Hybrid Physics-Informed UKF–Transformer Framework for Lithium-Ion Battery SOH Estimation

Abdel Rahman El Khatib[1,2], Khoury Boutrous [2], Ghaleb Hoblos [1], Kokou Langueh [2], Eric Duviella [2], and Jacques Boonaert [2]

[1] *University of Rouen Normandie, École Supérieure d'Ingénieurs en Génie Électrique (ESIGELEC), Normandie University, Institut de Recherche en Systèmes Électroniques Embarqués (IRSEEM), UR4353, F-76000 Rouen, France;*

[2] *IMT Nord Europe, Institute Mines-Télécom, University of Lille, Centre for Education, Research and Innovation in Digital Systems (CERI Digital Systems), F-59500 Douai, France;*

ABSTRACT

The accurate estimation of the State of Health (SOH) of lithium-ion batteries is essential for ensuring reliable operation and enabling prognosis in energy storage systems. Model-based approaches such as the Unscented Kalman Filter (UKF) provide physically interpretable estimates and stability properties that can be analyzed under standard modeling and noise assumptions. However, their performance is constrained by the need for an explicit and accurate battery model, as well as careful tuning of process and measurement noise covariances. As a result, standalone UKF implementations may struggle in the presence of nonlinear aging effects, parameter drift, and real-world operating variability. On the other hand, data-driven approaches—particularly transformer-based architectures—excel at modeling nonlinear systems and capturing long-range temporal dependencies. However, they typically require large and diverse datasets, are sensitive to the scenario distribution used during training, and may lack stability or physical interpretability. Despite these challenges, transformers can reduce the overall data requirement by efficiently learning both temporal relationships and cross-feature interactions within the input signals. In this context, we propose a hybrid physics-informed framework that combines a UKF with a transformer model. The UKF provides a physically grounded intermediate SOH estimate, while the transformer compensates for unmodeled dynamics and nonlinear degradation patterns. The transformer further improves the estimator's efficiency by capturing temporal dependencies and cross-feature relationships, which reduces the amount of training data required while maintaining robustness and generalization. Using the strengths of both techniques, the proposed hybrid approach improves estimation accuracy, robustness to noise, and generalization capability compared to either method used alone.

1. INTRODUCTION

Lithium-ion batteries have become the dominant energy storage systems in electric applications due to their higher power density and favorable cycling performance compared with other battery technologies. This increasing popularity is further driven by the growing use of electric vehicles and the integration of renewable energy sources in electrical grids (Yang, He, Ren, Gao, & Qi, 2025). Nevertheless, one of the main drawbacks of these batteries is that they require advanced management systems to avoid safety issues and to guarantee reliable operation under different operating conditions (Xiao et al., 2023). A battery management system (BMS) is mainly responsible for monitoring the battery states. Unfortunately, one of the key indicators—the state of health (SOH) of the battery—is not directly measured and must be estimated (Li, Song, & Wei, 2025). Several estimation techniques have been proposed in the literature for SOH and can be classified into three categories: model-based, data-driven, and hybrid techniques (Ahwiadi & Wang, 2025). While model-based techniques rely on explicit models that represent the electrochemical behavior of the battery, and data-driven techniques require extensive datasets for training, hybrid approaches aim to combine both. Given the difficulties of identifying the parameters of battery models and the lack of comprehensive datasets that cover all scenarios, hybrid models are gaining increasing popularity in the literature (El Khatib, Hoblos, Langueh, & Duviella, 2025).

An unscented Kalman filter has been proposed in (Liu et al., 2022) for SOH and SOC co-estimation. In this paper, a 2-RC Equivalent Circuit Model (ECM) is used, its parameters are identified online via a recursive least squares algorithm, and then SOC and SOH are estimated using the UKF. Al-

though this improves robustness compared to a standard UKF in their experiments, the approach still inherits the typical limitations of filter-based model-driven estimators: it relies on the validity of the equivalent circuit representation, depends on the identification quality of the ECM parameters, and its performance can degrade when real operating conditions introduce dynamics that are not captured by the ECM. In another paper (K. Zhang et al., 2025), a BIGRU–Transformer is proposed for SOH estimation to improve continuity capture in battery feature sequences, highlighting that a plain transformer may struggle with feature continuity. However, this model can be sensitive to noise in the training dataset, because it is relying solely on them. Other state of the art methods also use transformer for SOH estimation combining it other neural networks, such as Convolutional Neural Network (CNN) (Gui et al., 2025; Ding, Wu, & Li, 2025) and Long short term memory neural network (LSTM) (Gomez, Wang, & Chou, 2024). And such frameworks are also seen in RUL estimation (Zhao et al., 2025). Recent studies explicitly combined UKF-based filtering with transformer-style models in other domains. For instance, in (J. Tang, Sun, Zhang, Zhou, & Huang, 2025) introduced a Transformer–UKF algorithm for online positioning in horizontal directional drilling. In addition, a transformer is combined with a UKF for zero-shot state estimation in (Holtmann, Stenger, Posada-Moreno, Solowjow, & Trimpe, 2025), highlighting the benefit of pairing learned models with probabilistic filtering. Finally, in (J. Zhang, Dai, Zhou, Zhang, & Ma, 2025) a KUformer is proposed , which integrates a UKF-based component with a transformer architecture for battery SOH prediction. However, in this framework the UKF–transformer predictors often rely on the learned model as the primary estimator, which can reduce interpretability and make performance more sensitive to the training-data distribution under unseen operating conditions. On this basis, we propose a hybrid framework that relies on the UKF as a physically grounded baseline capacity estimator, which imposes the charge-balance (Coulomb-counting) equation as a constraint on the capacity evolution, and employs a four-attention-head transformer only as a residual correction module to compensate for model mismatch and reduce sensitivity to precise filter tuning.

The remainder of the paper is structured as follows: Section two presents the the Unscented Kalman Filter (UKF) in general, then its adaptation to the used battery model. Section three describes the transformer model. Section four is dedicated to the proposed hybrid framework and the obtained results. Finally, Section five concludes the paper.

2. THE UNSCENTED KALMAN FILTER AND PHYSICS-BASED CAPACITY MODELING

Filtering techniques have been widely used in the literature due to their recursive nature, and their ability to consider measurement and model uncertainties. The classical Kalman filter is used for linear systems, and given that battery degradation dynamics are not, Extended Kalman Filters were frequently used in the literature (Khatib, Hoblos, Langueh, & Duviella, 2025). Given that these EKF rely on the linearization of the model at each operating point using Jacobian transformations which complicates its application, and knowing that its accuracy decreases when nonlinearities are strong; unscented Kalman filters were proposed.

2.1. UKF Principles and Sigma-Point Transformation

The Unscented Kalman Filter (UKF) is a recursive Bayesian estimator designed for nonlinear state-space systems. Unlike the Extended Kalman Filter (EKF), which relies on local linearization via Jacobians, the UKF propagates a set of deterministically chosen samples (sigma points) through the nonlinear mappings without prior linerization. This procedure, known as the *Unscented Transformation* (UT), provides an accurate approximation of the transformed mean and covariance, typically up to second-order (and third-order for Gaussian variables), without requiring explicit derivatives. In this next subsections, the UKF will be presented in general, and later it will be adapted to our battery case.

2.1.1. Nonlinear state-space formulation

Let us consider the discrete nonlinear system:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k, \tag{1}$$

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{v}_k, \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{u}_k$ is the input, and $\mathbf{z}_k$ is the measured output of the system. The process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are assumed zero-mean with covariances $\mathbf{Q}$ and $\mathbf{R}$, respectively.

2.1.2. Sigma-point construction

Given the prior mean $\hat{\mathbf{x}}_k$ and covariance $\mathbf{P}_k$, the UKF constructs $2n+1$ sigma points $\{\boldsymbol{\chi}_k^{(i)}\}_{i=0}^{2n}$ as:

$$\boldsymbol{\chi}_k^{(0)} = \hat{\mathbf{x}}_k, \tag{3}$$

$$\boldsymbol{\chi}_k^{(i)} = \hat{\mathbf{x}}_k + \gamma \left[\mathbf{S}_k\right]_{(:,i)}, \quad i = 1, \ldots, n, \tag{4}$$

$$\boldsymbol{\chi}_k^{(i+n)} = \hat{\mathbf{x}}_k - \gamma \left[\mathbf{S}_k\right]_{(:,i)}, \quad i = 1, \ldots, n, \tag{5}$$

where $\mathbf{S}_k$ is the lower-triangular matrix from the Cholesky decomposition $\mathbf{P}_k = \mathbf{S}_k \mathbf{S}_k^\top$, and

$$\gamma = \sqrt{n + \lambda}, \qquad \lambda = \alpha^2(n + \kappa) - n. \tag{6}$$

The scaling parameters $\alpha \in (0, 1]$, κ, and β (used to calculate the weights in the next subsection) control the spread of sigma points and incorporate prior distribution knowledge (usually $\beta = 2$ for Gaussian variables).

2.1.3. Unscented weights

For each sigma point we assign a mean weight $W^{(m)}$ and a covariance weight $W^{(c)}$:

$$W_0^{(m)} = \frac{\lambda}{n+\lambda}, \tag{7}$$

$$W_0^{(c)} = \frac{\lambda}{n+\lambda} + (1-\alpha^2+\beta), \tag{8}$$

$$W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n+\lambda)}, \quad i = 1,\dots,2n. \tag{9}$$

2.1.4. Prediction through the process model

The sigma points are propagated through the nonlinear process model:

$$\boldsymbol{\chi}_{k+1|k}^{(i)} = f\left(\boldsymbol{\chi}_k^{(i)}, \mathbf{u}_k\right), \qquad i = 0,\dots,2n. \tag{10}$$

The predicted state mean and covariance are then computed as:

$$\hat{\mathbf{x}}_{k+1|k} = \sum_{i=0}^{2n} W_i^{(m)}\, \boldsymbol{\chi}_{k+1|k}^{(i)}, \tag{11}$$

$$\mathbf{P}_{k+1|k} = \sum_{i=0}^{2n} W_i^{(c)} \left(\boldsymbol{\chi}_{k+1|k}^{(i)} - \hat{\mathbf{x}}_{k+1|k}\right)\left(\boldsymbol{\chi}_{k+1|k}^{(i)} - \hat{\mathbf{x}}_{k+1|k}\right)^\top + \mathbf{Q}. \tag{12}$$

2.1.5. Prediction through the measurement model

The predicted measurement sigma points are obtained by passing the predicted sigma points through $h(\cdot)$:

$$\mathbf{z}_{k+1|k}^{(i)} = h\left(\boldsymbol{\chi}_{k+1|k}^{(i)}, \mathbf{u}_k\right), \qquad i = 0,\dots,2n. \tag{13}$$

Similar to before the predicted measurement mean and covariance are:

$$\hat{\mathbf{z}}_{k+1|k} = \sum_{i=0}^{2n} W_i^{(m)}\, \mathbf{z}_{k+1|k}^{(i)}, \tag{14}$$

$$\mathbf{P}_{zz} = \sum_{i=0}^{2n} W_i^{(c)} \left(\mathbf{z}_{k+1|k}^{(i)} - \hat{\mathbf{z}}_{k+1|k}\right)\left(\mathbf{z}_{k+1|k}^{(i)} - \hat{\mathbf{z}}_{k+1|k}\right)^\top + \mathbf{R}. \tag{15}$$

Then the cross-covariance between the state and measurement is obtained as:

$$\mathbf{P}_{xz} = \sum_{i=0}^{2n} W_i^{(c)} \left(\boldsymbol{\chi}_{k+1|k}^{(i)} - \hat{\mathbf{x}}_{k+1|k}\right)\left(\mathbf{z}_{k+1|k}^{(i)} - \hat{\mathbf{z}}_{k+1|k}\right)^\top. \tag{16}$$

2.1.6. Correction step

Finally, the Kalman gain, state update, and covariance update are given by:

$$\mathbf{K}_{k+1} = \mathbf{P}_{xz}\mathbf{P}_{zz}^{-1}, \tag{17}$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}\left(\mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k}\right), \tag{18}$$

$$\mathbf{P}_{k+1} = \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1}\mathbf{P}_{zz}\mathbf{K}_{k+1}^\top. \tag{19}$$

Finally, the sigma-point mechanism is used to propagate uncertainty (noises) through the nonlinear mappings involved in battery SOH estimation. The exact state definition and the selected process/measurement functions $f(\cdot)$ and $h(\cdot)$ depend on the adopted capacity/SOH model, which is presented in the next subsection.

2.2. State-Space Formulation for Capacity/SOH Estimation

In general, the State of Health (SOH) can be defined using the resistance or the capacity of the battery (K. Tang et al., 2025; Li et al., 2025). In our case, we are using a capacity-based SOH. For instance, let $Q_{\mathrm{cap},k}$ denote the available battery capacity at time step k (in Ah), and let Q_{nom} denote the nominal (rated) capacity. The SOH becomes:

$$\mathrm{SOH}_k = \frac{Q_{\mathrm{cap},k}}{Q_{\mathrm{nom}}}, \tag{20}$$

where $Q_{\mathrm{nom}} = 27\ \mathrm{Ah}$ in our case. The objective of the estimator is to track $Q_{\mathrm{cap},k}$ online, from which SOH_k is directly obtained via (20). In the proposed framework, SOC_k is assumed to be available (esimtated by an EKF) and is therefore used as an input rather than being co-estimated within our framework.

Based on that, the UKF state is chosen as the available capacity:

$$x_k = Q_{\mathrm{cap},k} \in \mathbb{R}^+, \tag{21}$$

and it is constrained to a physically meaningful interval, such as:

$$Q_{\min} \le Q_{\mathrm{cap},k} \le Q_{\max}, \tag{22}$$

with $Q_{\min} = 0.3Q_{\mathrm{nom}}$ and $Q_{\max} = Q_{\mathrm{nom}}$ which reflects the actual performance of the battery in real-world.

And because we know that the capacity fade is a slow phenomenon compared with the sampling period, the capacity dynamics are modeled as a random walk, instead of the equations in (12):

$$x_{k+1} = f(x_k) + w_k, \qquad f(x_k) = x_k, \tag{23}$$

where $w_k \sim \mathcal{N}(0, Q)$ is a zero-mean Gaussian process noise capturing gradual degradation and unmodeled capacity variations. This choice avoids enforcing a specific aging law and allows the estimator to adapt to different operating regimes.

To construct a measurement equation for capacity, the discrete-time SOC balance based on Coulomb counting is considered as:

$$\mathrm{SOC}_k = \mathrm{SOC}_{k-1} - \frac{I_k \Delta t}{3600\, Q_{\mathrm{cap},k}}, \tag{24}$$

where I_k is the current (A), Δt is the sampling time (s), and the factor 3600 converts ampere-seconds to ampere-hours. It is important to note that according to this equation a positive current is a discharge current. Defining the estimated SOC increment as:

$$z_k = \Delta\mathrm{SOC}_k = \mathrm{SOC}_k - \mathrm{SOC}_{k-1}, \tag{25}$$

yields the nonlinear measurement model as:

$$z_k = h(x_k, I_k) + v_k, \qquad h(x_k, I_k) = -\frac{I_k \Delta t}{3600\, x_k}, \tag{26}$$

where $v_k \sim \mathcal{N}(0, R)$ is a zero-mean Gaussian measurement noise affecting $\Delta\mathrm{SOC}_k$. Therefore, the UKF uses SOC_k and SOC_{k-1} to form z_k, and updates the capacity estimate by propagating sigma points through the nonlinear function $h(\cdot)$ in (26).

At each time step, the estimator inputs are SOC_k, I_k, and Δt, together with an initial capacity guess $Q_{\mathrm{cap},0}$ at initialization. The UKF outputs the updated capacity estimate $\hat{Q}_{\mathrm{cap},k}$, its associated covariance P_k, and the corresponding SOH estimation becomes:

$$\widehat{\mathrm{SOH}}_k = \frac{\hat{Q}_{\mathrm{cap},k}}{Q_{\mathrm{nom}}}. \tag{27}$$

In practice, $\Delta\mathrm{SOC}_k$ becomes less informative when SOC is saturated near 0 or 1, when SOC exhibits discontinuities (reset/initialization), or when the current magnitude is too small to provide sufficient excitation. For this reason, measurement updates are gated in such cases, and only the prediction step (23) is applied while the covariance is inflated by the process noise. Finally, the estimate is clamped to the bounds in (22) to preserve physical consistency and numerical stability.

3. TRANSFORMER-BASED SOH CORRECTION MODEL

While the UKF provides a physically grounded intermediate estimate of battery capacity (and thus SOH), its accuracy can degrade under model mismatch, nonlinear degradation effects, and operating variability. To compensate for these unmodeled dynamics, a transformer encoder is employed as a correction module. The transformer learns a data-driven mapping from a short temporal window of measured signals and UKF-related quantities to a refined capacity estimate. The corrected SOH is then obtained from the refined capacity through the capacity-based definition.

3.1. Input features and target.

At each time step k, the transformer receives a sequence of length L formed from a set of features that combine measurements, operating conditions, and UKF outputs. In the framework we chose at each time sample the feature vector as:

$$\mathbf{f}_k = \left[I_k,\ V_k,\ T_k,\ SOC_k,\ \widehat{Q}_k^{\mathrm{UKF}},\ P_k^{\mathrm{UKF}},\ c_k \right]^\top, \tag{28}$$

where I_k and V_k denote measured current and voltage, T_k is the measured temperature, SOC_k is the estimated SOC estimate (by an EKF), $\widehat{Q}_k^{\mathrm{UKF}}$ is the UKF capacity estimate, P_k^{UKF} is the corresponding UKF covariance (uncertainty), and c_k is the cycle. And as the transformer is not meant to look at a single instant and it learns from the history of the data, a sliding window is constructed as:

$$\mathbf{X}_k = [\mathbf{f}_{k-L+1}, \mathbf{f}_{k-L+2}, \ldots, \mathbf{f}_k] \in \mathbb{R}^{L \times d}, \tag{29}$$

with $L = 50$ and $d = 7$ in this work. The supervised learning target is the reference capacity at the end of the window,

$$y_k = Q_k^{\mathrm{true}}, \tag{30}$$

and the transformer is trained to get $\widehat{Q}_k^{\mathrm{Tr}} \approx Q_k^{\mathrm{true}}$ given the input $\mathbf{X}_k$. Now, when $k < L$ the transformer correction is not applied, therefore the UKF estimate is the same as the corrected one. In addition, it is important to note that to avoid conflict across different training scenarios, sequences are built independently within each dataset. More precisely, for each dataset group, samples are ordered by time, and a sliding window of length L is applied. Only windows that are fully contained within the same dataset are used for training, which ensures that the model does not get confused by data from different scenarios. And a final remark is that this procedure is meant to be done offline and the true capacity should be measured using Electrochemical spectroscopy (EIS) or any other direct measurement method to get the ground truth for the training phase.

3.2. The Architecture of the Transformer

Figure 1 illustrates the transformer used for the correction of the estimated capacity. The network is a lightweight transformer encoder followed by a pooling and regression head. At each time step, an input window $\mathbf{X}_k \in \mathbb{R}^{L \times d}$ (with $L = 50$ and $d = 7$ features) is first projected to an embedding space of dimension $d_{\text{model}} = 64$ through a linear layer. The choice of the window length $L = 50$ is made as a trade-off between providing sufficient temporal context for capturing short-term current/voltage response patterns and keeping the correction model lightweight; with a sampling time $T_s = 1$ s, this corresponds to a 50 s history window. A sinusoidal positional encoding is then added to retain temporal order which is a powerful characteristic of the transforms in general. The embedded sequence is processed by a single transformer en-

coder block composed of multi-head self-attention (queries Q, keys K, and values V are all derived from the same input sequence) and a position-wise feed-forward network, each wrapped with residual connections and layer normalization. Finally, global average pooling aggregates the temporal information into a fixed-length representation that is mapped by a linear regressor to the corrected capacity estimate $\widehat{Q}_k^{\mathrm{Tr}}$ (and hence $\widehat{\mathrm{SOH}}_k^{\mathrm{Tr}} = \widehat{Q}_k^{\mathrm{Tr}}/Q_{\mathrm{nom}}$). In this transformer, we chose

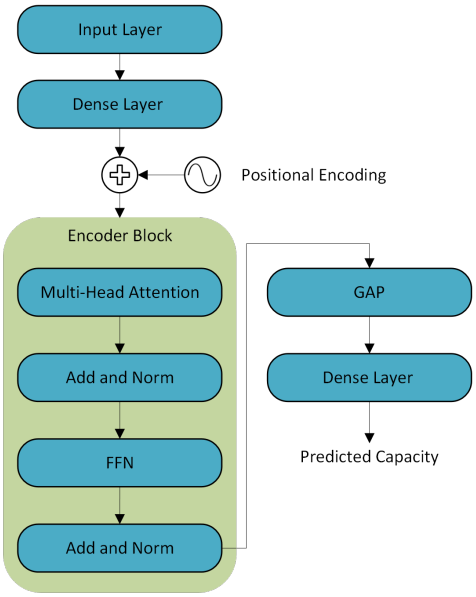

Figure 1. The architecture of the proposed transformer

empirically to put $H = 4$ attention heads, which provides a good trade-off between representational capacity and computational time. Given $d_{\mathrm{model}} = 64$, this corresponds to a per-head key dimension of $d_k = d_{\mathrm{model}}/H = 16$, which was sufficient to capture correlations among the input features without over-parameterizing the model. The feed-forward network (FFN) uses a hidden dimension of $d_{\mathrm{ff}} = 128$ with ReLU activation functions, enabling a nonlinear refinement of the attention output while keeping the overall architecture lightweight and stable for training. As the transformer is inspired from (Vaswani et al., 2023), and its parameters and size are chosen empirically to get the validation results for our case where we have timeseries as input, they can be further optimized.

3.3. Training objective and optimization

The transformer is trained using the classical mean-squared error (MSE) loss between the predicted and reference capacities:

$$\mathcal{L} = \frac{1}{N}\sum_{k=1}^{N}\left(\widehat{Q}_k^{\mathrm{Tr}} - Q_k^{\mathrm{true}}\right)^2. \tag{31}$$

The model is optimized using the Adam optimizer. The overall dataset contains $N = 1{,}619{,}559$ samples (windows). A $70/30$ train/validation split is used, with batch size 32 and a small number of epochs (4) as an initial configuration. This number can be also be optimized, but for the proof of concept we chose it because it allows us to see the improvement in the estimation while still requiring relatively small training time ($\approx$ less than 1 hour).

Once $\widehat{Q}_k^{\mathrm{Tr}}$ is obtained, the corrected SOH is computed directly as

$$\widehat{\mathrm{SOH}}_k^{\mathrm{Tr}} = \frac{\widehat{Q}_k^{\mathrm{Tr}}}{Q_{\mathrm{nom}}}. \tag{32}$$

This corrected estimate is used as the final SOH output of the transformer.

4. RESULTS OF THE HYBRID APPROACH

In this section, we will present results that show the validity of the proposed hybrid UKF–Transformer framework for capacity and SOH estimation. The concept is to compare the estimation obtained using the UKF alone with the ones corrected using the transformer. It is important to note that the final hybrid output is defined as

$$\widehat{Q}_k^{\mathrm{Hybrid}} = \begin{cases} \widehat{Q}_k^{\mathrm{UKF}}, & k < L, \\ \widehat{Q}_k^{\mathrm{Tr}}, & k \geq L, \end{cases} \tag{33}$$

where $L = 50$ is the transformer window length as discussed earlier. This switching reflects the fact that the transformer requires a full history window before producing a prediction; therefore, the hybrid estimate is identical to the UKF at the beginning of each sequence.

For the evaluation metric of this method, we use the root-mean-square error (RMSE). Given that Q_k^{true} denotes the reference capacity provided by the dataset and the estimation error is defined as $e_k = Q_k^{\mathrm{true}} - \widehat{Q}_k$, the RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} e_k^2}, \tag{34}$$

and is computed for both the standalone UKF and the hybrid framework. It is important to mention that to represent the battery we are using a physics table-based model from Matlab (MathWorks, 2026). Since the dataset is generated in simulation, it may not capture all real-world effects (sensor bias, hysteresis, and cell-to-cell variability); however, a major advantage is that it provides controlled and repeatable operating conditions, which allows systematic benchmarking of the proposed estimator under different current and temperature scenarios while considering the variations in the internal parameters of the battery.

For the UKF, we are considering three sigma points because $n = 1$ in the adopted capacity state-space model. The sigma-point parameters follow standard UKF practice: a small $\alpha =$

Table 1. Training operating scenarios used to build the transformer dataset.

Scenario	Temperature (°C)	Nominal current (A)
1	4.85	5
2	19	5
3	40	5
4	4.85	15
5	19	15
6	40	15
7	4.85	40
8	19	40
9	40	40

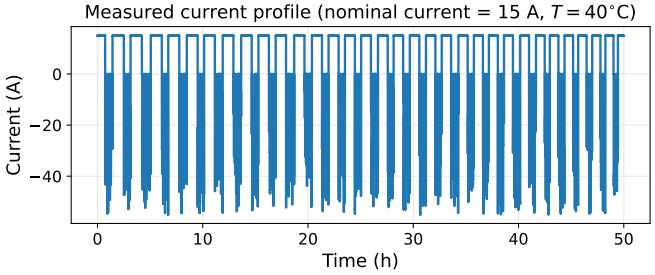

Figure 2. Measured current profile for the training scenario with nominal current 15 A at $T = 40°\text{C}$.

10^{-3} keeps the sigma points close to the mean to ensure a stable local approximation, $\beta = 2$ is a standard value under approximately Gaussian uncertainty assumptions (Wan & Van Der Merwe, 2000), and $\kappa = 0$ avoids excessively large spreads around the mean. Regarding the tuning of the noise covariances, the process noise $Q_{\text{proc}} = 3 \times 10^{-4}$ is selected small to reflect the fact that capacity fade is a slow phenomenon and is modeled as a random walk, while the measurement noise $R_{\text{meas}} = 5 \times 10^{-7}$ accounts for uncertainty in the SOC increment ΔSOC_k used as the measurement. Finally, the initial covariance $P_0 = 10^{-3}$ is intentionally chosen larger than extremely small values to prevent overconfidence at initialization and to ensure that the unscented Kalman gain remains sufficiently large for the estimate to move away from the initial capacity guess once new measurements are available.

After establishing this baseline, the transformer was trained on multiple scenarios that have different current profiles and temperature conditions. All data were normalized using the min-max standard method. These scenarios are summarized in Table 1. The current profile is produced by an SOC-based switching mechanism, where the values of the table 1 defines the nominal discharge current amplitude, and a complementary negative (charging) current is applied once SOC reaches a lower threshold to enforce repeated charge–discharge cycling, an example for a nominal current of 15 A is illustrated in Figure 2.

To ensure a fair evaluation of generalization, the test scenario corresponds to a distinct operating condition (nominal current

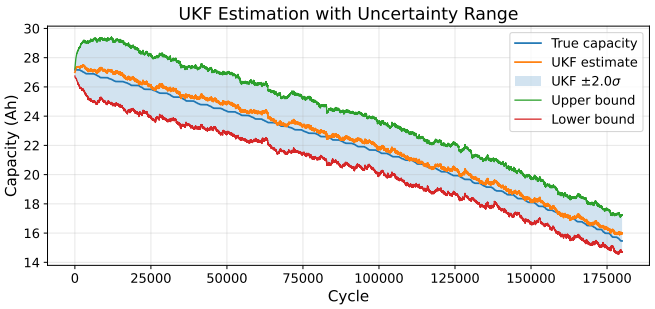

Figure 3. UKF capacity estimation with an uncertainty band computed from the UKF covariance P_k as $\pm 2\sigma$ around the estimate.

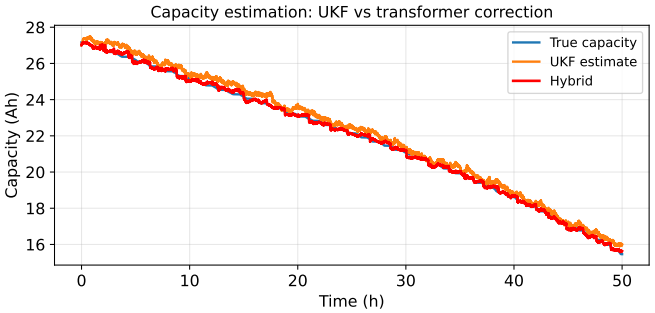

Figure 4. Capacity estimation results: true capacity, UKF estimate, and the UKF-transformer output.

15 A at $T = 40°\text{C}$) not used during training. The UKF parameters and nominal capacity were kept fixed across all scenarios, such that performance differences reflect the benefit of the learned correction rather than changes in the physics-based estimator. In Figure 3, the capacity estimated by the UKF is shown. The UKF managed to estimate the battery capacity with an RMSE of 0.3475 Ah, and the true capacity remains within the UKF uncertainty bounds throughout the entire test sequence.

For this very same test scenario, the Hybrid UKF-transformer is then applied. The Figure 4 shows the difference between the capacity estimated using the UKF and the UKF-Transformer. Finally, the Figure 5 shows that the variation of the estimation error of the hybrid framework is better than the one obtained using the UKF alone.

For both estimators, the RMSE is calculated as previously mentioned in Eq. (34). The obtained results are presented in Table 2. To assess generalization beyond a single operating condition, the hybrid estimator was evaluated on two additional unseen scenarios: a current-interpolation case $(25\ \text{A}, 40°\text{C})$ and a temperature-interpolation case $(40\ \text{A}, 25°\text{C})$. As shown in Table 2, the transformer correction consistently reduces the capacity estimation error across scenarios, yielding an RMSE reduction of approximately 70% relative to the UKF baseline. Taking into account the small number of training epochs and the limited amount of

Table 2. RMSE comparison between the standalone UKF and the hybrid UKF–Transformer estimator across unseen test scenarios (evaluated for $k \geq L$).

Test scenario	UKF RMSE (Ah)	Hybrid RMSE (Ah)
(25 A, 25°C)	0.3475	0.1045
(25 A, 8°C)	0.2477	0.1075
(40 A, 25°C)	0.3082	0.0846

used training scenarios shows that this method has great potentials in improving SOH estimation accuracy of batteries. An additional advantage of the proposed hybrid design is that it reduces the dependency on highly accurate UKF tuning. In a standalone UKF, the estimation quality is strongly influenced by the selection of the process and measurement noise covariances, since suboptimal $(Q_{\text{proc}}, R_{\text{meas}})$ choices can lead to slow adaptation, bias, or overconfidence. In the hybrid framework, the transformer exploits the availability of data collected under multiple operating scenarios to increase the robustness of the whole estimator against this problem. As a consequence, even when the UKF is tuned conservatively, the data-driven correction can recover part of the lost accuracy by compensating for model mismatch. This shifts the role of the UKF from being the sole source of accuracy to being a reliable informed baseline, while the transformer uses data to refine the estimate. Nevertheless, there are two main challenges that face this proposed approach. The first is the fact that the hybrid framework needs an Electrochemical spectroscopy (EIS) to get the real capacity before the training phase. Second, we know that data-driven techniques are not physically constrained; in other terms, it can give obsolete estimations. However, this can be fixed by imposing an external saturation to prevent unrealistic estimations or changes in the estimated capacity.

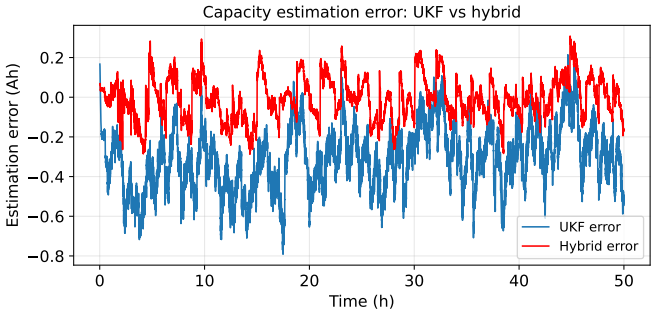

Figure 5. Capacity estimation error comparison between the UKF and the hybrid UKF–Transformer estimator.

5. CONCLUSION

This paper proposed a hybrid physics-informed framework for capacity-based battery SOH estimation that combines the stability and interpretability of a model-based Unscented Kalman Filter (UKF) with the learning capability of a lightweight transformer encoder. The UKF provides an intermediate capacity estimate based on a battery model and is maintained with fixed parameters across all operating scenarios, which preserves consistency and avoids scenario-specific retuning. Building on this baseline, the transformer is trained to correct residual errors by exploiting short temporal windows of measured signals together with UKF-related preliminary estimates, including the filter uncertainty.

The obtained results on an unseen operating condition confirm the validity of the hybrid estimator. While the UKF alone follows the global degradation trend, its accuracy is limited by model mismatch, indirect observability through SOC variations, and sensitivity to covariance tuning. In contrast, the proposed UKF-Transformer approach reduces the capacity estimation error and produces a corrected SOH trajectory that better matches the actual degradation of the battery. Importantly, the data-driven correction alleviates the need for aggressive or highly precise UKF tuning as well as the need for highly complicated physical models that add a heavy computational load, since systematic residual dynamics can be learned from data while the UKF uses relatively simplified model.

Future work will focus on extending the validation to larger and more diverse datasets, including different aging profiles and real experimental measurements, as well as investigating multi-cell generalization and transfer learning across chemistries. In addition, exploring the capacity of such method in module to pack health estimation is also interesting, as it can serve packs serviceability which is becoming more and more popular.

ACKNOWLEDGMENT

This Ph.D. work was funded by the Normandy Region and IMT Nord Europe.

REFERENCES

Ahwiadi, M., & Wang, W. (2025, January). Battery Health Monitoring and Remaining Useful Life Prediction Techniques: A Review of Technologies. *Batteries*, *11*(1), 31.

Ding, H., Wu, J., & Li, J. (2025, April). CNN-transformer based SOH prediction study for lithium batteries. In *2025 4th International Conference on Green Energy and Power Systems (ICGEPS)* (pp. 286–289). Hangzhou, China: IEEE.

El Khatib, A. R., Hoblos, G., Langueh, K., & Duviella, E. (2025, November). From First Life to Second Life: Advances and Research Gaps in Prognosis Techniques for Lithium-Ion Batteries. *Applied Sciences*, *15*(22), 12171.

Gomez, W., Wang, F.-K., & Chou, J.-H. (2024, June). Li-

ion battery capacity prediction using improved temporal fusion transformer model. *Energy*, *296*, 131114.

Gui, X., Du, J., Wang, Q., Zhao, H., Cheng, Y., & Zhao, J. (2025, September). Multi-modal data information alignment based SOH estimation for lithium-ion batteries using a local–global parallel CNN-Transformer Network. *Journal of Energy Storage*, *129*, 117178.

Holtmann, T., Stenger, D., Posada-Moreno, A., Solowjow, F., & Trimpe, S. (2025). *Sailing towards zero-shot state estimation using foundation models combined with a ukf.*

Khatib, A. R. E., Hoblos, G., Langueh, K., & Duviella, E. (2025, July). Robust Polytopic Luenberger Observer for SOC Estimation of a Battery Cell. In *2025 International Conference on Control, Automation and Diagnosis (ICCAD)* (pp. 1–6). Barcelona, Spain: IEEE.

Li, Q., Song, R., & Wei, Y. (2025, May). A review of state-of-health estimation for lithium-ion battery packs. *Journal of Energy Storage*, *118*, 116078.

Liu, S., Dong, X., Yu, X., Ren, X., Zhang, J., & Zhu, R. (2022, November). A method for state of charge and state of health estimation of lithium-ion battery based on adaptive unscented Kalman filter. *Energy Reports*, *8*, 426–436.

MathWorks. (2026). *Battery capacity estimation.* . (Accessed: 2026-02-10)

Tang, J., Sun, C., Zhang, Y., Zhou, W., & Huang, Y. (2025, September). Research on Online Positioning Method for Horizontal Directional Drilling Bits Based on Transformer-UKF Algorithm. *Journal of Physics: Conference Series*, *3106*(1), 012006.

Tang, K., Luo, B., Chen, D., Wang, C., Chen, L., Li, F., … Wang, C. (2025, August). The State of Health Estimation of Lithium-Ion Batteries: A Review of Health Indicators, Estimation Methods, Development Trends and Challenges. *World Electric Vehicle Journal*, *16*(8), 429.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2023, August). *Attention Is All You Need.* arXiv. (arXiv:1706.03762 [cs])

Wan, E., & Van Der Merwe, R. (2000). The unscented kalman filter for nonlinear estimation. In *Proceedings of the ieee 2000 adaptive systems for signal processing, communications, and control symposium (cat. no.00ex373)* (p. 153-158).

Xiao, Y., Wen, J., Yao, L., Zheng, J., Fang, Z., & Shen, Y. (2023). A comprehensive review of the lithium-ion battery state of health prognosis methods combining aging mechanism analysis. *Journal of Energy Storage*, *65*, 107347.

Yang, L., He, M., Ren, Y., Gao, B., & Qi, H. (2025, November). Physics-informed neural network for co-estimation of state of health, remaining useful life, and short-term degradation path in Lithium-ion batteries. *Applied Energy*, *398*, 126427.

Zhang, J., Dai, X., Zhou, R., Zhang, W., & Ma, H. (2025). Kuformer: Kan-ukf-based transformer for battery soh prediction. In *Iecon 2025 – 51st annual conference of the ieee industrial electronics society* (p. 1-6).

Zhang, K., Jiao, Z., Ma, J., Zhao, X., Wang, S., Yu, M., & Li, S. (2025, October). SOH estimation of battery based on the BIGRU-transformer algorithm considering the hierarchical voltage interval characterization of the discharge curve. *Journal of Energy Storage*, *132*, 117770.

Zhao, T., Zhang, Y., Wang, M., Feng, W., Cao, S., & Wang, G. (2025, August). A hybrid LSTM–transformer model for accurate remaining useful life prediction of lithium-ion batteries. *Frontiers in Electronics*, *6*, 1654344.