Comparison of Three Real-Time Implementable Energy Management Strategies for Multi-mode Electrified Powertrain

Comparison of Three Real-Time Implementable Energy Management Strategies for Multi-mode Electrified Powertrain

Atriya Biswas[1], Pier G. Anselma[2], Aashit Rathore[1], and Ali Emadi[1]

[1]McMaster Automotive Resource Centre (MARC), McMaster University, Hamilton ON, Canada
[2]Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, Torino, Italy
E-mail:biswaa4@mcmaster.ca

***Abstract*—Three real-time implementable energy management system (EMS) strategies have been articulated for forward simulation vehicle model with an electrified powertrain. Rule-based strategy and equivalent consumption minimization strategy (ECMS) have been profoundly used as a competent real-time implementable EMS strategy for electrified powertrain. Reinforcement learning (RL) is relatively new as a real-time EMS controller. All these three controllers have been articulated for model-in-the-loop (MIL) simulation. A comparison among state-of-the art RL-based controller, widely accredited ECMS, and rule-based control strategies is very crucial in order to analyze strengths and weaknesses of each of these strategies at the MIL and to make them apposite for the subsequent phases of utilitarian controller development.**

***Index Terms*—Actor-critic, Automotive systems, Deep reinforcement learning, electrified powertrains, energy management system, MIL, Q-learning, real-time.**

I. Introduction

WITH the advent of electrified powertrains, automotive industry has been witnessing gradual improvements in the fuel-economy and reduction in tailpipe emissions. The success of powertrain electrification hugely depends on the articulation of the EMS controller [1]. The EMS should satisfy the power demand from the vehicle and should constrain battery state of charge (SOC) within allowable limits.

Premeditated control strategies are the most convenient candidate for being embedded onto a utilitarian EMS since they are not laden with any mandatory real-time optimization. However, the rules extracted from the optimal solution can only yield optimal performance for the drive cycles whose global optimal result were used for extraction of rules.

Local optimization-based control policy can be implemented in real-time with a widely accredited ECMS which is an embodiment of Pontryagin's minimum principle (PMP) [2]. Optimality with close proximity to Dynamic programming (DP)can be achieved through ECMS with a few approximation and meticulous choice of equivalence factor (EF) [2]. However, the calculation of optimal equivalence factor in ECMS requires either full knowledge of future driving situations. The EF can be varied adaptively with an initial guess in MIL, software-in-the-loop (SIL), or even real-time simulation in absence of advance knowledge of the future driving situation.

While researchers are striving for a perfect mathematical tool ensuring autonomous as well as real-time control with near-global optimality under real-world driving situation, RL-based control is a lucrative option to explore. Application of RL was introduced for solving the energy management problem of a parallel hybrid electric vehicle in [3]. In a previous article [4] by the authors and only a handful of literature [5], [6] on application of RL have corroborated RL's prowess in optimal energy management of electrified powertrain. But all the RL controllers in literature were applied in backward simulation models which cannot represent real-time behavior of a physical system. Backward simulation platforms cannot emulate the prospective behavior of RL controllers in EMS of a real electrified powertrain for real driving condition. These motivated the authors of this article to develop an online RL-based EMS controller to interact with the real-time forward vehicle simulation model. Also, performance of the RL-based controller is compared with other two real-time implementable controller such as ECMS and rule-based controller (RBC) [7] with the same forward vehicle simulation platform to justify RL's dexterity in comparison with ECMS and RBC.

Rest of the brief is organized as follows: section II describes the fundamentals of EMS with brevity and elaborates the mathematical approach adopted for modeling the powertrain dynamics. Section III delineates articulation of a RBC specifically for the powertrain configuration adopted here. Section IV presents the conventional-ECMS framework and it's implementation. Section V posits an actor-critic based RL framework capable of autonomously learning the near-optimal values of EF as the control variable. Section VI presents the simulation results and juxtaposes the performances obtained through three real-time implementable control strategies. Conclusive remarks are drawn in section VII.

II. Fundamentals of Utilitarian EMS and Powertrain Modeling

A. Fundamentals of Utilitarian EMS

Working as a system level supervisory controller, EMS dictates operating points of the cardinal components of the

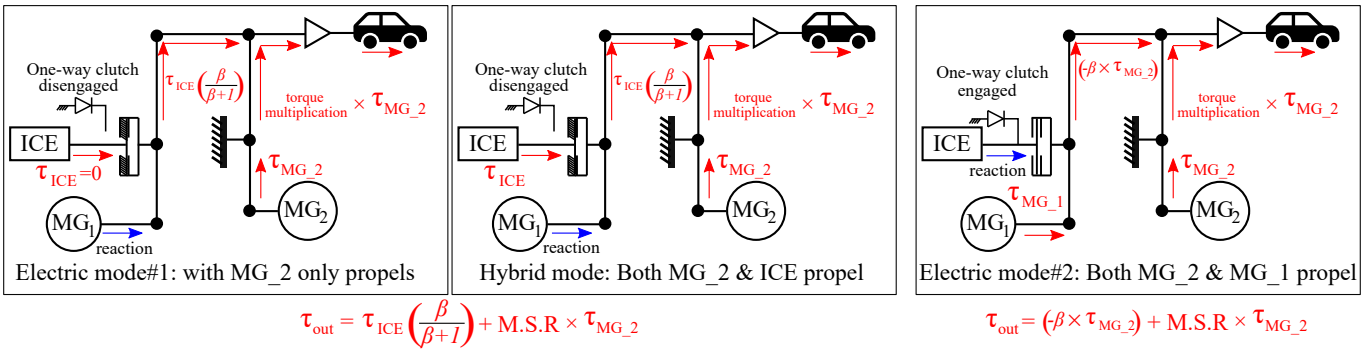

Fig. 1: Simple schematic diagrams showing power flow in different modes in the chosen e-CVT.

electrified powertrain within permissible limits of the respective components. The objectives of an utilitarian EMS can be chronologically divided in three stages, i.e., cardinal, adjunct, and utilitarian objectives. EMS should distribute the driver's requested power optimally among all the energy sources and minimize fuel consumption with a given condition that state-of-charge (SOC) of the electric energy storage system (EESS) remains within the permissible limits. Minimization of tailpipe emissions and drivability improvement can be considered as adjunct objectives. The simulation-based controller needs to abide by several constraints before being called apposite for real-time simulation. The EMS controller should be robust and computationally efficient for being apposite in an electronic control unit (ECU). Three real-time implementable control strategies, i.e., RBC, ECMS, and Deep reinforcement learning (DRL)-based agent will be employed in the EMS of an hybrid electrified powertrain (HEPT) and their performances will be juxtaposed based on all the aforementioned objectives.

B. Mathematical Modeling of Powertrain

A midsize 2500 kg representative passenger vehicle with multi-mode electronic continuously variable transmission (e-CVT) transaxle [8], whose schematic is shown in fig.1, is selected for this study. The transaxle is comprised of an internal combustion engine (ICE), two electric motor/generator (EMG)s, and a high-voltage battery pack.

Based on the chosen architecture, the powertrain can operate in three different transmission modes as shown in fig.1. The electric mode#1 does not offer any degree of freedom. The hybrid mode has two degrees of freedom. The ICE-locked electric mode#2 has one degree of freedom. The ICE is turned-off in this mode and it act as a rigid support with engagement of the one-way clutch.

1) ICE modeling: A medium-fidelity 3.3 Lt spark-ignition (SI) ICE is modeled through two most essential engine characteristic look-up tables, i.e., wide open throttle (WOT) torque and mass flow rate (MFR) of fuel.

2) EMG modeling: Low-fidelity models of both EMG1 and EMG2 are developed with a simple look-up table approach. Maximum torque ($\tau_{MG_{max.}}$) curves and efficiency (η_{MG}) table of both EMGs can be represented as following relations:

$$\begin{aligned} \tau_{MG_{max.}} &= f(\omega_{MG}, Voltage_{MG}) \\ \eta_{MG} &= f(\omega_{MG}, \tau_{MG}, Voltage_{MG}) \end{aligned} \tag{1}$$

3) High-Voltage Battery modeling: A low-fidelity high-voltage battery is simply modeled with an equivalent circuit approach and the model is comprised of an open circuit voltage (OCV) in series with internal resistance (IR) of the battery.

4) Transmission dynamics modeling: The system dynamics of the multi-mode e-CVT can be mathematically modeled using Eq.(2) and Eq.(3), with given methodology in [9].

$$\begin{bmatrix} \tau_{out} \\ \tau_{ice} \\ \tau_{mg1} \\ \tau_{mg2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} I_{oeq.} & 0 & 0 & 0 & -R_1 & -R_2 \\ 0 & I_{ice_{eq.}} & 0 & 0 & R_1+S_1 & 0 \\ 0 & 0 & I_{mg1_{eq.}} & 0 & -S_1 & 0 \\ 0 & 0 & 0 & I_{mg2_{eq.}} & 0 & -S_2 \\ -R_1 & R_1+S_1 & -S_1 & 0 & 0 & 0 \\ -R_2 & 0 & 0 & -S_2 & 0 & 0 \end{bmatrix} \begin{Bmatrix} \dot{\omega}_{out} \\ \dot{\omega}_{ice} \\ \dot{\omega}_{mg1} \\ \dot{\omega}_{mg2} \\ F_1 \\ F_2 \end{Bmatrix} \tag{2}$$

$$\begin{bmatrix} \tau_{out} \\ \tau_{mg1} \\ \tau_{mg2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} I_{oeq.} & 0 & 0 & -R_1 & -R_2 \\ 0 & I_{mg1_{eq.}} & 0 & -S_1 & 0 \\ 0 & 0 & I_{mg2_{eq.}} & 0 & -S_2 \\ -R_1 & -S_1 & 0 & 0 & 0 \\ -R_2 & 0 & -S_2 & 0 & 0 \end{bmatrix} \begin{Bmatrix} \dot{\omega}_{out} \\ \dot{\omega}_{mg1} \\ \dot{\omega}_{mg2} \\ F_1 \\ F_2 \end{Bmatrix} \tag{3}$$

III. Rule-based Control for Real-time EMS

The RBC under consideration is described in this section. RBC currently represents the real-time control approach typically implemented in on-board control units of commercially available e-CVT hybrid electric vehicle (HEV)s [10], [11]. Two levels of decisions need to be accomplished for this controller respectively regarding the status of the ICE (i.e. on/off) and the torque-split between power components. These are achieved in a sequential order in the vehicle controller module as illustrated in fig.(2). In general, at each time instant the vehicle states such as speeds, torques and battery SOC as example are evaluated in the vehicle plant. The proportional-integral (PI) controller modeling the driver then computes the overall vehicle power demand. This information, together with the vehicle states assessed in the vehicle plant model, represents the inputs to step 1 of the implemented RBC logic.

A. ICE status determination

The first step of the implemented RBC aims at determining whether the ICE should be operating or not. This is governed

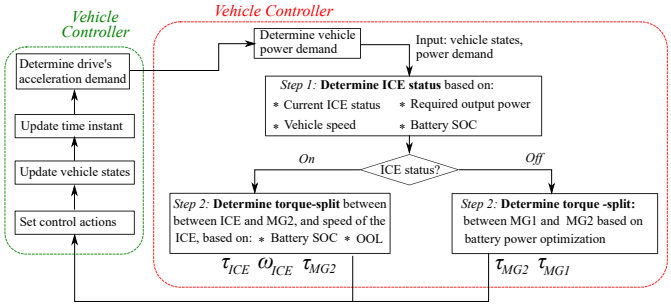

Fig. 2: Flowchart of the implemented RBC.

TABLE I: Decision table for ICE status

	Controlled ICE status	
Current ICE status	ON	OFF
ON	$V_{veh} >= V_{veh_{limit_{ON}}}$ $P_{load} >= P_{load_{limit_{ON}}}$ $SOC <= SOC_{limit_{ON}}$	$V_{veh} < V_{veh_{limit_{ON}}}$ $P_{load} < P_{load_{limit_{ON}}}$ $SOC > SOC_{limit_{ON}}$
OFF	$V_{veh} >= V_{veh_{limit_{OFF}}}$ $P_{load} >= P_{load_{limit_{OFF}}}$ $SOC <= SOC_{limit_{OFF}}$	$V_{veh} < V_{veh_{limit_{ON}}}$ $P_{load} < P_{load_{limit_{ON}}}$ $SOC > SOC_{limit_{ON}}$

by a set of rules that considers the current ICE status, the required output power and current values of vehicle speed v_{veh} and battery SOC as decision factors. Particularly, the ICE is set to operate in case at least one of the following conditions are satisfied including (1) the v_{veh} is above a certain threshold $v_{veh_{limit}}$, (2) the required output power P_{load} exceeds a predefined value $P_{load_{limit}}$ and (3) the battery SOC value is below a specific value SOC_{limit}. Table I summarizes the described decision process. In order to avoid frequent $stop - start$ of the ICE, different threshold values are considered depending on the current ICE status. These have been particularly selected to satisfy the following relationships: $v_{veh_{limit-ON}} > v_{veh_{limit-OFF}}$, $P_{load_{limit-ON}} > P_{load_{limit-OFF}}$ and $SOC_{limit_{ON}} < SOC_{limit_{OFF}}$.

B. Torque-split determination

Once the ICE status has been determined, the next step aims at evaluating the optimal torque-split. In case the ICE is set not to operate (i.e. pure electric), the torque-split between EMG1 and EMG2 can be determined by solving the optimization problem reported in Eq.(4).

$$Find\ \tau_{MG1}\ and\ \tau_{MG2}$$
$$P_{MG1} + P_{MG1} = P_{load}\ \&\&\ P_{batt} = min(P_{batt}) \quad (4)$$

A sweep of values for the EMG2 torque τ_{MG2} is particularly performed, and the corresponding value of EMG1 torque τ_{MG1} is selected in order to have the algebraic sum of powers (P) of the EMGs satisfying the power demand P_{load}. The optimal set of EMG torques corresponds to the one minimizing the overall value of the battery power P_{batt}. In this way, the overall efficiency of the electrical path of the powertrain including EMGs, power electronics and battery can be enhanced. In order to reduce the required computational power, the optimization problem illustrated in Eq.(4) could be solved off-line for all the possible values of P_{load} and then loaded in the on-board control unit in the form of lookup tables.

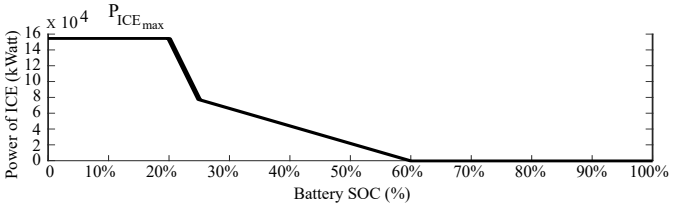

Fig. 3: Lookup table for selecting the ICE power.

On the other hand, in case hybrid operation has been selected at the previous step of the logic, the controlled ICE mechanical power $P_{ICE_{control}}$ is selected by interpolating in a 1D lookup table with battery SOC as independent variable as illustrated in fig.3. This table has particularly been calibrated in order to avoid excessively depleting the battery while simultaneously improve the fuel economy capability of the HEPT. After the value of $P_{ICE_{control}}$ has been selected, ICE speed ω_{ICE} and torque τ_{ICE} can be determined by solving the optimization problem associated to the well-known ICE optimal operating line (OOL) as reported in Eq.(5) [2].

$$Find\ \tau_{ICE}\ and\ \omega_{ICE}$$
$$\tau_{ICE} * \omega_{ICE} = P_{ICE_{control}}\ \&\&\ \dot{m}_{fuel} = min(\dot{m}_{fuel}) \quad (5)$$

ω_{ICE} and τ_{ICE} should particularly be able to deliver the mechanical power requested by the ICE, while simultaneously minimizing the instantaneous fuel consumption rate. The efficiency of the ICE operation can be enhanced in this way. Also in this case, the optimization problem illustrated in Eq.(5) could be solved off-line for different values of $P_{ICE_{control}}$ and then stored in the on-board control unit as computationally lightweight lookup tables. For the hybrid operation, the EMG2 torque τ_{MG2} can be finally determined in order to comply with the requested P_{load}. On the other hand, the reaction torque for the EMG1 τ_{MG1} can be obtained in the plant model by following the standard torque relationships for planetary gears.

IV. Equivalent Consumption Minimization Strategy for Real-time EMS

ECMS is based on the concept of converting electrical energy consumption into equivalent fuel consumption and then minimizing their sum [12]. It is an instantaneous optimization technique which yields sub-optimal solution. It develops the cost function which represents the total equivalent fuel consumption. ECMS was originally developed from heuristic idea that the energy utilized in the form of electrical energy is ultimately coming from chemical energy and has to be compensated in the future [2]. Although studies performed afterwards proved that an analytical derivation of the ECMS can be derived from the Pontryagin's Minimum Principle [2] [13] [14].

Following the description in [15], the generic formulation of the conventional ECMS objective function can be written as:

$$\{P_{(ice)}^{(opt)}(t), P_{(em)}^{(opt)}(t)\} = \underset{P_{(ice)}(t), P_{(em))}(t)}{\arg\min} J_t \tag{6}$$

where $P_{(ice)}^{(opt)}(t)$ is the optimum value of the power supplied by the ICE and $P_{(em))}^{(opt)}(t)$ is the optimum value of the power supplied by the EM. J_t is the objective function defined as.

$$J_t = \dot{m}_{ice}(P_{ice}(t)) + f(EF, P_{em}) \tag{7}$$

here, $\dot{m}_{ice}(P_{ice}(t))$ is the fuel consumed by the ICE and $f(EF, P_{em})$ is the equivalent fuel consumed by the electrical energy which is a function of EF and P_{em}. As already mentioned, the ECMS is highly sensitive to the value of EF and requires full information of drive cycle to optimally choose its value. In case of an HEV, the online implementaion of ECMS without prior information, only guarantees the charge sustainability criteria if the EF chosen is optimally constant value. Hence a lot of methods have been developed in literature to adapt the EF in real-time [16].

V. Deep Neural Network-based Reinforcement Learning Control for Real-time EMS

Tailored with deep neural network (DNN), advanced RL-agents have corroborated their prowess in obtaining near-optimal control through autonomous learning without prior information of the Markov decision model for different sectors such as robotic control, intelligent transport system.

A. Brief Fundamental of Reinforcement Learning

RL imitates the complex decision making capability exhibited predominantly by humans. RL encompasses a handful of learning algorithms which mathematically model an intricate characteristic of a sophisticated human capability, i.e., progressive learning to make sharper decisions with an objective of maximizing the long-term return [17]. The RL agent interacts with a Markov decision problem (MDP) which is mathematically modeled through the states $(S_t) \in \mathcal{S}$, actions$(A_t) \in \mathcal{A}$, and reward function$(r_t) \in \mathcal{R}$. The dynamics of MDP is defined by the probability of moving to state S' at time $t+1$ if action (A_t) is applied on S_t at time t $(\mathcal{P}_{S_t S'}^{A_t} = Pr(S'|S_t, A_t))$ [18]. The RL agent is comprised of three cardinal elements, i.e., it's policy function$(\pi(S))$, state value functions $(V(S_t))$, and action value functions $(Q(S_t, A_t))$.

Policy function, which governs the selection of an action at a certain state, is a mapping from state-space to action-space. The task of the RL agent is to find the optimal policy through iterative process with a sole objective of maximizing the long-term return which is denoted as follows:

$$\mathscr{R}_t = \sum_{k=1}^{T} \gamma^k r_{t+k} = r_t + \mathscr{R}_{t+1} = r_t + \gamma r_{t+1} + \mathscr{R}_{t+2} \tag{8}$$

The RL agent uses either $V(S_t)$ or $Q(S_t, A_t)$ as the main wield while iteratively searching for the optimal policy. If $\mathcal{P}_{S_t S'}^{A_t}$, which is also known as the model of MDP, is available

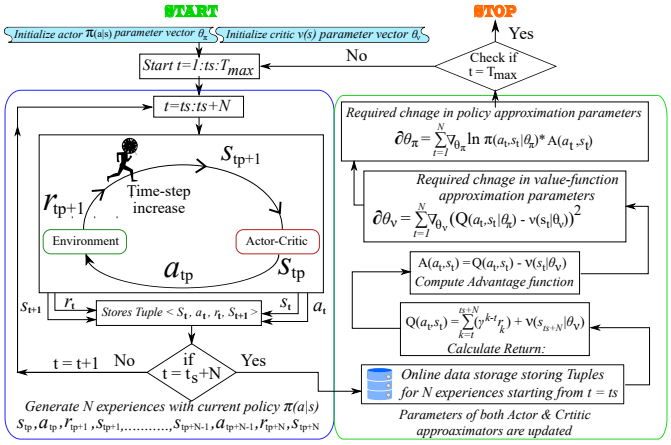

Fig. 4: Training algorithm used in A3C agent.

to the RL agent for the entire MDP, DP-based algorithms can be employed to find the global optimal policy for the MDP. However, in real-world situation, where the prior information of entire MDP model is not available to the RL agent, temporal difference (TD) learning algorithms are most apposite for finding near-optimal policy [17]. $Q-learning$ and $SARSA$ are the most abundantly used TD algorithms. When the environment has less number of discrete state and action variables, tabular approach is quite convenient for finding numerical values of near-optimal $V(S_t)$ and $Q(S_t, A_t)$ for all states and state-action combinations. But, if either state-space or action-space has copious number of discrete variables, tabular approach become highly inconvenient due to "$curse\ of\ dimensionality$" [19]. The most promising wield for tackling "$curse\ of\ dimensionality$" is the use of functional approximation for both value functions $(V(S_t)$ and $Q(S_t, A_t))$ and policy function $(\pi(S))$ [19]. Authors in [20] have corroborated guaranteed convergence of TD learning with linear function approximation of value functions. Although there was an ambiguity in the past decades over the guarantee of convergence and stability of TD learning with nonlinear approximator such multi-layer perceptron, authors in [21] first confirmed the convergence of TD learning with artificial neural network (ANN)-based approximator by using "$Experience\ Replay$". The linear and nonlinear approximation can be expressed as follows with Eq.(9) and (10) respectively:

$$\tilde{Q}(s, a; \psi) = \{\psi_1, \cdots, \psi_\Lambda\}^T \Big[\phi_1(s, a), \cdots, \phi_\Lambda(s, a)\Big] \tag{9}$$

$$\tilde{Q}(s, a; \psi) = g(\psi(\lambda)\Phi_d(s, a)) \tag{10}$$

$\psi \in \mathbb{R}^\Lambda$ and $\Phi \in \mathbb{R}^\Lambda$ are parameter and feature-function vectors respectively. $g(.)$ is a nonlinear function representing architecture of the nonlinear approximator. Assisted with functional approximation, the optimal policy can be obtained in three different ways, i.e., actor-only, critic-only, and actor-critic methods [22]. Actor-only and critic-only methods focus on optimizing policy function and value functions respectively. Both actor-only and critic-only methods

have their individual pros and cons. Whereas, actor-critic methods leverage advantages of both actor and critic methods and hence can achieve convergence faster than the former two when assisted by nonlinear approximator. For this brief, an advanced version of actor-critic method, i.e., "$Asynchronous\ Advantage\ Actor-Critic$"(A3C) [23] has been appointed as the RL agent for finding the near-optimal energy management policy for a multi-mode e-CVT powertrain. Fig.4 depicts the training algorithm employed in A3C agent.

VI. Simulation and Result

All these three aforementioned EMS controllers are implemented in Simulink® environment. Both plant model and controllers are developed in Simulink®. Performance of the three controllers is juxtaposed on the basis of reduced fuel consumption, ability of charge sustenance, and computational time. Two standard drive cycles are chosen for this comparison study. The vehicle speed profile of WLTP and highway cycles along with powertrain mode profiles for three different controllers are shown in fig.5 and fig.8 respectively. A Markov chain model (MCM) is employed to generate copious number of random drive cycle from the characteristics of a combined drive cycle comprised of WLTP, highway, UDDS cycles. The randomly generated drive cycles are used to train the RL agent before the agent is confronted with WLTP and highway cycles for performance validation. The numerical values of EF guaranteeing charge sustenance for both the cycles are found through trial-and-error method and hence they give perfect charge sustaining performance for both the cycles as shown in fig.6 and fig.10. The comparison of fuel consumption, drivability, and computational time among the controllers is furnished in Tab.II. These three metrics represent cardinal, adjunct, and utilitarian objectives of EMS respectively. Albeit rule-based controller seems to be ahead of others two controller based on adjunct and utilitarian objectives, it failed to satisfy the charge sustenance in highway cycle with a big difference.

TABLE II: Comparison of performance metrics

	ECMS		RL-agent		Rule-based	
Drive cycle	WLTP	HW	WLTP	HW	WLTP	HW
Fuel cons. (g)	1153	1007	1136	878	1432.4	1256
Drivability (no. of mode shift)	Worst	Worst	Better	Good	Best	Best
Computational time(s)	265	107	462	188	19	11

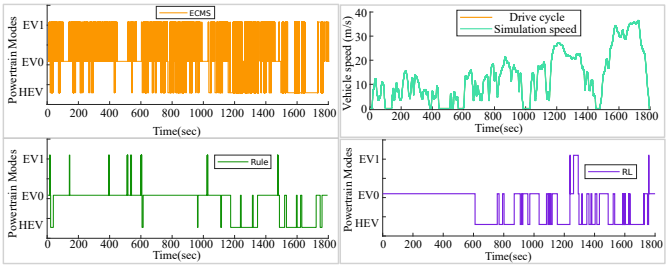

Fig. 5: Comparison of powertrain modes for WLTP cycle.

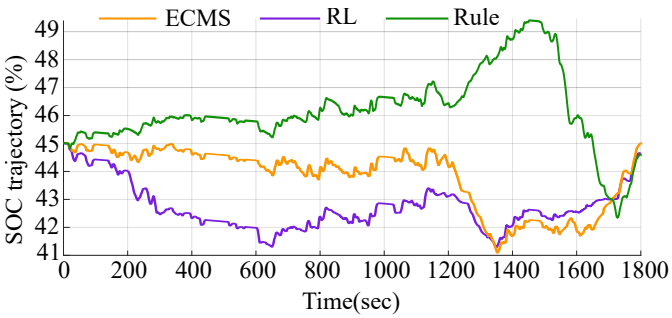

Fig. 6: Comparison of battery SOC profiles for WLTP cycle.

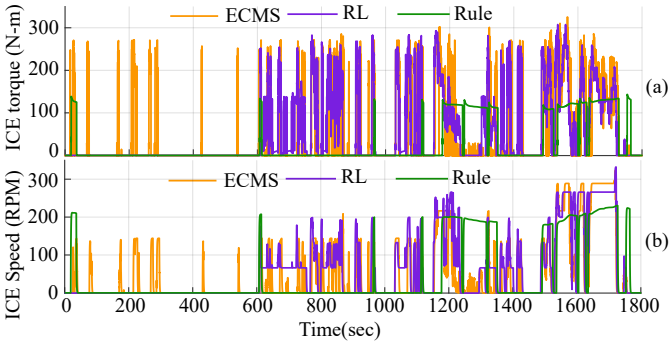

Fig. 7: Comparison of ICE torque & ICE speed for WLTP cycle.

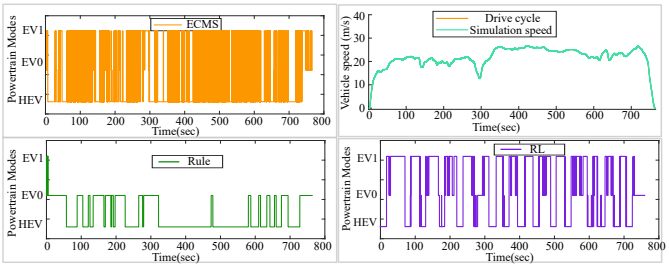

Fig. 8: Comparison of powertrain modes for highway cycle.

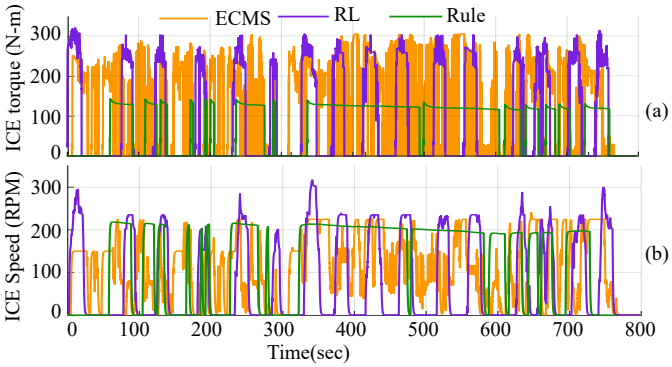

Fig. 9: Comparison of ICE torque & ICE speed for highway cycle.

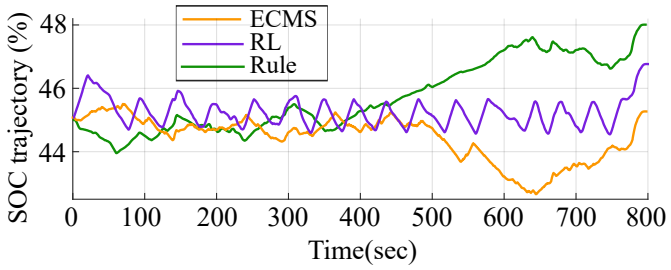

Fig. 10: Comparison of battery SOC trace for highway cycle.

VII. Conclusion

A brisk comparison between an emerging EMS controller and two already existing popular controller is conducted in this brief. Within it's short course, the article has indicated pros and cons of each of the controllers in controlling the energy management problem for a multi-mode e-CVT. The results predominantly corroborate a few areas, enumerated as follows:

- Any EMS controller should not only be evaluated based on it's competency of satisfying the primary objectives but the adjunct and utilitarian objectives should also be included in the performance evaluation criteria.
- Premeditated controllers such as rule-based control policies are far ahead of evaluative controllers [24] in terms of appropriateness in real-time implementation.
- Advanced RL agent-based controllers can be deployed as a judicious trade-off among cardinal, adjunct, and utilitarian objectives if more rigorous and comprehensive comparisons can be performed.

The RL agent has been trained "on-the-fly" with autonomously learning which indicates it's prospective capability as a strong contender of for real-time controller. More number of primary and adjunct objectives will be used as performance metric for the future study.

Acknowledgment

This research is supported, in part, thanks to funding from the Natural Sciences and Engineering Research Council of Canada (NSERC), NSERC Industrial Research Chair in Electrified Powertrains, and Canada Research Chair in Transportation Electrification and Smart Mobility.

References

[1] A. Emadi, *Advanced electric drive vehicles*. Boca Raton, FL: CRC Press, Oct. 2014.

[2] N. Kim, S. Cha, and H. Peng, "Optimal control of hybrid electric vehicles based on Pontryagin's minimum principle," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 5, pp. 1279–1287, Sep. 2011.

[3] A. Vogel, D. Ramachandran, R. Gupta, and A. Raux, "Improving hybrid vehicle fuel efficiency using inverse reinforcement learning," in *Proc. of the 26th AAAI Conf. on Artif. Intell.*, 2012, pp. 384–390.

[4] A. Biswas, P. G. Anselma, and A. Emadi, "Real-time optimal energy management of electrified powertrains with reinforcement learning," in *2019 IEEE Transport. Electrific. Conf. Expo (ITEC)*, June 2019, pp. 1–6.

[5] X. Qi, G. Wu, K. Boriboonsomsin, and M. J. Barth, "A novel blended real-time energy management strategy for plug-in hybrid electric vehicle commute trips," in *2015 IEEE 18th Int. Conf. Intel. Transp. Syst.*, Sep. 2015, pp. 1002–1007.

[6] Y. Li, H. He, J. Peng, and H. Zhang, "Power management for a plug-in hybrid electric vehicle based on reinforcement learning with continuous state and action spaces," *Energy Procedia*, vol. 142, pp. 2270 – 2275, 2017, proceedings of the 9th International Conf. Appl. Energy.

[7] P. G. Anselma, Y. Huo, J. Roeleveld, G. Belingardi, and A. Emadi, "Integration of on-line control in optimal design of multimode power-split hybrid electric vehicle powertrains," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3436–3445, April 2019.

[8] M. Pittel and D. Martin, "eFlite dedicated hybrid transmission for Chrysler Pacifica," in *SAE Technical Paper*. SAE International, 04 2018.

[9] P. G. Anselma, Y. Huo, J. Roeleveld, A. Emadi, and G. Belingardi, "Rapid optimal design of a multimode power split hybrid electric vehicle transmission," *Proc. Inst. Mech. Eng. D*, vol. 233, no. 3, pp. 740–762, 2019.

[10] J. Jeong, N. Kim, K. Stutenberg, and A. Rousseau, "Analysis and model validation of the toyota prius prime," in *SAE Technical Paper*. SAE International, 04 2019.

[11] M. A. Miller, A. G. Holmes, B. M. Conlon, and P. J. Savagian, "The GM Voltec 4ET50 multi-mode electric transaxle," *SAE Int. J. Engines*, vol. 4, no. 1, pp. 1102–1114, 2011.

[12] B. Gu and G. Rizzoni, "An adaptive algorithm for hybrid electric vehicle energy management based on driving pattern recognition," in *ASME Int. Mech. Engg. Cong. Expo.(IMECE)*, vol. Dynamic Systems and Control, Parts A and B, 11 2006, pp. 249–258.

[13] L. Serrao, S. Onori, and G. Rizzoni, "ECMS as a realization of Pontryagin's minimum principle for HEV control," in *2009 Proc Am. Control Conf.*, 2009, pp. 3964–3969.

[14] A. Chasse, A. Sciarretta, and J. Chauvin, "Online optimal control of a parallel hybrid with costate adaptation rule," *IFAC Proc. Volumes*, vol. 43, no. 7, pp. 99 – 104, 2010.

[15] C. Musardo, G. Rizzoni, Y. Guezennec, and B. Staccia, "A-ECMS: an adaptive algorithm for hybrid electric vehicle energy management," *Eur. J. Control*, vol. 11, no. 4, pp. 509 – 524, 2005.

[16] S. Onori, L. Serrao, and G. Rizzoni, "Adaptive equivalent consumption minimization strategy for hybrid electric vehicles," *Proc. ASME Dyn. Syst. and Control Conf. (DSCC)*, vol. 2010, no. 44175, pp. 499–505, 2010.

[17] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[18] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," in *Proc. 12th Int. Conf. Neural Info. Processing Systems*, ser. NIPS'99. Cambridge, MA, USA: MIT Press, 1999, p. 1057–1063.

[19] J. N. Tsitsiklis and B. van Roy, "Feature-based methods for large scale dynamic programming," *Machine Learning*, vol. 22, no. 1, pp. 59–94, Mar 1996.

[20] J. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Trans. Autom. Control*, vol. 42, no. 5, pp. 674–690, May 1997.

[21] M. Riedmiller, "Neural fitted Q iteration- first experiences with a data efficient neural reinforcement learning method," in *Machine Learning: ECML 2005*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 317–328.

[22] I. Grondman, L. Busoniu, G. A. D. Lopes, and R. Babuska, "A survey of actor-critic reinforcement learning: Standard and natural policy gradients," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1291–1307, Nov 2012.

[23] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," 2016.

[24] A. Biswas and A. Emadi, "Energy management systems for electrified powertrains: State-of-the-art review and future trends," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6453–6467, July 2019.