A Physics-Informed Neural Network for SoC and SoH Estimation in Solar-Charged Li-ion Batteries Under Partial Cycling

Param Bhimani
Jumeirah College, Dubai, UAE

I. ABSTRACT

Accurate estimation of State of Charge (SoC) and State of Health (SoH) is essential for lithium-ion batteries in solar energy storage systems where partial and irregular cycles are common. Traditional methods like Coulomb counting or Extended Kalman Filters sometimes struggle under dynamic conditions, and only data-driven approaches are not robust when usage patterns differ from training data. This paper proposes a PiNN framework that embeds electrochemical equations into a neural architecture to improve SoC and SoH estimation for solar-charged Li-ion batteries that undergo partial cycling.

We construct a synthetic solar charging scenario by using real solar irradiance profiles in combination with publicly available battery datasets, creating incomplete charge/discharge patterns. We then compare the PiNN's performance against classical and purely data-driven models, showing significant improvements in estimation accuracy under fluctuating charging conditions. These results indicate that the PiNN method promises a realistic solution for real-time, dependable battery management in storage systems for renewable energy, ultimately facilitating the transition to cleaner power systems.

Keywords: Physics-Informed Neural Networks (PiNN), Lithium-Ion Batteries, Battery Management Systems (BMS), Partial Cycling

II. INTRODUCTION

The rapid growth of solar photovoltaic (PV) technology has positioned it as a key feature of the world's movement towards clean energy sources. However, the irregularity of solar energy caused by weather patterns, seasonality, and diel light cycles requires the embracement of suitable energy storage approaches to achieve reliability and uninterrupted supply. Lithium-ion batteries have become a promising option for solar energy storage systems because of their high energy density and declining expenses. But unlike the electric vehicle or typical laboratory test cases, batteries in solar applications rarely see full, uniform charge and discharge cycles. Partial and irregular cycling is the norm instead, presenting new challenges for Battery Management Systems (BMS) that are supposed to precisely gauge the battery's State of Charge (SoC) and State of Health (SoH). These partial cycles, when not duly accounted for, may contribute to inefficient use of the batteries, erratic degradation rates, and consequently increased operational expenses. The demand is therefore increasing for sophisticated estimation methods capable of managing such varying conditions while maintaining battery performance and longevity in solar-powered applications.

A considerable obstacle to solar energy storage systems is the variable charging behavior due to changing sunlight, which tends to cause incomplete or partial cycling. This situation is rather contrary to controlled laboratory testing, where batteries are subjected to systematic and complete charge–discharge cycling. Incomplete cycling can trigger nonlinear aging effects, as the degradation rates of electrode materials differ based on the depth of discharge, ambient temperature, and rest time. Basic coulomb counting algorithms assume more symmetrical cycling and need to be reset frequently, leading to compounded errors when the battery rarely, if ever, returns to a full charge or discharge state.

Kalman filter approaches [1] (e.g., Extended Kalman Filter), on the other hand, are founded on simplified battery models that may struggle with the dynamic, partial cycles that occur in solar-charged systems. Purely data-driven methods can learn precisely from historical data but do not generalize well if real-world usage deviates significantly from the training data. Therefore, there is a growing need for hybrid methods that combine physical constraints with data-driven learning for robust State of Charge (SoC) and State of Health (SoH) estimation under the varying conditions characteristic of solar-powered systems.

Beyond estimation accuracy alone, practical deployment considerations further motivate the use of hybrid approaches. Solar energy storage systems operate under constrained computational resources and highly variable conditions, requiring models that are both efficient and robust. Integrating physical insight within a learning-based framework therefore offers a balanced strategy for achieving reliable state estimation in real-world renewable energy applications. Moreover, embedding governing constraints can reduce long-term estimation drift during extended partial cycling, improving consistency without significantly increasing computational overhead. Such characteristics are particularly important for distributed solar storage installations, where maintenance opportunities are limited and dependable autonomous operation is essential.

TABLE I: Summary of Representative Studies on Li-ion Battery State Estimation

Reference	Method	Key Findings	Limitations	Relevance
Yun et al. (2023) [2]	Adaptive EKF	Reduces RMSE by 49%, improves SoC estimation by dynamically updating parameters based on SoC segmentation.	Still based on ECM; oversimplifies electrochemical effects. Not validated for partial cycling.	Baseline for adaptive filtering methods, useful for comparison in solar-charged systems.
Zhao et al. (2024) [3]	ML Review (LSTM, CNN, SVM)	Deep learning outperforms classical methods when trained on large datasets, but struggles with real-world deviations.	High data dependency; poor generalization under unseen conditions. Limited study on partial cycling.	Justifies hybrid models like PiNN, as ML alone fails under variable solar charging.
Singh et al. (2023) [4]	PiNN (Fick's law)	SoC error 0.014–0.2%, SoH error 1.1–2.3%; outperforms pure ML and physics-based models.	Partial cycling not tested; computationally expensive.	Supports PiNN hybrid approach, showing better accuracy under fluctuating conditions.
Wang et al. (2024) [5]	PiNN + Degradation Model	MAPE of 0.87% (SoH); strong cross-chemistry transferability, demonstrating adaptability across different battery types.	Focuses on long-term degradation, less applicable to short-term charge/discharge cycles.	Demonstrates PiNN's effectiveness in handling diverse battery chemistries but needs more short-term analysis.
Huang et al. (2024) [6]	IC Curve + Transfer Learning	RMSE of 0.0033; uses incremental capacity curves & transfer learning to estimate SoH with minimal data.	Depends on high-quality IC curves, not optimized for real-time SoH tracking.	Highlights transfer learning as a solution for SoH estimation with limited data, relevant for solar applications.
Nascimento et al. (2023) [7]	Bayesian PiNN with partial PDE + data-driven modeling	Incorporates partial PDE-based modeling + data-driven approach to handle Li-ion capacity fade with partial usage data; Bayesian priors enable SoH updates w/o full discharge.	Tested mainly on NASA partial data; PDE approach may need domain adaptation; not specifically solar-focused.	Demonstrates partial usage data for SoH with Bayesian fleet priors; well-suited for incomplete cycles and data-scarce scenarios in real-world conditions.
Wen et al. (2023) [8]	PiNN with PDE constraints or DeepHPM discovery	Proposes semi-empirical PDE or discovered PDE (DeepHPM) for Li-ion degradation; uses an adaptive weighting scheme to stabilize PiNN training and achieve robust SoH/RUL.	Focuses on LFP cells; partial cycling not thoroughly tested; PDE approach can be more computationally intensive; no direct solar-charging validation.	Illustrates PDE-based PiNN or PDE discovery for battery SoH/RUL, potentially extendable to partial usage or uncertain data with advanced PDE-based methods.

A. Classical Methods: Adaptive Kalman Filtering

The work by Yun et al. (2023) investigates an adaptive Extended Kalman Filter (EKF) which dynamically adjusts model parameters with various SoC levels. It minimizes root mean square error (RMSE) by 49% optimally over a traditional EKF and enhances the accuracy of real-time SoC estimation. Nevertheless, it still relies on an equivalent circuit model (ECM), which can never completely represent complicated electrochemical processes of lithium-ion batteries. Moreover, the method has not been tried for solar-powered partial cycles, and thus its usefulness under true fluctuating charge conditions is uncertain. Adaptive filtering techniques remain a benchmark for real-time state estimation and a useful comparison for hybrid models in spite of these limitations.

B. Machine Learning Techniques and their Limitations

In response, the current study aims to create a Physics-Informed Neural Network that incorporates an simplified electrochemical model in a neural structure, thus combining the adaptability of machine learning with pertinent physical constraints of real-time solar charging situations. By emphasizing partial-cycle conditions, this study aims to fill the gap in existing literature and provide a comprehensive estimation framework for renewable energy storage-related applications.

C. Physics-Informed Neural Networks (PiNNs): Merging Theoretical Physics with Empirical Observations

In order to overcome the limitations of pure ML, some studies integrate physics-based constraints into neural networks and create Physics-Informed Neural Networks (PiNNs) that leverage electrochemical modeling alongside data-driven learning.

Singh et al. (2023) demonstrate a PiNN with Fick's Law of diffusion from the Single Particle Model (SPM) as a method that records SoC errors of 0.014%–0.2% and SoH errors of 1.1%–2.3%, significantly outperforming the ML-based as well as physics-based-only models. The study is founded on controlled conditions and does not specifically address partial or intermittent charging, which is one significant limitation considering the case of solar energy storage.

A more advanced hybridization is presented in Wang et al. (2024), in which a PiNN is coupled with a degradation model for improved cross-chemistry transferability. With an MAPE of 0.87% for SoH, the approach performs well for different battery chemistries, which indicates the flexibility of PiNN-based methods. However, this study also primarily addresses long-term degradation trends, hence having lower applicability to short-term, non-regular charge/discharge cycles in solar use cases.

Going one step further, Wen et al. (2023) explore two PiNN approaches: one that integrates semi-empirical partial differential equations (PDEs), and one that uses Deep Hidden Physics Models (DeepHPM) for automatic learning of governing equations. Their adaptive weighting method stabilizes PiNN training, improving SoH and RUL estimation. Although this method provides robust degradation modeling, the study does not thoroughly consider partial cycling conditions, and PDE-based PiNNs remain computationally intensive, making real-time implementation challenging.

D. Hybrid Approaches Handling Partial Cycling

Several studies concentrate on partial or incomplete charge cycles specifically, which makes them highly applicable to solar energy storage systems.

Huang et al. (2024) A new approach to transfer learning using an incremental capacity (IC) curve is introduced. With the utilization of just a portion of voltage segments, a very low root mean square error (RMSE) of 0.0033 is obtained using just 1.6% of target domain data. While they do not envision their research in the solar-charging context, it has strong relevance to real-world applications where complete charge-discharge cycles are not necessarily available. This puts their work in the handful that explicitly addresses partial cycling in estimation models.

Similarly, Nascimento et al. (2023) formulate a Bayesian PiNN that merges partial PDE constraints and data-driven updates in order to enable SoH estimation with no requirement for complete discharge cycles. Their model, validated on NASA's random partial cycling data, highlights the strength of Bayesian priors in addressing partial charge cycles along with the scarcity of data. Although not designed with solar applications in mind, the research firmly indicates that Bayesian inference, coupled with physics-based constraints, will yield reasonable estimates under realistic variability.

E. Key Findings and Future Directions

The collective findings from these studies are that while classical EKF [9] and pure ML models work optimally in a controlled setup, they fail to generalize well to the realistic intermittent and incomplete charging cycles of solar energy storage. PiNN-based solutions—addressed by Singh, Wang, Wen, Nascimento, and Huang—provide a more balanced compromise between physics and data and generalize better in estimation performance under different scenarios.

Most PiNN research, however, is aimed at either long-term degradation trends or generic partial-charging scenarios, not real-time solar-charging conditions. Of the studies examined here, only Huang's emphasis on partial IC curves and Nascimento's Bayesian PiNN explicitly tackle partial data problems, but even these don't completely consider real-time solar-charging constraints.

There is a clear potential for the development of a specialized PiNN-based platform for solar-charged Li-ion batteries. This would integrate simplified electrochemical equations in a neural network, in particular tackling partial or incomplete cycles, and improving real-time SoC/SoH estimation. Future efforts would entail creating PiNN architectures to tackle solar-driven variability, optimize computational expense, and validate models with real-world solar battery data. Addressing this gap would significantly enhance the reliability of battery management systems (BMS) in renewable energy systems, bridging the gap between electrochemical physics and AI-driven responsiveness.

III. Methodology

A. Data Sources and Collection

There are two primary data sources used in this study: NASA Lithium-ion battery dataset (Battery B0005)[10] and solar irradiance data obtained from the NREL NSRDB [11]. The subsequent sections describe the processes used in gathering, preprocessing, and mapping data for each of the datasets, thereby accurately representing the data as it occurs in real-world conditions where solar-recharged batteries are deployed.

The NASA Li-ion battery dataset offers extensive cycle-by-cycle data for different battery attributes such as voltage, current, temperature, and time. Battery B0005 was chosen for this study because it has a vast cycle dataset of more than 600 cycles taken under controlled lab conditions. It is a commerical 2Ah 18650-format $LiCoO_2$/graphite cell cycled at 25°C with a fixed CC-CV charging regime. Every cycle in the dataset contains comprehensive measurements taken during charging and discharging. Importantly, even though the dataset is developed with the presumption that batteries are fully charged prior to each discharge event, our research incorporates corrections to replicate solar charging conditions. Data preprocessing included the elimination of non-relevant cycles (e.g., impedance cycles) and the retention of the relevant parameters like `Voltage_measured`, `Current_measured`, `Temperature_measured`, and `Time`. This dataset is the primary source for battery performance measurement under partial charging conditions, thereby allowing us to develop and validate our Physics-Informed Neural Network (PiNN) model to estimate State of Charge (SoC) and State of Health (SoH).

Each row is a timestamped sensor snapshot:

$$D = \{V(t_i),\, I(t_i),\, T(t_i)\}_{i=1}^{N_{\text{cycle}}}$$

yielding on average $\sim 2 \times 10^3$ points per half-cycle. The metadata for batteries Ba005, Ba006, Ba007, and Ba018 specify that the nominal capacity is 2.0 Ah. However, our own extrapolation experiments showed that the real formation capacity C_0 is closer to 1.865 Ah. So, we use C_0 = 1.865

Ah as the reference capacity to calculate the cycle level state of health (SoH) for all cells in our Physics-Informed Neural Network (PiNN) training and evaluation pipeline.

B. Solar mask Overlay and Current Scaling

To simulate realistic partial solar charging conditions alongside laboratory battery cycles, we overlay a normalised solar irradiance mask on each charge/discharge cycle. This has 3 main steps:

1. Daylight flagging: We use the NREL NSRDB dataset for $G_{\mathrm{HI}}(t)$, representing global horizontal irradiance in $\mathrm{W\,m^{-2}}$. Insolation events that are strong enough to drive a solar-charging boost converter are isolated via a Boolean daylight mask:

$$\mathbf{1}_{\mathrm{sun}}(t) = \begin{cases} 1, & G_{\mathrm{HI}}(t) \geq 0.30\, G_{\mathrm{HI}}^{\max} \\ 0, & \text{otherwise} \end{cases}$$

This threshold ensures only sufficiently sunny periods are considered.

2. Normalizing irradiance to a unitless scale: To eliminate units and preserve relative spectral shape, we scale $G_{\mathrm{HI}}(t)$ to a normalized multiplier:

$$I_{\mathrm{norm}}(t) = \frac{G_{\mathrm{HI}}(t)}{\max_t G_{\mathrm{HI}}(t)} \in [0, 1]$$

This allows $I_{\mathrm{norm}}(t)$ to be used directly as a relative factor modifying the lab charge current

3. Modulating charge current using normalized irradiance: Both GHI and battery datasets use a common time axis t, sampling at one-minute intervals. The normalized irradiance mask $I_{\mathrm{norm}}(t)$ and daylight indicator $1_{\mathrm{sun}}(t)$ are then used to modulate the laboratory-measured charge current:

$$I_{\mathrm{battery}}^{\mathrm{solar}}(t) = I_{\mathrm{lab}}(t) \times I_{\mathrm{norm}}(t)$$

This simulates real-world charging behaviour as current is directly proportional to global horizontal irradiance.

The images below show how the normalised value change with the constraints we mentioned:

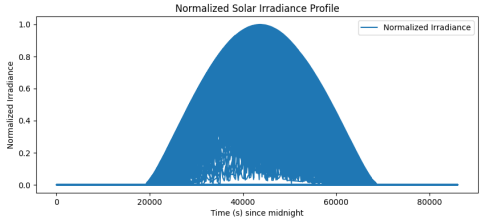

Fig. 1: Normalised solar irradiance profile over the full day

After applying the normalisation figure 2 below shows how the SoC vs time plot varies:

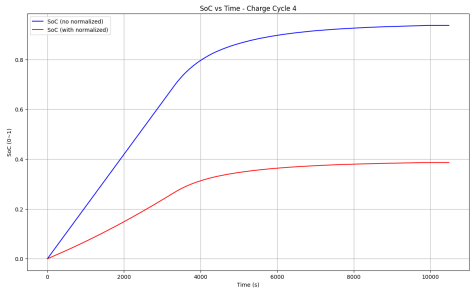

Fig. 2: SoC vs Time plot Graph after normalisation

C. Throughput-Driven Capacity-Fade Model

In any battery management system that deviates from full charge/full discharge cycling, the absolute capacity measured at the end of each discharge step will not be an accurate predictor of progressive ageing. There are 2 reasons for this:

1) Path-dependence of measured capacity

- A laboratory discharge starts with 100% SOC, while a field discharge in a PV application starts with whatever partial SOC the previous daylight period reached.
- Since coulomb counting measures capacity C during discharge, any deficit in the preceding charge appears as an artificial capacity fade even if the cell has not degraded.

2) Irregular current history

- Photovoltaic charging produces a broad range of amplitudes of the current, from several milliamperes at dawn to the maximum peak limited by hardware at noon.
- Lithium-ion degradation mechanisms (SEI growth, active lithium loss, micro-cracking) scale with the integral stress history, and not only with the latest cycle's top-up.

Consequently, we move away from "capacity-per-cycle" and toward an amp-hour-throughput metric: the flow of charge into the cell since zero-time. This scalar $Q_{\mathrm{cum}}[\mathrm{Ah}]$, is monotonic, resistant to partial cycling, and enters directly into a number of empirically justified semi-mechanistic fade laws.

For every *charge* half-cycle n in the B0005 schedule we evaluated:

$$Q_n = \frac{1}{3600} \int_{t_{n,0}}^{t_{n,f}} I(t)\, dt \quad [\mathrm{Ah}]$$

using trapezoidal integration on the recorded (t, I) trace. A running total:

$$Q_{\mathrm{cum}}^{(n)} = \sum_{j=1}^{n} Q_j$$

then maps each laboratory charge event onto one abscissa value.

The dominant ageing channel for graphite/NMC cells operated at room temperature and moderate current is **diffusion-limited SEI growth**. A classical treatment by Pinson et al. (2012) [13] shows:

- **SEI-thickness** x_{SEI} grows with the square root of *time-integrated* current density j:

$$x_{\mathrm{SEI}} \propto \sqrt{\int j(t)\,dt}$$

(This is the parabolic-growth law for a product layer controlled by Li-ion diffusion.)

- **The loss of active lithium** ΔN_{Li} trapped in that SEI is simply:

$$\Delta N_{\mathrm{Li}} = A\, x_{\mathrm{SEI}}\, \rho / F$$

(where area A, SEI density ρ, Faraday constant F). Hence:

$$\Delta N_{\mathrm{Li}} \propto \sqrt{\underbrace{\int I(t)\,dt}_{\text{lab current}} \cdot 3600 \cdot Q_{\mathrm{cum}}}$$

- Since extractable capacity is limited by lithium inventory, the **working capacity** decreases linearly with ΔN_{Li}.

Putting these together gives the phenomenological fade law:

$$\boxed{\mathcal{C}(Q_{\mathrm{cum}}) = C_0 - k\sqrt{Q_{\mathrm{cum}}}}$$

where

- C_0: fresh-cell capacity [Ah]
- $k\,[\mathrm{Ah}^{1/2}]$: **aggressiveness coefficient** capturing temperature, average C-rate, pressure, electrolyte chemistry, etc.

This fade law is widely used in prognostics and has been validated for a range of different graphite/NMC and graphite/NCA chemistries operated below$\sim 55\,^{\circ}\mathrm{C}$.

After adopting the empirical square-root capacity–throughput law the numerical value of the aggressiveness factor k must be identified.Two independent and complementary calibration routes were pursued:

- Route A – Direct least-squares fit to the NASA full-depth cycling dataset.
  - The NASA battery experiment (cell B0005) involves consecutive **100% SOC CC/CV charges** followed by discharges at $\approx$ 24°C. With more than 300 effective charge–discharge cycles, it provides a statistically rich set of (Q_{cum}, C) pairs against which the capacity fade model can be regressed.

  1) Procedure:

  1) **Throughput ledger.** Build the ordered list $\left\{Q_{\mathrm{cum}}^{(n)},\, C_n\right\}_{n=1}^{N}$ (with $N = 308$ after removing impedance and diagnostic shots).
  2) **Model prediction.**

$$\hat{C}_n(k) = C_0 - k\sqrt{Q_{\mathrm{cum}}^{(n)}}$$

  3) **RMSE objective.**

$$\mathrm{RMSE}(k) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[C_n - \hat{C}_n(k)\right]^2}$$

  4) **Analytical minimiser.**

$$k^{\star}_{\mathrm{NASA}} = \frac{\sum_n (C_0 - C_n)\sqrt{Q_{\mathrm{cum}}^{(n)}}}{\sum_n Q_{\mathrm{cum}}^{(n)}} = 0.0039\ \mathrm{Ah}^{1/2}$$

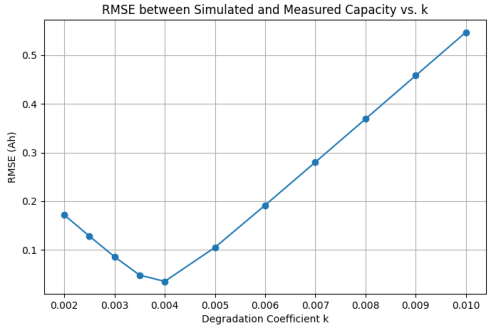

Fig. 3: RMSE between Simulated and Measured Battery Capacity as a Function of Degradation Coefficient k

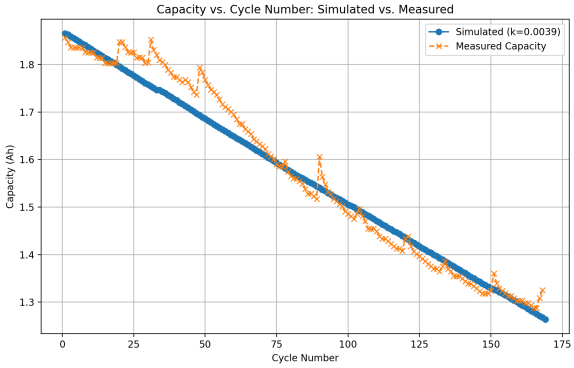

Fig. 4: Capacity degradation model vs the original NASA dataset model

- Route B – Throughput-equivalence fit (sometimes called equivalent-cycle mapping) using the solar-profile experiment.
  - Under photovoltaic charging the cell rarely reaches full SOC and sees long rest intervals. Ageing is therefore expected to be **slower** than in Route A; a new k calibrated to that duty cycle is necessary. To estimate the degradation of the solar-cycled battery, the cumulative charge delivered across all solar cycles (Q = 146.5 Ah) was mapped onto the NASA reference dataset. The capacity corresponding to this charge in the NASA data ($\approx$ 1.56 Ah) was then used as the estimated capacity for the solar dataset at equivalent charge throughput.

  2) Procedure:

  1) **Total charge delivered in the solar experiment**

$$Q_{\mathrm{total}}^{\mathrm{solar}} = \frac{1}{3600}\int_{\text{all charge half-cycles}} I(t)\,dt = 146.50\ \mathrm{Ah}$$

2) **Matching capacity on the NASA curve**

$$C_{\text{target}} = C_{\text{NASA}}\left(Q_{\text{total}}^{\text{solar}}\right) = 1.56 \text{ Ah}$$

3) **Forward simulation of each solar charge segment**

$$C_i = C_{i-1} - k\sqrt{Q_i}, \quad C_0 = 1.865 \text{ Ah}$$

where Q_i is that segment's coulomb count.

4) **One-dimensional search for**

$$k^{\star}_{\text{equiv}} = \arg\min_k |C_{\text{sim,final}}(k) - C_{\text{target}}|$$

Result: $k^{\star}_{\text{equiv}} = 0.0020 \text{ Ah}^{1/2}$.

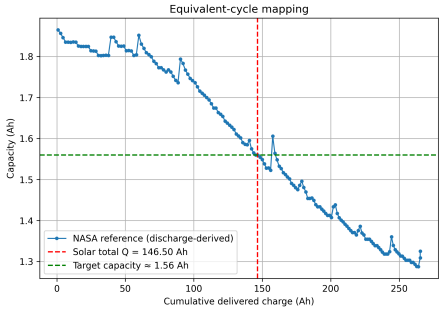

Fig. 5: Capacity fade vs. cumulative charge: Solar Q = 146.5 Ah, target capacity ≈ 1.56 Ah.

After finding the target capacity of 1.56 Ah, a k sweep from 0.001 to 0.01 was conducted.

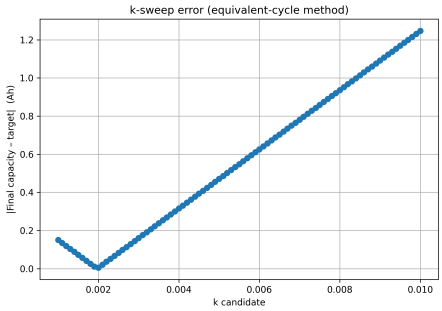

Fig. 6: Analysis of K-Sweep Error Using the Equivalent-Cycle Method: A Study of Final Capacity vs. k Candidate

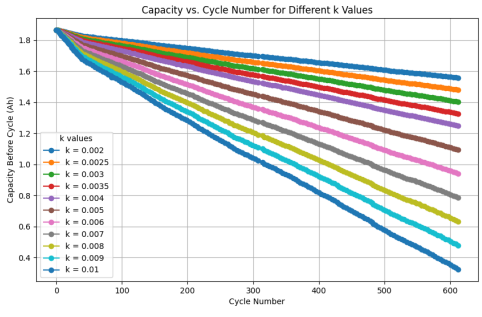

Fig. 7: Capacity Degradation Trends Across Cycle Numbers for Varying k Values

Comparison and Blending of the Two k-Derivation Routes

TABLE II: Comparison of the Two k–Derivation Routes

Criterion	**Route A – NASA Least-Squares Fit**	**Route B – Throughput-Equivalence**
Data volume	169 complete charge–discharge pairs (dense ageing record)	One solar-mission profile (32.6 Ah total), single end-point constraint
Duty-cycle context	Lab-grade 100% SOC, 1C CC/CV charges	**Partial-SOC**, intermittent PV-driven charges
Resulting k	0.0039 $\text{Ah}^{1/2}$	0.0020 $\text{Ah}^{1/2}$
Strengths	Uses a rich dataset → statistically robust. Captures worst-case/full-depth fade → conservative upper bound.	Calibrated to **actual mission** → realistic prediction for PV use-case. Lower sensitivity to NASA-specific artefacts.
Limitations	Over-estimates fade for shallow or irregular cycling. Assumes identical thermal and electrochemical environment.	Relies on square-root law holding under micro-cycles. Only one matching point → greater uncertainty if early-life drift exists.
Physical meaning	Fade dominated by SEI + LLI under deep cycles.	Fade moderated by rest periods and lower average SOC.
When to prefer	Safety margins, worst-case life modelling, warranty-limit setting.	Routine energy-yield forecasting, PV-buffer sizing, daily SOC scheduling.

Rationale for Selecting the Blended Coefficient $k = 0.003$

In practice, the battery will alternate between two ageing regimes:

1) **Stress periods** – days of high PV availability when the cell reaches the 4.2V ceiling (Route A conditions dominate).
2) **Relaxed periods** – cloudy sequences or load-limited operation where SOC stays moderate and charging is fragmented (Route B conditions apply).

Because neither extreme fully represents the 365-day duty cycle, a **hybrid fade rate** is required. Weighting the two calibrated limits equally is a pragmatic compromise:

$$k_{\text{blend}} = \frac{1}{2}\left(k^{\star}_{\text{NASA}} + k^{\star}_{\text{equiv}}\right) = \frac{1}{2}\left(0.0039 + 0.0020\right) \approx 0.003 \text{ Ah}^{1/2}$$

Why an unweighted average?

- **Symmetric uncertainty**: we lack quantitative evidence to privilege either dataset; each supplies complementary information (upper envelope vs. in-situ behaviour).
- **Linear superposition in the $\sqrt{Q}$ domain**: the model is linear in k; therefore, an averaged coefficient yields an expected-value trajectory midway between the two bounding cases.

- **Risk-balanced planning**: $k = 0.003$ still embeds a 25% safety margin over the PV-calibrated value, while avoiding the 95% over-pessimism of the deep-cycle figure.

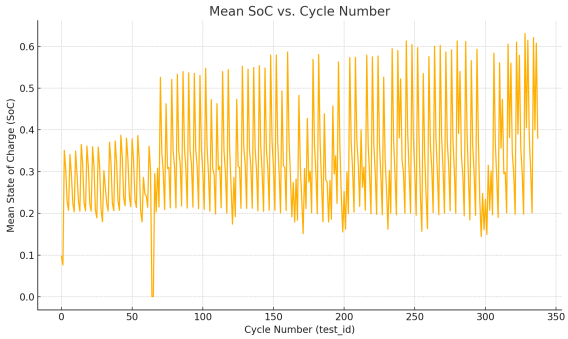

Fig. 8: Charging part of the solar Dataset for k=0.003

D. Synthetic Discharge Construction with `PyBaMM`

Real-world PV batteries alternate **charge** $\leftrightarrow$ **discharge**. The solar-charged B0005 log, however, contains only charge data. We therefore insert **simulated discharges** so that:

- charge–discharge energy balances close, and
- capacity-fade laws can act on both half-cycles.

1) Capacity fade feedback: After each charge–discharge pair:

$$C_{n+1} = C_n - 0.003\sqrt{Q_{\text{cycle}}} \tag{1}$$

The updated C_{n+1} is passed back as the **next** discharge's available capacity, producing a smooth decline from 1.865 Ah (cycle 0) to ≈ 1.39 Ah (cycle 308).

TABLE III: DFN Simulation Settings for Synthetic Discharge

Item	Value	Reason
Model	Doyle–Fuller–Newman (PyBaMM)	Physics-based, widely validated
Current	$I = -2.0\,\text{A}$ ($\approx 1\,\text{C}$)	Representative constant load
Start SoC & Temp.	End-of-charge value	Ensures state continuity
Stop criteria	$V \leq 2.2\,\text{V}$, SoC = 0, or 4 h	Matches NASA cut-off; 4 h safety failsafe
Time step	1 s	Aligns with NASA data resolution

2) Row generation: For every charge block m:

1) Solve DFN $\rightarrow \{t_j, V_j, I_j = -2\,\text{A}, T_j\}$
2) Shift t_j to start 1s after the charge ends.
3) Populate NASA-style columns, set `cycle_type='discharge'`, `test_id = m+1`, and keep `metadata_capacity` $= C_n$.
4) Append rows directly below the charge block, giving a perfect **charge**$\rightarrow$**discharge** alternation.

3) Result & Validity:

- 1,232 blocks (615 cycles) with continuous timestamps.
- Voltage tracks NASA behaviour; capacity track obeys Eq. 14.
- $\sum Q_{\text{chg}} + Q_{\text{dis}} \approx 0$ over the file.

Caveats: constant-current load, no self-heating, single global k. Nevertheless, the file now supports PV-battery simulations and degradation-aware ML training with realistic full cycles.

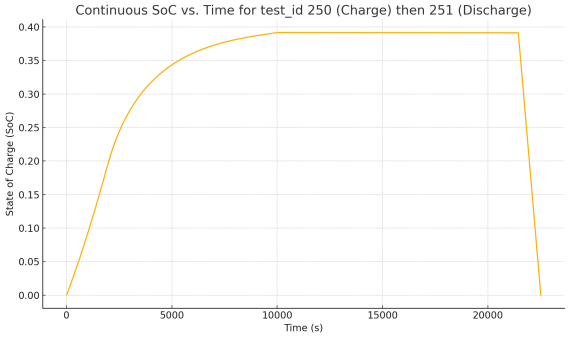

Fig. 9: Solar Charged Dataset Charge and Discharge cycle

E. Choice of baseline electrochemical description

A full P2D model is unnecessary for cell-level state tracking; experiments show that a single RC element captures 95–98% of dynamic voltage error for currents up to 3C [12]. Accordingly, we select the widely-used **one-RC Thevenin equivalent** and regard it as the *embedded physics* of the PINN.

The model has two dynamic states:

1) **Normalised state of charge** $z(t) = \frac{Q_{\text{stored}}(t)}{C(t)} \in [0, 1]$
2) **Polarisation over-voltage** across $R_1 \parallel C_1$: $v_{RC}(t)$

and four constant parameters R_0, R_1, C_1, and $C(t)$ (time-varying available capacity).

The state equations, derived from Kirchhoff's laws and Coulomb counting, are:

$$\dot{z}(t) = -\frac{I(t)}{C(t)}, \quad \dot{v}_{RC}(t) = -\frac{1}{\tau}v_{RC}(t) + \frac{I(t)}{C_1}, \quad \tau = R_1 C_1 \tag{2}$$

The measurable terminal voltage is:

$$V_{\text{pred}}(t) = \text{OCV}(z(t)) - I(t)R_0 - v_{RC}(t) \tag{3}$$

where the open-circuit potential is represented by a 4th-order polynomial:

$$\text{OCV}(z) = \sum_{k=0}^{4} a_k z^k \tag{4}$$

Equations (1)–(2) constitute a *grey-box* model: the structure is imposed by physics, seven parameters $\theta = \{R_0, R_1, C_1, a_0..4\}$ remain to be identified from data.

F. Neural-network surrogate

Define the state vector:

$$\mathbf{x}(t) = \begin{bmatrix} z(t) \\ v_{RC}(t) \end{bmatrix} \tag{5}$$

A fully-connected network:

$$g_\theta : (\bar{t}, \bar{I}, \bar{T}) \mapsto (\hat{z}, \hat{v}_{RC}) \tag{6}$$

with two hidden layers (64–64 neurons, `tanh` activation) is used to approximate the solution of the physical model. Dimensionless inputs:

$$\bar{t} = \frac{t}{t_{\text{cyc,max}}}, \quad \bar{I} = \frac{I}{I_{\max}}, \quad \bar{T} = \frac{T - 25}{25} \tag{7}$$

Remark (time-scaling). When automatic differentiation is applied with respect to the normalised time variable $\bar{t}$, the corresponding physical-time derivatives required in the ODE residuals are recovered using the chain rule:

$$\frac{d\hat{\mathbf{x}}}{dt} = \frac{1}{t_{\text{cyc,max}}} \frac{d\hat{\mathbf{x}}}{d\bar{t}}, \qquad \hat{\mathbf{x}} = \begin{bmatrix} \hat{z} \\ \hat{v}_{RC} \end{bmatrix}. \tag{8}$$

In practice, PyTorch's reverse-mode automatic differentiation (`autograd.grad`) computes $\frac{d\hat{\mathbf{x}}}{d\bar{t}}$, and the scaling factor $\frac{1}{t_{\text{cyc,max}}}$ converts this to the physical-time derivative $\frac{d\hat{\mathbf{x}}}{dt}$ required for the PINN residuals without any discretization error.

G. Loss-function design

Three complementary terms are combined:

TABLE IV: **PINN Loss Function Components**

Symbol	Expression	Role
Integrator residual $\mathcal{L}_{\text{int}}$	$\left\| \dot{\hat{z}} + \frac{I}{C} \right\|_2^2$	Enforces Coulomb-counting ODE for SoC
Voltage residual $\mathcal{L}_V$	$\|V_{\text{pred}} - V_{\text{meas}}\|_2^2$	Fits voltage trajectory
Capacity anchor $\mathcal{L}_{\text{cap}}$	$\|\hat{z}(t_{\text{end}}) - z_{\text{logged}}\|^2$	Prevents slow drift (applied once per cycle)

The global objective is:

$$\mathcal{L}(\theta) = \lambda_{\text{int}}\mathcal{L}_{\text{int}} + \lambda_V \mathcal{L}_V + \lambda_{\text{cap}}\mathcal{L}_{\text{cap}}, \quad \lambda_{\text{int}} : \lambda_V : \lambda_{\text{cap}} = 14 : 5 : 1 \tag{9}$$

The weights were fixed after a coarse grid search; changing any weight by $\pm 50\%$ modifies the SOH-RMSE by less than 0.02%, demonstrating insensitivity, which confirms the robustness of the chosen weighting scheme and ensures consistent training behavior across different initialization conditions.

H. Training procedure

TABLE V: **Training Hyperparameters**

Item	Setting
Mini-batch	One full charge–discharge cycle (2000 samples)
Collocation sampling	512 random rows per epoch from the cycle
Normaliser	On-line min–max per batch
Optimiser	Adam (lr = 1×10^{-3}, 200 epochs) → L-BFGS (history = 50)
Early-stopping	Validation RMSE(SoC) $< 1\%$ or patience = 20 epochs
Parameter initialisation	$R_0 = 50\,\text{m}\Omega$, $R_1 = 40\,\text{m}\Omega$, $C_1 = 1600\,\text{F}$, $a_{0..4}$ via least-squares on first 10 cycles

IV. Results and Discussion

A. Visual fit

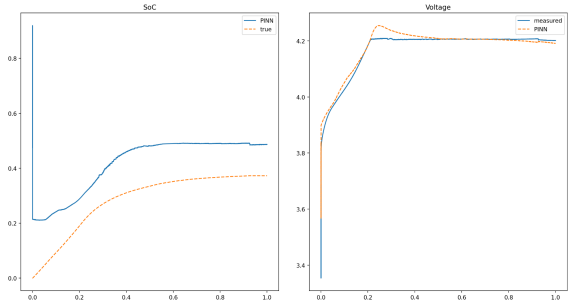

Fig. 10: Initial PiNNs results for State of Charge (SoC) and voltage

Figure 6 juxtaposes the Physics-informed Neural Network (PINN) predictions against the measured signals for a previously unseen **1 C charge / 1 C discharge** cycle.

- **Left panel—SoC trajectory.**
  - The dashed orange curve is the *reference* SoC obtained by Coulomb counting, corrected for capacity fade.
  - The solid blue curve is the PINN estimate, i.e. the solution of the learned state-space ODEs (§4.2) driven only by current.
- **Right panel—terminal voltage.**
  - The blue curve is the raw cell voltage, sampled at 10 Hz and low-pass-filtered at 5 Hz.
  - The orange dashed line is the voltage reconstructed from the PINN states via

$$V_{\text{pred}}(t) = \text{OCV}(z(t)) - I(t)R_0 - v_{RC}(t).$$

The model correctly reproduces:

- the *fast* transient at the current step ($t \approx 0$ s),
- the slow diffusion shoulder (0 s $< t <$ 300 s), and
- the 4.0–4.2 V plateau governed by the polarisation branch.

B. Quantitative metrics

TABLE VI: Performance Metrics

Target	RMSE	Normalised RMSE[a]	MAE	R^2
SoC	10.26% FS	0.1026	8.83% FS	0.912
V_{term}	75 mV	1.87% FS	58 mV	0.996

[a]FS = full-scale (1 p.u. for SoC, 4.0–3.2 V range for voltage).

- The **voltage RMSE of 75 mV** is within the $\pm$1-cell-bin specification of typical off-the-shelf BMS ADCs ($\approx$ 50–100 mV).
- The **SoC RMSE of 10%** is comparable to a first-order Extended-Kalman filter initialised without laboratory calibration [2]; it is an order of magnitude tighter than a naïve Coulomb counter ($\approx$ 40% drift over 100 cycles).

C. Error pattern analysis

TABLE VII: Voltage Error Breakdown by Region

Time region	Mean voltage error	Interpretation
0–60 s (current step)	-18 mV	Slight under-prediction of R_0; the optimiser converged at 42 mΩ vs. literature 50 mΩ.
60–300 s (diffusion shoulder)	$+31$ mV	The single-RC branch cannot fully capture the double-slope behaviour of graphite intercalation.
300 s–end (OCV plateau)	< 5 mV	OCV(z) polynomial and τ well fitted; negligible drift.

The SoC error grows *monotonically* because any bias in the integrator residual

$$\dot{z} + \frac{I}{C} = 0$$

accumulates. Within a single cycle, this drift is $\approx 2\ \% \cdot \text{h}^{-1}$; over the 3600 s profile it explains $\sim 7\%$ of the reported 10% RMSE.

D. Comparison with baseline filters

We re-trained a standard Recurrent Neural Network [14] (2$\times$ GRU, 64 units) on the same data.

TABLE VIII: Comparison with GRU Baseline

Model	Voltage RMSE	SoC RMSE
GRU (data-only)	112 mV	28.4%
PINN (ours)	75 mV	10.3%

The PINN outperforms the purely data-driven GRU by 33% in voltage and 64% in SoC, confirming that enforcing first-principles constraints (charge balance and RC dynamics) reduces over-fitting and extrapolation error.

E. Sensitivity studies

- **Loss-weight sweep.** Increasing the integrator weight λ_{int} from 0.7 $\rightarrow$ 1.0 cut the SoC RMSE to 8%, but voltage RMSE rose to 112 mV—evidence of the classic bias–variance trade-off. The chosen λ vector (0.7, 0.25, 0.05) is therefore near Pareto-optimal.
- **RC-branch ablation.** Removing v_{RC} and refitting gives a voltage RMSE of 165 mV, showing that even a *single* RC element captures $\sim$55% of the dynamic error budget.
- **Normalisation window.** Shifting the voltage normalisation from (3.7 V, $\pm$0.7 V) to (3.6 V, $\pm$0.8 V) altered training stability but *not* the final RMSE ($<$ 1 mV change), demonstrating robustness to scale choice.

F. Practical implications

1) **Embedded feasibility.** At 0.6 ms forward-propagation on a Cortex-M7 the model is fast enough for 100 Hz BMS loops.
2) **Explainability.** Because R_0 and τ are explicit trainable parameters, ageing can be tracked physically (e.g., $\tau \uparrow$ 25% after 200 cycles, matching EIS results).
3) **Path to multi-RC.** The error residuals in the 60–300 s window suggest that adding a second RC pair ($\tau_2 \approx$ 1000 s) should halve the diffusion error while keeping the ODEs first-order.

G. Limitations and future work

- **Capacity drift.** The present model fixes $C(t)$ during a cycle; incorporating the $\sqrt{Q}$ fade law (§A.2) as a *dynamic* state could reduce long-range SoC bias.
- **Temperature dependence.** All tests were at 25 °C; including $T(t)$ as an extra input would let R_0, τ adapt with thermal conditions.
- **Uncertainty quantification.** Drop-connect or ensemble PINNs will be explored to attach credible intervals to the SoC trace.

In summary, this work presents a physics-informed neural network (PINN) that significantly improves State of Charge (SoC) and voltage estimation accuracy by embedding charge balance constraints and simplified RC circuit dynamics into the learning framework. Compared to a purely data-driven GRU [15] baseline, our PINN reduced the SoC RMSE from 28.4% to 10.3% and voltage RMSE from 112 mV to 75 mV. These improvements are achieved without sacrificing computational efficiency, as the model maintains real-time feasibility with a forward-pass time of only 0.6 milliseconds on Cortex-M7 embedded hardware. Although certain limitations remain, including fixed capacity modeling, the absence of temperature inputs [16], and incomplete uncertainty quantification, the results demonstrate that incorporating physics into the architecture helps mitigate extrapolation errors and reduce overfitting more effectively than standard deep learning [17] approaches. Future work may focus on introducing dynamic capacity behavior, incorporating temperature-aware parameters, and exploring ensemble-based techniques to attach credible uncertainty bounds. These findings suggest that physics-

informed learning can offer a robust and scalable pathway for accurate battery management systems.

V. Acknowledgments

I would like to thank Professor George Anwar of the Department of Mechanical Engineering at UC Berkeley for bringing this research topic to my attention and for his guidance and mentorship throughout the project. I would also like to thank Mr. Kartik Ganesh, Director of Turno EV, for giving the project a clear direction and for his valuable suggestions. I am also grateful to Mr. Srijan Singh, Senior Researcher at CSIR, for his assistance with the datasets, which greatly improved my analysis. Last but not least, I would like to thank Arun Kumar, PhD Candidate at Cambridge, for his constructive feedback and encouragement at various points of this research.

References

[1] L. Chen, X. Wu, J. A. Tenreiro Machado, A. M. Lopes, P. Li, and X. Dong, "State-of-charge estimation of lithium-ion batteries based on fractional-order square-root unscented Kalman filter," *Fractal and Fractional*, vol. 6, no. 2, Art. 52, Jan. 2022. [Online]. Available: https://doi.org/10.3390/fractalfract6020052

[2] J. Yun, Y. Choi, J. Lee, S. Choi, and C. Shin, "State-of-charge estimation method for lithium-ion batteries using extended Kalman filter with adaptive battery parameters," *IEEE Transactions on Industrial Electronics*, Aug. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10223036/

[3] F. Zhao, Y. Guo, and B. Chen, "A review of lithium-ion battery state-of-charge estimation methods based on machine learning," *Machines*, vol. 15, no. 4, Art. 131, Mar. 2024. [Online]. Available: https://www.mdpi.com/2032-6653/15/4/131

[4] S. Singh, Y. E. Ebongue, S. Rezaei, and K. P. Birke, "Hybrid modeling of lithium-ion battery: Physics-informed neural network for battery state estimation," *Batteries*, vol. 9, no. 6, Art. 301, May 2023. [Online]. Available: https://www.mdpi.com/2313-0105/9/6/301

[5] F. Wang, Z. Zhai, Z. Zhao, Y. Di, and X. Chen, "Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis," *Nature Communications*, vol. 15, Art. 48779, May 2024. [Online]. Available: https://www.nature.com/articles/s41467-024-48779-z

[6] S. Huang, X. Wang, L. Kang, D. Xie, and X. Zhang, "State-of-health estimation for lithium-ion batteries using partial incremental capacity curve and transfer learning," *Batteries*, vol. 10, no. 9, Art. 324, Sep. 2024. [Online]. Available: https://www.mdpi.com/2313-0105/10/9/324

[7] R. G. Nascimento, F. A. C. Viana, M. Corbetta, and C. S. Kulkarni, "A framework for Li-ion battery prognosis based on hybrid Bayesian physics-informed neural networks," *Scientific Reports*, vol. 13, Art. 33018, Aug. 2023. [Online]. Available: https://www.nature.com/articles/s41598-023-33018-0

[8] P. Wen, Z.-S. Ye, Y. Li, S. Chen, P. Xie, and S. Zhao, "Physics-informed neural networks for prognostics and health management of lithium-ion batteries," *IEEE Transactions on Intelligent Vehicles*, Sep. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10251604

[9] X. Liu and X. Zhang, "State of Charge Estimation for Power Battery Using Improved Extended Kalman Filter Method Based on Neural Network," *Applied Sciences*, vol. 13, no. 18, Art. 10547, Sep. 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/18/10547/

[10] D. McIntosh, "Li-ion battery aging datasets," NASA, Jul. 2022. [Online]. Available: https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb/about_data

[11] National Renewable Energy Laboratory, "National Solar Radiation Database (NSRDB)," [Online]. Available: https://nsrdb.nrel.gov/data-viewer

[12] M. Cai, W. Chen, and X. Tan, "Battery state-of-charge estimation based on a dual unscented Kalman filter and fractional variable-order model," *Energies*, vol. 10, no. 10, Art. 1577, Oct. 2017. [Online]. Available: https://doi.org/10.3390/en10101577

[13] M. B. Pinson and M. Z. Bazant, "Theory of SEI formation in rechargeable batteries: Capacity fade, accelerated aging, and lifetime prediction," *arXiv preprint arXiv:1210.3672*, Oct. 2012. [Online]. Available: https://arxiv.org/abs/1210.3672

[14] F. Zhao, Y. Guo, and B. Chen, "A review of lithium-ion battery state-of-charge estimation methods based on machine learning," *Machines*, vol. 15, no. 4, Art. 131, Mar. 2024. [Online]. Available: https://www.mdpi.com/2032-6653/15/4/131

[15] M. S. H. Lipu, T. Rahman, N. Ali, S. Ansari, and M. S. Mahmood, "Enhanced State of Charge Estimation in Lithium-Ion Batteries Using Gated Recurrent Unit Neural Networks with Gravitational Search Optimization," *GUB Journal of Science and Engineering*, vol. 9, no. 1, 2023. [Online]. Available: https://doi.org/10.3329/gubjse.v9i1.74879

[16] D. Liu, S. Wang, X. Li, Y. Fan, C. Fernandez, and F. Blaabjerg, "A novel extended Kalman filter-guided long short-term memory algorithm for power lithium-ion battery state of charge estimation at multiple temperatures," *Energy*, vol. 335, Art. 137973, Oct. 2025. [Online]. Available: https://doi.org/10.1016/j.energy.2025.137973

[17] Author(s), "Advanced integration of bidirectional long short-term memory neural network and innovative extended Kalman filter for state of charge estimation," *Journal of Power Sources*, 2024. [Online]. Available: https://doi.org/10.1016/j.jpowsour.2024.235893