20th IMEKO TC10 Conference
*"Measurement for Diagnostics, Optimization and Control to Support Competitiveness and Innovation"*
Lisbon, Portugal, September 28–29, 2026

# State of Charge Estimation of Energy Storage Systems using a Flexible Reward based Reinforcement Learning Approach

Sadia Ali[1], Valentina Bianchi[1], Gabriele Patrizi[2], Lorenzo Ciani[2], Ilaria De Munari[1]

[1]*Università degli studi di Parma, Parma, Italy, sadia.ali@unipr.it, valentina.bianchi@unipr.it, ilaria.demunari@unipr.it*
[2]*Università degli studi di Firenze, Firenze, Italy, gabriele.patrizi@unifi.it, lorenzo.ciani@unifi.it*

***Abstract* – A flexible reward-based reinforcement learning (RL) approach for the state of charge (SoC) estimation of energy storage systems (ESSs) i.e. a Lithium-ion battery (LiB) and a hybrid supercapacitor (HSC), is presented in this research. The data acquisition for the Panasonic 18650 LiB is carried out through constant current – constant voltage (CC-CV) procedure whereas, for the cylindrical HSC electrochemical impedance spectroscopy (EIS) is used. ESS. A three-stage approach is presented i.e. reinforcement learning (RL) training, least square boosting (LSB) stage and three-point linear interpolation. RL incorporates a twin-delayed deep deterministic (TD3) neural network-based policy gradient agent along with a reward function specifically designed to take into account a flexible use in possible different contexts. The RMSE, along with other performance metric, is reported and analysed for each discharge cycle in the test datasets. The overall minimum average RMSE per cycle for the HSC and LiB are reported to be 0.33% and 0.87% whereas the overall average RMSEs are 0.71% and 1.10%, respectively. The ESSs results are progressive in comparison with the existing literature and validate the efficacy of the multi-term flexible reward function approach.**

***Keywords – Electrochemical impedance spectroscopy; Reinforcement learning; Lithium-ion battery; Ensemble learning; State-of-charge; Hybrid Supercapacitor; Sustainable Cities and Communities.***

## I. INTRODUCTION

The shift towards sustainable cities and communities requires minimum usage of fuel-based vehicles and an encouraging leap towards electric vehicles (EVs) [1]. Moreover, a rapid decline in the availability of conventional energy sources, makes this shift obligatory. The major component within the EV is the energy storage system (ESS), which requires an efficient management system to deal with parameter estimations and cell balancing [2]. One of the most inevitable factors in the transformation from fuel-based vehicles to EVs, is the accurate estimation of the state of charge (SoC) of the ESS, given that an inaccurate estimation can cause unexpected and potentially dangerous real-time situations [3].

Among the various ESSs, the lithium-ion battery (LiB) is the most widely used and experimented on device [4]. For the SoC estimation of a LiB, an improved algorithm based on Coulomb Counting and uncertainty evaluation, considering a period of ten years, is presented in [5]. This mathematical model-based approach results in a percentage error of 0.3%. In [6], another SoC estimation technique based on an improved bidirectional gated recurrent unit (BGRU) and an unscented Kalman filter (UKF) is proposed for the SoC estimation of a LiB. The approach is designed for different temperatures and results in an RMSE less than 1.12%.

Another ESS, the supercapacitor (SC) stands out due to its high energy density, high power density, fast charging, rapid discharge and long cycle life [7]. A sub-type of SC, the hybrid supercapacitor (HSC) sits between a battery and a traditional SC, naturally uniting the qualities of both these devices [8]. Due to its advantages, the HSC is now more aggressively researched on for the integration with EVs. For the SoC estimation of an HSC, a UKF based on a fractional order model is presented in [9]. The algorithm provides an error of less than 2% considering several limiting factors and a diverse range of temperatures. Another article [10] proposes a switched system approach designed specifically for the RC model of the supercapacitor (SC), to be used mainly while working with cell imbalances within an ESS. This approach results in a tracking error of less than 1%. An online scheme for an accurate estimation of SoC of an SC is presented in [11]. This approach considers the leakage effect to ensure proper and efficient management of the SC. The technique results in a minimum absolute average percentage error of 0.35%, with a 20% uncertainty introduced in the equivalent series resistance of the SC. A gaussian process regression (GPR) based fast approach incorporating the data collected for a HSC, through electrochemical impedance spectroscopy (EIS) is presented in [12]. The approach offers an optimal balance between accuracy and timing by incorporating two separate frequency points for SoC within [0,30] and [30,100] respectively. The technique results in an average root mean square error (RMSE) of 0.94%.

The RL is an agent-based technique that, based on a reward function, takes consecutive steps, interacting with the environment, to maximize the designed reward [13]. The custom reward function is formulated keeping in view the ultimate objectives derived from the problem statement. In [14], a deep RL based approach for a nickel-metal hybrid battery is presented, to estimate the SoC of the battery. The deep deterministic policy gradient-based approach utilizes a basic error-based reward function and results in an estimation accuracy of 98.8%. The initial prediction often contains a residual error mainly present due to structural complication and algorithmic limitations [15]. This residual error can be rectified during the second phase based on a least square boosting (LSB) algorithm. LSB utilizes the residual errors from the first phase of the training to correct the difference that was not captured during the initial phase i.e. RL [16]. Following the LSB, a three-point linear interpolation technique is deployed, which directs whether to keep the predicted value or replace it with a computed value by setting a reasonable threshold.

Majority of the frameworks presented in literature regarding the SoC estimation for LiB and HSC are model based. Moreover, the domain of RL, in terms of SoC, still provides a substantial opportunity to be explored. Finally, the possibility to create a flexible approach to be easily adapted to different ESS and applications field has not been studied yet.

Hence, in this paper, a flexible data driven approach encompassing RL, LSB and the three-point linear interpolation has been proposed and implemented for two different ESSs, a LiB and an HSC. The aim is to design a custom reward function that, by means of proper parameters, could be adapted to different discharge profiles and different ESS.

The paper is organized as follows: in Section II the methodology for acquiring the data used in the training and test processes is detailed; in Section III the proposed framework is described; in Section IV the results are presented and discussed, and in Section V the conclusions are drawn.

## II. DATA ACQUISITION

### *A. Lithium-Ion Battery*

The data acquisition for the Panasonic 18650 LiB is conducted using a constant current constant voltage (CC-CV) procedure [17]. The numerical values for the current, voltage and temperature across multiple discharge cycles were obtained at ambient temperatures. Within, the selected current range [2,3] A, five different current profiles were acquired for each current rate at 0.1A steps. The training and testing are performed by randomizing and creating five randomized datasets (RDs), then dividing each into 80% training and 20% test sub-datasets. For each RD, the proposed framework is deployed and tested, resulting in a total of 5 training instances.

### *B. Hybrid Supercapacitor*

The dataset used in this work includes also EIS measurements acquired from a battery-like Hybrid Supercapacitor (HSC) tested under standard conditions. The HSC is fully charged, then after a suitable resting period it is discharged using a step-wise constant current profile. The resting period is used to ensure electrochemical stability, while the particular discharge profile allows to perform EIS measurement every 2% of Depth of Discharge (DoD). Potentiostatic EIS is performed at five different frequencies (i.e., 50 mHz, 100 mHz, 500 mHz, 1 Hz, 5 Hz) for multiple discharge cycles and are then used for training the proposed network. Three different RDs, has been created, with a 70-30% ratio for the training and test subsets respectively, resulting in a total of 15 training instances.

## III. PROPOSED FRAMEWORK

In this research work, a three-stage framework for the accurate SoC estimation of ESSs is presented. During the primary stage, RL is utilized to train on the dataset (LiB/HSC) as the primary estimation technique. The RL training is conducted with the flexible reward function and a twin-delayed deep deterministic neural network (TD3NN) agent with 50 hidden units, an LSTM based architecture, an actor optimizer learning rate of 0.001 sec and a critic optimizer learning rate of 0.01 sec, within the MATLAB RL application.

The agent attempts to maximize the tailored flexible reward function mentioned below in eqs. (1–6).

$$Reward = -\frac{|SoC_{tr} - SoC_{est}|}{E_{\mathrm{m}}} + B_s + B_d \tag{1}$$

$$B_s, \quad \omega_L \leq |SoC_{est} - SoC_{pr}| \leq \omega_H \tag{2}$$

$$\omega_L = \Delta_{min} - 1, \; \omega_H = \Delta_{max} + 1 \tag{3}$$

$$\Delta_{min} = \min\left[abs(SoC_{tr}(t) - SoC_{tr}(t-1))\right] \quad (4)$$

$$\Delta_{max} = \max\left[abs(SoC_{tr}(t) - SoC_{tr}(t-1))\right] \quad (5)$$

$$B_d, \ if \ SoC_{est} < SoC_{pr} \quad (6)$$

Where $SoC_{tr}$ is the observed value, $SoC_{est}$ is the estimated value at t and $SoC_{pr}$ is the estimated value at t-1. $\Delta_{min}$ and $\Delta_{max}$ are the absolute minimum and maximum step change between the true SoC at time t and t-1. The $B_s$ and $B_d$ are scalar values that provide additional reward spikes when the difference between the current and previous value is within the tolerance window $(\omega_L, \omega_H)$ and current SoC value is smaller than the previous value, respectively. The reward function is flexible, since the tolerance window is dependent on the particular ESS, and the particular discharge cycle (i.e. application context). This flexibility allows for the reuse of the same framework for different ESSs. The tolerance window is designed based on the difference between the consecutive values of the observed SoC.

Following the RL training, the LSB stage in introduced to compensate for any residual error [12]. The LSB utilizes different number of maximum tree splits and learning cycles, to compensate for the inadequacies left during the RL phase. Ultimately, the three-point linear interpolation is performed. This interpolation is deployed mainly to rectify any substantial deviations from the otherwise consistent pattern.

## IV. RESULTS AND DISCUSSIONS

The proposed framework is deployed for the accurate SoC estimation of three RDs at five different frequencies for the HSC and 5 RDs for the LiB. The training and tests are performed on MATLAB R2023b on an 8th generation, Intel core i9 with 16 GB RAM system. The performance metrics such as RMSE and mean absolute error (MAE), mean squared error (MSE) and Coefficient of Determination (R2) are computed and analysed. The results are summarized in Table 1. The obtained minimum average RMSE per cycle was 0.33% and 0.87% for the HSC and LiB, respectively. To further show the results, a plot between the true SoC (SoCT) and the estimated SoC (SoCE) of a single RD (RD # 2 @ 5 mHz) for the HSC is presented in Fig. 1. Similarly, the comparison plot for the RD # 1 of the LiB is presented in Fig. 2 below. Both the plots show promising estimation capabilities of the proposed flexible approach.

*Table 1. Average per cycle performance metrics for ESSs*

| ESS | RMSE | MAE | MSE | $R^2$ |
|---|---|---|---|---|
| **LiB** | 1.1041 | 0.8509 | 0.9620 | 0.9980 |
| **HSC** | 0.7161 | 0.5282 | 0.5879 | 0.9993 |

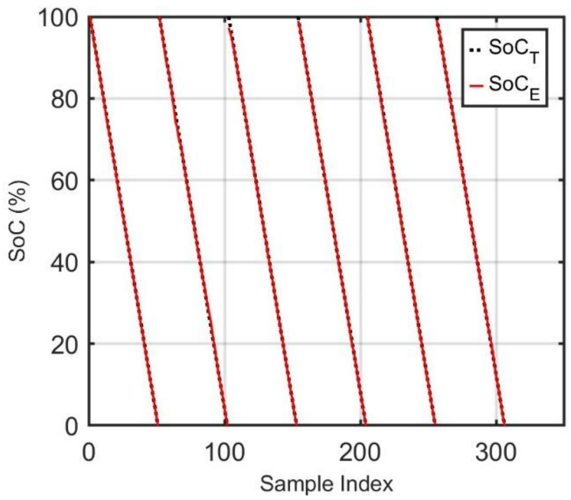


*Fig. 1. Example of the estimated SoC vs true SoC (HSC, RD # 2 @ 5 mHz)*

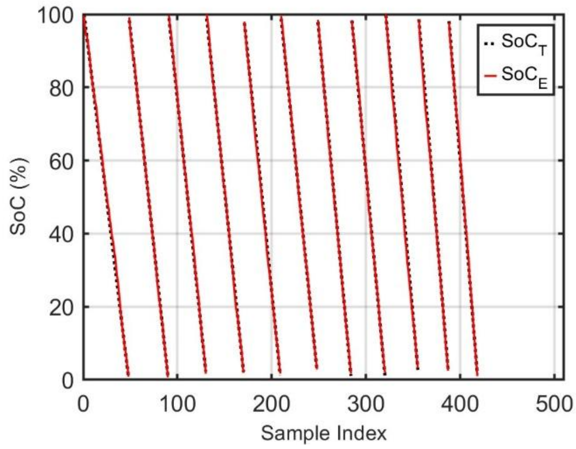


*Fig. 2. Example of the estimated SoC vs true SoC (LiB, RD # 1)*

*Table 2. Comparison – Existing Literature*

| ESS | Proposed Technique | RMSE |
|---|---|---|
| **LiB** | Flexible RL + LSB + interpolation | 1.10 |
| **LiB [6]** | BGRU + UKF | <1.12 |
| **HSC** | Flexible RL + LSB + interpolation | **0.71** |
| **HSC [9]** | UKF + particle swarm optimization | <2 |
| **HSC [12]** | RL + LSB + interpolation | 0.94 |

A comparison with existing literature is conducted to estimate the efficacy and efficiency of the proposed approach. The major details along with the performance metric are presented in Table 2.

The proposed framework provides the minimum RMSE for both the LiB and the HSC in comparison to the cited approaches. An interesting comparison, directly revealing the efficacy of the proposed framework is conducted with [12]. The cited approach utilizes a similar strategy and the same dataset, differing primarily in the reward function. It deploys an absolute negative reward, and the results reveal a difference of 0.23% in the RMSE, in favour of our proposed flexible reward approach.

## V. CONCLUSIONS AND OUTLOOK

This work presents a flexible three stage RL-based SoC estimation framework for ESSs, incorporating a tailored reward function. The presented framework is applied to one LiB and one HSC with different discharge cycles to prove its adaptable nature. The data extraction is performed through CC-CV and EIS for the LiB and HSC, respectively. The three stages include a primary RL prediction stage, a secondary LSB residual correction stage and an ultimate linear three-point interpolation stage. The RL training is performed with a tailored reward function consisting of additional bonus terms for a descending SoC pattern and to incorporate smoothness. The smoothness bonus is flexible in nature and dependent on the minimum and maximum difference between consecutive values of the observed SoC. A TD3NN agent is deployed for the RL training phase. The trained model from stage 1 is then used along with a LSB correction algorithm to ensure detection and correction of anomalies. Finally, a three-point linear interpolation is carried out to rule out any further discrepancies. 15 training and test runs for the HSC and 5 for the LiB were conducted. The resulting values for the performance metrics are recorded and analysed. This approach results in a minimum per cycle average RMSE of 0.33% and 0.87% for the HSC and LiB, respectively. The values of these performance metrics when compared with existing literature also prove that the proposed framework provides satisfactory results. The results validate the efficacy of the proposed flexible reward approach and provide a headstart towards replicating this technique for other ESSs as well. Moreover, this framework encourages future work and opens Potential extensions in the combined field of RL, LSB and ESSs.

## REFERENCES

[1] **R. Kumar, C. Bharatiraja, K. Udhayakumar, S. Devakirubakaran, S. Sathiya, and Lucian Mihet-Popa**, "Advances in Batteries, Battery Modeling, Battery Management System, Battery Thermal Management, SOC, SOH, and Charge/ Discharge Characteristics in EV Applications," *IEEE Access*, vol. 11, pp. 105761–105809, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3318121.

[2] **Popescu, A.**: *Electrical machines and drives*, Politehnium, Iasi, 2011, p. 234.

[3] **Curtley, M, Prince, O.A.**: A synchronous detector with improved parameters, *Proceeding of the International Symposium on Circuits and Systems*, Barcelona, July 15-17, 2009, pp. 255-260.

[4] **W. Chen, J. Liang, Z. Yang, and G. Li**, "A Review of Lithium-Ion Battery for Electric Vehicle Applications and Beyond," *Energy Procedia*, vol. 158, pp. 4363–4368, Feb. 2019, doi: https://doi.org/10.1016/j.egypro.2019.01.783.

[5] **F. Mohammadi**, "Lithium-ion battery State-of-Charge estimation based on an improved Coulomb-Counting algorithm and uncertainty evaluation," *Journal of Energy Storage*, vol. 48, p. 104061, Apr. 2022, doi: https://doi.org/10.1016/j.est.2022.104061.

[6] **Z. Cui, L. Kang, L. Li, L. Wang, and K. Wang**, "A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF," *Energy*, vol. 259, p. 124933, Nov. 2022, doi: https://doi.org/10.1016/j.energy.2022.124933.

[7] **J. Sun, B. Luo, and H. Li**, "A Review on the Conventional Capacitors, Supercapacitors, and Emerging Hybrid Ion Capacitors: Past, Present, and Future," *Advanced Energy and Sustainability Research*, vol. 3, no. 6, p. 2100191, Apr. 2022, doi: https://doi.org/10.1002/aesr.202100191.

[8] **D. P. Dubal, O. Ayyad, V. Ruiz, and P. Gómez-Romero**, "Hybrid energy storage: the merging of battery and supercapacitor chemistries," *Chemical Society Reviews*, vol. 44, no. 7, pp. 1777–1790, 2015, doi: https://doi.org/10.1039/c4cs00266k.

[9] **J. Zhang, B. Xiao, G. Niu, X. Xie, and S. Wu**, "Joint estimation of state-of-charge and state-of-power for hybrid supercapacitors using fractional-order adaptive unscented Kalman filter," *Energy*, vol. 294, p. 130942, May, doi: https://doi.org/10.1016/j.energy.2024.130942.

[10] **H. Li, W. He, X. Fang, and S. Li**, "State-of-Charge Estimation of Supercapacitors: A Switched Systems Approach," *IEEE Transactions on Transportation Electrification*, vol. 10, no. 3, pp. 5987–5996, Sep. 2024, doi: https://doi.org/10.1109/tte.2023.3328141.

[11] **P. Saha, S. Dey, and Munmun Khanra**, "Modeling and State-of-Charge Estimation of Supercapacitor Considering Leakage Effect," *IEEE transactions on industrial electronics*, vol. 67, no. 1, pp. 350–357, Jan. 2020, doi: https://doi.org/10.1109/tie.2019.2897506.

[12] **S. Ali, G. Patrizi, L. Ciani, I. De Munari, and V. Bianchi**, "Signal-to-Noise Ratio Analysis for a Reinforcement Learning based State of Charge Estimation of a Hybrid Supercapacitor," accepted for presentation at the *2026 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2026.

[13] **S. Gow, M. Niranjan, S. Kanza, and J. G. Frey**, "A review of reinforcement learning in chemistry," *Digital Discovery*, vol. 1, no. 5, pp. 551–567, 2022, doi: https://doi.org/10.1039/d2dd00047d.

[14] **I. Saba, M. Tariq, M. Ullah, and H. Vincent Poor**, "Deep reinforcement learning based state of charge estimation and management of electric vehicle batteries," *IET smart grid*, vol. 6, no. 4, pp. 422–431, Jun. 2023, doi: https://doi.org/10.1049/stg2.12110.

[15] **Emirhan Ilhan, A. B. Koc, and S. S. Kozat**, "Exploiting residual errors in nonlinear online prediction," *Machine Learning*, vol. 113, no. 9, pp. 6065–6091, May 2024, doi: https://doi.org/10.1007/s10994-024 06554-7.

[16] **V. Bianchi, F. Canzanella, S. Ali, I. De Munari, L. Ciani, and G. Patrizi**, "A single-point EIS measurement for SOC estimation of supercapacitor," *2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, May 2025, doi: https://doi.org/10.1109/i2mtc62753.2025.11078988.

[17] **V. Bianchi, M. Stighezza, A. Toscani, G. Chiorboli, and I. De Munari**, "An Improved Method Based on Support Vector Regression With Application Independent Training for State of Charge Estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023, doi: https://doi.org/10.1109/tim.2023.3306816.