# Ph.D. Thesis: Optimal Probing of Battery Cycles for Machine Learning-Based Model Development

## Abstract

This dissertation examines the problems of optimizing the selection of the datasets and experiments used for parameterizing machine learning-based electrochemical battery models. The key idea is that data selection, or "probing" can empower such models to achieve greater fidelity levels. The dissertation is motivated by the potential of battery models to enable both the prediction and optimization of battery performance and control strategies. The literature presents multiple battery modeling approaches, including equivalent circuit, physics-based, and machine learning models. Machine learning is particularly attractive in the battery systems domain, thanks to its flexibility and ability to model both battery performance and aging dynamics. Moreover, there is a growing interest in the literature in hybrid models that combine the benefits of machine learning with either the simplicity of equivalent circuit models or the predictiveness of physics-based models or both.

The focus of this dissertation is on both hybrid and purely data-driven battery models. Moreover, the overarching question guiding the dissertation is: how does the selection of the datasets and experiments used for parameterizing these models affect their fidelity and parameter identifiability? Parameter identifiability is a fundamental concept from information theory that refers to the degree to which one can accurately estimate a given model's parameters from input-output data. There is substantial existing research in the literature on battery parameter identifiability. An important lesson from this literature is that the design of a battery experiment can affect parameter identifiability significantly. Hence, test trajectory optimization has the potential to substantially improve model parameter identifiability. The literature explores this lesson for equivalent circuit and physics-based battery models. However, there is a noticeable gap in the literature regarding identifiability analysis and optimization for both machine learning-based and hybrid battery models.

To address the above gap, the dissertation makes four novel contributions to the literature. The first contribution is an extensive survey of the machine learning-based battery modeling literature, highlighting the critical need for information-rich and clean datasets for parameterizing data-driven battery models. The second contribution is a K-means clustering-based algorithm for detecting outlier patterns in experimental battery cycling data. This algorithm is used for pre-cleaning the experimental cycling datasets for laboratory-fabricated lithium-sulfur (Li-S) batteries, thereby enabling the higher-fidelity fitting of a neural network model to these datasets. The third contribution is a novel algorithm for optimizing the cycling of a lithium iron phosphate (LFP) to maximize the parameter identifiability of a hybrid model of this battery. This algorithm succeeds in improving the resulting model's Fisher identifiability significantly in simulation. The final contribution focuses on the application of such test trajectory optimization to the experimental cycling of commercial LFP cells. This work shows that test trajectory optimization is effective not just at improving parameter identifiability, but also at probing and uncovering higher-order battery dynamics not incorporated in the initial baseline model. Collectively, all four of these contributions show the degree to which the selection of battery cycling datasets and experiments for richness and cleanness can enable higher-fidelity data-driven and hybrid modeling, for multiple battery chemistries.