**Thesis Matei Dogariu**

# **1** **Introduction**

The study of post-industrial age human driven climate change has been well understood for the past few decades. Early on, the concept of energy decarbonization was proposed to be the long-term solution to humanities increasing greenhouse gas emissions ​(Nakicenovic, 1993)​.

# **2** **Related work**

# **3** **Research method**

To investigate whether a home battery can be used to aid in the decarbonization of household energy consumption, a computational modeling approach was employed. The method used in this paper involves the design of a household-like environment using the standard dictated by OpenAI's Gymnasium framework. On this standardized environment, multiple Reinforced Learning algorithms were trained with the goal to minimize the effective CO2 emissions using the battery's charge/discharge cycle.

## **3.1 Gymnasium Environment**

The basics of reinforcement learning place an agent inside an environment, who is tasked with making decisions while being given observations and rewards. The Gymnasium toolkit was selected in the creation of the custom environment because it has well-written documentation and is the most used Python environment-building library ​(Brockman et al., 2016)​. The framework's goal is to have a multitude of environments with the same interface, such that it acts both as a collection of benchmarks and as an accessible springboard for reinforcement learning applications. To expand on the latter point, apart from the pre-built environments, the framework also supports the building of custom environments using its default classes as building blocks.

All Gymnasium environments interact with the agent as follows: the agent makes an action for each step. Then it receives an observation and a reward. The agent's aim is to maximize some metric of the reward, depending on the type of agent.

 Reinforcement Learning theory states that environments must follow the Markov Property, where each of their states summarize completely all relevant information ​(Sutton & Barto, 2015).​ I.e., the system should provide concise information about the future of the system in a way that is independent of previous states. The response of such Markovian environment at time must only depend on the action taken and the environment's response at time . Sutton and Barto provide a mathematical representation of a Markovian environment:

(3.1)

Where is the probability of transitioning to state with reward , from state with action , and is the complete probability distribution. Note how atonly depends on the state and action taken at . In real-world applications the Markovian property is maintained only as an assumption. In Reinforced Learning, the state needs to be as good as possible of a basis to predict future actions and rewards ​(Sutton & Barto, 2015)​.

The environment has two collections, namely the action space and the observation space. In the case of our custom environment implementation the action space is discrete, having the values 0, 1 and 2. Each is mapped to the possible actions 'Discharge', 'Nothing' and 'Charge'. The observation space is continuous, and contains data about carbon intensity, electricity consumption, battery charge value and previous emissions delta.

Lastly, the simulated environment rewards agents when the absolute value of the difference between emissions emitted with the battery and baseline emissions increases. A formalized mathematical description of the proposed reward function:

(3.2)

(3.2)

## 3.2 Deep Q-Network Algorithm

## 3.3 Proximal Policy Optimization Algorithm

# 4 Results

# 5 Discussion

# 6 Conclusion