2026 IEEE Transportation Electrification Conference & Expo (ITEC) & Electric Aircraft Technologies Symposium (EATS) (ITEC+EATS) | 979-8-3315-8773-4/26/$31.00 ©2026 IEEE | DOI: 10.1109/ITECEATS66641.2026.11592974

Reinforcement Learning-Based Real-Time Balancing of Reconfigurable Battery Packs

Ali Irshayyid
ECE, Oakland University
Rochester, MI, USA
lukenuculaj@oakland.edu

Wanqun Yang
ECE, Oakland University
Rochester, MI, USA
junchen@oakland.edu

Tingjun Lei
EECS, University of North Dakota
Grand Forks, ND, USA
tingjun.lei@und.edu

Jun Chen*
ECE, Oakland University
Rochester, MI, USA
junchen@oakland.edu

***Abstract*—This paper presents a control framework based on reinforcement learning (RL) for achieving real-time state-of-charge (SOC) balancing within reconfigurable lithium-ion battery systems. Reconfigurable battery architectures provide enhanced flexibility via dynamically adjustable circuit topologies. Realizing their full capabilities requires an intelligent controller that can determine optimal topologies in real-time. This work addresses the problem by modeling the reconfiguration task as a Markov Decision Process and developing a policy optimization approach that directs switching actions according to cell states. The proposed RL agent operates on discrete action space and is guided by a multi-objective reward function. In particular, the controller is designed to balance the tradeoff among SOC equalization performance, switching effort, and runtime. To avoid the high cost associated with physical battery training, a surrogate model is proposed for efficient simulation-based learning. The surrogate model provides an accurate approximation of the battery pack dynamics while significantly reducing the computational burden during training. The proposed RL controller is evaluated against rule-based and greedy-based baselines, achieving superior SOC balancing performance with improved runtime. Specifically, the proposed controller achieves a final SOC spread of 0.0163 V, corresponding to more than 17% improvement over the greedy-based controller and 58% improvement over the rule-based controller.**

***Index Terms*—Reconfigurable battery packs, reinforcement learning, state-of-charge balancing, surrogate modeling.**

I. Introduction

ELECTRIC vehicles and many other devices rely on lithium-ion battery packs containing hundreds of cells [1], [2]. These multicell packs inevitably suffer from cell imbalance due to manufacturing variations and aging, with the weakest cell limiting pack performance [3]–[6]. Current balancing methods have fundamental limitations. Passive balancing wastes energy as heat [7], while active balancing requires complex, costly circuitry that continuously drains power [8]. While fixed topologies cannot dynamically manage individual cells, causing performance degradation [9], reconfigurable battery systems offer a solution by dynamically adjusting internal topology [10], [11]. However, controlling these systems is challenging due to nonlinear battery behavior and exponentially growing switch combinations [10]. Conventional approaches like dynamic programming become computationally prohibitive at scale [12], [13], while rule-based methods achieve only suboptimal performance [14], [15]. Deep reinforcement learning (DRL) presents a promising alternative [16]–[21]. However, training on hardware risks battery degradation and safety [22], while virtual training faces significant simulation time that is impractical for large battery systems.

Despite the promise of DRL for battery reconfiguration, existing approaches exhibit critical limitations. The study in [9] demonstrated a DRL-based controller that reduced voltage deviation by 60%, but it relied on a limited set of discrete topologies and a single-objective formulation that ignored switching cost and runtime efficiency. More recently, [23] explored AI-assisted reconfiguration for cell balancing, yet restricted their action space to only seven valid topologies, fundamentally limiting the controller's adaptability across diverse operating conditions. Furthermore, existing approaches lack rigorous benchmarking against provably optimal control strategies, making it difficult to quantify how closely learned policies approximate optimal performance. This paper directly addresses these gaps by proposing a multi-objective RL framework that explores all 256 valid configurations and benchmarks performance against both greedy controllers and dynamic programming solutions.

This paper proposes an RL-based control framework using an actor-critic architecture to learn switching strategies that balance SOC while minimizing switching frequency. We develop a linear surrogate model achieving 98.68% accuracy with 0.012% computational time compared to high-fidelity simulation. Our approach explores 256 valid topologies (versus 7 in [23]), reduces SOC imbalance by 91.6% over 105 minutes, and achieves 17.2% and 58.3% improvements over greedy and rule-based controllers respectively.

The remainder of the paper is organized as follows. Section II presents the battery packs and reconfiguration model. Section III introduces the surrogate model and the proposed RL control algorithm. Section IV provides simulation results. Section V concludes the paper with future work directions.

II. Dynamics of Battery Packs

A. Cell Dynamics

The cell model adopted in this work is based on the equivalent circuit model (ECM) [24], [25], which combines

This work is supported in part by National Science Foundation through Award #2237317. *Corresponding author: J. Chen.*

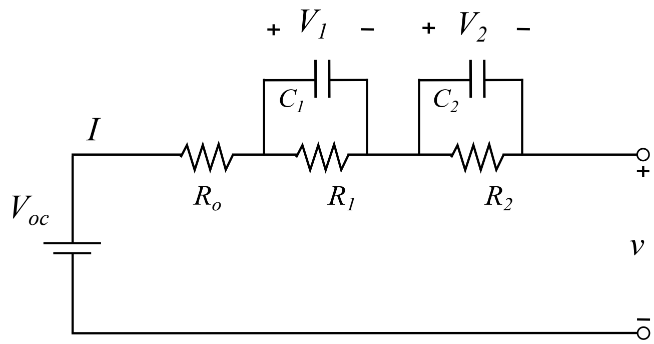

Fig. 1. The equivalent circuit model of a single battery cell.

an electrical circuit with a two-state thermal model to capture both the electrical and thermal dynamics of lithium-ion batteries. The electrical behavior is characterized by two RC pairs and a series resistance (see Fig. 1), with dynamics governed by:

$$\dot{V}_1 = -\frac{V_1}{R_1 C_1} + \frac{I}{C_1}, \tag{1a}$$

$$\dot{V}_2 = -\frac{V_2}{R_2 C_2} + \frac{I}{C_2}, \tag{1b}$$

$$v = V_{OC} - V_1 - V_2 - I R_o, \tag{1c}$$

where v is the terminal voltage, V_{OC} is the open-circuit voltage, and I is the current (positive for discharge, negative for charge). The RC pairs are characterized by voltages V_1 and V_2, resistance R_1 and R_2, and capacitance C_1 and C_2, with R_o representing the series resistance. The cell's SOC is govern by

$$\dot{s} = -\frac{\eta}{3600 C_n} I, \tag{2}$$

where s is the SOC, η is the coulombic efficiency and C_n is the nominal capacity of the cell in Amp-Hour. The thermal behavior is modeled using core and surface temperatures:

$$C_c \dot{T}_c = Q + \frac{T_s - T_c}{R_c}, \tag{3a}$$

$$C_s \dot{T}_s = \frac{T_c - T_s}{R_c} + \frac{T_f - T_s}{R_u}, \tag{3b}$$

where T_c is the core temperature, T_s is the surface temperature, T_f is the ambient temperature, C_c and C_s are the heat capacities of the core and surface respectively. R_c is the conduction resistance between T_c and T_s, while R_u is the convection resistance between T_f and T_s. The heat generation Q is given by

$$Q = I(V_{oc} - v) - I \frac{T_s + T_c}{2} \frac{dV_{oc}}{dT}. \tag{4}$$

All model parameters (V_{oc}, R_o, R_1, R_2, C_1, C_2) vary with SOC s and temperatures T_c and T_s, and can be expressed as $\sigma = f_\sigma(s, T_c, T_s)$, where $\sigma = \{V_{oc}, R_o, R_1, R_2, C_1, C_2\}$. In this paper, the experimentally validated parameters in [26], [27] are adopted.

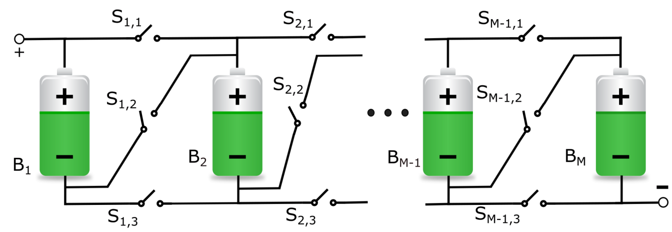

Fig. 2. Reconfigurable battery packs consisting of M cells.

B. Battery Reconfiguration

The individual cell models are integrated into a reconfigurable battery pack consisting of M cells interconnected through a three-switch-per-connection architecture (Fig. 2). The resultant battery pack architecture enables series/parallel reconfiguration between adjacent cells. Each cell B_i has three switches $(S_{i,1}, S_{i,2}, S_{i,3})$ that determine whether it connects to cell B_{i+1} in series ($sw_i = 1$) or parallel ($sw_i = 0$). Given pack current I_p and switch configuration $\mathbf{sw} = [sw_1, sw_2, \ldots, sw_{M-1}]$, the individual cell currents $\mathbf{I}$ are calculated by solving

$$\mathbf{AI} = \mathbf{C}, \tag{5}$$

where $\mathbf{A} \in \mathbb{R}^{M \times M}$ enforces the following: (1) $I_i = I_p$ for series cells, (2) current conservation $\sum_{j \in \mathcal{P}} I_j = I_p$ for any parallel group $\mathcal{P}$, (3) voltage consistency for parallel cells

$$R_{o,i} I_i - R_{o,j} I_j = \bar{V}_{\mathrm{OC},i} - \bar{V}_{\mathrm{OC},j}, \tag{6}$$

where $\bar{V}_{\mathrm{OC},i} = V_{\mathrm{OC},i} - V_{1,i} - V_{2,i}$ is the "modified" open-circuit voltage of cell i. The constraint vector $\mathbf{C}$ contains the corresponding values, and solving equation (5) yields the current distribution for updating battery states via equations (1)–(3).

III. RL-based Reconfiguration Control

A. Reconfigurable Battery Pack Surrogate Model

The high-fidelity model from Section II requires solving linear system (5) at each time step, making RL training computationally expensive. We address this by developing a linear surrogate model:

$$\mathbf{SOC}(k+1) = A \cdot \mathbf{SOC}(k) + B \cdot \mathbf{I}(k) + C \cdot \mathbf{sw}(k), \tag{7}$$

where $\mathbf{SOC}(k) \in \mathbb{R}^M$ represents cell SOCs, $\mathbf{I}$ is the current distribution obtained by solving (5), and $\mathbf{sw}(k)$ is the switch configuration. Using 5-minute time steps significantly reduces computational cost by requiring currents $\mathbf{I}$ calculation only at these intervals. Matrices A, B and C are obtained via least squares fitting to high-fidelity simulation data generated under random initial conditions. We assume constant current discharge for the pack within each time step, a valid assumption in cell balancing literature [13], [23]. To evaluate the proposed surrogate model, the high-fidelity model is used to generate training data, each with different initial SOC, different initial temperature, and different switch configuration. Using an 80/20 train-test split, the surrogate model achieves RMSE of 0.009051 on training dataset and 0.009187 on testing dataset.

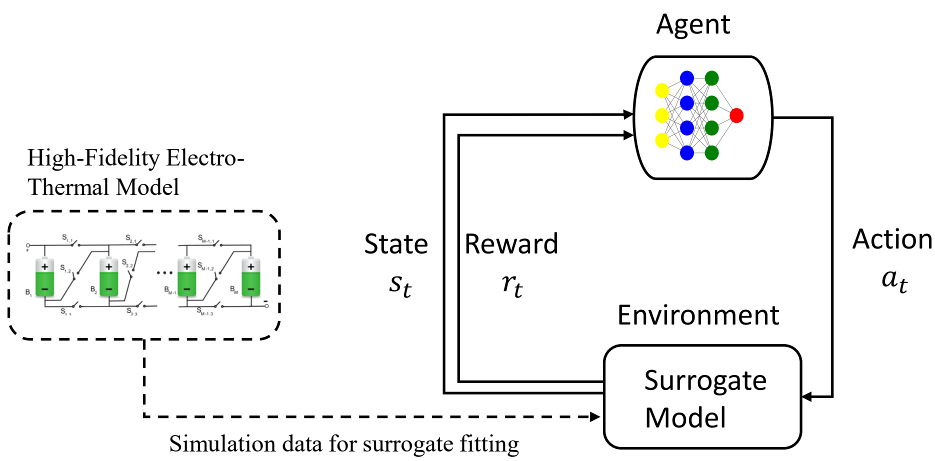

Fig. 3. RL training loop for reconfigurable battery pack balancing, where the agent interacts only with the surrogate model during training. The surrogate model environment provides an efficient approximation of pack dynamics for policy learning.

B. Proposed RL Framework

The reconfigurable battery pack control problem is formulated as a Markov Decision Process (MDP) and Proximal Policy Optimization (PPO) [28] is used as the foundation of our control framework (see Fig. 3). The MDP is defined by $(\mathcal{A}, \mathbf{S}, P, r, \gamma)$, where action space $\mathcal{A}$ represents switch configurations, $\mathbf{S}$ is the state space, r denotes immediate reward, and γ is the discount factor. The actor-critic architecture uses separate networks (256-128-64 hidden layers, Tanh activations) where the actor $\pi_\theta(s)$ outputs a distribution over 256 discrete switch configurations and the critic $V_\phi(s)$ estimates state values. The critic minimizes the temporal difference (TD) error as follows.

$$L^{VF}(\phi) = E\left[(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t))^2\right], \tag{8}$$

where r_t is the immediate reward at time t. The actor maximizes the clipped advantage-weighted loss as follows.

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t\Big[\min\big(r_t(\theta)\hat{A}_t,\, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\big) + \beta\mathcal{H}\Big], \tag{9}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio between the new and old policies, and ϵ controls the clipping range, β weights the entropy term which can be defined as:

$$\mathcal{H}(\pi_\theta(s_t)) = E\left[-\log \pi_\theta(s_t)\right]. \tag{10}$$

$\hat{A}_t$ represents the advantage estimation of action a_t compared to the average return achieved starting from state s_t, and it can be defind as:

$$\hat{A}_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t). \tag{11}$$

1) State and Action Spaces: The state at time t is represented by all cell SOCs at the current time step

$$s_t = [SOC_1^t, SOC_2^t, \ldots, SOC_M^t, sw_1^{t-1}, \ldots, sw_{M-1}^{t-1}], \tag{12}$$

where SOC_i represents the SOC of cell B_i. Episodes terminate when any cell SOC falls below 0.3 to ensure safe operation.

The actions represent the switch configurations

$$a_t = [sw_1^t, sw_2^t, \ldots, sw_{M-1}^t]. \tag{13}$$

To maintain minimum voltage requirements (19.8V), parallel connections are limited to 4, constraining the action space to $\mathcal{A} = \sum_{j=0}^{4}\binom{M-1}{j}$. The actor network outputs a discrete action index that maps to valid configurations (256) via the lookup table $\mathcal{L}$ such as:

$$a_t = \mathcal{L}(i_t). \tag{14}$$

2) Reward Function: The proposed multi-objective reward function balances three objectives SOC imbalance reduction, switching frequency reduction, and runtime maximization.

$$r_t = w_{SOC} \cdot R_{SOC} + w_{sw} \cdot R_{sw} + w_{time} \cdot 1, \tag{15}$$

where w_{SOC}, w_{sw}, and w_{time} controls reward function priorities. The SOC balancing term is defined as:

$$R_{\text{SOC}} = -\text{Var}\Big[SOC_1^t, \ldots, SOC_M^t\Big], \tag{16}$$

while the switching penalty is used to discourage frequent switch changes and it can be defined as

$$R_{sw} = -\sum_{i=1}^{M} \mathbf{1}\big[\mathcal{S}_i^{t-1} \neq \mathcal{S}_i^t\big]. \tag{17}$$

The three-term reward structure reflects practical tradeoffs inherent to battery management. The SOC variance term R_{SOC} penalizes charge spread across cells, directly incentivizing equalization throughout the discharge process. The switching penalty R_{sw} is included because frequent topology changes introduce mechanical wear on switching components and can induce transient disturbances in the current distribution. Finally, the constant runtime reward encourages the agent to sustain pack operation as long as possible before

TABLE I
COMPARISON OF THE PROPOSED APPROACH VS BASELINE CONTROLLERS

Metrics	Rule-based	Greedy-based	Proposed algorithm
ΔSOC	0.0391 ($\pm$0.008)	0.0197 ($\pm$0.004)	**0.0163** ($\pm$0.003)
Runtime	133.25 ($\pm$4.91)	105.3 ($\pm$4.09)	**133.45** ($\pm$4.6)
Switch changes	5.79 ($\pm$0.44)	**3.054** ($\pm$0.57)	4.3 ($\pm$0.33)

any cell reaches the discharge cutoff of 0.3 V, aligning the learned policy with the broader objective of maximizing usable energy delivery of the pack. The weights w_{SOC}, w_{sw}, and w_{time} allow tuning of the relative priority of each objective depending on the target application.

IV. RESULTS

This section evaluates the proposed RL-based reconfiguration control. The RL agent is trained on the surrogate model for computational efficiency, enabling thousands of episodes in minimal time, following by testing on the physics-based high fidelity model discussed in Section II. We use a 10-cell reconfigurable battery pack with cell-to-cell capacity variations, yielding 256 unique configurations under voltage constraints. The initial conditions are randomized, such that T_s, $T_c \in [17.5C, 27.5C]$ and SOCs $\in [0.8, 1.0]$. The battery pack undergoes a constant-current 1A discharge until any cell SOC drops below 0.3.

To evaluate the performance of the proposed RL controller, the best-performing RL policy, trained by interacting solely with the surrogate model, is tested on the physics-based simulator. A total of 47 independent test runs, each with random initial conditions, are performed. Two baselines are used to compare the proposed RL policy: a rule-based, and a model-based greedy controller. The rule-based baseline, inspired from [15], connects four pairs of adjacent cells with the lowest combined SOC in parallel at each time step. This allows the lower-SOC cells draw less current to help equalize the pack during discharge. The model-based greedy controller performs single-step optimization by evaluating all valid switch topologies through one-step-ahead SOC simulation using the physics-based model. The topology with the lowest ΔSOC is selected and executed. This baseline serves as a short-horizon near-optimal approach for comparison.

Table I summarizes the 47-run evaluation, where "ΔSOC" is the maximum-minimum SOC difference within the pack, "Runtime" is the time (minutes) until any cell's SOC drops below 0.3, and "Switch change" is the average number of series-parallel transitions per time step in each connection set (S_1, S_2, S_3). The proposed RL controller consistently outperforms both baselines. The RL controller achieves the lowest ΔSOC of 0.0163 ($\sigma = 0.003$), compared to 0.0391 ($\sigma = 0.0089$) and 0.0197 ($\sigma = 0.0048$) for the rule-based and greedy approaches, respectively. Notably, the RL controller maintains the lowest standard deviation, which shows robustness and stability. Additionally, it maintains a higher runtime of 133.45 minutes (compared to 133.25 and 105.3 for rule-based and greedy-based respectively) but with more switch changes (4.3), an acceptable tradeoff given the 17.2% improvement in SOC equalization. Fig. 4 shows the SOC evolution and ΔSOC during a test episode under the RL controller. All 10 cells converge smoothly, showing effective equalization.

These results collectively demonstrate the superiority of the proposed RL-based controller over both rule-based and greedy-based baselines.

V. CONCLUSION

This paper presents a reinforcement learning-based control framework for real-time SOC balancing in reconfigurable battery packs. The proposed approach successfully addresses the computational challenges of controlling reconfigurable battery systems by developing a linear surrogate model that achieved 98.68% accuracy while requiring only 0.012% computational time compared to high-fidelity simulation. The effectiveness of the proposed controller is validated through testing on a 10-cell reconfigurable battery pack under randomized initial SOC and temperature conditions. Compared with both rule-based and greedy-based baselines, the RL controller achieved the best balancing performance, producing the lowest ΔSOC while also maintaining strong runtime performance. Future work will focus on experimental validation on physical battery hardware to further validate the proposed controller. In addition, extending the framework to jointly consider temperature balancing, degradation-aware control, and larger battery packs with more complex switching architectures would further strengthen the practical applicability of the proposed approach.

REFERENCES

[1] A. Beaudet, F. Larouche, K. Amouzegar, P. Bouchard, and K. Zaghib, "Key challenges and opportunities for recycling electric vehicle battery materials," *Sustainability*, vol. 12, no. 14, p. 5837, 2020.

[2] J. Chen, A. Behal, and C. Li, "Active cell balancing by model predictive control for real time range extension," in *IEEE Conference on Decision and Control*, Austin, TX, USA, December 13–15, 2021.

[3] Z. B. Omariba, L. Zhang, and D. Sun, "Review of battery cell balancing methodologies for optimizing battery pack performance in electric vehicles," *Ieee Access*, vol. 7, pp. 129 335–129 352, 2019.

[4] J. Chen, L. Zhang, and W. Gao, "Reconfigurable model predictive control for large scale distributed systems," *IEEE Systems Journal*, vol. 18, no. 2, pp. 965–976, June 2024.

[5] J. Chen, A. Behal, and C. Li, "Active battery cell balancing by real time model predictive control for extending electric vehicle driving range," *IEEE Trans. Auto. Sci. Eng.*, vol. 21, no. 3, pp. 4003–4015, July 2024.

[6] D. Flessner and J. Chen, "Reinforcement learning-based event-triggered model predictive control for electric vehicle active battery cell balancing," *ASME Letters in Dynamic Systems and Control*, vol. 5, no. 2, pp. 1–6, April 2025.

[7] J. Cao, N. Schofield, and A. Emadi, "Battery balancing methods: A comprehensive review," in *2008 IEEE Vehicle Power and Propulsion Conference*. IEEE, 2008, pp. 1–6.

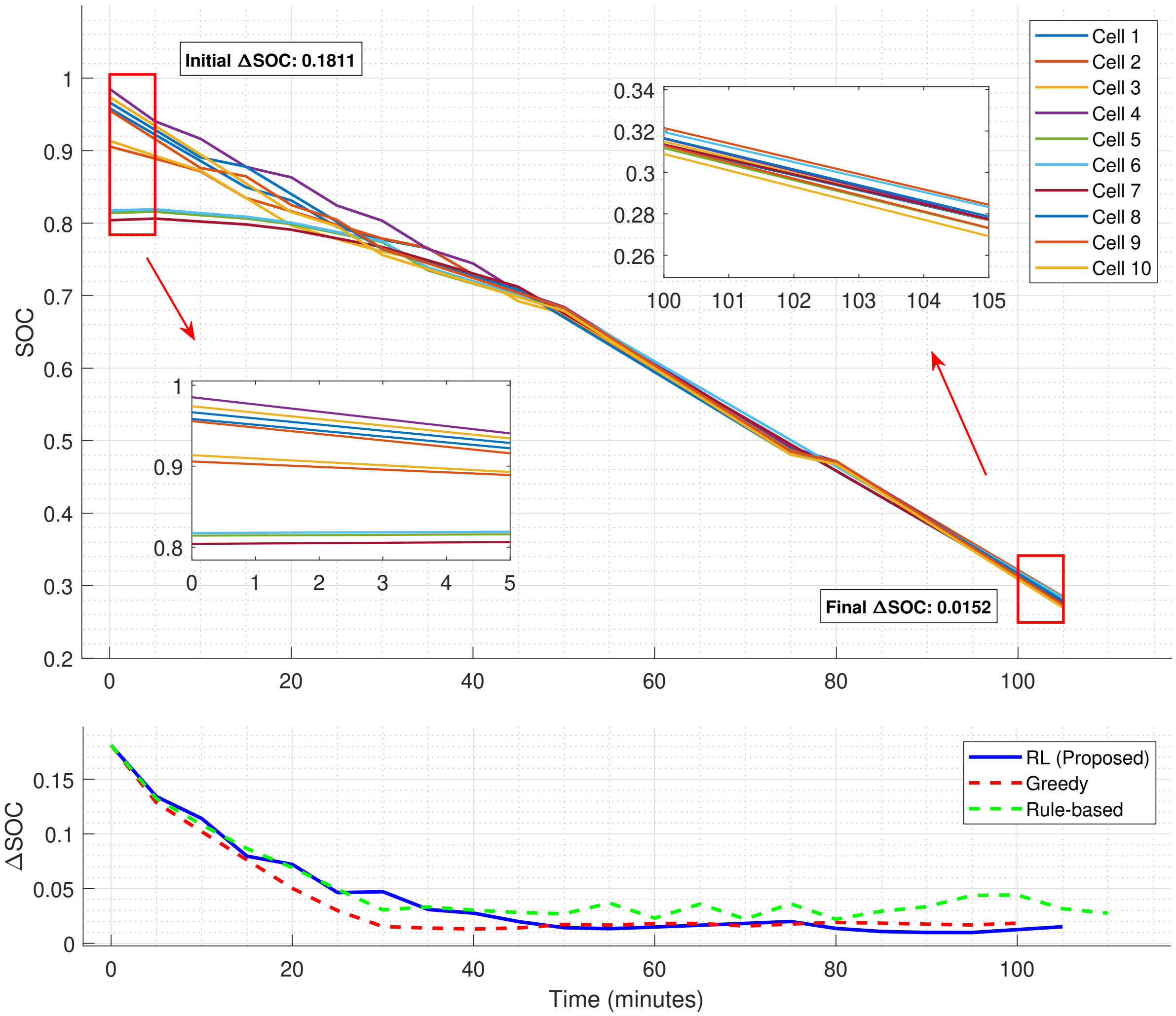

Fig. 4. Performance of RL controller balancing a 10-cell reconfigurable battery pack. *Top*: individual cell SOC trajectories over time. *Bottom*: the change in ΔSOC over time.

[8] A. Kelkar, Y. Dasari, and S. S. Williamson, "A comprehensive review of power electronics enabled active battery cell balancing for smart energy management," in *IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy*, 2020, pp. 1–6.

[9] F. Yang, F. Gao, B. Liu, and S. Ci, "An adaptive control framework for dynamically reconfigurable battery systems based on deep reinforcement learning," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 12, pp. 12 980–12 987, 2022.

[10] S. Jeon, J. Kim, J. Ahn, and H. Cha, "Optimizing discharge efficiency of reconfigurable battery with deep reinforcement learning," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 39, no. 11, pp. 3893–3905, 2020.

[11] A. Irshayyid and J. Chen, "Gnn-based surrogate model for reconfigurable battery packs," in *IEEE Conference on Control Technology and Applications*, San Diego, CA, August 25–27, 2025.

[12] N. Lin and S. Ci, "Toward dynamic programming-based management in reconfigurable battery packs," in *2017 IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, 2017, pp. 2136–2140.

[13] Y. Yang, J. He, C. Chen, and J. Wei, "Balancing awareness fast charging control for lithium-ion battery pack using deep reinforcement learning," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 4, pp. 3718–3727, 2023.

[14] T.-c. Chiueh, M.-C. Huang, K.-C. Juang, S.-H. Liang, and W. Ling, "Virtualizing energy storage management using RAIBA," in *2018 USENIX Annual Technical Conference (USENIX ATC 18)*, 2018, pp. 187–198.

[15] S. W. Moore and P. J. Schneider, "A review of cell equalization methods for lithium ion and lithium polymer battery systems," 2001.

[16] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[17] A. Irshayyid and J. Chen, "Comparative study of cooperative platoon merging control based on reinforcement learning," *Sensors*, vol. 23, no. 2, p. 990, 2023.

[18] A. Irshayyid, J. Chen, and G. Xiong, "A review on reinforcement learning-based highway autonomous vehicle control," *Green Energy and Intelligent Transportation*, vol. 3, no. 4, p. 100156, 2024.

[19] M. Ibrahim and R. Elhafiz, "Fleet management system's attack graph," in *IEEE Int. Conf. Electronics, Computers and AI*, 2025, pp. 1–7.

[20] M. Ibrahim, R. Elhafiz, and A. Al-Wadi, "Reinforcement learning-based attack graph analysis for wastewater treatment plant," *IEEE Transactions on Industry Applications*, vol. 59, no. 6, pp. 7858–7867, 2023.

[21] M. Ibrahim and R. Elhafiz, "Security assessment of industrial control system applying reinforcement learning," *Processes*, vol. 12, no. 4, p. 801, 2024.

[22] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep reinforcement learning-based energy storage arbitrage with accurate

lithium-ion battery degradation model," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4513–4521, 2020.

[23] Y. Weng and C. Ababei, "Ai-assisted reconfiguration of battery packs for cell balancing to extend driving runtime," *Journal of Energy Storage*, vol. 84, p. 110853, 2024.

[24] H. E. Perez, J. B. Siegel, X. Lin, A. G. Stefanopoulou, Y. Ding, and M. P. Castanier, "Parameterization and validation of an integrated electro-thermal cylindrical lfp battery model," in *Dynamic Systems and Control Conference*, vol. 45318, 2012, pp. 41–50.

[25] J. Chen, Z. Zhou, Z. Zhou, X. Wang, and B. Liaw, "Impact of battery cell imbalance on electric vehicle range," *Green Energy and Intelligent Transportation*, vol. 1, no. 3, p. 100025, 2022.

[26] X. Lin, H. E. Perez, S. Mohan, J. B. Siegel, A. G. Stefanopoulou, Y. Ding, and M. P. Castanier, "A lumped-parameter electro-thermal model for cylindrical batteries," *Journal of Power Sources*, vol. 257, pp. 1–11, 2014.

[27] C. Forgez, D. V. Do, G. Friedrich, M. Morcrette, and C. Delacourt, "Thermal modeling of a cylindrical lifepo4/graphite lithium-ion battery," *Journal of power sources*, vol. 195, no. 9, pp. 2961–2968, 2010.

[28] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 07 2017.