Energy Management for Airport Thermo-Electrical Microgrids: A Physics-Based RL Approach

Pablo Verdugo-Rivadeneira
Mehrdad Pirnia
Claudio A. Cañizares
Dept. of Electrical and Computer Engineering
University of Waterloo
Waterloo, Canada
{pverdugo, mpirnia, ccanizar}@uwaterloo.ca

Gonzalo E. Constante-Flores
Dept. of Electrical, Computer, and Energy Engineering
University of Colorado Boulder
Boulder, USA
gonzalo.constante@colorado.edu

***Abstract*—This paper proposes a model-free Reinforcement Learning framework for the optimal operation of multi-zone airport hangar thermo-electrical microgrids. The framework enforces constraint satisfaction through physics-based dynamic constraints and incorporates an implicit method to capture inter-zone heat exchanges. Simulation results based on the operation of an actual Canadian airport microgrid demonstrate the effectiveness of the proposed approach regarding thermal comfort and arbitrage performance. Furthermore, compared with conventional reward-driven methods, the proposed framework achieves faster convergence and more stable training behavior.**

***Index Terms*—Airport microgrid, Energy Management System, multi-zone building thermal model, reinforcement learning.**

I. Introduction

To address the climate impact of international aviation, the 184 member states of the International Civil Aviation Organization (ICAO) adopted in 2022 an ambitious target to reach net-zero carbon emissions by 2050 [1]. To this end, several alternatives for the energy supply of Electric Aircraft (EA) and airport buildings are being considered, including the use of alternative fuels such as hydrogen, the adoption of hybrid or fully EA, the electrification of ground transportation operations, the integration of Distributed Energy Resources (DERs), and the deployment of energy-efficient buildings with electrified thermal energy systems.

Research on airport Microgrids (MGs) has received notable attention. However, the integration of airport Thermo-Electrical (TE)-MG Energy Management Systems (EMSs) has only been addressed in a limited number of studies. Notably, [2] investigates the TE energy management in an airport building, using electricity, heating, and cooling demand profiles based on Monte Carlo simulations, and an optimization-based Model Predictive Control (MPC) approach is used in [3] to determine the optimal operation of a multi-zone airport hangar MG. Besides these works, most existing studies focus on other aspects of airport MG management, such as the provision of frequency regulation [4]–[6], airport MG planning [7], [8], and hybrid EA charging optimization schedules [9]. Furthermore, the impact of Electric Vehicle (EV) and EA charging and the integration of Photo-Voltaic (PV) and Battery Energy Storage System (BESS) into airport MGs have been assessed in [10], while [11]–[13] discuss the role of hydrogen in airport operations. Thus, a research gap remains in the development of EMS solutions that explicitly consider the coupling between thermal and electrical dynamics in airport buildings.

Reinforcement Learning (RL) has received considerable attention in recent years as an alternative to traditional optimization approaches for addressing EMS problems [14], particularly to circumvent the challenges of obtaining accurate system models. Unlike model-based methods, RL does not require an explicit representation of the system; instead, an agent learns through direct interaction with the environment and generates its own training data from experience, making RL especially valuable in situations where data are scarce, costly, or constantly evolving. Hence, significant progress in developing RL-based approaches for EMSs has been made, aiming to facilitate learning and ensure constraint satisfaction. However, when thermal comfort is considered, with the exception of [15], [16], and [17], most studies do not incorporate a building thermal model to capture the underlying thermal dynamics. Furthermore, only the latter includes a multi-zone configuration, although not within a MG framework, as it focuses solely on room temperature control. This underscores the limited application of RL to the energy management of airport TE-MGs and highlights a relevant direction for future research. Thus, aiming to address the identified gaps, the contributions of this paper are as follows:

- Propose a novel RL-based EMS framework for the optimal operation of an airport TE-MG incorporating real-world data for load demand, solar radiation profiles, ambient temperature, and e-plane charging, while employing a hybrid action space that combines discrete and continuous control variables to realistically capture the operational characteristics of the TE-MG.
- Integrate physics-based dynamic constraints into the RL environment to ensure safe and physically consistent control of both the BESS and indoor temperature, preventing infeasible charge/discharge BESS trajectories and unrealistic Heat Pump (HP) operations, based on a multi-zone building thermal model to properly represent heat dynamics and inter-zone heat exchange in an actual MG

979-8-3315-8138-1

not yet deployed, using an implicit method to efficiently handle variable coupling.

The rest of the paper is organized as follows: Section II provides a background overview on the RL framework formulated as a Markov Decision Process (MDP). Section III describes the proposed RL framework for the energy management of TE-MGs. Section IV comprises the application of the proposed RL framework to an actual hangar building being deployed for the Waterloo Wellington Flight Centre (WWFC), and a comparison of the agent's training performance against a reward-based RL method. Finally, the conclusions of the presented studies are provided in Section V.

II. Reinforcement Learning Framework Overview

The energy management problem can be formulated as an MDP. Thus the Markov property asserts that the probability of transitioning to the next state depends solely on the current state and action, and not on any prior states [18]. Formally, an MDP is defined as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is the state space, defining the environment's status at each time step k; $\mathcal{A}$ is the action space, denoting the available control actions; $\mathcal{P}(s_{k+1} \,|\, s_k, a_k)$ is the state transition probability function, which describes the system's evolution from state s_k to s_{k+1} under action a_k; $\mathcal{R}(s_k, a_k, s_{k+1})$ is the reward function, which quantifies the immediate cost or benefit resulting from taking action a_k in state s_k and transitioning to the next state s_{k+1}; and $\gamma \in [0,1]$ is the discount factor, which balances future versus immediate rewards. As per [18], at each time step, the agent follows a policy π that defines a mapping from states to probabilities of selecting each possible action.

RL algorithms can be broadly categorized into three classes: value-based, which estimate the expected return of each action and derive a policy implicitly; policy-based, which directly optimize the policy without relying on value estimates; and actor-critic methods, which combine the benefits of both approaches. For the latter, the actor updates the policy, while the critic evaluates the value function to guide the learning process. The current paper adopts the Proximal Policy Optimization (PPO) algorithm, which is known for its robustness and stable performance [19], compared to other actor–critic methods that often exhibit instability in hybrid action settings.

III. Proposed RL-Based EMS Framework

This section describes the RL framework proposed for the EMS of a multi-zone airport hangar TE-MG. The electric demand is assumed to be supplied with power from the main grid, PV generation, and BESS operation, while the thermal power is assumed to be provided by HPs, and an Electric Heater (EH) for the planes' hangar.

A. Action Space

The RL framework models the interaction of a single agent with the environment to control the operation of a BESS, multiple HPs, and an EH at each time step k. Thus, the hybrid action space a_k is defined as follows:

$$a_k = \left[a_k^{\text{bat}},\ a_{h,k}^{\text{HP}},\ a_k^{\text{EH}}\right] \tag{1}$$

where a_k^{bat} and $a_{h,k}^{\text{HP}}$ are continuous actions that can take positive or negative values representing charging and discharging for the former, and heating and cooling for the latter, respectively; and a_k^{EH} is the discrete control signal for the EH, taking values $\{0, 1\}$ to indicate off and on states, respectively.

B. State Space

The state space s_k of the TE-MG environment accounts for all the system variables the agent observes in order to achieve an optimal operation. In this context, the state of the system components at each time step is defined as follows:

$$s_k = \left[k, P_k^{\text{PV}}, P_k^{\text{d}}, P_k^{\text{ep}}, TOU_k, SOC_k, \theta_k^{\text{ext}}, \theta_{i,k}^{\text{in}}, \underline{\theta}_{i,k}^{\text{in}}\right] \tag{2}$$

which encompasses input data, namely, PV generation P_k^{PV}, building P_k^{d} and e-plane P_k^{ep} power demand, Time of Use (TOU) tariff, ambient temperature θ_k^{ext}, as well as a time-varying temperature limit $\underline{\theta}_{i,k}^{\text{in}}$ based on room comfort considerations. It also incorporates system variables, namely BESS State of Charge (SOC) and room i temperatures $\theta_{i,k}^{\text{in}}$.

C. Reinforcement Learning Environment

The RL environment comprises the modeling of the EMS elements, ensuring that the system variables are properly updated whenever an agent takes actions based on the observations at each time step, since measurements were not available for the TE-MG being considered. In practical applications, system observations can be obtained directly, without relying on equation-based models, thus reducing modeling complexity and errors. However, as highlighted throughout this section, certain equations remain necessary to enforce operational constraints, such as the power and capacity limits governing BESS and HP operation. Thus, the proposed model assumes that the electric demand is fully satisfied by the available generation as follows:

$$P_k^{\text{gp}} = P_k^{\text{d}} + P_k^{\text{ep}} + P_k^{\text{bat}} - P_k^{\text{PV}} + \sum_{h \in H} P_{h,k}^{\text{HP}} + a_k^{\text{EH}} P^{\text{EH}} \tag{3}$$

where P_k^{gp} corresponds to the power exchanged with the grid; P_k^{bat} is the BESS active power; $P_{h,k}^{\text{HP}}$ denotes the HP active power; and P^{EH} represents the rated power of the EH. In a real-world application, $P_k^{\text{d}}, P_k^{\text{ep}}$, and P_k^{PV} would be measured directly from load consumption and PV generation; $P_k^{\text{bat}}, P_{h,k}^{\text{HP}}$, and a_k^{EH} would be determined by the RL agent; and the resulting P_k^{gp} would be measured at the Point of Common Coupling (PCC).

The BESS operation at each time step k within the time horizon T is modeled as follows:

$$P_k^{\text{ch}} = \max\left(0, a_k^{\text{bat}}\right), \quad P_k^{\text{dch}} = \max\left(0, -a_k^{\text{bat}}\right) \tag{4}$$

$$SOC_{k+1} - SOC_k = \left(P_k^{\text{ch}} \eta^{\text{ch}} - \frac{P_k^{\text{dch}}}{\eta^{\text{dch}}}\right) \Delta t_k \tag{5}$$

$$\underline{SOC} \le SOC_k \le \overline{SOC} \tag{6}$$

$$\Phi_k(\delta_k) = 5.24 \times 10^{-4} \delta_k^{2.03} \tag{7}$$

979-8-3315-8138-1

where (4) decomposes the BESS action a_k^{bat} into two non-negative components, separately modeling the charging P_k^{ch} and discharging power P_k^{dch}; (5) represents the BESS energy balance; and the SOC limits are modeled in (6). Based on [20], $\Phi_k(\delta_k)$ in (7) represents BESS degradation, as a function of the Depth of Discharge (DOD) $\delta_k = 1 - SOC_k$.

Finally, the MG thermal dynamics, which ensure both comfort compliance and reliable operation, are modeled as follows, assuming known building thermal parameters:

$$C_i^{\text{in}} \frac{\theta_{i,k}^{\text{in}} - \theta_{i,k-1}^{\text{in}}}{3600\Delta t_k} = \mu_i Q_{h,k}^{\text{HP}} + Q_{i,k}^{\text{sun}} - Q_{i,k}^{\text{env}} - Q_{i,k}^{\text{vent}} - \sum_{j\in\mathcal{I}} Q_{ij,k}^{\text{trf}} + Q_{i,k}^{\text{ig}} - Q_{i,k}^{\text{gr}} \quad \forall i \in \mathcal{I}\backslash\{ph\},\ \forall k \in T \tag{8}$$

$$C_i^{\text{in}} \frac{\theta_{i,k}^{\text{in}} - \theta_{i,k-1}^{\text{in}}}{3600\Delta t_k} = a_k^{\text{EH}} P^{\text{EH}} + Q_{i,k}^{\text{sun}} - Q_{i,k}^{\text{env}} - Q_{i,k}^{\text{vent}} - \sum_{j\in\mathcal{I}} Q_{ij,k}^{\text{trf}} + Q_{i,k}^{\text{ig}} - Q_{i,k}^{\text{gr}} \quad \forall i \in \{ph\},\ \forall k \in T \tag{9}$$

$$Q_{ij,k}^{\text{trf}} = \sum_{\text{w}\in W} U_{ij}^{\text{w}} A_{ij}^{\text{w}} \left(\theta_{i,k}^{\text{in}} - \theta_{j,k}^{\text{in}}\right) \forall i,j \in \mathcal{I},\ \forall k \in T \tag{10}$$

$$Q_{i,k}^{\text{gr}} = U_i^{\text{f}} A_i^{\text{f}} \left(\theta_{i,k}^{\text{in}} - \theta^{\text{gr}}\right) \quad \forall i \in \mathcal{I}\backslash\{ng\},\ \forall k \in T \tag{11}$$

$$Q_{h,k}^{\text{HP}} = \psi_h^{\text{HP}} \frac{\theta_h^{\text{HP,out}}}{\theta_h^{\text{HP,out}} - \theta_k^{\text{ext}}} P_{h,k}^{\text{HP}} \quad \forall h \in H,\ \forall k \in T \tag{12}$$

$$Q_{i,k}^{\text{sun}} = \sum_{e\in\mathcal{B}} \sigma_{e,i} I_k A_{e,i} \xi_{e,i} \tau_{e,i} \quad \forall i \in \mathcal{I},\ \forall k \in T \tag{13}$$

$$Q_{i,k}^{\text{env}} = \sum_{e\in\mathcal{B}} U_{e,i} A_{e,i} \left(\theta_{i,k}^{\text{in}} - \theta_k^{\text{ext}}\right) \quad \forall i \in \mathcal{I},\ \forall k \in T \tag{14}$$

$$Q_{i,k}^{\text{vent}} = \frac{v_{i,k}^{\text{rate}} \rho^{\text{air}} Vol_i\, c^{\text{air}} \left(\theta_{i,k}^{\text{in}} - \theta_k^{\text{ext}}\right)}{3600} \quad \forall i \in \mathcal{I},\ \forall k \in T \tag{15}$$

$$Q_{i,k}^{\text{ig}} = \text{Np}_{i,k} Q_i^{\text{oc}} \text{CLF}_k + Q_{i,k}^{\text{lg}} + Q_{i,k}^{\text{eq}} \quad \forall i \in \mathcal{I},\ \forall k \in T \tag{16}$$

Here, C_i^{in} denotes the rooms' thermal capacitance, and (8) and (9) model the thermal dynamics of the building thermal zones $\mathcal{I}$ and the planes' hangar $\{ph\}$, with the respective thermal contribution from HPs and the EH. The heat transfer between rooms i and j $Q_{ij,k}^{\text{trf}}$ and between each room and the ground $Q_{i,k}^{\text{gr}}$ is defined in (10) and (11), excluding non-ground-level rooms $\{ng\}$ and considering wall w and floor f areas A with transmittance U. The HPs's thermal power $Q_{h,k}^{\text{HP}}$ is given by (12). Equations (13) to (15) describe the heat transfer due to solar irradiation $Q_{i,k}^{\text{sun}}$ with shadowing σ and transmission coefficient τ, and window surface ξ; the building envelope e $Q_{i,k}^{\text{env}}$; and ventilation $Q_{i,k}^{\text{vent}}$ with rate $v_{i,k}^{\text{rate}}$. Finally, (16) models internal heat gains $Q_{i,k}^{\text{ig}}$ from room occupants Q_i^{oc}, lighting $Q_{i,k}^{\text{lg}}$, and equipment $Q_{i,k}^{\text{eq}}$, assuming a cooling factor CLF_k and number of occupants $\text{Np}_{i,k}$. In an actual deployment, equations (5), (10), (11), (14), and (15) can be replaced by direct sensor readings (e.g., wall heat flux sensors, ventilation flow meters), whereas (12), (13), and (16) may be estimated using partial measurements combined with model-based representations.

D. Physics-Based Dynamic Constraints

The policy network outputs unconstrained actions, which are then clipped within the RL environment to enforce physical limits and yield deployable control policies, as follows:

$$\underline{Q}_{h,k}^{\text{HP}} = C_i^{\text{in}} \frac{\underline{\theta}_{i,k}^{\text{in}} - \theta_{i,k-1}^{\text{in}}}{3600\Delta t_k} - Q_{i,k}^{\text{sun}} + Q_{i,k}^{\text{env}} + Q_{i,k}^{\text{vent}} + \sum_{j\in\mathcal{I}} Q_{ij,k}^{\text{trf}} - Q_{i,k}^{\text{ig}} + Q_{i,k}^{\text{gr}} \quad \forall i \in \mathcal{I}\backslash\{nc\} \tag{17}$$

$$\overline{Q}_{h,k}^{\text{HP}} = C_i^{\text{in}} \frac{\overline{\theta}_{i,k}^{\text{in}} - \theta_{i,k-1}^{\text{in}}}{3600\Delta t_k} - Q_{i,k}^{\text{sun}} + Q_{i,k}^{\text{env}} + Q_{i,k}^{\text{vent}} + \sum_{j\in\mathcal{I}} Q_{ij,k}^{\text{trf}} - Q_{i,k}^{\text{ig}} + Q_{i,k}^{\text{gr}} \quad \forall i \in \mathcal{I}\backslash\{nc\} \tag{18}$$

$$P_{h,k}^{\text{HP}} = \text{clip}\left(a_{h,k}^{\text{HP}},\ \max\left(\underline{P}_h^{\text{HP}},\ \frac{\underline{Q}_{h,k}^{\text{HP}}}{COP_k}\right),\ \min\left(\overline{P}_h^{\text{HP}},\ \frac{\overline{Q}_{h,k}^{\text{HP}}}{COP_k}\right)\right) \tag{19}$$

where the clipping operator is defined as $\text{clip}(a, \underline{a}, \overline{a}) = \min(\max(a, \underline{a}), \overline{a})$, which projects a onto the interval $[\underline{a}, \overline{a}]$, and $\underline{Q}_{h,k}^{\text{HP}}$ and $\overline{Q}_{h,k}^{\text{HP}}$ in (17) and (18) denote the minimum and maximum HP thermal power limits required to ensure thermal comfort, excluding rooms with no temperature control $\{nc\}$. Notably, the HP limits are divided by the Coefficient of Performance (COP) to define the physics-based dynamic upper and lower bounds for HP operation in (19), with $\overline{P}_h^{\text{HP}}$ and $\underline{P}_h^{\text{HP}}$ defining the HP's rated power limits. Similarly, the dynamic limits for BESS operation can be determined as follows:

$$\overline{P}_k^{\text{bat}} = \max\left(0,\ \min\left(\overline{P}^{\text{bat}},\ \frac{\overline{SOC} - SOC_k}{\eta^{\text{ch}}\,\Delta t_k}\right)\right) \tag{20}$$

$$\underline{P}_k^{\text{bat}} = \min\left(0,\ \max\left(\underline{P}^{\text{bat}},\ \frac{(\underline{SOC} - SOC_k)\,\eta^{\text{dch}}}{\Delta t_k}\right)\right) \tag{21}$$

$$P_k^{\text{bat}} = \text{clip}\left(a_k^{\text{bat}},\ \underline{P}_k^{\text{bat}},\ \overline{P}_k^{\text{bat}}\right) \tag{22}$$

where $\underline{P}^{\text{bat}}$ and $\overline{P}^{\text{bat}}$ represent the BESS rated charge and discharge power limits, respectively. The second terms in (20) and (21) impose physics-based dynamic constraints to prevent overcharging or over-discharging accounting for the measured SOC_k and considered efficiency losses. Finally, (22) ensures that the selected action a_k^{bat} is clipped to the feasible range.

E. Multi-Zone Thermal Modeling

The building thermal model is used to compute zone temperatures, which in a real deployment would be directly measured and thus not modeled in the RL environment. Accordingly, the temperatures $\boldsymbol{\theta}_k \in \mathbb{R}^m$ of a building with m thermal zones are obtained by solving the following linear system:

$$\mathbf{M} \cdot \boldsymbol{\theta}_k = \mathbf{b}_k \tag{23}$$

where $\mathbf{M} \in \mathbb{R}^{m\times m}$ models the conductive couplings between zones and their surroundings, and $\mathbf{b}_k$ is the right-hand side vector that accounts for previous temperatures and internal and external thermal contributions. Thus:

$$M_{ii} = 1 + \frac{3600\Delta t_k}{C_i^{\text{in}}} \sum_{j\in\mathcal{I}} U_{ij}^{\text{w}} A_{ij}^{\text{w}}; \; M_{ij} = -\frac{3600\Delta t_k}{C_i^{\text{in}}} U_{ij}^{\text{w}} A_{ij}^{\text{w}} \tag{24}$$

$$b_{i,k} = \theta_{i,k-1}^{\text{in}} + \frac{3600\Delta t_k}{C_i^{\text{in}}} \Big(\mu_i Q_{h,k}^{\text{HP}} + Q_{i,k}^{\text{EH}} + Q_{i,k}^{\text{sun}} - Q_{i,k}^{\text{env}} - Q_{i,k}^{\text{vent}} + Q_{i,k}^{\text{ig}} - Q_{i,k}^{\text{gr}} \Big) \tag{25}$$

where M_{ij} defines the entries for $i \neq j$. The EH heat input $Q_{i,k}^{\text{EH}}$ is nonzero only for the hangar, while other zones may receive HP thermal power via μ_i.

F. Reward Function

The proposed reward scheme is as follows:

$$r_k = -c_k^{\text{gp}} P_k^{\text{gp}} \Delta t_k - \lambda_1 E_n RC \Phi_k - \lambda_2 c_k^{\text{temp}} - c_k^{\text{SOC}} \tag{26}$$

$$c_k^{\text{temp}} = \max\left(0, \underline{\theta}_{\{ph\}}^{\text{in}} - \theta_{\{ph\},k}^{\text{in}}\right)^2 \tag{27}$$

$$c_k^{\text{SOC}} = \mathbf{1}_{(k=T-1)} \left| SOC_k - SOC^{\text{tgt}} \right| \tag{28}$$

In (26), r_k denotes the reward, λ_1 and λ_2 are penalty coefficients associated with BESS degradation and temperature violations, respectively, and are determined through trial-and-error, c_k^{gp} is the electricity price, and E_n and RC represent the BESS rated capacity and replacement cost, respectively. In (27), c_k^{temp} accounts for temperature violations and is applied only to the planes' hangar, as the temperature in other rooms remains within limits due to the physics-based imposed constraints. Finally, c_k^{SOC} in (28) allows the BESS to reach a target SOC at the end of the operation.

IV. Results and Discussion

The proposed RL framework is applied to the MG, illustrated in Fig. 1, which is being deployed at the WWFC airport and comprises a PV system, a BESS, four HPs and an EH for indoor temperature control, and an e-plane charger. The hangar building has been distributed into ten distinct zones, as per [3]. The temperature inside each of the flight simulator rooms (Rooms 3, 4, and 5) is controlled independently by a HP, and an additional HP controls the temperature inside Room 8, supplying thermal power to the other rooms according to the thermal distribution factors μ_i defined in [3]. As for the planes' hangar, the indoor temperature is controlled through an EH with no direct thermal contribution from the HPs.

The performance of the proposed RL framework is evaluated under two scenarios, representing the hottest summer and the coldest winter days in Ontario, Canada. These scenarios are based on real data, using average hourly weather conditions in the region from 2007 to 2021, as depicted in Fig. 2 [3]. Furthermore, the predefined temperature setpoints are assumed to be a minimum of 21°C during working hours, i.e., from 8 am to 9 pm, and 15°C for the remaining hours, except for the planes' hangar, which maintains a constant setpoint of 10°C. The maximum temperature is capped at 24°C for all rooms, except for the hangar, which has no cooling system.

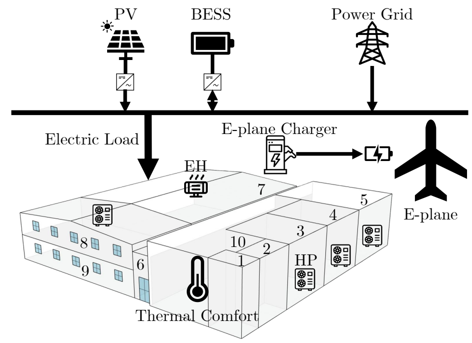

Fig. 1. Microgrid at WWFC Hangar 7.

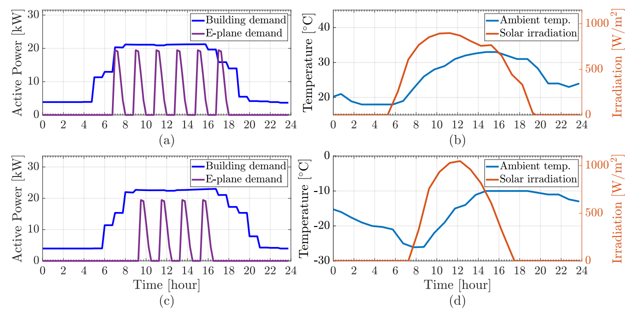

Fig. 2. Electric demand and environmental conditions for WWFC Hangar 7 for the hottest summer day (a) and (b) and the coldest winter day (c) and (d).

The policy network is a fully connected Deep Neural Network (NN) with four hidden layers of 256, 256, 256, and 128 neurons. The discount factor was set to $\gamma = 0.996$, with an entropy coefficient of 0.01 to encourage exploration and 15 training epochs. The penalty factors λ_1 and λ_2 are 0.1 and 0.25, respectively. It is worth noting that increasing λ_2 leads to more aggressive EH operation. As shown in Fig. 3, the proposed framework exhibits faster convergence and improved training stability compared to a reward-based RL approach, while achieving millisecond-scale policy evaluation.

The MG power balance for winter and summer days is illustrated in Fig. 4. Note that the demand is fully satisfied at all times, and that during summer no power export to the grid takes place, whereas in winter, power is exported between 10:30–11:15 am and 12:30–1:15 pm due to high solar availability. Fig. 5 illustrates the BESS SOC throughout the day, with the BESS following arbitrage behavior with conservative, incremental dispatch decisions that reflect the agent's learned policy.

Room temperatures for the coldest winter and hottest summer days are presented in Fig. 6. In rooms with temperature control, temperature violations are minimal, indicating the agents' effectiveness in managing the HPs and the EH. Furthermore, except for Room 9, where temperatures approach the set limits in both scenarios, the expected occupancy in rooms without temperature control is relatively low, limiting the impact on comfort levels.

V. Conclusion

In this paper, the implementation of an RL framework for the optimal operation of a multi-zone TE airport hangar MG was discussed, accounting for various thermal needs and

979-8-3315-8138-1

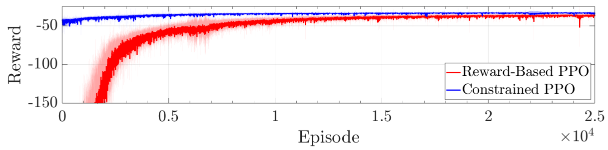

Fig. 3. Cumulative reward over training episodes for the winter scenario.

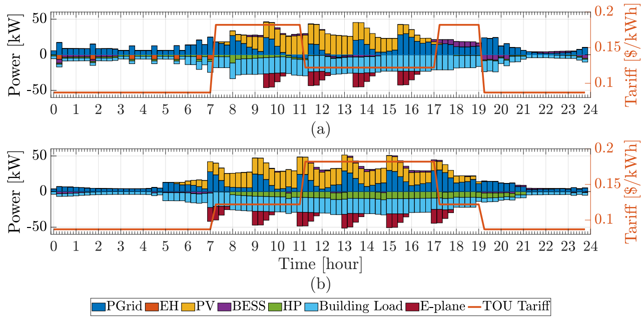

Fig. 4. Active power balance for the WWFC hangar MG for the (a) coldest winter and (b) hottest summer days.

resources, and e-plane charging based on actual data. Physics-based dynamic constraints enforced physically feasible operation of HPs and the BESS, while a multi-dimensional hybrid action space was adopted to enable the operation of an EH within a single-agent control strategy, thus reducing overall control complexity. An implicit numerical method was employed to capture heat dynamics and inter-zone thermal coupling in a computationally efficient manner. The developed EMS model was applied to an actual MG being deployed at the WWFC airport in Ontario, Canada, with simulation results demonstrating improved thermal comfort regulation and effective arbitrage performance. Compared with conventional reward-driven approaches, the proposed method exhibited faster convergence and more stable training behavior. Work currently underway includes comparisons with MPC-based formulations and potential real-time implementation challenges.

References

[1] I. Hannula, J. B. Menendez, T. Bosoni, A. Bressers, F. Briens, E. Connelly, J. Couse, L. Cret, M. Fajardy, C. Healya, J. Moorhouse, and Y. Om, "The Role of E-fuels in Decarbonising Transport," International Energy Agency, Tech. Rep., 2024.

[2] S. Jin and Y. Li, "Analyzing the performance of electricity, heating, and cooling supply nexus in a hybrid energy system of airport under uncertainty," *Energy*, vol. 272, p. 127138, 2023.

[3] P. Verdugo, C. Cañizares, and M. Pirnia, "Modeling and Energy Management of Hangar Thermo-Electrical Microgrid for Electric Plane Charging Considering Multiple Zones and Resources," *Applied Energy*, vol. 379, p. 124951, 2025.

[4] X. Liu, P. Fan, J. Yang, and U. Tahmeed, "A Flexible Frequency Control Strategy for Airport-Microgrid Through Vehicle-to-Grid and Aviation-to-Grid," *Journal of Energy Storage*, vol. 125, p. 116980, 2025.

[5] C. Haydaroğlu, "Chaos-Based Optimization for Load Frequency Control in Islanded Airport Microgrids with Hydrogen Energy and Electric Aircraft," *International Journal of Hydrogen Energy*, vol. 143, pp. 1198–1214, 2025.

[6] Z. Guo, J. Zhang, R. Zhang, and X. Zhang, "Aviation-to-Grid Flexibility Through Electric Aircraft Charging," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 8149–8159, 2022.

[7] Z. Guo, B. Li, G. Taylor, and X. Zhang, "Infrastructure Planning for Airport Microgrid Integrated with Electric Aircraft and Parking Lot Electric Vehicles," *eTransportation*, vol. 17, p. 100257, 2023.

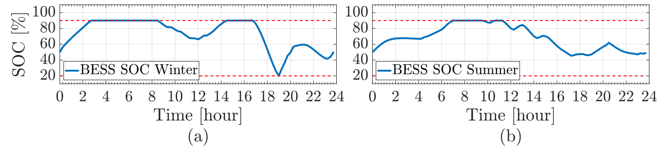

Fig. 5. BESS SOC for (a) coldest winter and (b) hottest summer days.

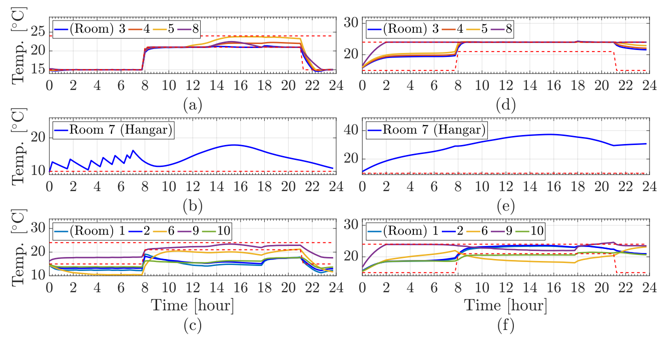

Fig. 6. Indoor temperature in (a) rooms with temperature control, (b) the planes' hangar, and (c) rooms without temperature control, in winter; and (d), (e), and (f), respectively, in summer.

[8] Z. Liu, Q. Wang, D. Sigler, A. Kotz, K. J. Kelly, M. Lunacek, C. Phillips, and V. Garikapati, "Data-Driven Simulation-Based Planning for Electric Airport Shuttle Systems: A Real-World Case Study," *Applied Energy*, vol. 332, p. 120483, 2023.

[9] B. Hou, S. Bose, L. Marla, and K. Haran, "Impact of Aviation Electrification on Airports: Flight Scheduling and Charging," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 2342–2354, 2024.

[10] P. Ollas, S. G. Sigarchian, H. Alfredsson, J. Leijon, J. S. Döhler, C. Aalhuizen, T. Thiringer, and K. Thomas, "Evaluating the Role of Solar Photovoltaic and Battery Storage in Supporting Electric Aviation and Vehicle Infrastructure at Visby Airport," *Applied Energy*, vol. 352, p. 121946, 2023.

[11] Y. Xiang, H. Cai, J. Liu, and X. Zhang, "Techno-Economic Design of Energy Systems for Airport Electrification: A Hydrogen-Solar-Storage Integrated Microgrid Solution," *Applied Energy*, vol. 283, p. 116374, 2021.

[12] H. Zhao, Y. Xiang, Y. Shen, Y. Guo, P. Xue, W. Sun, H. Cai, C. Gu, and J. Liu, "Resilience Assessment of Hydrogen-Integrated Energy System for Airport Electrification," *IEEE Transactions on Industry Applications*, vol. 58, no. 2, pp. 2812–2824, 2022.

[13] S. Rezaei, K. Alsamri, E. Simeoni, J. Huynh, and J. Brouwer, "Techno-Economic and Environmental Analysis of Clean Hydrogen Deployment: A Case Study of Los Angeles International Airport," *Energy Conversion and Management*, vol. 340, p. 119946, 2025.

[14] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[15] Y. Ye, H. Wang, P. Chen, Y. Tang, and G. Strbac, "Safe Deep Reinforcement Learning for Microgrid Energy Management in Distribution Networks With Leveraged Spatial–Temporal Perception," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3759–3775, 2023.

[16] X. Wang, P. Wang, R. Huang, X. Zhu, J. Arroyo, and N. Li, "Safe Deep Reinforcement Learning for Building Energy Management," *Applied Energy*, vol. 377, p. 124328, 2025.

[17] X. Ding, A. Cerpa, and W. Du, "Multi-Zone HVAC Control With Model-Based Deep Reinforcement Learning," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 4408–4426, 2025.

[18] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. The MIT Press, 2018.

[19] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[20] B. Xu, J. Zhao, T. Zheng, E. Litvinov, and D. S. Kirschen, "Factoring the Cycle Aging Cost of Batteries Participating in Electricity Markets," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2248–2259, 2018.

979-8-3315-8138-1