Strategy-proof local energy market with sequential stochastic decision process for battery control

Diego Kiedanski, Daniel Kofman, José Horta, David Menga

Strategy-proof local energy market with sequential stochastic decision process for battery control

Diego Kiedanski*, Daniel Kofman*, José Horta*, David Menga†

*Telecom Paristech, 23 Avenue d'Italie, Paris, France
Emails: {diego.kiedanski, jose.horta, daniel.kofman}@telecom-paristech.fr

‡EDF R&D, EDF Lab Paris-Saclay, 91120 Palaiseau, France
Email: david.menga@edf.fr

***Abstract*—Low voltage distribution networks were not designed to support massive deployment of distributed energy resources (DER) such as solar panels, which is currently hindering the Energy Transition. Recent research contributions have shown that local energy markets improve the capacity of distribution grids to host DER. In parallel, distributed models were created to deal with the control of batteries in presence of stochastic demand and production, as well as variable electricity prices. The combination of the two techniques has received little attention until now, from both the literature and the industry. In this paper we extend the traditional approach to sequential stochastic decision processes by also modeling the interaction with the neighborhood in addition to the utility. The model is then solved using reinforcement learning techniques. For the local energy market we use MUDA: a strategy-proof multi-unit double auction.**

The performance of the proposed system is evaluated through simulations, demonstrating its capacity to effectively decrease the overall exchange of energy with the grid and the monetary cost for users.

***Index Terms*—ADP, Battery, Smart Grid, Energy Market, Auction**

I. Introduction

Massive deployment of renewable energy sources at residential premises, mainly solar panels, together with local storage, are becoming an important element of the Energy Transition. Distribution electricity networks have not been designed for supporting such massive deployment and therefore novel paradigms are required to avoid expensive investments that would delay exploiting the benefits of the mentioned deployments for both the prosumers and for the distribution networks operators.

The generation of renewable energy on its own might be more detrimental than beneficial if there is not a proper shift in the demand curve to accompany the production. Local batteries can be used to mitigate, at least partially, this problem. Nevertheless, an autonomic architecture is required at the customer side to decide and schedule the different possible flows of energy. What is more, households with intelligent architectures and batteries can perform load shaping without modifying the their energy demand, i.e: without compromising comfort.

The deployment of local markets at the neighborhood level has several benefits in addition to enabling the provision of advanced services. Such markets, and the possibility of distribution network operators (DSO) to induce prosumers behavior, when adequately designed, can provide a high performing distributed scheduling. It is important to note that the utility still plays a role in the market (prosumers can buy and sell energy from/to the utility). Local markets can reduce the energy and power through the last Medium Voltage/ Low Voltage (MV/LV) transformer and maximize the benefits of local renewable energy deployment, while respecting quality of electricity supply constraints and guaranteeing the stability of the network. There are three main limitations of previous proposals: either they are only strategy proof [1] (SP) in the price and not in the quantity, they assume that agents are deterministic, the stochastic optimization is centralized or some combination of the above.

Our contribution consists of a framework that combines and extends two techniques that had been applied with great success in the Smart Grid but that had not been used in conjunction before: a sequential stochastic decision process (SSDP) and a multi-unit double auction. In particular: a) we use an auction that is SP in the quantity bid as well as the price, b) the proposed SSDP is decentralized, requires no information about other prosumers or complex forecasts and seamlessly integrates the interaction with other households through a market, c) through simulations we show that both prosumers and the DSO benefit from its use.

A. Related Work

An et.al. [1] propose a framework for the exchange of renewable energy among microgrids (MG). Their approach consists of two-stage stochastic programming optimization followed by a distributed double auction. The optimization problem is solved by a centralized controller, which needs to have knowledge of energy production and demand of each MG. With respect to their proposed auction, it's built as a distributed algorithm that maximizes social welfare.

[1]In a strategy-proof auction, there is no incentive to lie when bidding.

In [2], Etesami et. al. use stochastic game theory to model the interactions between different prosumers and a utility company. Prosumers are not able to exchange energy among them but might buy from the utility using their closest substation. Although they take into account the stochasticity of the problem, the flow of energy back to the grid is forbidden.

In the work of Horta et.al. [3], a double auction mechanism is proposed for the exchange of energy among household. Prosumers optimize the use of their battery based on a linear programming problem assuming a perfect forecast, trade their missing/surplus of energy in the market and re-adjust the battery usage according with the traded energy and it's price.

Bessler [4] presents an algorithm to replace double auctions in local energy markets. In the algorithm, producers propose an energy price and consumers pay the price if it is below a private threshold. Analogously to the ultimatum game, the prosumer is forced to offer competitive prices or sell for cheap at the Feed in Tariff (FIT). Because the only subgame-perfect equilibrium in the ultimatum game is to offer an unequal trade [5], it is unclear whether the proposed mechanism is SP.

Several Approximate Dynamic Programming (ADP) approaches have been proposed. Most of them resemble the basic structure proposed in [6] and give good results but they either do not allow selling energy back to the grid or do not model the interaction between households. In [7], a multi-battery arbitrage problem is solved in presence of wind supply and with the ability to sell back to the grid. They do not allow however to sell the extra supply of production to the grid without storing it first in the battery. [8] uses the same approach that [7] for several controllable devices and solar energy but does not model the ability to sell back to the grid. Wei et.al. [9] propose a new algorithm to solve the ADP problem for the same scenario as above, but forbids the flow from the battery to the grid.

II. Proposed Framework

Let $\mathcal{N}$ denote a given set of prosumers connected to the same Low Voltage (LV) network. Each of them might have a battery and/or a photovoltaic panel (PV). For each timeslot t in a fixed time horizon $\mathcal{T}$, each prosumer will decide: how much she is going to charge/discharge his battery and how much energy she has to buy or sell.

Prosumers can buy (sell) energy from (to) the utility or by participating in the local energy market. The only objective of each household is to reduce their electricity bill. Households do not have access to any information about other prosumers. The only exception being the result of the auction. Even then, it is the result of an aggregated response and not an individual action. During this study we considered a time horizon of 1 day and timeslots of 10 minutes each.

A. Stochastic Sequential Decision Process Structure

The SSDP follows the basic structure of an ADP program as described in [6]. For each prosumer p and every timeslot t, Figure 1 shows the sequence of events involved in the model. First, p observes the current state (I) and decides how

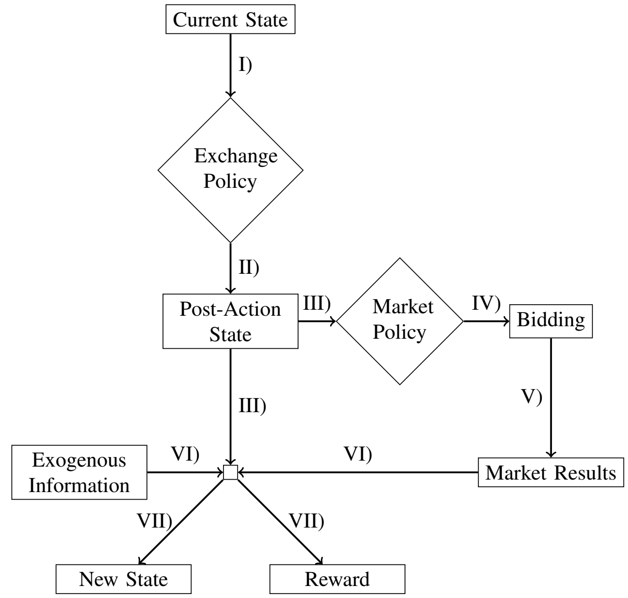

Fig. 1. Flowchart of the different actions taken and information received by a given prosumer in a timeslot.

much energy to exchange with the market, the utility and the battery in the current timeslot by following the *exchange policy* (II). After taking the action, observes a post-action state and decides how much to offer in the market (III) that will occur at the end of the timeslot (and for which the energy will be exchanged in timeslot $t + 1$). This is achieved by following the *market policy*. Finally, the prosumer waits for the results for the market and external information (VI) that arrives just before the end of t. With this information computes the reward of the period and the new state (VII). This process is repeated for each time step (144 times).

The state stores several pieces of information: the battery state of charge (SoC) at the beginning of timeslot t (B_t), the energy available from production P_t, the energy to be consumed L_t, the price of selling and buying energy to the utility, a bound on the energy to be exchanged in the market mq_t, the market price for that quantity and a forecast of P_{t+1} and L_{t+1}. Of the above, all but B_t are considered exogenous information (which we will denote W_{t+1}), because it needs to be provided from an outside source such as a smart meter (B_t is the exception because it can be computed from B_{t-1} and the action taken). It also encodes information about the prosumer that does not change with time such as bounds on the energy that can be traded in a given period with the battery and the grid (C_b and C_g), the maximum and minimum SoC and the reservation price, which is a measure of how much money a prosumer wants to obtain from trading in the market (it is used to compute the desired price offered at bidding time).

An action is a tuple of the form: $x_t = (b_t, m_t, g_t)$, where b_t, m_t, g_t is the amount of energy exchanged with the battery, the market and the grid respectively. A negative value of b_t means that the battery is charging, while positive values of m_t and g_t represent energy that is sold. Figure 2 depicts the

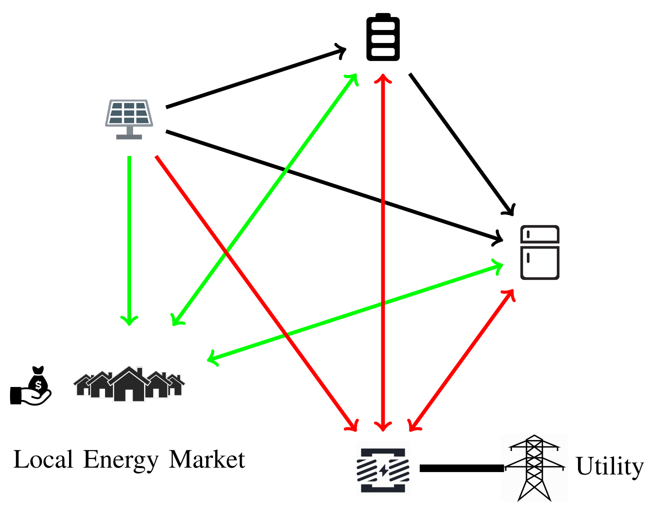

Fig. 2. The arrows indicate the direction in which energy flows are allowed. Black arrows are for a household internal flows, green arrows are for flows exchanged with the neighborhood and red arrows are flows exchanged with the utility through the last MV/LV transformer. The fridge represents consumption in the household.

allowed energy flows without taking into account constraints. Actions are chosen by following a policy; a deterministic mapping $X_t^\pi : \mathcal{S} \to \mathcal{X}_t$ given by equation 1:

$$X_t^\pi(S_t) = \arg\max_{x_t \in \mathcal{X}_t} \{r_1(S_t, x_t) + \gamma V_t^\pi(S_t^x)\} \tag{1}$$

In equation 1, $V_t^\pi(S)$ represents the *value-to-go* to be obtained by being in state S the next timeslot and γ is a discount factor. S_t^x is the post-action state. It can be seen as an intermediate state between S_t and S_{t+1}. If B_t was the battery SoC at the beginning of the timeslot, the battery level in the post-action state (B_{xt}) already reflects the effect of taking action b_t: $B_{xt} = B_{t+1} = B_t - b_t$. On the other hand, information such as the production of next period (P_{t+1}) is not available. $r_1(S_t, x_t)$ is a reward (or penalty) for taking action x_t while being in state S_t. It is the product of the traded energy by its price.

Finally, $\mathcal{X}_t$ is the space of feasible actions which is defined by the following constraints:

$$\begin{cases} 0 \le m_t \le mq_t & \text{if } mq_t > 0 \\ 0 \ge m_t \ge mq_t & \text{otherwise} \end{cases} \tag{2}$$

$$P_t - L_t + b_t = m_t + g_t, \tag{3}$$

$$|b_t| \le C_b, \quad |m_t + g_t| \le C_g, \quad m_t g_t \ge 0 \tag{4}$$

The ideal quantity to offer in the market is m_{t+1} (so that the bound mq_{t+1} is tight). There is not enough information to compute this value, $X_{t+1}^\pi(S_{t+1})$, but can be approximated. This can be done by using the value of m_{xt} in $X_{t+1}^\pi(S_t^x)$ (In equation 3, production and consumption are replaced by their respective forecast, and in r_1 the real price is replaced by a desired or wished price).

The exogenous information W_{t+1} together with the market results become available to the prosumer just before the end of the period. To make the model more realistic, we account for the small errors in P_t and L_t, by including a random normal distributed noise in W_{t+1}. No decision is taken regarding the noise: it is absorbed by the battery if it does not violate any constraints or it is sold/bought from the utility at the appropriate cost.

With knowledge of the exogenous information, the transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ and the reward function $r_t = r(S_t, x_t, W_{t+1})$ are computed. The transition function solely updates the post-action state with the new reading from the smart meter, as well as the forecasts and energy prices for the next period. The reward function is the exact economic cost incurred by the prosumer in the timeslot.

1) Objective and Learning: The objective of each prosumer is to maximize the discounted reward obtained:

$$\max_{\pi \in \Pi} \mathbb{E}^\pi \{\sum_{t=0}^{T} \gamma^t r(S_t, x_t, W_{t+1})\} \tag{5}$$

where maximizing in Π represents choosing the best policy. In turn, that means to find the optimal value of V_t^π. Because it is impossible to compute its value exactly, a standard technique in ADP known as Value Function Approximation (VFA) is used. Several kinds of VFAs exist. In the present work we use a lookup table approximation around the post-action state. In order to learn this approximation the reinforcement learning algorithm named *temporal difference with* $\lambda = 0$ is used. The process of learning the VFAs is carried out by simulating the interaction of several households during N days (called episodes). After every timeslot t of every episode n, each prosumer updates their VFA according to equation 6.

$$V_{t-1}^n(S_t^x) = (1 - \alpha_t)V_{t-1}^{n-1}(S_t^x) + \alpha_t \hat{v}_t \tag{6}$$

where α_n is the learning rate and $\hat{v}_t$ is the value of the maximum in $X^\pi(S_t)$. Pseudo code of the algorithm is available at [6, p. 391]. The non-stationarity property of the market result makes the application of models requiring known probability distributions hard to apply. There might be, however, simpler approaches than reinforcement learning which are applicable in this same scenario. The optimality of those techniques is outside of the scope of this work.

B. Auction

A multi-unit double auction is a mechanism for allocating goods (energy in this context) based on several offers by both: buyers and sellers. Every prosumer might place an offer to buy a quantity q of energy at price p, therefore becoming a buyer. The same holds for selling. After receiving all the offers, the market clears and the auctioneer notifies each participant of the result: a quantity and price. In this framework, we use MUDA, an auction proposed by Segal-Halevi [10]. In [10] they prove that the following properties hold:

- Individually Rational (IR): prosumers never lose money by participating in the auction.

- Weakly Budget Balanced (WBB): the auctioneer never loses and might earn money (which can be used cover the expenses of running the auction). This is achieved by charing small fees only to the winners. The fees are small enough that IR is preserved.
- Asymptotically Efficient: the efficiency of the auction increases with the number of participants.
- Strategy-proof: participants do not have an incentive to lie on the quantity bid nor in the price asked. Strategy-proofness is an important property because it eliminates the need to learn the best bidding strategy. The only optimal strategy is to tell the truth. Previous works such as [3] used auctions which where strategy proof only on the price, but not the quantity.

The main idea for guaranteeing strategy-proofness in both quantities is to include a stochasticity in the algorithm.

III. Numerical experiments

ADP has already given good results for the arbitrage problem in presence of stochastic errors in load and production forecasts. Therefore, we focus on the evaluation of the proposed auction in a random environment. To achieve this, we ran several simulations with and without market, to quantify the effect the later has in the performance of the model. Four cases were considered varying the number of prosumers and the usable size of their batteries:

- Case I: 100 prosumers, 3 kWh battery
- Case II: 100 prosumers, 6 kWh battery
- Case III: 49 prosumers, 3 kWh battery
- Case IV: 49 prosumers, 6 kWh battery

In each of the cases, 50% of the prosumers had a battery and a PV, 25% only a PV and 25% only a battery. PV panels had a rated peak of production at 6kWh.

For each prosumer, realistic synthetic load and production curves where used. The consumption data was obtained from SMACH [11].

Utility price consist of two levels, a cheaper one (c€0.15) in the early hours of the day ($t < 90$) and more expensive (c€0.3) during the rest of the day. The FIT was fixed at c€0.1. Reservations prices where taken equidistant in the interval $r \in [0, 0.3]$ and randomly distributed among prosumers. As a result, a prosumer p with reservation price r_p, offers in the market at price $FIT \times (1 + r_p)$ if selling and $pU_t \times (1 - r_p)$ if buying.

At each timeslot t, the error in the forecast of a prosumer p is modeled as $X_{p,t} \sim \mathcal{N}(0, \frac{1}{288})$ i.i.d. By doing this, we can guarantee that the overall error $\sum_{t=0}^{144} X_p, t \sim \mathcal{N}(0, \frac{1}{2})$ is small (less than 10% of the rated peak of production).

Because the outcome of the auction is random, for each time step the auction is executed 20 times and the instance with higher total profit is used.

Because batteries are allowed to end up in a state different than the starting one, we assume that just before the end of the time horizon, batteries are returned to their original state. If t_1 is chosen appropriately (for example, 1am), it's safe to assume that the process of restoring the batteries will not cause negative side effects. To account for this situation, all plots shown in this section include the amount of energy needed for the restoration process. In particular for Subsection III-B, the price at which this energy was traded was assumed to be the worst possible.

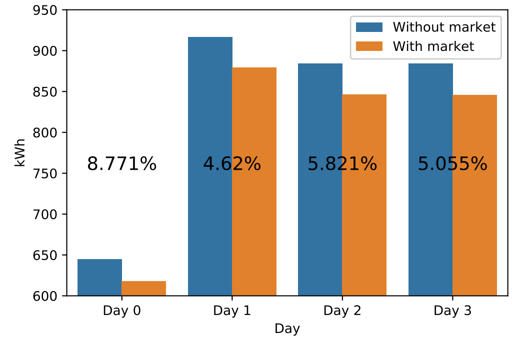

Fig. 3. TG reduction in the neighborhood at four different days. For each day the result is shown when the market is present and when it's not. On top of each pair of bars is the overall reduction of the day.

A. Reduction of energy exchanged through the last MV/LV transformer

Let $D_{t,p} = P_{t,p} - L_{t,p}$ be the demand of a given prosumer p at time t (positive if there is an excess of production and negative otherwise). If batteries are to be used only as temporary storage and their energy level should remain the same at the end of the time frame, the total amount of energy that must be exchanged with the grid is $\mathcal{D} = \sum_t \sum_p D_{t,p}$.

If $G_{t,p}$ denotes the amount of energy traded with the grid at time t by prosumer p, the energy exchanged with the grid at time t is given by $G_t = \sum_p G_{t,p}$. Then, the net energy exchanged with the grid (irrespective of the direction of the flow) is given by $TG = \sum_t |G_t|$. In an ideal case, TG would be bounded by $\mathcal{D}$ (it might be lower if there is excess of production and it's stored on batteries), but because load and production curves are not necessarily synchronized, it might be necessary for the system to *loan* from or to the grid in other to deal with a peak at a given time. As the energy market matches production and demand inside the neighborhood, it has a positive impact in the amount of TG. As an example, Figure 3 depicts the TG of a neighborhood under Case IV). Although there is some variance as the consumption of each day is different, in every case, the proposed mechanism effectively reduces the total TG.

To understand how some of the parameters impact the performance, Figure 4 shows the TG with and without market for the four different cases.

Battery size has an important role in TG as it can be seen that fixing the amount of prosumers and increasing their battery size (Case I $\rightarrow$ Case II, Case III $\rightarrow$ Case IV) achieves a reduction, which is almost twice the amount for the Cases I/II. This fact is probably due to the ability of prosumers to

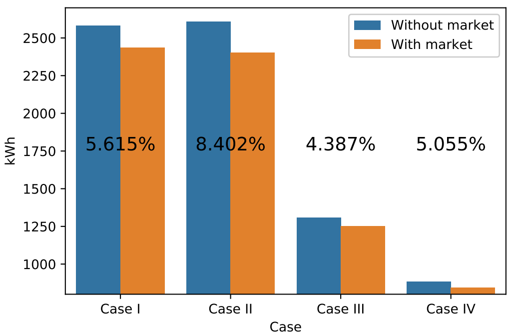

Fig. 4. TG reduction in four different scenarios with and without market. Shown on top of each pair of bars is the overall reduction of the case.

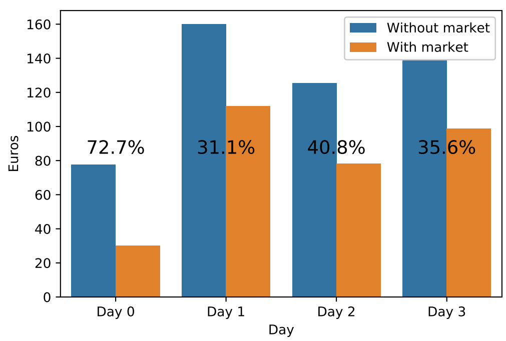

Fig. 5. Cost reductions in the neighborhood at four different days. For each day a scenario with and another one without market are considered. Shown on top of each pair of bars is the overall reduction of the day.

better store the produced energy. What is more, increasing the amount of prosumers also yields a reduction in TG (Case III $\rightarrow$ Case I, Case IV $\rightarrow$ Case II). In this case, the effect is caused by the efficiency of the market, which increases with the amount of participants.

As it is impossible to design an auction that is IR, strategy proof, budget balanced and efficient [12], mechanism tend to be efficient only asymptotically. By including only 50 prosumers, the mechanism is far from the limit and improvements are smaller than expected.

B. Financial viability

Although it is IR to participate in the market, it might be possible that in a less trained version of the model, chosen actions result in an overall increased cost. This might happen, for example, if a greedy prosumer sold all her production in the market, but has to later buy the same amount she sold to the grid at a higher price. For Case IV) the total cost incurred by prosumers with and without market is plotted in Figure 5. In every day, there is a considerable reduction in the overall cost incurred by the neighborhood. This fact will ease the deployment of the proposed model, as users can directly benefit from its presence.

IV. Concluding remarks

Local energy markets represent an important enabler for the energy transition, easing the integration of renewable energies at residential distribution grids by exploiting demand flexibility from elastic loads or storage technologies. In this paper we provide considerable improvements to these markets in terms of mechanism design, by using a strategy-proof auction with respect to both price and quantity, and in terms of demand flexibility allocation under uncertainty, by proposing a stochastic model for the control of batteries, which can be solved using reinforcement learning techniques. We show by simulation under realistic scenarios that the proposed system achieves important reductions in the net energy exchanged with the grid as well as in monetary cost for the prosumers. These results could be further improved by considering more advanced learning techniques, Deep Q-Networks being one example. A natural extension of the model is to allow for two batteries, one of which belongs to an Electric Vehicle. The proposed framework has the advantage that for such an inclusion, only small changes must be done. One of which is the replacement of lookup tables (which are known to suffer from the curse of dimensionality) for other already existing approaches such as concave function approximation.

References

[1] D. An, Q. Yang, W. Yu, X. Yang, X. Fu, and W. Zhao, “Sto2Auc: A Stochastic Optimal Bidding Strategy for Microgrids,” *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2260–2274, 2017.

[2] S. R. Etesami, W. Saad, N. Mandayam, and H. V. Poor, “Stochastic Games for Smart Grid Energy Management with Prospect Prosumers.”

[3] J. Horta, D. Kofman, D. Menga, and A. Silva, “Novel market approach for locally balancing renewable energy production and flexible demand,” no. October, pp. 533–539, 2017. [Online]. Available: http://arxiv.org/abs/1711.09565

[4] S. Bessler, “An algorithm for renewable energy allocation and trading in a microgrid,” in *2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2018, pp. 1–5.

[5] M. A. Nowak, K. M. Page, and K. Sigmund, “Fairness versus reason in the ultimatum game,” *Science*, vol. 289, no. 5485, pp. 1773–1775, 2000. [Online]. Available: http://science.sciencemag.org/content/289/5485/1773

[6] W. B. Powell, *Approximate Dynamic Programming*, 2011, vol. 2. [Online]. Available: http://web.mit.edu/dimitrib/www/dpchapter.pdf

[7] D. Salas and W. B. Powell, “Benchmarking a Scalable Approximate Dynamic Programming Algorithm for Stochastic Control of Multidimensional Energy Storage Problems,” *Informs J. Comput.*, no. 2004, pp. 1–41, 2015. [Online]. Available: http://www.castlelab.princeton.edu/Papers/SalasPowell-BenchmarkingADPformultidimensionalenergystorageproblems.pdf

[8] C. Keerthisinghe, G. Verbič, and A. C. Chapman, “Energy management of PV-storage systems: ADP approach with temporal difference learning,” *19th Power Syst. Comput. Conf. PSCC 2016*, pp. 1–7, 2016.

[9] Q. Wei, G. Shi, R. Song, and Y. Liu, “Adaptive Dynamic Programming-Based Optimal Control Scheme for Energy Storage Systems with Solar Renewable Energy,” *IEEE Trans. Ind. Electron.*, vol. 64, no. 7, pp. 5468–5478, 2017.

[10] E. Segal-Halevi, A. Hassidim, and Y. Aumann, “MUDA: A Truthful Multi-Unit Double-Auction Mechanism,” pp. 1–18, 2017. [Online]. Available: http://arxiv.org/abs/1712.06848

[11] E. Amouroux, T. Huraux, and F. Sempé, “Simulating human activities to investigate household energy consumption,” *Int. Conf. Agents Artif. Intell.*, 2013. [Online]. Available: http://perso.limsi.fr/sabouret/ps/icaart2013.pdf

[12] R. B. Myerson and M. A. Satterthwaite, “Efficient Mechanisms for Level- k Bilateral Trading,” vol. 1, no. 339179, p. 2012, 2015.