Reinforcement Learning of Optimal Input Excitation for Parameter Estimation with Application to Li-ion Battery

Rui Huang, Jackson Fogelquist, and Xinfan Lin, *Senior Member, IEEE*

***Abstract*— Estimation and diagnostics of system states and parameters is ubiquitous in industrial applications. Estimation is often performed using input and output data, and the quality of input excitation has critical impact on the accuracy of the results. Therefore, optimal input excitation design has been receiving increasing research attention. Previously, input design is formulated as an optimization problem to find a sequence of excitation, which maximizes a certain criterion associated with estimation accuracy, e.g., the information content of the data. However, the practice suffers from several major drawbacks, including the susceptibility to uncertainty (especially that in target parameter) and tractability of solution. In this research, a reinforcement learning (RL) framework is proposed as a new approach for input design. We envision the input generation procedure as a Markov Decision Process, and leverage reinforcement learning to learn an optimal policy for generating the input excitation. The new approach improves the robustness of the generated input sequence through the feedback mechanism of the policy, and tractability through the learning mechanism of RL. The methodology is applied to optimal excitation design for estimating critical lithium-ion battery electrochemical parameters in simulation and experiments. Results show that the new RL-based framework significantly outperforms the conventional direct optimization approach (with one order-of-magnitude higher information level) under the presence of uncertainty in the target parameter for estimation, and achieves substantially smaller estimation error compared with other profiles in experiment. The obtained RL policy could be used for battery health diagnostics and testing of second-life batteries for repurposing applications.**

***Index Terms*— lithium-ion battery, parameter estimation, information content of data, excitation optimization, reinforcement learning.**

I. Introduction

ESTIMATION of unknown system internal variables, such as states and parameters, is an important topic in modeling and control research. Model parameter estimation, in particular, is critical for system monitoring, diagnostics and control, as the accuracy of parameters is essential to guarantee the fidelity of the model and the effectiveness of the model-based practice of the above functionalities. For example, in the area of battery management and control [1], [2], parameter estimation has been widely applied to state of health monitoring [3], [4], and model calibration [5]–[7] to enable real-time state estimation [8], [9] and optimal control [10], [11] among others. The practice of estimation typically determines the values of parameters by fitting a model to the measured input-output data using a certain algorithm. The topic has been studied extensively with the primary focus on the model development and algorithm design, while the importance of data has been less investigated [2]. However, the quality of data has a significant impact on parameter estimation, as data containing low information about the target parameter may impose a fundamental limit on the estimation accuracy [12], while data with undesirable structures could substantially amplify the errors caused by system uncertainties [13].

The critical role of data has motivated the research on the design of input excitation to generate the optimal data for estimation. In statistics, input excitation design is a particular case of optimal experimental design (OED), which aims at determining experiment conditions/settings that maximizes a certain criterion related to the data quality. Commonly used criteria in OED include information content of the data, e.g., Fisher information (FI) [14], and information gain by using data for estimation, such as the Kullback-Leibler divergence [15], [16] among others. One challenge with input excitation design is the large number of optimization variables (typically > 1000) due to the long input sequence. Take the field of battery management and control as an example, FI (or its variant) is commonly used in input optimization for battery parameter estimation due to its more favorable (yet still high) computational complexity compared with other criteria. Existing research in this area includes early works on the battery equivalent circuit model [7], [17] and later ones on electrochemical model [5], [18]. These works considered input current excitation subject to a certain pattern, such as sinusoidal, pulse, and (piece-wise) constant current, and optimized the features of the patterns, e.g., frequency and magnitude. Alternative objective such as the global sensitivity index has also been considered [18]. Recent works have explored optimal current excitation with no imposed pattern for estimating battery electrochemical parameters [19]–[21], enabled by efficient sensitivity computation technique [20]. Similar study was performed in [19], with sensitivity computed by perturbation and input sequence divided into smaller segments to facilitate computation. Most recently, the study was further extended to

This work is supported by the NSF CAREER Program (Grant No.2046292) and the NASA HOME Space Technology Research Institute (Grant No.80NSSC19K1052).

R. Huang, J. Fogelquist and X. Lin are with the Department of Mechanical and Aerospace Engineering, University of California, Davis, Davis, CA 95616 USA (e-mail: rhhuang, jbfogelquist, lxflin@ucdavis.edu).

design inputs that could mitigate the propagation of system uncertainties to the estimation error [13].

In the state of art, input excitation design is typically formulated as an optimization problem to determine an optimal input sequence based on the selected design criterion and the underlying system dynamics. Such approach, however, suffers from several notable drawbacks. First, the practice faces an intrinsic dilemma regarding the values of the target parameter to be estimated. Specifically, the design criterion, e.g., FI, is typically dependent on system parameters, including the target parameter, whose value is the goal of estimation and remains unknown during the input design stage. As a result, a certain $a\,priori$ value needs to be assumed, which would inevitably deviate from the true value. In fact, a small deviation of the $a\,priori$ parameter value could lead to input sequences that are far from optimal under the actual parameter value, which will be demonstrated with an example later in this paper. Fundamentally, this reflects the susceptibility of input (time) sequence optimization to system uncertainties and disturbances. Second, the optimization procedures are computationally intensive, involving iterative computation and inversion of the design criterion and system dynamics, as well as their Jacobian/gradient. This issue is especially prominent for optimization with no imposed pattern, where the optimization variables consist of the input at each instant of the time sequence, yielding a high dimension. For instance, in order to make the optimization tractable for battery electrochemical models (single particle model [20] or reduced order Doyle-Fuller-Newman (DFN) model [21]), significant efforts have been made to simplify and reformulate the model and the problem. Nevertheless, extension to more complicated full-order battery models still appears to be intractable [5]. In addition, global minimum/optimality is always difficult to guarantee for conventional gradient-driven optimization, especially for the highly non-convex input design problems involving nonlinear models and objectives.

To overcome the limitations of the conventional approach, in this paper, input excitation design is formulated as an optimal control problem with the goal of finding a closed-loop control policy for input excitation generation. We envision input generation as a Markov Decision Process (MDP) with dynamics characterized by certain state(s), which are directly related to battery physics. The objective is then to derive an optimal control policy that can maximize the design criterion, e.g., FI, over the considered time horizon. The reinforcement learning method is used to learn the policy, which is guided by rewards specified based on the design criterion. The contributions and advantages of the proposed new method include the following aspects. First, the derived control policy could significantly reduce the impact of system uncertainties on the performance of the designed input excitation. At each time instant, the policy generates an input (current) action based on certain states of the system. Compared with the conventional practice of simply giving a time sequence of inputs at the beginning [20], the feedback of states in real time would correct for the deviation caused by system uncertainties and disturbances. A recent work also leveraged RL to generate current to improve the identifiability of battery stoichiometric parameters [22]. However, instead of using battery physical states, they adopted the past current and voltage trajectories as the RL states, yielding a very high dimensional state space (200 total for both policy and value functions). This leads to substantially higher computational and memory load for training and implementation of the policy, compared to our approach where only 1-2 states are needed for RL. In addition, there also lacks the benefit of using physical state feedback to counter uncertainties. We will demonstrate a case where an input control policy informed by the state estimates of an observer could generate an input control sequence that is robust to the deviation in the assumed $a\,priori$ value of the target parameter for estimation, hence resolving the aforementioned dilemma facing the conventional approach. Second, the use of reinforcement learning could also substantially improve the tractability of input excitation design. Specifically, the policy is learned using the reward generated at each time instant, and the procedures only involve the forward computation of the rewards and states based on the model. The associated computational complexity is much more favorable than both the direct optimization of the time sequence and the traditional optimal control, which essentially needs to not only compute but also invert the system dynamics and the Jacobians of the design objective. Therefore, the reinforcement learning-based approach provides a promising solution to enable input optimization for nonlinear systems with complicated and high-dimensional dynamics. The proposed framework is applied to excitation design for the single-variate estimation of critical lithium-ion battery electrochemical parameters in simulation and experiments, showing one order-of-magnitude higher FI under the presence of uncertainty in the target parameter for estimation, compared with the sequence obtained by conventional direct optimization. Detailed analysis is provided to explain the improvement enabled by the proposed framework. Experimental results are presented to validate the accuracy of estimation using the optimized input excitation. To the best of the authors' knowledge, this research is the first to explore input optimization for system parameter estimation by combining reinforcement learning with system dynamics.

II. Input Excitation Design

In this section, the input design problem is first formulated by using FI as the criterion. Under this formulation, we introduce the general procedure of the conventional input excitation design and, more importantly, the proposed reinforcement learning-based excitation optimization framework.

A. Design Criterion

Consider a single-input-single-output discrete-time system with dynamics modeled as

$$\begin{aligned} \boldsymbol{x}_k &= f_k(\boldsymbol{x}_{k-1}, \theta, u_{k-1}) \\ y_k &= g_k(\boldsymbol{x}_k, \theta, u_k), \end{aligned} \tag{1}$$

in which $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$ are the states of the model, and u and y are the scalar input and output of the model respectively. In this work, we consider the estimation of a single parameter θ of the system. The output sensitivity of

TABLE I
NOMENCLATURE

a	Action of RL agent
A	Electrode area
c_e	Electrolyte lithium concentration
c_{se}	Electrode particle surface lithium concentration
D_s	Electrode lithium diffusion coefficient
F	Faraday constant
F_{info}	Fisher information
I	Current
r	Reward of RL
R_l	Lumped battery ohmic resistance
R_s	Radius of electrode particle
S	RL states
u	System input
U	Open circuit potential
V	Battery voltage
x	System physical states
y	System output
α	Learning rate of RL
β	Electrode lithium stoichiometry
γ	Decaying factor of RL reward
δ	Electrode thickness
ϵ	Greedy factor of RL exploration
ε_s	Active material volume fraction
η	Overpotential
θ	Model parameters
σ^2	Measurement noise variance
ϕ_e	Electrolyte potential

the parameter θ is the partial derivative of the output y to θ, $\frac{\partial y}{\partial \theta}$, which reflects the impact of parameter change on the output variation. High sensitivity indicates strong correlation between y and θ, and hence the easiness of detecting parameter change from output measurement. The output sensitivity could also formally define the quality of data through the FI [14], which is a metric to evaluate the information content about θ contained in a data series $y_1, \cdots, y_k, \cdots, y_N$ measured over time. Under i.i.d. Gaussian output measurement noise, FI takes the form [23]

$$F_{info} = \frac{1}{\sigma_y^2} \sum_{k=1}^{N} \left(\frac{\partial y_k}{\partial \theta} \right)^2, \tag{2}$$

where σ_y^2 is the variance of the measurement error in y. FI is directly related to the accuracy of the estimation results. Specifically, the inverse of FI gives the Cramér-Rao bound, which quantifies a lower bound of the estimation error variance of an unbiased estimator of θ [24],

$$\sigma^2(\hat{\theta}) \geq F_{info}^{-1}. \tag{3}$$

Due to its correlation with data quality and estimation accuracy, FI has been frequently adopted as the "gold standard" for the optimization of input excitation.

B. Conventional Approach for Input Excitation Design

The conventional approach aims at designing the input excitation sequence directly by maximizing the FI

$$\min_{[u_1, u_2, \ldots u_N]} - \sum_{k=1}^{N} \left(\frac{\partial y_k}{\partial \theta} \right)^2, \tag{4}$$

subject to the constraints imposed by system dynamics, and the bounds on state variables, input and output. The variables to be optimized are the series of inputs u_k's applied to the system over discrete time instances k's, which will change the state trajectory and parameter sensitivity, and hence the FI (to be discussed in more details in the next section). The formulated excitation optimization problem needs to be solved by nonlinear optimization methods, such as in the practice of lithium-ion battery parameter estimation [13], [20]. However, as discussed previously, there are several major drawbacks with the conventional approach. First, in order to maximize the objective (4), an $a\,priori$ value of the target parameter θ needs to be assumed, as the FI typically depends on θ. A small deviation of the $a\,priori$ value could significantly affect the optimality of the designed excitation sequence when applied to the actual system with a different parameter value. An example will be provided in Section IV for illustration. In addition, conventional nonlinear optimization approaches, especially the gradient-based algorithms, are computationally intensive due to the need of iterative evaluation and inversion of system dynamics and Jacobians during the optimization procedure. This is especially challenging for the input optimization problem, where the number of optimization variables, i.e., input instances, is typically very large ($>$1000). Consequently, the practice would be intractable for complicated systems with high order, nonlinearity, and complexity. The drawbacks of the conventional approach inspires the exploration for a new methodology for input excitation optimization in this work.

C. Formulation of Reinforcement Learning Framework

As an emerging alternative to the conventional optimization approach, a RL-based optimization framework is proposed that can offer a fundamentally different and innovative way to address the drawbacks and challenges facing the former. In this framework, the generation of input excitation is considered as a Markov Decision Process (MDP) with dynamics described by certain state(s) S, which can be (some of) the physical states of the system or information states related to the estimation/learning process, e.g., sensitivity. The goal is then to find an optimal policy π that generates the optimal action a^* at every time step: $a^* = \pi(S)$, to maximize the objective of the excitation design over the whole data sequence, e.g., FI. Reinforcement learning learns the policy π by exploring the state-action space to solve the Bellman Equation [25], in the form of which most optimal control problems can be formulated. In this work, the classic Q-learning scheme is adopted as the basic RL algorithm as a demonstration due to its simplicity [26], while other RL algorithms can be adopted alternatively under the framework. Specifically, a state-action value function, known as the Q function $Q(S, a)$, is used to represent the maximum expected total return of the sequence when taking action a at state S, where the expected return is the objective of input generation (e.g., maximal FI). The purpose of learning is to train the accurate Q function. During the training phase, learning is performed in episodes, which consist of a series of time steps. At each time step k, an action a_k is taken with the state at S_k, and the Q function is updated

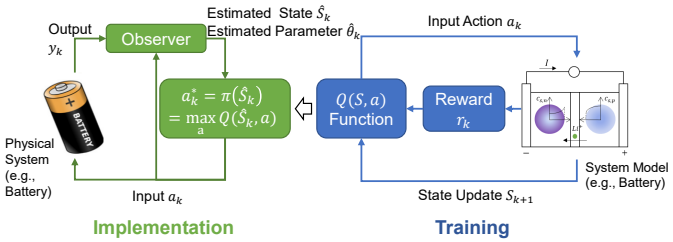


Fig. 1. Proposed RL-based excitation optimization and generation framework

according to

$$Q(S_k, a_k) \leftarrow Q(S_k, a_k) + \alpha[r_k + \gamma \max_a Q(S_{k+1}, a) - Q(S_k, a_k)], \quad (5)$$

which is derived from the Bellman Equation [26]. In (5), r_k is the reward that is directly related to the learning objective. In the case of FI in (2), which is the sum of the squared output sensitivity over time steps, the single-step reward r_k can be conveniently formulated as:

$$r_k = \begin{cases} (\frac{\partial y_k}{\partial \theta})^2, & \text{normal condition} \\ \text{negative penalty}, & \text{constraint violation} \end{cases} \quad (6)$$

in which a negative reward (penalty) is applied when the constraints on states and output are violated to enforce the safety limits. Additionally, regarding the updating rule in (5), α is the learning rate governing the learning speed, and γ is the decaying factor accounting for the discounting effect of rewards over time. Meanwhile, during training, the action a_k is chosen based on the ϵ-greedy policy in (7), which has a probability of $1-\epsilon$ to take the action that exploits the maximum return based on the Q function, and a probability of ϵ to explore a random action for learning purpose:

$$a_k \leftarrow \begin{cases} \text{argmax}_a\, Q(S_k, a) & \text{with probability } 1-\epsilon \\ \text{a random action} & \text{with probability } \epsilon \end{cases}. \quad (7)$$

After the training is finished, the obtained Q function is used to generate the excitation in the implementation stage, i.e., $a_k^* = \pi(S_k) = \text{argmax}_a\, Q(S_k, a)$ to maximize the return, based on the feedback of the states or state estimates by an observer. The schematic of the RL-based excitation optimization and generation framework is shown in Fig. 1. In the context of input design for battery parameter estimation, the action a is the current I, and a reduced-order battery electrochemical model is used to compute the state transition and the reward (output sensitivity) for reinforcement learning. The state S is chosen as a certain physical battery state and/or information state, e.g., sensitivity. A brief discussion of the model and battery physics will be included in the next section.

III. Battery Model and Sensitivity Computation

This section first introduces the battery model used for optimization and reinforcement learning. The computation of parameter sensitivity will be discussed next, which is needed for evaluating the FI as the objective of input optimization and the reward for reinforcement learning.

A. SPMe Battery Model

A reduced-order electrochemical model, i.e., the single particle model with electrolyte dynamics (SPMe) [8], is used for describing battery physics in this work. The model captures the key electrochemical processes occurring in the anode, separator, cathode, and electrolyte of the battery during operation, including the diffusion of lithium ions in the electrode and the electrolyte, and the intercalation/deintercalation kinetics of lithium ion between the electrode and the electrolyte. The model is a simplification of the full-order DFN model based on the single particle assumption, which assumes uniform current density across the electrode and uses one single particle in each electrode to represent the storage and transfer of lithium ion. The SPMe offers adequate balance of model fidelity and complexity, and therefore has been widely used in battery control research [13], [18], [19], [27].

In the context of SPMe, the output is the voltage, and input is the current, i.e., $y = V$ and $u = I$. The terminal voltage V consists of 4 different parts as described in (8),

$$\begin{aligned} V =& (U_p(c_{se,p}) - U_n(c_{se,n})) + (\phi_{e,p}(c_{e,p}) - \phi_{e,n}(c_{e,n})) \\ &+ (\eta_p(c_{se,p}, c_{e,p}) - \eta_n(c_{se,n}, c_{e,n})) - IR_l, \end{aligned} \quad (8)$$

where the subscripts p and n denote the positive and negative electrode respectively. The first part is the difference between the two electrodes in the open circuit potential (OCP) U, which is a nonlinear function of the particle surface lithium concentration c_{se} and reflects the equilibrium potential of the electrodes. The second part represents the potential difference in electrolyte (ϕ_e) due to the gradient in electrolyte lithium concentration c_e caused by diffusion. The third part is the difference in the overpotential η at the electrode particle surface and electrolyte interface, which drives the lithium flux at the particle surface due to the intercalation/deintercalation reaction. Finally, R_l accounts for the lumped ohmic resistance, including that of the electrolyte, SEI film, and current collectors. The model further includes diffusion equations to describe the evolution of c_{se} and c_e, and a nonlinear Butler-Vollmer equation to capture η. Procedures including Laplace transform and Padé approximation have been applied to convert the original diffusion equations in the form of PDE to a low-order rational transfer function to facilitate control and optimization. For example, the transfer function for c_{se} from the input current I is obtained as

$$c_{se}(s) = -\frac{7R_s^4 s^2 + 420 D_s R_s^2 s + 3465 D_s^2}{s(R_s^4 s^2 + 189 D_s R_s^2 s + 3465 D_s^2)} \cdot \frac{I(s)}{F \varepsilon_s A \delta}, \quad (9)$$

where R_s is the radius of the electrode particle, D_s is the diffusion coefficient, F is the Faraday constant, ε_s is the active material volume fraction, and A and δ are the area and thickness of the electrode respectively. A detailed description of the model can be found in [28].

B. Parameter Sensitivity Computation

Based on SPMe, a methodology has been derived to efficiently calculate the sensitivity of battery parameters [28]. Specifically, the sensitivity of the battery voltage to a certain parameter can be obtained by applying the chain rule of

differentiation to the voltage equation in (8). In this work, the parameters targeted for input excitation optimization are ε_s and D_s respectively, which are the key parameters related to the battery state of health [3], [29], [30]. Among them, ε_s represents the portion of the electrode material that could actively store and transport Li-ion, and is an electrode parameter affecting c_{se}. Therefore, the sensitivity of ε_s is

$$\frac{\partial V}{\partial \varepsilon_s}(t) = \frac{\partial \eta}{\partial \varepsilon_s} + \left(\frac{\partial \eta}{\partial c_{se}} + \frac{\partial U}{\partial c_{se}}\right) \cdot \frac{\partial c_{se}}{\partial \varepsilon_s}(t), \tag{10}$$

where $\frac{\partial \eta}{\partial \varepsilon_s}$, $\frac{\partial \eta}{\partial c_{se}}$, and $\frac{\partial U}{\partial c_{se}}$ are nonlinear coefficients dependent on current I that can be easily obtained based on the model. Meanwhile, $\frac{\partial c_{se}}{\partial \varepsilon_s}$ is dynamic and will change over time even under constant input, posing major difficulties for sensitivity calculation. To characterize the dynamic nature of this term, a sensitivity transfer function (STF) can be derived by taking the derivative of c_{se} in (9) to ε_s,

$$\frac{\partial c_{se}}{\partial \varepsilon_s}(s) = \frac{7R_s^4 s^2 + 420 D_s R_s^2 s + 3465 D_s^2}{s(R_s^4 s^2 + 189 D_s R_s^2 s + 3465 D_s^2)} \cdot \frac{I(s)}{F \varepsilon_s^2 A \delta}. \tag{11}$$

The obtained STF can be easily implemented in time domain, e.g., through the state space representation, and used to compute the sensitivity under arbitrary input or for optimization of input [13], [20]. Same approach can be applied to compute the sensitivity for other parameters. For example, the sensitivity of D_s can be derived as

$$\frac{\partial V}{\partial D_s}(t) = \left(\frac{\partial \eta}{\partial c_{se}} + \frac{\partial U}{\partial c_{se}}\right) \cdot \frac{\partial c_{se}}{\partial D_s}(t), \tag{12}$$

with a STF

$$\frac{\partial c_{se}}{\partial D_s}(s) = \frac{43R_s^4 s^2 + 1980 D_s R_s^2 s + 38115 D_s^2}{(R_s^4 s^2 + 189 D_s R_s^2 s + 3465 D_s^2)^2} \cdot \frac{21 R_s^2 I(s)}{F \varepsilon_s A \delta}. \tag{13}$$

IV. Optimization Results and Simulation Verification

In this section, the current excitations optimized for estimating ε_s and D_s of the cathode, i.e., $\varepsilon_{s,p}$ and $D_{s,p}$, using the proposed RL-based approach are presented and analyzed. The optimization task aims at maximizing the FI of the target parameters with a battery starting SOC of 0.5 over a 1800s time horizon. Comparison to the results of conventional optimization is also discussed to demonstrate the efficacy and robustness of the new approach.

For the RL-based approach, each training episode starts at 0.5 SOC and terminates after 1800 steps or violating constraints, with a step size of 1 second. Training has been attempted with various number of episodes ranging from 0.1 to 1 million. A single particle battery model with parameters adopted from [31] is used for data generation and simulation study. To account for the physical constraints of the considered battery, the current and voltage are bounded according to the specifications. The input/output ranges, penalty for constraint violation, and algorithmic hyper-parameters of RL are shown in Table II. The negative penalty is set by analyzing the scale of the reward $(\frac{\partial y_k}{\partial \theta})^2$ and tuning in sumulation trials. Noted that the exploration probability ϵ decays linearly over episodes to prioritize exploration at the beginning and exploitation at the end. For $\varepsilon_{s,p}$, the RL state S is the battery state of charge (SOC) defined based on the cathode particle surface concentration $c_{se,p}$,

TABLE II
Setup and Hyperparameters of RL

Input Range	[-77A,77A]	
Output Range	[2.75V,4.4V]	
Penalty for Constraint Violation	-1	
Learning Rate α	0.5	
Decaying Factor γ	0.999	
Greedy Factor ϵ	[1,0]	
RL States $S(k)$	For $\varepsilon_{s,p}$	For $D_{s,p}$
	$\mathrm{SOC}(k)$	$[\mathrm{SOC}(k), \frac{\partial c_{se}}{\partial D_s}(k)]$

$$\mathrm{SOC} = \frac{\beta - \beta_{0\%}}{\beta_{100\%} - \beta_{0\%}}, \tag{14}$$

where β is the lithium stoichiometry ratio of $c_{se,p}$ to the maximum allowed lithium concentration $c_{s,p}^{\max}$ by the cathode material, i.e., $\beta = \frac{c_{se,p}}{c_{s,p}^{\max}}$, and $\beta_{0\%}$ and $\beta_{100\%}$ are the lithium stoichiometry at 0% and 100% SOC. The state and input are digitized to discrete values to index the Q function. Specifically, the state SOC is discretized to 101 intervals between 0% and 100%, and the input current is discretized to 45 between –77 A and 77 A (-3 C and 3 C) respectively. The states S for $D_{s,p}$ learning will be discussed later.

As benchmark, the conventional direct optimization is implemented using the General Purpose OPtimal Control Software (GPOPS) package under same settings.

A. Analysis of Nominal Optimization Results

We will first evaluate the optimization results of the new and conventional approaches under the nominal conditions. Specifically, we will demonstrate the optimized current excitation, analyze and explain the observed patterns, and compare the FI of the obtained policy/sequence under the nominal training/optimization conditions.

1) Optimization Results for $\varepsilon_{s,p}$: The current excitation for estimating $\varepsilon_{s,p}$ generated by using the RL-trained policy and that optimized by the conventional approach, along with the resultant battery voltage V and SOC evolution, are shown in Fig. 2.

As shown in Fig. 2-1 and 2-2, similar patterns in current excitation can be observed between the two approaches. Qualitatively, both current profiles consist of three stages as labeled. The first one is high-current charging, during which the battery SOC rises rapidly. The second stage is reduced-current charging after the voltage limit is reached, featured by decreased terminal voltage and near-constant SOC. The last stage is the SOC-sustaining pulse current, where SOC maintains at around 0.78 under alternating charging and discharging pulses. Explanation can be provided for the presented excitation pattern by correlating to physics underlying the dynamic equations that govern the parameter sensitivity, which is related to the reward/objective of learning/optimization. As shown in (10), the output voltage sensitivity of ε_s is

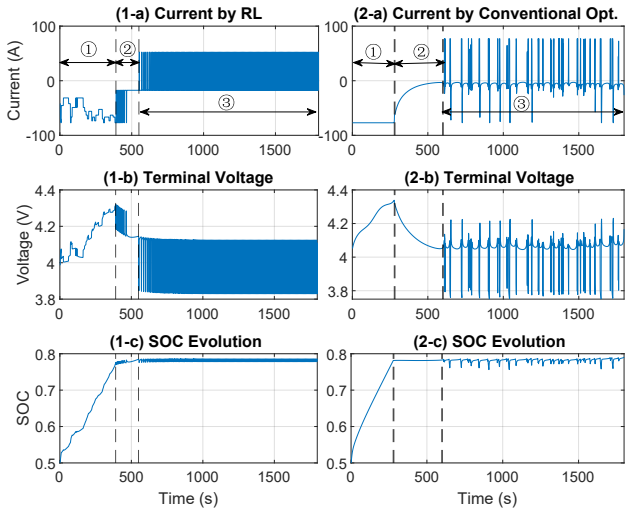


Fig. 2. Optimized current, resultant voltage and SOC for $\boldsymbol{\varepsilon_{s,p}}$ by RL (1-a,b,c) and conventional approach (2-a,b,c)

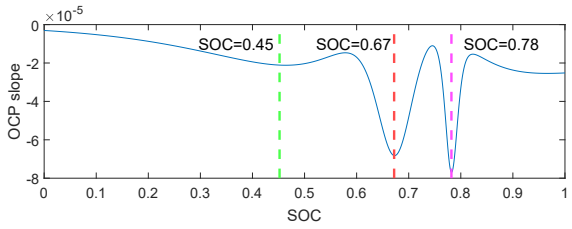


Fig. 3. Slope of cathode OCP curve versus SOC

comprised by the non-dynamic term $\frac{\partial \eta}{\partial \varepsilon_s}$ and the dynamic term $(\frac{\partial \eta}{\partial c_{se}} + \frac{\partial U}{\partial c_{se}}) \cdot \frac{\partial c_{se}}{\partial \varepsilon_s}$, where $\frac{\partial U}{\partial c_{se}} \cdot \frac{\partial c_{se}}{\partial \varepsilon_s}$ is the dominant part dependent on both the slope of cathode open circuit potential $\frac{\partial U}{\partial c_{se}}$ and the state sensitivity $\frac{\partial c_{se}}{\partial \varepsilon_s}$. The slope of cathode OCP with respect to SOC is plotted in Fig. 3, which achieves the maximum in magnitude at SOC $= 0.78$. For $\frac{\partial c_{se}}{\partial \varepsilon_s}$, as characterized by (11), its transfer function features a pole at $s = 0$, representing an integrator $1/s$. Therefore, its magnitude will increase under continuous charging/discharging current, and remain steady under zero or alternating excitation. As a result, intuitively the optimal current excitation is to charge the battery and accumulate $\frac{\partial c_{se}}{\partial \varepsilon_s}$ until the SOC reaches the highest OCP slope, i.e., SOC $= 0.78$, to yield the highest $\frac{\partial U}{\partial c_{se}} \cdot \frac{\partial c_{se}}{\partial \varepsilon_s}$, and then stay around the SOC to maintain high sensitivity, which is as exactly the three-stage pattern of both the RL-based and conventional optimization results. It is noted that the pulse in the last stage marginally contributes to the output sensitivity by increasing the non-dynamic term $\frac{\partial \eta}{\partial \varepsilon_s}$ while keeping the SOC around 0.78.

For quantitative comparison, as shown in Table III, the optimal current learned by RL yields a normalized FI of 282, which is very close to that of the conventional method at 300. The learning curve of RL is presented in Fig. 4, with two subplots showing the evolution of the return per episode and the value of one Q table entry $Q(50, 22)$, i.e., the expected total return when taking the $22nd$ action ($I = -3.5$A) at the $50th$ state (SOC $= 0.49$) as an example. It is seen that the return shows a rising trend over episodes until the

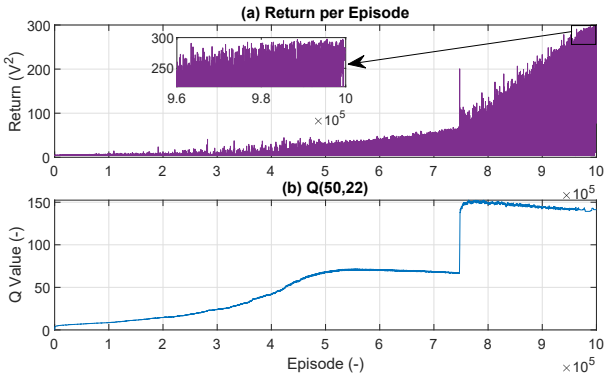


Fig. 4. Learning curve of RL with (a) the evolution of return per episode and (b) the evolution of $\boldsymbol{Q}$(50,22) over episodes

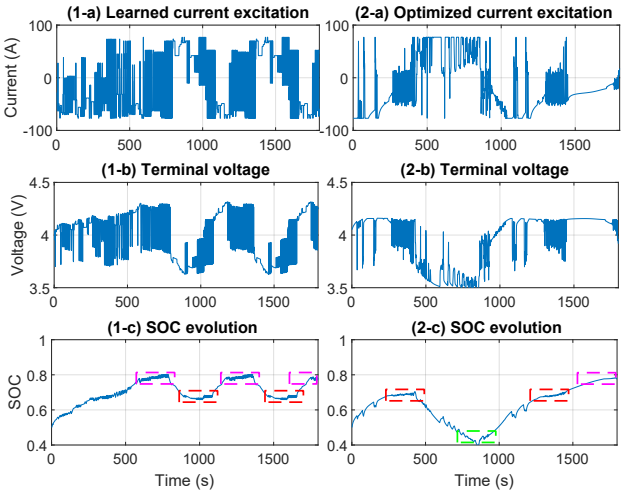


Fig. 5. Optimized current, resultant voltage and SOC for $\boldsymbol{D_{s,p}}$ by RL (1-a,b,c) and conventional approach (2-a,b,c)

end due to the exploration for the optimal policy at the beginning with high exploration rate ϵ, and the exploitation of the refined policy at the end as ϵ is gradually reduced to 0. The convergence of the RL policy is further demonstrated by the evolution of the Q table entry, represented by $Q(50, 22)$, which reaches a constant value over training process. Both qualitative/quantitative analyses and the convergence of the learning curve show that the proposed RL-based approach is capable of finding the optimal excitation.

2) Optimization Results for $\mathrm{D_{s,p}}$*:* The excitation optimized for the estimation of another battery electrochemical parameter, i.e., cathode diffusion coefficient $D_{s,p}$, is also explored using the proposed RL-based approach. The RL states S consist of two terms. The first one is the battery SOC, same as in the case of $\varepsilon_{s,p}$, and the second one is the sensitivity of the diffusion state c_{se}, i.e., $\frac{\partial c_{se}}{\partial D_s}$, representing the information state. The inclusion of the second state is because of the non-unique mapping from SOC alone to the optimal policy due to the critical role of $\frac{\partial c_{se}}{\partial D_s}$ in sensitivity dynamics, as detailed in the subsequent analysis. The optimized current excitation, resulting terminal voltage and battery SOC of both the RL-based and the conventional approach are shown in Fig. 5.

Similar patterns in current can be observed between the two

methods, which is most evident in the SOC evolution shown in (1-c) and (2-c) of Fig. 5. Both the excitations from the RL-trained policy and conventional optimization feature multiple stages of alternating charging/discharging current, during which the SOC swings between different levels, as marked by color boxes, corresponding to different peak locations of OCP slope labeled in Fig. 3. Analysis can be performed by examining the dynamics of the sensitivity of D_s to explain the patterns of the excitation and the difference from the case of ε_s. The output sensitivity of D_s shown in (12) is dominated by $\frac{\partial U}{\partial c_{se}} \cdot \frac{\partial c_{se}}{\partial D_s}$, which is the product of cathode OCP slope $\frac{\partial U}{\partial c_{se}}$ and the state sensitivity $\frac{\partial c_{se}}{\partial D_s}$. The dynamics of $\frac{\partial c_{se}}{\partial D_s}$, of which the STF is given in (13), are stable with all negative poles. As a result, the state sensitivity will die out under zero excitation, and persistent input excitation is needed to maintain a certain level of state sensitivity. Therefore, in order to achieve high OCP slope and state sensitivity concurrently to maximize the sensitivity and FI over the data sequence, the input policy should apply continuous current over the time horizon to drive the battery SOC to different peak locations of the OCP slope, which explain the difference from the patterns of ε_s whose STF has an integrator. Regarding the quantitative performance, the RL policy is able to yield a normalized FI of 1.12, which is better than that of the conventional approach at 0.831, showing that the RL-based approach is able to generate qualitative similar but quantitatively even better performance compared with the direct optimization counterpart. The reason for the better performance of RL is that it manages to consistently drive the SOC to the OCP peak with the largest magnitude (marked with the magenta dashed box), while the conventional approach sometimes drives the SOC to the smallest peak (marked with the green dashed box).

B. Performance under Uncertainty in Parameter

In order to evaluate the performance of the optimized current excitation in practical scenario, another simulation was conducted under the uncertainty in the target and other parameters. The case of estimating $\varepsilon_{s,p}$ is shown here for illustration. Specifically, the previous RL policy and open-loop sequence trained/optimized for $\varepsilon_{s,p}$ under the nominal $\varepsilon_{s,p} = 0.5$ and $D_{s,p} = 10^{-13}$ are applied to a battery (simulated by SPMe) with actual $\varepsilon_{s,p}$ = 0.4 and $D_{s,p} = 1.2 \times 10^{-13}$, both with 20% deviation (uncertainty) from the nominal values. It is noted that the 20% variation of $\varepsilon_{s,p}$ reflects the same amount of deviation in the cathode capacity, which can be caused by either degradation or manufacturing variability. This mismatch emulates the aforementioned uncertainty in the target and other parameters, which are not known exactly at the input optimization phase. For the conventional optimization approach, the excitation sequence is applied directly, while the RL-trained policy generates the excitation based on the feedback of battery SOC as shown in the implementation part of Fig. 1. The battery SOC is estimated based on the SPMe and the current and voltage measurement using an Extended Kalman Filter, which is a standard practice in battery management [9], [32].

The resultant FI of $\varepsilon_{s,p}$ is summarized in Table III. It is shown that the RL-trained policy gives a FI of 284 when evaluated at the actual $\varepsilon_{s,p} = 0.4$ and $D_{s,p} = 1.2 \times 10^{-13}$, which is close to that under the nominal $\varepsilon_{s,p}$ and $D_{s,p}$, while the FI of the open-loop sequence drops substantially to only 51. In addition, we further applied the optimized excitations to batteries with a series of $\varepsilon_{s,p}$ values in simulation, ranging from 0.4 to 0.6 (-20% to +20% around the nominal value). The resultant FI of the excitations obtained by the two approaches is shown in Fig. 6. It is seen that the conventional approach yields slightly higher FI when the actual $\varepsilon_{s,p}$ is close to the nominal $\varepsilon_{s,p} = 0.5$. However, the FI drops significantly when the actual $\varepsilon_{s,p}$ deviates from the nominal. The RL-based method, on the other hand, maintains a high FI over the whole range of actual $\varepsilon_{s,p}$. The significant difference in performance can be explained by the SOC evolution under the two excitations, shown in Fig. 7, for the case when applying to estimate $\varepsilon_{s,p} = 0.4$. It is seen that the RL-trained policy still manages to drive the SOC to the desired peak of the OCP slope, i.e., SOC $= 0.78$, while the open-loop sequence reaches SOC $= 0.85$ instead, where the OCP slope is much smaller according to Fig. 3. The SOC deviation of the latter is caused by the fact that the open-loop sequence simply applies the pre-determined current values without referring to the actual battery states, especially the SOC which is important for maximizing sensitivity/FI. The battery SOC is proportional to the stoichiometry ratio β as shown in (14), which is defined based on the particle surface concentration c_{se}. According to (9), the transfer function of c_{se} has a pole at $s = 0$, representing an integrator, and is (inversely) proportional to ε_s. Therefore, during optimization, SOC is projected to reach 0.78 under the nominal $\varepsilon_{s,p}$ driven by the designed current sequence. However, when applied to battery with a different $\varepsilon_{s,p}$, the actual SOC will drift, causing the miss of the OCP slope peak and substantial reduction in FI. The RL policy, on the contrary, establishes a mapping between the SOC (state) and current (action), and uses the estimate of the actual SOC as feedback to inform the generation of current excitation. Therefore, the policy is valid for different values of actual $\varepsilon_{s,p}$, as long as the estimated SOC is accurate. By using a closed-loop observer with voltage as feedback, the impact of the imprecise knowledge of $\varepsilon_{s,p}$ in the model can be mitigated to yield accurate estimate of SOC [33].

The advantage of the RL-based approach is further demonstrated by applying the designed input excitation to parameter estimation in simulation. Specifically, estimation is conducted using the designed current and the generated voltage output based on the least squares method. To emulate the practical measurement uncertainties, the current excitation data are added with a Gaussian white noise of 0 mean and 0.1A standard deviation, and the voltage output is injected with a Gaussian white noise of 0.01 V mean bias and 0.01V standard deviation. The resultant $\varepsilon_{s,p}$ estimation errors are shown in Table III. It is seen that under the nominal conditions with no uncertainty in $\varepsilon_{s,p}$ and $D_{s,p}$, the RL and the conventional approaches yield similar errors, i.e., -2.50% vs. -2.38%, which is in accordance with their similar FI. However, when subject to the 20% uncertainty in $\varepsilon_{s,p}$ and $D_{s,p}$, the estimation error by RL slightly increases to -2.95%, while that of the

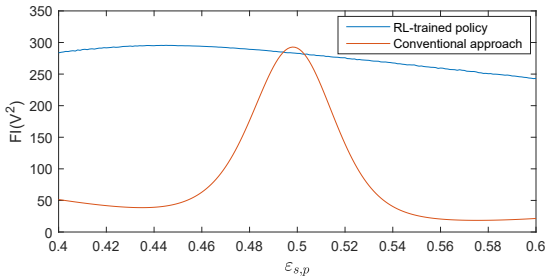


Fig. 6. FI of RL policy and sequence optimized by conventional method (under nominal $\boldsymbol{\varepsilon_{s,p} = 0.5}$, $\boldsymbol{D_{s,p} = 10^{-13}}$) applied to batteries with $\boldsymbol{\varepsilon_{s,p}}$ from 0.4 to 0.6 and $\boldsymbol{D_{s,p} = 1.2 \times 10^{-13}}$ in Simulation

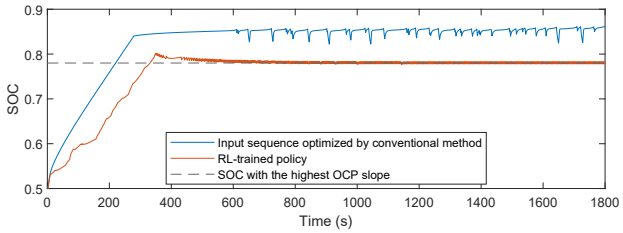


Fig. 7. Actual SOC evolution when current excitations obtained by RL policy and conventional direct sequence optimization (under nominal $\boldsymbol{\varepsilon_{s,p} = 0.5}$, $\boldsymbol{D_{s,p} = 10^{-13}}$) are applied to battery with $\boldsymbol{\varepsilon_{s,p} = 0.4}$, $\boldsymbol{D_{s,p} = 1.2 \times 10^{-13}}$

conventional approach grows substantially to -7.08%, due to the significant decrease in FI evaluated at the actual battery parameters. These results demonstrate that the proposed RL-based input excitation optimization is fundamentally more robust to the model uncertainty, and could address the intrinsic dilemma facing the conventional excitation design regarding the uncertainty of the target parameter.

Regarding the computational complexity of the two methods, on one hand, RL takes more time in training compared with the conventional direct sequence optimization (17000 s versus 500 s on the same computing platform) due to the need for exploration. On the other hand, during implementation, the computation load for applying the RL policy is light and comparable to that of applying the open-loop sequence (0.1s versus 0.07s), as it only involves reading the state feedback and looking up the optimal action. However, it is noted that due to its robustness to parameter uncertainty, the RL policy trained under one parameter (set) can be applied to batteries with different parameters, while the conventional optimization requires re-generating the sequence under different parameter values. Therefore, in this sense, the proposed RL-based method is much more efficient. The robustness of the direct sequence optimization could also be improved by adopting the Model Predictive Control (MPC), which solves the sequence optimization problem online repeatedly over receding horizon and uses the feedback of output/states. The associated computational and memory load, however, are substantial for the real-time implementation phase, which makes it less appealing compared to the RL policy.

TABLE III
FISHER INFO AND ESTIMATION ERROR OF $\boldsymbol{\varepsilon_{s,p}}$ FOR EXCITATIONS OBTAINED BY TWO APPROACHES UNDER NOMINAL/ACTUAL $\boldsymbol{\varepsilon_{s,p}}$ AND $\boldsymbol{D_{s,p}}$ IN SIMULATION

Approach	at nominal $\varepsilon_{s,p} = 0.5$, $D_{s,p} = 10^{-13}$		at $\varepsilon_{s,p} = 0.4$, $D_{s,p} = 1.2 \times 10^{-13}$	
	FI	Est. Err.	FI	Est. Err.
Conventional	300	-2.38%	51	-7.08%
RL	282	-2.50%	284	-2.95%

V. EXPERIMENTAL VALIDATION

In this section, the RL methodology is applied to generate current excitation for estimating the parameter of a physical battery. The results are validated by applying the learned policy to the actual battery in experiment and conducting parameter estimation with current and voltage measurements. The battery used for experimental validation is a LG50M Li Nickle-Manganese-Cobalt (NMC) battery, with most parameters measured in [34] (using a series of invasive measurement techniques) as reference for validation. The OCP-related parameters are identified using low current testing data from the actual battery following procedures in [35]. The battery model with the identified (reference) parameters has been shown to achieve accurate voltage prediction, exemplified by the small model errors shown in Table IV, which is computed as the difference between the measured voltage and the prediction of the model using the reference parameters. Specifically, the mean absolute error (MAE) of voltage prediction under the Federal Urban Driving Schedule (FUDS) and 1C Pulse profile are 5.7 mV and 20.4 mV respectively. It is noted that the reference parameters established this way may still not be the "ground-truth". Nevertheless, they provide an adequate benchmark for evaluating the estimation errors as the parameter values are obtained from an approach different from data fitting. These results, combined with the simulation results discussed in Section IV where the true parameter values are controlled and known, provide comprehensive validation of the proposed methodology. The bounds for current and voltage are $[-5\text{A}, 5\text{A}]$ and $[2.6\text{V}, 4.16\text{V}]$ respectively in accordance with this specific battery chemistry. The target parameter for estimation is $\varepsilon_{s,p}$. In order to emulate the actual estimation scenario, the $\varepsilon_{s,p}$ used in the SPMe for training/optimizing the excitations via RL and the conventional approach deviates from the reference value by 20%, while other parameters use the reference values. The current optimized for estimating $\varepsilon_{s,p}$ using the two approaches are shown in Fig. 8.

It is seen that the optimal excitation for the experimental battery resembles that for the simulated one previously shown in Fig. 2. Both profiles feature 3 major stages, and the resultant SOC converges to similar values, due to the underlying sensitivity dynamics analyzed in Section IV.

The performance of the obtained current excitation is evaluated in terms of the FI and the error when applying to estimate the parameter of the battery in experiment. Specifically, the FI was first computed using the SPMe model with the

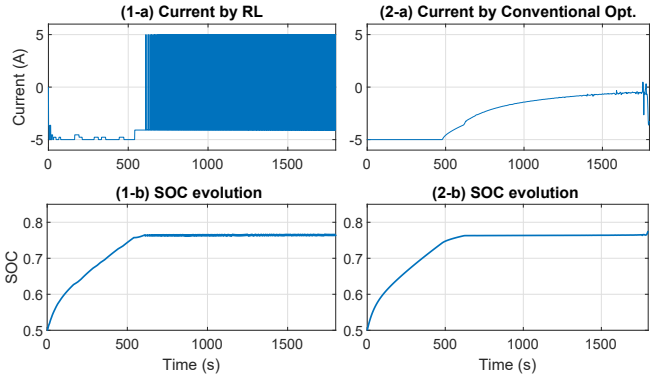


Fig. 8. Optimized current for estimating $\varepsilon_{s,p}$ and resultant SOC by RL (1-a,b) and conventional approach (2-a,b) for experimental battery

TABLE IV
MODELING ERROR, FI AND $\varepsilon_{s,p}$ ESTIMATION ERROR OF VARIOUS EXCITATIONS IN EXPERIMENT

	RL	Conventional	FUDS	1C CC	1C Pulse
Model Err. (MAE, mV)	7.0	8.0	5.7	44.4	20.4
FI	94.5	94.2	2.3	94.0	0.6
Esti. Err.	1.05%	2.48%	2.91%	-9.00%	-37.4%

actual/reference parameters. The current excitation was then applied to the battery using an Arbin LBT 21084 cycler, and the measured voltage was collected and used to estimate the parameter based on SPMe. The least squares algorithm, which is one of the most widely used methods for parameter estimation, is adopted for estimation under all excitations to ensure a fair comparison. The performance is also compared to those of several other current profiles, including FUDS, 1C constant current (CC) discharging, and 1C pulse current, which have been frequently used for battery parameter estimation in current practice. The results are summarized in Table IV.

As shown in the table, the RL-based current excitation yields the highest FI of around 95 and the lowest $\varepsilon_{s,p}$ estimation error of 1.05% when applied to the actual battery. The error is less than 1/2 of that under the profile optimized by conventional approach (2.48%) and that under FUDS (2.91%), and around one order of magnitude smaller than that under 1C pulse (37%) and 1C CC (9%). The excellent estimation error of the proposed approach can be explained by correlating to the FI and an estimation error formula derived in Eqn. (10) of [13]. Specifically, estimation errors are caused by the inevitable system uncertainties in the estimation process, including measurement noises, model uncertainty (e.g., unmodeled/neglected system dynamics), and parameter mismatch (imprecise knowledge of other model parameters). The error formula quantifies the estimation error of the least squares algorithm caused by these uncertainties. It is noted that FI is the denominator of the error formula, meaning that higher FI could effectively reduce the errors. This explains the larger error under the 1C pulse and FUDS current, whose FI is significantly smaller than that of the current optimized by RL as shown in Table IV. For the current optimized by the conventional method and 1C CC current, on the other hand, although their FI is comparable to that of the RL-optimized current, the model uncertainty, which depends on operating conditions, are larger (MAE of 8 mV and 44 mV vs. 7 mV). According to Eqn. (10) in [13], the modeling uncertainty shows up as a term in the numerator of the error formula (multiplied with sensitivity), and hence results in larger estimation error. The substantially smaller estimation error demonstrates the advantage of using the current excitation optimized by the RL approach for estimating battery electrochemical parameter.

VI. CONCLUSION

In this paper, we study the optimization of input excitation for parameter estimation by proposing a new reinforcement learning framework, which addresses the drawbacks of the conventional direct optimization in robustness and tractability. The objective is to obtain an optimal policy for input excitation generation by learning through rewards, which are related to the design criterion, i.e., FI of the data, and computed based on battery physics. The developed RL framework has been applied to estimate two critical battery parameters, i.e., the cathode active material volume fraction $\varepsilon_{s,p}$ and electrode lithium diffusion coefficient $D_{s,p}$. It is shown that the RL approach manages to learn a unique set of input patterns that would yield maximum FI, which can be explained by the underlying dynamics of parameter sensitivity. More importantly, by using the state estimate as feedback for input generation, the RL policy could maintain high FI when applied to systems (batteries) with a target parameter ($\varepsilon_{s,p}$) different from the one used for training, whereas the open-loop sequence optimized by the conventional approach sees significant drop in FI. These results demonstrate the robustness of the RL approach to the uncertainty in target parameter, which is an intrinsic dilemma facing the conventional direct sequence optimization. Experiment validation has also been performed with a physical battery, where the current excitation optimized by the RL approach shows significantly smaller estimation error over other baseline profiles thanks to the higher FI. The obtained RL policy could be used for battery health diagnostics, as well as testing of retired electric vehicle batteries before second-life applications, since the designed excitation could enable more accurate estimation results of battery health-related parameter (e.g., ε_s) under a short testing time as compared to the current practice relying mostly on empirical testing profiles. Such capability is critical for enabling battery repurposing at a large scale. In future works, we plan to apply the methodology to input optimization for multi-parameter joint estimation and subject to more types of system uncertainties, where the design objective is more complicated to evaluate and optimize. We will also apply the framework to more complicated high-order systems, e.g., the full-order DFN model for batteries, for which input optimization using the conventional direct optimization is intractable, and explore the benefits of using other RL algorithms, e.g., the Actor-Critic method.

REFERENCES

[1] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," *Journal of power sources*, vol. 226, pp. 272–288, 2013.

[2] X. Lin, Y. Kim, S. Mohan, J. B. Siegel, and A. G. Stefanopoulou, "Modeling and estimation for advanced battery management," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, pp. 393–426, 2019.

[3] S. Lee, P. Mohtat, J. B. Siegel, A. G. Stefanopoulou, J.-W. Lee, and T.-K. Lee, "Estimation error bound of battery electrode parameters with limited data window," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3376–3386, 2019.

[4] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, "Gaussian process regression for in situ capacity estimation of lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 127–138, 2018.

[5] S. Park, D. Kato, Z. Gima, R. Klein, and S. Moura, "Optimal experimental design for parameterization of an electrochemical lithium-ion battery model," *Journal of The Electrochemical Society*, vol. 165, no. 7, p. A1309, 2018.

[6] Z. Wei, H. He, J. Pou, K.-L. Tsui, Z. Quan, and Y. Li, "Signal-disturbance interfacing elimination for unbiased model parameter identification of lithium-ion battery," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5887–5897, 2020.

[7] Z. Song, H. Hofmann, X. Lin, X. Han, and J. Hou, "Parameter identification of lithium-ion battery pack for different applications based on cramer-rao bound analysis and experimental study," *Applied Energy*, vol. 231, pp. 1307–1318, 2018.

[8] S. J. Moura, F. B. Argomedo, R. Klein, A. Mirtabatabaei, and M. Krstic, "Battery state estimation for a single particle model with electrolyte dynamics," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 2, pp. 453–468, 2016.

[9] Y. Li, B. Xiong, D. M. Vilathgamuwa, Z. Wei, C. Xie, and C. Zou, "Constrained ensemble kalman filter for distributed electrochemical state estimation of lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 240–250, 2020.

[10] Y. Gao, X. Zhang, B. Guo, C. Zhu, J. Wiedemann, L. Wang, and J. Cao, "Health-aware multiobjective optimal charging strategy with coupled electrochemical-thermal-aging model for lithium-ion battery," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3417–3429, 2019.

[11] J. Wu, Z. Wei, W. Li, Y. Wang, Y. Li, and D. U. Sauer, "Battery thermal-and health-constrained energy management for hybrid electric bus based on soft actor-critic drl algorithm," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 3751–3761, 2020.

[12] X. Lin and A. G. Stefanopoulou, "Analytic bound on accuracy of battery state and parameter estimation," *Journal of The Electrochemical Society*, vol. 162, no. 9, p. A1879, 2015.

[13] Q. Lai, H. J. Ahn, Y. Kim, Y. N. Kim, and X. Lin, "New data optimization framework for parameter estimation under uncertainties with application to lithium-ion battery," *Applied Energy*, vol. 295, p. 117034, 2021.

[14] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 2006.

[15] D. V. Lindley, "On a measure of the information provided by an experiment," *The Annals of Mathematical Statistics*, vol. 27, no. 4, pp. 986–1005, 1956.

[16] W. Shen and X. Huan, "Bayesian sequential optimal experimental design for nonlinear models using policy gradient reinforcement learning," *arXiv preprint arXiv:2110.15335*, 2021.

[17] M. J. Rothenberger, D. J. Docimo, M. Ghanaatpishe, and H. K. Fathy, "Genetic optimization and experimental validation of a test cycle that maximizes parameter identifiability for a li-ion equivalent-circuit battery model," *Journal of Energy Storage*, vol. 4, pp. 156–166, 2015.

[18] A. Pozzi, X. Xie, D. M. Raimondo, and R. Schenkendorf, "Global sensitivity methods for design of experiments in lithium-ion battery context," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 7248–7255, 2020.

[19] A. Pozzi, G. Ciaramella, S. Volkwein, and D. M. Raimondo, "Optimal design of experiments for a lithium-ion cell: parameters identification of an isothermal single particle model with electrolyte dynamics," *Industrial & Engineering Chemistry Research*, vol. 58, no. 3, pp. 1286–1299, 2018.

[20] Q. Lai, H. J. Ahn, G. Kim, W. T. Joe, and X. Lin, "Optimization of current excitation for identification of battery electrochemical parameters based on analytic sensitivity expression," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 346–351.

[21] J. Forman, J. Stein, and H. Fathy, "Optimization of dynamic battery paramter characterization experiments via differential evolution," in *2013 American Control Conference*. IEEE, 2013, pp. 867–874.

[22] H. Chun, K. Yoon, J. Kim, and S. Han, "Improving aging identifiability of lithium-ion batteries using deep reinforcement learning," *IEEE Transactions on Transportation Electrification*, 2022.

[23] L. L. Scharf and L. T. McWhorter, "Geometry of the cramer-rao bound," *Signal Processing*, vol. 31, no. 3, pp. 301–311, 1993.

[24] H. Cramér, "Mathematical methods of statistics (pms-9), volume 9," in *Mathematical Methods of Statistics (PMS-9), Volume 9*. Princeton university press, 2016.

[25] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[26] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279–292, 1992.

[27] H. Perez, S. Dey, X. Hu, and S. Moura, "Optimal charging of li-ion batteries via a single particle model with electrolyte and thermal dynamics," *Journal of The Electrochemical Society*, vol. 164, no. 7, p. A1679, 2017.

[28] Q. Lai, S. Jangra, H. J. Ahn, G. Kim, W. T. Joe, and X. Lin, "Analytical derivation and analysis of parameter sensitivity for battery electrochemical dynamics," *Journal of Power Sources*, vol. 472, p. 228338, 2020.

[29] S. Dey, B. Ayalew, and P. Pisu, "Nonlinear robust observers for state-of-charge estimation of lithium-ion cells based on a reduced electrochemical model," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1935–1942, 2015.

[30] R. Xiong, L. Li, Z. Li, Q. Yu, and H. Mu, "An electrochemical model based degradation state identification method of lithium-ion battery for all-climate electric vehicles application," *Applied energy*, vol. 219, pp. 264–275, 2018.

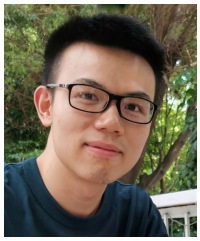
[31] S. Moura, "Single particle model with electrolyte and temperature: An electrochemical battery model," 2016.

[32] X. Lin, "Theoretical analysis of battery soc estimation errors under sensor bias and variance," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7138–7148, 2018.

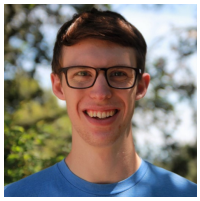
[33] X. Lin, A. Stefanopoulou, P. Laskowsky, J. Freudenberg, Y. Li, and R. D. Anderson, "State of charge estimation error due to parameter mismatch in a generalized explicit lithium ion battery model," in *Dynamic Systems and Control Conference*, vol. 54754, 2011, pp. 393–400.

[34] C.-H. Chen, F. B. Planella, K. O'regan, D. Gastol, W. D. Widanage, and E. Kendrick, "Development of experimental techniques for parameterization of multi-scale lithium-ion battery models," *Journal of The Electrochemical Society*, vol. 167, no. 8, p. 080534, 2020.

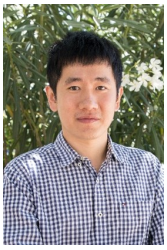
[35] Q. Lai, J. Fogelquist, and X. Lin, "System identification of battery single particle model parameters using new data optimization approach," in *American Control Conference*. IEEE, 2022.



Rui Huang received the B.S. degree in automotive engineering and the M.S. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2017 and 2020 respectively. He is currently a Ph.D. student in the Department of Mechanical and Aerospace Engineering at the University of California, Davis, CA, USA. His research interest include the modeling, control, disgnostics and data analysis of battery systems.



Jackson Fogelquist received B.S. and M.S. degrees in mechanical engineering from San José State University, San José, CA, USA, in 2017 and 2019, respectively. He is currently a Ph.D. student in the Department of Mechanical and Aerospace Engineering at the University of California, Davis, CA, USA, and a Fellow of the California State University Chancellor's Doctoral Incentive Program. His research interests include modeling, estimation, data analysis, and data optimization with applications in battery systems.



Xinfan Lin (SM'22, M'17) received B.S. and M.S. degrees in automotive engineering from Tsinghua University, Beijing, China, in 2007 and 2009, respectively, and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2014.

He is currently an Associate Professor with the Department of Mechanical and Aerospace Engineering, University of California, Davis, CA, USA. He is also a faculty member of the Institute of Transportation Studies (ITS) at UC Davis. From 2014–2016, he was a Research Engineer at Ford Motor Company, responsible for development of the battery management system and other EV components and systems. He has published over 50 peer-reviewed papers, and has been granted 5 patents. He has received several notable awards, including the NSF CAREER Award, LG Global Innovation Award, and LG Battery Innovation Award among others. He has also been serving as the Secretary, Publicity Chair, and Award Committee Chair of the Energy Systems Technical Committee (ESTC) of the ASME Dynamic Systems and Control Division (DSCD). His research interests include modeling, estimation, control, data analysis, and machine learning with applications in energy, automotive, and aerospace systems.