**Title: Data-driven Lithium-ion Battery Capacity Degradation Early Prediction and Modeling**

**Abstract**: With the increasing deployment of lithium-ion (Li-ion) batteries in various applications, their inevitable capacity degradation is becoming a significant issue. It is essential for a variety of Li-ion battery market participants to accurately model and predict the capacity degradation to improve the awareness of battery aging behaviors in making battery planning and operating decisions. However, there are two challenges in incorporating capacity degradation into the decision-making problems: 1) the accurate and efficient capacity degradation modeling and prediction across different battery chemistries and operating conditions, and 2) the decision-making model complexity increases with the nonlinear model and non-analytical procedures in capacity degradation calculation. This dissertation creates novel optimization- and learning-based methodologies that aim to help address the above two challenges. Chapters 1 and 2 elaborate on the Li-ion capacity degradation review and this dissertation's research objective and boundary. The main innovative works of the dissertation are summarized below.

First, we present an empirical model-augmented end-to-end learning framework in chapter 3 to predict the entire capacity-fade trajectory of a cell using only early-life data, e.g., data from the first 100 cycles or the first week of test. The end-to-end learning framework simultaneously fits an empirical model to estimate the capacity trajectory and trains a machine learning model to estimate the parameters of the empirical model using early-life data. We prove that the end-to-end learning framework yields a lower training error and a tighter generalization gap compared to a baseline sequential optimization framework. We also extend the end-to-end framework to enable probabilistic predictions, allowing us to further understand predictive uncertainty. The comparison with state-of-the-art machine/deep learning and physics model demonstrates the advance of the proposed end-to-end learning models. Its effectiveness has been validated on three large battery datasets across battery chemistries and operating conditions.

Based on the above end-to-end model, early prediction-driven capacity degradation modeling is presented in chapter 4. The Gaussian process regression (GP) model is used to predict the empirical model parameters using the inputs of test conditions (i.e., charge current, discharge current, and depth of discharge). The GP model is selected because of its probabilistic predictions to model the cell-to-variation under identical test conditions. The end-to-end model is introduced to generate the target empirical model parameters, significantly reducing the necessary battery aging test time cost by more than 90%. In addition, three different sampling algorithms for battery aging experimental design are compared (i.e., random search, Latin hypercube sampling, and Bayesian optimization). The Bayesian optimization using a custom acquisition function achieved the highest predictive accuracy and the lowest run-to-run variation. The acquisition function is designed to select new test conditions where the GP capacity degradation model has a higher predictive uncertainty (exploration) and lower cross-validation predictive error (exploitation). Comparative case studies on 240 cells are conducted to evaluate the performance of early prediction-assisted capacity degradation modeling.

Chapter 5 presented a novel use case on how to incorporate capacity degradation into sizing a behind-the-meter Li-ion BESS. We first define a two-stage stochastic programming problem to size a Li-ion BESS to maximize life-cycle profit while also considering the nonlinear capacity degradation. We propose a model-based optimal planning method and a model-free reinforcement learning method to deal with the two-stage stochastic programming problem. Also, a new Li-ion BESS life-cycle operating environment is established to evaluate sizing and operation solutions. In the model-based method, the stochastic problem is initially formulated as a deterministic equivalent mixed-integer nonlinear program that is challenging to solve. We then convert it to a more solvable continuous nonlinear program without loss of accuracy through equivalent simplifications. The model-free method combining Bayesian optimization and reinforcement-learning algorithms is investigated, primarily serving as a comparison model. A case study is conducted using one-year load, solar generation, temperature, and price data in California.