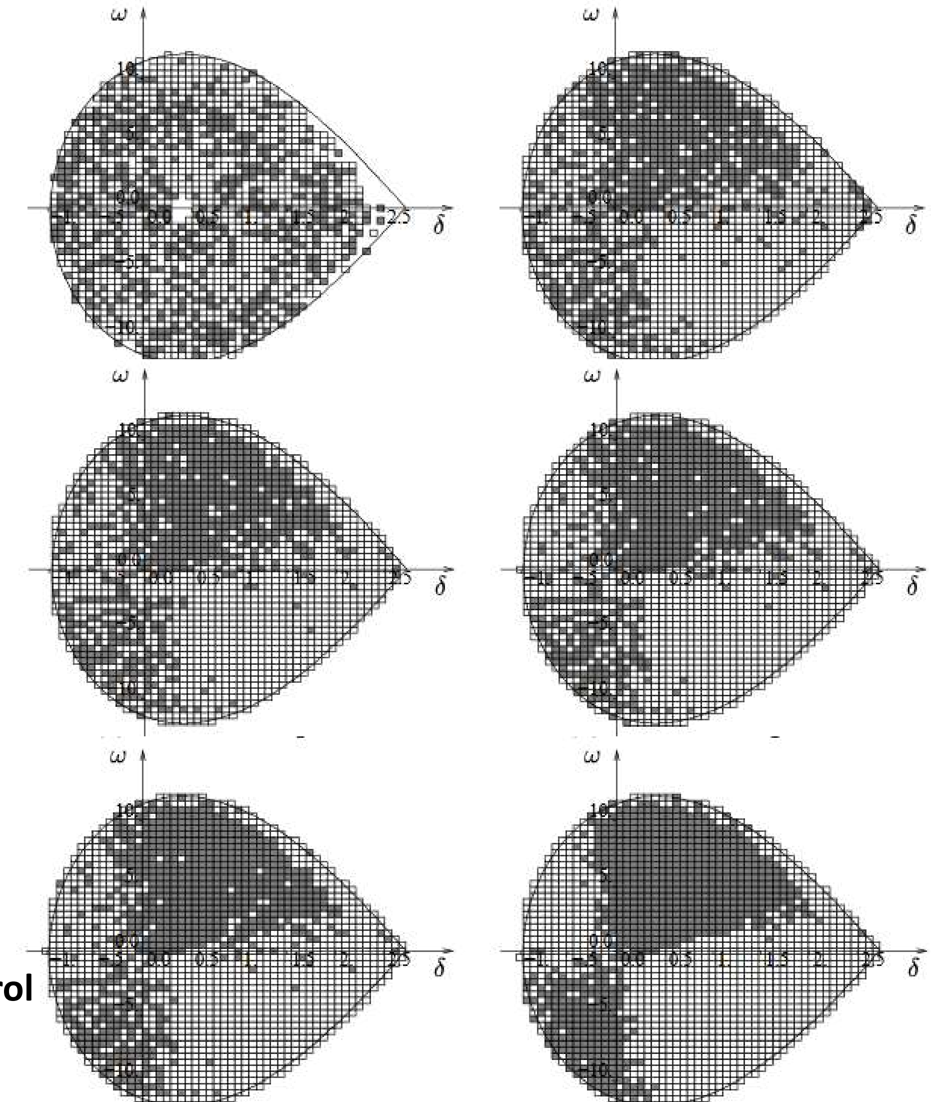



nput layer
hidden layer 1
hidden layer 2
hidden layer 3
output layer

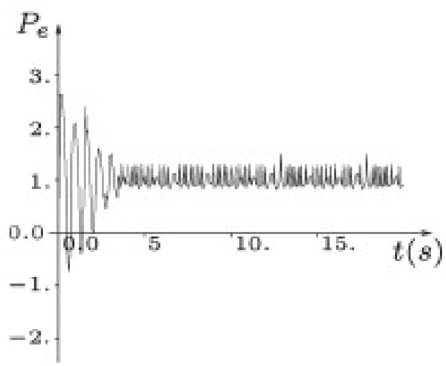

80
OXYGEN



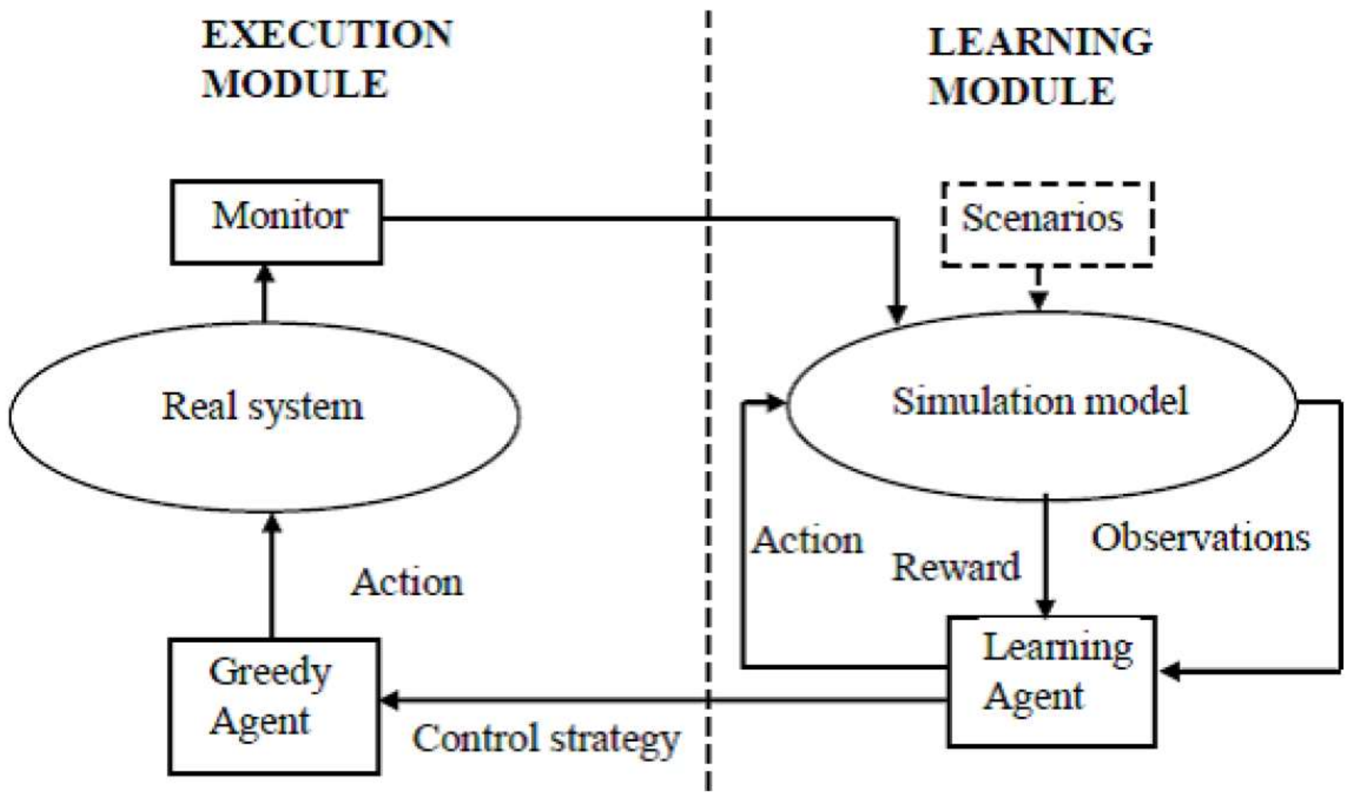
Learning phase



Effect of the resulting control policy



First control law for stabilizing power systems every computed using reinforcement learning. More at: "Reinforcement Learning Versus Model Predictive Control: A Comparison on a Power System Problem". D. Ernst, M. Glavic, F.Capitanescu, and L. Wehenkel. IEEE Transactions on Syestems, Man, An Cybernetics—PART B: Cybernetics, Vol. 39, No. 2, April 2009.


EXECUTION MODULE
LEARNING MODULE
Monitor
Scenarios
Real system
Simulation model
Action
Reward
Observations
Action
Greedy Agent
Learning Agent
Control strategy

Reinforcement learning for trading in the intraday market

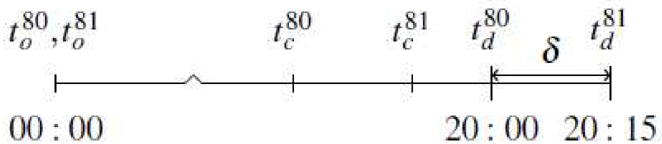


Figure 1: Trading time-line for products Q-80 and Q-81

Table 1: Order Book for Q-80 and time-slot 20:00-20:15

i	Type	v [MW]	p [€/MWh]	
4	"Sell"	6.25	36.3	
2	"Sell"	2.35	34.5	⟵ ask
1	"Buy"	3.15	33.8	⟵ bid
3	"Buy"	1.125	29.3	
5	"Buy"	2.5	15.9	

Complex problem:
- Adversarial environment
- Highly dimensional
- Partially observable

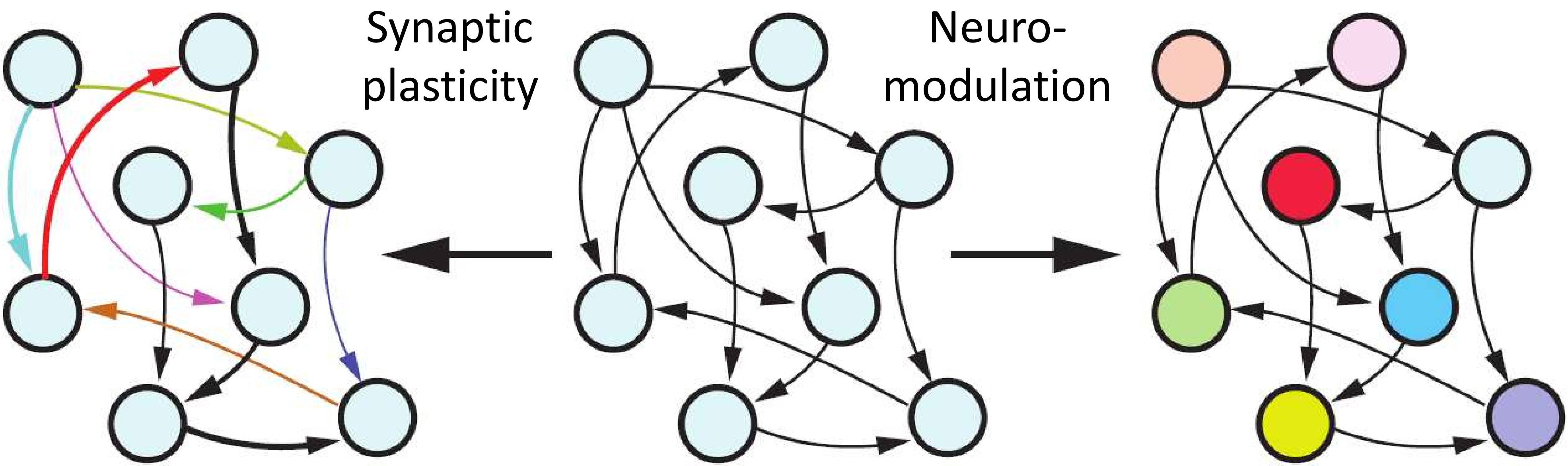
Best results obtained with optimisation of strategies based on past data together with supervised learning to learn from the optimised strategies (imitative-learning type of approach)

More: "Intra-day Bidding Strategies for Storage Devices Using Deep Reinforcement". I. Boukas, D. Ernst, A. Papavasiliou, and B. Cornélusse. Proceedings of the 2018 15th International Conference on the European Energy Market (EEM).

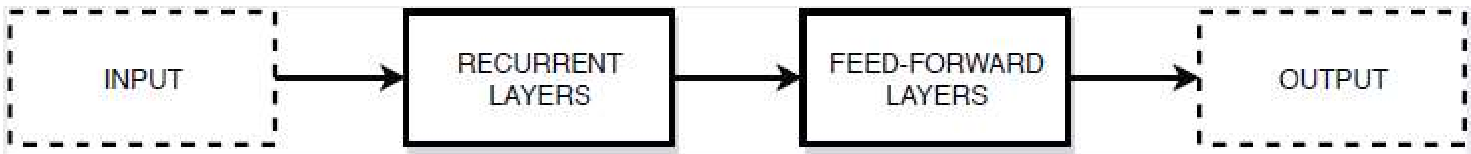
“A critical present objective is to develop deep RL methods that that can adapt rapidly to new tasks.”

Deepmind, “Learning to reinforcement learn.” (2016).

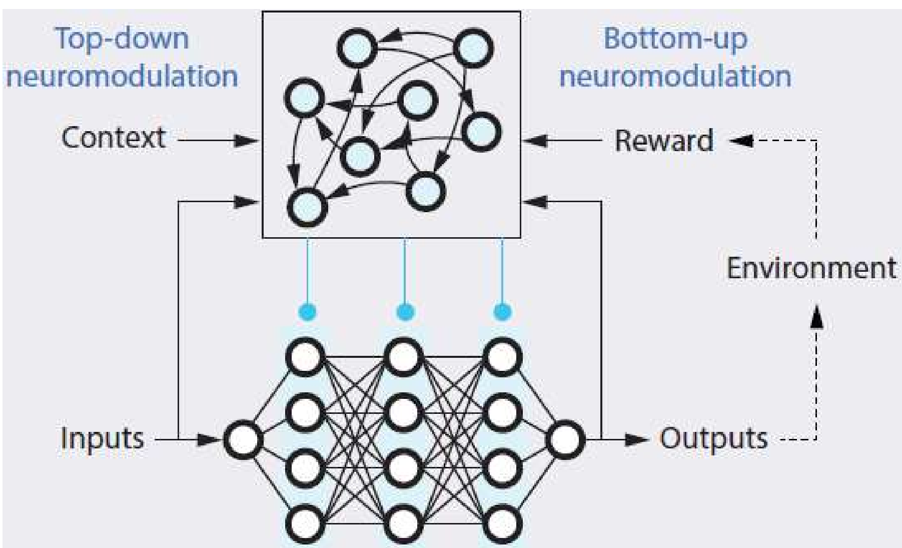
Walking: a meta-RL problem solved through synaptic plasticity and neuro-modulation

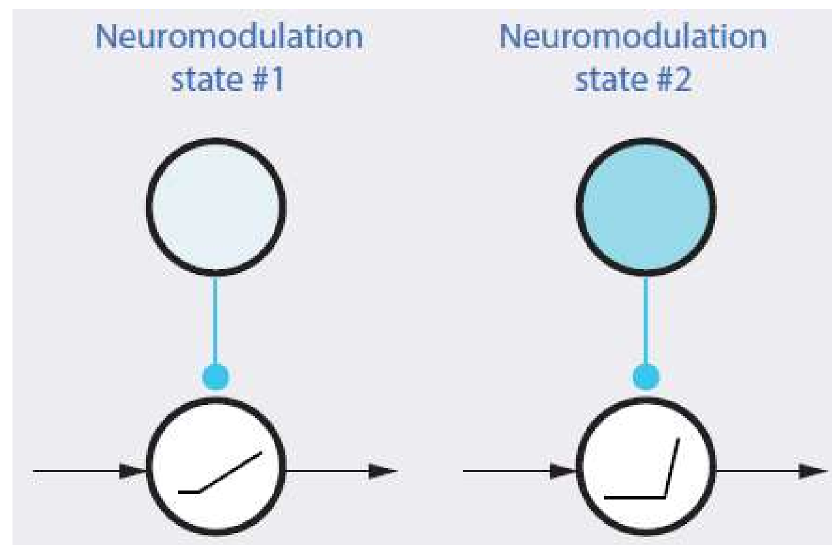


Classical architecture for solving meta-RL problems:



Our new architecture:



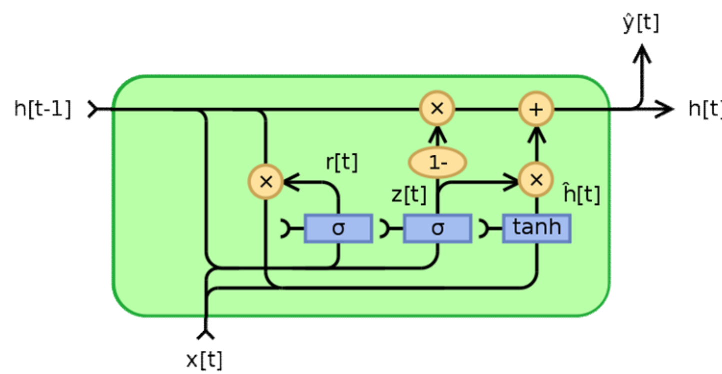


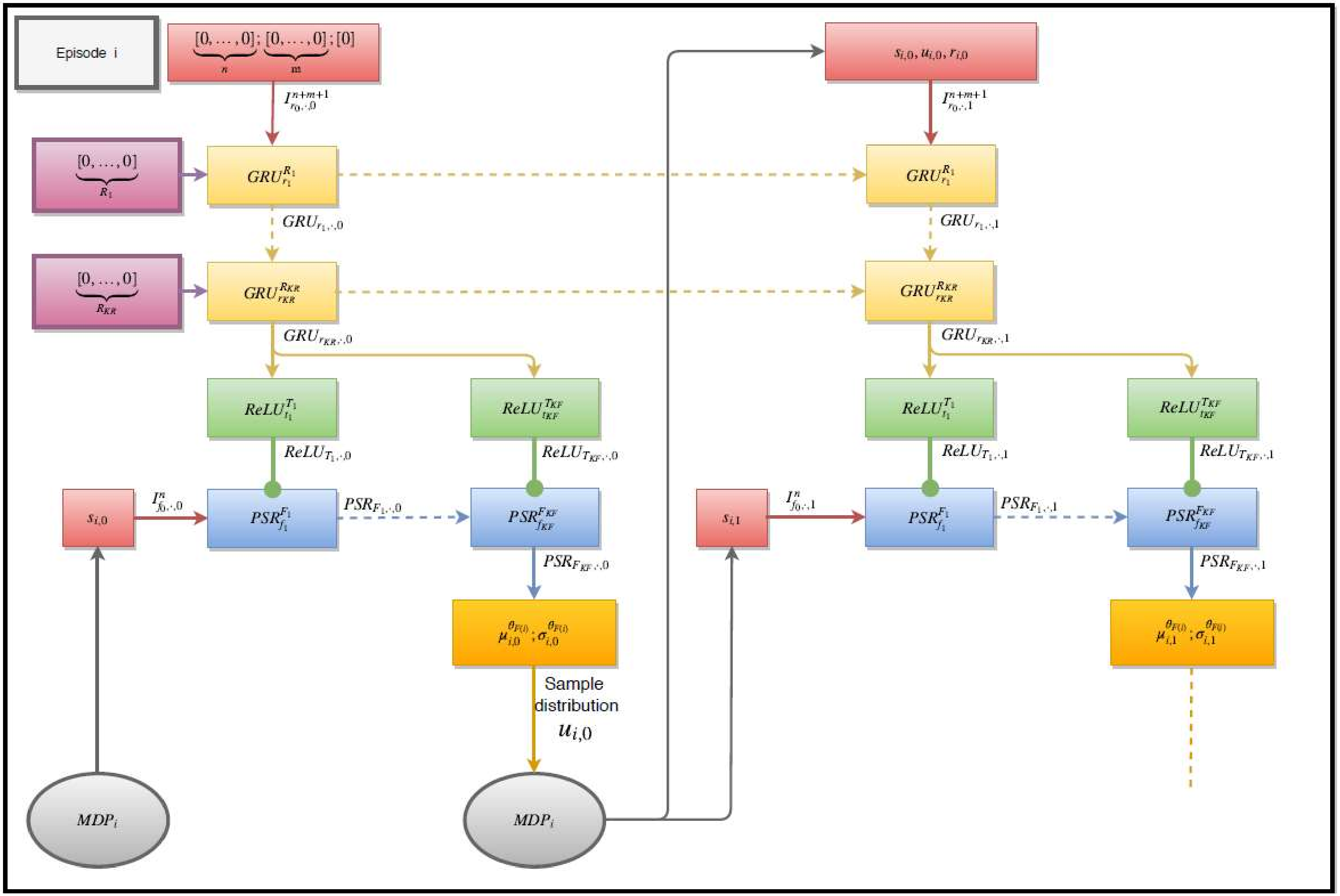
Rectified Linear Unit: $ReLU(x) = max(0, x)$

Saturated Relu: $sReLU(x) = min(max(-1, x), 1)$

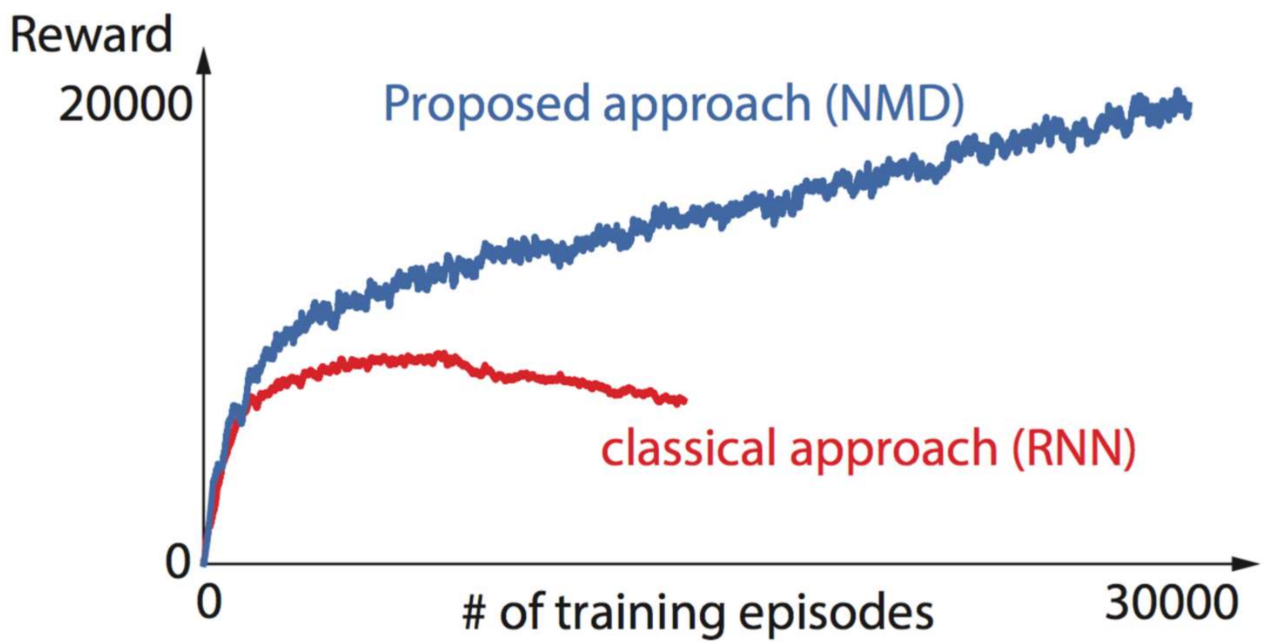
Parametrized *sRelu*: $PSR(x) = sReLU(m * x + a)$

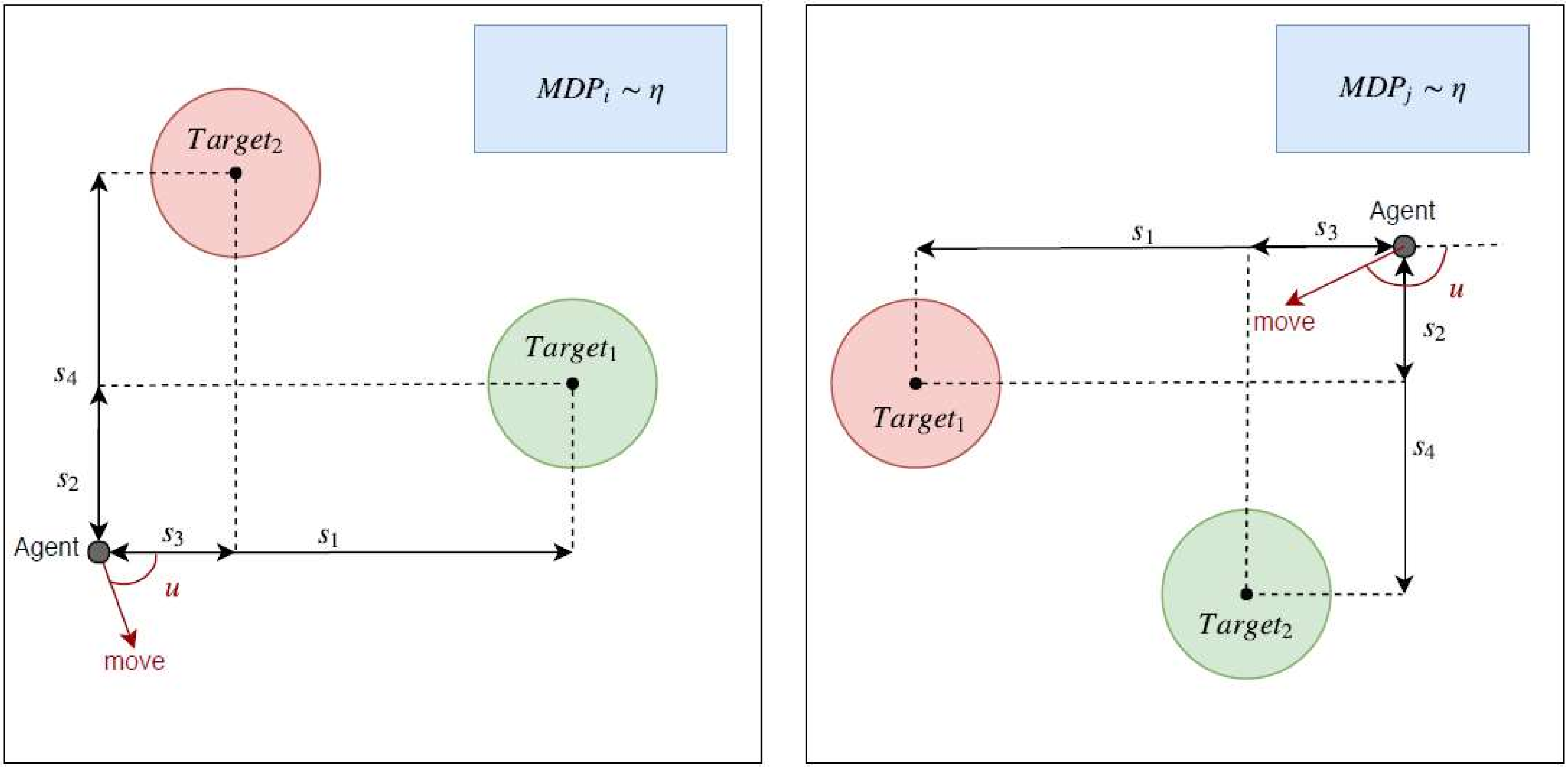
Gated Recurrent Unit: GRU

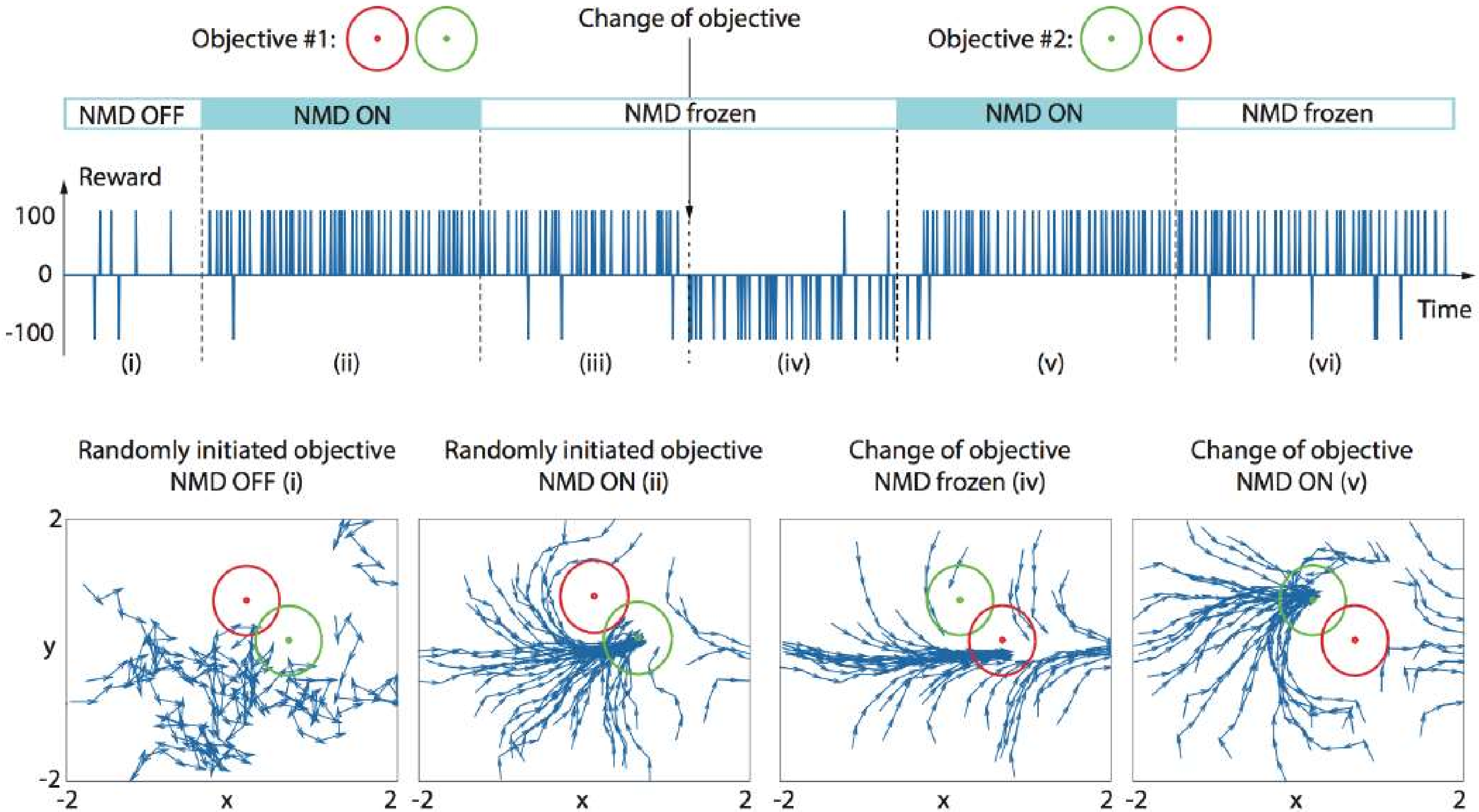




Episode i
Sample
distribution
$u_{i,0}$
MDP_i
MDP_i


Reward
20000
Proposed approach (NMD)
classical approach (RNN)
0
0
# of training episodes
30000


$MDP_i \sim \eta$
$Target_2$
$Target_1$
s_4
s_2
Agent
s_3
s_1
u
move
$MDP_j \sim \eta$
Agent
s_1
s_3
u
move
s_2
$Target_1$
s_4
$Target_2$



More: "Introducing neuromodulation in deep neural networks to learn adaptive behaviours". N. Vecoven, D. Ernst, A. Wehenkel and G. Drion. Download at: https://arxiv.org/abs/1812.09113