Flexibility Trading Through Load Reshaping in Energy Communities Using Deep Reinforcement Learning

M. Asim Amin*, **Renato Procopio**, **Marco Invernizzi**
Department of Electrical, Electronics, Telecommunication Engineering and Naval Architecture, University of Genova, Genova I-16145, Italy, ***asim.amin@edu.unige.it***
Milutin Petronijević
Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva St. 4, RS-18000 Niš, Serbia
Ivana Radonjic Mitic
Faculty of Sciences and Mathematics, University of Niš, Višegradska St. 33, RS-18000 Niš, Serbia

Introduction

The Energy Community (EC) plays an essential role in the active engagement of end-users and deployment of Renewable Energy (RE) and contributes to net-zero carbon emissions. In EC, several entities/homes join to exchange a surplus or demand reduction during peak hours with the grid. In the proposed work, the end user offers flexibility by adjusting the load demand by following different load-shaping patterns and avoiding sharing flexibility during critical hours. Therefore, load reshaping is proposed in EC to provide a surplus to the aggregator and further share the surplus with the grid to overcome the uncertainty and intermittency caused by the growing number of RE and inadequate power system flexibility. Hence, the paper focuses on developing a Deep Reinforcement Learning (DRL) based flexibility aggregation from the end-user and offering it to meet the grid demand by following the limit constraints and increasing the economic benefits of the end-user. In addition, the proposed approach is validated with the case study by comparing several algorithms to see individual entity/home contributions and meet the overall cost minimization objective.

The EC is essential in energy decentralization by integrating end-users and providing benefits for the environment, economy, and society [1]. The flexibility contribution by end-users ensures the stability of grids, and critical issues such as capacity adjustment, scheduling, and market mechanisms must be considered.

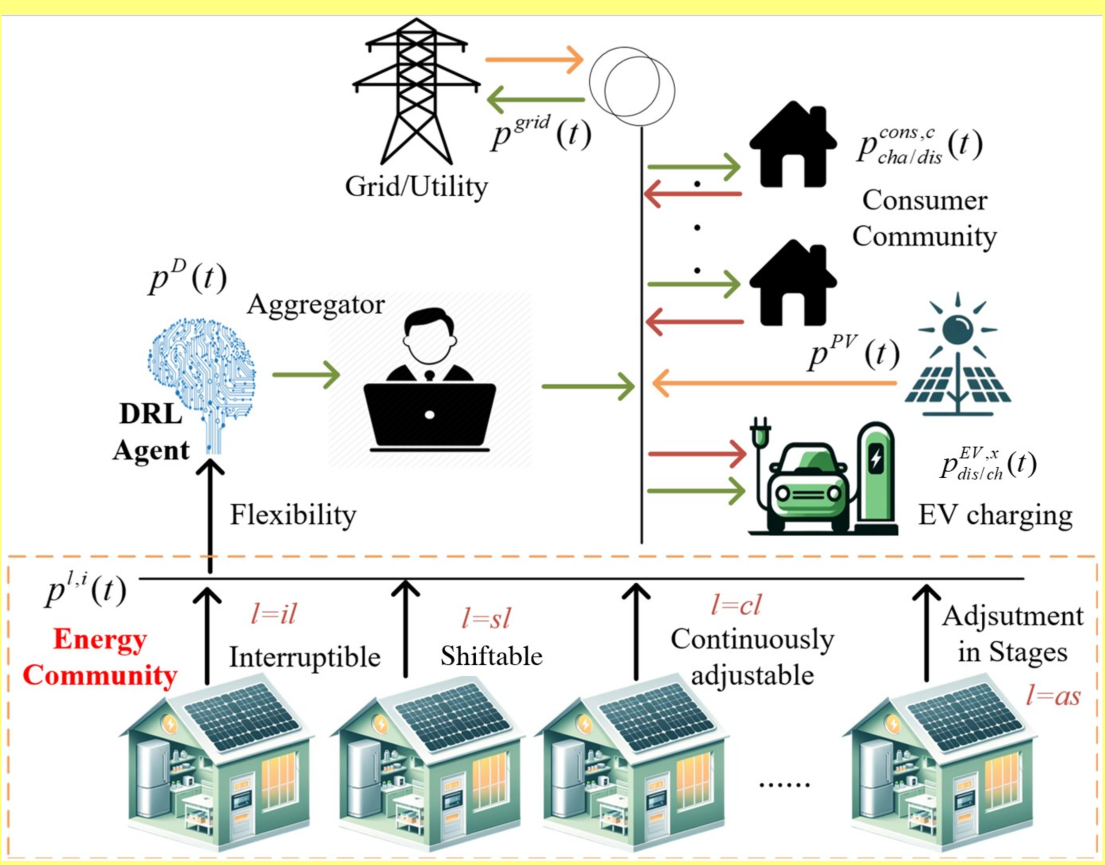

Fig. 1. EC load reshaping DRL-based framework

Proposed Approach

In the proposed work, the DRL-based framework is developed to integrate several load types to offer flexibility to the aggregator, as given in **Fig.1.**

The considered model for implementing different load types in EC from the ref [2] includes modelling household appliances, battery storage, and photovoltaic (PV) for the formulation. Likewise, for the EV formulation, the model from [3] uses the EV demand charging time, followed by further battery charging/discharging constraints. Hence, the flexibility offered by each home with specific load type $\boldsymbol{l \in il, sl, cl, as}$, and **D is a total number of homes *i*** attached in EC, and it is expressed as:

$$p^{D}(t)=\sum_{i=1}^{D}\sum_{l} p^{l,i}(t),\ \forall l,t\in T$$

In addition, the power balance for each house is given as:

$$p^{bat,i}_{ch}(t)-p^{bat,i}_{dis}(t)+p^{PV,i}(t)-p^{adj,i}(t)-p^{noadj,i}(t)=p^{l,i}(t)$$

Hence, the overall power balance equation at the ECs and aggregator level is expressed as:

$$p^{dem}(t)=p^{EV,x}_{ch}(t)-p^{EV,x}_{dis}(t)-p^{cons,c}_{cha}(t)+p^{cons,c}_{dis}(t)+p^{grid}(t)+p^{PV}(t)$$

where EV charging/discharging $p^{EV,x}_{ch}(t)/p^{EV,x}_{dis}(t)$ constraints are considered. In addition, the offered flexibility by each home $p^{l,i}(t)$, PV production $p^{PV,i}(t)$, adjustable/non-adjustable $p^{adj,i}(t)/p^{noadj,i}(t)$, battery charge/discharge $p^{bat,i}_{ch}(t)/p^{bat,i}_{dis}(t)$, consumer community charging/discharging $p^{cons,c}_{ch}(t)/p^{cons,c}_{dis}(t)$, and c is the number of attached communities. In addition, the overall objective for EC is cost minimization and the associated costs.

$$Obj=\min\sum_{t}\left[\lambda_1 p^{D}(t)-\lambda_2 p^{dem}(t)\right]$$

where λ_1/λ_2 are the associated costs for demand balance between ECs and the power grid.

Deep RL Based Model

DRL-based methods are flexible in learning the optimal policy or actions by interacting with the agent and the environment, and the objective of the Markov Decision Process (MDP) is to find the optimal policy that maximizes the overall reward function. In the MDP environment, the agent gets states-spaces $s(t)\in S(t)$ from the environment, takes optimal action-space:

$$s(t)=\left[p^{dem}(t),p^{D}(t)\right],\quad \forall D,t\in T$$

$$a(t)=\left[p^{D}(t)\right],\quad \forall D,t\in T$$

In the reward function, the first term represents a demand penalty and the second term for the flexibility reward with weighting factor (α and β) tunned through the DRL-based training.

$$r(t)=\max_{t}\left[-\alpha\left(\frac{p^{dem}(t)-p^{D}(t)}{p^{dem}(t)}\right)^2+\beta\times p^{D}(t)\right]$$

To validate the DRL, PPO (Proximal Policy Optimization) offers a clipped substitute for the stable update, A2C (Advantage Actor-Critic) is an actor-critic network and less robust, and TD3 (Twin Delayed Deep Deterministic) shows continuous control by dealing with overestimation bias [4].

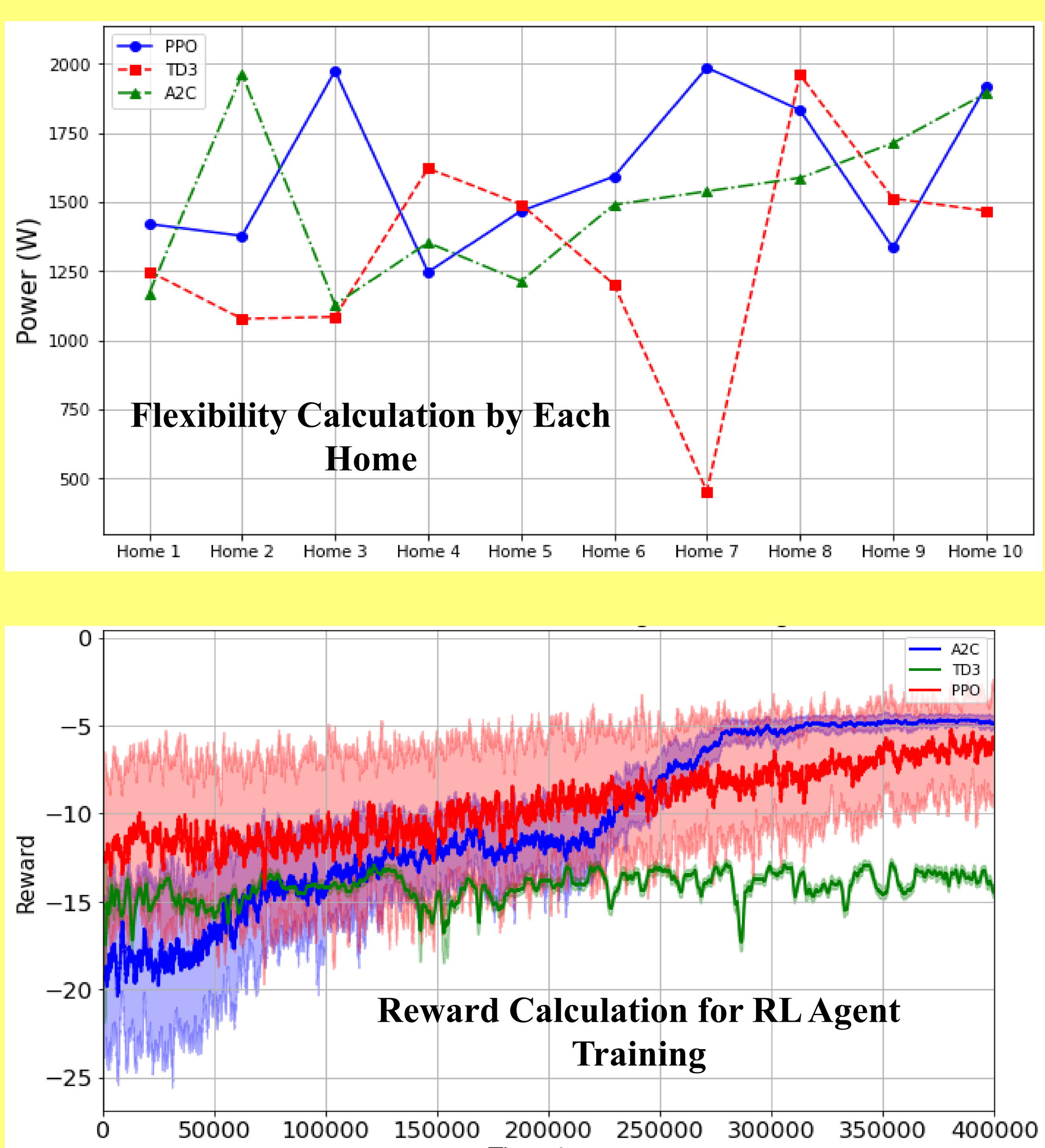

Fig. 2. Reward and flexibility calculation using DRL-based training over 400,000 timesteps (A2C, TD3, PPO)

In the comparison of the proposed algorithms, the PPO performs better in achieving the goal, and it can be seen from the flexibility contribution. From the average reward profile, PPO proves to have continuous learning and makes the balance between exploration and exploitation compared to A2C, and TD3. For the proposed approach, we have considered **10 ECs with 8 EVs** to be charged with different arrival/ departure times as **[3, 5, 7, 9, 13, 15, 18, 21]/ [11, 13, 15, 18, 21, 24, 27, 29]** hrs. Furthermore, all the ECs contribute to managing EV charging and share the surplus amount with the utility/grid, community charging/discharging, and PV usage by meeting the given objective function.

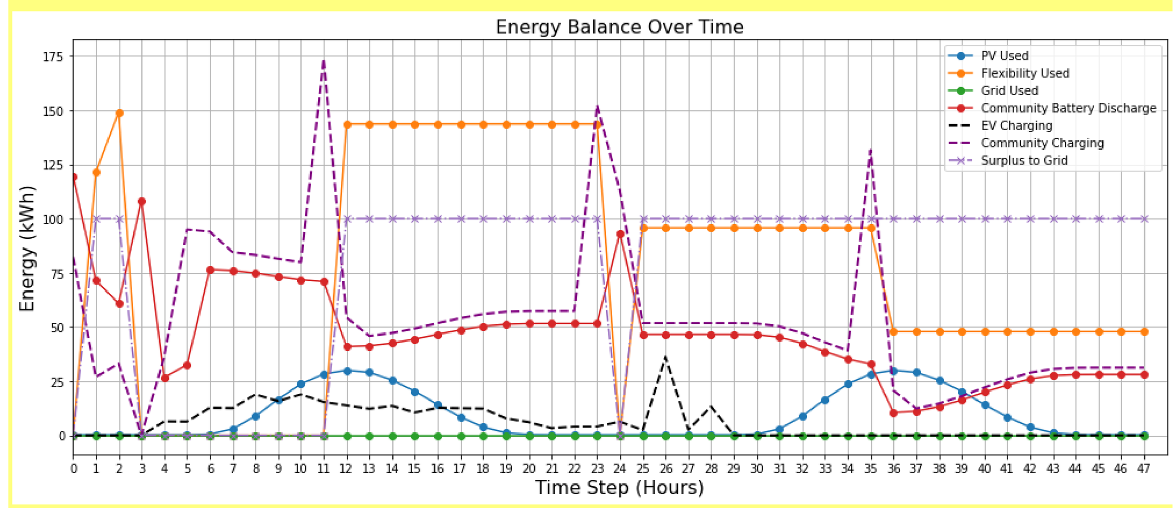

Fig. 3. Overall energy balance over time using DRL-based objective for 48 hrs.

As the PPO-based model performs better than the other RL-based algorithms, as it shows consistent learning over time, and the overall maximum contribution by each home is compared in the reward and flexibility contribution profiles.

PPO achieves a remarkable result over the other. From **Fig. 2**, it can be seen that the agent has learned the model and offers as much support as possible to maintain the balance between the available PV power and EV charging, and consistent support to the grid.

Meanwhile, try not to affect the user's comfort by giving them time to readjust their consumption. It is important to note that the agent resets over the day, and during the reset, the backup support is offered to maintain the balance through the consumer community to maintain the overall balance.

Conclusions

The main conclusions drawn from this work are:

- The work proposed a flexibility calculation from the end-user participation for remote ECs using DRL.
- In the environment, the flexibility calculation is optimized based on the load demand from consumer and EVs charging based on their arrival/departure time.
- Furthermore, the proposed framework is trained on several algorithms, and PPO is best performed compared to the other algorithms.

References

[1] M. A. Amin, R. Procopio, M. Invernizzi, A. Bonfiglio, and Y. Jia. Exploring the Role of Energy Communities: A Comprehensive Review. Energy Convers. Manag. X, 25, 100883 (2025).

[2] X. Wei et al., Two-Stage Cooperative Intelligent Home Energy Management System for Optimal Scheduling. IEEE Trans. Ind. Appl. 58, 5423–5437 (2022)

[3] M. A. Amin et al., Profit Maximization by Integrating Demand Response in Multiple VPPs Optimal Scheduling. Proc. of 2022 Int. Conf. Electr. Eng. Sustain. Technol. ICEEST, Pakistan (2022).

[4] A. Raffin et al., Stable-Baselines3: Reliable Reinforcement Learning Implementations. Journal of Machine Learning Research, 22, 1–8, (2021).