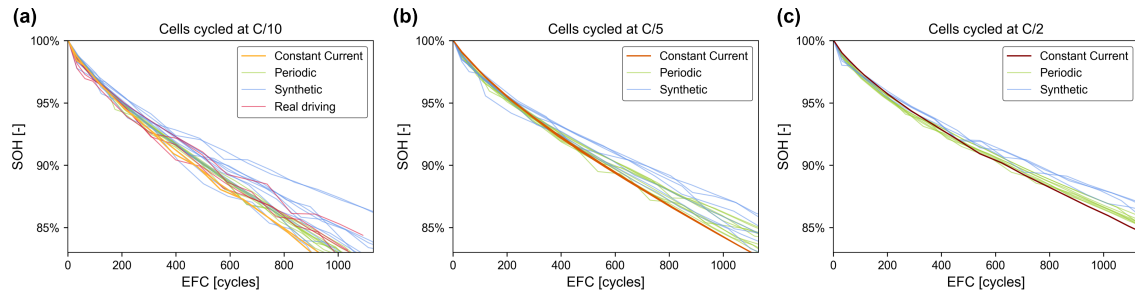

Dynamic cycling enhances battery lifetime

In the format provided by the
authors and unedited

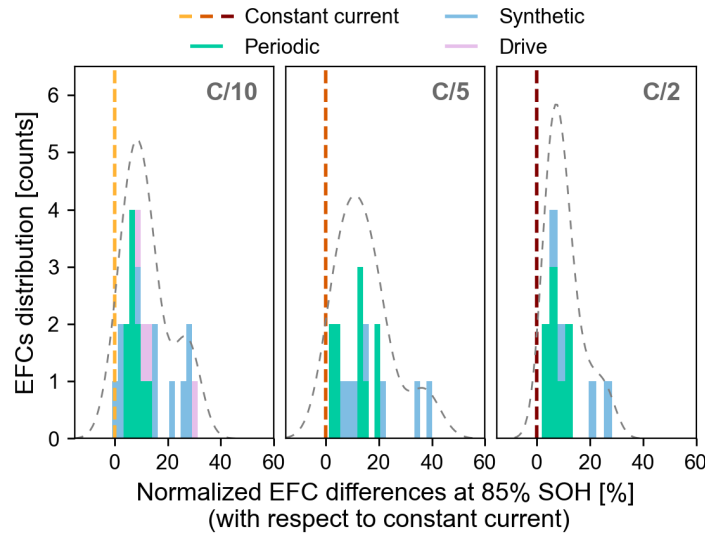
Contents

Supplementary Figures	S2
Supplementary Tables	S24
Supplementary Methods	S27
Supplementary Notes	S35
Supplementary References	S35

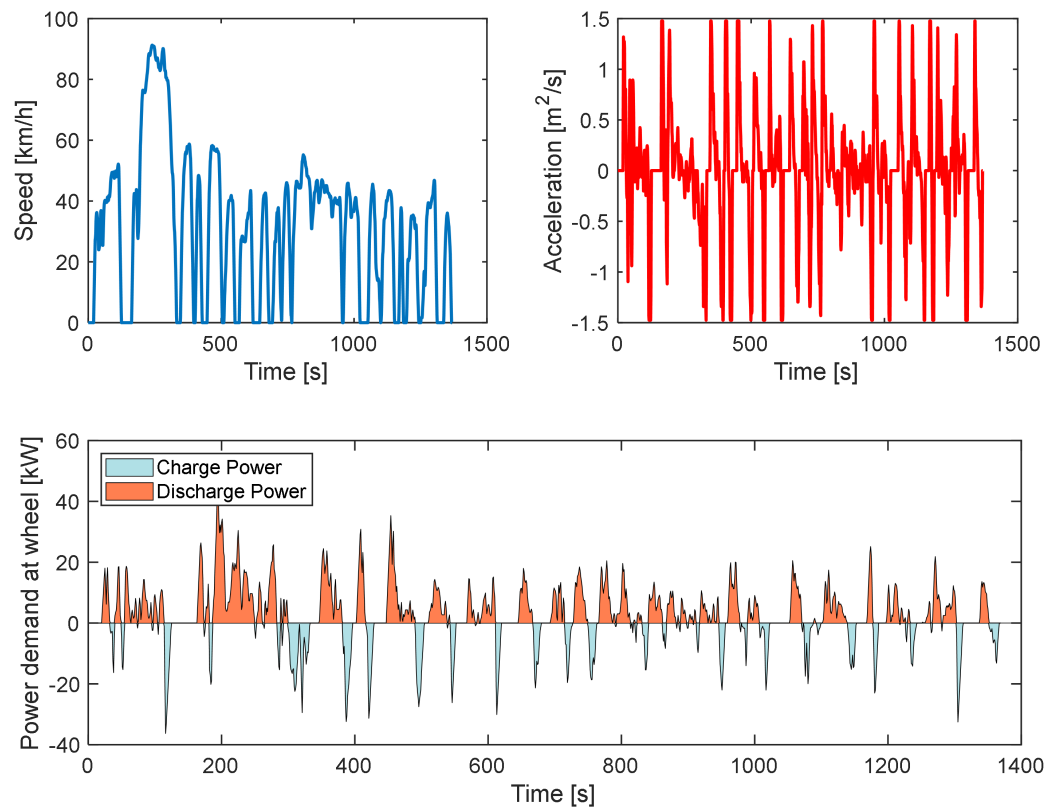
Supplementary Figures



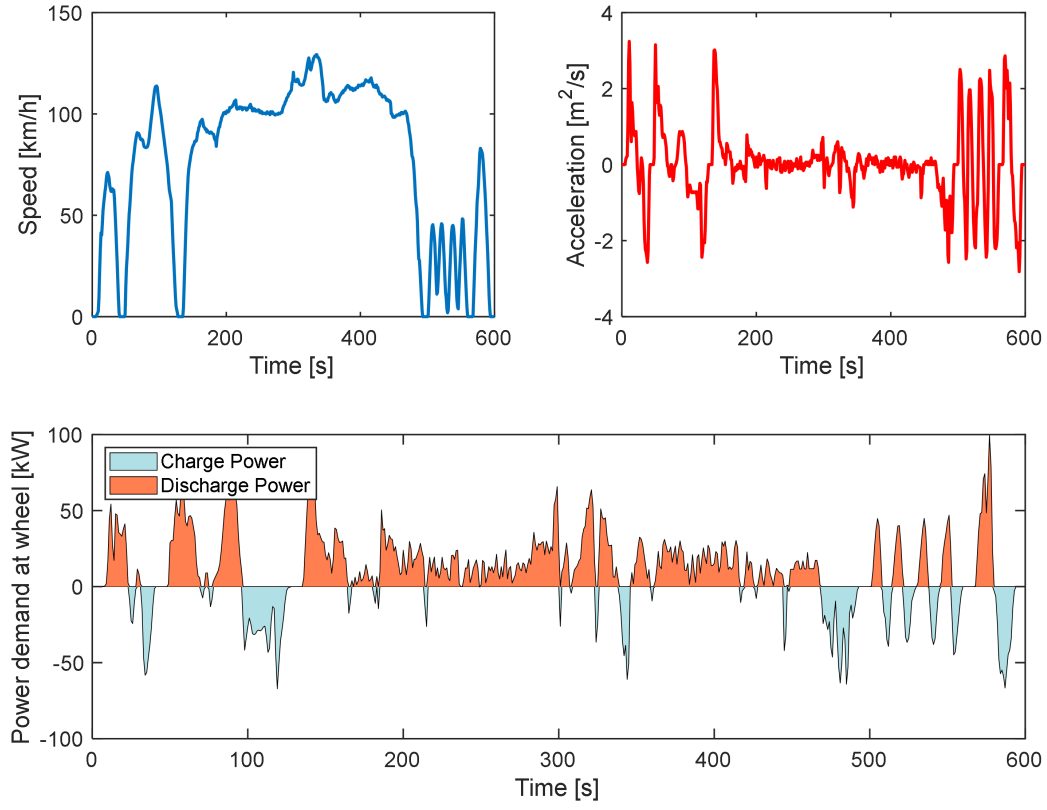
Supplementary Figure 1: Degradation trajectories as a function of EFC for cells cycled at C/10 (a), C/5 (b), and C/2 (c). Protocol types are represented by colors. From constant current to periodic profiles to synthetic profiles, the more realistic the profile, the more gain in lifetime. For C/10, real driving trajectories are varying in the range of the synthetic profiles.



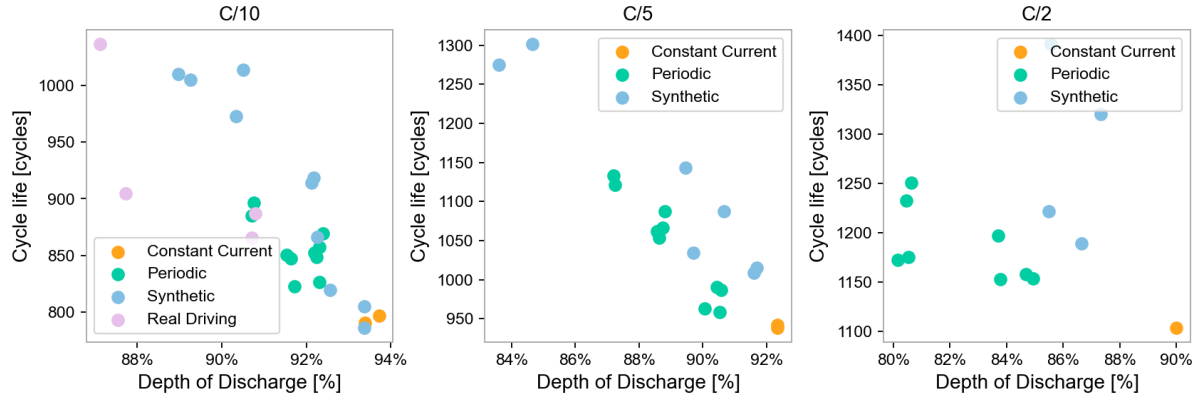
Supplementary Figure 2: Distribution of EFC differences with respect to constant current cycling at 85 %. Protocol types are represented by colors.



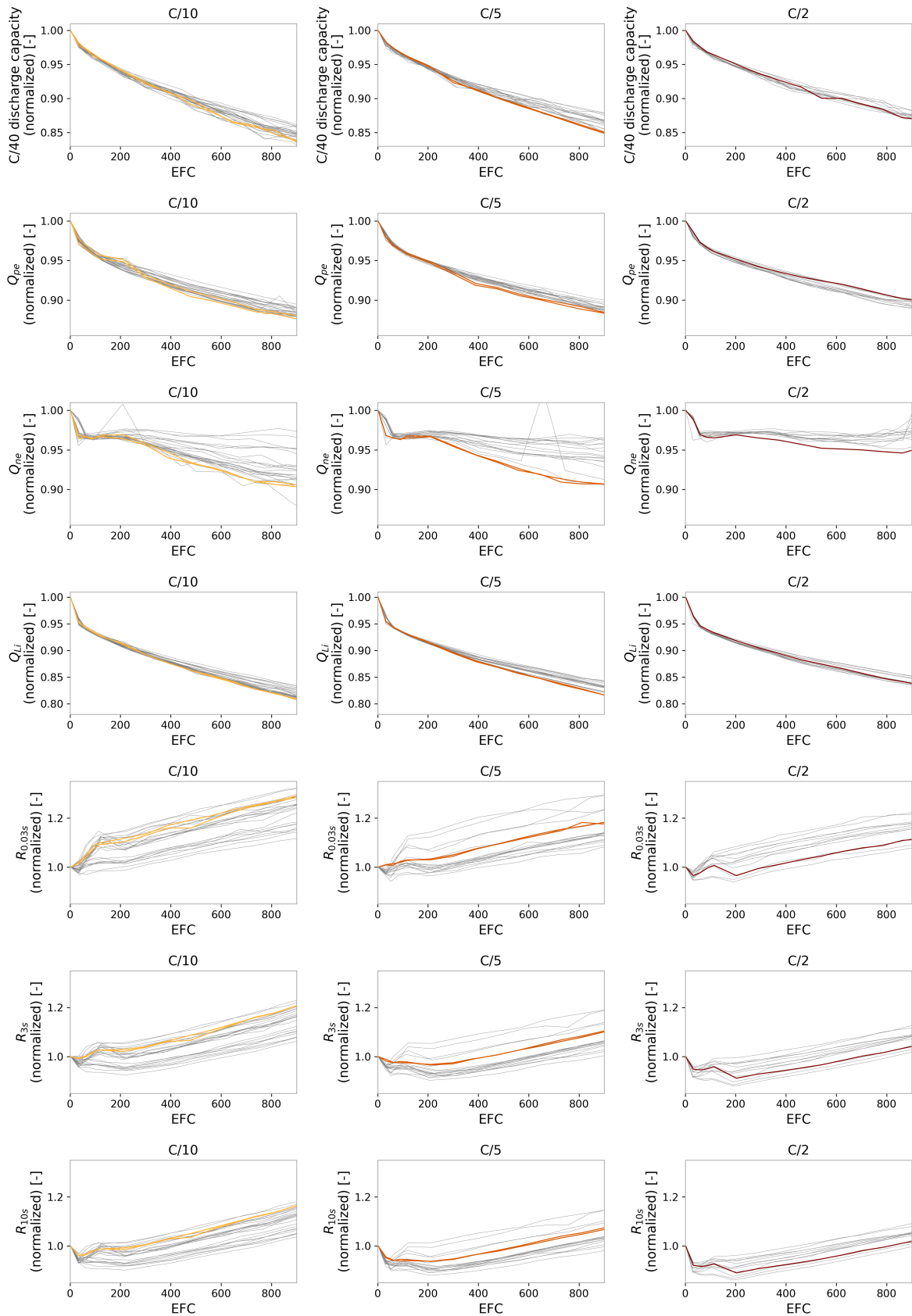
Supplementary Figure 3: Driving profile and power demand for UDDS.



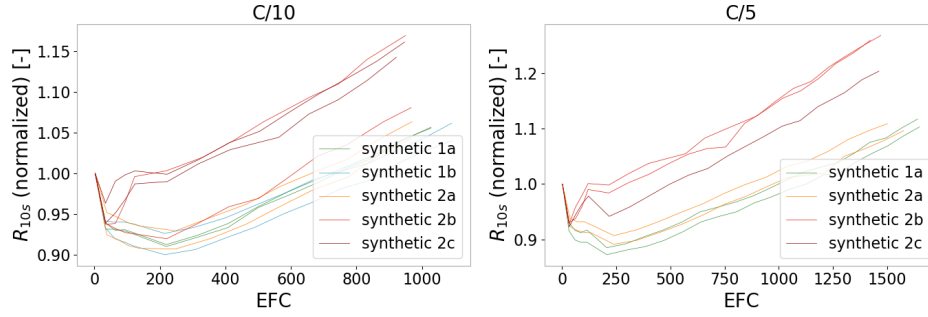
Supplementary Figure 4: Driving profile and power demand for US06.



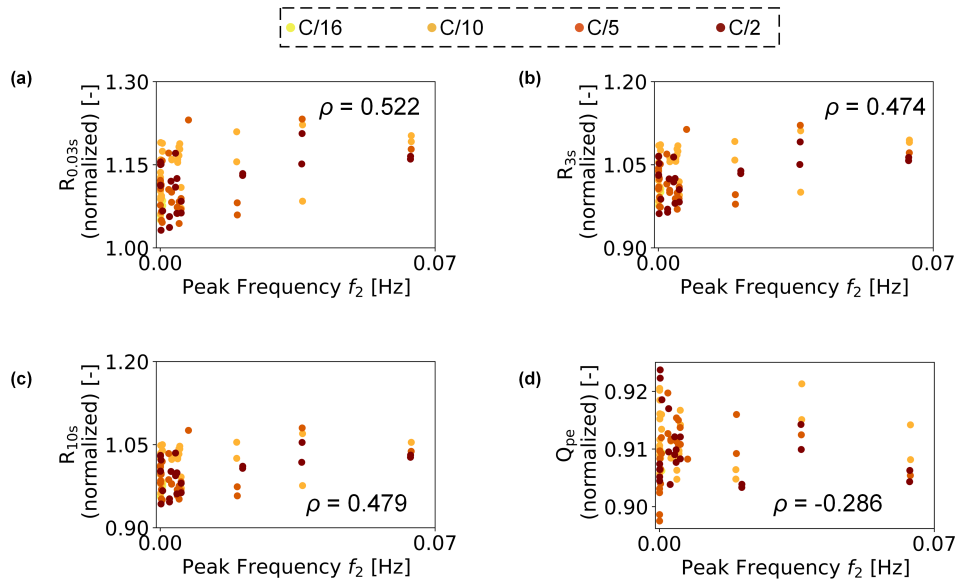
Supplementary Figure 5: Cycle life (at 85 % SOH) versus DoD for cells cycled at the same C-rate. At high C-rate, no correlation is observed. At lower C-rates, a correlation trend emerges, indicative of silicon degradation. However, when isolating cells within the same DoD (DoD window of 4%), the spread in EFC with respect to constant current reaches 28 % and 22 % for C/10 and C/5 respectively, in line with results in Figure 2b.



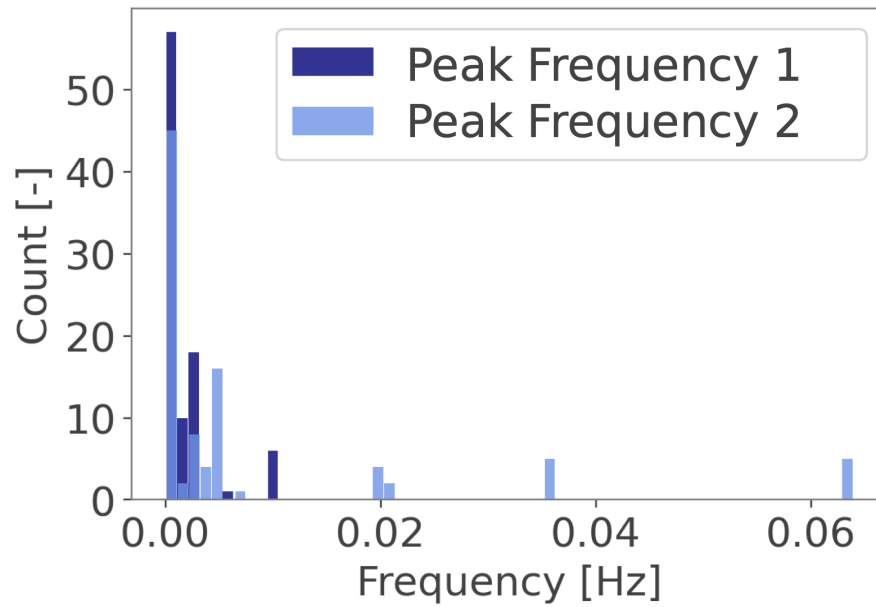
Supplementary Figure 6: Low rate capacity, electrode specific capacities, lithium inventory and resistances also exhibit a large spread in degradation trends across the dataset. Constant current protocols are shown in color.



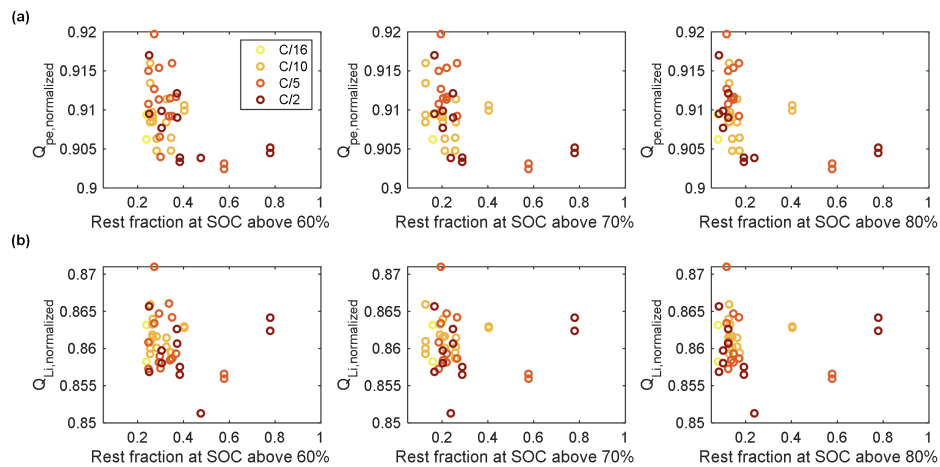
Supplementary Figure 7: Within the synthetic duty cycles, resistance growth varies widely. In particular, synthetic profile 2b and 2c which are derived from city driving, appear to have more resistance growth than profiles generated from highway driving.



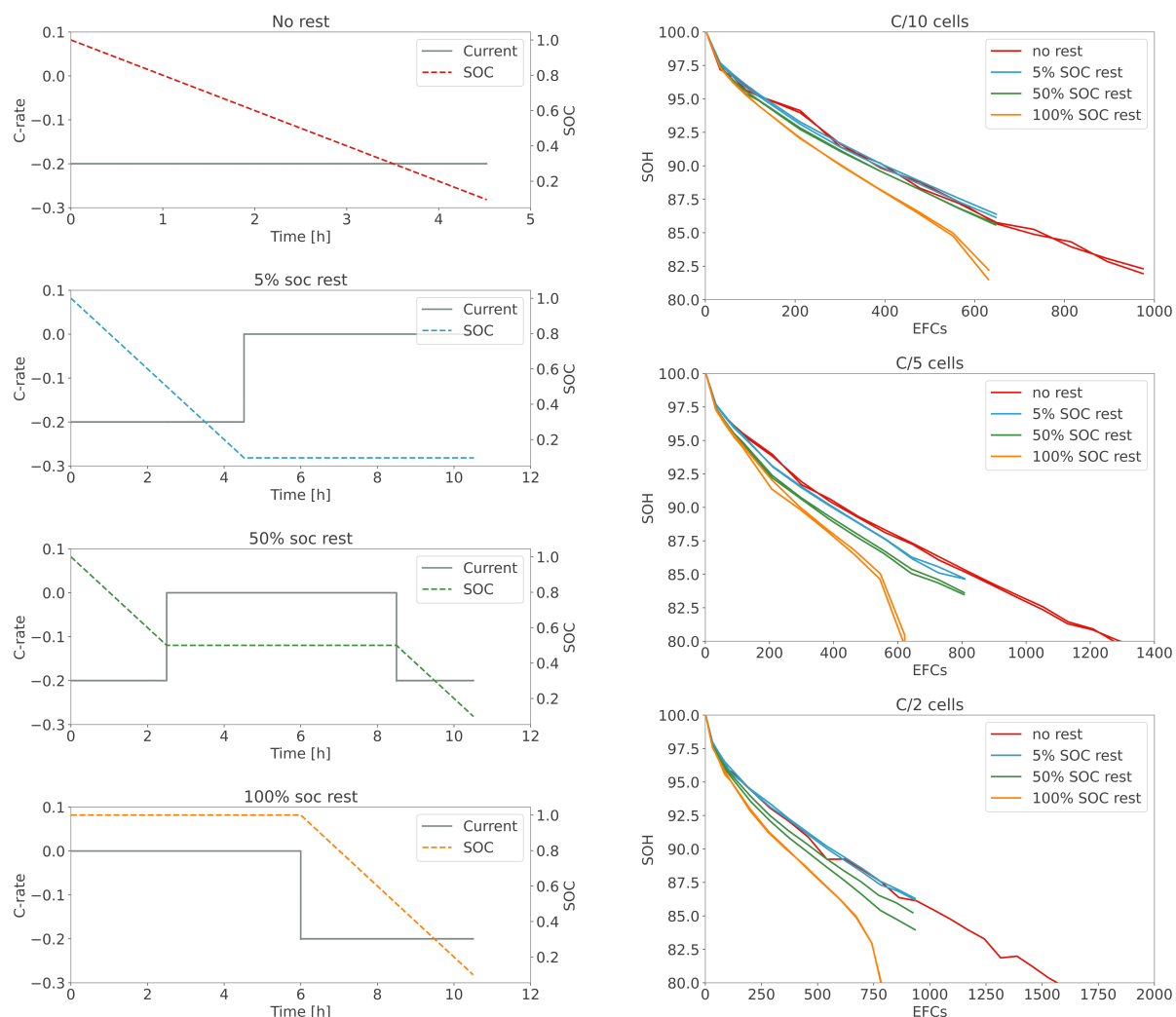
Supplementary Figure 8: Correlations between relevant features and f_2 . (a) $R_{0.03s}$ and f_2 , (b) R_{3s} and f_2 , (c) R_{10s} and f_2 , and (d) Q_{pe} and f_2 .



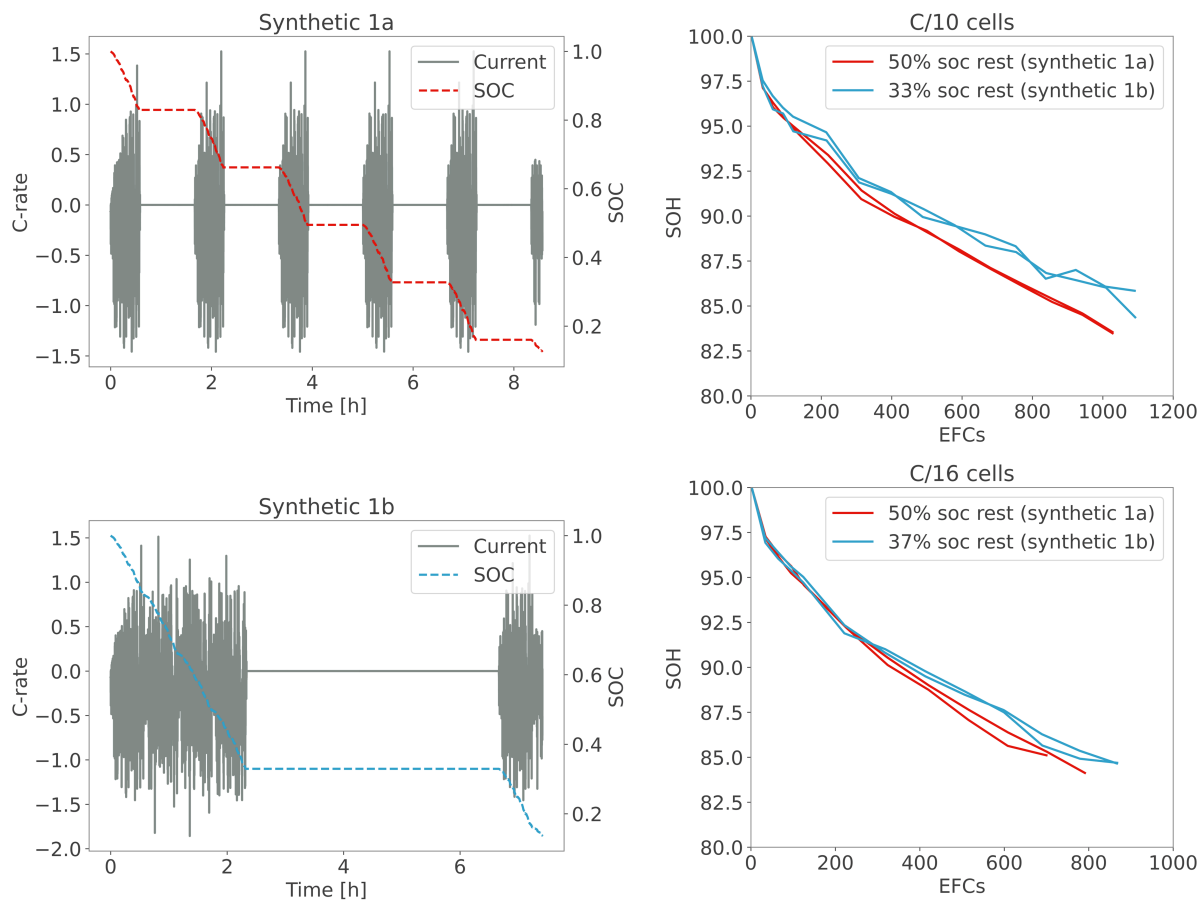
Supplementary Figure 9: Distributions of peak frequencies 1 & 2 from aging cycle protocol parametrization.



Supplementary Figure 10: **a)** Q_{pe} versus rest fraction at elevated SOC, either above 60 % (left), above 70 % (middle), or 80 % (right). **b)** Q_{Li} versus rest fraction at elevated SOC, either above 60 % (left), above 70 % (middle), or 80 % (right).



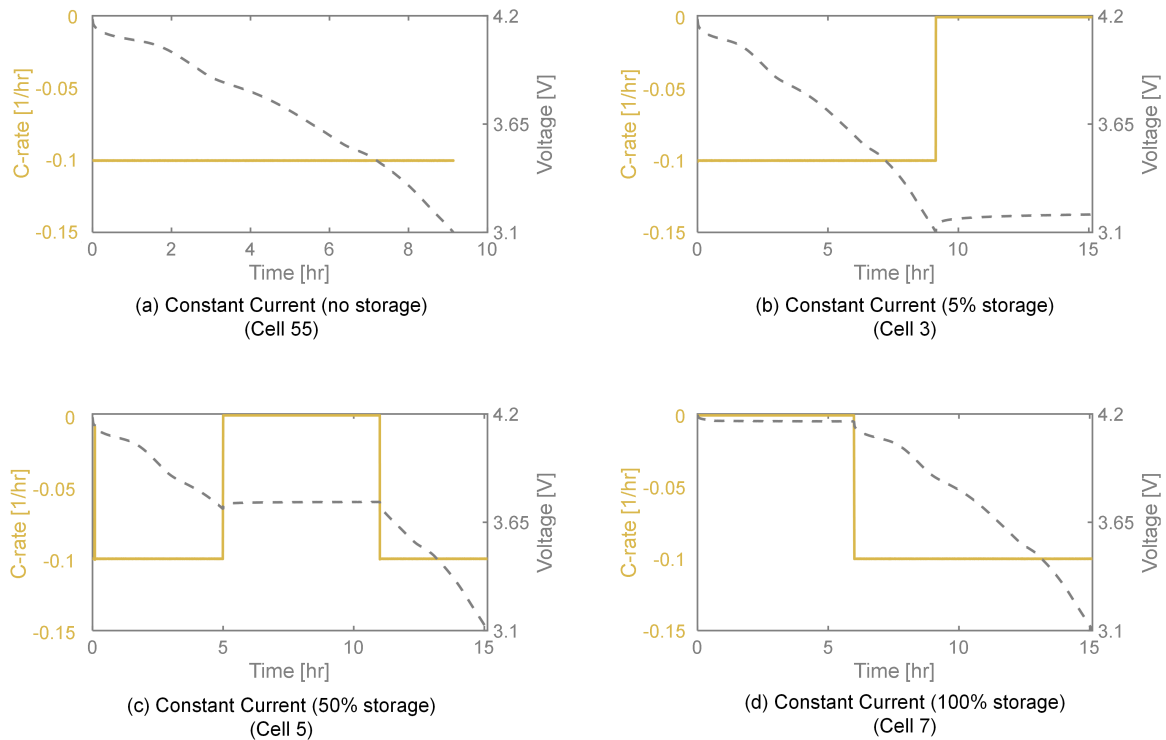
Supplementary Figure 11: Differences in degradation trajectories between cells cycled with or without a rest period, at various SOC levels. The dependence of calendar aging on SOC is also strong when combined with cycling. The higher the SOC during rest period, the faster the degradation, despite cycling.



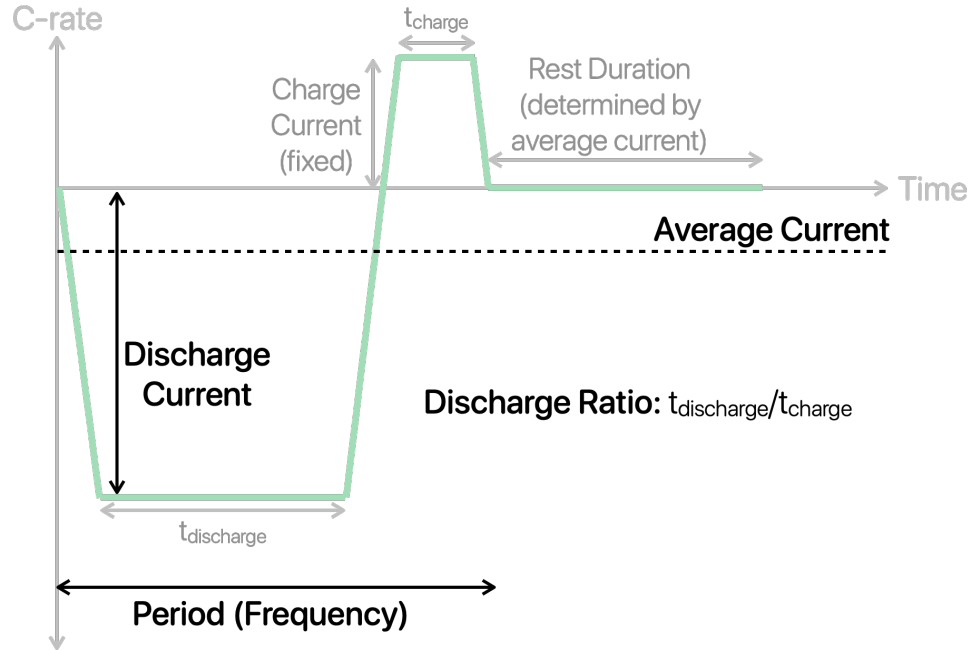
Supplementary Figure 12: The path-dependence observed by Gering *et al.*[1] is also observed here. Rest periods at higher SOC levels lead to accelerated degradation.



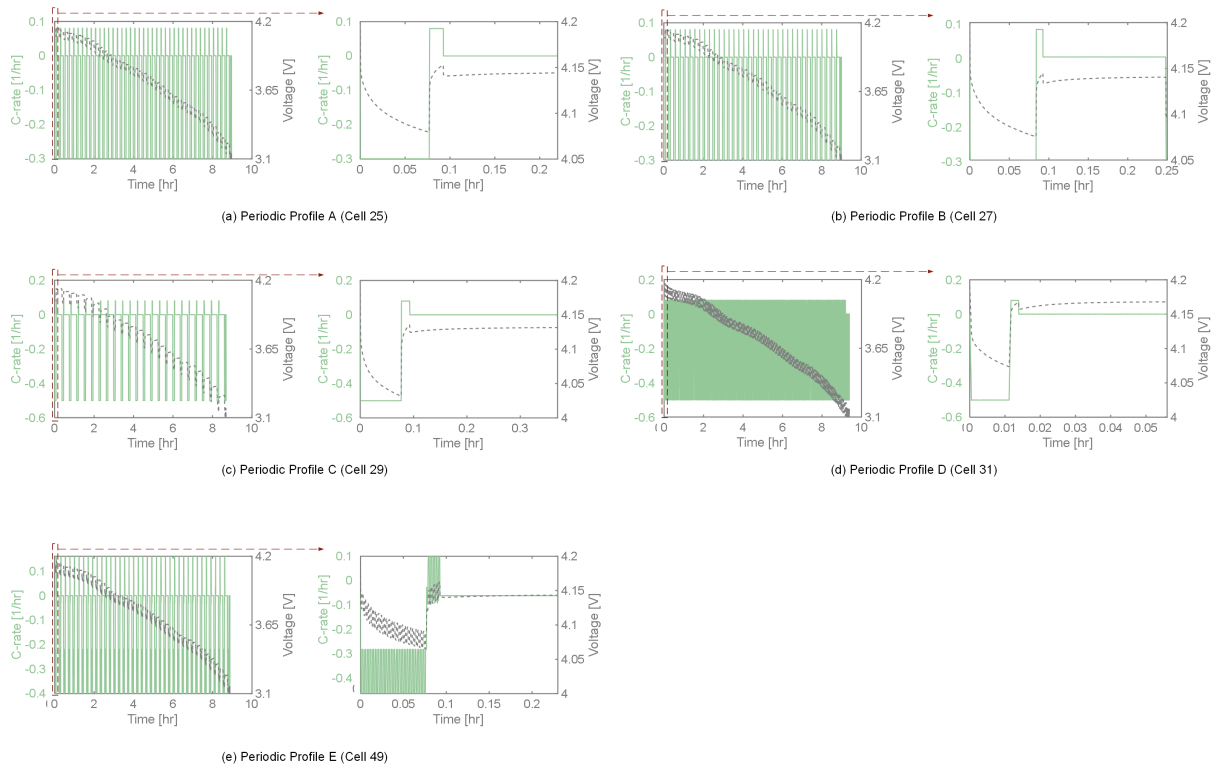
Supplementary Figure 13: Photos of battery cyclers used for battery cycling experiments. The photo shows the CSZ temperature chamber (front-most object, off-white instrument with blue marking in the middle) and Maccor battery cycler (directly to the left, black instrument with white cables coming out) used for the experiments.



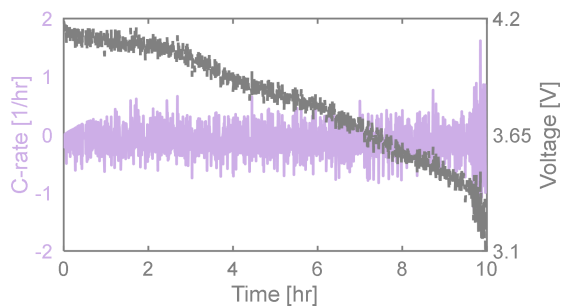
Supplementary Figure 14: Realistic study constant current profiles (shown for C/10 average current only).



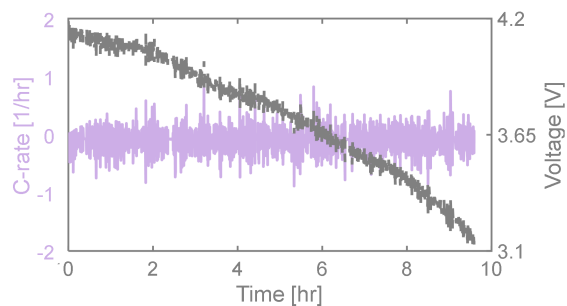
Supplementary Figure 15: Periodic protocol structure. All the parameters that are used to describe the periodic protocols are labeled on the plot. The “free” parameters are shown in black, while the fixed or dependent parameters are shown in grey.



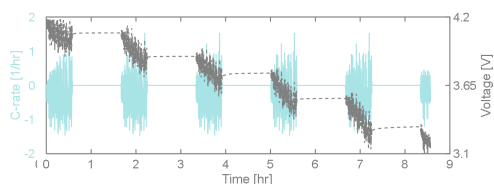
Supplementary Figure 16: Realistic study periodic profiles (shown for C/10 average current only).



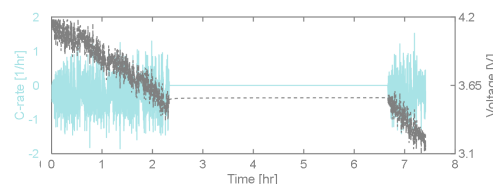
(a) City 1 Drive Profile (Cell 91)



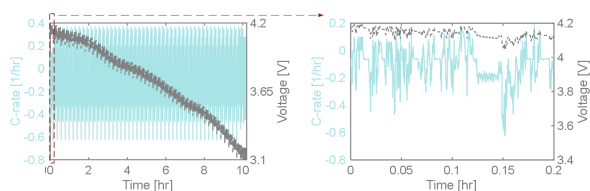
(b) City 2 Drive Profile (Cell 95)

Supplementary Figure 17: Realistic study drive profiles (shown for C/10 average current only).

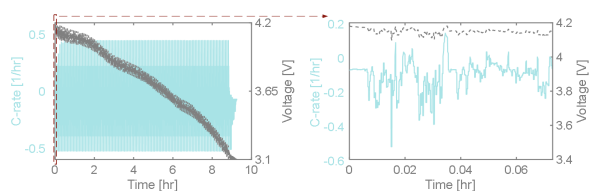
(a) Synthetic Profile 1a (Cell 59)



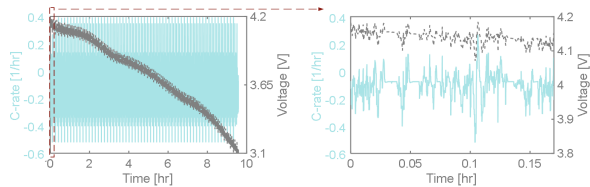
(b) Synthetic Profile 1b (Cell 71)



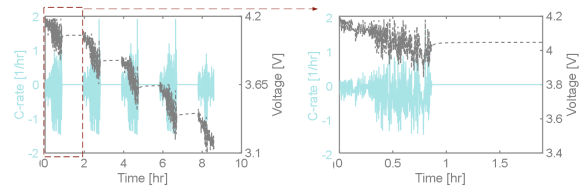
(c) Synthetic Profile 2a (Cell 65)



(d) Synthetic Profile 2b (Cell 73)

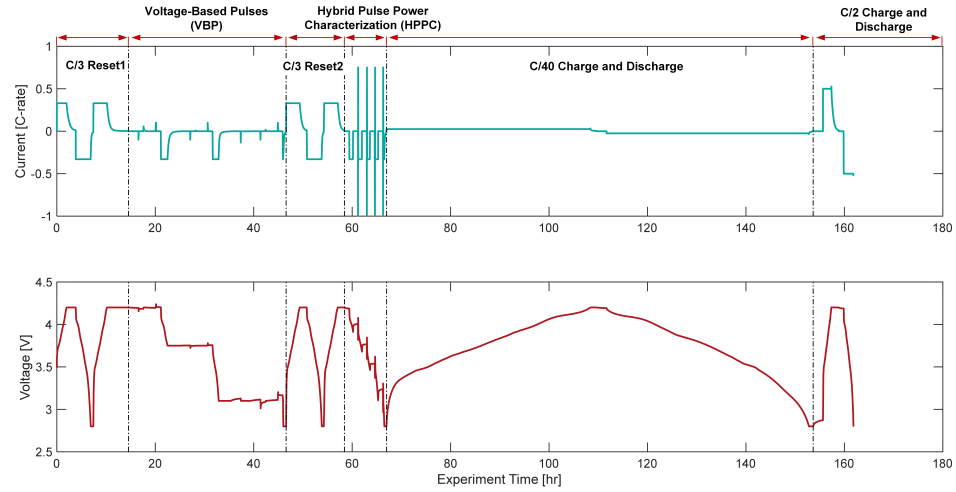


(e) Synthetic Profile 2c (Cell 81)

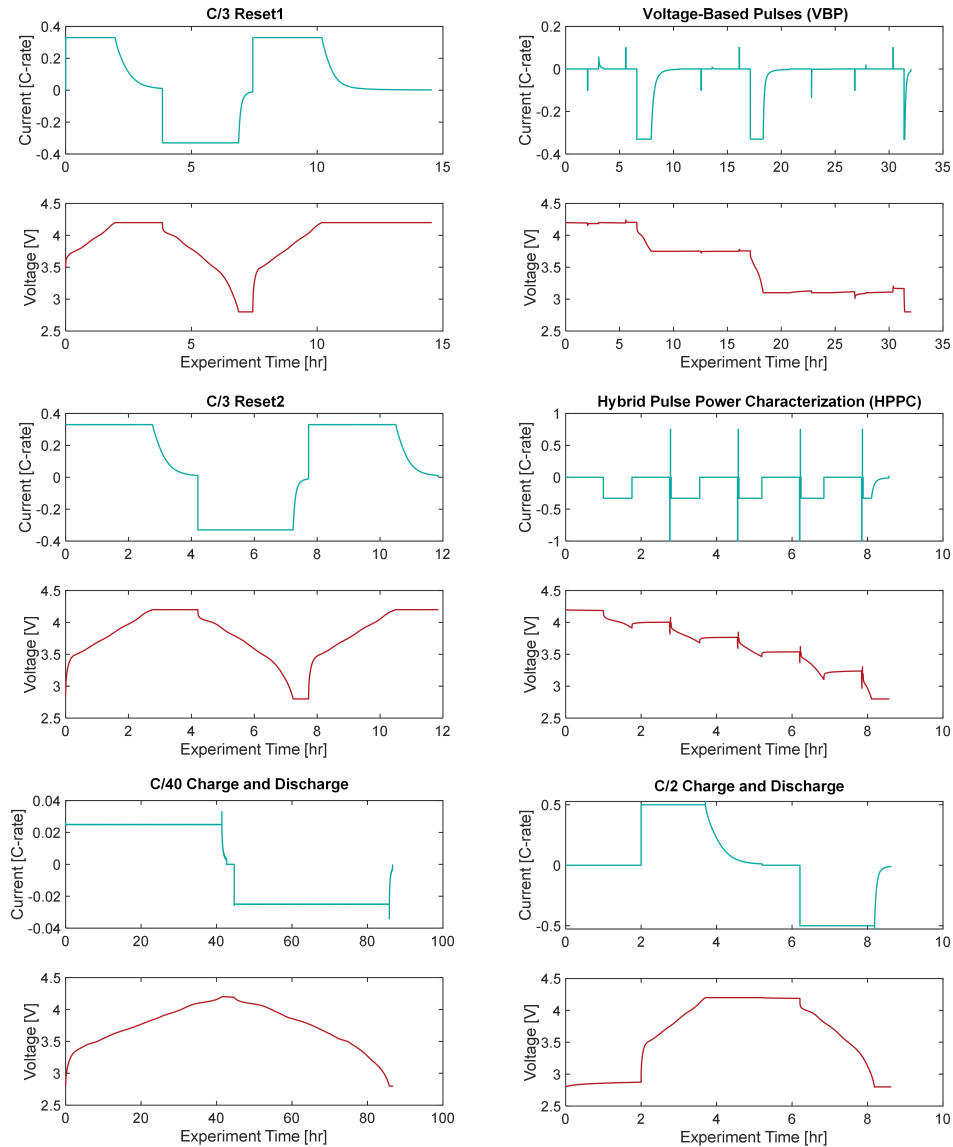


(f) Synthetic Profile 3 (Cell 79)

Supplementary Figure 18: Realistic study synthetic profiles (shown for C/10 average current only).

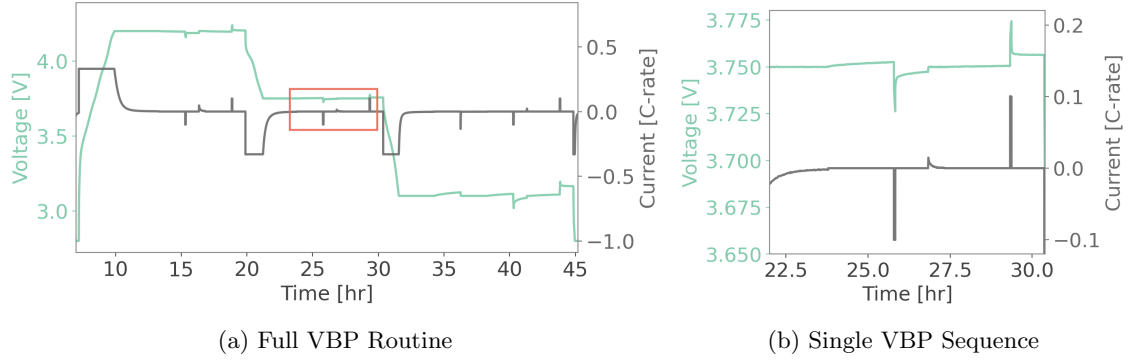


(a) Entire diagnostic procedure

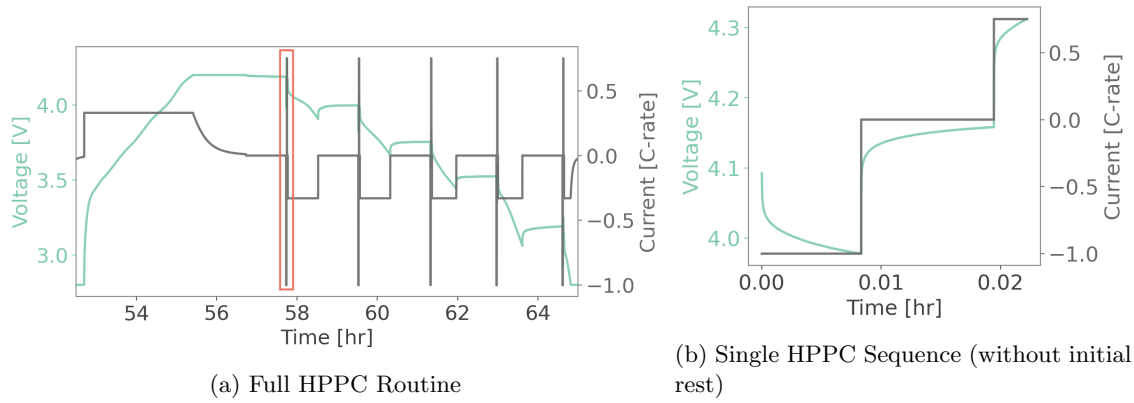


(b) Individual components of the diagnostic protocol

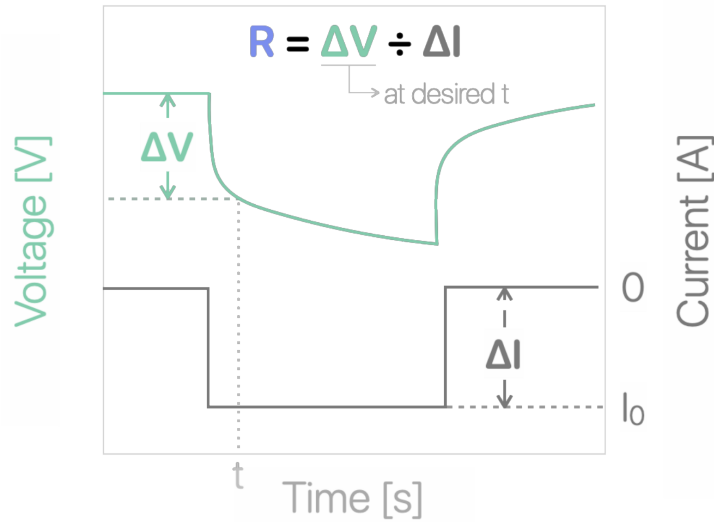
Supplementary Figure 19: Current and voltage for the diagnostic cycle protocol (a) and individual components (b).



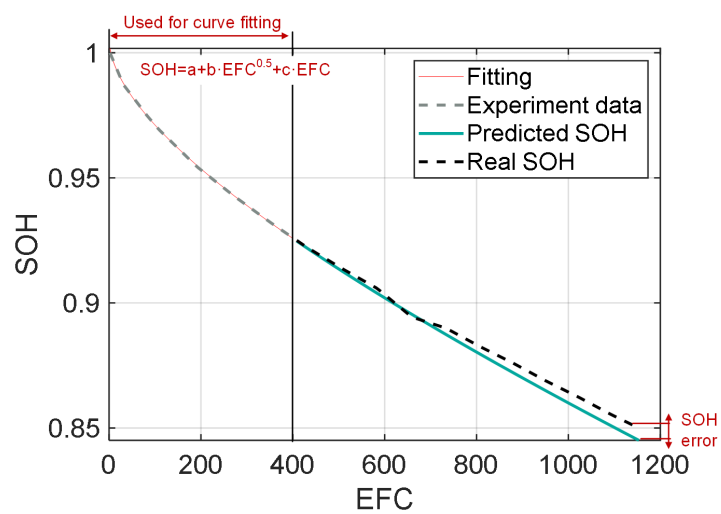
Supplementary Figure 20: VBP routine and single sequence for a single cell near beginning of life. The right figure is zoomed in to the region outlined in the red box in the left figure.



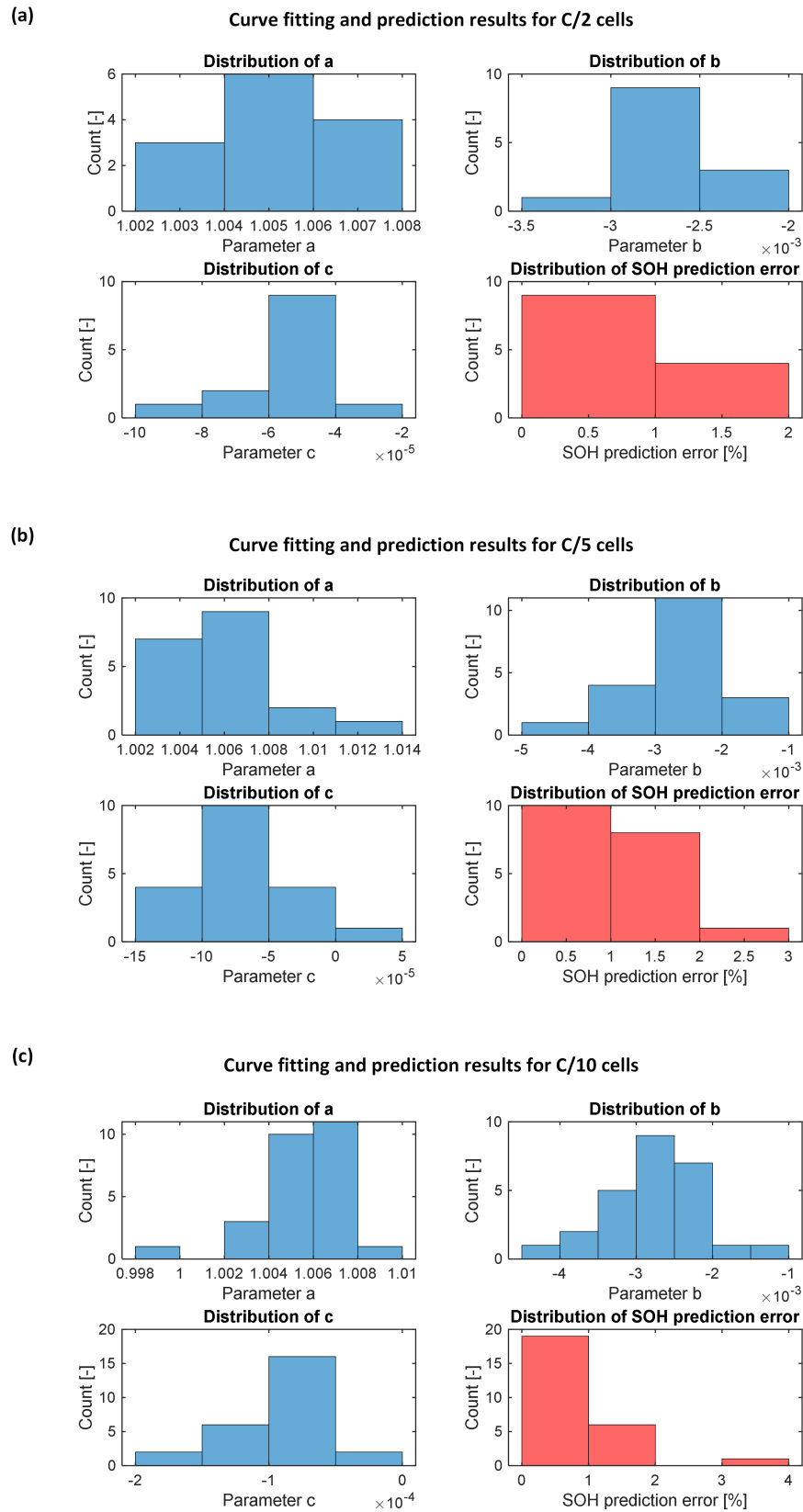
Supplementary Figure 21: a) HPPC routine and single sequence for a single cell at BoL. b) Zoom into the region outlined in the red box in the left figure.



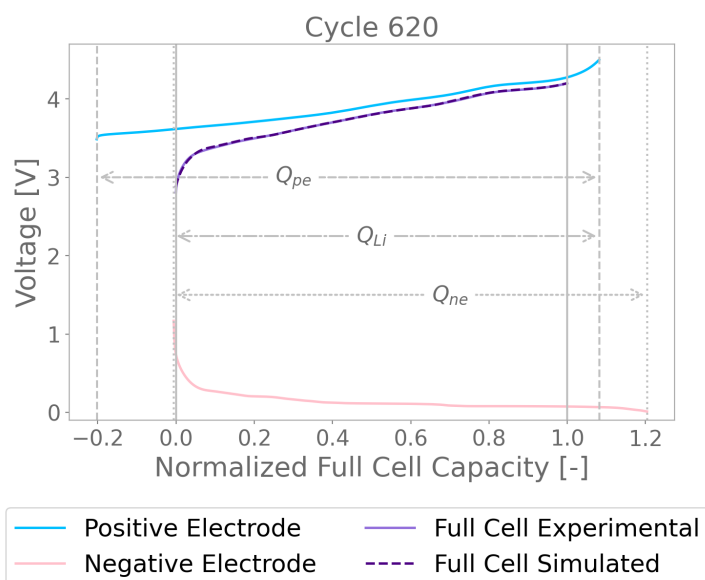
Supplementary Figure 22: Resistance calculation from current pulses assuming simple ohmic relationship, adapted from [2].



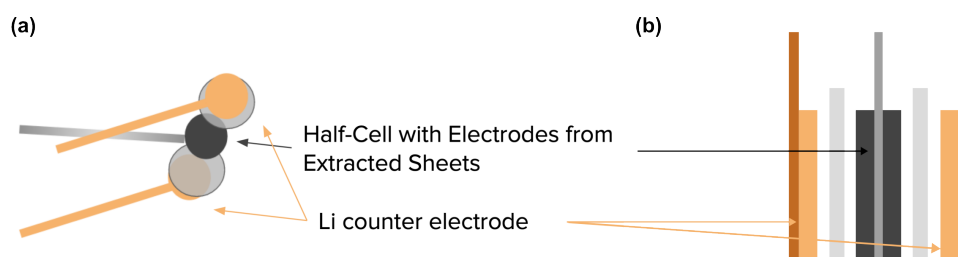
Supplementary Figure 23: SOH degradation trajectory fitting example for cell 041.



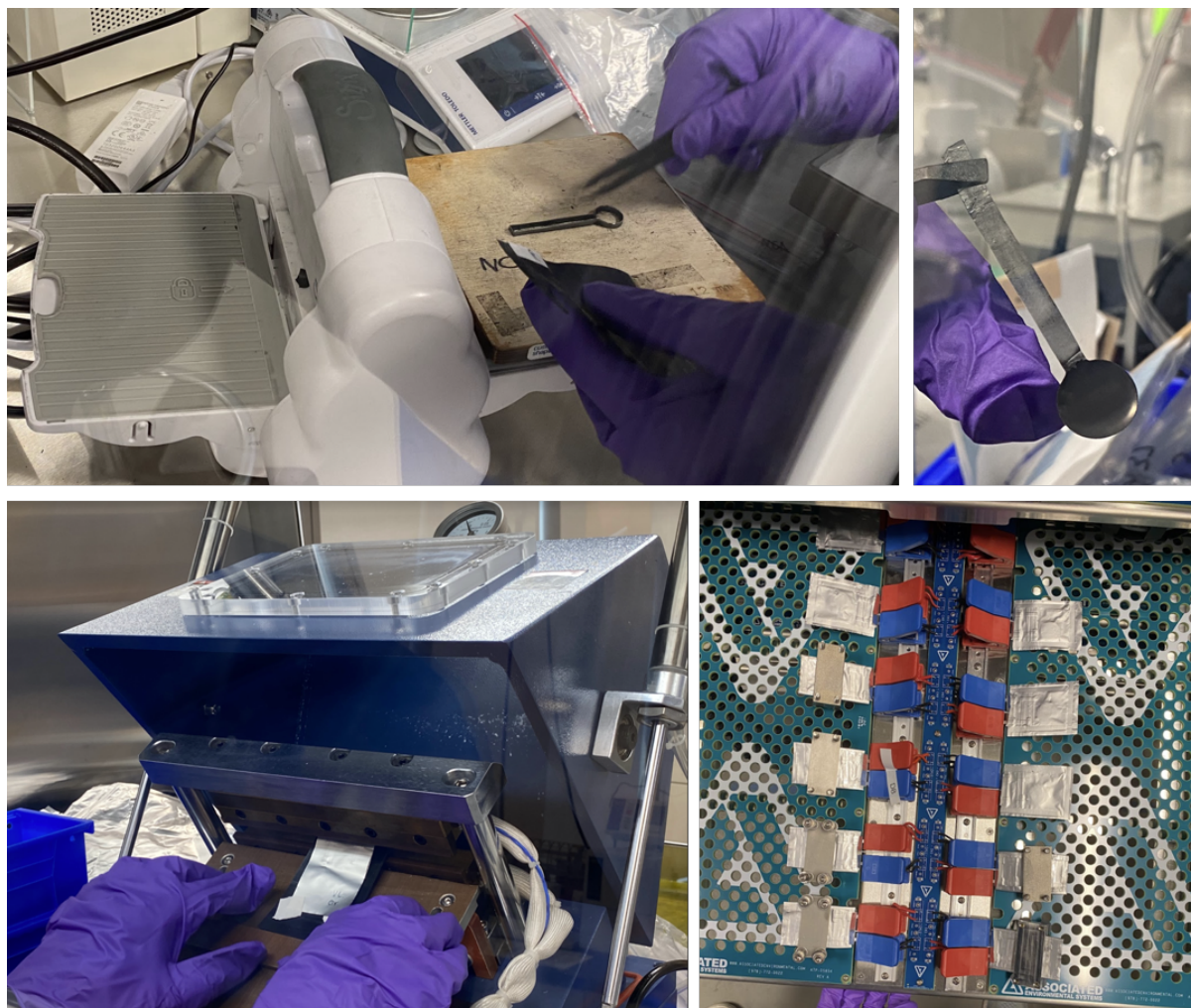
Supplementary Figure 24: Fitting results to predict EOL based on the first 400 EFCs.



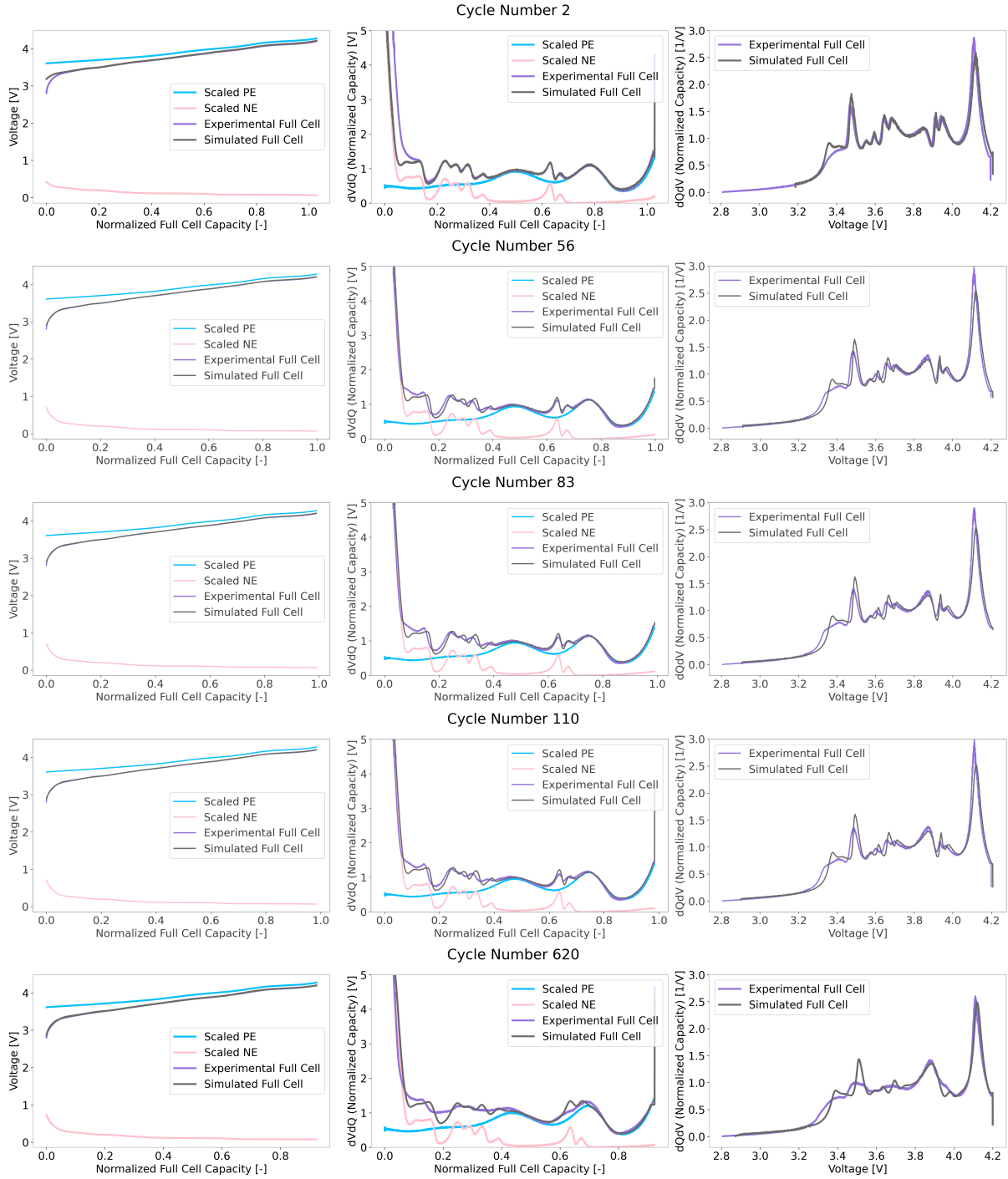
Supplementary Figure 25: Example of the results of the half-cell differential voltage fitting analysis and its outputs for one aged cell (periodic profile D - C/2).



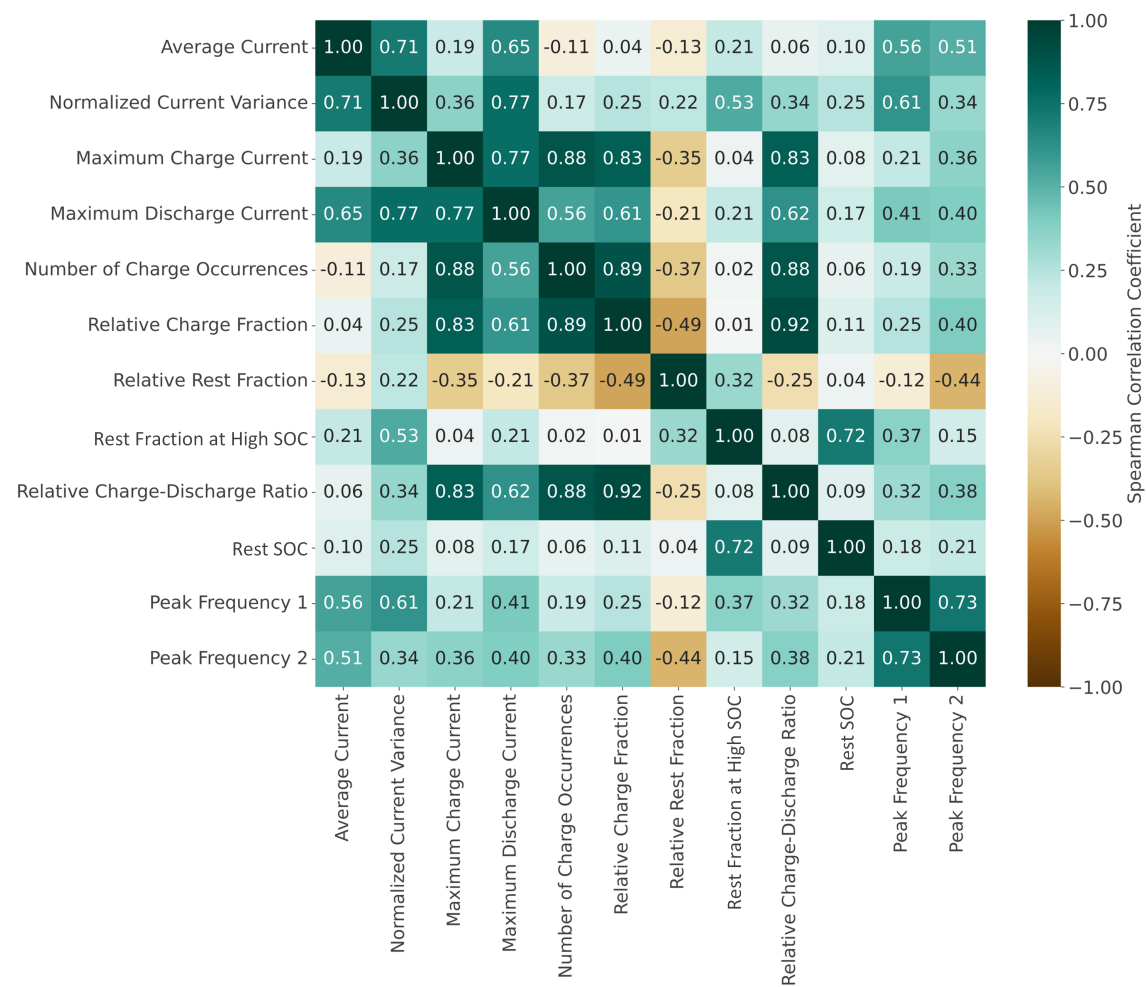
Supplementary Figure 26: Schematic of half-cell assembly structure (adapted from [3]). **a)** Top view. **b)** Side view.



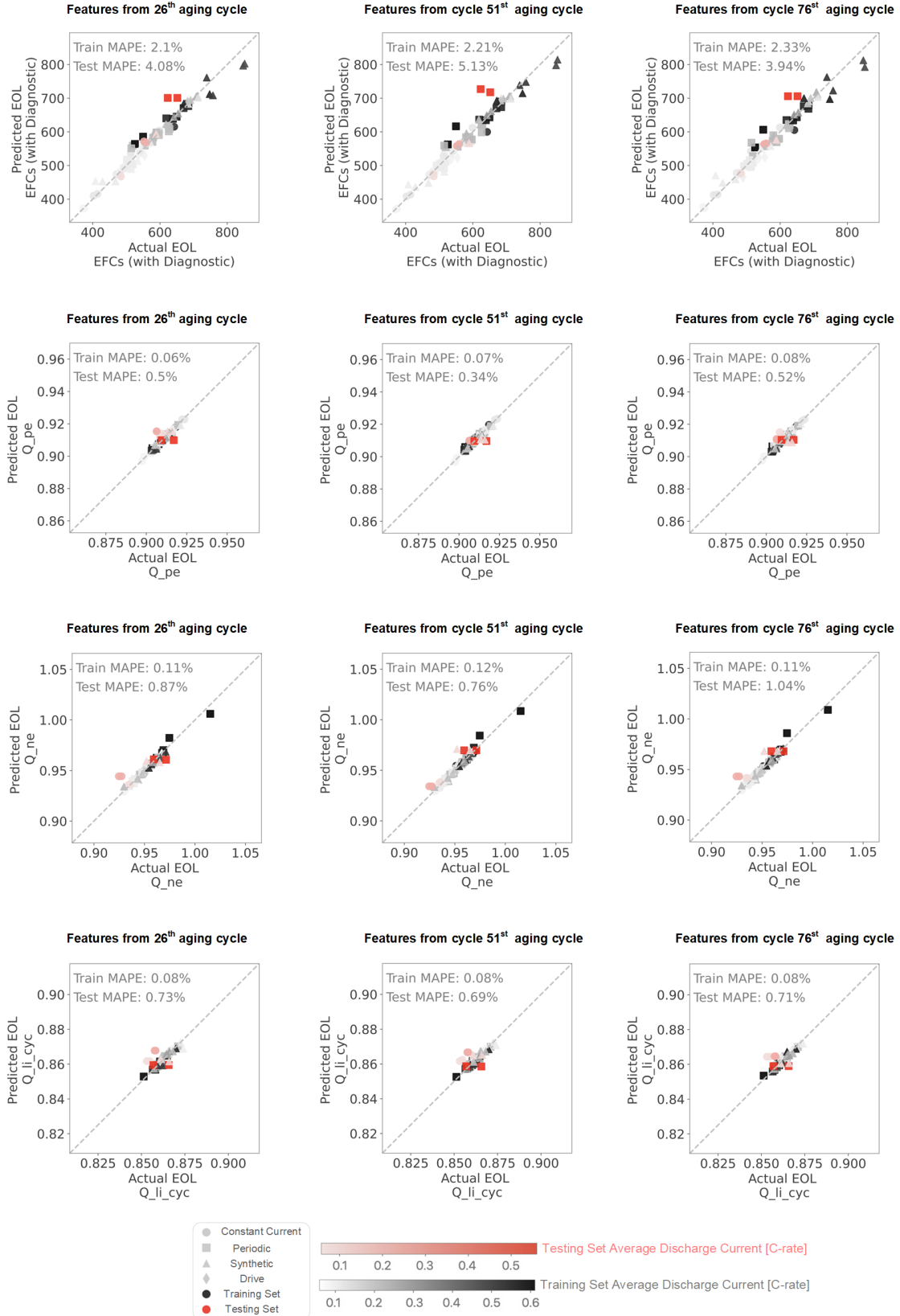
Supplementary Figure 27: Images from half cell fabrication process. The upper left image shows us punching out electrodes in the correct shape for the half cell from the original electrode sheet. The upper right image shows a punched electrode. The lower left image shows us sealing a pouch cell created from the punched electrodes with lithium metal counter electrodes. The lower right image shows the pouch cells cycling in a battery cycler to collect low-rate voltage profiles.



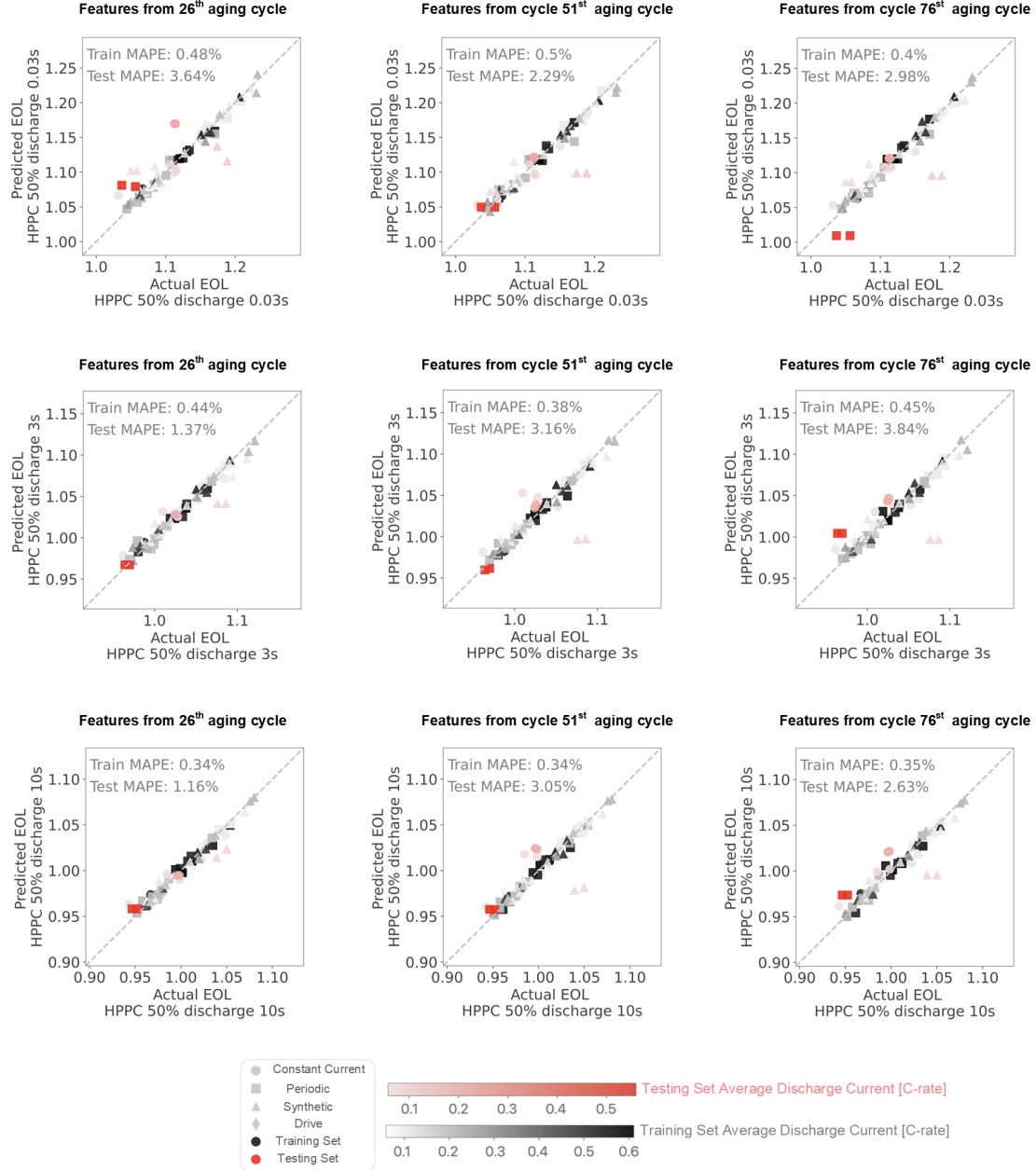
Supplementary Figure 28: Half-cell differential voltage fitting example plots for a pristine and aged cell (periodic profile D - C/2). The $V - Q$ curves (left), $\frac{dV}{dQ}$ curves (center) and $\frac{dQ}{dV}$ curves (right) are represented, normalized by the cell capacity.



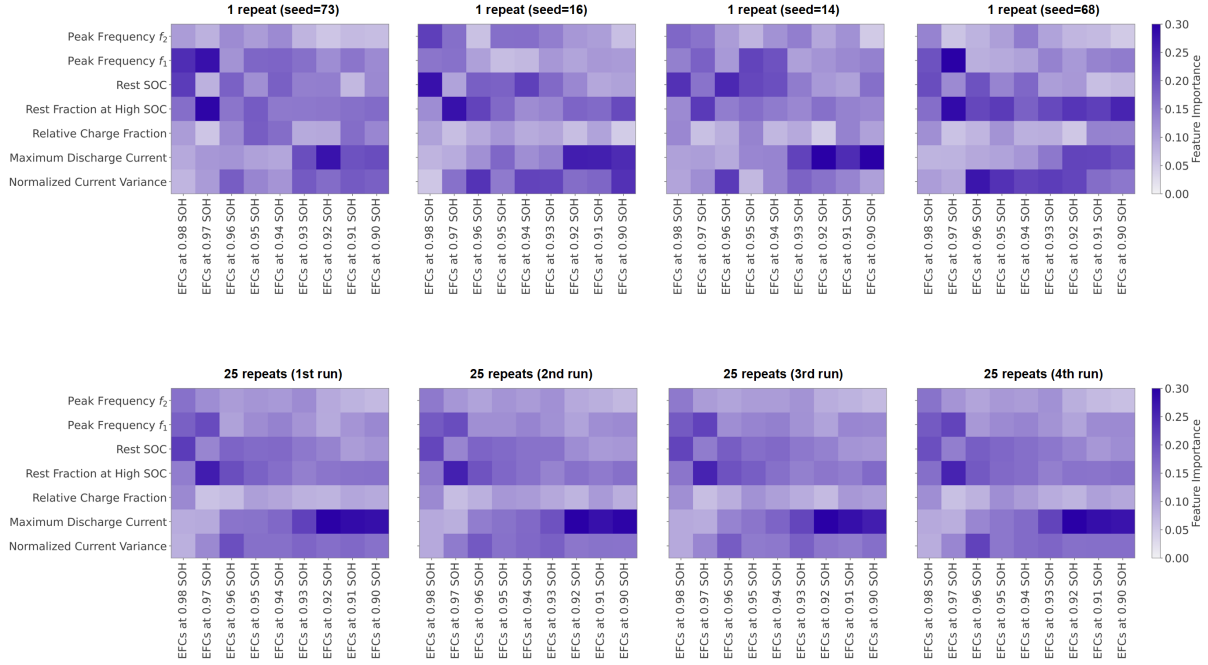
Supplementary Figure 29: Realistic study aging cycle parameter correlations.



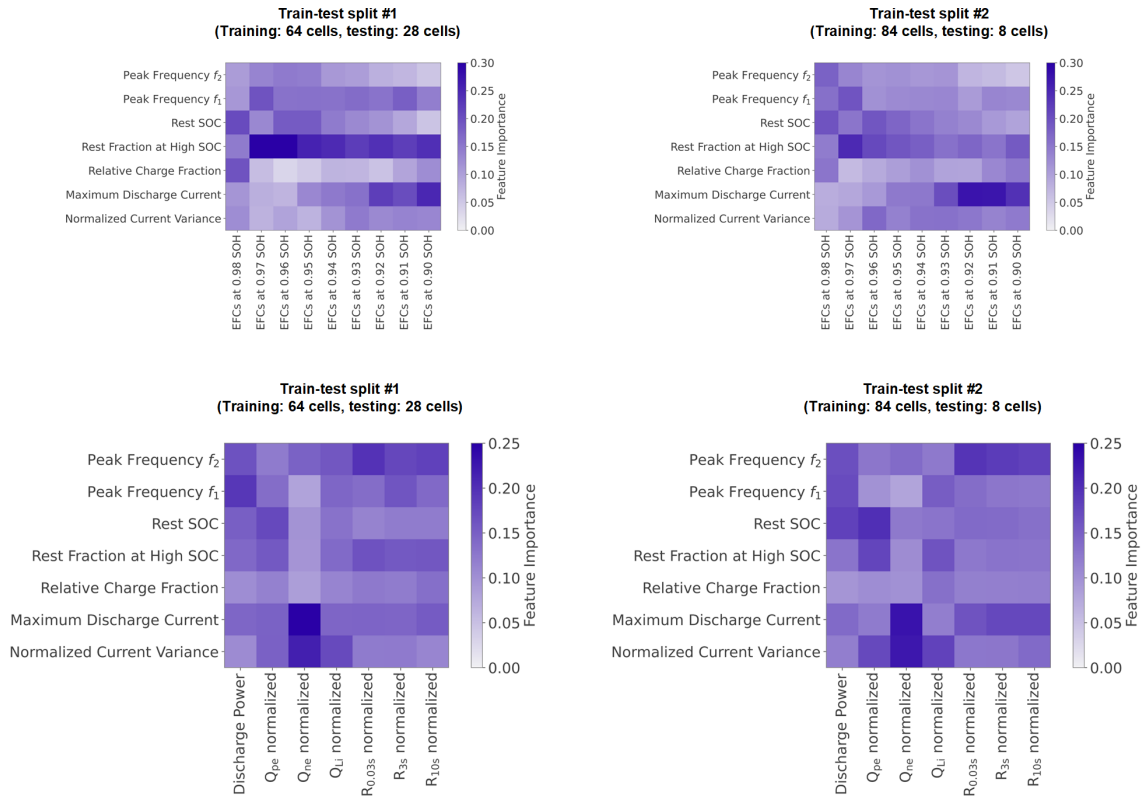
Supplementary Figure 30: ML model parity plots. Training and test set errors for EFC, Q_{pe} , Q_{ne} and Q_{Li} are shown for each plot.



Supplementary Figure 31: ML model parity plots. Training and test set errors for $R_{0.03s}$, R_{3s} and R_{10s} are shown for each plot.



Supplementary Figure 32: Repeatability and consistency of SHAP results require multiple random seed runs.



Supplementary Figure 33: Results of Figure 3 are consistent across different train-test splits.

Supplementary Tables

Description	value
m_v vehicle mass [kg]	2087
c_r rolling friction coefficient [-]	0.01
A_f Frontal Area [m^2]	2.34
ρ_{air} air density [kg/m^3]	1
c_d drag coefficient [-]	0.24

Supplementary Table 1: Parameters for a typical electric vehicle.

	UDDS (mild)		US06 (aggressive)	
	CC baseline [miles]	Dynamic cycling gain [miles]	CC baseline [miles]	Dynamic cycling gain [miles]
C/10	457,608	71,178	319,477	49,740
C/5	539,163	121,540	376,395	84,848
C/2	628,081	104,638	438,501	73,073

Supplementary Table 2: Maximum mileage gain for dynamic cycling versus constant current cycling.

Parameter	Meaning	Used in ML Model?
Average Current	Overall protocol average current	✓
Normalized Current Variance	Overall protocol current variance divided by average current	✓
Maximum Charge Current	Absolute charge current maximum	✓
Maximum Discharge Current	Absolute discharge current maximum	
Number of Charge Occurrences	Number of times current set point switches to charge mode from either rest or discharge	✓
Relative Charge Fraction	Fraction of protocol time spent charging ($I > 0$)	
Relative Rest Fraction	Fraction of protocol time spent resting ($I = 0$)	
Rest Fraction at High SOC	Fraction of protocol time spent resting ($I = 0$ and $SOC > 0.6$)	✓
Charge-Discharge Ratio	Charging time divided by discharging time	✓
Rest SOC	Weighted (time-based) average of SOC's spent at rest (capacity-calculated)	
Peak Frequency 1	Intermediate Peak Frequency	✓
Peak Frequency 2	Highest Peak Frequency	✓

Supplementary Table 3: Parameters calculated from realistic aging cycle protocols. Not all parameters are used in model development. Parameters that are used are indicated with a ✓.

Protocol Type	Profile	Description	Average Current			
			C/16	C/10	C/5	C/2
Constant Current	No Rest			55, 56	9, 10	17, 18
	5 % SOC Rest			3, 4,	11, 12	19, 20
	50 % SOC Rest			5, 6	13, 14	21, 22
	100 % SOC Rest			7, 8	15, 16	23, 24
Periodic	A	Reference		25, 26	33, 34	41, 42
	B	Charge-Discharge Ratio		27, 28	35, 36	43, 44
	C	High Current Amplitude		29, 30	37, 38	45, 46
	D	Frequency		31, 32	39, 40	47, 48
	E	Superposition		49, 50	51, 52	53, 54
Synthetic	1a	Highway (short trips)	57, 58	59, 60	61, 62	
	1b	Highway (long trips)	63, 64	71, 72		
	2a	Urban (City 1)		65, 66	67, 68	69, 70
	2b	Urban (City 2)		73, 74	75, 76	77, 78
	2c	Urban (City 1 + 2)		81, 82,	83, 84	85, 86
	3	Highway + Urban		79, 80	87, 88	
Drive	City 1	Full Day	89, 90	91, 92		
	City 2	Full Day	93, 94	95, 96		

Supplementary Table 4: List of all cells and protocols in realistic usage conditions dataset. Rows indicate protocol types and profiles. Columns indicate average currents at which those profiles are cycled. The entry in each location indicates the cell/channel number that was cycled with that protocol at that average current. Cell 18 lost current control (likely due to a bad connection) and Cell 83 shorted. Cells 1 & 2 were used for preliminary testing and calendar aging experiments.

Basis Signal Number	Frequency	Discharge Current	Discharge Ratio
1	0.003 Hz	3*(average C-rate)	5
2	0.003 Hz	3*(average C-rate)	10
3	0.003 Hz	5*(average C-rate)	5
4	0.02 Hz	5*(average C-rate)	5
5	0.09 Hz	1*(average C-rate)	1

Supplementary Table 5: Basis set of signals used to design periodic profiles.

Periodic Profile	Basis Signal(s) Number(s)
A: Reference	1
B: Discharge Ratio Effect	2
C: Discharge Current Effect	3
D: Frequency Effect	4
E: Superposition	1 + 5

Supplementary Table 6: Description of periodic profiles used as aging cycles in realistic usage condition study using basis set signals from Supplementary Table 5.

Item	Hyperparameters	Explanation
1	booster='gbtree'	Specify that the XGBoost will use the gradient tree boosting method
2	max_depth=7	Maximum depth of a tree. Used to make the model more complex and more expressive.
3	eta=0.12	Learning rate. Tuning it can help the convergence speed, but may impact the performance.
4	reg_alpha=0.87	L1 regularization term on weights. Used to control overfitting.
5	reg_lambda=20	L2 regularization term on weights. Used to control overfitting.
6	subsample=0.86	The fraction of samples to be used to fit the individual base learners.
7	min_child_weight=4	The minimum sum of instance weight to further partition a leaf node. Also used to control overfitting.
8	colsample_bytree=0.87	The ratio of columns being subsampled for each tree. Used to control overfitting, by increasing the diversity of trees.
9	random_state=user_seed	Sets the seed used by XGBoost. A different random seed will be used each time.

Supplementary Table 7: Hyperparameters and their explanations for XGBoost.

Supplementary Methods

Supplementary Method 1 : Calculation of the driving range

Based on the total cumulative energy from the battery aging campaign, we use two representative driving testing profiles, UDDS and US06, to calculate the total driving range [4]. The US06 is an aggressive driving schedule with high acceleration and high speed while the UDDS is a milder schedule that represents city driving conditions. For our analysis, the total cumulative energy from a single cell reported by the Maccor cycler is scaled up to a 96S74P configuration (96 cells connected in series, and 74 cells connected in parallel) battery pack used in a typical EV:

$$E_{pack} = n \times m \times E_{cell} \quad (\text{Supplementary Equation 1})$$

Where E_{cell} is the total cumulative energy for a single cell. n and m are the number of cells connected in series and parallel, respectively.

To calculate the power demand for a typical EV over different drive cycles, the following vehicle longitudinal dynamics equation is used:

$$P_w(t) = \left(c_r m_v g \cos(\theta) + m_v g \sin(\theta) + \frac{1}{2} \rho_{air} c_d A_f v^2 + \delta m_v \frac{dv}{dt} \right) v(t) \quad (\text{Supplementary Equation 2})$$

$$\delta m_v = m_v + m_r$$

Where m_v is the vehicle mass, m_r is the equivalent mass of the rotating parts, c_r is the rolling resistance coefficient, g is gravity acceleration, θ is the slope of the road, A_f is the frontal area, c_d is the air drag coefficient, ρ_{air} is air density, $\delta > 1$ is the mass factor caused by the rotating inertia, and v is the vehicle speed.

The rotating mass includes various components [5]. Here, the wheels, driveshaft, and electric motor are considered:

$$m_r = m_{w,r} + m_{d,r} + m_{m,r} \quad (\text{Supplementary Equation 3})$$

where $m_{w,r}$, $m_{d,r}$ and $m_{m,r}$ are the rotating mass of the wheels, driveshaft, and motor, respectively. The rotating mass for each of the rotating components is calculated as [5]:

$$m_{i,r} = \frac{\Theta_i}{r_{i,r}} \quad (\text{Supplementary Equation 4})$$

where Θ_i is the moment of inertia, and $r_{i,r}$ is the equivalent radius.

The exact values of Θ_i and $r_{i,r}$ depend on the specific vehicle model. The values used here are for general cases. Considering a Tesla Model S with both a front motor and a rear motor, the total rotating mass is calculated as follows:

$$m_r = \frac{\Theta_w}{r_{w,r}^2} + \frac{\Theta_d}{r_{d,r}^2} + \frac{\Theta_m}{r_{m,r}^2}$$

$$= 4 \times \frac{1.53}{0.352^2} + 2 \times \frac{0.1}{0.1^2} + 2 \times \frac{0.05}{0.2^2} \quad (\text{Supplementary Equation 5})$$

Therefore, the mass factor δ is calculated as:

$$\delta = \frac{m_v + m_r}{m_v} = \frac{2087 + 72.5}{2087} = 1.035 \quad (\text{Supplementary Equation 6})$$

The parameters in Supplementary Equation 2 are listed in Supplementary Table 1.

Using Supplementary Equation 2, the power demand under UDDS and US06 driving profiles are shown in Supplementary Figure 3 and Supplementary Figure 4, respectively. The area in orange is the power demand at the wheel, and the area in blue is the power obtained from the regenerative braking system. Then the energy consumption for the testing profile is given by:

$$\begin{aligned} E_{profile} &= \int_{t_0}^{t_f} P_w(t) dt \\ &= E_{profile}^{dis} + E_{profile}^{ch} \end{aligned} \quad (\text{Supplementary Equation 7})$$

Where t_0 and t_f are the initial and final times of the driving testing profile. $E_{profile}$ represents the cumulative energy demand, and can be further divided into cumulative charge energy $E_{profile}^{ch}$ and cumulative discharge energy $E_{profile}^{dis}$.

Then, the total energy consumption for the EV $E_{vehicle}$ is calculated as follows:

$$E_{vehicle} = \left(E_{profile}^{ch} \times \eta + \frac{E_{profile}^{dis}}{\eta} \right) \times (1 + c_{aux}) \quad (\text{Supplementary Equation 8})$$

Where $\eta = 0.8$ is the lumped average powertrain efficiency (i.e., from the battery to the wheel), and $c_{aux}=0.12$ is the auxiliary coefficient accounting for auxiliary energy consumption (i.e., headlights, air conditioner, and multimedia).

For the driving testing profile, the total driving range is given by:

$$D = \int_{t_0}^{t_f} v(t) dt \quad (\text{Supplementary Equation 9})$$

Finally, the energy consumption per kilometer is calculated as:

$$E_{km} = \frac{E_{vehicle}}{D} \quad (\text{Supplementary Equation 10})$$

Using the vehicle parameters above, the energy consumption is 142.1166 Wh/km for UDDS, and is 203.567 Wh/km for US06. Then, using the scaled-up total cumulative energy E_{pack} from our battery aging campaign, and these energy consumptions, we can calculate the mileage at end of life for all profiles. The average constant current lifetime mileage along with the maximum mileage gain using a dynamic profile are reported in Supplementary Table 2.

Supplementary Method 2 : Data generation and collection

The experimental setup is identical to the one described in Moy et al. [6]. Batteries were cycled from inside the CSZ temperature chamber shown in Supplementary Figure 13 with channels from the Maccor battery cycler. Data was acquired on a “Maccor computer” not connected to the internet to avoid routine updates which would interrupt the experiments. The MIMS Server software automatically converted the data to ASCII txt files and transferred it to a “server computer”, connected via ethernet. A python script periodically uploaded the data to an S3 bucket hosted by Amazon Web Services (AWS). From the S3 bucket, the ASCII battery cycling data files could be accessed and analyzed remotely through an AWS Sagemaker Jupyter Notebook instance. All analysis code and data structuring functions were written and run in the Sagemaker instance.

Supplementary Method 3 : Aging cycles

The channels and cells that were used for the experiments, along with their respective aging cycles, are shown in Supplementary Table 4. Two cells failed during the experiments. The first (Cell 83) was one of the cells running synthetic protocol 2c at a C/5 average current. This cell appeared to short very early on, near the beginning of the cycling experiments. The second (Cell 18) was a cell running the constant

current protocol with no rest at a $C/2$ average current, likely due to a bad connection between the battery and the cycler channel. In this work, negative current indicates discharge and positive current indicates charge for the full cell.

Constant current profiles used in this work are illustrated in Supplementary Figure 14. As far as periodic profiles, the overall design structure is represented in Supplementary Figure 15 and Supplementary Figure 16 represents the resulting profiles. For all profiles, the discharge and charge pulses were trapezoidal waves with ramp rate (to discharge and charge segments) of 11 C-rate units/s. We held the charge current fixed at $0.08C$, and the discharge current was a chosen value. The discharge ratio was another chosen variable, defined as the time spent at the maximum discharge current divided by the time spent at the minimum discharge current. These two times were not independent, as they were constrained by the length of the period, or inverse frequency, which was another chosen variable. The rest duration was constrained by the desired average current, and was thus adjusted accordingly. A summary of the basis set of the signals used to create the periodic profiles is shown in Supplementary Table 5. These basis signals were then used as shown in Supplementary Table 6 to generate the periodic profiles for the aging cycle experiments. All periodic profiles were run at average currents of $C/10$, $C/5$, and $C/2$. For the periodic sequences the average current included the rest time.

We previously tested a larger number of repetitions (ex. 10) within a sequence, but most higher current profiles reached the lower cutoff voltage limit before finishing a single cycle. For example, the charge and discharge pulses shown in Supplementary Figure 15 were repeated 10 times before going into a rest. As a result, we chose to include solely one unique repetition per sequence (Supplementary Figure 16) and loop over the sequence until reaching the lower cut-off voltage. We chose a fixed charge current of $0.08C$ from calculating the average charge current from one of the real driving profiles.

For driving protocol generation, velocity profile data from EV driving is collected from two cities (referred to as City 1 and City 2). From the velocity of the vehicle, the current profile for one individual cell is generated. City 1 contains a region of highway-driving. The C-rate and voltage response is depicted in Supplementary Figure 17.

From the real driving duty cycles, synthetic protocols are generated using an algorithmic procedure described in [6]. The purpose of this algorithm is to create synthetic duty cycles that are representative of real driving. Supplementary Figure 18 shows the current and voltage profiles of the resulting protocols. Due to experimental limitations, not all synthetic duty cycles could be run at all average currents run for the constant current and periodic protocols. The originally planned experiments included all six synthetic duty cycles each at average C-rate $C/10$, $C/5$, and $C/2$. We changed the average currents of the synthetic profiles simply by multiplying all the current set points by a scaling factor to achieve the desired average current. However, instantaneous currents in highway-based synthetic duty cycles at $C/2$ exceeded hardware limits, so synthetic duty cycles 1a and 1b ($C/2$ average) were excluded and replaced by $C/16$ average. Synthetic duty cycle 3 ($C/2$ average) was also excluded for the same reason. The 1b $C/5$ average current profile did not complete a full repetition of the synthetic duty cycle before reaching the lower voltage cutoff (in Supplementary Figure 18, the long dynamic part of the profile 1b discharges the cell to 3.1 V before the rest can occur), so this protocol was also excluded from the experiment.

Supplementary Method 4 : Diagnostic cycles

Diagnostic cycles are run at Cycle 0 (before any aging cycles), then after 25, 50, 75, and 100 aging cycles to provide higher granularity in diagnostic metric trajectories at beginning of life (BoL) and then every 100 cycles.

Each diagnostic cycle starts with a reset cycle, following a CCCV charge protocol: CC at $C/3$ until 4.2V, then CV at 4.2V until 0.01C cutoff, followed by a CCCV discharge protocol: CC at $C/3$ until 2.8V, then CV at 2.8V until 0.01C cutoff.

To calculate resistance-based degradation metrics, we included a hybrid pulse power characterization sequence (HPPC), a standard sequence detailed in the Idaho National Laboratory Battery Test Manual for Electric Vehicles [7], which allows us to calculate resistances at various SOC and timescales. The HPPC sequence begins with a one hour rest period to allow the cell to equilibrate and establish a stable open circuit voltage (OCV). Then a 30 second discharge pulse (1C current), followed by a 40 second rest, and a 10 second charge pulse (0.75C current) are applied to the battery. This sequence was carried out at 100 %, 75 %, 50 %, 30 %, and 10 % SOC increments for the realistic usage study. The HPPC defines SOC using capacity, starting at the top of charge. For the HPPC in the realistic usage condition study, the batteries were charged with a C/3 constant current (CC) to 4.2V, then held at 4.2V constant voltage (CV) until a current cutoff of 0.01C was reached. This point was defined as 100 % SOC. To reach the next SOC, a C/3 constant discharge current was applied with a capacity cutoff of 25 % of full cell initial capacity, rather than a voltage cutoff (as was used to define fully charged). Similarly, the remaining SOC were defined using capacity cutoffs referenced to the full cell initial capacity. Supplementary Figure 21 shows the full HPPC sequence within a diagnostic cycle (Supplementary Figure 21a) and a single HPPC sequence at 100 % SOC (Supplementary Figure 21b). The HPPC standardized sequence has some drawbacks, such as asymmetric charge and discharge current pulse magnitudes and durations. To remedy these drawbacks, we further designed the voltage-based pulse (VBP) routine described in the following section.

The voltage-based pulses also probe resistance and were carried out at three voltages: 4.2V, 3.75V, and 3.1V, which correspond approximately with voltage-defined SOC: 100 %, 50 % and 5 %. This voltage-based SOC definition remains constant throughout all the diagnostic cycles, despite battery aging. Within the VBP sequence, the battery was first charged to 4.2V with a C/3 constant current, then held at 4.2V until a 0.001C current cutoff was reached. For the VBPs, establishing a stable, consistent OCV was essential, so lower current cutoff and extended rest periods were used. The 4.2V and 3.75V VBP sequences consisted of a 2 hour rest following the charge/discharge end condition. Following that, a 150 second 0.1C discharge pulse was applied, followed by a 1 hour rest (to extract relaxation data). A CV hold was then applied at the desired voltage for 30 minutes to re-equilibrate the cell to the correct voltage, followed by a 2 hour rest. Then a 150 second 0.1C charge pulse was applied, followed by a 1 hour rest (for relaxation data [8]). The cell was discharged from 4.2V to 3.75V with a C/3 - 3.7V - 0.001C CCCV procedure. At low voltages, it was very difficult to reach a stable OCV due to high battery resistances and OCV drift. We applied a two-step discharging procedure for the 3.1V VBP sequence. The first step was again a C/3 - 3.1V - 0.001C CCCV step, followed by a 2 hour rest to allow for cell equilibration. The second discharge step was a 3.1V CV hold for two hours, to attempt to establish a stable cell voltage. Despite extensive process, we still observed OCV drift at the lower voltage, as can be seen from 30 - 40 hours in Supplementary Figure 20a. After the two-step procedure, the same VBP sequence as described previously was followed at 3.1V. Supplementary Figure 20 shows the full VBP sequence within a diagnostic cycle (Supplementary Figure 20a) and a single VBP sequence at 3.75V (Supplementary Figure 20b).

Resistances at various timescales were calculated from the VBP sequences (at specific voltages) and HPPC sequences (at specific SOC, defined by full cell nominal capacity) within the diagnostic cycle. A current pulse with known magnitude I_0 is applied to the battery. At a desired time after the pulse begins, the voltage response compared to when the battery was at rest, ΔV , is calculated, where $\Delta V = V_{x \text{ seconds}} - \text{OCV}$. We use an ohmic relationship to calculate resistance by dividing the voltage response of the battery at a specific time (relative to the rest OCV) by the magnitude of the current pulse, as illustrated in Supplementary Figure 22. The timescale at which the overpotential is calculated can correspond to processes that occur in the battery at different timescales [2]. For example, ohmic resistance prevails at short timescales, charge transfer resistance is observed at intermediate timescales and polarization resistance dominates at longer timescales.

Finally, two rate performance tests (RPT) are performed as part of the diagnostic cycle to extract charge and discharge capacity metrics. The first RPT is performed at C/40 (charge CCCV to 4.2 V,

cutoff C/400, discharge CCCV to 2.8 V, cutoff C/400) while the second RPT is performed at C/2 (charge CCCV to 4.2 V, C/100, discharge CCCV to 2.8 V, cutoff C/100).

Supplementary Method 5 : Degradation trajectory curving fitting

Based on the aging trajectories of all 92 cells, we built a curve-fitting model (including a $X^{0.5}$ dependence to a linear function) for SOH prediction from early cycles:

$$SOH = a + b \cdot EFC^{0.5} + c \cdot EFC \quad (\text{Supplementary Equation 11})$$

We note that adding the $X^{0.5}$ dependence to obtain a good fit is in agreement with results in Figure 4a (main manuscript), where we show, using an independent analysis, that early degradation is dominated by the loss of lithium inventory Q_{Li} , synonymous of side reactions (such as solid electrolyte interphase growth).

As shown in Supplementary Figure 23, the data from the first 400 EFCs was used to fit the model parameters a, b, and c. The model was then extrapolated to predict SOH.

The distribution of fitting parameters and the SOH prediction errors (defined at 85 % SOH) for all 92 cells are shown in Supplementary Figure 24. The results demonstrate that the curve-fitting model has good prediction accuracy.

The following function form (including a square-root dependence) is used to fit the degradation curves of Q_{Li} .

$$Q_{Li, \text{fitted}} = 1 + a \cdot EFC^{0.5} + b \cdot EFC \quad (\text{Supplementary Equation 12})$$

where a and b are the fitted parameters.

The optimization problem is formulated as follows:

$$\arg \min_{\lambda_{\min} \leq \lambda_{iden} \leq \lambda_{\max}} J = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(\frac{abs(Q_{Li, \text{exp}}(k) - Q_{Li, \text{fitted}}(k))}{Q_{Li, \text{exp}}(k)} \right)} \quad (\text{Supplementary Equation 13})$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the vectors containing the lower and upper bounds for the identified parameters vector $\lambda_{iden} = [a, b]$, respectively. $Q_{Li, \text{exp}}$ and $Q_{Li, \text{fitted}}$ are the lithium inventory obtained from experimental data and curve fitting, respectively. N is the total number of $Q_{Li, \text{exp}}$ before reaching 90 % SOH.

In this study, the Matlab PSO toolbox is used, where the population size was set to 200, and the self-adjustment weight and social adjustment weight were set as 2 and 2, respectively. The mean absolute percentage error (MAPE) shown in Figure 4h is defined as:

$$MAPE = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(\frac{abs(Q_{Li, \text{exp}}(k) - Q_{Li, \text{fitted}}(k))}{Q_{Li, \text{exp}}(k)} \right)} \quad (\text{Supplementary Equation 14})$$

Supplementary Method 6 : Half-cell differential voltage model

The half-cell differential voltage fitting method uses the potential profiles of the positive and negative electrodes, individually acquired experimentally (at a C/40 rate), to simulate a full cell voltage profile that matches the experimental one. This method relies on a few key assumptions: 1) that currents are low enough that resistance is not significant, 2) that the shape of the individual electrode potential profiles does not change during aging, and 3) that the experimental half cells made from electrodes harvested from the commercial full cell are representative of full cell behavior. The optimization procedure described below enables the extraction of the electrode-specific capacities Q_{PE} and Q_{NE} , lithium inventory Q_{Li} , as shown in Supplementary Figure 25 along with the SOC (relative to the individual electrode's full range) and potentials of both electrodes at the charged and discharged states of the full cell.

First, the individual electrode open-circuit potential (OCP) profiles were collected by disassembling a full cell and assembling half-cells (Supplementary Figure 26). Images from the process are shown in Supplementary Figure 27. The entire disassembly process was carried out in an argon-filled glovebox. The full cell was first carefully cut and peeled open using a dremel and pliers to remove the outer casing. The encased jelly roll was then unrolled and measured to calculate the total area of the full cell. A custom-made lollipop-shaped punch was used to punch out half-cell electrodes from the harvested full cell electrodes. These electrodes were double-sided, leaving an active area of approximately 3.5 cm^2 after the tabs were cleaned of active material (scraping graphite with a plastic razor or using NMP to wipe off positive electrode active material). As the electrodes were double-sided, we used two lithium foil counter electrodes and two separators (Celgard), one set for each side of the electrode sheet (Supplementary Figure 26) when assembling the half cells. The half cells were made using pouches that were filled with 500 μL of LP40 electrolyte (1M LiPF₆ in EC:DEC 1:1 by wt) and vacuum sealed. Once the cells were fully sealed, they were removed from the glovebox, formed and cycled on a Biologic BCS-815 battery cycler at 35 °C (the same temperature as the full-cell cycling experiments). Pressure plates were used to apply a slight pressure (finger tight) to the pouch cells to ensure good contact between internal components. We collected the open-circuit potentials for both electrodes at C/40 to match the rate of the full cell RPT from which the mechanistic degradation modes are extracted.

The half-cell differential voltage fitting algorithm consists in scaling and shifting the anode and cathode OCP curves to reconstruct the measured full-cell voltage. The governing equations in this process are described below.

$$\overrightarrow{Q_{PE,scaled}} = \frac{\overrightarrow{Q_{PE,raw}}}{\max(\overrightarrow{Q_{PE,raw}})} \times Q_{PE} + \text{offset}_{PE} \quad (\text{Supplementary Equation 15})$$

$$\overrightarrow{Q_{NE,scaled}} = \frac{\overrightarrow{Q_{NE,raw}}}{\max(\overrightarrow{Q_{NE,raw}})} \times Q_{NE} + \text{offset}_{NE} \quad (\text{Supplementary Equation 16})$$

$$\overrightarrow{V_{FC,sim}(Q_{FC})} = \overrightarrow{V_{PE}(Q_{PE,scaled})} - \overrightarrow{V_{NE}(Q_{NE,scaled})} \quad (\text{Supplementary Equation 17})$$

where $\overrightarrow{Q_{PE,scaled}}$, $\overrightarrow{Q_{PE,raw}}$, $\overrightarrow{Q_{NE,scaled}}$, and $\overrightarrow{Q_{NE,raw}}$ are numpy arrays of capacity values that map to the voltage profiles. The raw capacities are the capacity values collected from the half-cell data, while the scaled variables are the capacities of the positive and negative electrodes scaled in reference to the full-cell capacity.

The optimization problem is then:

$$\min_{Q_{PE}, \text{offset}_{PE}, Q_{NE}, \text{offset}_{NE}} \left(\frac{dV}{dQ} \right)_{err}^3 \times V_{err} \quad (\text{Supplementary Equation 18})$$

$$V_{err} = \text{mean}(\text{abs}(\overrightarrow{V_{FC,sim}(Q_{FC})} - \overrightarrow{V_{FC,exp}(Q_{FC})})) \quad (\text{Supplementary Equation 19})$$

$$\left(\frac{dV}{dQ} \right)_{err} = \text{mean} \left(\text{abs} \left(\frac{d\overrightarrow{V_{FC,sim}(Q_{FC})}}{d\overrightarrow{Q_{FC}}} - \frac{d\overrightarrow{V_{FC,exp}(Q_{FC})}}{d\overrightarrow{Q_{FC}}} \right) \right) \quad (\text{Supplementary Equation 20})$$

A higher weight (cubic power) was placed on the differential voltage error to ensure that the physical processes (such as phase transformation and solid-solution electrode behavior) were being accurately captured by the simulated voltage profiles. Only $\frac{dV}{dQ}$ values under a threshold are considered to calculate the $\left(\frac{dV}{dQ} \right)_{err}$ error to prevent overfitting the large slope regions at SOC extrema ($\frac{dV}{dQ} < 1.75$).

Examples of this fitting applied to a single cell (periodic profile D - C/2) in the pristine state (cycle 2) , early states (cycle 29-110) and in an aged state (cycle 620) are shown in Supplementary Figure 28. The negative electrode $\frac{dV}{dQ}$ is flipped for visualization purposes. The $V - Q$ and $\frac{dV}{dQ}$ curves are used for the error optimization while the $\frac{dQ}{dV}$ curve is used for validation.

We note that the fitting results do not capture the low SOC region as accurately as the rest of the curves, especially in the first few cycles. This could be explained by the silicon behaving differently than graphite, in particular considering that the half-cell system is different than the commercial full-cell system. However, after initial cycling, as soon as cycle 29 (Supplementary Figure 28), we observe a better matching of the low SOC region, in particular in the large slope region.

Additionally, the negative electrode $\frac{dV}{dQ}$ graphite features are lost fairly rapidly, in particular, the three sequential peaks in the 0.2 - 0.4 SOC region. This known phenomenon is commonly attributed to heterogeneities in the electrode [9]. This may be exacerbated by a difference in degradation rate between the silicon oxide and the graphite as evidenced by Kirkaldy *et al.* [10]. This leads to challenges in ensuring reliable fits for the negative electrode, as it becomes more difficult to accurately match these features between the simulated and experimental data for aged cells. On the other hand, we see that the two main positive electrode features (broad $\frac{dV}{dQ}$ peaks around 0.5 and 0.8 SOC) appear to be relatively well captured by the simulation throughout aging. Thus, the estimated Q_{PE} may be more reliable than the estimated Q_{NE} , although further experimental validation would be necessary to confirm this. The fitting procedure failed throughout aging for one cell (cell 45), which has been excluded from the analysis.

Supplementary Method 7 : Machine Learning pipeline

To extract the aging cycle protocol parameters, we used the current profile data from Aging Cycle 26, the first high-resolution aging cycle after the second diagnostic cycle (for experimental troubleshooting reasons, aging Cycle 26 is the first cycle to provide 1s resolution for all cells). The time profile was generated to be in evenly spaced intervals and the current profile was consequently interpolated. The capacity profile was subsequently calculated by coulomb counting.

The calculation of most of the parameters in Supplementary Table 3 was fairly straightforward with standard python and/or pandas functions. A few parameters required custom functions and analysis, such as the ones related to the number of charge/discharge/rest occurrences, frequency, or rest SOC.

For the frequency metrics, we calculated the discrete sine transform of the interpolated protocol current profile values using the Scipy Fast Fourier Transform library[11] (`scipy.fft`, using function: `dst`). The squared amplitudes (to ensure positive values) of these frequencies were sorted in descending order, and the frequencies associated with the first three frequencies were identified. These frequencies were then sorted in ascending order, to ensure that the lowest frequencies were being compared to the lowest and the highest frequencies being compared to the highest between different protocols. The zeroth frequency was almost always 0 across all cells, which was expected since the discharge protocols have a negative bias. For the protocols in which the zeroth frequency was not zero, a non-number filler value was added at the beginning so that the zeroth identified frequency then became the first identified frequency. As our goal was to investigate the effect of the dynamic behavior in discharge protocols, we only used the (nonzero) first and second peak frequencies in our analysis. Peak frequency 1 was the lower peak frequency and the peak frequency 2 as the higher peak frequency. A distribution of these frequencies is shown in Supplementary Figure 9.

We next investigated the (Spearman) correlation (ρ) between the parameters that we used in the ML model development, shown in Supplementary Figure 29. It can be found some features are highly correlated, which means they influence the prediction output simultaneously. To facilitate the SHAP analysis used in this study, we remove these correlated features in a pre-processing step. The threshold of $\rho > 0.8$ is used, and 4 features are deleted according to this threshold as shown in Supplementary Table 3. Notably, we do not see strong correlations with any other metrics for the last three parameters: Rest SOC, Lower Key Frequency, and Higher Key Frequency. This means that any importance that is ascribed to those parameters can be attributed solely to those parameters. For the more highly correlated features, it is important to remember that the feature importance assigned to one parameter might be partially caused by importances of correlated input parameters. We use these extracted parameters to

further develop our input-output relationship mapping using the data-driven degradation diagnostics framework.

For our XGBoost model, we used a booster of “gbtree”, importance type of “gain”, and subsample ratio of 0.2 (to mitigate overfitting) for all output metrics. EFCs at EoL had more spread in the output values, so we additionally implemented regularization hyperparameters (reg_alpha and reg_lambda). We split the dataset into a training set and validation set (90-10 split) by shuffling the protocols randomly, and then assigning the cells associated with each protocol to the respective set. Thus, cells with identical protocols were constrained to either the validation set or the training set, but couldn’t be split across both. We performed a “pseudo” cross-validation by using repeated shuffles of the training set and validation set to select the model hyperparameters. The final parameters used are described in Supplementary Table 7 and the final test split contains 10 cells.

We used the “shap” python package “TreeExplainer” function [12] to interpret the relative importances of the input features on predicting the target output. The SHAP values of both the training and test sets were combined when finding the average SHAP value for a given input feature and predicted output. Figure 3 in the main text provides a visualization of these SHAP values, where the color intensity represents the SHAP value of the relative importance between features and outputs.

Importantly, the goal of this ML pipeline is not to develop a model to make the most accurate predictions (prognosis) but rather to build a model that fits our data well enough such that SHAP analysis can determine the impact of the various discharge features (diagnosis). Achieving this would indicate that the feature importances identified by the model, or the impact of each feature on improving model performance, were truly meaningful. Thus, we chose to leverage the entirety of our data to tune the hyperparameters of our model through the various validation iterations. Finally, the train-test split results dependence was also investigated. Supplementary Figure 33 shows that Figure 3 results are consistent across different train-test splits. Test cells are chosen uniformly across the C-rates and the protocols. Duplicate cells are always placed together in either the train or the test set to prevent overfitting. 8 test cells were used for train-test split #2 while train-test splits #1 contains 28 test cells. Parity plots of the training and test set used for the main results are shown in Supplementary Figure 30 and Supplementary Figure 31, and overfitting was not observed. The python code used for the machine learning pipeline will be made publicly available on Github at the time of publication. <https://github.com/geslina>.

To improve the consistency of the SHAP results, the SHAP analysis is run 25 times, each time with a different random seed. All the 25 results were then averaged, and the mean values were used. Supplementary Figure 32 compares the SHAP results from using a single random seed to the averaged results of 25 random seeds across 4 different runs. It can be seen that by using the average value of 25 seeds, the consistency of the SHAP results is improved.

Supplementary Notes

The SOC during rest periods is critical in determining the degradation trajectory (Supplementary Figure 11). Specifically two highway-based protocols (1a and 1b) were designed to study the path-dependence in dynamic conditions [1]. Both protocols consist in the same discharge segment, with rests at different SOC (Supplementary Figure 18). Supplementary Figure 12 shows that Protocol 1a has rest periods at higher SOC, resulting in faster degradation, illustrating a clear path-dependence. This path-dependence is a practical example where vehicle-to-grid integration would minimize the time at rest at elevated SOC, thus maximizing lifetime [13].

Supplementary Figure 10 shows the evolution of Q_{pe} and Q_{Li} with respect to the rest fraction at elevated SOC either above 60 % (left) (same feature used in Figure 3), above 70 % (middle) or above 80 % (right) for cells with rest fractions at elevated SOC. Ignoring the two most aggressively-cycled cells (C/2, 6h storage at 100 % SOC - the two points further on the right, close to 0.8 on the x-axis), we observe a decent correlation between the rest fraction at elevated SOC and Q_{Li} (Supplementary Figure 10b), although not as strong as with Q_{pe} (Supplementary Figure 10a). While loss of lithium inventory due to side reaction is likely to play a role, positive electrode instability dominates, particularly for the two most aggressively cycled cells (C/2, 100 % storage). This data suggests that while electrolyte oxidation at the positive electrode and side reactions contribute to lithium loss, cathode structural instabilities (such as disordering, or oxygen evolution) is more sensitive to elevated SOC. However, electrode-level experimental characterization would be needed to validate these insights.

Supplementary References

- [1] Gering, K. L. *et al.* Investigation of path dependence in commercial lithium-ion cells chosen for plug-in hybrid vehicle duty cycle protocols. *Journal of Power Sources* **196**, 3395–3403 (2011).
- [2] Barai, A., Uddin, K., Widanage, W. D., McGordon, A. & Jennings, P. A study of the influence of measurement timescale on internal resistance characterisation methodologies for lithium-ion cells. *Scientific Reports* **8**, 21 (2018).
- [3] Geise, N. R. *Quantifying Degradation in Lithium-Metal and Lithium-Ion Batteries*. Ph.D. thesis, Stanford University, California (2021).
- [4] Dynamometer drive schedules (2022). URL <https://www.epa.gov/vehicle-and-fuel-emissions-testing/dynamometer-drive-schedules>. Accessed February 22, 2023.
- [5] Guzzella, L., Sciarretta, A. *et al.* *Vehicle propulsion systems*, vol. 1 (Springer, 2007).
- [6] Moy, K., Ganapathi, D., Geslin, A., Chueh, W. & Onori, S. Synthetic duty cycles from real-world autonomous electric vehicle driving. *Cell Reports Physical Science* **4**, 101536 (2023).
- [7] Christopherson, J. P. *Battery Test Manual For Electric Vehicles* (U.S. Department of Energy, 2015). URL <https://inldigitallibrary.inl.gov/sites/sti/sti/6492291.pdf>.
- [8] Kang, S. D. & Chueh, W. C. Galvanostatic intermittent titration technique reinvented: Part i. a critical review. *Journal of The Electrochemical Society* **168**, 120504 (2021).
- [9] Sieg, J. *et al.* Local degradation and differential voltage analysis of aged lithium-ion pouch cells. *Journal of Energy Storage* **30**, 101582 (2020).
- [10] Kirkaldy, N., Samieian, M. A., Offer, G. J., Marinescu, M. & Patel, Y. Lithium-ion battery degradation: Measuring rapid loss of active silicon in silicon–graphite composite electrodes. *ACS Applied*

Energy Materials **5**, 13367–13376 (2022).

- [11] Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
- [12] Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* **2**, 2522–5839 (2020).
- [13] Gong, J. *et al.* Quantifying the impact of V2X operation on electric vehicle battery degradation: An experimental evaluation. *eTransportation* **20**, 100316 (2024).