Journal of The Electrochemical Society

Predicting Capacity Fading Behaviors of Lithium Ion Batteries: An Electrochemical Protocol-Integrated Digital-Twin Solution

To cite this article: Hang Li *et al* 2022 *J. Electrochem. Soc.* **169** 100504

Journal of The Electrochemical Society, 2022 **169** 100504
1945-7111/2022/169(10)/100504/9/$40.00 © 2022 The Electrochemical Society ("ECS"). Published on behalf of ECS by IOP Publishing Limited

Predicting Capacity Fading Behaviors of Lithium Ion Batteries: An Electrochemical Protocol-Integrated Digital-Twin Solution

Hang Li,[1] Jianxing Huang,[1] Weijie Ji,[1] Zheng He,[2] Jun Cheng,[1] Peng Zhang,[2] and Jinbao Zhao[1,z]

[1] State Key Lab of Physical Chemistry of Solid Surfaces, Collaborative Innovation Centre of Chemistry for Energy Materials, State-Province Joint Engineering Laboratory of Power Source Technology for New Energy Vehicle, Engineering Research Center of Electrochemical Technology, Ministry of Education, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, 361005, People's Republic of China
[2] College of Energy & School of Energy Research, Xiamen University, Xiamen 361102, People's Republic of China

The capacity degradation and occurrence of safety hazards of lithium ion batteries are closely associated with various adverse side electrochemical reactions. Nevertheless, these side reactions are non-linearly intertwined with each other and evolve dynamically with increasing cycles, imposing a major barrier for fast prediction of capacity decay of lithium ion batteries. By treating the battery as a black box, the machine-learning-oriented approach can achieve prediction with promising accuracy. Herein, a numerical-simulation—based machine learning model is developed for predicting battery capacity before failure. Based on the deterioration mechanism of the battery, numerical model was applied to test data from only 25 batterie to extend 144 groups data, resulting in the digital-twin datasets, which can reliably predict the maximum total accumulative capacity of the lithium ion batteries, with an error less than 2%. The workflow with iterative training dramatically accelerates the capacity prediction process and saves 99% of the experimental cost.

Manuscript submitted August 11, 2022; revised manuscript received September 7, 2022. Published October 6, 2022. *This paper is part of the JES Focus Issue on Heterogeneous Functional Materials for Energy Conversion and Storage II.*

Supplementary material for this article is available online

Commercial lithium ion batteries have been widely applied in mobile devices, which will contribute to carbon neutrality.[1–5] Along with the development of high performance materials of lithium ion battery, the battery manage system (BMS) is also of great importance which governs the steady operation of the batteries under various working conditions.[6–9] One of the greatest challenges of BMS is associated with the prediction of the evolution of the battery capacity.[10–13]

The capacity of a battery is one of the key performance parameters for battery system,[14,15] which plays an important role in regulating the health and safe operation of the battery.[16,17] In a rechargeable lithium ion battery, the energy conversion and storage process is fundamentally built on a reversible electrochemical reaction taking place at the interface between electrode and electrolyte as well as the bulk electrode materials. Therefore, at atomic-level, the lithium ion diffusion kinetics both across the solid electrolyte interphase and in the bulk phase of the active materials on both cathode and anode, and electron conduction and charge transfer efficiency across the various solid-solid interfaces within the three-dimensional electrode network play a profound role in determining the battery performance. Theoretically, the kinetics of these electrochemical processes depend primarily on the inherent quality of the battery which is an assembly of the electrode materials, the electrolyte, the separator and the current collector, which eventually governs the performance of a battery. Nevertheless, in reality, the operando parameters under which a battery works is also of great significance in determining its actual energy conversion efficiency.[18] For instance, among various parameters, the working temperature, according to the Arrhenius equation, plays an important role in determining the electrochemical reaction kinetics[19]; other factors such as the charge or discharge current density affects the polarization, which also imposes important impacts on the capacity.

The capacity degradation of a lithium ion battery occurs inevitably. Overall, it is a consequence of the irreversibility of the electrochemical reactions. In detail, it originates from different side electrochemical reactions, which slows down the interfacial lithium intercalation and deintercalation kinetics and decreases the energy conversion efficiency; it is often associated with the loss of active cathode or anode material and the deterioration of ionic kinetics. For instance, the irreversible structural changes caused by high degree of delithiation and the bulky solid-liquid interface from excessive decomposition of electrolyte at high cut-off voltage in cathode leads to the severe capacity loss[20–22]; the deposition of "dead lithium" or lithium dendrite occurred in the anode not only consume active lithium which directly correlates to the capacity and coulombic efficiency, but also causes safety hazards.

In view of the complicacy of a battery system, they will be used more safely when the capacity fade mode of a battery is available. Nevertheless, the above-mentioned side reactions occur both in the interface and in the bulk of the material are intricately interlaced with each other, which pose a grand barrier for accurate capacity prediction.[20,23,24] For instance, the increasing of thickness of solid electrolyte interphase (SEI), generated at the surface of anode, will affect the ion transport and increase the electrochemical polarization.[25,26] Meanwhile, severer polarization will increase the chance for occurrence of lithium dendrite,[27] which leads to more SEI and causes safety hazards.[28–30] Additionally, both the electrolyte and the active material suffer from varying degrees of degradation during the cycling life.[31–37] In this scenario, thanks to the development of high performance computers and breakthroughs in algorithms, the numerical simulations based on pseudo-two-dimensions (P2D) model, which takes the side reactions into consideration, have been applied for battery capacity prediction.[27,30,38–40] However, it remains a challenge to construct a realistic simulation model that could simultaneously consider the multiple parameters widely ranging from physicochemical factors to even mechanical property.[41–44] What's more, the side reactions in P2D model lack an unifying perspective, particularly in terms of the equations and intertwine effects.

In line with the above issue, machine learning (ML), as a powerful technology, turns out to be an effective tool for addressing the aforementioned problems.[38,45–51] The major merit of this method is that it can function properly even without the prerequisite knowledge of degradation mechanisms.[38,52,53] Machine learning that treat the battery as a black box can achieve good performance has become the interest in the battery capacity prediction. It thus has

[z] E-mail: jbzhao@xmu.edu.cn

become the interest in the field of battery capacity prediction. According to some successful machine learning researches, the features which have statistically significant correlations with the capacity can be screened and selected for capacity predicting.[45,46] However, the drawback of this approach is that massive real datasets are needed for training. This could be a real issue when it comes to a battery, since the long cycle life span of the battery would require equivalently long testing time for construction of necessary databases. Besides, the cost of the experimental testing to obtain the necessary dataset would be enormous and make it cannot to be a routine way for research. Obviously, the scarcity of real experimental data is destined to be a main obstacles to carry out machine learning in pursuing reliable prediction of the total capacity throughout the whole cycle life.[54–56]

In complement to machine learning, the digital-twin emerging as an alternative have also been applied for capacity prediction.[38,57,58] This method, based on limited experimental data and electrochemical theory, can ensure high-quality extension of training datasets from numerical simulation. That is, the digital-twin method can fuse both merits of the above-mentioned model-based P2D method and data-driven machine learning, which thus enable easier and more reliable performance monitoring of a battery.[59] Digital twin technology can reasonably solve the problem of high test cost for acquiring massive experimental data, while ensuring the fidelity of training data for machine learning, so as to achieve reliable and stable behavior monitoring, prediction, evaluation and optimization during the battery life cycle.[57,60] So far as we know, such digital-twin method has not been deployed for capacity prediction of lithium ion battery.

In this work, a novel workflow based on limited experimental test data is proposed to predict the capacity of a battery built with $LiNi_{0.8}Co_{0.1}Mn_{0.1}O_2$/Graphite@$SiO_x$ battery. To identify the optimal operation condition which ensures the maximum total accumulative capacity releasing during cycle life, a numerical-simulation-based machine learning model is built to predict the battery degeneration behaviors. Based on the deterioration mechanism, the features highly related with the capacity fading were screened and adopted for the simulation model building. Based on the electrochemical theory, the numerical-simulation-based models with side reactions are calibrated and employed to create batches of high-quality training data for the machine learning training. The resulting digital-twin datasets were deployed for machine learning to provide high-precision predictions of total accumulative capacity throughout the cycle life; the prediction error was found to be less than 2%, and the cost can be reduced by 99%.

Experimental

Battery capacity fade test data.—25 experimental high capacity density $LiNi_{0.8}Co_{0.1}Mn_{0.1}O_2$/Graphite@$SiO_x$ lithium ion batteries were measured in this work. The battery was designed with the capacity 3Ah during the voltage windows of 2.75–4.2 V. All the battery tests are proceeded under the protocol of constant current-constant voltage (CC-CV). More detail information of the battery was listed in the supporting information.

The total capacity is calculated as the follows describe

$$Q_{sum} = \sum_{1}^{n} \hat{y}(\hat{y} > 80\% \cdot y_0)$$

.—Where y_0 is the initial discharge capacity during the first cycle. Note that, some batteries cycle life before failure bigger than 1000 cycles. In order to simplify the result, the total capacity calculation obeys another rule that all the capacity summed during the 1000 cycles.

FEM simulation.—The pseudo-two-dimensions model is built with the COMSOL Multiphysics platform. The 1D lithium ion battery model adds side reactions to simulate the capacity fade. Four different reactions contain negative electrode SEI formation (Graphite, SiO_x), active particle fragmentation (NCM811, SiO_x). The SEI formation is expressed as the cathodic Tafel equation. In addition, particle fragmentation is described as the Arrhenius equation. The temperature, rate, working voltage windows affect the reaction rate constant and result in different forms of capacity fade. The model calibration process relies on experience to adjust parameters to achieve consistency with experimental data.

Machine-learning model development.—The algorithms used in this research are Neural Network (NN) and random forest trees (RF). The NN is used to solve the regression problems (capacity fade fitting and prediction), whereas, the RF method is used for the classification (feature importance analysis). The NN model has the multi-layer perceptron (MLP) that trains using backpropagation which has the advantage in solving the non-linear problems. The NN model takes the form as follow proposed:

$$f(x) = W_2 g(W_1^T x + b_1) + b_2$$

Where W_1, W_2 represent the weights of the input layer and hidden layer, b_1, b_2 are model parameters. $g(x)$ is the activation function. The weights of NN are calculated in the training process by minimizing the square error loss function. Mathematically, the model train with ℓ_2 regularization by penalizing weights with large magnitudes to avoid overfitting. The loss function written as,

$Loss(\hat{y},\ y,\ W) = \frac{1}{2}\hat{y} - y_2^2 + \frac{\alpha}{2}W_2^2$Where the $\hat{y}$ is the predicted capacity, y is the test capacity. The weight was updated by backward-propagation from the output layer to the former layers. The backward-propagation is based on the stochastic gradient descent technique.

The predicted error refers to the actual measurement results, and the calculation process is as follows:

$$\text{Error} = \sum_{1}^{n} \frac{|\hat{y} - y|}{yn} * 100\%$$

Where the $\hat{y}$ is the prediction capacity, the y is the test capacity. The n is the cycle number. The computation process has been added in the manuscript.

All features in NN train data are standardized to have mean 0 and variance 1 for fasting converges and getting better solutions. To avoid overfitting, the data are randomly split into training and test sets with the ratio of 0.8:0.2. When the reporting performance measures on the test set, we instead choose to focus on the mean absolute error that is more intuitive than the root means squared error. RMSE is defined as follow:

$$RMSE = \sqrt{\frac{1}{n} i = 1 \sum_{i=1}^{n} (\hat{y} - y)^2}$$

The result predicted by NN is put in the RF model for features importance analysis. RF is the ensemble methods which generates hundreds of decision trees. The model output is dependent on the average of all the trees. We control the number of trees to get excellent performance and accuracy. The relative importance of the features could be got from the relative depth (i.e., expected fraction of the samples) of a feature used as a decision node in a tree. Permutation feature importance is computed in the RF model. The features are shuffled 10 times and the model is refitted to estimate the importance.

A step-by-step introduction. First of all, the data collected from the experiment was used to calibrate the COMSOL simulation model, and a high-precision simulation model achieved by controlling the side reaction parameters. Then, used the scanning function of COMSOL to output batch simulation data. Next, these simulation data was applied for machine learning training and prediction. The above continuous behavior is an iterative process. When the machine learning outputs the target results, the verify experiments was used to check the result. The error between the experimental data and the predicted data determine the reliable of digital twin output. If the results are not reliable, the results of the verification experiments are re-input into COMSOL for calibration, and the next iteration process is carried out until the results of the digital twin are reliable.

The data processing and machine learning (regression and feature importance analysis) are played in Python with Anaconda, Scikit-learn package, NumPy, Panda. The visualization were illustrated with Python and MATLAB. The FEM simulation model of lithium ion battery is built in COMSOL Multiphysics.

Machine Learning Model

Battery performance.—Several factors may impact the performance of battery cycling capacity, such as cycle rate, temperature, cut-off voltage (COV), operation windows and so on. The capacity fading exhibits diverse modes under different operation conditions. Herein, these factors have varying influences on the total energy release of the battery during the life cycle (Fig. 1A). In our previous research,[39] the processes of battery degradation has the highly non-linear and coupled natures (Fig. S1). Remarkably, the difference of the total accumulative capacity released during the life cycle may vary by up to 5 times (Fig. S2). As a result, the industry's primary concern should be the battery's operation condition for the maximum total accumulative capacity releasing. Further research and engineering optimization may be of great scientific significance in addressing this concern and improving the better's performance.

Our research attempts to address the challenge: prediction of the best performance with limited number of experimental battery data. Only 25 batteries cycling data serves as the starting point for predicting the battery life cycle. The 25 batteries are performed with 25 different cycling protocols, contains different temperatures (25 °C, 35 °C, 60 °C), cut-off voltage (2.75–4.0 V, 2.75–4.2 V, 3–4.05 V, 3–4.25 V), rate (1 C, 2 C, 3 C). The important features related to the capacity fade is adopted to get reliable result (Fig. 1B). Those data used in prediction model is not sufficient so can't ensure the accuracy of the result. Digital-twins as the powerful magic weapon in the dilemma of lacking real data is used to extend high-quality datasets. The sight is drawn to the finite element model (FEM) simulation to extend the training sets for digital-twins. Based on the theory of electrochemical and fading, the numerical simulation model could output batches of data which ensure adequate sample for the machine learning model training.

FEM simulation.—The machine learning model based on limited experimental datasets has limited confidence interval. Some results could show good agreement with experiments, prediction results which outside the training datasets condition cannot achieve appropriate accuracy (Fig. 2). This point can be checked based on experience or operation. The strategy of extending training datasets via FEM simulation doesn't imply that any output data of the numerical model is adopted. The inaccuracy of the numerical model will lead to weak generalization ability of the training data, which makes it difficult to ensure the consistency between the prediction results of machine learning and the experimental results. We have built a numerical simulation model in our prior research, but the model primarily focuses on a single temperature and barely consists with test data under different temperatures. The simulation model built based on the theory of degradation and electrochemical can solve the availability of the data from the root. Moreover, the accurate simulation model that consistent with the real test further enhance the confidence of output data. Herein, calibrating the numerical model will provide quality data.

In comparison with the previous work that focused on the working voltage windows, the real test data in this research adds two more features (i.e., temperature, rate). Increasing the numbers of features will be more favorable for the prediction in regression model while avoiding redundant and irrelevant features (Fig. 1B).

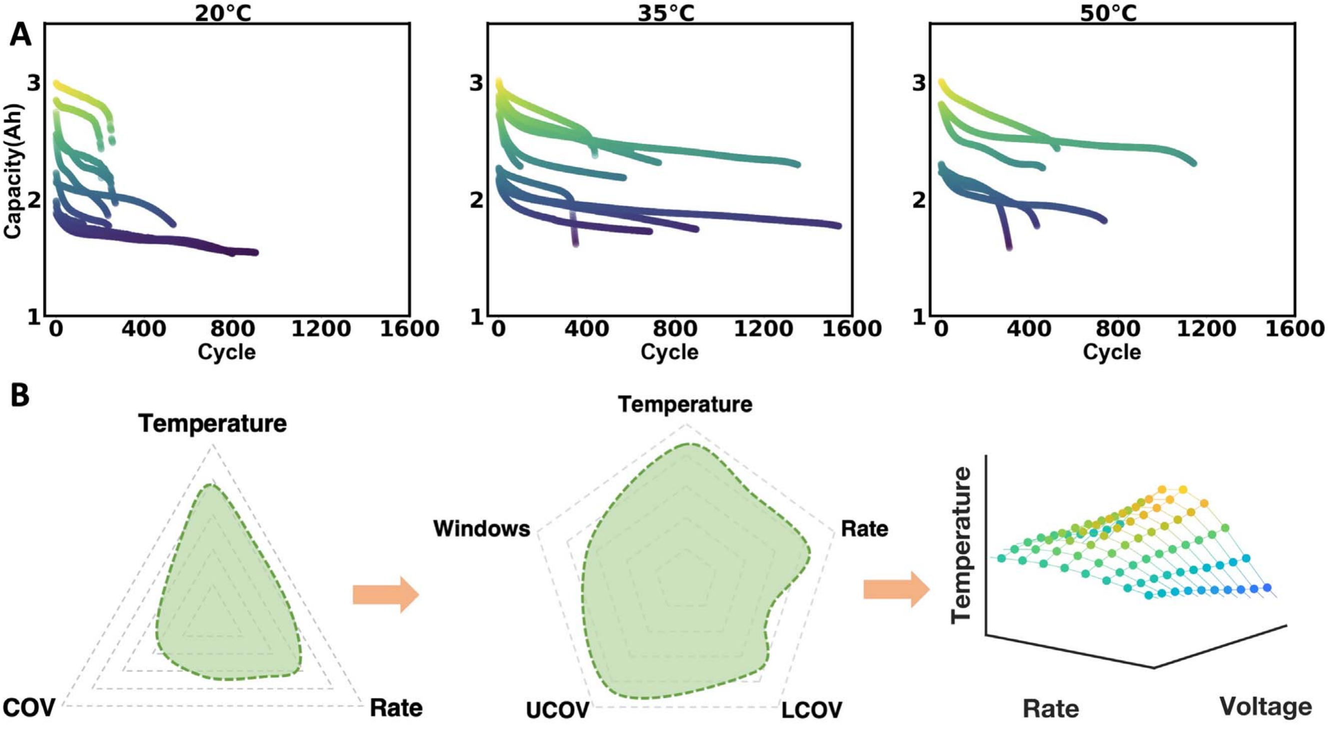

Figure 1. Different capacity fade mode of lithium ion battery and schematic diagram of the characterizes related to the capacity fading. (A) Different batteries capacity fade mode under different temperature shows above the figure. The different modes are relative with the different working voltage windows (2.75–4.0 V, 2.75–4.2 V, 3–4.05 V, 3–4.25 V) and different rate (1 C, 2 C, 3 C). Some batteries sharply decrease to failure between 500 cycles. However, some battery life can reach 1000 cycles. (B) More factors (features) are selected for the precise capacity degeneration research. The cut-off voltage is subdivided into upper cut-off voltage (UCOV) and low cut-off voltage (LCOV). The windows represent as the range of working voltage.

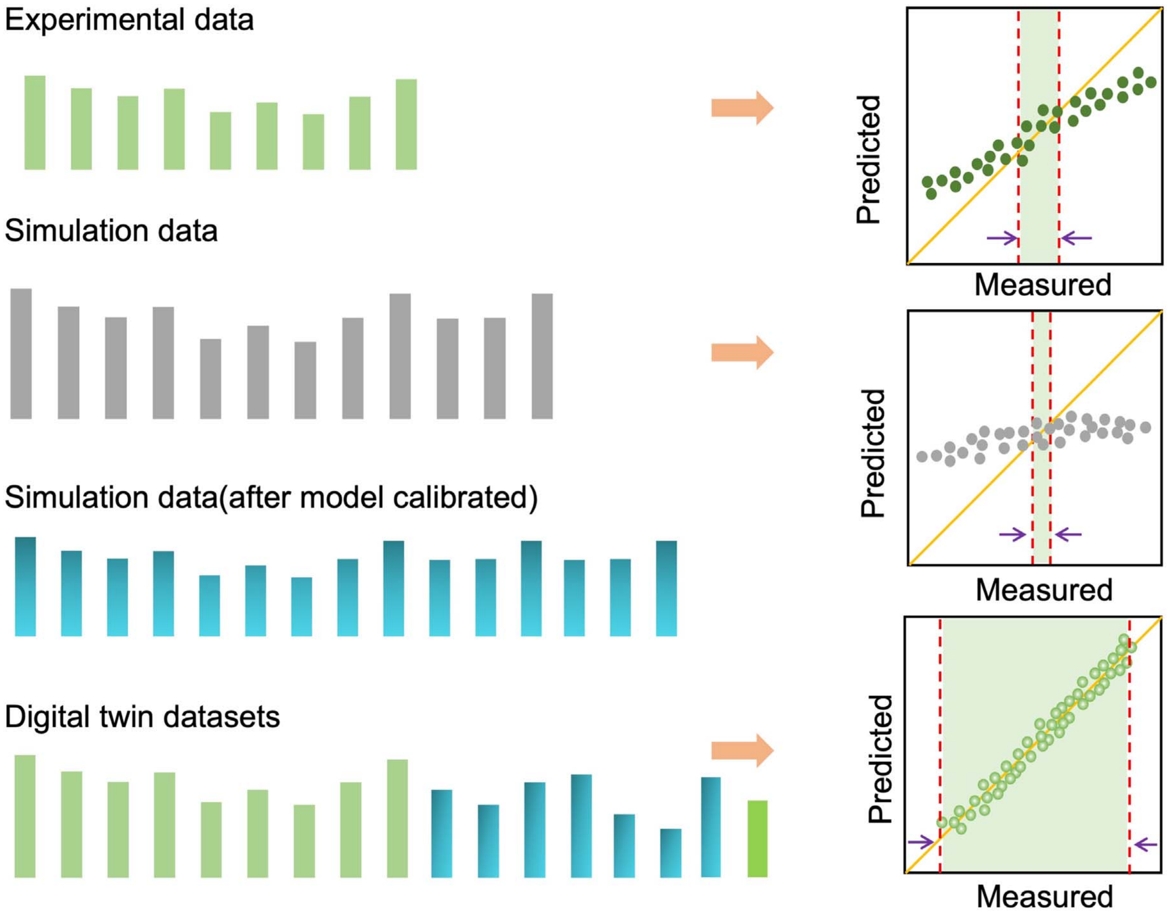

Figure 2. Schematic diagram of confidence interval of machine learning using different data. The experimental data is combined from test data under limited cycling protocols. The simulation model could extend the training data, especially add the data that under different cut-off voltage. The calibration process is necessary to ensure the generalizability of numerical data. After calibrating with real test data, the numerical simulation can output batches of high-quality data for machine learning training. The digital-twins datasets combined with limited test data and massive simulation data can effectively broaden the confidence interval and improve the accuracy of machine learning.

This real data is utilized to calibrate the numerical model to high-accuracy (Figs. 3A, 3B). In fact, considerable effort and time are spent in the calibration of the model.

The battery capacity is affected by four side reactions in the numerical model, all of which are based on the physicochemical reactions.[39] The calibration procedure are accomplished through controlling side reactions, and the cycling data of 25 batteries are sat as the standard reference (Figs. 3A, 3B). The calibration model is used to get high-quality numerical data from sweeping different temperatures, rates and working voltage windows (Fig. S3).

This strategy does not require massive experiment but get high-quality data, has considerable advantages and accelerates the speed of machine learning. Four side reactions are involved in the COMSOL simulation model, including the growth of SEI, the deterioration of positive and negative active materials, and the increase of interface impedance. These reactions vary with operating conditions, resulting in nonlinear changes in capacity degradation. However, the law of the side reactions varying with operating conditions have been grasped. Such as the thickness of SEI will rise with the temperature and the deterioration of the positive active particles are aggravated with the increase of the cut-off voltage. Therefore, the model was calibrated by change the reaction rate constants based on experience, experimental data, and relevant literature.

The error between simulation result and test data is reduced to whining 3%, ensure that the quality data applied for the machine learning. Numerical data and the experimental data are combined as the initial digital-twins datasets. The digital-twins datasets are used to train and deploy machine learning models. Whereas, we are well aware of our approach's shortcoming, which is that the quality of digital-twins datasets cannot be totally consistent with real-world test data. In this case, machine learning has to include important iterative training process (Fig. S4). In each iteration, the target experimental results are compared and collected into the datasets. When the predicted result agrees with the experimental result, the iterative training stops and thus a machine learning model with high accuracy is obtained.

ML model training.—The digital-twins datasets are the center of the workflow for battery capacity prediction. Furthermore, powerful algorithms have been developed and employed successfully, including Deep Neural Network, Support Vector, Kernel Ridge and Naive Bayes and so on.[45–47,61–64] Herein, we preferentially pick the machine learning algorithms which are widely and generally applied. Neural network based estimator, which use nonlinear mapping to predict target value, are preferred for highly nonlinear applications.[61,65–69] Alternatively, they can be called black-box-based method because it does not require professional knowledge about the internal dynamics. In other words, this algorithm enhances the workflow's universality and consistency.

The digital-twins datasets are split into training and test datasets for the neural network building which has multi-layer perceptron (MLP). In terms of non-linear model, the multi-layer perceptron, a supervised learning algorithm, is particularly efficient (Fig. S5). The K-fold cross validation method is used to evaluate the estimator while choosing the high-performance structure for the multi-layer

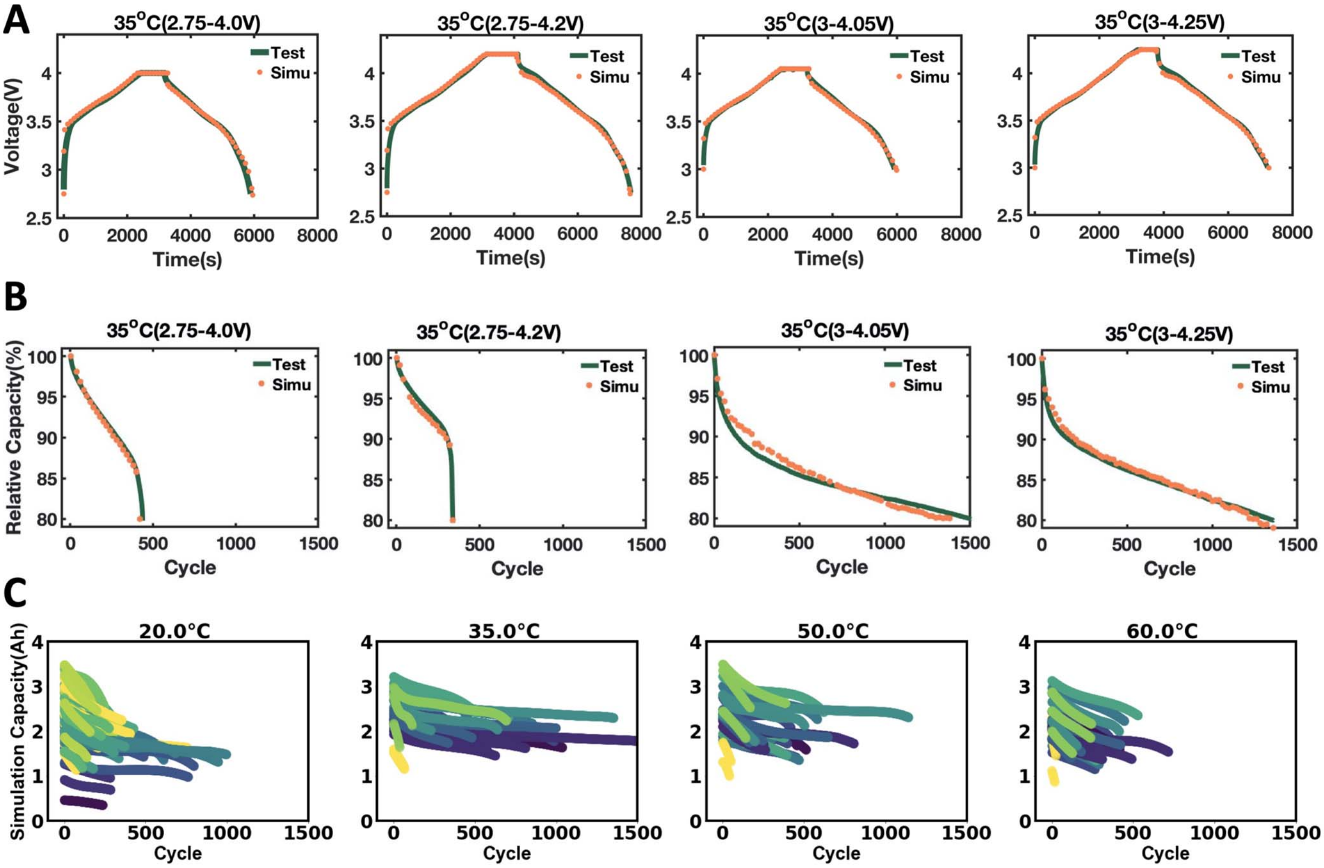

Figure 3. Schematic diagram of calibration process of numerical simulation based model. The simulation data get from sweeping the calibrated numerical model. (A), (B) The calibration process of the numerical model is corresponded with the real test data, no matter the cycle voltage or the capacity. (C) After the simulation model calibrated with real test data, the output numerical data can be of sufficient quality for the training. Meanwhile, the sweep concentrates on the test condition's surroundings in order to avoid numerical inaccuracies.

perceptron. In the basic approach, the K-fold cross validation involves 10 mutually exclusive subsets. Meanwhile, the root mean squared error (RMSE) is introduced as the fitting accuracy to the model evaluation system. The data points that does not overfit is selected under the premise of accuracy. (Fig. 4). Naturally, the computer resources also are concerned in the workflow. Finally, the neural network is designed with 3 hidden-layer and each layer has 50 neurons.

Before the data are put in the MLP, the digital-twins datasets are normalized. With the help of the gradient descent rule, the parameters of MLP are updated by the partial derivative of loss function.[70–72] This loop iterates continues until set accuracy or the maximum iterations (loss value is minimum) is reached. The learning rate is set to adaptive mode in the loop, and the solver's maximum number of iterations is set to 10^{-6}. The solver display result of converging through limited epochs (Fig. S6). The neural network model has a great prediction ability regardless of the training or test datasets (Fig. 5A). Meanwhile, the neural network's parameters have been properly confirmed.

Prediction result.—The machine learning workflow gives a reasonable result after six iterations (Table I). The maximum capacity released under the restrained condition is found by accumulate of all the battery discharge capacity during the life cycling. If distributing the test evenly within the constraints, the times of iterations will reduce. However, it does not affect the prediction result but the prediction speed, and can be solved through times of iterations. After the last iteration, the prediction result is verified by the experiment (Fig. 5C). Surprisingly, the average percentage error of the battery capacity prediction on the training dataset is lower than 1% (Fig. 5C (3.0–4.3 V)). An additional experiment after the last iteration is designed to verify the prediction result (Fig. 5C (3.2–4.3 V). Note that This set of experimental data was not applied to the training data.). The prediction result under other condition can be accepted because average error is around 1%. In our opinion, the prediction result which condition among the datasets has more rationale value than the condition beyond the datasets. Mathematically, compared with the result of interpolation, the extrapolated values have relatively more risk and lower confidence interval. As mentioned, distributing the test evenly may promote the accuracy of prediction result. Increasing test samples dispersed inside the boundary for the training may alleviate the difficulty of lower confidence interval.

Detailly, the condition includes the various temperatures (i.e., 15 °C–60 °C, step is 5 °C), rates (i.e., 1–3 C, step is 0.5 C, where the C is the charge and discharge current,1 C = 3 A), the LCOV(i.e., 2.7–3.5 V, step is 0.1 V), the UCOV (i.e., 3.7–4.5 V, step is 0.1 V). In where, the temperature is below the 60 °C and the rate set lower than 3 C for preventing the thermal runaway from occurring. Besides, the working voltage windows which the difference between UCOV and LCOV is set more than 1 V to assure energy density of battery in operating. All the condition sat are constringed by the electrochemical. A higher UCOV will result in severe interfacial side reaction such as lithium deposition, particle fragmentation and so on. The anode collect (Copper) will dissolute under lower LCOV and deposition in the cathode interface. These side reactions not only result in a larger capacity loss but also introduce serious safety risks.[73–79] In summary, 2250 conditions are found that fulfill the constrains. In this work, 25 sets of experimental data were used in the initial study, and 6 sets of experimental data were used in the

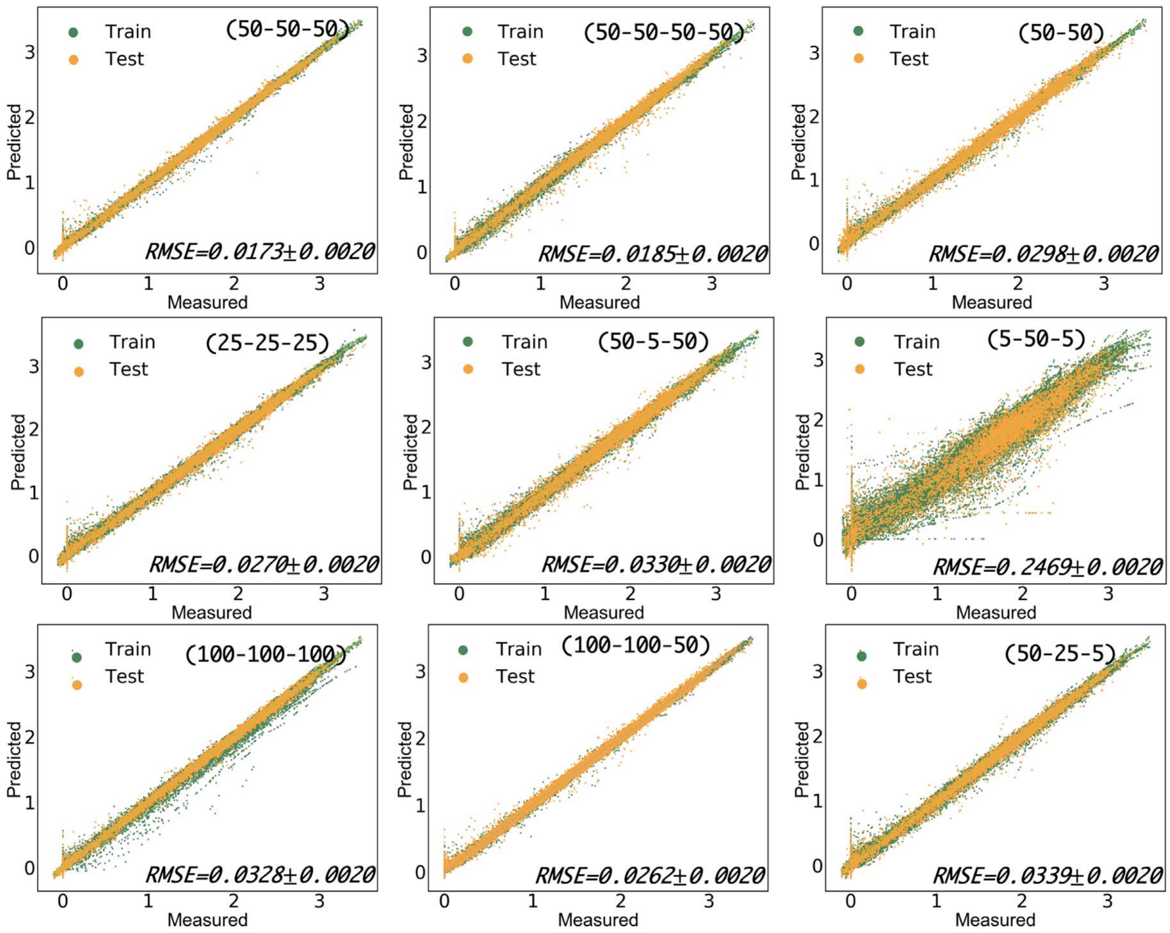

Figure 4. Cross-validation comparison framework. The parameters of the MLP in neural network is determined by the K-fold cross-validation and RMSE (i.e., the numbers of hidden layers and nodes). Noting that the numbers in the upper right represent the structure of neural network.

iterative process. The 99% calculation in this research is the basis on substituting all conditions experimentally.

The target condition is obtained by the workflow through limited times of checks and iterations. Bright colors are used for specifically identify, and it can be concluded that the relatively high capacity released during the life cycle is focused in one area (Fig. 6A). Taking the temperature for example, the condition which total accumulative capacity higher than 2 kAh is between 35 °C–50 °C (Fig. 6B). This fact signs the target we obtained is the global maximum. With the data of all the condition prediction results, a random forest trees model is developed for the importance analysis (Fig. 6C). According to the result of RF model, the UCOV has a greater impact on battery capacity fading than other factors. However, the temperature is as important as the rate. The battery appears to be slightly affected by the working voltage windows (WV-windows).

The feature importance result is excellently agree with our experiments and test. Based on the knowledge of electrochemistry, extreme temperature(i.e., low and high temperature) and rate(i.e., heavy current) will inevitably cause more side reaction. Furthermore, the UCOV not only will accelerate the side reaction but also aggravate the material failure. Whereas, the LCOV appeared to be less important for the capacity fade due to the sharp voltage change at the charge start and discharge end. Taken together, the analysis highlights the importance of the features and quantify the impacts. Deeply speaking, the result enables for the monitoring of training and battery performance evaluation.

As we stated at the beginning, the feature selected based on the electrochemical will help to obtain promised result. For the capacity degeneration prediction, all the important features affect the cycle are considered in the model not only contribute to the speed of iteration but also the accuracy of prediction. In the last, the datasets after iteration in mentioned workflow are used to test the features we selected (Fig. S8). However, before used for training, the datasets are added or reduced some feature to verify. The change of the feature based on the relationship with capacity. The standard of the verification is based on the test data (45 °C, 1 C, 3.2–4.3 V, it needs to be addressed that this data is not conclude in the training datasets) (Figs. S8C, S8D). The RMSE of the model which numbers of features reduced is drastically increased. Whereas, the accuracy is slightly improved with the number of features added. From the accuracy view, the more feature in the training datasets will increased the reliability of the model. Nevertheless, this model is prone to suffer from overfitting and consume more computer resource. Besides, the cost of training is likely to dramatically rise. Collectively, the features we adopt from the electrochemical theory are demonstrated to be appropriate for the model and training.

Conclusions

Data-driven approach for the battery capacity prediction is obstructed by the amount of training data. In this research, the numerical simulation model built on electrochemical knowledge used as the core to effectively extend the training datasets. Digital-twins datasets combine the simulation data and limited test data is used for the machine learning training. The features based on the deterioration mechanism are scientifically adopted for the model building. Numerical simulation based machine learning model is built for the capacity prediction before degradation. The machine learning workflow which uses the test data for iteration can reach

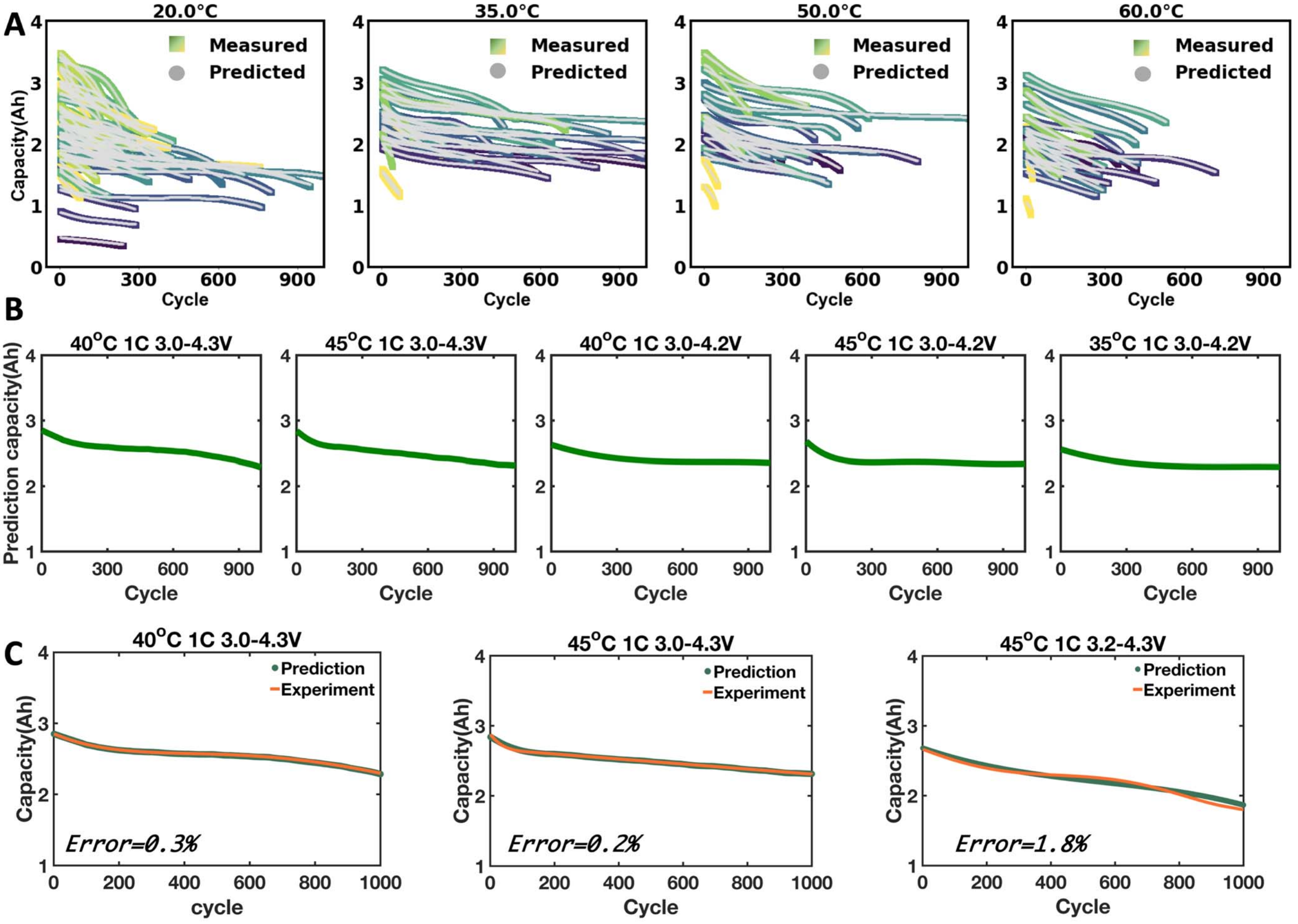

Figure 5. Prediction result of the machine learning. (A) training and the model built process of the neural network. The result is classified by temperature. (B) top five batteries by ranking the sum capacity during the life cycle, where the operating condition is signed above of figure. The data statistics are performed from the prediction result of all conditions. (C) validate the predictive accuracy of the machine learning model using experimental data.

Table I. Prediction results of the iteration.

Iteration	Temperature(°C)	LCOV(V)	UCOV(V)	Rate(C)	Sum capacity (kAh)	Error (kAh)
1	15	2.7	3.7	1.5	3.119	2.144
2	15	2.7	4.5	1	2.963	2.640
3	55	3.1	4.5	1	2.850	1.558
4	45	3	4.3	1	2.835	0.341
5	40	3	4.3	1	2.560	0.007
6	40	3	4.3	1	2.550	0.003

[*] Noting that the table ranked by the iteration times. The conditions in table is corresponded with the prediction of the maximum capacity released during the cycle life in each iteration. Besides, the 6th iteration isn't verified because the target is same as 5th iteration. The errors are calculated by comparing the prediction with experimental data.

high accuracy result. The innovation of this work lies in the machine learning workflow is trained on the digital-twins datasets coming from the numerical simulation to predict battery cycle life.

The focus of this research, in our opinion, should be on the electrochemical method of building digital-twins datasets. Even though the high precision result and quick approach have been achieved, there are still some efforts spent on the simulation model calibration. To be honest, the ability for simulation and the knowledge of electrochemical is vital for this workflow, and will hinder it generally applying. The machine learning workflow which can automatically calibration the numerical simulation model extremely attracts us to develop and build.

Acknowledgments

We gratefully acknowledge the financial support of National Key Research and Development Program of China (2017YFB0102000), National Natural Science Foundation of China (21875195), the Key Project of Science and Technology of Xiamen (3502Z20201013), the Fundamental Research Funds for the Central Universities (20720190040). We would like to thank Prof. Yao Zhou for his valuable guidance and suggestions. We are grateful to Feng Wang, Chaoyue Liu and Yuanyuan Li for their comments on earlier versions of the manuscript.

ORCID

Jinbao Zhao https://orcid.org/0000-0002-2753-7508

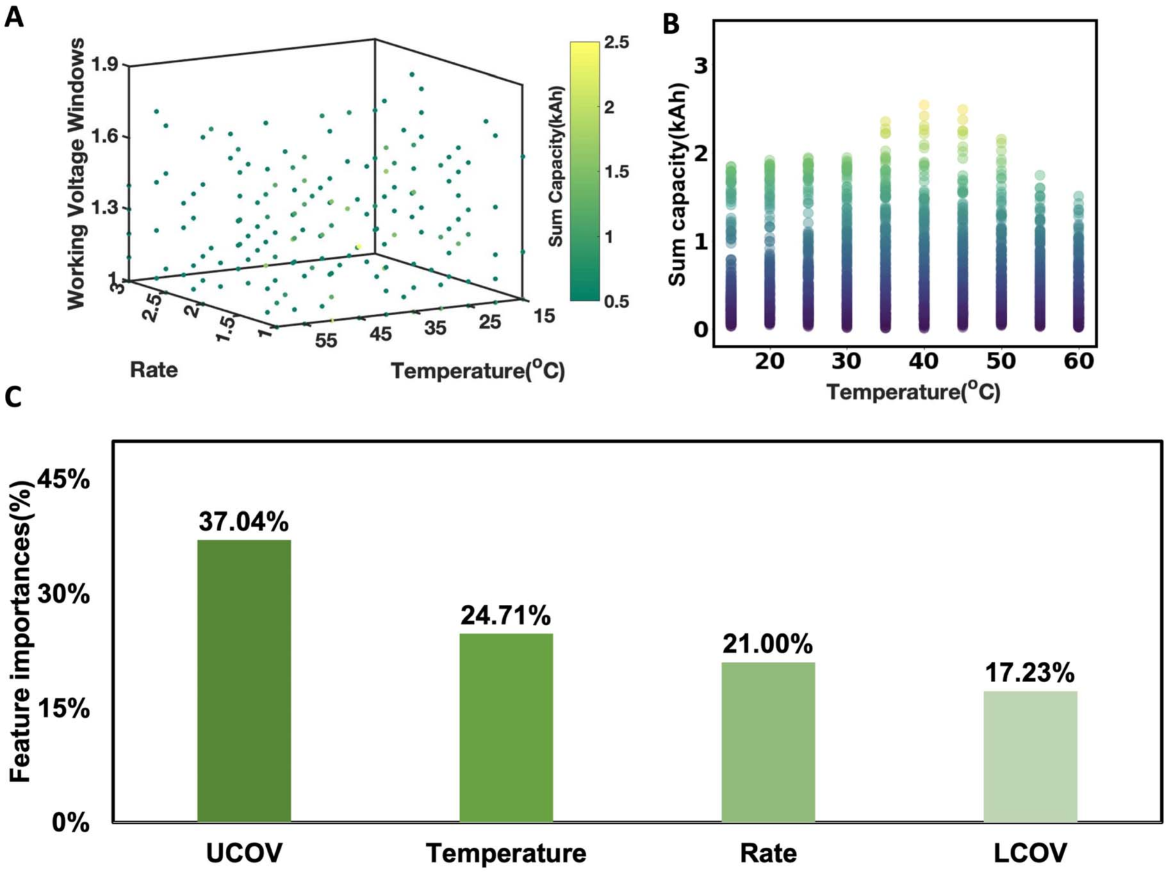

Figure 6. Feature importance of the prediction result analyze by the random forest algorithm. (A) The sum capacity distribution in different temperatures, rates, working voltage windows. (B) The sum capacity is classified by the temperature, and the result show the maximum sum capacity at the temperature of 40 °C. (A) and (B) means the prediction result was the global maximum value. (C) Normalized importance of the feature in the input value.

References

1. F. Duffner, N. Kronemeyer, J. Tübke, J. Leker, M. Winter, and R. Schmuch, *Nat. Energy*, **6**, 123 (2021).
2. P. R. Shearing and L. R. Johnson, *Joule*, **4**, 1359 (2020).
3. J. W. Choi and D. Aurbach, *Nat. Rev. Mater.*, **1**, 16013 (2016).
4. M. Li, J. Lu, Z. Chen, and K. Amine, *Adv. Mater.*, **30**, 1800561 (2018).
5. T. M. Gür, *Energy Environ. Sci.*, **11**, 2696 (2018).
6. J. Kim, J. Oh, and H. Lee, *Appl. Therm. Eng.*, **149**, 192 (2019).
7. S. Li, C. Gu, P. Zhao, and S. Cheng, *Energy Convers. Manag.*, **235**, 114004 (2021).
8. S. Li and P. Zhao, *J. Energy Storage*, **33**, 102064 (2021).
9. C. Lyu, Y. Jia, and Z. Xu, *Appl. Energy*, **299**, 117243 (2021).
10. V. Sulzer et al., *Joule*, **5**, 1934 (2021).
11. K. Liu, X. Hu, Z. Wei, Y. Li, and Y. Jiang, *IEEE Trans. Transp. Electrification*, **5**, 1225 (2019).
12. R. Xiong, Y. Zhang, J. Wang, H. He, S. Peng, and M. Pecht, *IEEE Trans. Veh. Technol.*, **68**, 4110 (2019).
13. L. Zheng, J. Zhu, D. D. C. Lu, G. Wang, and T. He, *Energy*, **150**, 759 (2018).
14. R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter, *Nat. Energy*, **3**, 267 (2018).
15. S. Randau et al., *Nat. Energy*, **5**, 259 (2020).
16. X. Li, C. Yuan, X. Li, and Z. Wang, *Energy*, **190**, 116467 (2020).
17. H. Li, X. Kong, C. Liu, and J. Zhao, *Appl. Therm. Eng.*, **161**, 114144 (2019).
18. S. F. Schuster, T. Bach, E. Fleder, J. Müller, M. Brand, G. Sextl, and A. Jossen, *J. Energy Storage*, **5**, 212 (2016).
19. W. Chang, C. Bommier, T. Fair, J. Yeung, S. Patil, and D. Steingart, *J. Electrochem. Soc.*, **167**, 090503 (2020).
20. J. P. Pender et al., *ACS Nano*, **14**, 1243 (2020).
21. S. K. Jung, H. Gwon, J. Hong, K. Y. Park, D. H. Seo, H. Kim, J. Hyun, W. Yang, and K. Kang, *Adv. Energy Mater.*, **4**, 1300787 (2014).
22. J. Y. Wang, S. N. Guo, X. Wang, L. Gu, and D. Su, *J. Electrochem.*, **28**, 2108431 (2022).
23. K. Liu, T. R. Ashwin, X. Hu, M. Lucu, and W. D. Widanage, *Renew. Sustain. Energy Rev.*, **131**, 110017 (2020).
24. S. Atalay, M. Sheikh, A. Mariani, Y. Merla, E. Bower, and W. D. Widanage, *J. Power Sources*, **478**, 229026 (2020).
25. L. Yang, X. Cheng, Y. Ma, S. Lou, Y. Cui, T. Guan, and G. Yin, *J. Electrochem. Soc.*, **160**, A2093 (2013).
26. F. Nobili, S. Dsoke, M. Mancini, R. Tossici, and R. Marassi, *J. Power Sources*, **180**, 845 (2008).
27. T. Waldmann, B.-I. Hogg, and M. Wohlfahrt-Mehrens, *J. Power Sources*, **384**, 107 (2018).
28. B. Wu, J. Lochala, T. Taverne, and J. Xiao, *Nano Energy*, **40**, 34 (2017).
29. G. Bieker, M. Winter, and P. Bieker, *Phys. Chem. Chem. Phys.*, **17**, 8670 (2015).
30. G. Liu and W. Lu, *J. Electrochem. Soc.*, **164**, A1826 (2017).
31. H. H. Ryu, K. J. Park, C. S. Yoon, and Y. K. Sun, *Chem. Mater.*, **30**, 1155 (2018).
32. G. Gachot, P. Ribière, D. Mathiron, S. Grugeon, M. Armand, J. B. Leriche, S. Pilard, and S. Laruelle, *Anal. Chem.*, **83**, 478 (2011).
33. P. Barnes et al., *J. Power Sources*, **447**, 227363 (2020).
34. H. Li, A. Liu, N. Zhang, Y. Wang, S. Yin, H. Wu, and J. R. Dahn, *Chem. Mater.*, **31**, 7574 (2019).
35. M. Miyachi, H. Yamamoto, H. Kawai, T. Ohta, and M. Shirakata, *J. Electrochem. Soc.*, **152**, A2089 (2005).
36. H. Yamamura, K. Nobuhara, S. Nakanishi, H. Iba, and S. Okada, *J. Ceram. Soc. Jpn.*, **11**, 119 (2011).
37. S. S. Zhang, *Energy Storage Mater.*, **24**, 247 (2020).
38. X. Hu, L. Xu, X. Lin, and M. Pecht, *Joule*, **4**, 310 (2020).
39. H. Li, W. Ji, Z. He, Y. Zhang, and J. Zhao, *J. Energy Storage*, **47**, 103830 (2022).
40. Y. Yang, L. Chen, L. Yang, X. Du, and Y. Yang, *Energy*, **206**, 118155 (2020).
41. T. Waldmann, S. Gorse, T. Samtleben, G. Schneider, V. Knoblauch, and M. Wohlfahrt-Mehrens, *J. Electrochem. Soc.*, **161**, A1742 (2014).
42. T. C. Bach, S. F. Schuster, E. Fleder, J. Müller, M. J. Brand, H. Lorrmann, A. Jossen, and G. Sextl, *J. Energy Storage*, **5**, 212 (2016).
43. T. Waldmann, G. Bisle, B.-I. Hogg, S. Stumpp, M. A. Danzer, M. Kasper, P. Axmann, and M. Wohlfahrt-Mehrens, *J. Electrochem. Soc.*, **162**, A921 (2015).
44. M. Shen and Q. Gao, *Int. J. Energy Res.*, **43**, 5042 (2019).
45. K. A. Severson et al., *Nat. Energy*, **4**, 383 (2019).
46. Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee, *Nat. Commun.*, **11**, 1706 (2020).
47. B. Jiang et al., *Joule*, **5**, 3187 (2021).
48. P. M. Attia et al., *Nature*, **578**, 397 (2020).
49. Q. Deng and B. Lin, *Energy Mater.*, **1**, 100006 (2021).
50. Y. Liu, Q. Zhou, and G. Cui, *Small Methods*, **5**, 2100442 (2021).
51. Y. Liu, B. Guo, X. Zou, and S. Shi, *Energy Storage Mater.*, **31**, 434 (2020).
52. M. S. Hosen, J. Jaguemont, J. Van Mierlo, and M. Berecibar, *iScience*, **24**, 102060 (2021).

53. H. Valladares, T. Li, L. Zhu, H. El-Mounayri, A. M. Hashem, A. E. Abdel-Ghany, and A. Tovar, *J. Power Sources*, **528**, 231026 (2022).
54. C. Lv et al., *Adv. Mater.*, **34**, 2101474 (2021).
55. M. A. Hannan, M. S. H. Lipu, A. Hussain, P. J. Ker, T. M. I. Mahlia, M. Mansor, A. Ayob, M. H. Saad, and Z. Y. Dong, *Sci Rep.*, **10**, 4687 (2020).
56. A. Samanta, S. Chowdhuri, and S. S. Williamson, *Electronics*, **10**, 1309 (2021).
57. B. Wu, W. D. Widanage, S. Yang, and X. Liu, *Energy AI*, **1**, 100016 (2020).
58. X. Qu, Y. Song, D. Liu, X. Cui, and Y. Peng, *Microelectron. Reliab.*, **114**, 113857 (2020).
59. N. G. Panwar, S. Singh, A. Garg, A. K. Gupta, and L. Gao, *Energy Technol.*, **9**, 2000984 (2021).
60. N. Dawson-Elli, S. B. Lee, M. Pathak, K. Mitra, and V. R. Subramanian, *J. Electrochem. Soc.*, **165**, A1 (2018).
61. Y. Zhang, R. Xiong, H. He, and M. G. Pecht, *IEEE Trans. Veh. Technol.*, **67**, 5695 (2018).
62. J. Mao, J. Miao, Y. Lu, and Z. Tong, *Chin. J. Chem. Eng.*, **37**, 1 (2021).
63. A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, *J. Power Sources*, **239**, 680 (2013).
64. J. Li, Q. Pan, and P. Duan, *IEEE Trans. Cybern.*, **46**, 1311 (2016).
65. A. Kruger, W. F. Krajewski, J. J. Niemeier, D. L. Ceynar, and R. Goska, *IEEE Access*, **4**, 8948 (2016).
66. L. S. Bruckman, N. R. Wheeler, J. Ma, E. Wang, C. K. Wang, I. Chou, J. Sun, and R. H. French, *IEEE Access*, **1**, 384 (2013).
67. S. Khaleghi, D. Karimi, S. H. Beheshti, M. S. Hosen, H. Behi, M. Berecibar, and J. Van Mierlo, *Appl. Energy*, **282**, 116159 (2021).
68. S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, *Appl. Energy*, **260**, 114296 (2020).
69. J. Tian, R. Xiong, W. Shen, and J. Lu, *Appl. Energy*, **291**, 116812 (2021).
70. X. Dang, L. Yan, K. Xu, X. Wu, H. Jiang, and H. Sun, *Electrochim. Acta*, **11**, 356 (2016).
71. J. Wu, C. Zhang, and Z. Chen, *Appl. Energy*, **173**, 134 (2016).
72. M. Ibrahim, S. Jemei, and G. Wimmer, *Electr. Power Syst. Res.*, **136**, 262 (2016).
73. X. Feng, M. Ouyang, X. Liu, L. Lu, Y. Xia, and X. He, *Energy Storage Mater.*, **10**, 246 (2018).
74. G. Xu, L. Huang, C. Lu, X. Zhou, and G. Cui, *Energy Storage Mater.*, **31**, 72 (2020).
75. S. Zhang, *Energy Environ. Mater.*, 1 (2021).
76. X. Liu et al., *Joule*, **2**, 2047 (2018).
77. K. Liu, Y. Liu, D. Lin, A. Pei, and Y. Cui, *Sci. Adv.*, **12** (2018).
78. L. Li, S. Basu, Y. Wang, Z. Chen, P. Hundekar, B. Wang, J. Shi, Y. Shi, S. Narayanan, and N. Koratkar, *Science*, **359**, 1513 (2018).
79. Q. Wang, P. Ping, X. Zhao, G. Chu, J. Sun, and C. Chen, *J. Power Sources*, **208**, 210 (2012).