Balancing Cost and Self-Consumption in Home Energy Management: A Deep Reinforcement Learning and Model-Based Perspective with EV and Storage Systems

Ilias Kappos
Intracom Telecom
Research Collaboration Dpt
Athens, Greece

Sergio Martinez-Hermida
CITCEA-UPC
Universitat Politècnica Catalunya
Barcelona, Spain

Sara Barja-Martinez
CITCEA-UPC
Universitat Politècnica Catalunya
Barcelona, Spain

Christoforos Menos-Aikateriniadis
National Technical University of Athens
Intracom Telecom, Research Collaboration Dpt
Athens, Greece

Monica Aragüés-Peñalba
CITCEA-UPC
Universitat Politècnica Catalunya
Barcelona, Spain

***Abstract*—This study presents a comparative analysis of two home energy management system (HEMS) strategies: a model-based mixed-integer linear programming (MILP) approach and a reinforcement learning-based Deep Q-Network (DQN), under EV-PV and EV-PV-Battery Energy Storage System (BESS) configurations. The MILP approach serves as the optimality benchmark to assess the effectiveness of the proposed DQN strategy. A multi-objective function is introduced to balance cost minimization and self-consumption, allowing user-defined prioritization. Results show that MILP consistently outperforms DQN in cost savings and self-consumption. In the EV-PV setup, both methods achieve around 20% cost savings, though weekday EV charging restrictions limit self-consumption gains. Adding a BESS improves performance, with MILP reaching 28.16% cost savings and 21.59% self-consumption, compared to DQN's 19.77% and 10.05%. While MILP offers superior performance, DQN's adaptability to uncertainty makes it well-suited for real-time, dynamic environments.**

***Index Terms*—Battery energy storage system, deep q-network, deep reinforcement learning, electric vehicle, home energy management system, mixed-integer linear programming**

I. Introduction

A. Background and motivation

The variability of renewable energy sources (RES) and changing consumption patterns make it difficult to maintain a stable and reliable grid. Demand-side flexibility is essential in this context, as it enables adjustments in electricity consumption to match supply, helping to stabilize the grid and reducing reliance on fossil-fuel peaking plants. Smart home solutions, such as home energy management systems (HEMS), enable residential flexibility by optimizing scheduling and dispatching control signals to flexible assets like electric vehicles (EVs) and battery energy storage systems (BESSs).

Model-based optimization methods like mixed integer linear programming (MILP) provide optimal solutions but require full foresight or frequent re-optimization, making them computationally intensive. In contrast, deep reinforcement learning (DRL) methods have recently emerged as effective alternatives, thanks to their dynamic nature and near real-time decision-making capabilities, reducing computational time through pre-trained models. However, DRL solutions may be suboptimal due to reward-shaping limitations, while MILP ensures optimality when uncertainties like EV driver behavior are not considered. Additionally, DQN's lack of future foresight limits its performance in deterministic settings. This study uses MILP as a benchmark to assess the performance of the proposed DRL strategy.

B. Related work

Recent research has investigated model-based and DRL methods for intelligent EV charging, focusing on reducing electricity costs and optimizing PV self-consumption. Studies [1], [2] introduce MILP methods to optimize daily electricity costs by coordinating appliances, PV systems, and BESS under dynamic pricing. In [3], a multi-objective hybrid energy management system minimizes electricity expenses and the household greenhouse gas emissions. Similarly, [4] introduces an N-step DRL approach that prioritizes solar energy utilization for EV charging while reducing costs. End-user preferences have been considered in work [5] that proposes a DQN-based EV charging strategy for cost savings optimization without jeopardizing the time-related habits of end-users. Additionally, [6] introduces a safe reinforcement learning approach for multi-energy management, including EV and BESS, with dy-

This work was supported by the ODEON project (Horizon Europe Grant Agreement ID 101136128).

979-8-3315-2503-3/25/$31.00 ©2025 IEEE

namic price forecasting and cost minimization while satisfying constraints.

Despite their broad applicability, most research focuses on either model-based or DRL-based methods without evaluating and comparing their relative effectiveness in a HEMS environment. A few studies [7] [8] adopt a two-stage optimization approach: the first stage employs a DRL-based approach to schedule a specific asset (e.g., EV charging load). In contrast, the second stage incorporates this schedule into a MILP-based HEMS to manage additional flexible appliances and BESS. These hybrid methods combine MILP's optimality with DRL's adaptability, showing promise in reducing costs and improving system efficiency. However, neither of these studies uses MILP as benchmark to evaluate DRL performance under a flexible home environment using flexible key resources such as EVs and BESS.

C. Contributions

This work proposes an energy optimization framework for smart-connected homes, considering minimization of electricity costs and self-consumption maximization. The contributions of this paper can be summarized as follows:

- Proposes and evaluates two distinct optimization strategies: a two-step MILP-based and DQN-oriented approaches. Both energy optimization methods formulate a multi-objective problem, aiming to balance the trade-off between minimizing electricity consumption costs and maximizing self-consumption.
- Explores a flexible smart-connected home setup that integrates solar PV, EV, and BESS, operating under a dynamic real-time pricing scheme that supports home-to-grid interaction.
- Presents a realistic modeling framework that incorporates real-world household data while accounting for end-users historical charging preferences.

II. Home Energy Management System Framework

This work focuses on the optimal operation of a HEMS connected to the main grid, incorporating an EV, PV panels, a BESS and inflexible household loads. The EV charges or remains idle, while the BESS charges and discharges as needed. The HEMS's optimal decisions are modeled as continuous variables. The objective is to determine the optimal operation of the flexible assets based on dynamic pricing tariffs and PV generation.

A. Optimization Goals

The proposed method minimizes electricity costs and maximizes daily self-consumption (SC), formulating a multi-objective function that enhances energy savings for end-users compared to historical costs, thereby reducing household reliance on the power grid. These optimization goals are quantified through the Key Performance Indices (KPIs) shown in Eq. (1) and Eq. (2), respectively.

$$\text{Savings } (\%) = \sum_{t=0}^{T} \left(\frac{\text{Cost}_t^{\text{Baseline}} - \text{Cost}_t^{\text{Optimal}}}{|\text{Cost}_t^{\text{Baseline}}|} \right) \cdot 100 \quad (1)$$

$$\text{SC } (\%) = \left(\frac{1}{T} \cdot \sum_{t=0}^{T} \left[1 - \frac{\text{Energy}_t^{\text{buy}}}{\text{Demand}_t^{\text{total}}} \right] \right) \cdot 100 \quad (2)$$

B. MILP-Based Optimization Strategy

This method introduces a two-step optimization process. First, each objective is independently optimized to determine its optimal value. Then, the HEMS operates using the proposed multi-objective function.

- **Step 1a) Minimize costs**

$$\begin{aligned} f^{\text{cost }*} &= \min f^{\text{cost}} \\ &= \min \sum_{t}^{T} \left(C_t^{buy} \cdot \chi_t^{buy} \cdot \delta_t^{buy} - C_t^{sell} \cdot \chi_t^{sell} \cdot \delta_t^{sell} \right) \end{aligned} \quad (3)$$

- **Step 1b) Maximize self-consumption**

$$\begin{aligned} f^{\text{SC}*} &= \max f^{\text{SC}} \\ &= \min \sum_{t}^{T} \left(\chi_t^{buy} \cdot \delta_t^{buy} + \chi_t^{sell} \cdot \delta_t^{sell} \right) \end{aligned} \quad (4)$$

- **Step 2) Multi-objective function**

$$\min \left(\alpha \cdot \frac{f^{\text{cost}}}{f^{\text{cost }*}} + (1-\alpha) \cdot \frac{f^{\text{SC}}}{f^{\text{SC}*}} \right) \quad , \quad \alpha \in [0,1] \quad (5)$$

where C_t^{buy}, C_t^{sell} represent electricity purchase and sale prices, $\chi_t^{buy}, \chi_t^{sell}$ denote the amounts of electricity bought and sold, $\delta_t^{buy}, \delta_t^{sell}$ are binary variables restricting the system to either buy or sell at time t. The values obtained in steps 1a and 1b normalize each term of the multi-objective function, which is then optimized with a user-defined weight α.

Constraints: The three optimizations are performed with a series of constraints that must be met:

1) Energy balance:

$$\begin{aligned} P_t^{inflex} - P_t^{PV} &= \left(\chi_t^{buy} \cdot \delta_t^{buy} - \chi_t^{sell} \cdot \delta_t^{sell} \right) \\ &+ \left(\sigma_t^{BESS,dis} \cdot \delta_t^{dis} - \sigma_t^{BESS,ch} \cdot \delta_t^{ch} \right) - \sigma_t^{EV,ch} \end{aligned} \quad (6)$$

where P_t^{inflex} is the household inflexible power load, P_t^{PV} is the PV generation, $\sigma_t^{BESS,ch}, \sigma_t^{BESS,dis}$ are the charging and discharging power of the BESS, $\sigma_t^{EV,ch}$ the charging power of the EV (only allowed to charge), and $\delta_t^{ch}, \delta_t^{dis}$ are binary variables that prevent simultaneous BESS charging and discharging.

2) Maximum contracted power from/to the grid:

$$0 \leq \chi_t^{buy} \leq P_{max}^{buy}, \quad 0 \leq \chi_t^{sell} \leq P_{max}^{sell} \quad \forall t \quad (7)$$

3) EV SoC update:

$$SoC_t^{EV} = SoC_{t-1}^{EV} + \eta^{EV} \cdot \frac{\sigma_t^{EV,ch}}{N^{hour}} \quad (8)$$

4) BESS SoC update:

$$SoC_t^{BESS} = SoC_{t-1}^{BESS} + \frac{\sigma_t^{BESS,ch}}{N^{hour}} \cdot A^{ch} - \frac{\sigma_t^{BESS,dis}}{A^{dis} \cdot N^{hour}} \tag{9}$$

5) BESS SoC at the end of the day:

$$SoC_{t=0}^{BESS} = SoC_{t=T}^{BESS} \tag{10}$$

6) BESS Maximum charging/discharging power:

$$0 \leq \sigma_t^{BESS,ch}, \sigma_t^{BESS,dis} \leq P_{max}^{BESS} \quad \forall t \tag{11}$$

7) BESS Maximum and Minimum SoC allowed:

$$SoC_{min}^{BESS} \leq SoC_t^{BESS} \leq SoC_{max}^{BESS} \quad \forall t \tag{12}$$

8) EV starting SoC:

$$SoC_{t=0}^{EV} = SoC_{\max}^{EV} - \eta^{EV} \cdot \frac{P_{day}^{EV}}{N^{hour}} \tag{13}$$

9) EV daily load to match historical:

$$P_T^{EV} = \sum_{t=1}^{T} \sigma_t^{EV,ch} \in [0.97 \cdot P_{day}^{EV}, 1.03 \cdot P_{day}^{EV}] \tag{14}$$

10) EV charging preferences:

$$\sigma_t^{EV,ch} = 0 \quad \forall t \in H \tag{15}$$

where P_{max}^{buy} and P_{max}^{sell} denote the maximum buying and selling power to the grid, N^{hour} the time granularity factor, A^{ch} and A^{dis} the charging and discharging efficiencies of the BESS, P_{max}^{BESS} the BESS maximum charging/discharging power, SoC_{min}^{BESS} and SoC_{max}^{BESS} the BESS state of charge (SoC) limits, η^{EV} is the EV charging efficiency, P_{day}^{EV} the daily historical EV load, $SoC_{\max}^{EV}$ the EV's max SoC in kWh and H the hours when the EV is unavailable for charging. Analysis of the Pecan Street dataset [9] revealed that most users avoid charging between 9:30 and 17:00 on weekdays, with no similar pattern observed on weekends.

C. DQN-based Strategy

Two Deep Q-Network (DQN) models are proposed as alternatives to the MILP-based approach—one trained for weekdays and one for weekends, reflecting different historical charging patterns. The HEMS optimization is modeled as a 15-minute interval Markov Decision Process with the tuple $(\mathcal{S}, \mathcal{A}, R, P)$. At each 24-hour episode, the Agent observes the current state and selects actions for the EV (charge/idle) and BESS (charge/idle/discharge), receiving rewards based on performance.

1) State Space: Each state s_t encapsulates all the information needed to select an optimal action. It includes the household's inflexible demand P_t^{inflex}, PV generation P_t^{PV}, electricity prices C_t^{buy} and C_t^{sell}, state of charge of the EV and BESS (SoC_t^{EV}, SoC_t^{BESS}), and cumulative EV charging load P_t^{EV} up to time t. A binary flag u_t^{BESS} indicates whether the BESS can still reach its final SoC by the end of the time horizon. Temporal context is captured via $\sin\left(\frac{t}{T}\right)$, $\cos\left(\frac{t}{T}\right)$, and a day-type indicator where $day = 1$ for weekends and 0 for weekdays. Thus, the state space $\mathcal{S}$ is defined as:

TABLE I
SUB-REWARDS OF THE PROPOSED DQN

Symbol	Condition	Value
$r_t^{Day_{EV}}$	$a_t^{EV} > 0$, $P_t^{EV} \leq 1.03 \times P_{day}^{EV}$	$\frac{a_t^{EV}}{P_{max}^{EV}}$
	$a_t^{EV} > 0$, $P_t^{EV} > 1.03 \times P_{day}^{EV}$	-10
	$a_t^{EV} = 0$, $P_t^{EV} \leq 1.03 \times P_{day}^{EV}$	-1
	$a_t^{EV} = 0$, $P_t^{EV} > 1.03 \times P_{day}^{EV}$	0
$r_t^{C_{EV}}$	$a_t^{EV} > 0$, $C_t \leq Q_{cost}(0.25/0.5/0.75)$	+2/ +1 / -1
	$a_t^{EV} > 0$, $C_t > Q_{cost}(0.75)$	-2
$r_t^{C_{BESS}}$	$a_t^{BESS} > 0$, $C_t \leq Q_{cost}(0.25/0.5/0.75)$	+2/ +1 / -1
	$a_t^{BESS} > 0$, $C_t > Q_{cost}(0.75)$	-2
	$a_t^{BESS} < 0$, $C_t \leq Q_{cost}(0.25/0.5/0.75)$	-2/ -1 / +1
	$a_t^{BESS} < 0$, $C_t > Q_{cost}(0.75)$	+2
$r_t^{SC_{EV}}$	$a_t^{EV} > 0$, $P_t^{PV} > 0$ and $P_t^{net} > 0$	0.5
	$a_t^{EV} > 0$, $P_t^{PV} > 0$ and $P_t^{net} \leq 0$	1.5
r_t^{SC}	$r_{A,t}^{SC}$: $P_t^{net} + a_t^{EV} < 0$, $a_t^{BESS} > 0$	$+\frac{a_t^{BESS}}{P_{max}^{BESS}}$
	$r_{B,t}^{SC}$: $P_t^{net} > 0$, $P_t^{flex} < 0$, $P_t^{net} + P_t^{flex} = 0$	+1.5
	$r_{B,t}^{SC}$: $P_t^{net} \leq 0$, $P_t^{flex} \geq 0$, $P_t^{net} + P_t^{flex} = 0$	+1.5
	$r_{B,t}^{SC}$: Otherwise	-1
$r_t^{SoC_{EV}}$	$a_t^{EV} > 0$, $SoC_t^{EV} \geq SoC_{max}^{EV}$	-10
$r_t^{SoC_{BESS}}$	$a_t^{BESS} > 0$, $SoC_t^{BESS} \geq SoC_{max}^{BESS}$	-10
	$a_t^{BESS} < 0$, $SoC_t^{BESS} \leq SoC_{min}^{BESS}$	-10
$r_t^{Day_{BESS}}$	$\lvert E_{LB}^{BESS} \rvert > E_{t->T,max}^{BESS}$ or $\lvert E_{UB}^{BESS} \rvert > E_{t->T,max}^{BESS}$	-10
r_t^{pref}	weekday, $38 \leq t \leq 60$, $a_t^{EV} > 0$	-20

$$s_t = \Big(P_t^{inflex}, P_t^{PV}, C_t^{buy}, C_t^{sell}, SoC_t^{EV}, SoC_t^{BESS}, P_t^{EV}, u_t^{BESS}, \sin\left(\frac{t}{T}\right), \cos\left(\frac{t}{T}\right), day) \quad \forall t \in \mathcal{T} \tag{16}$$

2) Action Space: In this work, the Agent operates within a discretized action space since DQN does not support continuous actions. At each timestep t, the joint action is defined as:

$$\alpha_t = (\alpha_t^{EV}, \alpha_t^{BESS}), \quad \forall t \in T \tag{17}$$

The components are discretized as follows:

$$\alpha_t^{EV} \in \mathcal{A}_{EV} = \{0, 1.2, 1.3, \ldots, 3.3\} \text{ kW} \tag{18}$$

$$\alpha_t^{BESS} \in \mathcal{A}_{BESS} = \{-3.0, -2.9, \ldots, 3.0\} \text{ kW} \tag{19}$$

Here, $\alpha_t^{EV} = 0$ means no EV charging, and BESS actions are positive for charging, negative for discharging, and 0 for idle. The full joint action space has size:

$$|\mathcal{A}| = |\mathcal{A}_{EV}| \times |\mathcal{A}_{BESS}| = 23 \times 61 = 1403 \tag{20}$$

3) Set of Rewards: Different rewards have been defined in an energy system with or without BESS (only EV), as outlined in Table I and described below for the general case in which a BESS is available:

$$r_t = r_t^{C_{EV}} + r_t^{C_{BESS}} + r_t^{SC} + r_t^{SoC_{EV}} + r_t^{SoC_{BESS}} + r_t^{pref} + r_t^{Day_{EV}} + r_t^{Day_{BESS}} \tag{21}$$

Cost-related Rewards ($r_t^{C_{EV}}$ and $r_t^{C_{BESS}}$): Historical cost percentiles are used to identify daily periods of low or high electricity prices, following a similar methodology to that presented in [5]. These percentiles are derived by combining dynamic hourly electricity tariffs with historical demand.

Self-Consumption Reward (r_t^{SC} or $r_t^{SC_{EV}}$ only for EV): As presented in Table I, the self-consumption reward r_t^{SC} in the case a BESS is available comprises two components added together: $r_{A,t}^{SC}$, which rewards excess PV energy stored in the BESS, and $r_{B,t}^{SC}$, which strongly rewards cases where PV generation perfectly meets demand. These depend on P_t^{net}, the net load after subtracting PV generation P_t^{PV} from the inflexible load P_t^{inflex}, and P_t^{flex}, the combined EV and BESS actions $a_t^{EV} + a_t^{BESS}$. When only an EV is scheduled, the simplified reward $r_t^{SC_{EV}}$ is used.

State of Charge Constraints ($r_t^{SoC_{EV}}, r_t^{SoC_{BESS}}$): Ensure EV and BESS operate within their storage limit.

EV charging preferences Reward (r_t^{pref}): Based on Eq. (15), this reward reflects typical weekday/weekend charging behavior. The Agent is penalized for charging between 09:30–17:00 on weekdays, while weekends allow unrestricted EV charging.

EV daily load Constraint ($r_t^{Day_{EV}}$): As defined in Eq. (14), this reward encourages the Agent to match historical daily EV consumption. Rewards are proportional to the action when the daily load stays within the limit, normalized by P_{max}^{EV}, which is the maximum EV charging power as listed in Table II.

BESS Final SoC Reward ($r_t^{Day_{BESS}}$): Encourages the BESS to end the day with a SoC close to its initial value (±3% tolerance). To check feasibility, we compute:

- Max Charge/Discharge Energy:

$$E_{t\to T,\max}^{BESS} = (96 - t) \times \frac{P_{\max}^{BESS}}{4}$$

  where $P_{\max}^{BESS}$ is the maximum (dis)charging power of the BESS, as listed in Table II.
- Lower/Upper Bound Energy (E_{LB}^{BESS}, E_{UB}^{BESS}): Energy required to reach the final SoC within tolerance from the current SoC.

III. Case study

This work examines two case studies: I) An EV-PV system where BESS does not exist and II) an EV-PV-BESS. In the first one, the state space of the DQN-based strategy given by (16) is defined without SoC_t^{BESS} and u_t^{BESS}. The action space is defined solely from (18) and the reward function given by the equation (21) is defined without the terms $r_t^{C_{BESS}}$, $r_t^{SoC_{BESS}}$, $r_t^{Day_{BESS}}$ while the self-consumption sub-reward r_t^{SC} is substituted with $r_t^{SC_{EV}}$.

We use data from house 4373 of the Pecan Street dataset [9], applying the preprocessing methodology from [5] to ensure consistency in data cleaning. Table II presents the EV and BESS parameters. The maximum contracted power for buy P_{max}^{buy} and sell P_{max}^{sell} is 4.5 kW. The α parameter is set to 0.7. For the DQN approach, data is split 80%-20% for training and testing, with the test set comprising 64 days. Hourly Spanish electricity prices from 2018 (excluding taxes), corresponding to the same consumption dates, are applied for buying and selling. The granularity is 15 minutes, so N^{hour} is set to 4.

TABLE II
BESS and EV Parameters

BESS Parameter (unit)	Value	EV Parameter (unit)	Value
SoC_{max}^{BESS} (kWh)	6	$SoC_{\max}^{EV}$ (kWh)	24
SoC_{min}^{BESS} (kWh)	1	SoC_{min}^{EV} (kWh)	4
P_{max}^{BESS} (kW)	3	P_{max}^{EV} (kW)	3.3
$SoC_{t=0}^{BESS}$ (kWh)	3	$SoC_{t=0}^{EV}$ (kWh)	Eq. (13)
A^{ch}, A^{dis} (–)	0.95	η^{EV} (–)	0.92

TABLE III
Comparison of Weighted Average Cost Savings and SC Improvement

	EV-PV		EV-PV-BESS	
	Cost Savings	SC improvement	Cost Savings	SC improvement
DQN	19.40%	-0.45%	19.77%	10.05%
MILP	20.07%	3.11%	28.16%	21.59%

IV. Results and performance comparison

The results are summarized in Table III. The performance of both MILP and DQN methods is evaluated across 64 days (the DRL test set days) in terms of cost savings and self-consumption improvement. The results are computed as weighted averages, ensuring a fair comparison by normalizing with absolute cost and grid power usage before optimization to prevent distortions from minor variations.

For the EV-PV configuration, the proposed approach results in 19.40% cost savings with DQN and 20.07% with MILP. Regarding SC improvement, DQN shows a slight decrease of -0.45%, whereas MILP achieves an improvement of 3.11%, indicating a more favorable outcome for the MILP approach in both cost and SC performance metrics. The relatively limited SC improvement in the EV-PV configuration is related to the constraint that the EV cannot charge during weekdays from 9:30 to 15:00, when solar generation is typically high. This constraint restricts solar utilization during peak hours, limiting SC improvements for both DRL and MILP.

In the second use case, where the BESS is integrated into the HEMS, both optimization strategies showcase an enhanced performance. MILP continues to outperform DQN, with 28.16% cost savings and a 21.59% SC improvement, compared to 19.77% and 10.05%, respectively, for DQN. The addition of the BESS improves SC by storing excess solar energy during periods of high generation, when EV charging is restricted, and releasing it later. This enables more efficient use of solar energy and enhances overall system performance.

Figures 1 and 2 show that both strategies avoid EV charging during the restricted weekday period (9:30-15:00), as defined by Eq. (15). In the no-BESS scenario (Fig. (1)), charging occurs at night to exploit lower electricity prices. DQN schedules additional EV charging from 08:00 and 09:30 and 15:00 and 16:30, attempting to utilize solar generation during these periods. However, when solar generation is higher, the MILP strategy achieves lower costs and higher SC, scheduling the EV charging between 15:00 and 16:30.

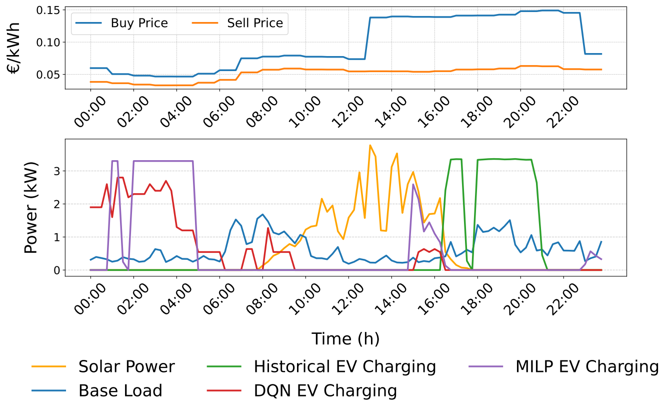

Fig. 1. Case study I) EV-PV: Electricity market prices (top) and comparison of DQN and MILP scheduling strategies (bottom) for day 11/01/2018.

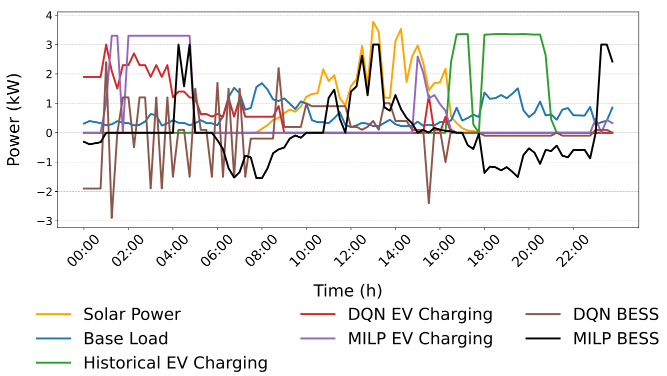

Fig. 2. Case study II) EV-PV-BESS: Comparison of DQN and MILP scheduling strategies for day 11/01/2018. Price profile identical to Fig. 1).

The EV-PV-BESS case (Fig. (2) shows significant EV charging between 00:00 and 05:00, with the DQN strategy exhibiting oscillations between charging and discharging for the BESS to cover both EV and inflexible base load, likely due to lower electricity prices at night. The MILP strategy charges the BESS primarily between 04:00 and 06:00, when the grid covers a larger portion of the load due to lower prices. Later, both strategies use solar power to charge the BESS, but DQN underutilizes it, showing a cautious charging profile. MILP stores more energy and discharges it in the final hours when electricity prices are higher, ensuring better base load coverage. This comparison shows that MILP offers a more balanced strategy, while DQN exhibits greater fluctuations and a more conservative charging/discharging approach.

V. Conclusions

This paper compares DQN and MILP strategies for HEMS under two configurations: EV-PV and EV-PV-BESS. Real-world data from the Pecan Street dataset are processed and used to evaluate both strategies over a 64-day test period. Results show that MILP consistently outperforms DQN regarding cost savings and self-consumption. In the EV-PV case, both methods suffer from limited charging flexibility during solar peak hours, resulting in poor self-consumption, though MILP still performs better. The addition of a BESS enhances both approaches by improving solar energy utilization and boosting self-consumption, but MILP retains a clear edge in economic and energy efficiency. DQN, however, offers advantages in adaptability and lower operational complexity, making it well-suited for real-time deployment in dynamic environments.

Future work could enhance MILP strategies using rolling horizon optimization and problem decomposition, for scalable, real-time deployment. For DRL, the research could focus on refining the reward function to better guide agent behavior and exploring multi-agent environments to optimize coordination among HEMS devices.

References

[1] J. Lyu, T. Ye, M. Xu, G. Ma, Y. Wang, and M. Li, "Price-sensitive home energy management method based on pareto optimisation," *International Journal of Sustainable Engineering*, vol. 14, no. 3, pp. 433–441, 2020.

[2] D. F. U. Putra, M. A. Fikri, A. R. Hidayatullah, and A. Soeprijanto, "Design of home power energy management system using mixed integer linear programming (milp) based on extreme learning machine (elm) dynamic pricing," *International Review of Automatic Control*, vol. 12, no. 4, pp. 174–181, 2019.

[3] S. Barja-Martinez, F. Rücker, M. Aragüés-Peñalba, R. Villafafila-Robles, Í. Munné-Collado, and P. Lloret-Gallego, "A novel hybrid home energy management system considering electricity cost and greenhouse gas emissions minimization," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2782–2790, 2021.

[4] S. Sykiotis, C. Menos-Aikateriniadis, A. Doulamis, N. Doulamis, and P. S. Georgilakis, "A self-sustained ev charging framework with n-step deep reinforcement learning," *Sustainable Energy, Grids and Networks*, vol. 35, p. 101 124, 2023, ISSN: 2352-4677.

[5] C. Menos-Aikateriniadis, S. Sykiotis, and P. S. Georgilakis, "Unlocking the potential of smart ev charging: A user-oriented control system based on deep reinforcement learning," *Electric Power Systems Research*, vol. 230, p. 110 255, 2024, ISSN: 0378-7796.

[6] H. Ding, Y. Xu, B. Chew Si Hao, Q. Li, and A. Lentzakis, "A safe reinforcement learning approach for multi-energy management of smart home," *Electric Power Systems Research*, vol. 210, p. 108 120, 2022, ISSN: 0378-7796.

[7] T. H. B. Huy, H. Truong Dinh, D. Ngoc Vo, and D. Kim, "Real-time energy scheduling for home energy management systems with an energy storage system and electric vehicle based on a supervised-learning-based strategy," *Energy Conversion and Management*, vol. 292, p. 117 340, 2023, ISSN: 0196-8904.

[8] S. H. Mansour, S. M. Azzam, H. M. Hasanien, M. Tostado-Véliz, A. Alkuhayli, and F. Jurado, "Deep reinforcement learning-based plug-in electric vehicle charging/discharging scheduling in a home energy management system," *Energy*, vol. 316, p. 134 420, 2025, ISSN: 0360-5442.

[9] *Dataport - Pecan Street Inc.* https://www.pecanstreet.org/dataport/, [Online; accessed 10-February-2025], 2009.