

Deep Reinforcement Learning Based Fast Charging and Thermal Management Optimization of An Electric Vehicle Battery Pack

Mohammad Hossein Abbasi^a, Ziba Arjmandzadeh^b, Jiangfeng Zhang^{a,*}, Bin Xu^b, Venkat Krovi^a

^aThe Department of Automotive Engineering, Clemson University, Greenville, SC 29607, USA.

^bThe Department of Aerospace and Mechanical Engineering, the University of Oklahoma, Norman, OK 73019, USA.

Abstract

Existing optimal strategies for fast-charging electric vehicle batteries are predominantly at the cell level. The proposed high-fidelity methods for extending available cell models to pack level are associated with a large computational burden, making real-world implementation impossible. Further, fast charging optimization and thermal management problems are dependent. That is, the cooling system reduces battery temperature, allowing for higher charging current, while optimal current minimizes the need for cooling and, in turn, reduces thermal system power consumption. There is a lack of studies where fast charging optimization and battery thermal management problems are jointly solved. Therefore, this paper proposes a simulation study using a deep reinforcement learning (RL) approach that concurrently solves fast charging and thermal management problems for a battery pack with low computational complexity. In this regard, we formulate each cell using an electro-thermal-aging model, which accounts for the heat exchange between adjacent cells. The electro-thermal-aging model plays the role of the environment for RL and is not the focus of novelty in this work. The RL agent is then trained to output the optimal charging current and coolant mass flow rate. Moreover, the proposed methodology is examined through a numerical study where the outperformance of our model is showcased by comparing it with baseline algorithms of model predictive control (MPC) and CC-CV. Consequently, three battery packs comprising 20, 444, and 7104 cells are used, respectively. We demonstrate that RL requires less than a second to finish the simulation for 20 cells, while MPC requires more than 80 minutes. In addition, RL keeps the cells' core temperature below 33°C, but MPC results reach 40°C. The RL performance in charging packs with 444 and 7104 cells is then compared with that of CC-CV. In terms of computation time, RL and CC-CV are nearly the same, while regarding the average core and surface temperature of the cells as well as the cell aging, RL attains better outcomes, extending the battery pack's life by up to two years after 1000 fast charging cycles.

Keywords: Electric Vehicle Fast Charging, Battery Thermal Management, Deep Reinforcement Learning, Proximal Policy Optimization, Model Predictive Control

1. Introduction

Increasing greenhouse gas emissions and fossil fuel depletion have triggered efforts toward mitigating our carbon footprint and promoting carbon neutrality [1]. Hence, a global consensus has been reached on phasing out internal combustion engine vehicles and replacing them with electric vehicles (EVs) [2]. However, the main bottleneck of EV development and popularity lies in the slow battery charging speed compared with the short refueling time of internal combustion vehicles [3], which negatively affects the service quality experienced by the consumer [4]. Furthermore, the fast charging of the lithium-ion batteries used in EVs is limited by battery degradation, which depends on temperature resulting from the heat generated by charging current [5]. The suitable battery operating temperature is said to be below 40°C, while the temperature gradient within the pack should be less than 5°C [6]. Nonuniform temperature rise, hand in hand with high current density, aggravates lithium concentration gradients, increasing mechanical stress

*Corresponding author

Email address: jiangfz@clemson.edu (Jiangfeng Zhang)

Nomenclature

Abbreviations:

BMS	Battery Management System
CC-CV	Constant current-constant voltage
ECM	Equivalent circuit model
MDP	Markov Decision Process
MPC	Model predictive control
OCV	Open circuit voltage
PPO	Proximal policy optimization
RL	Reinforcement learning
SOC	State of charge
TMS	Thermal Management System

Equivalent Circuit Model:

i	Index of cell
$I(t)$	Applied current
$I_i(t)$	Current passing through cell
$V_i(t)$	Pack terminal voltage
$V_{1,i}(t)$	First polarization voltage
$V_{2,i}(t)$	Second polarization voltage
$V_{t,i}(t)$	Cell terminal voltage
$z_i(t)$	Cell state of charge
$C_{1,i}(z, T_c)$	First polarization capacitance
$C_{2,i}(z, T_c)$	Second polarization capacitance
$R_{0,i}(z, T_c)$	Cell internal resistance
$R_{1,i}(z, T_c)$	First polarization resistance
$R_{2,i}(z, T_c)$	Second polarization resistance
$V_{OCV,i}(z)$	Cell open circuit voltage
$C_{b,i}$	Cell capacity

Thermal Model:

$\dot{m}_{cool}(t)$	Cooling fluid mass flow rate
$T_j^{NBR}(t)$	The surface temperature of j -th neighbor cell
$T_{cool,i}^{in}(t)$	Coolant temperature before reaching the cell

$T_{cool,i}^{out}(t)$	Coolant temperature after touching the cell
$T_{c,i}(t)$	Cell core temperature
$T_{f,i}(t)$	Ambient temperature
$T_{s,i}(t)$	Cell surface temperature
A_{cool}	Heat transfer area
C_{cool}	Cooling fluid specific heat capacity
C_c	Cell core heat capacity
C_s	Cell surface heat capacity
h_{cool}	Heat transfer coefficient
m_{cool}	Coolant mass
N	Number of neighboring cells
R_c	Heat conduction resistance
R_m	Cell-to-cell heat transfer resistance
R_u	Convection resistance

Aging Model:

$\Delta Q_i(t)$	Cell capacity loss percentage
$A_{h,i}(t)$	Cell ampere-hour throughput
$R_{rise,i}(t)$	Cell internal resistance increase percentage
E_1^a, E_2^a	Activation energy
R	Universal gas constant

Optimization Algorithms

$\dot{m}^{\max}$	Coolant mass flow rate upper-bound
ω_i	Weighting parameter
c	MPC control horizon
$I^{\max}$	Applied current upper-bound
N_p	Number of parallel connected cells
N_s	Number of series connected cells
p	MPC prediction horizon
$T^{\max}$	Temperature upper-bound
T_c^{thr}	Core temperature level for penalizing RL agent
z^*	Target SOC level

on the cell [7]. Besides, the battery management system (BMS) handles fast-charging current, and the thermal management system (TMS) controls the battery temperature. From the optimization point of view, it is useful to consider the mutual impacts of BMS and TMS on each other, as TMS enables higher charging current by cooling the battery. In contrast, current optimization can reduce the need for battery cooling, which cuts back the incurred TMS power consumption. Thus, this work proposes an algorithm based on deep reinforcement learning (RL) to optimally fast charge a battery pack and control the underlying thermal management system.

EV battery fast charging is a multifaceted problem in terms of battery model selection, aging and thermal characterization, range of charging speed, and battery pack size. Either equivalent circuit models (ECMs) or electrochemical models are utilized for the battery model. ECMs enjoy fast and straightforward implementation but lack the inclusion of internal battery states, while electrochemical models are too slow for onboard applications [8]. Ref. [9] uses a thermoelectric battery model to optimize the constant current-constant voltage (CC-CV) charging algorithm. In [10], a Bayesian optimization model is developed to fast charge a battery in the presence of voltage and temperature constraints. Although the thermal dynamics of the battery are considered in [10], the study focuses on one Li-ion battery

in the absence of a cooling system. When charging a single cell, the strategy in [9] finds the optimal current for the CC stage without considering the aging characteristics. Ref. [11] extends the work in [9] by adding a battery aging model and optimizing the currents of the multistage constant-current charging profile. The study in [11] also concentrates on a single cell and considers battery capacity loss in the aging model, ignoring the internal resistance rise mechanism. In [8], ECM coupled with thermal and aging models is employed to optimally fast charge a battery pack while minimizing the incurred costs of the fast charging station. However, the ECM model is extended into a pack model, ignoring the differences in cells' aging and charging currents. Likewise, [12] and [13] use an electro-thermal-aging model to fast charge the battery via a model-based optimization algorithm and RL, respectively, but both studies focus on a single cell rather than a battery pack. Ref. [14] proposes an adaptive boosting charging mechanism that uses a capacity estimation method that adaptively adjusts the charging strategy. Even though a coolant is considered in [14], the study is limited to one cell, and the coolant behavior is not optimized. Electrochemical-thermal modeling is used in [15] to fast charge a battery through actor-critic RL. However, the work is restricted to one cell and lacks consideration of TMS. The research in [16] finds the optimal charging strategy for a battery pack using stochastic model predictive control (MPC), where the underlying model comprises a simplified electrochemical model based on the single particle model with electrolyte, a lumped thermal model, and an aging formula describing capacity loss and solid electrolyte interphase layer growth. The stochastic MPC in [16] is cast into deterministic MPC and solved using polynomial chaos expansion; however, the case study is conducted on a battery pack with four cells, and the simulation time is not reported.

Thermal management in EVs is essential for extending EV range, rendering energy-efficient vehicle operation, and maintaining safety by preventing component overheating [17]. The types of thermal cooling systems include air cooling, liquid cooling, phase change material cooling, and heat pipe cooling, where liquid cooling is believed to be more effective and efficient [18]. Nevertheless, most of the research in this area concentrates on thermal management alone and disregards its influence on fast charging optimization and vice versa. The work in [17] conducts a parametric study on the vehicle level thermal management and its impacts on the Tesla Model S and Model 3. The thermal system incorporates air conditioning refrigerant, indirect battery liquid cooling, and cabin cooling air loops. Five PID controllers are designed to control the temperature of the battery pack, cabin climate, condenser subcool, evaporator superheat, and chiller superheat, respectively. However, the study does not optimize the coolant mass flow rate, and the performance of the proposed model under battery fast charging is not investigated. Ref. [19] proposes an algorithm based on neural networks to find the optimal charging current and coolant mass flow rate from among a finite set of current and coolant candidates. Moreover, the experimental setting comprises eight series-connected cells. The study in [19] is validated by a few mass flow rates and a maximum charging current of 2.5C for a small battery module. The research in [2] presents a temperature-difference-aware fast charging approach based on a genetic algorithm considering a lumped thermal model cooled by a liquid coolant. Despite optimizing the charging current, the mass flow rate is fixed. In addition, the study in [2] is at the module level with an experimental setup of 16 cells.

RL is a technique for solving sequential decision-making problems where an agent learns how to act by interacting with the environment. The learning process is guided by a reward signal, which should be maximized. The main advantages of RL are as follows. In the case of a model-free RL, the approach does not require system dynamics. Furthermore, RL is adaptive to uncertainties since it learns directly by interacting with the environment, and it does not rely on a model prone to noises stemming from model inaccuracies. Deep deterministic policy gradient (DDPG) is applied in [13] to fast charge a battery. The proposed RL algorithm in [13] outperforms CC-CV and MPC when charging a single cell. Similarly, [15, 20, 21] exploit RL to fast charge a single cell using an electrochemical model. Further, an adaptive model-based RL is proposed in [22] to fast charge a lithium-ion cell with minimal degradation, where a probabilistic surrogate model of differential Gaussian process describes cell aging. Ref. [23] investigates a balancing-aware battery pack fast charging based on Deep Q-Network. However, the case study utilizes a pack of four cells far smaller than EV battery packs. In [24], soft actor-critic RL is used to fast charge a Li-ion battery cell using an electro-thermal model. However, the study in [24] lacks an aging model and does not consider the impact of a cooling system. The work in [25] employs RL to train a neural network for fast charging an EV battery pack. However, the EV battery behavior is simplified into the change in the state of charge (SOC) during fast charging, disregarding the battery's internal electrical, thermal, and aging dynamics.

The above literature study identifies two research gaps: i) The existing literature lacks simultaneous optimization of fast charging and thermal management for a battery pack, and ii) Previous studies predominantly analyze a single cell or a small battery module usually made up of less than 20 cells; hence, the analysis of a large battery pack is

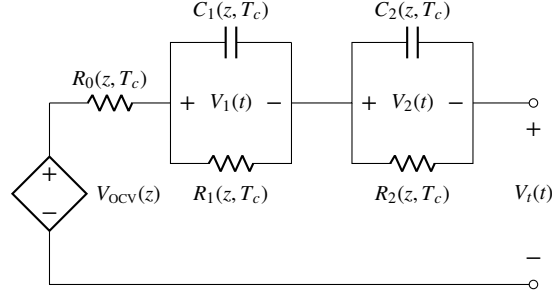


Figure 1: Second-order equivalent circuit model of each cell

missing. It is useful to study concurrent optimization of fast charging and thermal management to assess their mutual impact on one another. Moreover, considering a large battery pack rather than a cell or small module integrates the implications of cell-to-cell heat transfer, the difference in cell aging, cell imbalance, and the increased computation time into the model. Therefore, we propose an algorithm leveraging the proximal policy optimization (PPO) method to fast charge, in a simulation environment, a large battery pack with 7104 cells while controlling its TMS behavior. PPO was proposed to address the following issues of RL. Unlike supervised learning, the agent in RL creates its own data by interacting with its environment rather than utilizing a static dataset. Hence, as the RL agent learns, the data changes because the agent's actions influence the data, and the change in data can lead to instability. In addition, RL performance is generally very sensitive to hyperparameters. Consequently, PPO was devised to address the abovementioned issues and bring about the benefits of easy implementation and tuning. Further, PPO is a policy gradient method (on-policy) that, unlike Q-learning, does not learn from stored off-line data and, thus, does not have a replay buffer to store past experiences. Instead, PPO directly learns from what the agent encounters in the environment. Ultimately, the contributions of this manuscript are as follows:

- Using the PPO method, a model-free RL-based algorithm is proposed to design a simulation tool for larger pack concurrent fast charge and thermal management.
- Each cell of the battery pack is modeled through an electro-thermal-aging model accounting for cell-to-cell heat transfer and the impact of the cooling fluid.
- Using battery packs comprising 20, 444, and 7104 cells, the performance of the proposed approach is benchmarked against MPC and CC-CV, and the superiority of the PPO-based strategy is demonstrated.

The remaining sections of the paper are structured as follows. Section 2 introduces the problem and provides a detailed formulation of the electrical, thermal, and aging models. In Section 3, the proposed algorithm of PPO is elaborated. The simulation results are explored in Section 4, and Section 5 concludes the paper.

2. Problem Description

This paper simultaneously solves the fast charging optimization and thermal management of an EV battery pack in a simulated environment. The optimization problem finds the charging current and coolant mass flow rate in such a way that the battery pack is charged quickly with minimum degradation. To formulate the problem, each cell in the battery pack is characterized by an electro-thermal-aging model comprising a second-order ECM (see Fig. 1), a two-state thermal model augmented by considering interconnection between adjacent cells, and an aging model consisting of a lumped capacity fade formula along with an internal resistance rise function. Since exploiting a model-based approach would be intractable due to the nonlinearity and a large number of differential equations of the problem, we solve the problem through a model-free deep RL approach. Moreover, the performance of the developed RL method is assessed by comparing it with MPC for a small battery module. Once the superiority of the developed model-free strategy is established, it is used for optimal charging of a large battery pack.

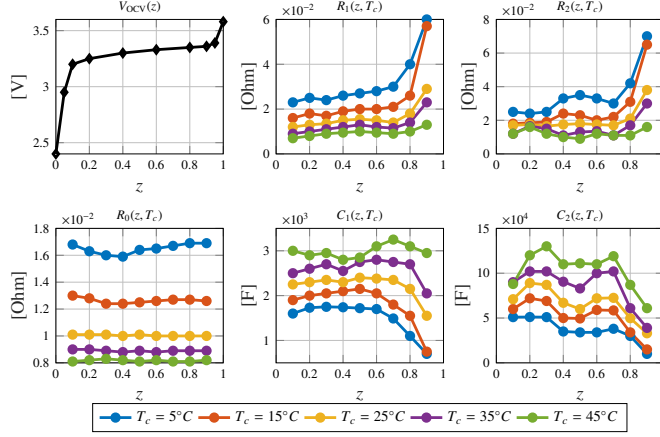


Figure 2: Equivalent circuit model data [26]

2.1. Equivalent Circuit Model

The electrical behavior of each cell of the battery pack is modeled by a second-order ECM as depicted in Fig. 1, where R_0 and RC pairs are functions of the cell's core temperature, $T_c(t)$, and SOC, $z(t)$, while the open circuit voltage is only a function of SOC. The variations of the said parameters with $z(t)$ and/or $T_c(t)$ are illustrated in Fig. 2, which only describe the changes of parameter values when the cell is charging; the cell's discharging characteristics follow different curves [26]. However, we only focus on cell fast charging in this work. In the following, the governing equations of the ECM are presented. Hereafter, for conciseness, the parameter $X(z, T_c)$ is written as X , where X can be any of the parameters given in Fig. 2. For the i -th cell in the pack, we have [8]

$$\frac{dz_i(t)}{dt} = \frac{I_i(t)}{C_{b,i}}, \quad (1a)$$

$$\frac{dV_{1,i}(t)}{dt} = -\frac{V_{1,i}(t)}{R_{1,i}C_{1,i}} + \frac{I_i(t)}{C_{1,i}}, \quad (1b)$$

$$\frac{dV_{2,i}(t)}{dt} = -\frac{V_{2,i}(t)}{R_{2,i}C_{2,i}} + \frac{I_i(t)}{C_{2,i}}, \quad (1c)$$

$$V_{t,i}(t) = V_{OCV,i} + V_{1,i}(t) + V_{2,i}(t) + R_{0,i}I_i(t). \quad (1d)$$

The data in Fig. 2 are obtained for core temperature values of $T_c(t) \in \{5, 15, 25, 35, 45\}$ degree Celsius and the following SOC, $z(t) \in \{0.1, 0.2, \dots, 0.9\}$. Our study employs the data as a lookup table where any temperature value other than $\{5, 15, 25, 35, 45\}$ and any SOC value other than $\{0.1, 0.2, \dots, 0.9\}$ is interpolated using nearest neighbor interpolation. In addition, the proposed algorithms are designed to keep the maximum temperature below 40°C while SOC increases from 20% to 80%.

2.2. Thermal Model

A two-state thermal model is used to describe the cell's thermal behavior [26]. Furthermore, the model is augmented to capture thermal coupling between adjacent cells [17, 27]. The battery cells can be organized in compact or regular cylindrical configurations in the pack [28]. We select the regular configuration, where a maximum of eight neighbor cells surround each cell [28]. For cell i we have

$$\frac{dT_{c,i}(t)}{dt} = \frac{T_{s,i}(t) - T_{c,i}(t)}{R_c C_c} + \frac{|I_i(t)(V_{OCV,i} - V_{t,i}(t))|}{C_c}, \quad (2a)$$

$$\frac{dT_{s,i}(t)}{dt} = \frac{T_{f,i}(t) - T_{s,i}(t)}{R_u C_s} - \frac{T_{s,i}(t) - T_{c,i}(t)}{R_c C_s} - \frac{1}{R_m C_s} \sum_{j=1}^N (T_{s,i}(t) - T_j^{\text{NBR}}(t)) - \frac{T_{s,i}(t) - T_{\text{cool},i}^{\text{in}}(t)}{R_{\text{cool}} C_s}, \quad (2b)$$

Table 1: The Values of B_1 For Different C-rates				
C-rate	C/2	2C	6C	10C
B_1	31630	21681	12934	15512

$$\frac{dT_{\text{cool},i}^{\text{out}}(t)}{dt} = \frac{\dot{m}_{\text{cool}}(t)}{m_{\text{cool}}} (T_{\text{cool},i}^{\text{in}}(t) - T_{\text{cool},i}^{\text{out}}(t)) + \frac{1}{m_{\text{cool}} R_{\text{cool}} C_{\text{cool}}} (T_{s,i}(t) - T_{\text{cool},i}^{\text{in}}(t)), \quad (2c)$$

where (2a) states how cell core temperature changes based on its surface temperature and the joule heat generated as current passes through the cell. In (2a), R_c indicates conduction resistance against heat transfer between the core and surface of the cell. The heat exchange rate is controlled by the core's heat capacity, C_c . Moreover, Eq. (2b) represents cell surface temperature alteration as a function of ambient temperature, cell core temperature, the surface temperature of adjacent cells, and coolant temperature, respectively. The parameters R_u and R_m state the convection resistance against cell surface cooling and the cell-to-cell heat transfer resistance, respectively, while surface heat capacity, C_s , alters the incidence of heat exchange through the cell's surface. Furthermore, $R_{\text{cool}} = 1/(h_{\text{cool}} A_{\text{cool}})$ displays the heat transfer resistance against the coolant and the surface of the cell. On the other hand, the equation in (2c) expresses the dynamic of coolant temperature as it exchanges heat with the cell. The variable $T_{\text{cool}}^{\text{in}}(t)$ represents the coolant temperature before reaching the cell, and $T_{\text{cool}}^{\text{out}}(t)$ is the coolant temperature after it passes by the cell and absorbs heat from it. Lastly, the model parameter values are available in [17, 26].

2.3. Aging Model

The aging model comprises capacity loss and internal resistance rise, which are presented in the following, respectively [29, 30].

$$\Delta Q_i(t) = B_1 \exp\left(-\frac{E_1^a}{RT_{c,i}}\right) (A_{h,i}(t))^\gamma, \quad (3)$$

$$R_{\text{rise},i}(t) = B_2 \exp\left(-\frac{E_2^a}{RT_{c,i}}\right) A_{h,i}(t), \quad (4)$$

where (3) denotes capacity loss percentage and (4) indicates the percentage increase of internal resistance. The parameter B_1 in (3) is a function of C-rate, and its experimental values are shown in Table 1. The four points in Table 1 are used to interpolate B_1 for any other C-rate values via nearest neighbor interpolation. Additionally, the activation energy E_1^a follows C-rate through the following equation, $E_1^a = -31700 + 370.3 \times \text{C-rate}$, and the power law factor is $\gamma = 0.55$ [29]. Moreover, the parameter $B_2 = \alpha_R + \gamma_R \exp(4d)$, where $\alpha_R = 3.2053 \times 10^5$, $\gamma_R = 3.6342 \times 10^3$, and $d = 0.9179$. Besides, the activation energy in (4) is $E_2^a = 51800$ [J/mol] [30]. Finally, $R = 8.314$ [J/K/mol] is the universal gas constant, and $A_h(t)$ exhibits the cell's ampere-hour throughput, equal to the integral of charging current from the beginning of the cell's life.

2.4. Current Distribution Among Parallel Cells

Due to the temperature dependency of the ECM model and discrepancy in cell aging, the current passing through parallel cells is not necessarily equally divided. Consequently, we find the relationship by which the current is distributed among parallel cells in the pack. Note that this relation is only used to simulate the environment and is not exploited by the RL algorithm. Eq. (1d), can be rewritten for the i -th cell of N_p parallel cells as follows,

$$V_t(t) = V_{\text{ocv},i} + V_{1,i}(t) + V_{2,i}(t) + R_{0,i} I_i(t), \quad (5)$$

where $I_i(t)$ is to be found. Further, since $I_i(t)$ is unknown, $V_t(t)$ is unknown too. Hence, by rearranging (5), we obtain

$$I_i(t) = \frac{V_t(t) - V_{\text{ocv},i} - V_{1,i}(t) - V_{2,i}(t)}{R_{0,i}}. \quad (6)$$

The currents of all parallel cells sum to $I(t)$, which is the applied current to the pack. Therefore,

$$I(t) = \sum_{i=1}^{N_p} I_i(t). \quad (7)$$

By substituting (6) in (7), we have

$$I(t) = \sum_{i=1}^{N_p} \left(\frac{V_t(t) - V_{ocv,i} - V_{1,i}(t) - V_{2,i}(t)}{R_{0,i}} \right), \quad (8)$$

$$I(t) = V_t(t) \sum_{i=1}^{N_p} \frac{1}{R_{0,i}} - \sum_{i=1}^{N_p} \left(\frac{V_{ocv,i} + V_{1,i}(t) + V_{2,i}(t)}{R_{0,i}} \right). \quad (9)$$

From (9), $V(t)$ can be calculated as

$$V_t(t) = \frac{I(t) + \beta}{\alpha}, \quad (10)$$

$$\alpha = \sum_{i=1}^{N_p} \frac{1}{R_{0,i}}, \quad (11)$$

$$\beta = \sum_{i=1}^{N_p} \left(\frac{V_{ocv,i} + V_{1,i}(t) + V_{2,i}(t)}{R_{0,i}} \right). \quad (12)$$

Finally, the current for each parallel cell is obtained by substituting (10) in (6). It is worth mentioning that for $i \in \{1, 2, \dots, N_p\}$, the variables $V_{1,i}(t)$ and $V_{2,i}(t)$ are recursively calculated for each time step after proper discretization of the problem.

3. Fast Charging Algorithms

This section is dedicated to elaborating the RL algorithm used in this study.

3.1. Markov Decision Process (MPD)

An MDP is a mathematical tool that provides a framework for modeling sequential decision-making problems where the outcomes are partly dependent on control actions and partly random. MPDs are effective in solving optimization problems, and if solved by dynamic programming, they can find the optimal solution. The MDP sequence flow is as follows. First, the system or environment is at state s . The agent, a.k.a. the decision maker, chooses action a based on an action selection algorithm. This action takes the system to state s' and generates a reward $r(s, a, s')$. The probability that executing action a at state s would take the system to state s' and generate reward $r(s, a, s')$ is $p(s', r|s, a)$, which is called transition probability. It merits mentioning that dynamic programming is only applicable to MPD for small-scale problems with known transition probabilities [31]. Accordingly, RL is devised to relax the need for transition probabilities.

3.2. Modeling EV Battery Fast Charging with MDP

The EV fast-charging problem can be characterized as an agent who completes the charging duty through sequential decisions on charging current C-rate and cooling fluid mass flow rate. In this respect, the EV fast charging problem is cast into an MDP model which is explained in the following.

3.2.1. Fast Charging Environment

The environment or system is the battery pack comprising N_s series strings of N_p parallel lithium-ion cells (A123 ANR26650M1). An RL agent interacts with the system such that the agent takes action and observes the results. Hence, to model the battery pack accurately, the decision variables in (1)-(4) are considered for each cell in the pack. In other words, the current passing through each cell is calculated, which is used to compute each cell's terminal voltage, core temperature, surface temperature, capacity loss, and internal resistance rise each time step. Moreover, the coolant temperature near each cell is also integrated into the model. The RL agent takes continuous actions, i.e., the C-rate of applied current and the mass flow rate of coolant, that are within limited boundaries. Subsequently, the actions are passed into the system, and feedback is returned in the form of a reward.

3.2.2. Observation Space

This paper considers the maximum cell SOC, $z^{\max}(t)$, and the surface temperature of the hottest cell in the pack, $T_s^{\max}(t)$, as states. To locate the hottest cell, we consider a bottom-cooling system as the topology of this study's thermal management system (TMS). In the bottom-cooling setting, the hottest cell is roughly situated at the center of one side of the module [32]. We assume the BMS performs perfect cell balancing and, therefore, the SOC of the pack is equal to that of each cell or $z_i(t) = z^{\max}(t)$, $\forall i \in \{1, 2, \dots, N_s N_p\}$. The observation space is a two-dimensional continuous space where the first and second state boundaries are defined as $T_s^{\max}(t) \in [5^\circ\text{C}, 45^\circ\text{C}]$ and $z^{\max}(t) \in [0, 1]$. The boundaries for $T_s^{\max}(t)$ are selected based on the adopted data from [26].

3.2.3. Action Space

The actions are taken by the RL agent in order to maximize the cumulative reward. In other words, the actions are control signals aimed to orchestrate the environment in a desirable direction. In our problem, the actions are the C-rate of charging/applied current as well as the mass flow rate of cooling fluid. In addition, the action space is a two-dimensional continuous space where C-rate values can be anything between 0 and 10 and mass flow rate is limited to 0.1, i.e., $\dot{m}_{\text{cool}}(t) \in (0, 0.1]$ [kg/s] [17, 33]. The C-rate upper-bound is selected as a large value by which an empty battery can be fully charged in less than 10 minutes [34]. Moreover, the mass flow rate upper limit is chosen as 0.1 [kg/s] such that the battery can be cooled up to 5 degrees while the TMS power consumption is kept at a low to moderate level.

3.2.4. Reward

Reward signal indicates the desirability of transitioning from state s to s' by taking action a . In essence, the reward function plays the role of the objective function of an optimization problem. In this manuscript, the problem's goal is to minimize the charging time and battery aging. Thus, we define the reward signal in the following form

$$R(t) = -\omega_1 |z^{\max}(t) - z^*| - \omega_2 \max\{T_c^{\max}(t) - T_c^{\text{thr}}, 0\} - \omega_3 \Delta Q^{\max}(t) - \omega_4 \Delta R_{\text{rise}}^{\max}(t) - \omega_5 |I(t)| - \omega_6 \dot{m}_{\text{cool}}(t), \quad (13)$$

where the first term penalizes the agent if the SOC level is not equal to the target SOC of $z^* = 0.8$. Subsequently, the second term reduces the reward if the core temperature of the hottest cell in the pack exceeds the threshold level of $T_c^{\text{thr}} = 32^\circ\text{C}$. Next, the reward is decreased by two aging terms, which indicate the percentage capacity loss and internal resistance rise of the cell with maximum degradation. Lastly, the two inputs or actions are added to the reward function to investigate whether their presence improves the results. Note that the reward function is constructed in such a way as to be non-positive, $R(t) \leq 0$. In the following, the logic behind each term in the reward function is elaborated:

- $|z^{\max}(t) - z^*|$ is added to ensure that the RL agent learns to charge the battery.
- $\max\{T_c^{\max}(t) - T_c^{\text{thr}}, 0\}$ penalizes the reward function if the maximum temperature in the pack becomes greater than T_c^{thr} . This term helps the RL agent learn to respond quicker to the rise of battery temperature. The goal is to control the battery temperature to minimize degradation during fast charging. It is challenging to rely only on the battery capacity loss as a metric to find an optimal charging current to minimize degradation because of the following reasons: 1) battery capacity cannot be accurately measured, and 2) battery capacity may not change significantly during one charging session. Thus, the second term is added to the reward function.
- $\Delta Q^{\max}(t)$ and $\Delta R_{\text{rise}}^{\max}(t)$, where maximum capacity drop and maximum resistance rise are considered. These terms account for aging, which helps the RL agent learn how effective its choices were in the long run.
- $|I(t)|$ and $\dot{m}_{\text{cool}}(t)$, which are the inputs. Adding these terms helps find an optimal solution with minimal effort or energy consumption.

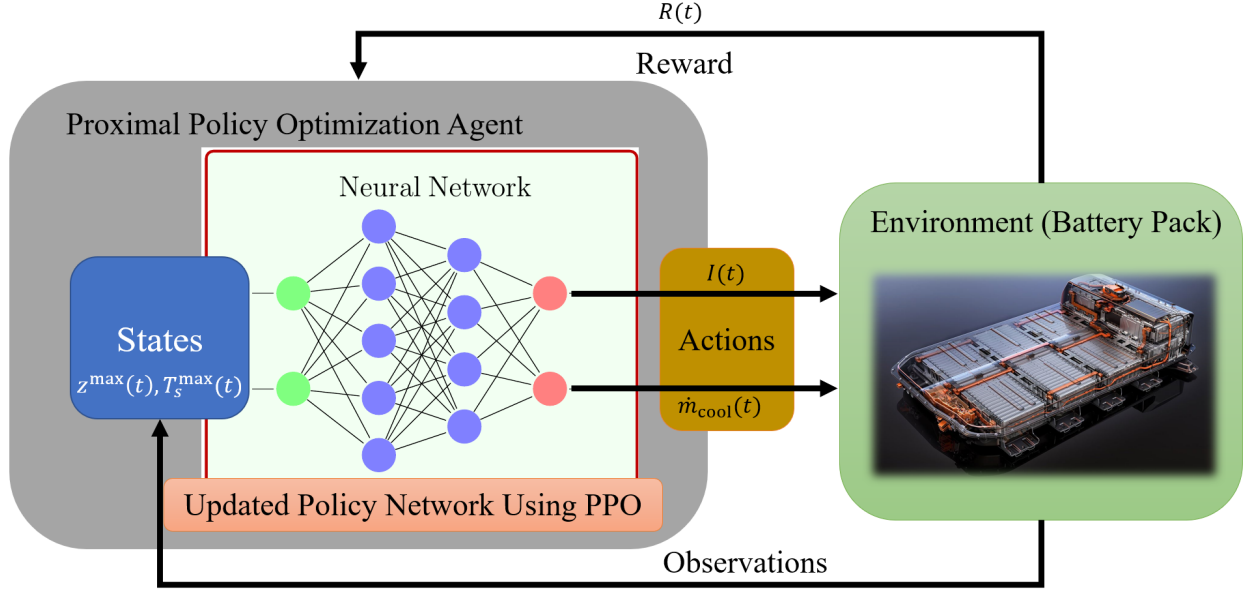


Figure 3: PPO Block Diagram

3.2.5. Proximal Policy Optimization

The optimal solution to this paper's fast charging and thermal management problems is achieved by finding the optimal policy π^* . In MDP, a policy $\pi : s \rightarrow a$ determines what actions should be taken when in state s . Finding the optimal policy is attained by training the underlying deep neural network. In this paper, the PPO algorithm is selected as the model-free RL strategy after examining some other RL methods, such as DQN and DDPG. The advantages of PPO include ease of implementation, efficient sampling, and straightforward tuning. PPO maximizes the objective function $L(\theta)$ in the following [35]

$$\theta_{k+1} = \arg \max_{\theta} \hat{\mathbb{E}}_t [L(\theta)], \quad (14)$$

$$L(\theta) = \min \left(r_t(\theta) \hat{A}_t, \text{clip} \left(r_t(\theta), 1 - \epsilon, 1 + \epsilon \right) \hat{A}_t \right), \quad (15)$$

$$r_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}, \quad (16)$$

where θ denotes the vector of policy parameters, $\hat{\mathbb{E}}_t[\cdot]$ indicates expectation, $r_t(\theta)$ shows the probability ratio between the updated policy outputs and those of the old policy, $\hat{A}_t$ represents the estimation of the advantage function at t , and ϵ is a hyperparameter used to generate a truncated version of $r_t(\theta)$ in $\text{clip}(\cdot)$ (for more detail see [35]).

4. Simulation Results and Analysis

This section presents the results of implementing the proposed approach and evaluates its efficiency. We consider a battery pack formed by the regular configuration of cylindrical A123 ANR26650M1 cells arranged in N_s series strings of N_p parallel cells. The pack is cooled off by ethylene-glycol/water 50–50% with the help of a bottom-cooling system. Although the selected cooling system is not as efficient as the current liquid cooling systems used in Tesla, our proposed RL approach to such a system achieved satisfactory cooling performance. On top of that, it is worth noting that bottom-cooling is used in many light-duty EVs, such as Audi e-tron and Ford Mach-E [36, 37]. Moreover, we assume the BMS performs near-perfect cell balancing, which results in a similar SOC level for all cells throughout a charging session. Also, the battery pack is equipped with at least two sensors, one voltmeter that measures the pack's terminal voltage and one temperature sensor that reads the hottest cell's temperature. In

Table 2: Coolant Temperature Impact on Core & Surface Temperatures While Mass Flow Rate Is Constant

Coolant Temp.*	$\dot{m}_{\text{cool}}$	T_c^{max}	T_c^{min}	ΔT_c	T_s^{max}	T_s^{min}	ΔT_s
N/A	N/A	44.00	42.93	1.07	40.41	39.02	1.39
25	0.1	42.39	41.55	0.84	37.91	36.83	1.08
20	0.1	41.57	40.78	0.79	36.87	35.87	1.00
15	0.1	40.75	40.04	0.71	35.83	34.92	0.91
10	0.1	39.93	39.31	0.62	34.80	33.97	0.83
5	0.1	39.12	38.57	0.55	33.76	33.02	0.74
0	0.1	38.32	37.83	0.49	32.74	32.08	0.66

* Temperatures are in [°C] and mass flow rates are in [kg/s]

other words, the optimization problem is controlled by the cell with the highest degradation. Since the degradation is directly affected by temperature, we assume the cell with the highest temperature faces the fastest capacity loss. Finally, the simulations are executed on a computer with a 3.00 GHz octa-core CPU and 16 GB RAM.

4.1. Sensitivity Study of Coolant Temperature

Our control strategy focuses on the mass flow rate of the coolant. We assume that the coolant temperature upon entering the battery pack is fixed. The said temperature is output by the chiller, a component within TMS connected to the battery pack. Note that the coolant flows from the chiller to the pack [17]. On the other hand, both the coolant temperature upon entering the pack and its mass flow rate impact the pack's temperature. The designed optimization problem controls the mass flow rate with a pump. However, a fixed coolant temperature should be selected.

In this regard, we investigate the effect of multiple coolant temperatures on the cell with maximum core and surface temperatures as shown in Table 2, where $\Delta T_c = T_c^{\text{max}} - T_c^{\text{min}}$ and $\Delta T_s = T_s^{\text{max}} - T_s^{\text{min}}$. Six candidate temperatures of 25, 20, 15, 10, 5, and 0 [°C] are examined. The results in Table 2 are obtained by charging a pack of 20 cells comprising four series strings of five parallel cells (4S5P) for 450 seconds. The cells are charged from 20% to nearly full charge with 6 C-rate constant current. Accordingly, the first row of Table 2 expresses the results without coolant. The subsequent rows, however, display the results when the coolant is utilized at different temperatures while its mass flow rate is kept at 0.1 [kg/s], which is the upper-bound of $\dot{m}_{\text{cool}}(t)$. Note that the values shown in the table are the final ones at the end of the charging process. Further, it is seen that lower coolant temperature yields more heat absorption from the cells. Besides, the differences between maximum and minimum temperatures, i.e., ΔT_c and ΔT_s , shrink as coolant temperature drops. However, the difference between each cell's core and surface temperatures has grown. Ultimately, we select the coolant temperature of 0°C as it yields the highest heat absorption. It is worth stating that choosing 0°C aligns with the findings of [17].

4.2. Hyperparameter Study of PPO

Tuning the PPO begins by selecting a small battery pack with 20 cells formed by four series strings of five parallel cells. The small pack enables us to run the simulations faster, which in turn expedites the tuning process. Besides, it paves the way for a fair comparison with MPC seeing that running MPC for a large pack is intractable. Once we demonstrate the outperformance of PPO over MPC, the best-tuned PPO will be used to charge a large pack.

Fig. 4 illustrates the results of charging the pack obtained by a well-tuned PPO. The underlying RL algorithm applies a deep neural network with two fully connected hidden layers with 64 neurons per layer, which is trained over 50,000 time steps with ten epochs. The utilized weights are $\omega_1 = 0.05$, $\omega_2 = 0.015$, $\omega_3 = 28.46$, $\omega_4 = 93.75$, $\omega_5 = \omega_6 = 0$, where it is observed that incorporating inputs into the reward function does not improve the results. The weights are selected through trial and error. The scale of the weights depends on the values in the reward function. For example, ω_1 and ω_2 are small values, as SOC and temperature are large numbers in the first two terms of the reward function. On the other hand, ω_3 and ω_4 are large numbers because the maximum capacity drop and maximum resistance rise are very small. Finally, $\omega_5 = \omega_6 = 0$ since considering inputs in the objective function does not change the results. One may also normalize the values of the six terms in the objective function and then select the weights such that the summation is equal to one, i.e., $\sum_i \omega_i = 1$. As can be seen, the reward converges to around -1. As

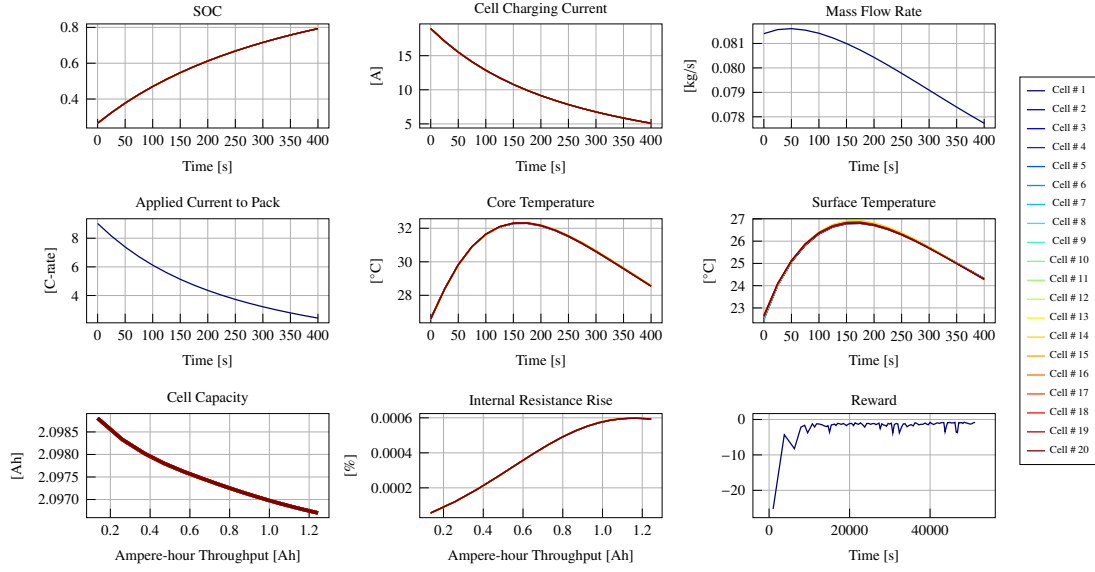


Figure 4: The results of optimizing fast charging and thermal management of a pack with 20 cells (4S5P) by PPO. The control inputs are cell charging current and mass flow rate.

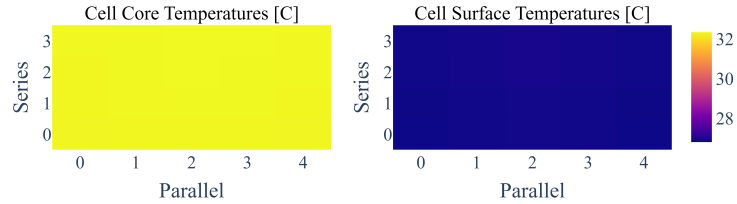


Figure 5: Heatmap of cell core and surface temperatures for a pack with 20 cells when charged by the PPO charging current displayed in Fig. 4. The pack is cooled by the mass flow rate shown in Fig. 4. The colorbar indicates temperature in $^{\circ}\text{C}$.

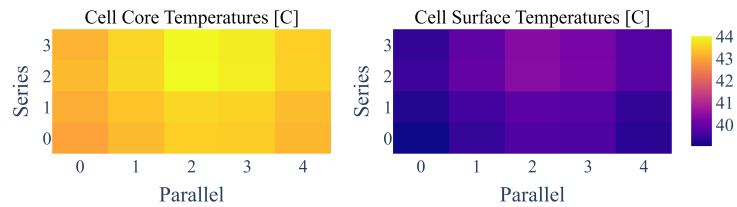


Figure 6: Heatmap of a pack with 20 cells (4S5P) when charged by a constant current at 6 C-rate without cooling. The colorbar indicates temperature in $^{\circ}\text{C}$.

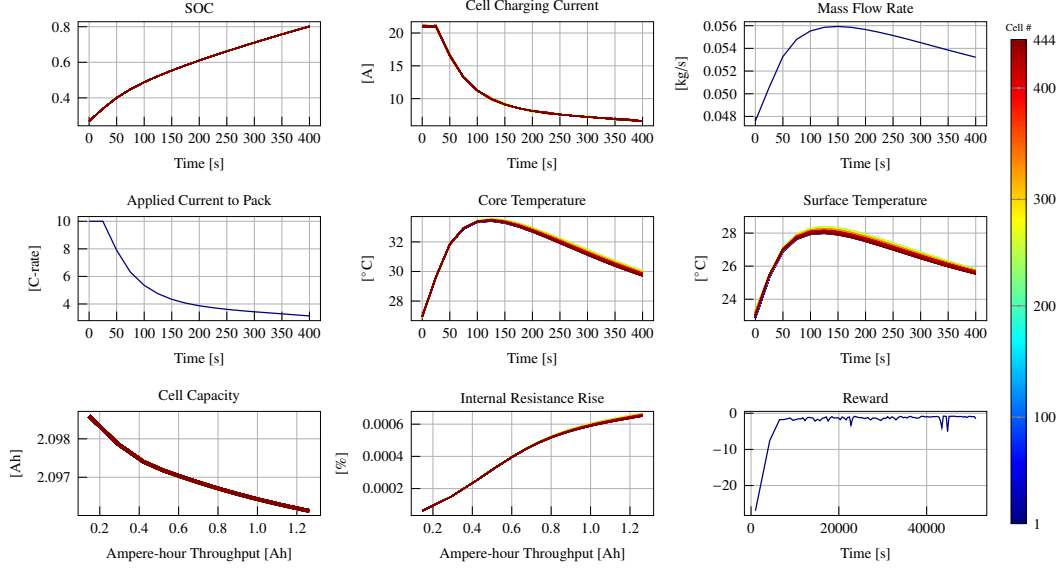


Figure 7: The results of optimizing fast charging and thermal management of a pack with 444 cells (6S74P) by PPO. The control inputs are cell charging current and mass flow rate. Each cell is shown by a color represented in the colorbar.

previously mentioned, the reward is always non-positive, i.e., $R(t) \leq 0$. However, based on the reward definition in (13), its cumulative value is always negative if the initial SOC is smaller than z^* . Furthermore, the 20 cells are charged, from 20% to 80%, in approximately 400 seconds. In addition, the cell core temperatures are elevated to just above $T_c^{\text{thr}} = 32^\circ\text{C}$, while surface temperatures rise to near 27°C . On the other hand, both the charging current and the mass flow rate have decreasing trends. The charging C-rate starts at 9.02C and ends at 2.45C.

Fig. 5 depicts the heatmap of the cells' core and surface temperatures, where a colored square presents each cell with dark squares showing cooler cells as opposed to light colors that stand for warmer cells. Additionally, each heatmap can be considered a bird-eye view of the pack where parallel and series cells are aligned along the horizontal and vertical axes, respectively. Moreover, Fig. 6 is plotted as a reference of comparison with Fig. 5. In Fig. 6, the pack is charged at 6C from 20% to 80% without a cooling mechanism. The 6C current has been opted because it is roughly the average charging current in Fig. 4. It can be seen that most core and surface temperatures exceed 40°C with higher temperature gradients in Fig. 6. In contrast, temperature differences in heatmaps of Fig. 5 are negligible, where the maximum difference of core temperatures is 0.05°C and that of surface temperatures is 0.05°C .

4.3. Cell to Pack Level Study of PPO

In Section 4.2, we demonstrated the advantages of PPO over MPC for fast charging a battery pack of 20 cells. As such, in this section, we use the same deep neural network architecture to be trained with PPO to control fast charging and thermal management of a module comprising 444 cells arranged in six series strings of 74 parallel cells (6S74P). It is noteworthy that this module is the building block of Tesla's 85 kWh battery pack [8]. The PPO training takes 94.95 minutes to charge the pack from 20% to 80%. In addition, the evaluation of the trained PPO requires ~ 3 seconds to compute the applied current C-rate and the mass flow rate of the cooling fluid.

Fig. 7 illustrates the results of charging the pack with 444 cells (6S74P). Judging by the reward subplot and comparing it with that of the 20-cell scenario in Fig. 4, the convergence speed during the training process is close in both cases, with the accumulative reward tending to around -1 . Analogous to charging 20 cells, the charging current and the mass flow rate have decreasing trends. Moreover, the charging process finishes in 400 seconds, identical to the results in Fig. 4. In terms of cell temperatures, core temperatures peak at just above 32°C with temperature gradients of 0.33°C at the end of charge, while surface temperatures peak at around 28°C with gradients of 0.28°C at the end of the charging session. Higher temperature gradients in Fig. 7 compared to those in Fig. 4 are due to the larger number of cells surrounding the center of the module that increases temperature differences across the pack owing to the cell-to-cell heat transfer.

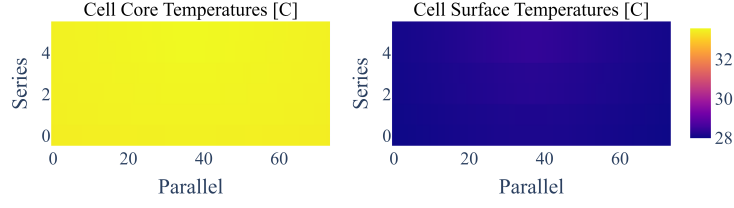


Figure 8: Heatmap of cell core and surface temperatures for a pack with 444 cells when charged by the PPO charging current displayed in Fig. 7. The pack is cooled by the mass flow rate shown in Fig. 7. The colorbar indicates temperature in $^{\circ}\text{C}$.

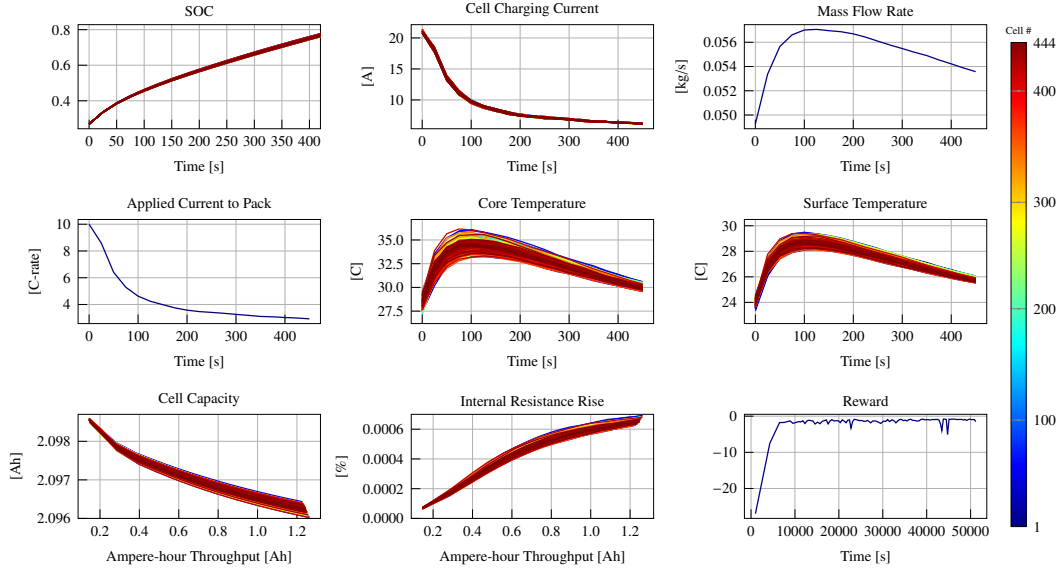


Figure 9: The results of optimizing fast charging and thermal management of a pack with 444 cells (6S74P) by PPO when the internal resistance has disturbance.

Finally, Fig. 8 depicts the heatmap of the cell core and surface temperatures from a top-down perspective. The figure shows that the central cells experience the most temperature rise, which verifies our selection of the middle top cell's surface temperature as one of the RL states discussed in Section 3.2.2. Note that the bottom-cooling system influences the outcomes in Fig. 8.

Lastly, since the working environment of battery packs in actual situations can be more complex than the model in this paper or other references, Gaussian noise is added to the internal resistance of each cell in the battery pack to examine the performance of the proposed algorithm. The results are plotted in Fig. 9. Because the noise randomly increases the internal resistances, the cell surface and core temperatures are slightly higher than the results in Fig. 7. To prevent further increase in temperatures, the coolant mass flow rate increases while the current drops faster (see Fig. 9), resulting in an increase of about one minute in the charging time. Consequently, the proposed methodology responded to the disturbance in internal resistance by increasing the coolant mass flow rate and reducing the charging current.

4.4. Comparison to Baseline Methods

To investigate the efficacy of MPC, it is used to charge 20 cells, and the corresponding inputs and states are plotted in Fig. 10. It takes 86.16 minutes to run the underlying MPC and obtain the results in Fig. 10. On the other hand, as opposed to RL, MPC leads to higher core and surface temperature gradients at the end of charging. In Fig. 10, cells' maximum core and surface temperature gradients at the end of charging with MPC are 0.73°C and 0.72°C , respectively. However, according to Fig. 4, the highest cells' core and surface temperature differences at the end

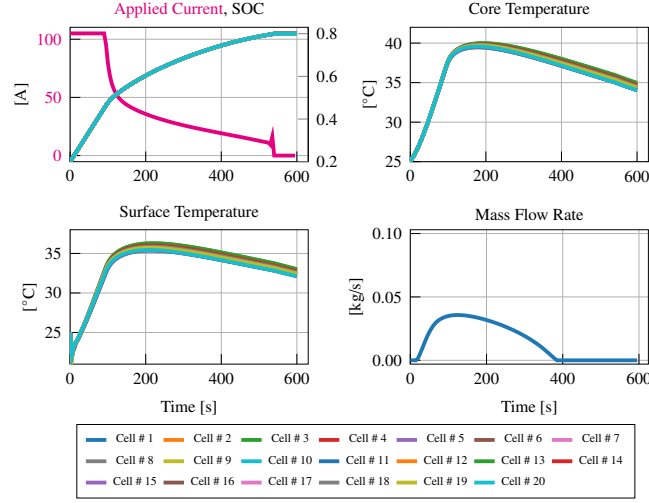


Figure 10: The results of charging a pack of 20 cells with the optimal charging current and cooling mass flow rate obtained by MPC.

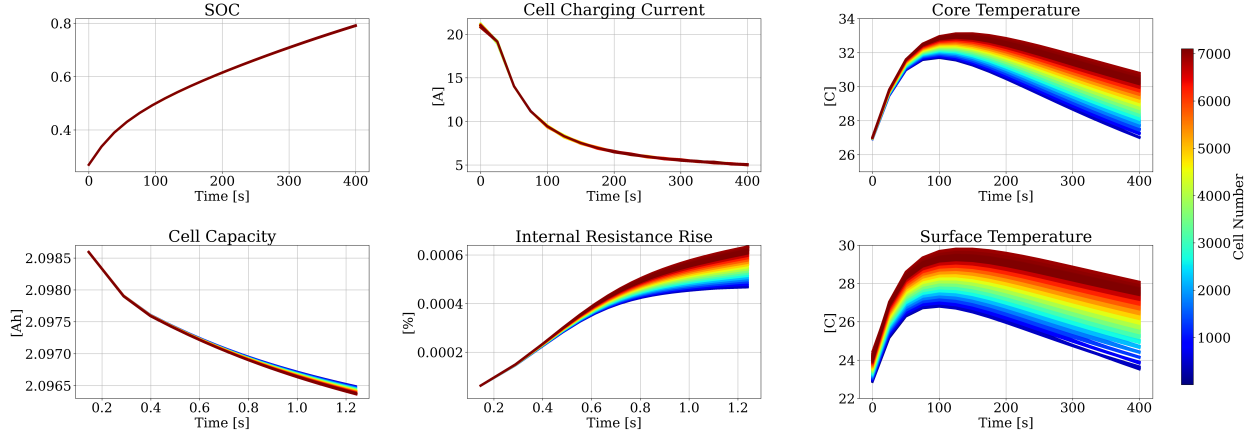


Figure 11: The results of optimizing fast charging and thermal management of a pack with 7104 cells by PPO (16 series modules of 6S74P cells or a pack of 96S74P cells). Each cell is shown by a color represented in the colorbar.

of charging with RL are 0.05°C and 0.05°C , respectively. Consequently, we conclude that PPO can achieve better outcomes compared to MPC.

Subsequently, the tuned PPO in Section 4.3 is used to charge a battery pack comprising 16 series-connected modules. Each module consists of six series strings of 74 parallel cells. In other words, the pack can be considered a large 96S74P module consisting of 7104 cells, which is identical to Tesla's 85 kWh pack [8].

As discussed in Section 2, the developed RL approach finds the charging current and mass flow rate applied to the pack. Since the 16 modules are wired in series, the same current flows through each module. Additionally, the PPO algorithm requires the surface temperature of the hottest cell in the pack. Therefore, the hottest cell in the hottest module should be used to train the PPO. Assuming perfect cell balancing is available, we only need the information from the hottest module to train the PPO. Hence, training PPO to charge the pack is similar to training it to optimize the fast charging of the hottest module, except that in the case of the pack, we are dealing with higher temperatures. The results in Fig. 11 and Fig. 12 are obtained after training the PPO utilizing the hottest module in the pack, like what we did in Section 4.3.

Fig. 11 represents the outcome of charging a battery pack with 16 modules of 6S74P cells, i.e., the pack with 7104 cells. As seen, the pack is charged from 20% to 80% in 400 seconds. Compared to Fig. 7, the core and

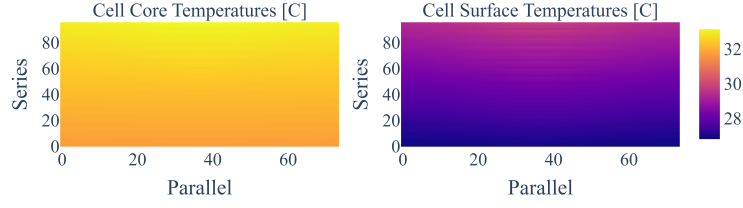


Figure 12: Heatmap of cell core and surface temperatures for a pack with 7104 cells when charged by the PPO charging current displayed in Fig. 11. The colorbar indicates temperature in [°C].

Table 3: The Maximum And Minimum Values in Fig.11

	Surface Temp.	Core Temp.	Capacity	R_0 Rise
Min. End of Charge	23.22°C	26.62°C	2.0977 Ah	4.672e-4%
Max. End of Charge	27.90°C	30.55°C	2.0979 Ah	6.356e-4%
Peak Maximum	29.82°C	33.13°C	N/A	N/A

surface temperature gradients have increased because of the rise in coolant temperature as it flows through the cells. The cooling fluid enters the pack from one side and exits from the other, moving along the 16 modules. Therefore, as Fig. 12 displays, the temperatures increase along the series axis, while the middle part is warmer in each pack. Accordingly, surface temperatures undergo larger temperature differences on account of direct contact with the cooling plate. On the other hand, although core temperatures face smaller temperature gradients, the average core temperature is greater than the average surface temperature. In addition, the maximum difference in core and surface temperatures at the end of charge are 3.93°C and 4.68°C, respectively. Lastly, Table 3 reports the maximum and minimum values at the peak and at the end of the charging session.

Next, the performance of the proposed PPO is compared with that of CC-CV, seeing that CC-CV is widely adopted in practice. Thereby, two cases are investigated: the case with a 6S74P module consisting of 444 cells and the one where a pack with 7104 cells is charged. In both cases, the battery is charged from 20% to 80% three times, where a CC-CV with 6C-rate is tested at the CC stage in the first case, then a CC-CV with a 4C-rate at the CC stage is used for the second case, and finally, a CC-CV with 5C-rate is utilized in the third case. Moreover, the mass flow rate of the cooling fluid is increased, in a step-wise fashion, as cell temperatures rise [28].

Firstly, Table 4 compares the computation time of each approach, where MPC is the slowest while PPO and CC-CV require almost the same run time. PPO and CC-CV are much faster than MPC for single cells, thus their results for a single cell are not included in Table 4. In addition, Fig. 13 demonstrates core and surface temperature evolution when a module of 444 cells is charged through PPO and the three abovementioned CC-CVs. Each color represents one of the cases where the solid lines illustrate the average cell temperature, while the dashed lines show the actual maximum and minimum temperatures from different cells. As seen, the proposed approach charges the module quicker than CC-CV with 4C while keeping the peak cell temperature below that of CC-CV with a 6C and a 5C constant current. Similarly, Fig. 14 depicts the results where a pack with 7104 cells is charged. The results follow the same trend as in Fig. 13 where PPO finishes charging faster than CC-CV with 4C while maintaining the peak cell temperature smaller than peak temperatures when the pack is charged with CC-CV with a 6C or a 5C. On the other hand, in Fig. 14, the cell temperature differences in PPO case is greater than those of CC-CV cases. Nonetheless, the peak temperatures in the PPO case are still smaller than those of the CC-CV 5C case.

Lastly, Fig. 15 shows the pack's capacity fade over 1000 charge/discharge cycles when charged with CC-CV 5C and PPO, respectively. Through 1000 cycles, the pack has outputted roughly 1270 [Ah] energy, leading to 6.47% and

Table 4: Simulation Time Comparison

Approach	1 Cell	20 Cells	444 Cells	7104 Cells
MPC	56 [s]	86.16 [min]	N/A	N/A
PPO	N/A	0.23 [s]	3.39 [s]	77.79 [s]
CC-CV	N/A	0.22 [s]	4.10 [s]	71.27 [s]

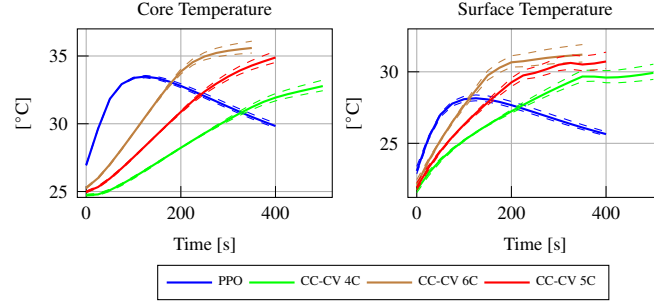


Figure 13: The fast charging and thermal management results when a module of 444 cells is charged by PPO versus CC-CV

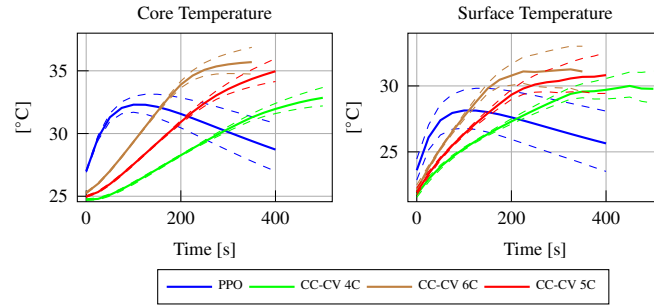


Figure 14: The fast charging and thermal management results when a pack of 7104 cells is charged by PPO versus CC-CV

5.87% capacity fade for CC-CV 5C and PPO, respectively. Note that the 0.60% difference in capacity loss between the two approaches accounts for approximately 100 charge cycles. Assuming the battery is not slow charged and only fast charged once a week, 100 cycles extends the battery's life by nearly two years.

5. Conclusion

This paper conducts a simulation study on fast charging and thermal management of a battery pack through RL. Each battery pack cell is modeled using an electro-thermal-aging model, which considers thermal coupling among adjacent cells. The RL agent is trained to charge the battery as fast as possible while minimizing battery aging in terms of capacity fade and internal resistance rise. In addition, MPC and CC-CV approaches are employed as baselines, and their results and computation times are compared with those of RL. In this regard, a small module comprising 20 cells is utilized to demonstrate the outperformance of RL over MPC. The trained RL's simulation time is faster than that of MPC, and it also achieves lower battery temperatures. On the other hand, MPC becomes too slow for larger battery module sizes. Next, the RL is tuned for a module with 444 cells, and the same tuned hyperparameters are used to charge a pack with 7104 cells. The module with 444 cells is merely used to expedite the tuning stage. Moreover, the

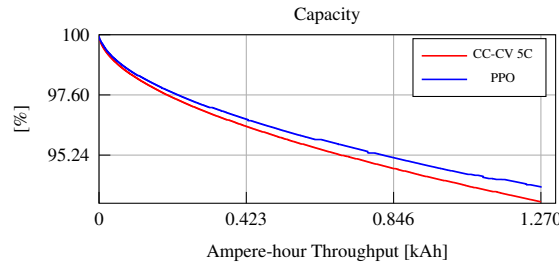


Figure 15: Capacity fade after 1000 charge cycles

module with 444 cells is also exploited to verify that the proposed approach works well under the complex working environment of battery packs. The case study for the pack with 7104 cells shows that the average charging current is around 5 to 6 C-rate. Therefore, the 444-cell module and 7104-cell pack are charged with three CC-CV algorithms with 4C, 5C, and 6C constant current stages, respectively. The results reveal that RL charges the battery pack as fast as CC-CV 5C while the cell's core and surface temperatures remain cooler than the temperatures under CC-CV 5C or 6C. Thus, the pack's degradation in terms of capacity loss is slower in the case of RL compared to the CC-CV 5C and 6C cases, and the battery pack's cycle life is extended by nearly two years over 1000 charge cycles.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-2312196.

References

- [1] H. Sun, F. Tao, Z. Fu, A. Gao, and L. Jiao, "Driving-behavior-aware optimal energy management strategy for multi-source fuel cell hybrid electric vehicles based on adaptive soft deep-reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [2] Z. Fan, Y. Fu, H. Liang, R. Gao, and S. Liu, "A module-level charging optimization method of lithium-ion battery considering temperature gradient effect of liquid cooling and charging time," *Energy*, vol. 265, p. 126331, 2023.
- [3] M. H. Abbasi, J. Zhang, and V. Krovi, "A lyapunov optimization approach to the quality of service for electric vehicle fast charging stations," in *2022 IEEE Vehicle Power and Propulsion Conference (VPPC)*. IEEE, 2022, pp. 1–6.
- [4] J. Sieg, A. U. Schmid, L. Rau, A. Gesterkamp, M. Storch, B. Spier, K. P. Birke, and D. U. Sauer, "Fast-charging capability of lithium-ion cells: Influence of electrode aging and electrolyte consumption," *Applied Energy*, vol. 305, p. 117747, 2022.
- [5] J. Kleiner, R. Singh, M. Schmid, L. Komsijska, G. Elger, and C. Endisch, "Influence of heat pipe assisted terminal cooling on the thermal behavior of a large prismatic lithium-ion cell during fast charging in electric vehicles," *Applied Thermal Engineering*, vol. 188, p. 116328, 2021.
- [6] Z. Fan, R. Gao, and S. Liu, "Thermal conductivity enhancement and thermal saturation elimination designs of battery thermal management system for phase change materials based on triply periodic minimal surface," *Energy*, vol. 259, p. 125091, 2022.
- [7] W. Xie, X. Liu, R. He, Y. Li, X. Gao, X. Li, Z. Peng, S. Feng, X. Feng, and S. Yang, "Challenges and opportunities toward fast-charging of lithium-ion batteries," *Journal of Energy Storage*, vol. 32, p. 101837, 2020.
- [8] M. H. Abbasi and J. Zhang, "Joint optimization of electric vehicle fast charging and dc fast charging station," in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2021, pp. 1–6.
- [9] K. Liu, K. Li, H. Ma, J. Zhang, and Q. Peng, "Multi-objective optimization of charging patterns for lithium-ion battery management," *Energy Conversion and Management*, vol. 159, pp. 151–162, 2018.
- [10] B. Jiang, M. D. Berliner, K. Lai, P. A. Asinger, H. Zhao, P. K. Herring, M. Z. Bazant, and R. D. Braatz, "Fast charging design for lithium-ion batteries via bayesian optimization," *Applied Energy*, vol. 307, p. 118244, 2022.
- [11] K. Liu, X. Hu, Z. Yang, Y. Xie, and S. Feng, "Lithium-ion battery charging management considering economic costs of electrical energy loss and battery degradation," *Energy conversion and management*, vol. 195, pp. 167–179, 2019.
- [12] H. E. Perez, X. Hu, S. Dey, and S. J. Moura, "Optimal charging of li-ion batteries with coupled electro-thermal-aging dynamics," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7761–7770, 2017.
- [13] Z. Wei, Z. Quan, J. Wu, Y. Li, J. Pou, and H. Zhong, "Deep deterministic policy gradient-drl enabled multiphysics-constrained fast charging of lithium-ion battery," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 3, pp. 2588–2598, 2021.
- [14] S. Su, W. Li, A. Garg, and L. Gao, "An adaptive boosting charging strategy optimization based on thermoelectric-aging model, surrogates and multi-objective optimization," *Applied Energy*, vol. 312, p. 118795, 2022.
- [15] S. Park, A. Pozzi, M. Whitmeyer, H. Perez, A. Kandel, G. Kim, Y. Choi, W. T. Joe, D. M. Raimondo, and S. Moura, "A deep reinforcement learning framework for fast charging of li-ion batteries," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 2770–2784, 2022.
- [16] A. Pozzi and D. M. Raimondo, "Stochastic model predictive control for optimal charging of electric vehicles battery packs," *Journal of Energy Storage*, vol. 55, p. 105332, 2022.
- [17] B. Xu and Z. Arjmandzadeh, "Parametric study on thermal management system for the range of full (tesla model s)/compact-size (tesla model 3) electric vehicles," *Energy Conversion and Management*, vol. 278, p. 116753, 2023.
- [18] X. Tan, P. Lyu, Y. Fan, J. Rao, and K. Ouyang, "Numerical investigation of the direct liquid cooling of a fast-charging lithium-ion battery pack in hydrofluoroether," *Applied Thermal Engineering*, vol. 196, p. 117279, 2021.
- [19] S. Chen, N. Bao, A. Garg, X. Peng, and L. Gao, "A fast charging-cooling coupled scheduling method for a liquid cooling-based thermal management system for lithium-ion batteries," *Engineering*, vol. 7, no. 8, pp. 1165–1176, 2021.
- [20] S. Park, A. Pozzi, M. Whitmeyer, H. Perez, W. T. Joe, D. M. Raimondo, and S. Moura, "Reinforcement learning-based fast charging control strategy for li-ion batteries," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2020, pp. 100–107.
- [21] Z. Wei, X. Yang, Y. Li, H. He, W. Li, and D. U. Sauer, "Machine learning-based fast charging of lithium-ion battery by perceiving and regulating internal microscopic states," *Energy Storage Materials*, 2023.
- [22] Y. Hao, Q. Lu, X. Wang, and B. Jiang, "Adaptive model-based reinforcement learning for fast charging optimization of lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, 2023.

- [23] 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*, 2023.
- [24] X. Yang, H. He, Z. Wei, R. Wang, K. Xu, and D. Zhang, "Enabling safety-enhanced fast charging of electric vehicles via soft actor critic-lagrange drl algorithm in a cyber-physical system," *Applied Energy*, vol. 329, p. 120272, 2023.
- [25] K. Wang, H. Wang, Z. Yang, J. Feng, Y. Li, J. Yang, and Z. Chen, "A transfer learning method for electric vehicles charging strategy based on deep reinforcement learning," *Applied Energy*, vol. 343, p. 121186, 2023.
- [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] A. Allam, S. Onori, S. Marelli, and C. Taborelli, "Battery health management system for automotive applications: A retroactivity-based aging propagation study," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 703–716.
- [28] "Battery pack thermal management," <https://www.mathworks.com/help/sps/ug/lithium-pack-cooling.html>, 2022.
- [29] J. Wang, P. Liu, J. Hicks-Garner, E. Sherman, S. Soukiazian, M. Verbrugge, H. Tataria, J. Musser, and P. Finamore, "Cycle-life model for graphite-lifepo4 cells," *Journal of power sources*, vol. 196, no. 8, pp. 3942–3948, 2011.
- [30] A. Cordoba-Arenas, S. Onori, Y. Guezennec, and G. Rizzoni, "Capacity and power fade cycle-life model for plug-in hybrid electric vehicle lithium-ion battery cells containing blended spinel and layered-oxide positive electrodes," *Journal of Power Sources*, vol. 278, pp. 473–483, 2015.
- [31] C. Fang, H. Lu, Y. Hong, S. Liu, and J. Chang, "Dynamic pricing for electric vehicle extreme fast charging," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 531–541, 2020.
- [32] H. Moghaddam and S. Mazzyar, "Designing battery thermal management systems (btms) for cylindrical lithium-ion battery modules using cfd," *KTH School of Industrial Engineering and Management: Stockholm, Sweden*, 2019.
- [33] K. Yang, M. Grill, and M. Bargende, "Simulative investigation of the influence of a rankine cycle based waste heat utilization system on fuel consumption and emissions for heavy duty utility vehicles," in *Energy and Thermal Management, Air-Conditioning, and Waste Heat Utilization: 2nd ETA Conference, November 22-23, 2018, Berlin, Germany 2*. Springer, 2019, pp. 172–193.
- [34] S. Ahmed, I. Bloom, A. N. Jansen, T. Tanim, E. J. Dufek, A. Pesaran, A. Burnham, R. B. Carlson, F. Dias, K. Hardy *et al.*, "Enabling fast charging—a battery technology gap assessment," *Journal of Power Sources*, vol. 367, pp. 250–262, 2017.
- [35] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [36] "Battery and thermal management," <https://www.audi-mediacycenter.com/en/emotive-design-and-revolutionary-technologythe-audi-e-tron-gt-quattro-and-the-audi-rs-e-tron-gt-13655/battery-and-thermal-management-13784>, 2021.
- [37] "Mach-e: Battery tray and battery cell features," <https://www.youtube.com/watch?v=4vRjo0gaG1g>, 2022.