Preprints, 10th IFAC International Symposium on
Advances in Automotive Control
Columbus, Ohio, USA, August 28-30, 2022

# Reinforcement learning based EV energy management for integrated traction and cabin thermal management considering battery aging

**Ibrahim Haskara, Bharatkumar Hegde, Chen-Fang Chang**

*General Motors Company, Detroit, MI
USA (Tel: 586-291-6838; e-mail: ibrahim.haskara@gm.com).*

Abstract: Energy management in electric vehicles plays a significant role in both reducing energy consumption and limiting the rate of battery capacity degradation. The work summarized in this paper explores machine-learning techniques for electrified propulsion control in designing energy management (EM) controllers. The role of the EM is to coordinate delivery of multiple power requests from a modular battery of an electric vehicle (EV) to improve range and battery longevity. Reinforcement learning is adopted for integrated EV traction and HVAC controls. The EM acts as a supervisory controller augmenting the HVAC controls. It is designed to adjust internal HVAC control parameters based on current drive parameters to improve energy efficiency and battery state of health (SoH) without affecting driver demand and cabin comfort. An empirical battery aging model is incorporated into the problem formulation to address long-term battery capacity degradation. Reduced energy consumption and battery aging are demonstrated.

*Keywords:* Electric vehicles, Energy management, Health-aware battery management, Artificial intelligence, Machine learning, Reinforcement learning

## 1. INTRODUCTION

This paper concerns with developing learning-based strategies for electric vehicle (EV) energy management. Effective energy management strategy (EMS) is a key enabler to improve EV range and long-term battery state of health given a particular hardware configuration. Energy management can specifically address the capacity fading of battery and energy loss during operation by optimizing operating points for various energy demands.

In general, EV operation involves multiple loads or “energy consumers”, like traction, HVAC, battery thermals etc. that draw energy from battery. Simultaneous high-power consumption events reduce efficiency and accelerate battery aging. Traction request interpreted from driver often needs to be met without any modification, however, the other loads can be adjusted owing to their different time scales and dynamic responses without compromising their individual objectives. The focus of this paper is to develop a dynamic, real-time, battery-health-conscious load shaping mechanism and demonstrate it on the integrated traction and HVAC controls considering the traction and HVAC as the two major loads on the battery.

We have built on our previous work where we utilized machine learning (ML) techniques; in particular, reinforcement learning (RL), in constructing these supervisory operating strategies. In (Haskara et al. 2021), we developed RL-based strategies for optimal operation of a hybrid energy storage system (HESS), that include a super-capacitor (SC) and a battery. The main objective of that energy management strategy (EMS) is to determine optimal power-split strategy which, for a given total traction power request, distributes this request between battery and SC in both drive and regen modes so that the overall energy efficiency and battery aging are improved. Through this optimized operating strategy, battery throughput and capacity loss have been reduced significantly through the entire vehicle and battery life. RL has been demonstrated to produce comparable results to that of using dynamic programming in hybrid vehicle applications (Lee et al. 2020). Deep Q-learning (Li et al. 2021) and double deep Q-learning (Han et al. 2019) have been successfully used for electrified vehicle EM.

In this work, we consider a modular battery as the single power source. However, we include a secondary power demand in the form of a closed-loop controlled HVAC system together with the main traction demand. Cabin comfort is included as an additional performance variable, which is quantified through cabin temperature response. Supervisory control to minimize total energy usage through HVAC temperature set-point variation has been used in conjunction with model predictive control (Wang et al. 2018). Similar methodology has been shown to reduce battery aging by 13% (Vatanparvar et al. 2018). Dynamic programming has been used to regulate cabin temperature while also minimizing total energy use (Sakhdari et al. 2015).

In this paper, RL is used to coordinate the HVAC operation with the traction demand. The primary reason for this choice is that RL offers a data-driven optimization scheme without requiring analytical representations of the dependence of HVAC power consumption with the traction states. Specifically, RL-based EM controller is designed as an additional optimization wrapper to improve the system-level energy efficiency instead of replacing the baseline HVAC controller, which could still employ a model-based structure.

The remainder of the paper is organized as follows: Section 2 provides the scope of integrated traction and HVAC control, the summary of the developed models as well as the proposed reinforcement learning formulation. Section 3 presents the development process of the proposed designs with their performance on energy consumption and battery aging. Finally, Section 4 provides conclusions regarding the general applicability of the methods for other applications.

## 2. METHODOLGY

The proposed EM strategy is demonstrated on integrated EV traction and HVAC controls.

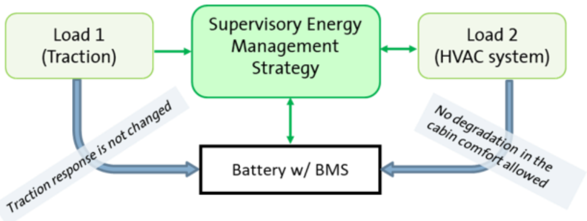


*Figure 2. Integrated EV traction and HVAC control scope*

Fig. 2 depicts power requests from the traction system (e.g., vehicle drive and regen) and the HVAC system (e.g., total power drawn by the HVAC actuators like blower motor, chiller etc.), which are controlled independently in default EV operation. As shown in Fig. 2, “the supervisory energy management strategy” is an add-on component that monitors instantaneous traction states, like vehicle speed and acceleration. It is tasked to modify a reference variable within the HVAC control system to improve the performance in terms of long-term battery aging. This is to be achieved without any change in traction response by design and no perceivable degradation in the driver’s desired cabin comfort.

The proposed design includes formulation of EMS objectives in the reinforcement learning framework followed by a training and validation exercise to calibrate those strategies via data-driven structures. Training and validation steps use a system-level dynamic EV energy flow model and large-scale real-world drive profiles in a high-performance computing (HPC) environment.

### *2.1 Model development for energy management*

The vehicle model and simulation architecture (Fig. 3) represents the power flows during the EV operation. Starting from a given driver speed/acceleration request; it dynamically generates the mechanical traction power using vehicle longitudinal dynamics and the corresponding electrical dynamics using an EV drive (motor and inverter) model. This power demand is then converted to a battery power demand and other battery states like current, voltage, state-of-charge (SoC) and temperature. An empirical battery capacity model is used to capture battery aging impact of the EV operation. A model of HVAC system is also included to determine the instantaneous HVAC power request from the battery associated with the HVAC operation. HVAC system model includes a cabin model that keeps track of the cabin heat flow, cabin temperature and corresponding power drawn from battery as well as a closed-loop control system to maintain a desired cabin temperature.

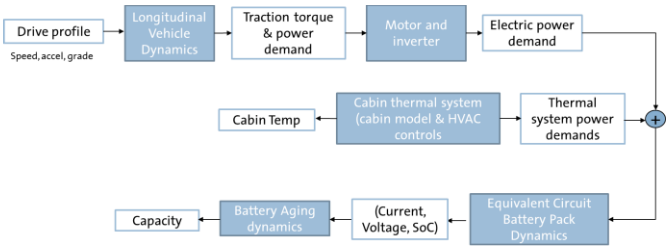

Drive profile
Speed,accel,grade
Longitudinal Vehicle Dynamics
Traction torque & power demand
Motor and inverter
Electric power demand
Cabin Temp
Cabin thermal system (cabin model & HVAC controls
Thermal system power demands
+
Capacity
Battery Aging dynamics
(Current, Voltage, SoC)
Equivalent Circuit Battery Pack Dynamics

*Figure 3. Electric Vehicle Model and Simulation Architecture*

The vehicle is modelled as point mass with longitudinal dynamics. Equation (1) describes the power demand $P_{traction}$ of the vehicle while travelling at velocity $V_{veh}$ and acceleration $Acc_{veh}$ on a road segment with grade $\theta_{grade}$.

$$P_{traction} = V_{veh}.\Big(M_{total}Acc_{veh} + M_{curb}\; g\; sin\big(\theta_{grade}\big) + M_{curb}C_{rr}\; g\; V_{veh} + \tfrac{1}{2}\rho_{air}V_{veh}^2\; C_d\; A_f\Big) \qquad (1)$$

The parameters of the model include the curb weight and equivalent inertia $M_{curb}$ and $M_{total}$ respectively; rolling resistance $C_{rr}$; air density $\rho_{air}$; aerodynamic drag coefficient of $C_d$; and effective frontal area $A_f$.

The total power demand from the battery, $P_{batt}$, is the sum of traction power $P_{EM} = \eta_{EM}\, P_{traction}$ and miscellaneous accessory power consumptions including a constant nominal power for instrument clusters, HVAC, battery thermal control etc. The battery is modelled with three primary states: state of charge (SOC); temperature ($T_{batt}$); and capacity degradation ($Q_{loss}$).

$$I_{batt} = \frac{V_{oc} - \sqrt{V_{oc}^2 - 4\, R_{int}\, P_{batt}\, n_{batt}}}{2\, R_{int}\, n_{batt}} \qquad (2)$$

$$SOC(k+1) = SOC(k) - \frac{1}{Q_{batt}} \frac{n_{batt}\, I_{batt}}{3600}\, \Delta t \qquad (3)$$

The electrical dynamics of the battery is modelled as a zeroth order equivalent circuit using (2-3). The battery current ($I_{batt}$) is a function of the open circuit voltage of the battery pack ($V_{OC}$), internal resistance ($R_{int}$), and battery efficiency ($\eta_{batt}$). SOC dynamics using (3) models the energy content of the battery based on the current drawn from battery and the capacity of the battery $Q_{batt}$. Further, the parameters of the battery such as internal resistance, efficiency and open circuit voltage depend on temperature and SOC of the battery.

Temperature dynamics of the battery is modelled as a lumped system. The electrical energy loss at the battery heats up the battery.

$$Q_{loss}(\%) = a_c(.)\, exp\left(-\frac{E_{ac}}{R_g T_{batt}}\right). Ah^n \qquad (4)$$

The capacity degradation of the battery is represented with a heuristic model reported by Cordoba et. al. (Cordoba-Arenas 2015) using (4). The model incorporates capacity degradation effects of using the battery at various temperatures, electrical throughput of the battery ($Ah$), and power drawn from the battery. The capacity loss $Q_{loss}$ is expressed as a percentage degradation from the nominal battery capacity at the beginning of the battery life.

In Equation (4), $n$ is a calibration constant, $a_c$ captures effects of battery temperature, depth of charge/discharge, and battery current, $E_{ac}$ is the cell activation energy, $R_g$ is the gas constant. These parameters require a large dataset of battery experiments to calibrate and fully validate. In this work, we use the parameter set from the original paper for lithium-ion batteries, which is deemed sufficient to compare different strategies.

$$\Delta Q_{loss} = \frac{\partial Q_{loss}}{\partial Ah}.\Delta Ah = n.a_c(.)\,exp\left(-\frac{E_{ac}}{R_g T_{batt}}\right).Ah^{n-1}.\Delta Ah \quad (5)$$

We further define an instantaneous battery capacity loss rate, $\Delta Q_{loss}$, using (5). The $\Delta Q_{loss}$ term plays a significant role in defining the reward function for reinforcement learning and helps weigh the impact of control actions on battery aging.

The cabin temperature dynamics are modelled as a lumped 2-state system. The states, cabin temperature ($T_{cabin}$) and panels or body temperature ($T_{panel}$), exchange heat with each other and the ambient air. The heat transfer coefficients between panel and ambient air is $k_{amb}$ and it increases with vehicle speed. The cabin receives energy from solar irradiation which is modeled as a heat input $Q_{solar}$. Cabin HVAC controller relies on airflow rate to cabin and the temperature of the air being delivered to the cabin to regulate cabin air temperature. In this work, the HVAC controller is simplified to a cabin heat input $Q_{HVAC}$ generation, which combines the two control levers, i.e., the flow rate ($\dot{m}_b$) and the temperature of the supplied air into the cabin ($T_s$) by (6)

$$Q_{HVAC} = \dot{m}_b c_p (T_{cabin} - T_s) \quad (6)$$

$$P_{HVAC} = \eta_{HVAC}(Q_{HVAC}, V_{veh}) Q_{HVAC} \quad (7)$$

The cabin heat input is modulated in closed-loop to maintain cabin temperature around its desired set-point. The corresponding flow rate and the flow temperature are then controlled by actuator-level controls. Cabin heat flow determines the instantaneous power drawn ($P_{HVAC}$) from battery. Efficiency of delivering heat to cabin is represented by $\eta_{HVAC}$ dependent on cabin heat flow and vehicle speed.

### 2.2 Reinforcement Learning formulation

Fig. 4 shows the general architecture of the RL-based EMS architecture on the integrated EV traction and HVAC controls. The EMS can receive and utilize a variety of signals including driver demand and current traction states (e.g., vehicle speed, accel, traction demand etc.), cabin comfort settings (e.g., cabin temperature set-point) and any available preview information (e.g., road grade). It has also access to real-time signals from the battery and cabin that are used to quantify a real-time performance reward for the reinforcement learning controls. The output of the EMS is a design variable to be selected as a control knob for HVAC power modulation.

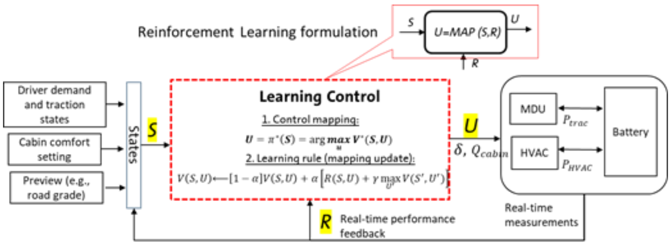

Reinforcement Learning formulation
S
U=MAP (S,R)
U
R
Driver demand and traction states
Cabin comfort setting
Preview (e.g., road grade)
States
S
Learning Control
1. Control mapping:
$U = \pi^*(S) = \arg\max_u V^*(S,U)$
2. Learning rule (mapping update):
$V(S,U) \longleftarrow [1-\alpha]V(S,U) + \alpha\left[R(S,U) + \gamma \max_{U'} V(S',U')\right]$
U
$\delta$, $Q_{cabin}$
MDU
$P_{trac}$
HVAC
$P_{HVAC}$
Battery
R
Real-time performance feedback
Real-time measurements

*Figure 4. Reinforcement learning formulation for EMS*

There are several challenges that need to be resolved through the design. First, the EMS interacts with the local HVAC controls so the variable that EMS modulates needs to be compatible and not cause adverse interactions with the subsystem level control operations. Second, dependencies and varying time scales among various objectives need to be considered. For example, the traction dynamics are much faster than the cabin temperature dynamics whereas the battery aging is affected by both power, and it is a cumulative variable.

An example case of how the EMS interacts with the subsystem-level HVAC controls is shown in Fig. 5. The EMS augments the cabin temperature setpoint; tweaking it around a nominal value determined by cabin comfort metrics. The HVAC controller then uses this modified reference signal in modulating the HVAC actuators. A variety of regulation controller has been used for the baseline cabin temperature regulation.

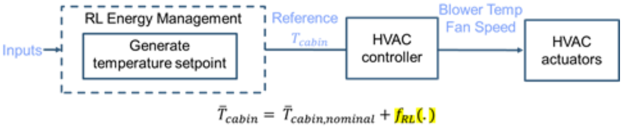


*Figure 5. Example of EM subsystem controls interaction*

Another design aspect is to construct a reward signal that incorporates energy consumption, battery aging, and cabin comfort objectives. In the current design, traction power is to be delivered as it is, so there is no torque shaping allowed. For HVAC, cabin comfort is quantified through the level of variations in cabin temperature and Predicted Mean Vote (PMV) model. PMV is a common index that aims to predict the cabin comfort on a seven-point thermal sensation scale (Lahlaou et al. 2020). The cabin comfort described using a PMV range can be translated into an allowable cabin temperature range for a given ambient temperature and assuming that the passengers belong to an average scenario in terms of metabolism, clothing etc. In this paper, we determine the comfortable temperature range independent of the EMS design so that the remaining control objective is to maintain the cabin temperature within this band during EMS operation. Finally, for the battery aging, we used the battery aging model as described earlier to penalize the delta capacity loss.

The EMS is in essence an optimal control problem defined by (8) solved by reinforcement learning techniques. The function $r(.)$ is the reward parameterized by states of the system, disturbance, and optimization metrics. The control variable, $a$, is an additive term; together with the cabin comfort temperature ($T_{cabin,comf}$), it defines the cabin temperature reference ($T_{cabin,ref}$). Magnitude of $a$ is limited to an acceptable range $\overline{T_{PMV}}$ based on the PMV comfort metrics.

$$\underset{a}{argmax}\, J = \sum_{0}^{\infty} r(SOC(k), Q_{loss}(k), P_{dem}(k)) \quad (8)$$

*such that,*

$$|a| \leq \overline{T_{PMV}};\; T_{cabin,ref} = T_{cabin,comf} + a$$
$$SOC(k), Q_{loss}(k), T_{cabin}(k) \ldots \in X(k)$$
$$X(k+1) = f(X(k), f_{RL}(a), P_{dem})$$

The states of the system are denoted $X$ and the states transition is affected by the control through the transformation $f_{RL}(a)$.

## 3. DESIGN PROCESS AND RESULTS

The design process tailors the RL algorithms into a formulation equivalent to the optimal control problem described previously and solve it by learning the input-output mapping of the RL strategy. The objective is to maximize the design reward by running through representative drive cycles with traction and HVAC controls.

### *3.1 Reinforcement learning design*

The RL formulation considered in this work as EMS controller augments the cabin temperature setpoint (e.g., RL action); the main inputs to the RL (or RL states) are vehicle power demand, vehicle acceleration and vehicle speed. The specific role of RL in all formulations is to populate this defined input-output mapping. Formally, the state-action and reward are described below:
RL States: $s = \{V_{veh}(k), Acc_{veh}(k), P_{dem}(k)\}$
RL Action: A = {Cabin temperature setpoint modifier $a$(k)}
RL Policy: $\pi(s) \rightarrow a$
RL Reward function: $r(\Delta Q_{loss}, T_{cabin,comf}, T_{cabin})$
The reward function for the reinforcement learning is defined using a weighted sum of $\Delta Q_{loss}$ and $T_{error}$ which measures deviation of cabin temperature from $T_{cabin,comf}$.

$$\Delta\, \underset{a}{argmax}\; r = -\left(w_1\, \Delta Q_{loss} + w_2\, T_{error}^2\right) \quad (9)$$

The aging term $\Delta Q_{loss}$ quantifies instantaneous battery aging at the current time based on the total throughput, battery current and temperature and $T_{error}$ is the cabin temperature response to capture the cabin comfort in the reward. The temperature term ensures that the RL does not converge to a trivial solution of choosing the furthest cabin temperature allowed by design.

For implementations, a custom RL toolbox has been developed in a modular form in MATLAB and used to explore different EMS formulations. During the implementations, RL adjustment has also been applied at sampled data points and held constant over a defined window. The instantaneous reward is also integrated over this window to determine a cumulative reward, which is then used in updating the Q-value for Q-Leaning and SARSA learning updates by (10-11):

Q-learning (off-policy algorithm):

$$Q(s,a) \leftarrow\!\cdot\, Q(s,a) + \sigma \cdot (r + \gamma \cdot max_a\, Q(s',a) - Q(s,a)) \quad (10)$$

SARSA (on-policy algorithm):

$$Q(s,a) \leftarrow\!\cdot\, Q(s,a) + \sigma \cdot (r + \gamma \cdot Q(s',a') - Q(s,a)) \quad (11)$$

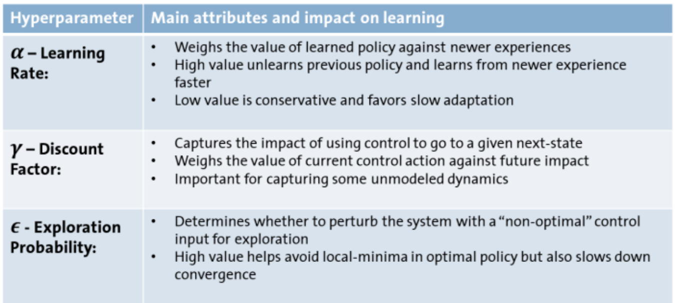
where $a$ is the current action (RL action), $s$ are the states, $r$ is the reward function, $max_a\, Q(s', a,\ \epsilon)$ is called greedy policy, $\sigma$ is the learning rate, $\gamma$ is the discount factor and $\epsilon$ is the exploration probability in the greedy policy. The parameters of the RL algorithms have been generated using a novel adaptation scheme as will be described in Section 3.2.

The Q-value is represented as an N-dimensional look-up table for ease of access and interpretability. In storing and updating the Q-values over the input-action space, a quantization scheme is applied. In general, the quantization step should be fine enough to capture effects of changing states and control variable smoothly; and coarse enough to reduce the size of the Q-value mapping. Sparsity of trained policy increases with finer discretization and number of states, which may improve in turn the optimality of final policy.

### *3.2 Hyperparameter selection*

Another key design aspect is selection of RL hyperparameters. Hyperparameters define how learning is executed and affects convergence & optimality of the final policy, which is essential for success of RL algorithm implementation.

*Table. 1. Main design hyperparameters*



| Hyperparameter | Main attributes and impact on learning |
|---|---|
| **$\alpha$ – Learning Rate:** | • Weighs the value of learned policy against newer experiences<br>• High value unlearns previous policy and learns from newer experience faster<br>• Low value is conservative and favors slow adaptation |
| **$\gamma$ – Discount Factor:** | • Captures the impact of using control to go to a given next-state<br>• Weighs the value of current control action against future impact<br>• Important for capturing some unmodeled dynamics |
| **$\epsilon$ - Exploration Probability:** | • Determines whether to perturb the system with a “non-optimal” control input for exploration<br>• High value helps avoid local-minima in optimal policy but also slows down convergence |

Table 1 shows the three main hyperparameters with their main description and impact on learning. There is a strong coupling among those parameters and how to select them is usually ad-hoc and involves some level of trial and error. In this paper, we used an adaptation scheme that adjust them over the course of training to change the training behaviour as the learning continues.

Among them, exploration probability ($\epsilon$) is the proportion of “exploration: trying something new” to “exploitation: building on what is already learned” at a given learning step. In the adaptive scheme, exploration probability has been chosen to have a decaying characteristic with the episode number, which is set high in the initial phase of training to encourage exploration and then set to decay as the learning improves. $\epsilon$ is adapted from the initial $\epsilon_0$ over the training using (12), where D is the decay rate which has to be chosen appropriately.

$$\epsilon = \epsilon_0 \, D^{episode} \tag{12}$$

Learning rate is adjusted to reflect the level of confidence in current policy over new experiences. A typical desired learning rate curve can be generated using (13). The weighting term $w_\alpha$determines the adaptation rate. As illustrated in Fig. 6, the value of $\alpha$ is set to be bounded so that the RL policy continues to get updated.

$$\alpha = \frac{1}{1+w_\alpha \text{ episode}} \tag{13}$$

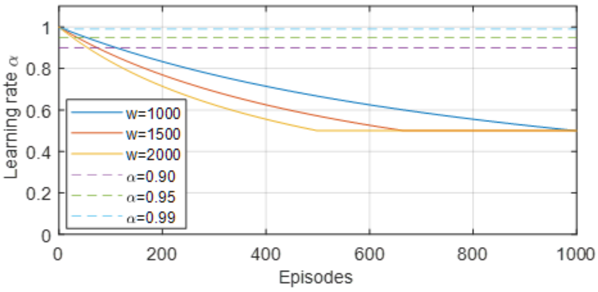


*Figure 6. Adaptive learning rate selection*

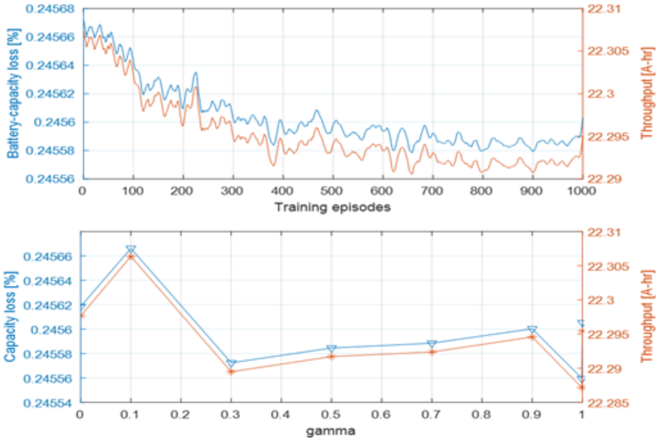


*Figure 7. Discount factor impact on performance*

These hyperparameters have influence on the optimality of the resulting solution as well. Fig. 7 illustrates the impact of discount factor on battery capacity loss and battery throughput.

### 3.3 Results

This section summarizes example results from the training and validation process through driving the EV models by various drive profiles. An energy consumed breakdown during an initial training profile, which was an FTP cycle, is shown in Fig. 8 where episode number indicates the training iteration and cumulative energy consumed for each load in Joules is plotted per episode number. Here, the consumed traction energy stays the same since we don't allow any torque/speed shaping in the design, yet the total HVAC energy, and in turn the total energy, has a reducing trend as the training evolves.

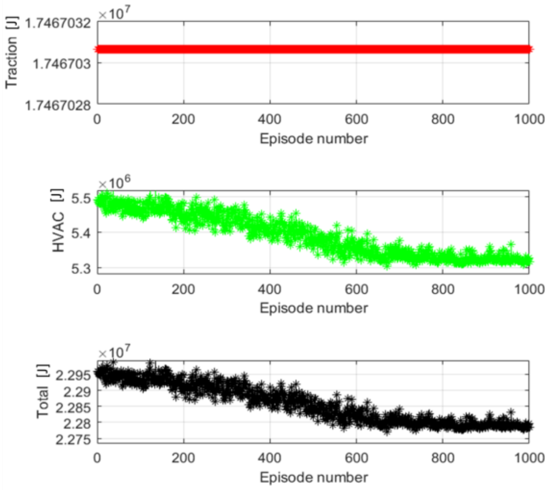


*Figure 8. Convergence during RL training*

Drive cycle is the primary disturbance to the system. Statistical properties of the drive profile are key in achieving the general optimality and robustness of the data-driven solution. Hence, we have used long real-world drive profiles. Figs. 9-11 represent example validation results on these real-world drive profiles where RL was adjusting the nominal cabin heat input controls dynamically via perturbations to the temperature set-point. Fig. 9 shows the heat flow per time for baseline controls (nominal) and after adjusted by the RL-based EMS. A PI controller has been designed to achieve the cabin temperature regulation that adjusts the heat delivery into the cabin. The nominal case uses this control only with the default temperature set-point. Fig. 10 shows the resulting cabin temperature response and the vehicle speed. Note that, EMS maintains the cumulative HVAC heat input the same in order not to impact the cabin comfort, yet applies small dithers in both directions.

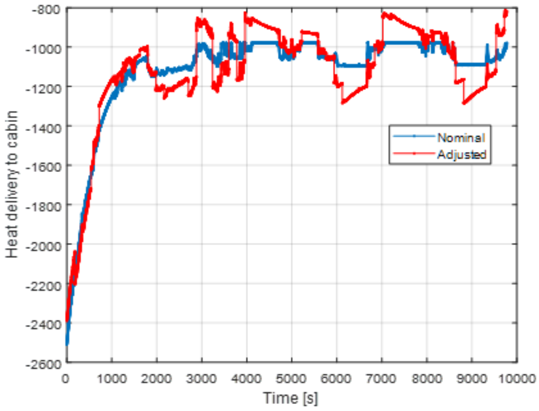


*Figure 9. Heat flow input to cabin*

For the current example, HVAC was running in cooling mode, and EMS, in essence, was adjusting the cooling level based on current traction states. Specifically, it slightly increased cooling at higher speeds and decreased at lower vehicle speeds leveraging the efficiency dependence of the AC operation on the vehicle speed. A similar mechanism can also be interpreted through the traction power demands where the EMS adjusts the HVAC power to reduce the simultaneous high demands from both systems to improve battery aging.

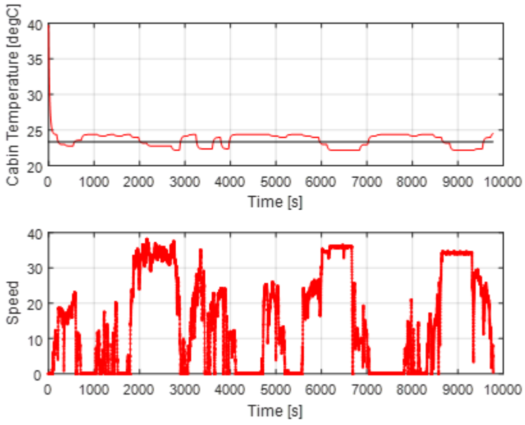


*Figure 10. Cabin temperature response and vehicle speed*

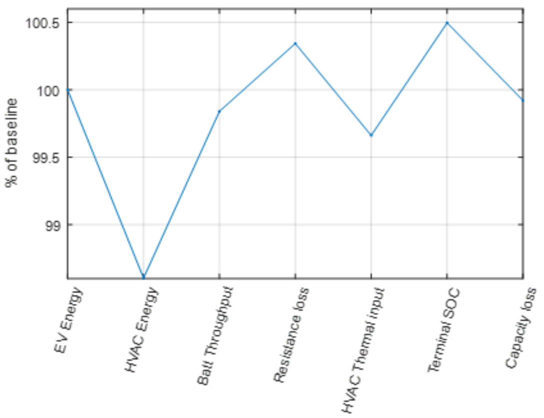


*Figure 11. Example cycle-end performance*

Finally, Fig. 11 includes the cumulative cycle-end values for various variables normalized with respect to their nominal values demonstrating energy efficiency and battery aging improvement with the current EMS while maintaining about the same total heating flow through the drive profile.

## 4. CONCLUSIONS

More efficient BEV energy management can be achieved using a supervisory controller that better coordinates multiple power requests (drive, thermal, HVAC, etc.). In this paper, we developed such an energy management control strategy and demonstrated it on the integrated EV traction and HVAC controls. Optimal strategy modifies an HVAC set-point in response to vehicle states and power demands. The design process included creating a large-scale simulation platform, developing custom RL toolbox and EV models for fast training and performance evaluations in an HPC environment.

It was shown that the battery aging and energy efficiency can be improved without affecting the total heat flow to the cabin. Note that EMS did not use torque shaping or rely on an additional energy storage system, instead generated a small dynamic perturbation near the nominal cabin temperature set point with zero mean. Further improvement in the overall performance could be realized by increasing the control degree of freedom in the previously mentioned aspects. Another contribution of this work was to devise a novel adaptive hyperparameter selection scheme, which simplifies the RL calibration and convergence. In general, reinforcement learning to design energy management controllers both reduces calibration effort during design phase by enabling incorporation of various optimization objectives into the final calibration and generates a real-time self-learning mechanism deployable in the field.

## REFERENCES

Cordoba-Arenas, A., Onori, S., Guezennec, Y., and Rizzoni, G. (2015). 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, 473–483.

Ermon, S., Xue, Y., Gomes, C., and Selman, B. (2013). Learning policies for battery usage optimization in electric vehicles, Machine Learning, 177–194.

Han, X., He, H., Wu, J., Peng, J. and Li, Y., 2019. Energy management based on reinforcement learning with double deep Q-learning for a hybrid electric tracked vehicle. Applied Energy, 254, p.113708.

Haskara, I., Hegde, B., and Chang, C. (2021). "Reinforcement Learning Based Energy Management of Hybrid Energy Storage Systems in Electric Vehicles," SAE Technical Paper 2021-01-0197.

Lahlou, A., Ossart, F., Boudard, E., Roy, F., and Bakhouya, M. (2020), Optimal Management of Thermal Comfort and Driving Range in Electric Vehicles, Energies, 13, issue 17, 1-31.

Lee, H., Song, C., Kim, N. and Cha, S.W., 2020. Comparative analysis of energy management strategies for HEV: Dynamic programming and reinforcement learning. IEEE Access, 8, 67112-67123.

Li, W., Cui, H., Nemeth, T., Jansen, J., Ünlübayir, C., Wei, Z., Zhang, L., Wang, Z., Ruan, J., Dai, H. and Wei, X., 2021. Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles. Journal of Energy Storage, 36, p.102355.

Sakhdari, B., and Azad, N.L. (2015). An Optimal Energy Management System for Battery Electric Vehicles, *IFAC-Papers On-Line*, volume (48), issue (15), 86-92.

Vatanparvar, K., and Al Faruque, M.A. (2018). Design and Analysis of Battery-Aware Automotive Climate Control for Electric Vehicles, ACM Transactions on Embedded Computing Systems, volume (17), issue (4), 1-22.

Wang, H., Kolmanovsky, I., Amini, M.R., and Sun, J. (2018). Model Predictive Climate Control of Connected and Automated Vehicles for Improved Energy Efficiency, 2018 Annual American Control Conference (ACC), Milwaukee, USA, 828-833.