Exploratory Analysis of Strain-Derived Features for Machine-Learning Estimation of State of Health in 18650 Lithium-Ion Cells

Malichi Flemming II[a], George Anthony[a], Ryan L. Limbaugh[a], Austin R. J. Downey[a,b]

[a]Department of Mechanical Engineering, University of South Carolina

[b]Department of Civil & Environmental Engineering, University of South Carolina

UNIVERSITY OF
South Carolina

Molinaroli College of
Engineering and Computing

Outline

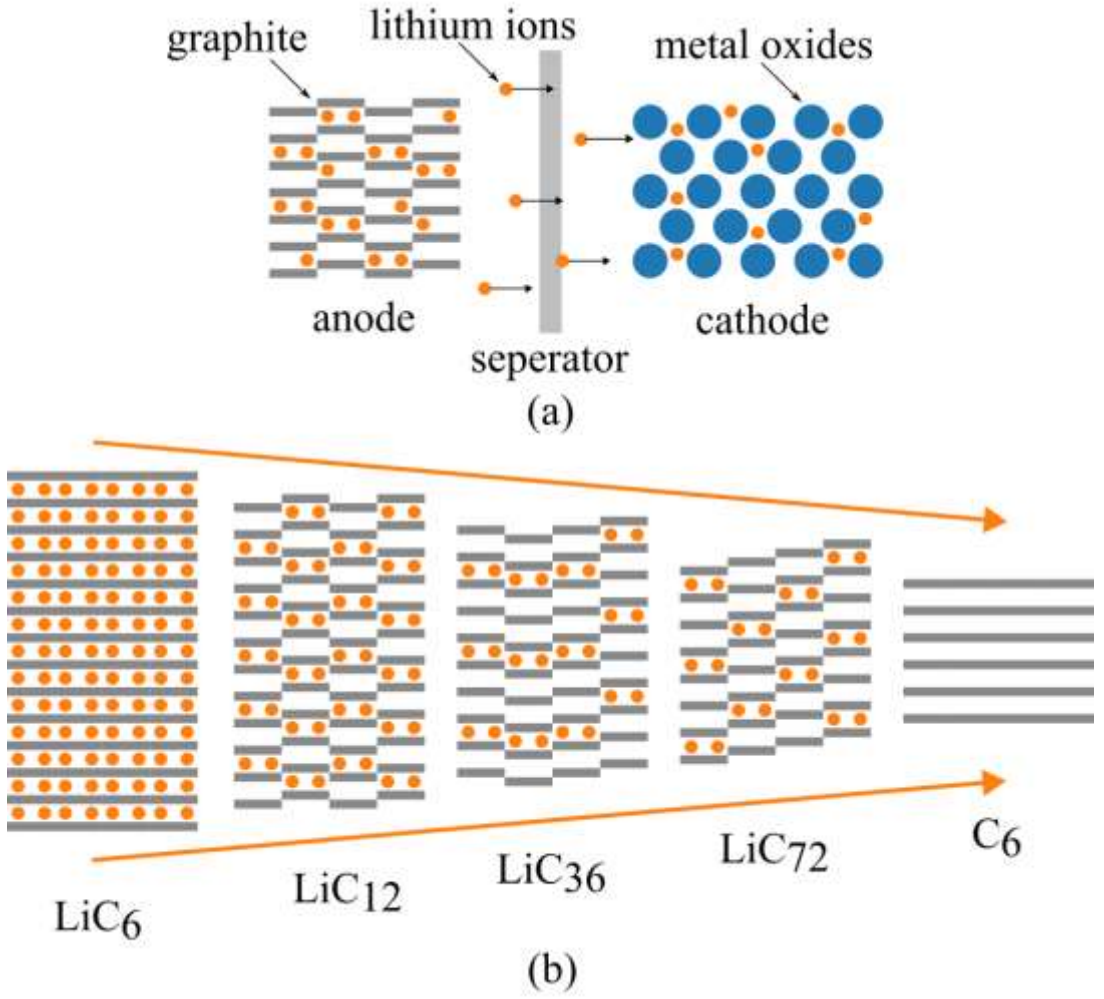
- Investigative motivation
- Strain-based cycling setup and strain detrending
- Feature extraction and ranking
- Feature dimension reduction and analysis

Investigative Motivation

- Rise in demand of electric transportation, renewable energy, robotics, and battery-powered consumer electronics.
- Increasing battery efficiency makes it a more attractive option for storing and supplying power.
- Advanced BMS utilize battery state estimation models to increase battery efficiency, prolong battery lifespan, and improve battery safety.
- Recent research demonstrates that strain sensing improves battery state prediction accuracy when coupled to traditional electrical and thermal measurements [1].

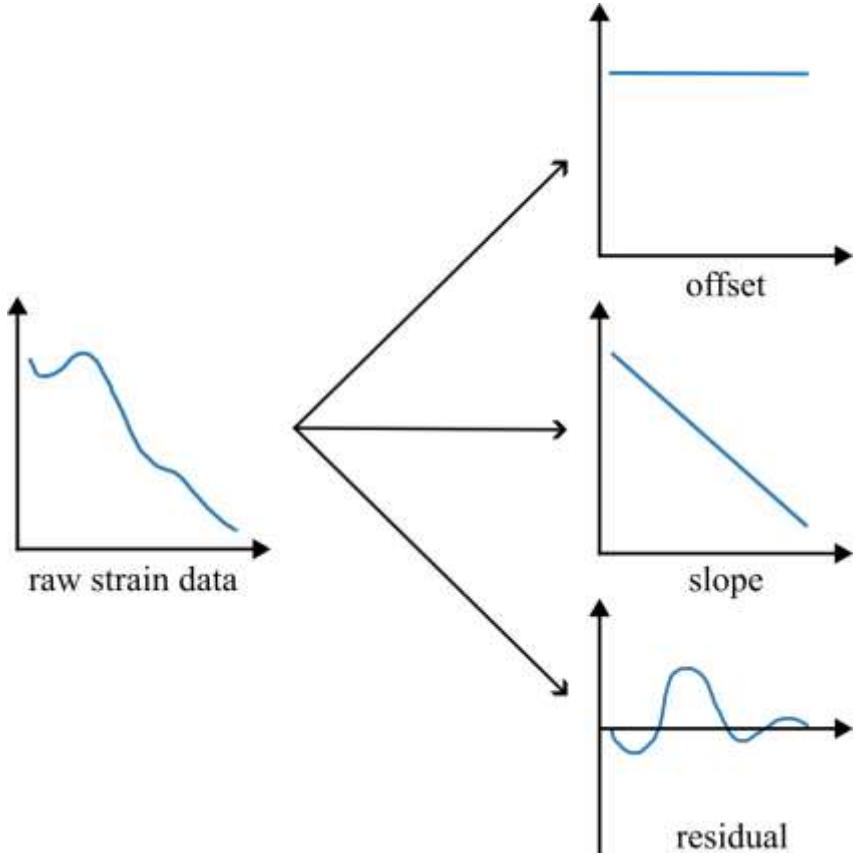
Why Strain Sensing

- Allows us to monitor how the battery deforms physically as it ages.
- Battery deformation is directly correlated with battery degradation.
- Battery degradation is primarily caused by internal electrochemical reactions.
  - SEI growth
  - Lithium plating
- Avoiding or minimizing these processes would extend battery lifespan.



Strain Decomposition Framework

- Strain signals for each cycle were decomposed into three components: offset, slope, and residual.
- The offset, defined as the initial strain at the start of discharge, was subtracted from the full strain profile.
- The slope, representing the linear trend between the initial and final strain values, was also removed.
- The remaining signal, the residual strain, served as the basis for detailed mechanical analysis.

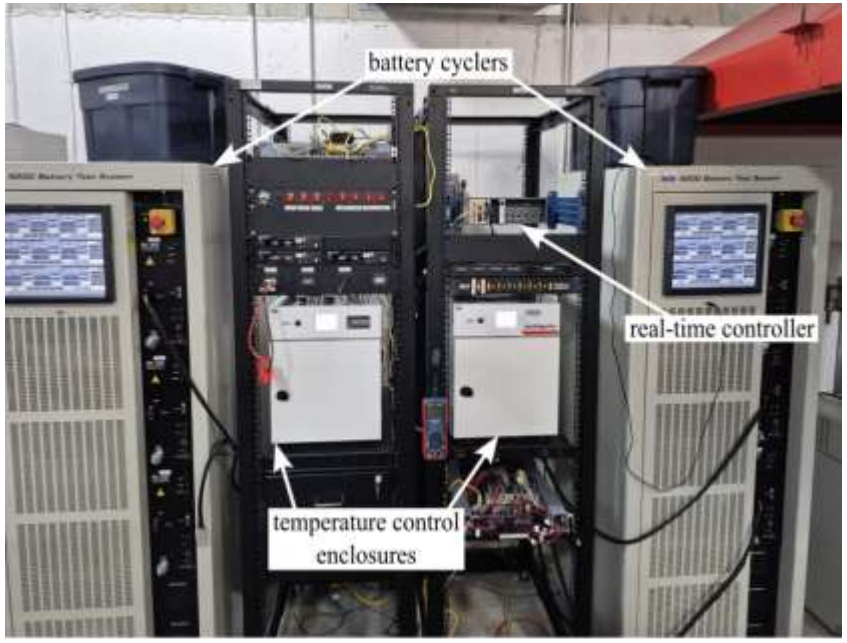


Strain-based cycling setup and strain detrending

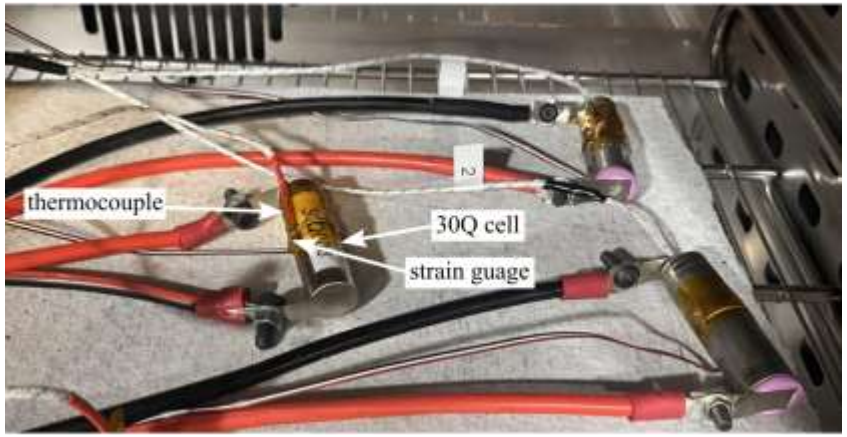
Molinaroli College of Engineering and Computing
UNIVERSITY OF SOUTH CAROLINA

Cycling And Strain Monitoring Procedure

- Three Samsung 30Q lithium-ion cells (3000 mAh) were cycled using 1C CC-CV charging followed by 1C CC discharge in a temperature-controlled environment.
- Each cycle included a one-hour rest period between charge and discharge phases.
- During charge and discharge, hoop strain was recorded with a surface-mounted strain gauge, along with current, voltage, and temperature data.
- Cycling continued for 450 cycles, reaching roughly 80 % of the initial capacity.
- Capacity per cycle was computed via coulomb counting, and SOH was expressed as the percentage of rated capacity.



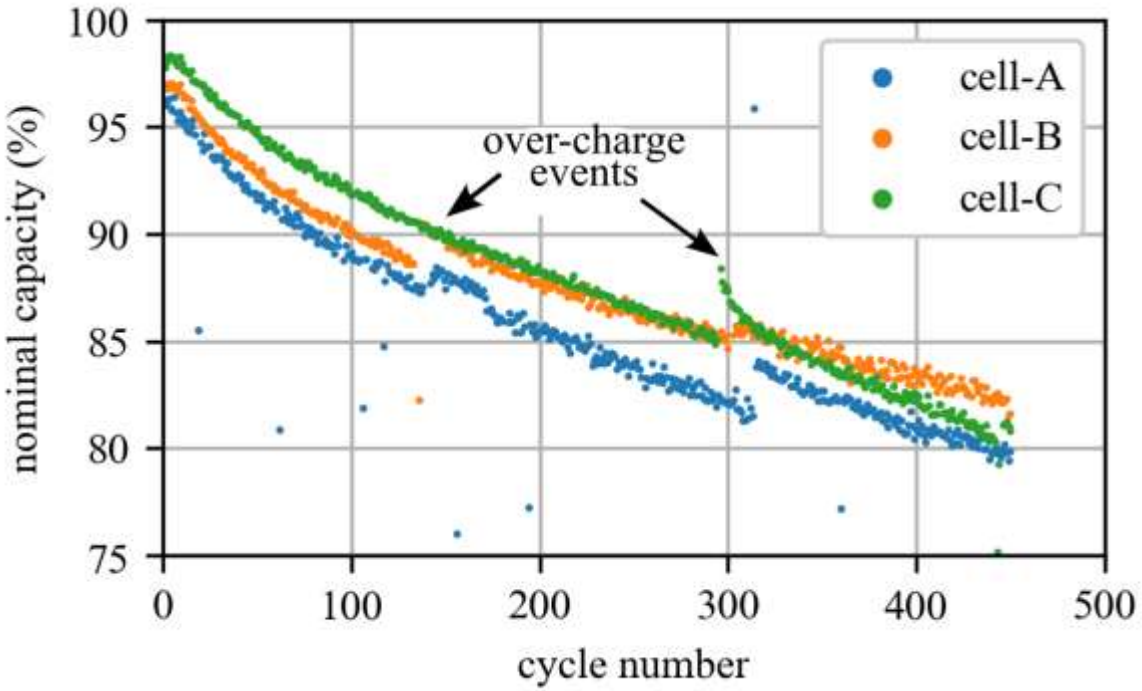
(a)



(b)

Capacity Degradation of 30Q Cells with Overcharge Events

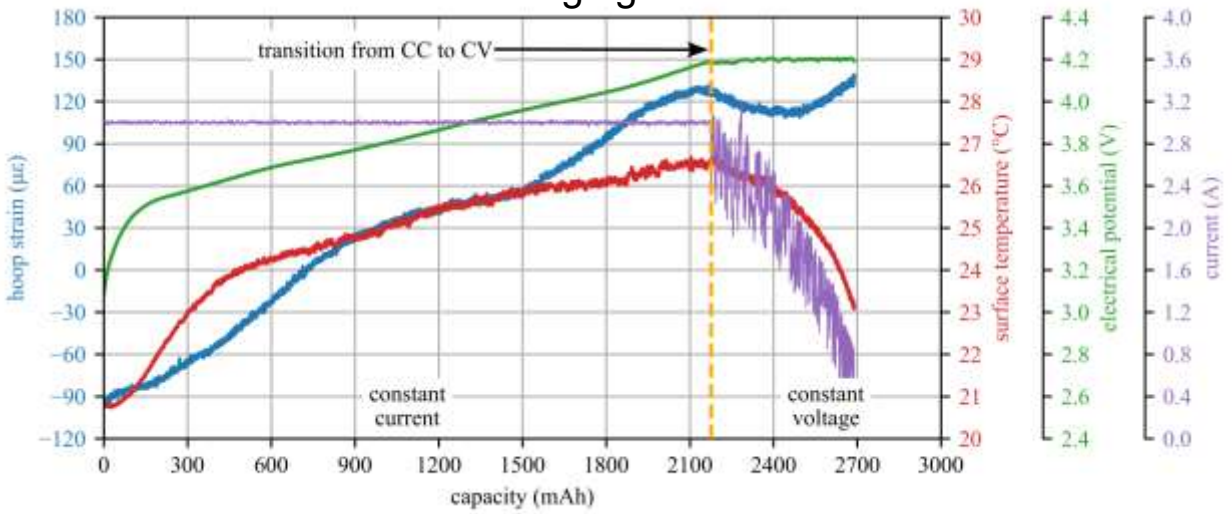
- The capacity trajectories of three Samsung 30Q cells over ~450 cycles show the expected gradual decline toward end-of-life near 80% SOH.
- Two marked overcharge events act as strong deviations, producing abrupt step-changes in the degradation curve.
- These overcharge incidents accelerate capacity fade and create clear alterations from the nominal aging trend.
- The events provide a natural benchmark for comparing how electrical and mechanical signals respond to abnormal cycling.
- The pronounced shifts following overcharge can illustrate if off-nominal conditions can significantly influence subsequent strain behavior.



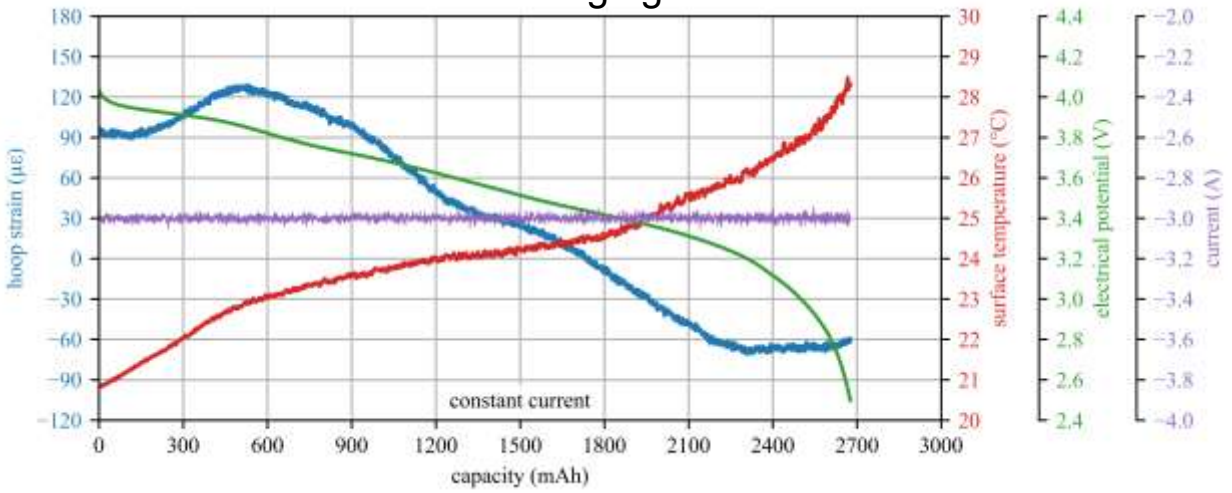
Cycle 100 Strain And Thermal Behavior

- During the CC portion of charging, both strain and temperature rise steadily, reflecting lithium intercalation into the graphite anode.
- In the CV region, strain and temperature initially decrease as current drops, but as it approaches full charge strain rises again while temperature falls rapidly, indicating that strain is governed by intercalation staging rather than thermal effects.
- During discharge, strain increases early alongside a mild temperature rise, consistent with the onset of deintercalation.
- As discharge progresses, strain returns toward its baseline while temperature continues to elevate from resistive heating, showing that strain primarily tracks reversible structural changes rather than thermal conditions.

charging

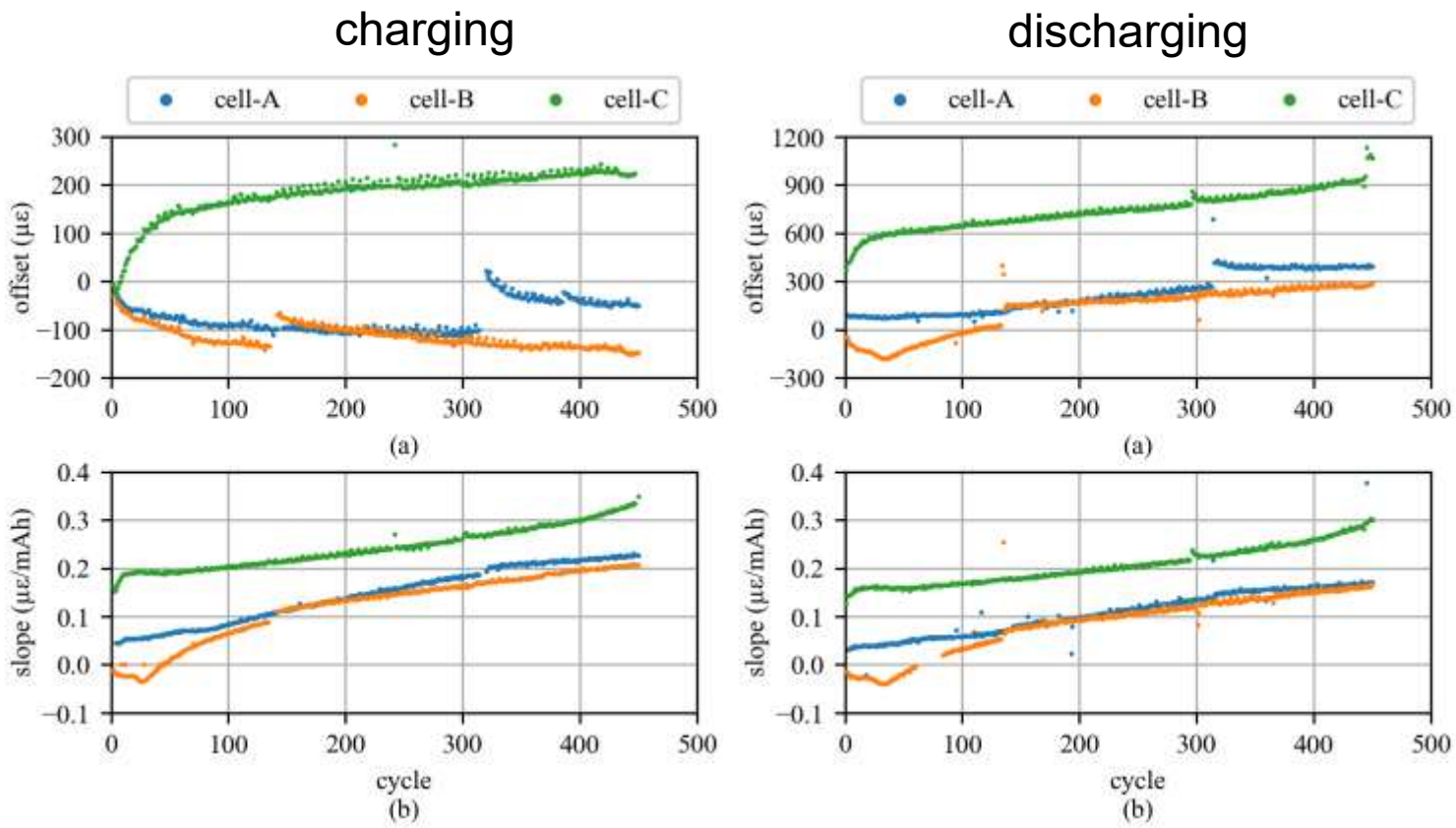


discharging



Strain-derived Offset And Slope Metrics

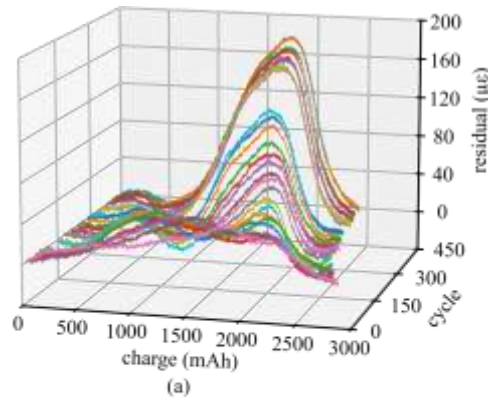
- Offset and slope features were extracted for charge and discharge to capture baseline drift and intra-cycle trends.
- Offsets represent quasi-static residual-strain drift while slopes capture dynamic, rate-dependent behavior.
- Early cycles show clustered offsets and near-zero slopes, indicating stable operation.
- Aging causes offsets to diverge and slopes to grow, with charge–discharge asymmetry revealing distinct inflection points.
- Combined offset–slope evolution provides low-cost indicators of long-term drift, emerging kinetics, and degradation onset.

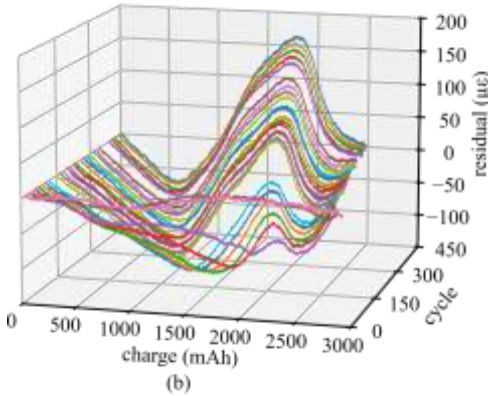


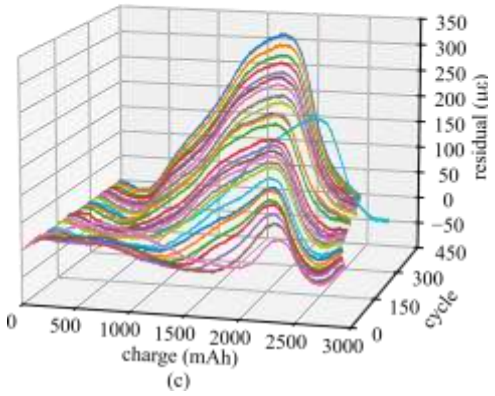
Residual Strain Evolution

- Residual strain signals for charge and discharge are shown on matched axes to enable direct comparison after cycle-wise detrending.
- Early cycles cluster near zero, indicating nearly linear expansion–contraction behavior in fresh cells.
- With cycling, residuals broaden and develop structure, reflecting growing non-stationarity and irreversible mechanical contributions.
- Variance increases smoothly in both segments, with outliers corresponding to previously identified overcharge events.

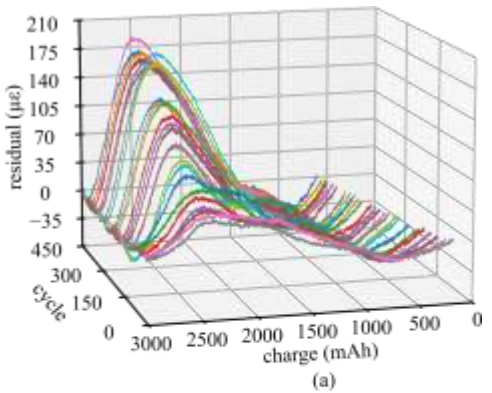
charging

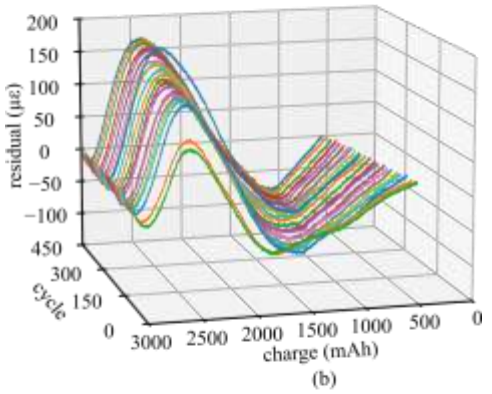


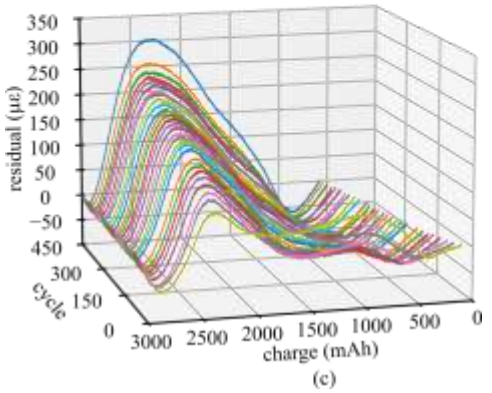




discharging







Feature extraction and ranking

Molinaroli College of Engineering and Computing
UNIVERSITY OF SOUTH CAROLINA

Distribution Shape (Non-gaussian)

Captures asymmetry and tail heaviness in the residual strain, making it sensitive to early regime changes and departures from Gaussian behavior that often precede shifts in averages or variability.

• Skewness is defined as

$$g_1 = \frac{m_3}{m_2^{3/2}}, (1)$$

where $(m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3)$is the third central moment, $(m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)$ is the second central moment (variance), n is the sample size, and $\bar{x}$ is the sample mean.

• Kurtosis is computed as

$$g_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}, (2)$$

Where s is the sample standard deviation. The first term is a scaled and normalized fourth moment and the second term subtracts the bias correction and 3 to yield excess kurtosis.

Time-domain Amplitude Statistics

Summarize level and range over a cycle, providing a stable, low-cost readout of load severity and slow baseline drift in the strain response.

- Maximum strain ($\varepsilon_{\max}$) is the highest strain value experienced, typically the tensile peak during a loading cycle.
- Minimum strain ($\varepsilon_{\min}$) is the lowest strain value experienced, typically the compressive peak.
- Mean strain ($\varepsilon_{\mathrm{mean}}$) Is the average strain value, representing the midpoint of the strain cycle and is expressed as

$$\varepsilon_{\mathrm{mean}} = \frac{\varepsilon \mathrm{max} + \varepsilon \mathrm{min}}{2}. \quad (3)$$

Time-domain Impulsiveness And Shape Ratios

Emphasize spike-like events and waveform forms via peak-to-average relationships, highlighting short, high-energy transients that basic amplitude summaries may miss.

- Shape factor is the ratio of overall energy to average magnitude, capturing waveform form relative to its typical level. Computed as

$$\text{shape factor} = \frac{RMS(x)}{\frac{1}{n}\sum_{i=1}^{n}|x_i|}. \quad (4)$$

- Impulse factor is the ratio of the highest excursion to the average magnitude, emphasizing short, high-amplitude events. Computed as

$$\text{impulse factor} = \frac{\max_i |x_i|}{\frac{1}{n}\sum_{i=1}^{n}|x_i|}. \quad (5)$$

- Crest Factor is widely used in vibration analysis and fault detection, and it indicates the presence of sharp, high-energy spikes. It is defined as

$$\text{crest factor} = \frac{\text{peak value}}{\text{RMS}}. \quad (6)$$

Spectral Localization

Characterizes where strain energy concentrates in frequency, revealing resonance shifts and energy reallocation associated with protocol changes and evolving mechanical state.

- Root Mean Square Frequency (RMSF) provides a measure of the distribution of signal energy across frequency components. It is a weighted average frequency, useful in vibration and audio analysis. Calculated as

$$\mathrm{RMSF} = \sqrt{\frac{\sum_k f_k^2 P_k}{\sum_k P_k}}, \quad (7)$$

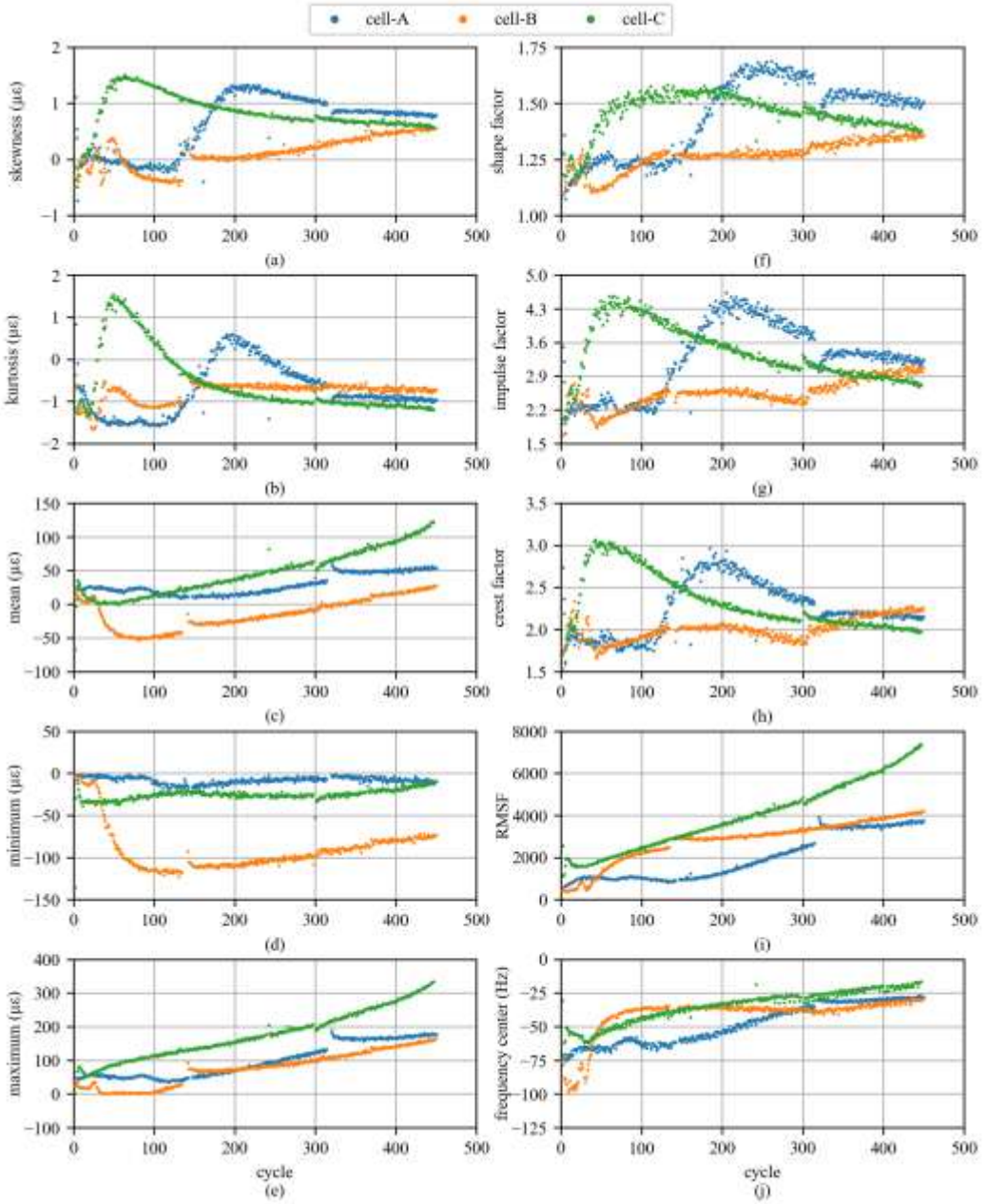
  where f_k is the frequency associated with the k^{th} FFT bin (in Hz).Represents the discrete frequency values over which the spectrum is defined and P_k is the spectral power at frequency f_k.

- Center Frequency represents the central or resonant frequency around which a signal or filter is most active. It is defined as the midpoint between the lower cutoff frequency $(f_1)$and upper cutoff frequency (f_2). Computed as

$$f_{\text{center}} = \frac{f_1 + f_2}{2}. \quad (8)$$

Charge Features

- All strain-based features for charge and discharge are plotted on common axes, enabling direct cross-cycle comparison.
- Distribution-shape features rise sharply early in life, then transition to slower drift, indicating increasing non-Gaussian behavior. Cell-A evolves more gradually.
- Time-domain amplitude statistics drift steadily across lifespan, reflecting slow baseline shifts with minimal cycle-to-cycle noise.
- Waveform-ratio features spike early and show localized excursions around disturbances, capturing impulsive, high-energy events that amplitude metrics mute.
- Spectral localization features stabilize after ~50 cycles; RMSF detects overcharge-related excursions, while center frequency remains steady, separating transient events from long-term resonance trends.



Spearman Consistency Metric Across Cells

- Per-feature Spearman summaries (Fisher average $\bar{\rho}$ and median $|\rho|$) quantify redundancy and cross-cell consistency for both charge and discharge.
- Maximum strain, RMSF, mean strain, and center frequency show high cross-cell correlations with low variability, indicating strong generalizability.
- Crest factor, impulse factor, shape factor, and skewness exhibit low correlations and large spreads, revealing strong sensitivity to cell-specific effects and poor transferability.
- Maximum strain and RMSF emerge as the most reliable cross-cell features, showing the highest average correlations and lowest median absolute correlations across both operating segments.

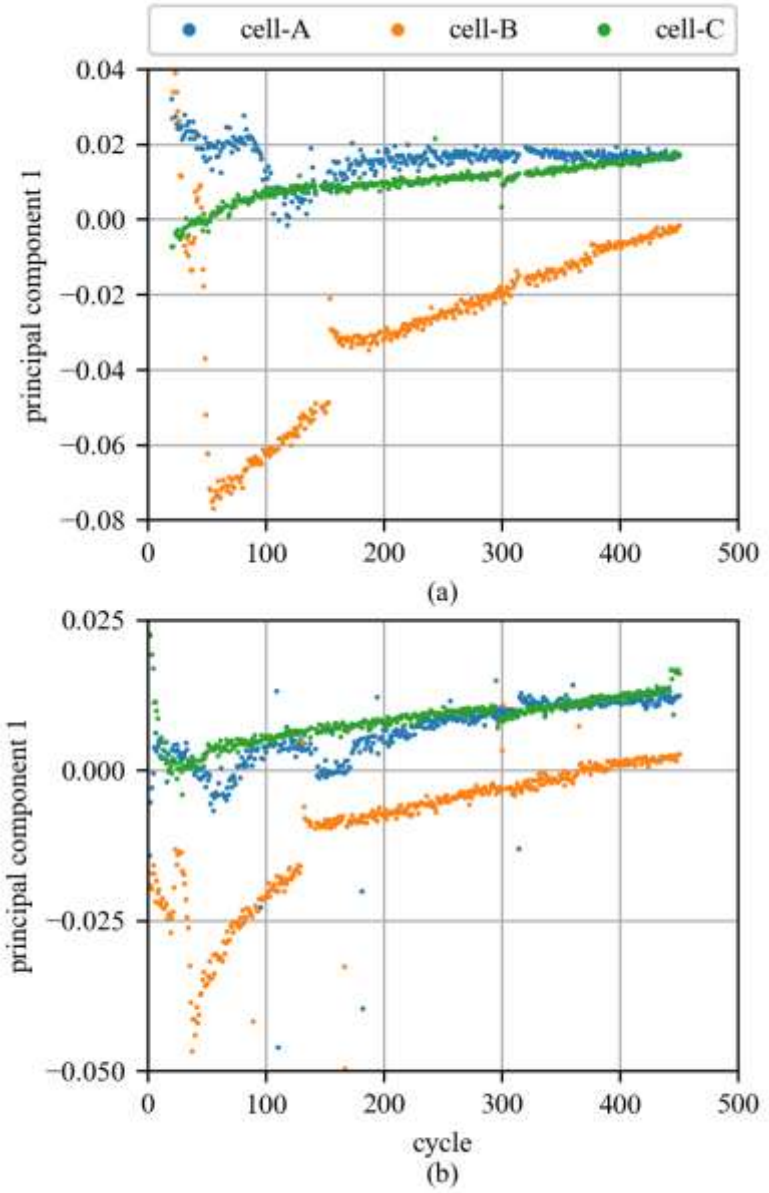
State	Charging		Discharging	
Feature	$\bar{\rho}$**(z-avg)**	**median**$\lvert\rho\rvert$	$\bar{\rho}$**(z-avg)**	**median**$\lvert\rho\rvert$
Mean Strain	0.8406	0.1185	0.9154	0.0688
Minimum Strain	0.3483	0.5669	0.5693	0.4071
Maximum Strain	0.9673	0.0271	0.9585	0.0361
Crest Factor	0.1822	0.6410	0.1365	0.6642
Impulse Factor	0.1432	0.6931	0.1143	0.6693
Shape Factor	0.3147	0.5756	0.2069	0.6366
Skewness	0.1214	0.7138	0.4645	0.5179
Kurtosis	0.3577	0.5941	0.6384	0.2990
RMSF	0.9646	0.0356	0.9566	0.0361
Center Frequency	0.8141	0.1896	0.9012	0.0730

Feature dimension reduction and analysis

Molinaroli College of Engineering and Computing
UNIVERSITY OF SOUTH CAROLINA

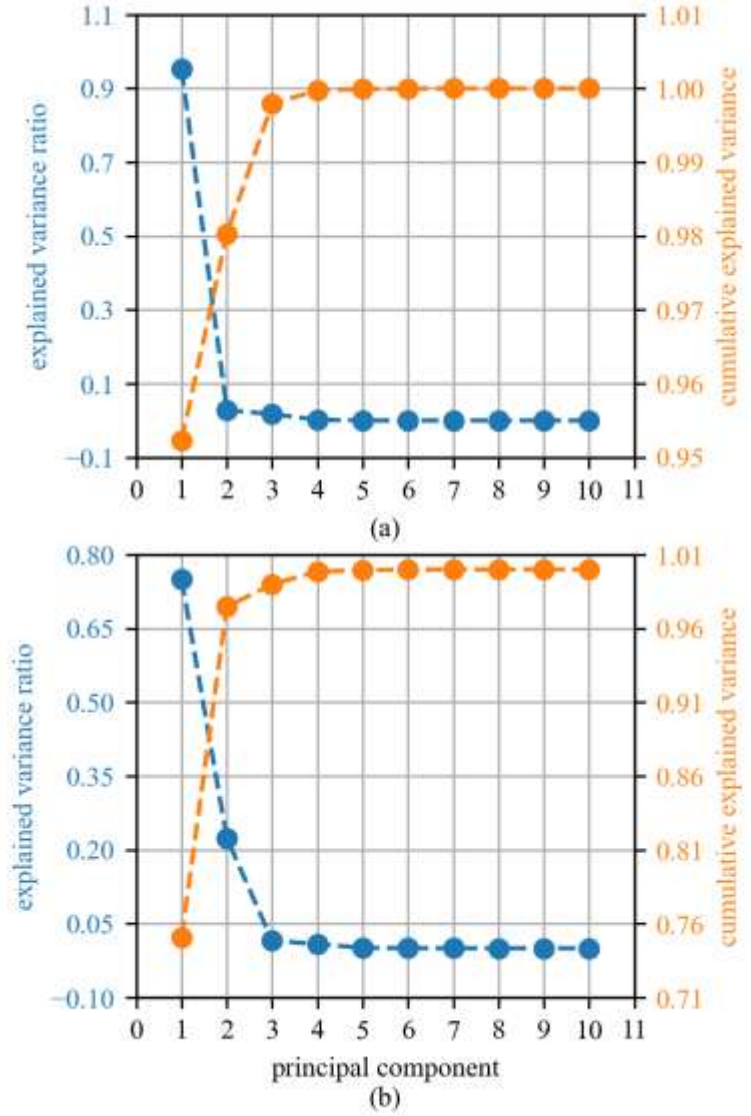
PC1 Cycle-Wise Analysis

- PC1 is plotted versus cycle index for both charging and discharging because it carries the largest share of variance.
- PC1 evolves smoothly with cycling across all cells, acting as a low-dimensional aging coordinate.
- The drift direction of PC1 stabilizes after the initial ~50-cycle break-in period.
- Cells A and C show nearly identical PC1 trajectories in shape.
- Cell B follows the same trend but with a scaled amplitude, appearing as a stretched version of A and C.
- Overcharge events near cycles ~150 and ~315 appear as clear excursions in PC1, especially for Cells A and B.
- Cells A and C follow similar discharge trends but with different scaling.



Scree And Cumulative Variance Plot Analysis

- Scree and cumulative explained-variance results confirm PC1 dominates the information content.
- For charging, PC1 alone explains ~95% of the variance and PC1+PC2 ~99% of the variance.
- For discharging, PC1 explains ~75% and PC1+PC2 ~98% of the variance



Conclusion

- Maximum strain and RMSF show the strongest cross-cell consistency, with Spearman correlations above 95% and median absolute correlations below 4%. **Preferred when interpretability is required.**
- These two features likely dominate the leading principal components, consistent with PCA results showing ~98% variance captured by the first two components. Reduces 10 features down to 2 components while losing minimum information. **Preferred when computational efficiency and model robustness is critical.**
- Post-overcharge cycles exhibit increased residual strain and altered slopes, indicating irreversible structural damage aligned with degraded electrical performance.
- The findings establish a foundation for incorporating mechanical strain diagnostics into battery-management-system frameworks.

Discussion

This material is partially supported by the the National Science Foundation (NSF) grant number CPS- 2237696. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Molinaroli College of Engineering and Computing
UNIVERSITY OF SOUTH CAROLINA

Thank You for Your Time

Name: Malichi Flemming II
Title: Ph.D. Student
Email: MALICHI@email.sc.edu
Lab GitHub: github.com/arts-laboratory

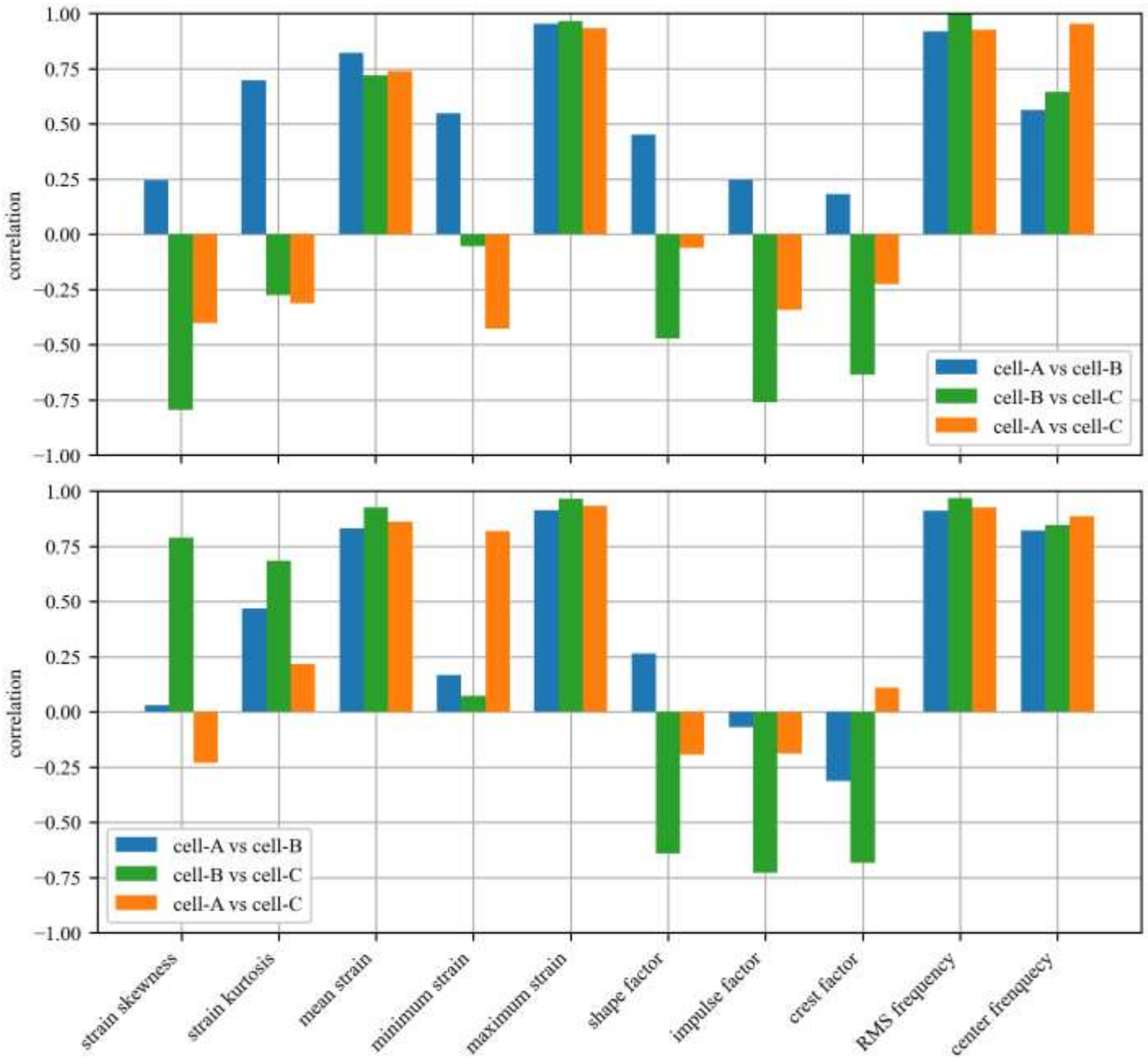
Molinaroli College of Engineering and Computing
UNIVERSITY OF SOUTH CAROLINA

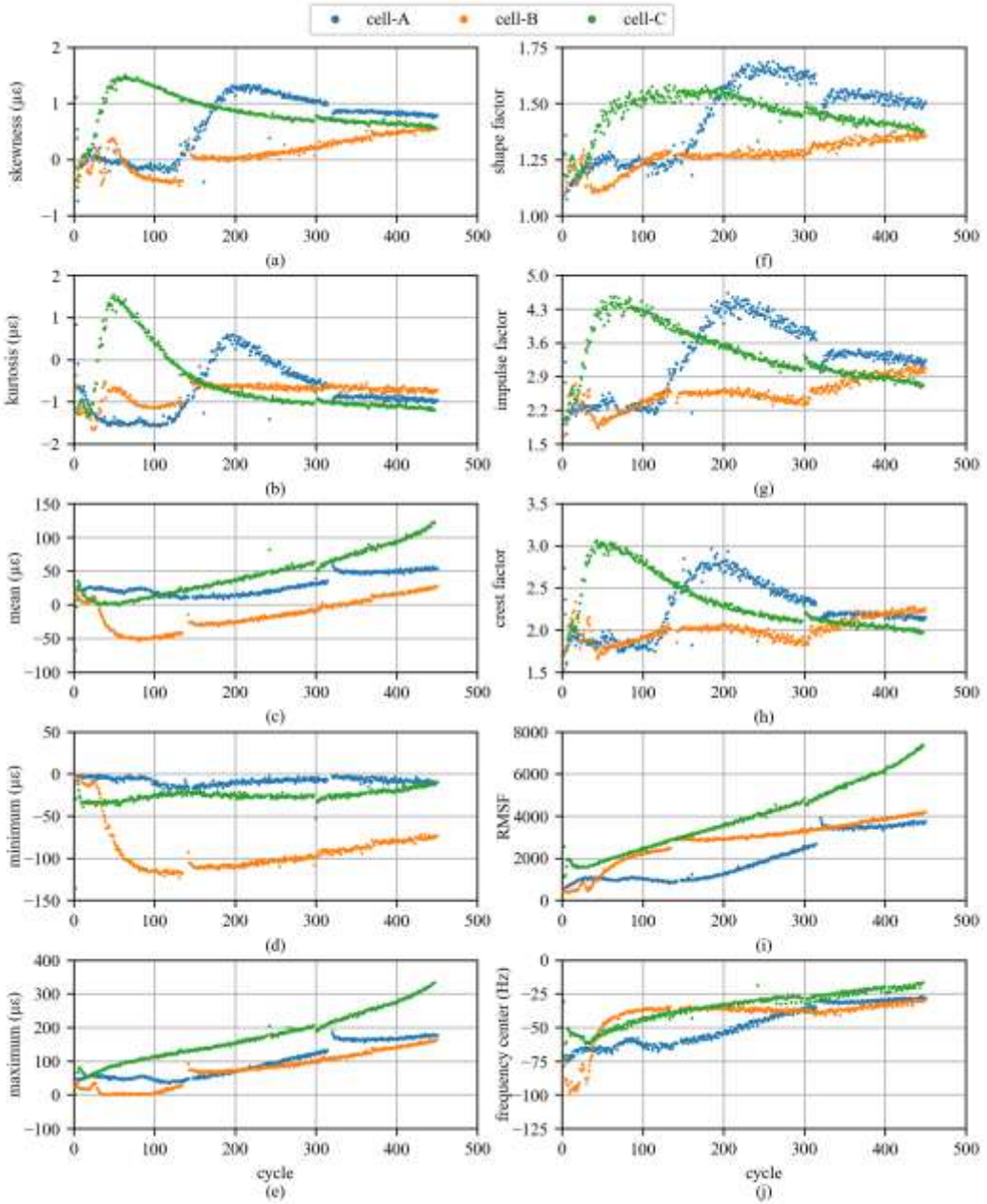
References

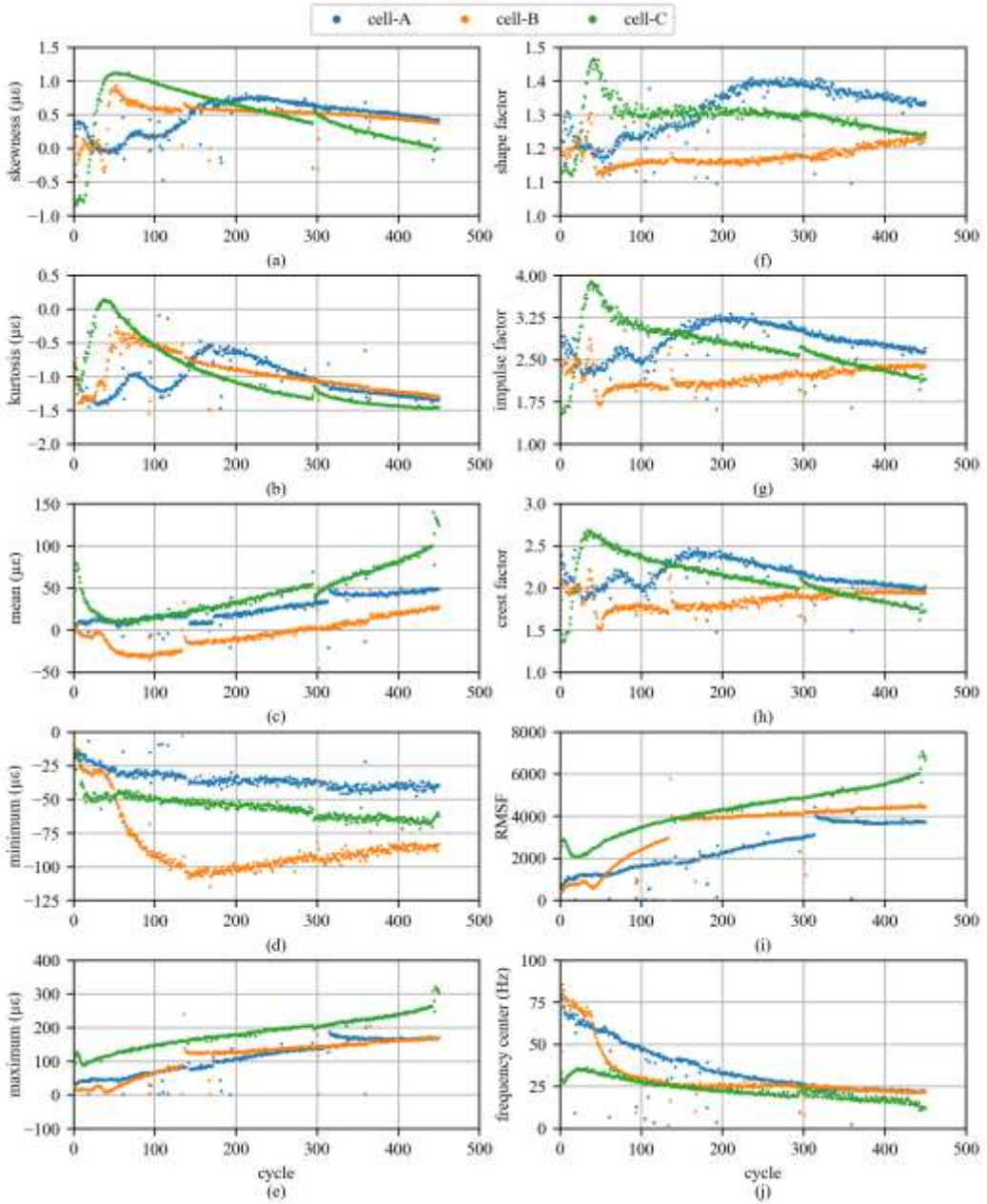
1. Yihuan Li, Kang Li, Xuan Liu, Xiang Li, Li Zhang, Bruno Rente, Tong Sun, and Kenneth T.V. Grattan. A hybrid machine learning framework for joint soc and soh estimation of lithium-ion batteries assisted with fiber sensor measurements. Applied Energy, 325:119787, November 2022.

charging

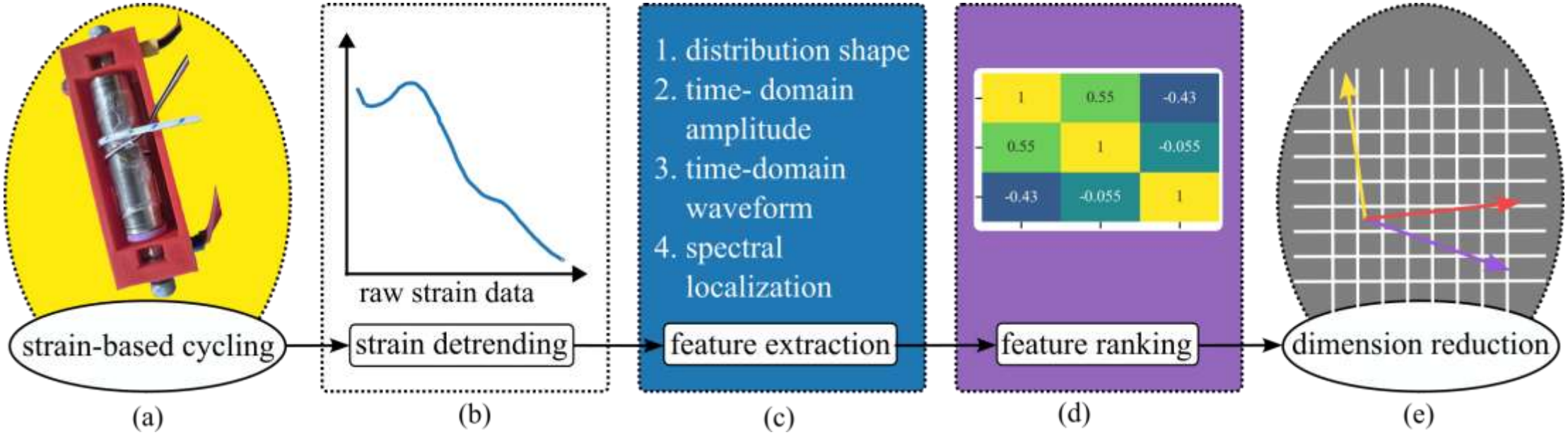
discharging




cell-A
cell-B
cell-C
skewness (με)
shape factor
kurtosis (με)
impulse factor
mean (με)
crest factor
minimum (με)
RMSF
maximum (με)
frequency center (Hz)
cycle
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)

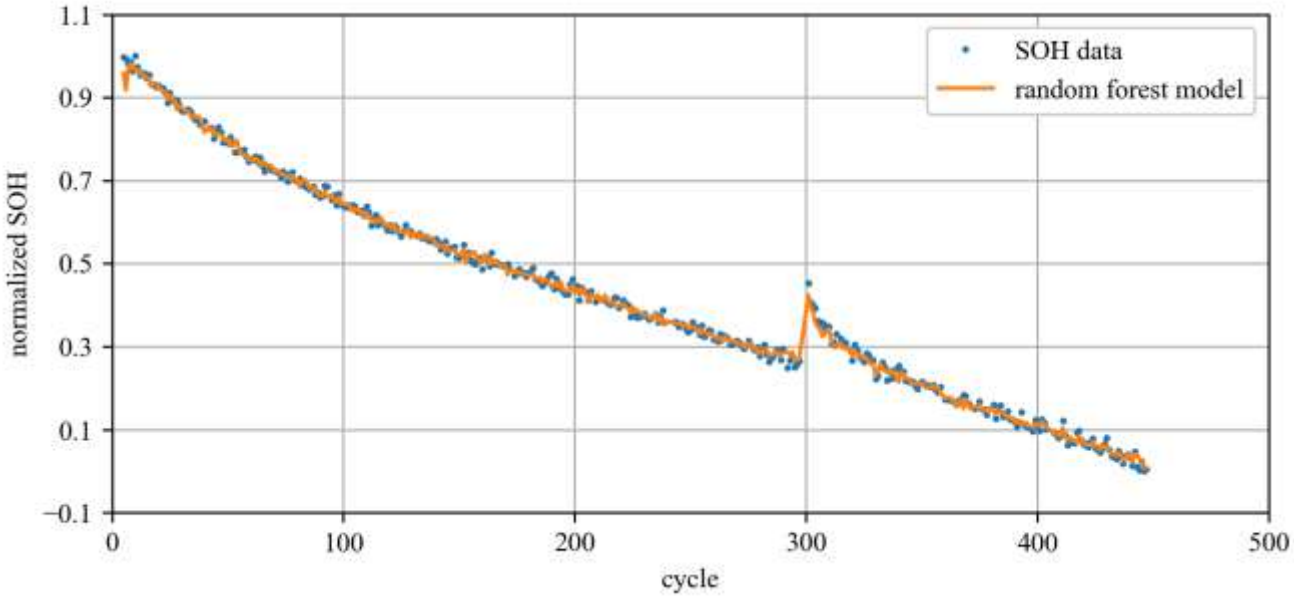

cell-A
cell-B
cell-C
skewness (με)
shape factor
kurtosis (με)
impulse factor
mean (με)
crest factor
minimum (με)
RMSF
maximum (με)
frequency center (Hz)
cycle
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)

Methodology



Random-forest Model For Strain-based SOH Prediction

- The Random Forest model achieved strong predictive performance, with an R2 of 0.9949 explaining over 99% of SOH variance.
- Error metrics were very low (MAE = 0.0151, MSE = 0.00035, RMSE = 0.0187), indicating consistently small prediction deviations.
- The model effectively captured nonlinear relationships among strain, temperature, and cycle-progression features, closely matching experimental SOH trends.
- Figure shows measured SOH (blue markers) aligning closely with the model's predicted trajectory (orange line).



PC1-PC2 Plane Analysis

- PCA trajectories with aging-direction arrows illustrate how features evolve in reduced space.
- Discharge trajectories trace a clear aging path in the PC1–PC2 plane.
- Charging trajectories therefore show a strong one-dimensional progression along PC1
- Cell B shows a clean, well-separated progression in the 2D discharge space.
- Overall, a 1D indicator suffices for charging, while a 2D indicator is more appropriate for discharging SOH evolution