Giovani Michel
(Me)

Towards Q-learning-based control using a spiking Neuromorphic network and sparse encoding

Giovanni Michel[1], Alpha Renner[2], Gerd J. Kunde[3], Andrew T. Sornborger[1]

[1]Information Sciences (CCS-3), Los Alamos National Laboratory, Los Alamos, New Mexico, USA

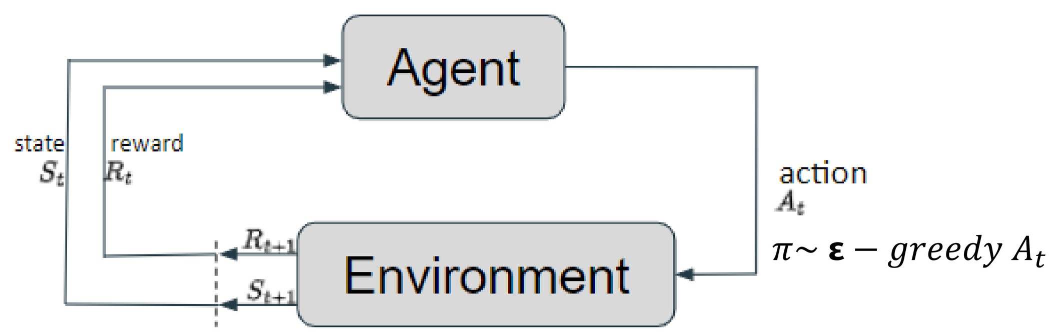
[2]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland

[3]Nuclear and Particle Physics Applications (P-3), Los Alamos National Laboratory, Los Alamos, New Mexico, USA

[4]College of Engineering and Computer Science, Florida Atlantic University, Boca Raton, Florida, USA

Highlights

- Neuromorphic processors represent an emerging alternative to current standard Von Neumann architectures by orders of magnitude in reduced power consumption.
- We present the first step towards a fully-online reinforcement learning implementation by presenting proof of concept that neuromorphic algorithms can solve complex control tasks.
- In our research we present a closed-loop, neuromorphic implementation of a proof-of-principle problem: cartpole balancing [1]. Our neuromorphic solution uses OpenAI gym to simulate the cart pole dynamics off-chip.
- For now, the Q matrix is learned off-chip. The inference and actual control problem is implemented on Intel's neuromorphic processor Loihi 2 [3, 4].

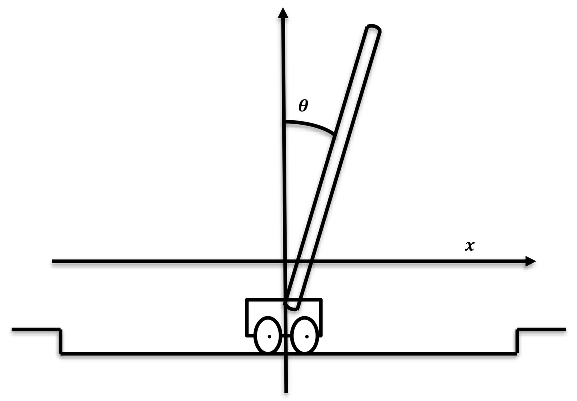


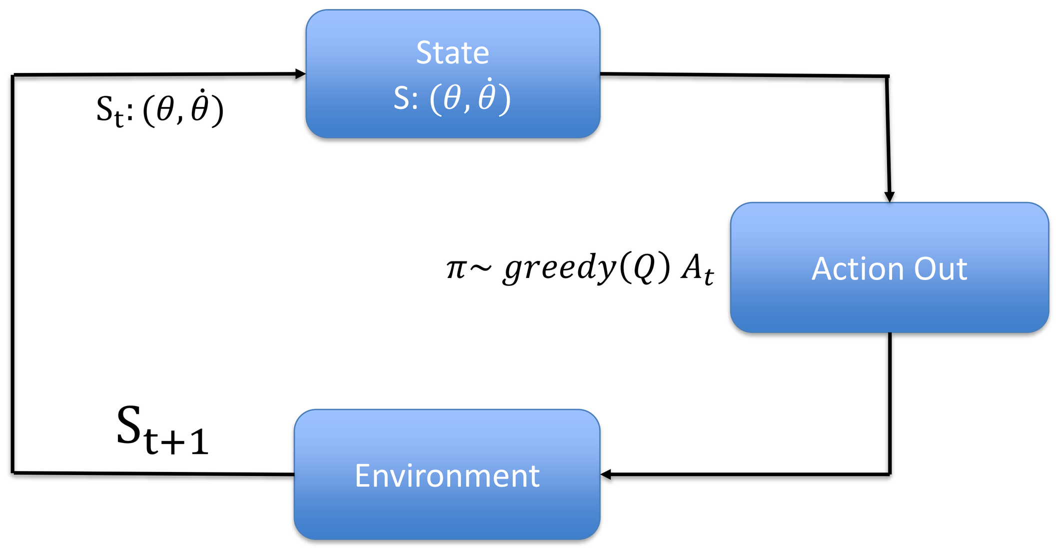
Introduction

- The reinforcement learning algorithm learns to compute the best action based on the pole state to output the correct force required to keep the pole upright.
- The agent uses the reward signal to select actions from the Q-matrix and demonstrates a model-free architecture using Off-Policy Temporal Difference (TD) Learning. The TD control update rule for the Q-learning, [1]

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \boldsymbol{\alpha}[R_{t+1} + \boldsymbol{\gamma} \max Q(S_{t+1}, a) - Q(S_t, A_t)]$$

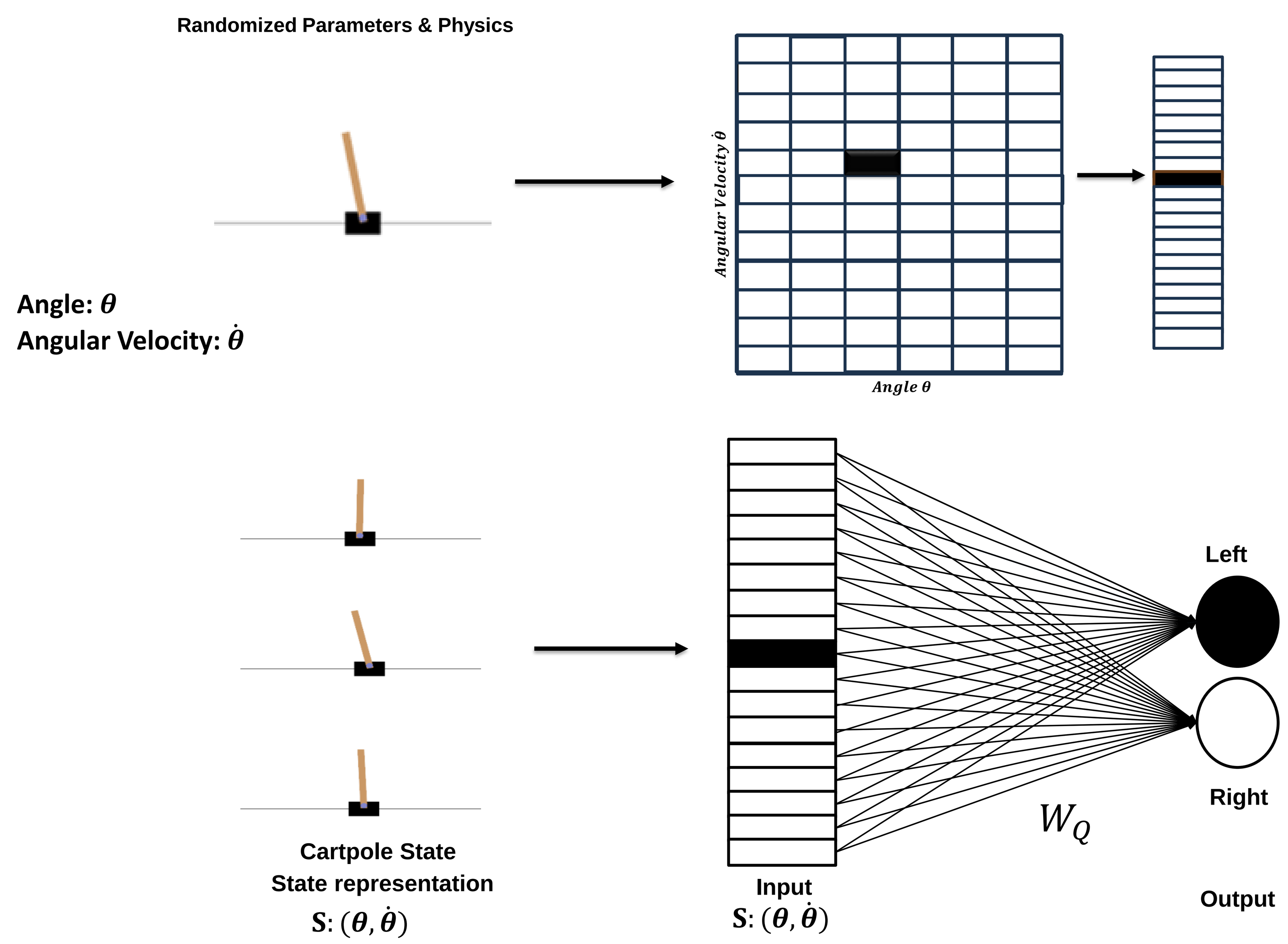
- We use the pole's angle, angular velocity, and Q-learning to solve the cart pole balancing problem:





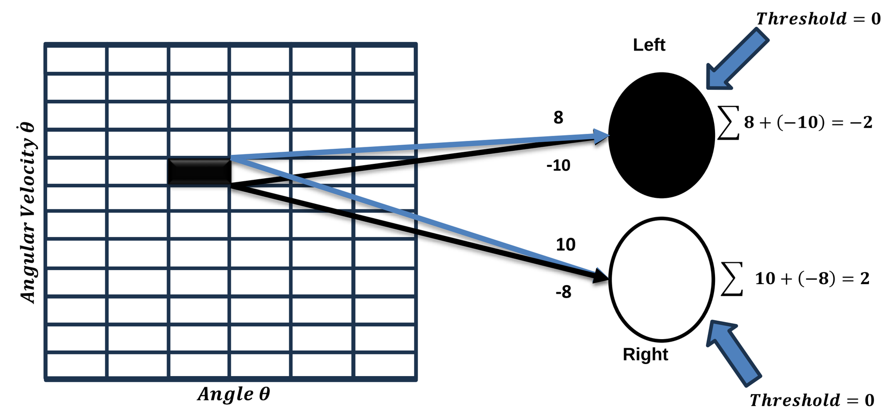
Methods

- The cartpole has a continuous state space. Therefore, we discretize the cartpole angles (θ), the angular velocities ($\dot{\theta}$), representing the state space as a matrix we then unroll the matrix into a flattened vector and represent the discretized parameters as sparse binary vectors that is sent through the pre-synaptic layer.

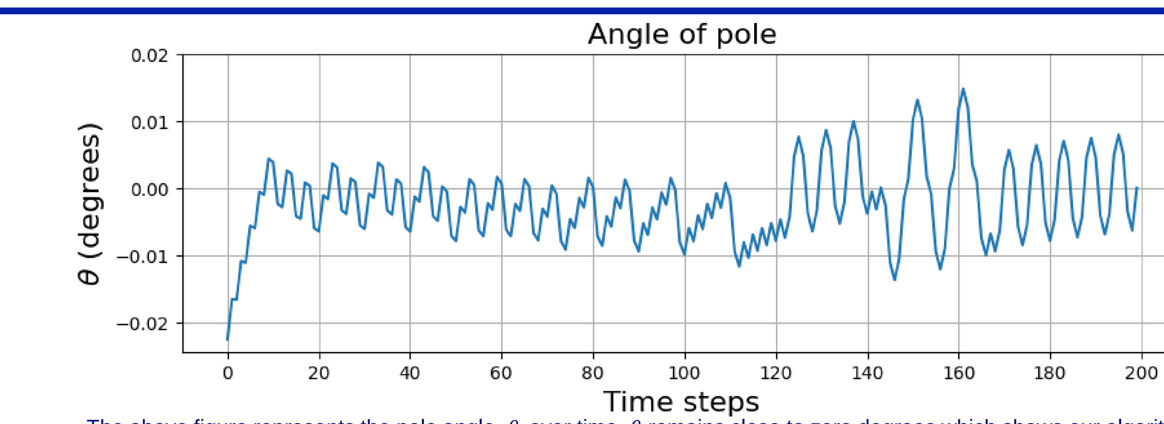


The neuromorphic circuit that is implemented on Loihi 2

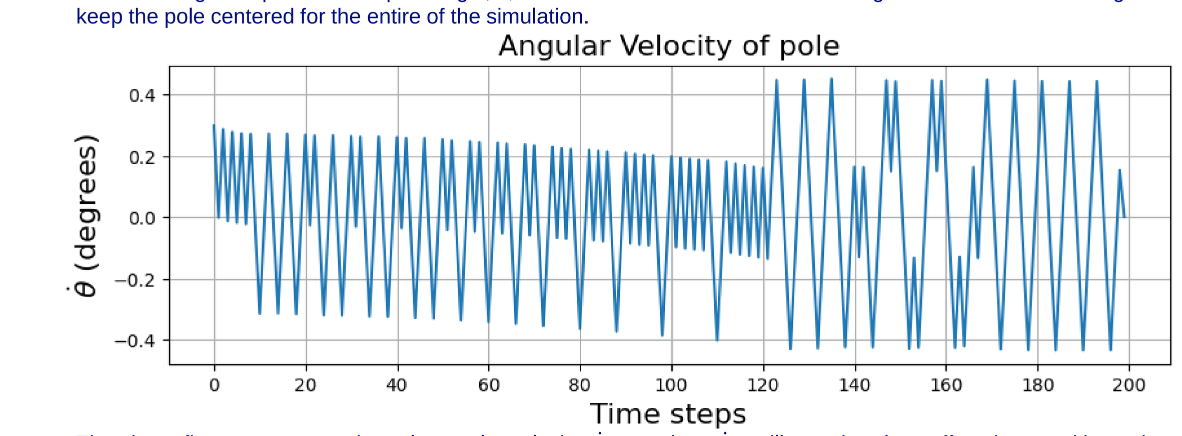
- We synaptically-encoded the Q-matrix that takes these vectors as input and outputs an action to move the base of the cartpole left or right.
- We use the spikes from the neuromorphic hardware implemented on Intel's neuromorphic research processor [3, 4] to perform the action selection to control the cart pole. The actions are used in conjunction with OpenAI gym to demonstrate closed-loop control.
- Synaptic weight matrix W_Q is calculated from the original Q matrix by subtracting from each weight (w_i, w_j) counterpart to the other action to ensure only one action neuron fires in a timestep, for instance: $w_{i,right} = q_{i,right} - q_{i,left}$.



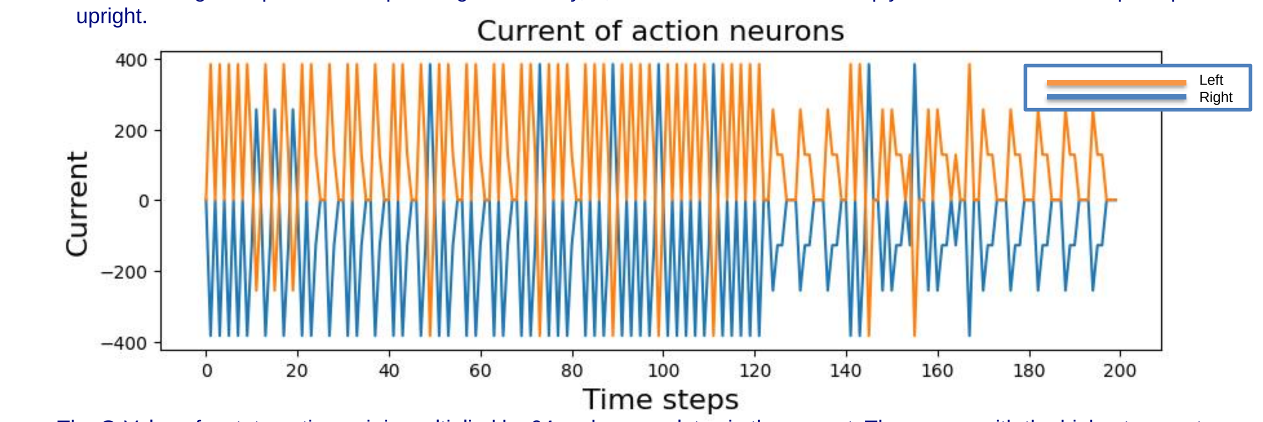
Argmax $(Q(S_t, A_t)) = W_Q - flip(W_Q)$



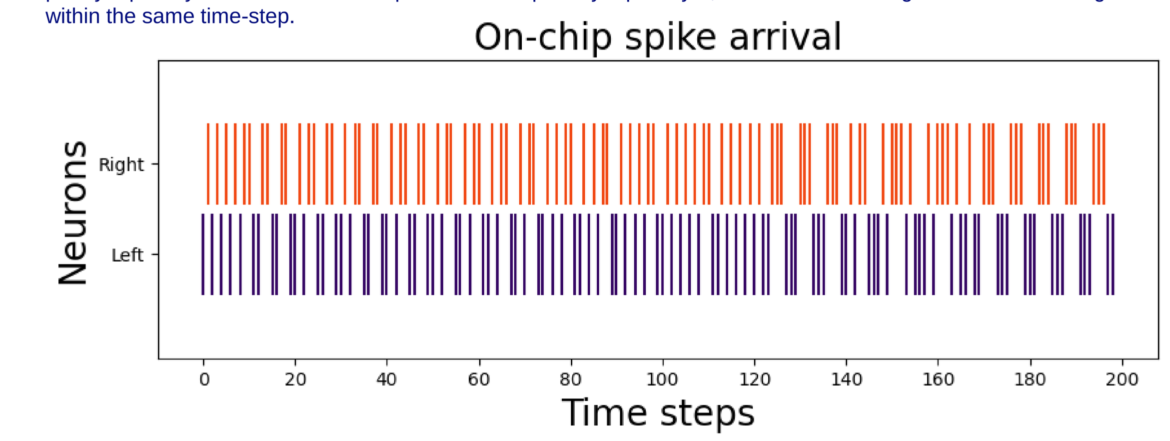
- The above figure represents the pole angle, θ, over time. θ remains close to zero degrees which shows our algorithm can keep the pole centered for the entire of the simulation.



- The above figure represents the pole angular velocity, $\dot{\theta}$, over time. $\dot{\theta}$ oscillates sharply to affect theta and keep the pole upright.



- The Q-Value of a state-action pair is multiplied by 64 and accumulates in the current. The neuron with the highest current represents the best action. At time step 120, there is a shift in the membrane potential that initially appears anomalous. However, this change results from the pole angle and angular velocity, which produces a new state sent through the pre-synaptic layer. A spike from the pre-synaptic layer takes one time step to reach the post-synaptic layer, but the current integration into the voltage occurs instantly within the same time-step.



- Loihi 2 on chip output spike pattern that is sent from the neuromorphic hardware to OpenAI gym. The spike pattern represents the action selected by the post-synaptic neurons over time.

Discussion

- We demonstrate in our work that reinforcement learning algorithms implemented on neuromorphic hardware can solve complex dynamical systems like the cart pole problem using a towards Q-learning implementation. To measure the performance of the algorithm we calculated the average reward over 100 trials. We achieve a mean reward of 200.00. Future work will involve implementing the *Q-learning* fully on-chip and learning the weights on neuromorphic hardware.
- The on-chip learning implementation will utilize neuronal mechanisms that are then executed on neuromorphic hardware. For example, the learning process necessitates an epsilon-greedy approach to explore all possible states, which can be achieved using firing neurons.
- The learning rule uses the max state-action value received at S_{t+1} $\max(Q(S_{t+1}, a))$, which makes updating the q-value for that state-action pair $Q(S_t, A_t)$ a difficult problem. One possible solution involves using graded spikes from the neuromorphic hardware which will store $\max(Q(S_{t+1}, a))$ as a spike.

References

[1] Barto, Andrew G., Richard S. Sutton, and Charles W. Anderson. "Neuronlike adaptive elements that can solve difficult learning control problems." *IEEE transactions on systems, man, and cybernetics* 5 (1983): 834-846.

[2] Brockman, Greg, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. "Openai gym." *arXiv preprint arXiv:1606.01540* (2016).

[3] Davies, Mike, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou et al. "Loihi: A neuromorphic manycore processor with on-chip learning." *Ieee Micro* 38, no. 1 (2018): 82-99.

[4] Davies, Mike, Andreas Wild, Garrick Orchard, Yulia Sandamirskaya, Gabriel A. Fonseca Guerra, Prasad Joshi, Philipp Plank, and Sumedh R. Risbud. "Advancing neuromorphic computing with loihi: A survey of results and outlook." *Proceedings of the IEEE* 109, no. 5 (2021): 911-934.