

# Multi-Agent Reinforcement Learning for Strategic Bidding in Two Stage Electricity Markets

Francesco Morri<sup>1</sup>,  
Hélène Le Cadre<sup>1</sup>,  
Luce Brotcorne<sup>1</sup>,  
Pierre Gruet<sup>2</sup>

<sup>1</sup>Inria Center of Lille University, France

<sup>2</sup> EDF R&D and FiME, France

*E-mail addresses:* francesco.morri@inria.fr

Our goal is to study the dynamics of electricity markets involving multiple competitive generators through multi-agent reinforcement learning (MARL) approaches. We start by formulating the electricity market as a two-stage stochastic game, involving a finite set of conventional and renewable energy producers, which bid on the day-ahead market, and an Independent System Operator (ISO), which is responsible for the clearing of the market. We assume that a predetermined part of the producers are non-strategic, bidding at their marginal costs, while the others might bid strategically trying to learn the outcome of the clearing. In the first stage, the strategic producers optimize simultaneously their bids by minimizing their expected costs (opposite of the expected profits), which is the difference between their production cost and the payment they receive from the ISO. The renewable energy producers' objective functions include a penalty assigning a cost to the imbalances caused by their forecast errors. In the second stage, the ISO receives the bids of all the producers. It clears the market by determining the activated volumes for each producer and a price minimizing the total cost under capacity constraints, including a conditional value at risk (CVaR) constraint for the renewable producers, capturing the risk aversion level that the requested volume violates their uncertain capacity. We derive closed form expressions for the producers' best-responses considering pay-as-clear and pay-as-bid as pricing schemes, and simulate the market dynamics, using MARL. To that purpose, we rely on modified versions of two actor-critic algorithms: Deep Deterministic Policy Gradient and Soft Actor-Critic. The simulations show how the producers adapt dynamically their strategies to learn the best bidding strategy, under limited information exchange. Finally, we identify conditions for the convergence of MARL algorithms to local equilibria of the stochastic game.