

Research Papers

A comprehensive approach to sparse identification of linear parameter-varying models for lithium-ion batteries using improved experimental design

A.M.A. Sheikh ^{a,*}, M.C.F. Donkers ^a, H.J. Bergveld ^{a,b}

^a Department of Electrical Engineering, Eindhoven University of Technology, Groene Loper 19, 5612AP, Eindhoven, Netherlands

^b NXP Semiconductors, High Tech Campus 46, 5656AE Eindhoven, Netherlands



ARTICLE INFO

Dataset link: <https://github.com/tue-battery/PyBatteryID>

Keywords:

Lithium-ion batteries
Equivalent-circuit model
LPV framework
Sparse identification

ABSTRACT

This paper proposes a comprehensive approach to the identification of battery models in the linear parameter-varying (LPV) framework inspired by the equivalent-circuit model structures. The proposed LPV model structure is formulated using the input–output representation, wherein the model parameters are considered to depend on the state-of-charge, the current magnitude, and the current direction. The aforementioned dependence is explained using a suitable set of basis functions motivated with appropriate physical insight. Furthermore, the optimal experimental design problem is discussed to propose an improved input design for the model identification procedure. Subsequently, the Least Absolute Shrinkage and Selection Operator (LASSO) along with the ridge regression are employed for the selection of significant model terms and parameter estimation to yield a sparse model with adequate simulation capabilities. Finally, several battery models with varying model order and basis-function complexity are identified, which are subsequently validated and compared using a real drive-cycle dataset. The corresponding voltage simulation results yield the root-mean-squared error (RMSE) value for the best-performing model to be around 24.5 mV when a 1 A h NMC battery gets discharged with the lower voltage cutoff as low as 2.5 V.

1. Introduction

Owing to the increased global focus on reducing carbon emissions, the electrification of various transport systems using lithium-ion batteries has been at the forefront of recent research. To meet the desired safety and performance criteria in such battery-powered systems, battery management systems (BMS) are employed to accurately describe the nonlinear battery behaviour using appropriate algorithms given a particular set of measurements, such as current, voltage, and temperature. Quite generally, model-based techniques have gained much traction in understanding various battery behaviours under dynamic operating conditions. The commonly studied battery models can be mainly classified into physics-based analytical models and data-driven empirical models. The physics-based models utilize mathematical formulations of the electrochemistry and thermodynamics principles, involving a system of partial differential equations. A few examples of such models include (i) the Doyle–Fuller–Newman (DFN) model [1] based on porous-electrode theory, (ii) Single-Particle Model (SPM) [2] which is a simplification of the DFN model, (iii) lattice-gas models [3, 4], and (iv) the models based on the theory of nonequilibrium thermodynamics [5,6]. Although the physics-based models can accurately

describe various battery processes with high interpretability, their implementation in a BMS environment is still a bane due to their high computational complexity and laborious parameterization procedures. On the other hand, the data-driven empirical battery models attempt to learn the dynamics of interest from the given experimental data with no prespecified model structure — also referred to as *black-box* models. Such modelling approaches which include Fuzzy Logic [7], Neural Networks [8], and Gaussian Process Regression [9], can also achieve adequate prediction accuracy, but the questions regarding the interpretability of the identified models remain unaddressed.

The equivalent-circuit models (ECMs) constitute another important class of battery models in the literature, which can be more appropriately categorized as lower-fidelity physics-based models, that is, the *semi-empirical* models. Essentially, the circuit components are incorporated into the model structure using physical insight, for instance, the resistor–capacitor networks represent the electrochemical interfacial dynamics inside the battery (see Ch. 4 in [10]). Accordingly, if the identification procedure uses experimental data that sufficiently spans such battery dynamics (e.g., over a broad frequency spectrum), then the resulting parameters can become representative of the underlying

* Corresponding author.

E-mail addresses: a.m.a.sheikh@tue.nl (A.M.A. Sheikh), m.c.f.donkers@tue.nl (M.C.F. Donkers), h.j.bergveld@tue.nl (H.J. Bergveld).

physical processes, rendering the resulting ECM physically insightful. For instance, an ECM can be identified using the so-called theory of the distribution of relaxation times (DRT) along with the Electrochemical Impedance Spectroscopy (EIS) measurements [11]. Reportedly, the ECM parameters can then represent the battery internal resistance, the SEI layer, the charge-transfer kinetics, and the diffusion processes inside the battery. In addition, an application-specific approach to the identification of an ECM also exists, which involves experimental data representing a subset of the battery dynamics in a limited frequency range, for instance, related to a drive-cycle application [12]. The estimated parameters of such an ECM become less representative of the underlying physical processes as the size of the chosen subset of the battery dynamics shrinks. Nevertheless, such ECMs have been the preferred choice from the practical perspective due to their adequate accuracy, ease of parameterization, and reduced computational complexity.

Primordially, the ECMs have been represented within a linear time-invariant (LTI) framework, wherein the model parameters remain constant during the battery operation [13–15]. Since the battery is better described as a nonlinear system, the LTI framework cannot capture the time-varying dynamics, even if a higher-order ECM is chosen [16]. To mitigate several issues with the LTI framework, an intermediate paradigm between the LTI and the nonlinear systems has been presented in the literature giving rise to a class of *linear parameter-varying* (LPV) models [17]. Essentially, the nonlinear battery dynamics are approximated by a measurable time-varying signal – referred to as the scheduling variable – over the entire battery operating regime. Such mathematical description leads to the time-varying but linear relations between the inputs and the outputs of the battery, thereby allowing better generalizability for the LPV-ECMs than their LTI counterparts.

A brief literature review of the ECMs suggests an increased focus on the incorporation of unique battery characteristics in the prescribed LPV models [18–20], such as the state-of-charge (SOC) dependence, the current-direction dependence, and the temperature dependence of the model parameters. Such an approach effectively provides additional degrees of freedom to capture the complex battery dynamics. Yet the question of generalizability persists whether the adopted model identification approach can lead to better approximative models of the battery, thereby enhancing the model applicability over a wide range of the intended battery operation. Often, the principle of parsimony gets adopted while modelling various physical processes based on the presumption that nature admits simplistic descriptions in general [21]. Accordingly, a parsimonious or sparse model gets preferred over a complex model even if the latter performs *slightly* better than the former [22]. Furthermore, the quest to obtain sparse models relies profoundly on the informativity of the experimental datasets used for the model identification. Remarkably, the battery literature is devoid of any work that focuses on the identification of sparse battery models while emphasizing informative identification experiments. As such, there is a need to emphasize a proper structure while prescribing novel battery models along with informative experiments and efficient model identification procedures to obtain generalizable models with adequate prediction (or simulation) capabilities.

In this paper, we demonstrate a comprehensive approach to identifying a sparse battery model by extending the equivalent-circuit model structure in the LTI framework to the LPV framework. Firstly, we motivate the dependence of various model components on the SOC, the current magnitude and the current direction in the form of the so-called scheduling variable. Secondly, the functional form of the scheduling-variable dependence has been explored using a set of suitable basis functions with appropriate physical insight. Moreover, we briefly discuss the required informativity of the experimental datasets used in the model identification procedure and subsequently propose a suitable design for the input current profile. Next, we employ the Least Absolute Shrinkage and Selection Operator (LASSO) and the ridge regression for the selection of significant model terms and the

parameter estimation. Finally, we validate several models with varying model orders and basis-function complexities using a real drive-cycle dataset and compare the respective voltage simulation performances.

This paper has been organized as follows. Section 2 provides preliminaries regarding common linear parameter-varying system representations. Section 3 introduces the proposed battery model structure in the LPV framework. Section 4 motivates a suitable input signal design from the perspective of the optimal experimental design problem. Section 5 describes a model selection and estimation procedure to yield a sparse battery model. Section 6 presents the conditions regarding the battery experiments presented in this paper. Section 7 discusses the obtained results regarding the identified battery models. Section 8 concludes the work presented herein.

2. LPV model representations

This section gives a brief introduction to the two commonly studied model representations of linear parameter-varying systems, namely the state-space and the input–output representations. The respective subsections introduce some notation required in the context of this paper, which has been presented in a similar fashion as in the literature related to the linear parameter-varying systems (see [17] for a detailed presentation).

2.1. LPV state-space representation

In contrast to linear time-invariant (LTI) systems, the state-space representation of a linear parameter-varying (LPV) system introduces a so-called scheduling variable in the system matrices. More formally, the discrete-time LPV state-space representation of a system in a single-input single-output (SISO) setting can be given by

$$x_{k+1} = (\mathcal{A} \diamond p)_k x_k + (\mathcal{B} \diamond p)_k u_k \quad (1a)$$

$$y_k = (\mathcal{C} \diamond p)_k x_k + (\mathcal{D} \diamond p)_k u_k \quad (1b)$$

where $x : \mathbb{Z} \rightarrow \mathbb{R}^{n_x}$ is the state vector, $u : \mathbb{Z} \rightarrow \mathbb{R}$ the input, $y : \mathbb{Z} \rightarrow \mathbb{R}$ the output, $p : \mathbb{Z} \rightarrow \mathbb{P} \subseteq \mathbb{R}^{n_p}$ the scheduling variable, $k \in \mathbb{Z}$ the discrete time instant, and $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ the collection of p -dependent matrices which represent a *state-space realization* of the LPV system. Note that the notation $(\square \diamond p)_k$ corresponds to having an arbitrary functional dependence on the variable p at different time instants, e.g., $(f \diamond p)_k = f(\dots, p_{k-1}, p_k, p_{k+1}, \dots)$.

2.2. LPV input–output representation

In addition to the state-space representation, the discrete-time LPV representation of a system can be given using a difference equation relating only the system inputs and the outputs

$$y_k + \sum_{i=1}^{n_a} (a_i \diamond p)_k y_{k-i} = \sum_{i=0}^{n_b} (b_i \diamond p)_k u_{k-i} \quad (2)$$

where $(a_i \diamond p) : \mathbb{P} \rightarrow \mathbb{R}$ and $(b_i \diamond p) : \mathbb{P} \rightarrow \mathbb{R}$ are the p -dependent model parameters, and n_a and n_b the model orders with $n_a \geq n_b \geq 0$. An important point to note here is that a parameter a_i is considered to have *static* dependence on p if it depends only on the instantaneous value of p , i.e., $(a_i \diamond p)_k = a_i(p_k)$, otherwise, the *dynamic* dependence is implied. The same holds for a matrix in the state-space representation above, for instance, $(\mathcal{A} \diamond p)_k = \mathcal{A}(p_k)$ implies static p -dependence.

3. Proposed model structure

The proposed battery model structure decomposes the battery terminal voltage V into (1) an *a priori* known component function V_{EMF}

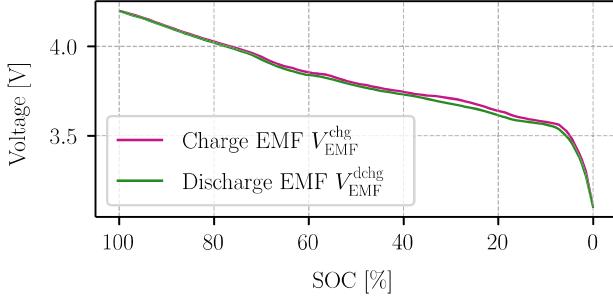


Fig. 1. Charge and discharge EMF curves as a function of SOC for the 1 A h lithium-ion NMC/graphite battery cell studied in this paper (see Section 7) obtained using GITT experiments.

representing the battery electromotive-force — also referred to as open-circuit equilibrium voltage, and (2) an unknown component function F describing the battery overpotentials. Concretely,

$$V_k = V_{\text{EMF}}(s_k) + \underbrace{F(y_{k-1}, \dots, u_k, u_{k-1}, \dots, p_k, p_{k-1}, \dots)}_{=y_k} \quad (3)$$

where s is the battery SOC, y the total battery overpotential, u the battery input current, and p the scheduling variable.

3.1. Motivation behind model components

The above model equation can be motivated using the physical intuition behind the component functions V_{EMF} , and F , which would also facilitate the direct formulation of the input–output representation of an equivalent-circuit model. As mentioned earlier, V_{EMF} represents the battery EMF which, in this paper, has been considered to be a static nonlinear function depending only on the battery SOC s , and assumed to be available *a priori* through various techniques [23], such as Galvanostatic Intermittent Titration Technique (GITT), extrapolation to zero current, or low-current cycling. More precisely, three trivial choices for the function V_{EMF} exist, namely the charge $V_{\text{EMF}}^{\text{chg}}$, the discharge $V_{\text{EMF}}^{\text{dchg}}$, or the average $V_{\text{EMF}}^{\text{avg}}$ EMF function; the latter being the average of the first two functions. In this paper, $V_{\text{EMF}}^{\text{avg}}$ has been chosen to represent the battery EMF assuming negligible hysteresis, which has been obtained using the GITT experiments in both the charge and the discharge directions (see Fig. 1).

The component function F essentially represents the overpotentials associated mainly with the battery dynamics due to (i) the mass transport (diffusion) limitations inside the battery resulting in a *slow* overpotential, (ii) the battery internal ohmic resistance resulting in a *fast* overpotential, and (iii) the charge-transfer resistance at the electrode–electrolyte interface resulting in an *intermediate* overpotential compared to the first two overpotentials. Note that the term *slow* (respectively *fast*) refers to a physical process having the characteristic time constant sufficiently larger (respectively smaller) than the model sampling time. For most practical purposes, such as during the drive-cycle application, it suffices to consider a finite-order difference equation as given by (2) to approximate the underlying overpotentials, namely

$$y_k + a_1 y_{k-1} + \dots + a_n y_{k-n} = b_0 u_k + b_1 u_{k-1} + \dots + b_n u_{k-n} \quad (4)$$

where n is the order of the difference equation, and a_i , b_i the parameters of the difference equation. Note that the explicit p -dependence of the parameters $(a_i \diamond p)$, $(b_i \diamond p)$ has been omitted for the sake of readability. An interesting idea is to describe the functional dependence of the parameters using the SOC s , the magnitude of the input current $|u|$, and the direction of the input current δ . Such an argument can be easily motivated by the reported dependence of the model parameters on the aforementioned variables [24–26]. Note that the dependence of battery

behaviour on the temperature and hysteresis phenomenon can also be incorporated into the model structure using appropriate experimental design and basis-function selection, but has not been considered in this paper for the sake of simplicity; see [27] for strategies to incorporate the hysteresis phenomenon into the proposed model structure.

3.2. Scheduling-variable dependence

The SOC s , the current magnitude $|u|$, and the current direction δ dependence introduced in the model component function F can be combined into a single variable — the so-called scheduling variable. Formally, the scheduling variable p can be defined as a three-dimensional vector $p : \mathbb{Z} \rightarrow \mathbb{P} \subset \mathbb{R}^3$ given by

$$p_k := [s_k \quad |u_k| \quad \delta_k]^T \quad (5)$$

where the scheduling space \mathbb{P} is a subset of \mathbb{R}^3 , namely $\mathbb{P} = [0, 1] \times [0, \infty) \times [-1, 1]$. The corresponding SOC trajectory s_k can be given as a solution to the following widely accepted coulomb-counting difference equation in the literature, namely

$$s_{k+1} = s_k + (\tau/Q)u_k \quad (6)$$

where τ corresponds to the model sampling time, and Q the battery capacity. Moreover, the trajectory δ_k can be obtained by defining δ using the sign of the input current through a first-order difference equation as given by

$$\delta_k = \varepsilon(u_k)\delta_{k-1} + [1 - \varepsilon(u_k)]\text{sgn}(u_k), \quad \text{where } \varepsilon(u_k) = \begin{cases} \varepsilon_0 & \text{for } u_k \neq 0 \\ \varepsilon_1 & \text{for } u_k = 0 \end{cases} \quad (7)$$

and $\text{sgn}(\cdot)$ corresponds to the signum function. In addition, $\varepsilon : \mathbb{R} \rightarrow \mathbb{E} \subseteq \mathbb{R}$ can be considered as a *hybrid* parameter which takes on different values depending on the current u_k , where $\varepsilon_0, \varepsilon_1 \in [0, 1]$. Note that the notation $\delta_k^{[\varepsilon_0, \varepsilon_1]}$ has been used to denote a specific current-direction trajectory for the given values of ε_0 and ε_1 . For instance, the values $\varepsilon_0 = 0$ and $\varepsilon_1 = 1$ yield $\delta_k^{[0,1]} = \text{sgn}(u_k)$ when the current is nonzero, and $\delta_k^{[0,1]} = \delta_{k-1}^{[0,1]}$ otherwise. In effect, δ can be considered to have a memory such that it retains its previous value when the current becomes zero. Another interesting case can correspond to choosing ε_0 and ε_1 close but not equal to 0 and 1, respectively. In such a case, δ has two notable properties, namely (i) it has a *finite memory* so that the value of δ slowly decays to zero when the current becomes zero, which can model extremely slow relaxation effects inside the battery, and (ii) it can be considered as a continuous analogue of the otherwise discontinuous $\text{sgn}(\cdot)$ function when the current is nonzero, which effectively renders the trajectories of the scheduling variable p in the scheduling space continuous. Simply put, the model parameters change their values “smoothly” when the current direction is changed, which is based on the practical assumption that the real physical processes are continuous.

3.3. Input–output representation

Quite generally, the static nature of the V_{EMF} function in (3) gets exploited in the literature to explain the battery terminal voltage by modelling the battery overpotentials separately. Concretely,

$$y_k = V_k - V_{\text{EMF}}(s_k) = F(y_{k-1}, \dots, y_{k-n}, u_k, \dots, u_{k-n}, p_k, \dots, p_{k-n}) \quad (8)$$

where y_k represents the total effective overpotential at time instant k , which can be regarded as the output of the overpotential model, and n the model order. In addition, the equivalent-circuit models for battery overpotentials have been ubiquitously realized using the modal state-space representation, wherein the A -matrix takes the diagonal form, and the state vector x represents the collection of voltages across the resistor–capacitor networks. In this paper, the input–output model representation has been readily adopted as the starting point by noting the fact that an n th-order state-space representation of an

equivalent-circuit model in the LTI framework can be transformed into an n th-order difference (or differential) equation of the form given by (4) using transfer-function-based approaches (see [28] as an example). Now, the scheduling-variable dependence discussed previously can be formally incorporated in the model structure by rendering the parameters $\{a_i\}_{i=1}^n, \{b_i\}_{i=0}^n$ in (4) as functions of the variable p . In this regard, several possibilities exist to introduce the p -dependence of model parameters in the LTI framework, thereby yielding LPV model structures that can involve varying degrees of mathematical complexity but with little practical difference in the present context. Accordingly, the so-called *shifted-form* input–output representation will be considered in this paper, which incorporates a special dynamic p -dependence as can be given by

$$y_k + \sum_{i=1}^n a_i(p_{k-i})y_{k-i} = \sum_{i=0}^n b_i(p_{k-i})u_{k-i} \quad (9)$$

It is worth mentioning that the chosen shifted-form for the LPV input–output representation in (9) simplifies the derivation of the equivalent LPV state-space representation for an arbitrary model order n , which otherwise may become intractable as the model order is increased (see [29] for an introduction to LPV equivalence transformations).

3.4. State-space representation

Even though the proposed methodology does not rely on the state-space representation for model identification, it may be beneficial to present the equivalent state-space representation of the proposed model structure for the sake of completeness. A possible application of such representation can be given during the design of an extended Kalman Filter for the estimation of the battery SOC [30]. Accordingly, the state-space representation of the shifted-form input–output representation in (9) can be obtained using the so-called cut-and-shift procedure as detailed in [31]. The resulting state-space representation of the proposed model can be given by the matrix quadruple $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ as follows

$$\left[\begin{array}{c|c} \mathcal{A}(p_k) & \mathcal{B}(p_k) \\ \hline \mathcal{C}(p_k) & \mathcal{D}(p_k) \end{array} \right] = \left[\begin{array}{ccccc|c} -a_1(p_k) & 1 & 0 & \cdots & 0 & b_1(p_k) - a_1(p_k)b_0(p_k) \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & 0 & 0 & \cdots & 1 & \vdots \\ -a_n(p_k) & 0 & 0 & \cdots & 0 & b_n(p_k) - a_n(p_k)b_0(p_k) \\ \hline 1 & 0 & \cdots & \cdots & 0 & b_0(p_k) \end{array} \right] \quad (10)$$

where the special form of the dynamic p -dependence considered in the input–output representation gets converted into the static dependence in the state-space representation. Moreover, since the above state-space representation is not in the modal (diagonal) form, the corresponding state vector does not directly represent the collection of voltages across the resistor–capacitor networks of an ECM. An important point to note here is that a physical interpretation of the identified resistors and capacitors can be quite misleading since there do not exist any established guarantees that such resistors and capacitors would not take “physically impossible” values during a black-box model identification procedure. Nevertheless, the existence of such a diagonal form can reveal the analytical relations describing the functional dependence of the corresponding resistors and capacitors on the scheduling variable p . More formally, given the state-space representation $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$, the corresponding modal representation requires the formulation of the so-called dynamic eigenvalue problem [32] involving a possibly parameter-varying transformation matrix $(\mathcal{L} \diamond p)_k$ as given by

$$(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}) \xrightarrow{(\mathcal{L} \diamond p)_k} (\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \tilde{\mathcal{D}}) \quad (11)$$

where the diagonalized matrix $\tilde{\mathcal{A}}$ should contain elements on the diagonal as real-valued functions, which can then be related to the resistor–capacitor networks of an LPV-ECM. Unfortunately, the above problem formulation requires diagonalization of a time-varying matrix, which is a non-trivial task, and therefore has been considered out of the scope of this paper.

3.5. Functional dependence of model parameters

Even though the p -dependence of the model parameters $\{a_i\}$ and $\{b_i\}$ has been well recognized using either experimental observations or appropriate physical reasoning (see Section 3.2), the exact form of the functional dependence remains unknown. This inherently constitutes a black-box identification problem. In effect, the parameters $\{a_i\}$ and $\{b_i\}$ can then be expanded in a completely flexible manner by allowing numerous many possibilities for the so-called basis functions, namely polynomials, exponentials, (hyperbolic) trigonometric functions, and so on. Often, it is advisable to blend some physical reasoning with the choice of the basis functions to avoid the black-box approach and rather adopt a somewhat *dark-greyish*¹ approach. Fortunately, some physical insight can be given in this case to relate the behaviour of the model parameters with the individual components of p , that is, the SOC s , the current magnitude $|u|$, and the current direction δ . Accordingly, the basis function expansion in terms of each variable can be motivated as follows.

SOC s . Generally, the LPV battery models proposed in the literature assume polynomial basis functions to explain the SOC-dependence of the model parameters. A bit of careful experimental observations suggests that certain battery parameters, such as the internal resistance, may exhibit asymptotic *blow-up* behaviour as the SOC approaches zero. Therefore, the considered basis functions should exhibit a slope approaching infinity as the SOC approaches zero to model such behaviour — a property that the polynomials do not possess. Consequently, the basis function expansion in terms of polynomials would require several higher-order polynomial terms to model the battery behaviours at low SOCs. To overcome this issue, an extended class of polynomials referred to as fractional polynomials [34], is considered in this paper, which can include the terms containing fractional and negative exponents. More formally, the family of fractional polynomials can be given by

$$\{\dots, s^{-m_1}, s^{-m_2}, \ln(s), s^{m_3}, s^{m_4}, \dots\}$$

where $m_i \in \mathbb{R}$. The above set can be significantly restricted to only a few terms, namely $\{s^{-1}, \ln(s), s\}$, to essentially simplify the model identification problem. Such a restriction can be motivated using a hopeful argument that the inclusion of the s^{-1} term would lead to the generation of sufficiently representative candidate model terms, thereby allowing sparse model identification.

Current magnitude $|u|$. In a similar fashion as the SOC, it can be argued that the model parameters may exhibit increased nonlinear behaviour as the current magnitude is increased. Again, a careful examination of various function classes suggests that the introduction of exponential terms may result in a more parsimonious representation than using only the polynomial terms. For instance, using a term $|u|^p$ for $p \geq 1$ as a possible basis function would result in a candidate model term $|u|^p u$ which is less than u for the values of u between 0 and 1. Consequently, at least two terms are needed to explain the nonlinear behaviour related to higher current values, for instance, $\alpha u + \beta |u|u$ for some values of $\alpha, \beta \in \mathbb{R}$. Ideally, a single candidate term that behaves linearly for smaller current values and increases nonlinearly for higher current values may be more representative. Such a functional behaviour can be achieved by considering a term of the form $e^{|u|}u^p$ as a basis function, which would result in a candidate term $e^{|u|}u^p u$ possessing the desired behaviour, that is, $e^{|u|}u^p u > u$ for $u = (0, \infty]$ and $\gamma, p > 0$. Now, it can be easily examined that for $\gamma = p = 1$, (1) the function $e^{|u|}u$ exhibits severe nonlinear growth, and (2) the

¹ Refer to [33] for a perspective on model structures and the palette of grey shades.

range of u for which the function can be considered practically linear is quite small, which may not be sufficient to explain the linear battery operating regime. A possible solution can be given by using a comparatively less nonlinear function, such as $e^{\gamma\sqrt{|u|}u}$, where γ acts as the normalization factor such that the linear battery operating regime can be matched with that of the function itself. Accordingly, a suitable basis function set chosen in this paper can be given by $\{e^{\gamma\sqrt{|u|}u}\}$ with $\gamma = \frac{1}{20}$ to explain the current-magnitude dependence of the model parameters. Note that when the battery models are expected to explain severe nonlinear behaviour, which can be the case for an aged battery, the functions such as $e^{\gamma|u|^p}$ with $p \geq 1$ may serve as appropriate basis-function candidates.

Current direction δ . A relatively straightforward approach has been adopted to explain the current-direction dependence of the model parameters, that is, only the linear terms of the form $\delta^{[\epsilon_0, \epsilon_1]}$ have been studied. In principle, a two-dimensional manifold can be searched to generate the δ -trajectories for various values of ϵ_0 and ϵ_1 . However, only the values $(\epsilon_0, \epsilon_1) = (0.01, 0.99)$ have been considered herein to approximate the *obvious* direction-dependent battery behaviour. Accordingly, the basis function set can be given by $\{\delta^{[0.01, 0.99]}\}$.

Once the suitable basis functions have been selected for each component of the scheduling variable, an extra step can be taken to generate the combinations of the individual functions with replacement. More formally, considering l as the *nonlinearity order* and taking an arbitrary value $l = 3$, a dictionary (set) of basis functions which includes the combinations can be given by

$$\mathcal{G} \triangleq \{1, s^{-1}, \dots, s \ln(s), \dots, se^{\frac{1}{20}\sqrt{|u|}}, \dots, s^3, \dots, se^{\frac{1}{20}\sqrt{|u|}}\delta^{[0.01, 0.99]}\} \quad (12)$$

where the cardinality $|\mathcal{G}| = r$ represents the number of elements in \mathcal{G} . Note that the redundant combinations in \mathcal{G} can be discarded, such as $s^{-1}s$, $s^{-1}\ln(s)s$, and so on. Furthermore, the combinations containing multiple δ -trajectories have also been ignored for the sake of simplicity. It can be noted that even a quite conservative selection of the individual basis functions can lead to a massive dictionary of the possible basis functions when the cross-terms (or combinations) are allowed. Finally, the parameters $\{a_i\}$ and $\{b_i\}$ can be expanded in terms of the basis functions as given by

$$a_i(p) = \sum_{j=1}^r a_{ij}g^{(j)}, \quad b_i(p) = \sum_{j=1}^r b_{ij}g^{(j)} \quad (13)$$

where $\{a_{ij}\}$ and $\{b_{ij}\}$ represent the to-be-determined *free* model parameters, and $g^{(j)}$ the j th element in \mathcal{G} .

Remark. The incorporation of physical insight during the formulation of the dictionary \mathcal{G} reduces the flexibility (or complexity) of the model class through the restrictive inclusion of certain basis functions, thereby rendering the black-box identification setting into *dark-greyish*. As pointed out in [33], system identification is not just a science but an art of solving real problems, and using adequate physical insights “to shrink the model set as much as possible” may protect against the failure of a model identification attempt.

4. Optimal experimental design

During the model identification procedure, an important concern related to designing optimal experiments arises so that the quality of the resulting parameter estimate can be maximized. Several design variables can be considered while formulating the optimal experimental design problem, which include the input frequency spectrum, the sampling interval, and the size of the collected dataset [35]. In this paper, an analytically motivated suggestion presented in [36] (see Result 4.1) has been readily adopted as the guiding principle, which advocates for the design of an identification experiment to resemble the intended

model application “as much as possible”. Herein, the problem has been confined specifically to designing an input signal with appropriate frequency-domain and time-domain properties, while considering the model application as in a practical battery management system.

The explicit realization of the desired input signal for the identification experiment can be motivated by formulating several arguments based on the frequency spectrum and the amplitude distribution of a typical drive cycle. Ideally, the input signal should appropriately excite all the battery dynamics corresponding to the expected frequencies during a practical drive cycle at each point in the scheduling space. From the time-domain perspective, since the practical drive cycles involve varying current amplitudes, the amplitude distribution of the input signal should resemble that of the drive cycle to capture the nonlinear effects to the relevant extent. Likewise, the input signal should also contain sufficiently large regions of zero current ($u_k \sim 0$) to explore the necessary diffusive dynamics inside the battery. It may be worth mentioning that even though some works have quite rightly advocated against the use of pulse current tests during the model identification intended for practical BMS applications [16,37], the absence of appropriate zero-current regions in the input signal can suppress the relevant dynamics during the identification experiment, thereby affecting the quality of the parameter estimate.

In light of the above statements, a possible input signal design can be given using a rectangular pulse train (RPT) $_{\tau, \tau_0}$ with period τ and zero-input regions of length τ_0 as follows

$$\mathcal{I} \triangleq \alpha \cdot (\alpha_0 \cdot (\text{RPT})_{\tau, \tau_0} + \xi + v) \quad (14)$$

where \mathcal{I} represents the stochastic input signal having the following characteristics. Namely, (i) the initial amplitude of the pulse train can be specified using a fixed value α_0 , (ii) the time period of each pulse in the pulse train can be given by a random variable sampled from a normal distribution $\tau \sim \mathcal{N}(\mu_\tau, \sigma_\tau)$ such that the frequency spectrum of sufficiently many pulses resembles that of a practical drive cycle, (iii) a zero-current region of length given by a random variable $\tau_0 \sim \mathcal{N}(\mu_{\tau_0}, \sigma_{\tau_0})$ can be inserted following each phase of the pulses in the pulse train, (iv) a zero-mean random signal $\xi \sim \mathcal{N}(0, \sigma_\xi)$ can be superimposed on the pulse train to excite the high-frequency dynamics, (v) a fixed DC offset v can be introduced so that the pulse train effectively discharges or charges the battery, thereby spanning the entire SOC range during the identification experiment, and (vi) the amplitude of the final signal can be modified by a random variable $\alpha \sim \mathcal{U}(\alpha_-, \alpha_+)$ to match the resulting amplitude distribution with that of the drive cycle. Note that the random signal ξ and the offset v become piecewise zero during the zero-current regions, thereby not affecting the rest period following each phase of the pulses in the pulse train. Moreover, from the practical perspective, the offset v should be small enough so that the battery experiences a sufficient number of pulses for an SOC range wherein the model parameters $\{a_i\}, \{b_i\}$ as functions of only the SOC can be regarded constant. Admittedly, such nonanalytical arguments cannot lead to a unique optimal input signal, and several other signal designs can be motivated, for instance, $\mathcal{I}' \triangleq \alpha \cdot (\alpha_0 \cdot (\text{RPT})_{\tau, \tau_0} + v) + \xi$. Nevertheless, this paper demonstrates the utility of the signal \mathcal{I} as it leads to adequate identification results, which will be discussed in Section 7.

5. Model selection and estimation

Given an informative experimental dataset and a set of suitable basis functions \mathcal{G} as given in (12), the subsequent steps in a model identification procedure can be given by the model selection and the estimation. The former step includes the model-order selection as well as the selection of the significant model terms — also referred to as the variable selection, feature selection, or support recovery [38]. Indeed, the model selection procedure becomes redundant when the model structure is known *a priori*, that is, in the case of white-box or *light-greyish* identification settings where appropriate physical insights are

available. However, in the case of black-box identification, the model-selection step becomes important to ensure only the relevant terms in the model structure while respecting the principle of parsimony. In the following, the model selection problem will be limited to variable selection while assuming the model order n to be given.

Various methods have been proposed in the literature to tackle the variable selection problem, which include (i) the exhaustive subset selection methods which identify the best subset by comparing all possible combinations of the candidate model terms using a suitable quantitative measure, such as the Akaike or the Bayesian information criterion, (ii) the stepwise selection methods which iteratively add the significant terms or drop the insignificant terms from the model structure based on some quantitative measure, and (iii) the shrinkage methods which solve a penalized regression problem forcing the estimated parameters corresponding to the insignificant terms towards zero. Understandably, the best subset selection approach quickly becomes intractable as the number of candidate model terms increases. Furthermore, it has been reported that the shrinkage methods, such as the Least Absolute Shrinkage and Selection Operator (LASSO), may parallel with or even outperform the subset selection and the stepwise selection methods for variable selection [39,40]. In this paper, the LASSO has been employed, which shrinks the parameter vector such that several parameters turn out to be exactly zero, thereby performing the desired variable selection automatically. Accordingly, the input-output representation in (9) can be used to consider a linear regression setting as given by,

$$Y = \Phi\theta + \epsilon \quad (15)$$

where $Y \in \mathbb{R}^{N \times 1}$ represents the collection of the output measurements $\{y_k\}_{k=1}^N$, $\Phi \in \mathbb{R}^{N \times M}$ the collection of the regression vectors $\{\phi_k\}_{k=1}^N$, $\theta \in \mathbb{R}^M$ the unknown parameter vector, and $\epsilon \in \mathbb{R}^{N \times 1}$ the error component. In addition, the regression vector $\phi_k : \mathbb{Z} \rightarrow \mathbb{R}^M$ contains all the candidate model terms which can be generated using the set \mathcal{C} , that is,

$$\phi_k = f_k \otimes g_k$$

with

$$\begin{aligned} f_k &\triangleq [y_{k-1} \quad \cdots \quad y_{k-n} \quad u_k \quad \cdots \quad u_{k-n}] \\ g_k &\triangleq \begin{bmatrix} g_k^{(1)} & \cdots & g_k^{(r)} \end{bmatrix} \\ &= \begin{bmatrix} 1 & s_k^{-1} & \cdots & e^{\frac{1}{20}\sqrt{|u_k|}} s_k & \cdots \end{bmatrix} \end{aligned}$$

where \otimes denotes the Kronecker product, $f_k : \mathbb{Z} \rightarrow \mathbb{R}^{2n+1}$ represents the vector containing the output and the input terms, $g_k : \mathbb{Z} \rightarrow \mathbb{R}^r$ the vector containing the basis functions, and $M = r(2n + 1)$. Furthermore, the parameter vector θ can be expressed in terms of the free model parameters as given by

$$\theta = [-a_{11} \quad \cdots \quad -a_{1r} \quad -a_{21} \quad \cdots \quad b_{n1} \quad \cdots \quad b_{nr}]^T \quad (16)$$

where $\{a_{ij}\}$, $\{b_{ij}\}$ correspond to the free model parameters. Accordingly, the variable selection involves filtering out the columns of the matrix Φ corresponding to the insignificant model terms, which can be carried out using LASSO resulting in several parameters a_{ij} and b_{ij} in (13) to be exactly zero. More formally, the LASSO regression can be given by

$$\hat{\theta}_{\text{LASSO}} = \arg \min_{\theta} \{ \|Y - \Phi\theta\|_2 + \lambda_1 \|\theta\|_1 \} \quad (17)$$

where λ_1 is the tuning parameter controlling the amount of shrinkage (or the regularization) via the l_1 -norm, and $\hat{\theta}_{\text{LASSO}}$ the LASSO estimate. Note that a suitable value of the parameter λ_1 can be determined using the k -fold cross-validation approach [41]. Furthermore, given $M' < M$ nonzero values in the obtained LASSO estimate $\hat{\theta}_{\text{LASSO}}$, the corresponding M' model terms can be regarded as the *selected* model terms. Unfortunately, several issues have been reported which render the LASSO estimate as unsuited for the final model estimate, namely

(1) the selection of noisy terms in the model structure [42], and (2) the large bias in the estimate $\hat{\theta}_{\text{LASSO}}$ due to constant shrinkage, which can affect the model prediction capabilities [43]. Accordingly, several LASSO variants have been proposed in the literature which augment the original LASSO with a second estimator, such as relaxed LASSO [42], adaptive LASSO [44], LARS-OLS hybrid [45], and LASSO-Ridge [46]. Herein, the ridge regression has been carried out following LASSO to mitigate the aforementioned issues, which has then been considered to yield the final model estimate. Formally, a *filtered* regression vector $\phi'_k : \mathbb{Z} \rightarrow \mathbb{R}^{M'}$ can be constructed using the selected model terms leading to the following model estimate,

$$\hat{\theta}_{\text{Ridge}} = \arg \min_{\theta} \{ \|Y - \Phi'\theta\|_2 + \lambda_2 \|\theta\|_2 \} \quad (18)$$

where λ_2 is the tuning parameter controlling the amount of shrinkage via the l_2 -norm, $\Phi' \in \mathbb{R}^{N \times M'}$ the collection of the regression vectors $\{\phi'_k\}_{k=1}^N$, and $\hat{\theta}_{\text{Ridge}}$ the final model estimate. Note that the parameter λ_2 can also be determined similarly to λ_1 using the cross-validation approach.

6. Experimental setup

In this paper, a non-commercial lithium-ion NMC/graphite pouch-shaped battery cell with 1 Ah rated capacity has been studied. The battery was kept in an isothermal environment at 25 °C using an Espec SU-642 environmental chamber having temperature control of 0.1 °C. In addition, the battery was connected simultaneously with a Delta Elektronika SM 30–200 power supply and an H&H ZS electronic load to either charge or discharge the battery, respectively. Furthermore, GITT experiments to characterize the battery EMF function V_{EMF} were carried out using 0.1 C current pulses corresponding to roughly 1% ΔSOC followed by adaptive rest periods, such that the variation in the battery terminal voltage ΔV at the end of each rest period was $\Delta V < 1$ mV/h. Moreover, the upper and lower voltage cutoff limits were set at 4.2 V and 2.5 V, respectively, during the application of both the identification as well as the test-input current profile. Finally, the battery current and voltage measurements were recorded with the sampling time $\tau = 1$ s.

7. Results and discussion

Several battery models with varying nonlinearity order l and model order n have been identified to generate a two-dimensional grid of the battery models $\mathcal{M}_{l,n}$ with the corresponding values $l, n \in \{1, 2, 3, 4\}$. Fig. 2 shows the current profile used to identify the models $\mathcal{M}_{l,n}$, which can be considered as a realization of the input signal design \mathcal{I} given in (14) with the associated parameters given as $\alpha_0 = 0.4$ A, $\mu_\tau = 64$ s, $\sigma_\tau = 32$ s, $\mu_{\tau_0} = 16$ s, $\sigma_{\tau_0} = 8$ s, $\sigma_\xi = 0.1$ A, $\alpha_+ = 1$, $\alpha_- = 0.1$, and $v = -0.2$ A. Note that the parameters have been chosen roughly keeping in mind the specifications of the studied battery and the arguments presented in Section 4. Subsequently, the voltage simulation capabilities of the identified models have been compared using a test dataset comprising of a real drive-cycle current profile as shown in Fig. 3. Table 1 summarizes the obtained results quantitatively using the root-mean-square error (RMSE) and the mean absolute error (MAE) corresponding to each model. In addition, the RMSE and MAE values for certain models $\{\mathcal{M}_{2,1}, \mathcal{M}_{2,2}, \mathcal{M}_{3,3}, \mathcal{M}_{3,4}, \mathcal{M}_{4,3}, \mathcal{M}_{4,4}\}$ have been highlighted in Table 1, whose corresponding simulated voltage curves are shown in Fig. 4. Note that the grey regions in Fig. 4 represent severely nonlinear battery operating regimes since the lower cutoff voltage limit had been set purposefully lower than usual to demonstrate the usefulness of the proposed model structure.

Several remarks can be given in light of the results presented in Table 1 and Fig. 4. It can be seen that the increase in the model order n tends to improve the model simulation performance in general, which can be attributed to the application of efficient model selection and estimation techniques to mitigate the multicollinearity issues in an ill-conditioned regression setting. Another important thing to consider

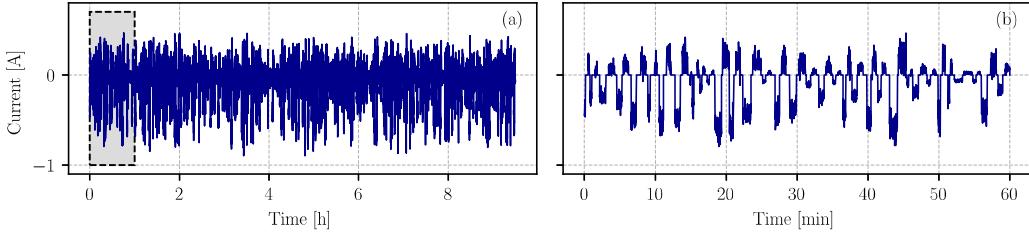


Fig. 2. A realization of the proposed input signal design I used to identify the battery models; (b) shows the highlighted portion of the current profile in (a).

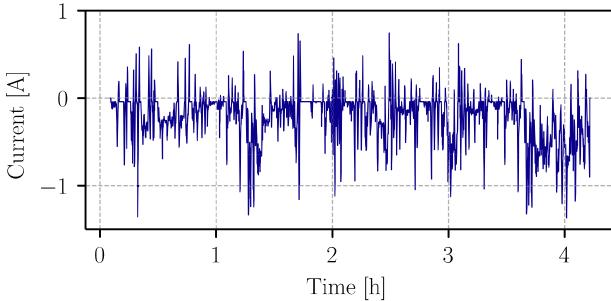


Fig. 3. A real drive-cycle current profile used as the test input to validate the identified battery models.

Table 1
Error description of the simulation results corresponding to the battery models $\mathcal{M}_{l,n}$ with model order n and nonlinearity order l .

	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$l = 1$	RMSE = 77.184 mV	155.066	127.852	38.205
	MAE = 37.104 mV	64.341	54.744	17.634
$l = 2$	56.588	36.179	62.526	493.292
	25.898	21.837	24.124	153.638
$l = 3$	42.702	66.963	3.320 × 10¹⁵	30.953
	26.821	27.137	3.656 × 10¹⁵	14.715
$l = 4$	59.737	6.012 × 10¹⁴	24.513	25.283
	45.691	1.178 × 10¹⁴	15.105	17.697

Table 2
Error description of the simulation results corresponding to the battery models $\mathcal{M}_{l,n}^{\text{poly}}$ with model order n and nonlinearity order l .

	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$l = 1$	RMSE = 5261.715 mV	5268.986	49565.242	2.934×10^7
	MAE = 1631.922 mV	1554.620	11172.700	5.505×10^6
$l = 2$	2.060×10^7	1551.968	6.836×10^{24}	2.323×10^{11}
	3.493×10^6	525.733	8.413×10^{23}	3.707×10^{10}
$l = 3$	6.055×10^{15}	1373.677	1.893×10^{22}	9.949×10^{29}
	6.410×10^{14}	473.604	2.393×10^{21}	1.183×10^{29}
$l = 4$	3.889×10^{17}	1.292×10^{12}	2.481×10^{14}	1.109×10^{30}
	3.931×10^{16}	1.862×10^{11}	3.400×10^{13}	1.325×10^{29}

is the possibility of the identification of unstable models which can diverge during the simulation, for instance, the models $\{\mathcal{M}_{2,4}, \mathcal{M}_{3,3}, \mathcal{M}_{4,2}\}$ in this case. A possible solution can be given by the restriction of basis functions for the autoregressive terms (the past output terms) to enforce stability during the model identification. Nevertheless, the stability of simulations has been deemed as an open area in system identification [33] and is not explored further in this paper.

The utility of selecting the basis functions based on physical insight can be further emphasized by identifying another set of battery models $\mathcal{M}_{l,n}^{\text{poly}}$ using conventional polynomial basis functions for explaining the SOC-dependence of the model parameters, similar to the approach adopted in [12]. More precisely, the models $\mathcal{M}_{l,n}^{\text{poly}}$ can be identified in a similar fashion as the models $\mathcal{M}_{l,n}$ using the same basis functions for the current magnitude $|\mu|$ and the current direction δ but with different basis functions for the SOC s , that is, $\{s\}$. Note that the

Table 3

Error description of the simulation results corresponding to the battery models $\mathcal{M}_{l,n}^{\text{poly,ridge}}$ with model order n and nonlinearity order l .

	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$l = 1$	RMSE = 26145.386 mV	3.707×10^5	5.249×10^5	3.628×10^5
	MAE = 5550.079 mV	62059.579	84531.335	58912.698
$l = 2$	17870.774	1257.378	1048.016	1663.901
	2891.048	272.868	203.783	306.649
$l = 3$	700.172	4536.609	10159.878	11076.394
	157.537	759.866	1662.790	1853.021
$l = 4$	194.707	764.849	1533.452	1869.786
	59.172	153.933	282.086	342.421

nonlinearity order l allows the generation of the higher-order SOC terms up to the degree l , for instance, the term s^3 can be generated using $l \geq 3$. Table 2 summarizes the simulation performance of the identified models quantitatively using the test input shown in Fig. 3, which can be directly compared with Table 1. It can be seen that the error values corresponding to the models $\mathcal{M}_{l,n}^{\text{poly}}$ are multiple orders of magnitude larger than that for the models $\mathcal{M}_{l,n}$ due to the inability of the terms s^m for $m \geq 1$ to sparsely explain the battery voltage behaviour. Furthermore, the increase in the nonlinearity order l and the model order n does not necessarily improve the error values but instead can result in unstable model behaviours. Indeed, a battery model containing several higher-order SOC terms of the form s^m for $m \geq 1$ can be identified in a non-sparse fashion (in this case, by skipping the LASSO regression), which may result in improved voltage simulation accuracy. Such an argument can be confirmed by identifying another set of models $\mathcal{M}_{l,n}^{\text{poly,ridge}}$ in a similar fashion as $\mathcal{M}_{l,n}^{\text{poly}}$ but employing only the ridge regression, thereby skipping the variable selection using LASSO. Table 3 summarizes the corresponding simulation performances of the models $\mathcal{M}_{l,n}^{\text{poly,ridge}}$ showing that (1) the corresponding error values are significantly lower than that for the models $\mathcal{M}_{l,n}^{\text{poly}}$, and (2) the increase in the nonlinearity order l generally improves the simulation performance. Moreover, it can be argued that further increase in the value of l beyond $l = 4$ may lead to additional improvements in the voltage simulation accuracy for the models $\mathcal{M}_{l,n}^{\text{poly,ridge}}$. However, based on our initial presumption about the principle of parsimony, the models identified in a non-sparse fashion may not be close to the true mathematical description of the battery as compared to a well-performing sparse model. Thus, it can be concluded that the usage of conventional polynomial basis functions to explain the SOC-dependence cannot lead to sparse battery models having appropriate simulation capabilities.

Another important contribution of this paper, namely the design of a synthetic identification input signal, can be analysed by comparing the models $\mathcal{M}_{l,n}$ with another set of battery models $\mathcal{M}_{l,n}^{\text{dc}}$ identified using a real drive-cycle current profile as considered a common practice in the literature [12,16,47]. Fig. 5 shows the drive-cycle current profile used to identify the models $\mathcal{M}_{l,n}^{\text{dc}}$. Table 4 summarizes the error values between the simulated and the measured voltage curves for the models $\mathcal{M}_{l,n}^{\text{dc}}$ using the test input shown in Fig. 3. Using Tables 1 and 4, it can be seen that the models $\mathcal{M}_{l,n}^{\text{dc}}$ perform quite poorly as compared to the models $\mathcal{M}_{l,n}$. In essence, the usage of real drive-cycle current profiles can be risky since the informativity of different drive-cycle datasets

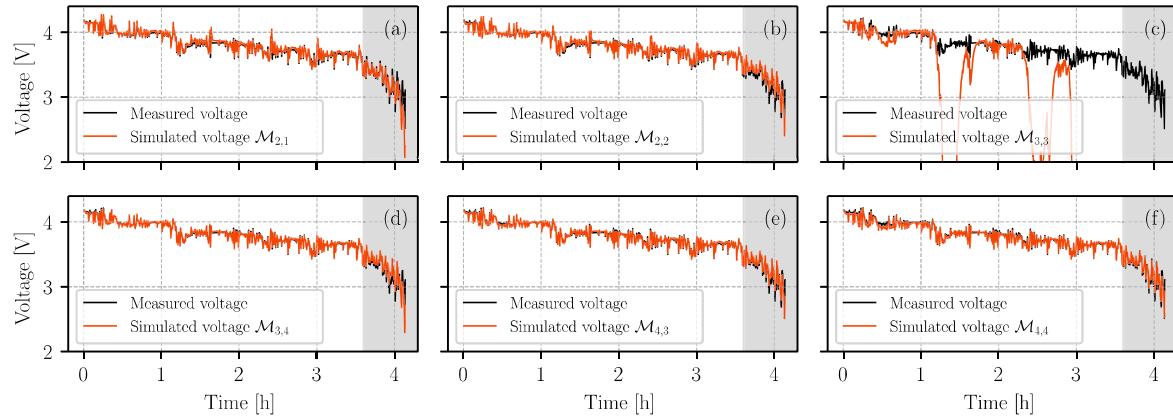


Fig. 4. Comparison of the simulated voltage curves with the measured voltage for select identified models; (a) $\mathcal{M}_{2,1}$ (b) $\mathcal{M}_{2,2}$ (c) $\mathcal{M}_{3,3}$ (d) $\mathcal{M}_{3,4}$ (e) $\mathcal{M}_{4,3}$ (f) $\mathcal{M}_{4,4}$.

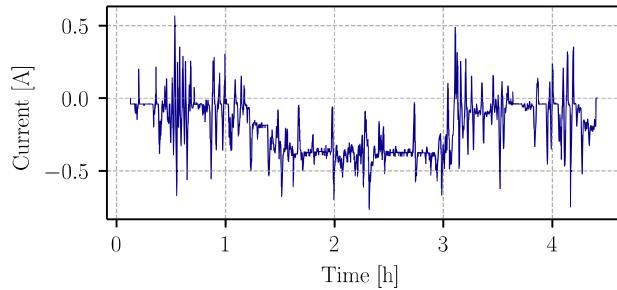


Fig. 5. A real drive-cycle current profile used to identify the battery models $\mathcal{M}_{l,n}^{dc}$.

Table 4
Error description of the simulation results corresponding to the battery models $\mathcal{M}_{l,n}^{dc}$ with model order n and nonlinearly order l .

	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$l = 1$	RMSE = 116.151 mV MAE = 71.951 mV	79.854 56.747	72.024 52.403	73.496 53.153
$l = 2$	140.283 76.056	86.832 53.422	4567.661 484.284	1067.718 178.074
$l = 3$	98.904 61.786	81.299 62.439	5.473 $\times 10^5$ 2.891 $\times 10^4$	69.088 49.228
$l = 4$	98.855 59.646	4.777 $\times 10^{16}$ 5.522 $\times 10^{15}$	7.153 $\times 10^{26}$ 7.072 $\times 10^{25}$	4.497 $\times 10^{31}$ 5.034 $\times 10^{30}$

can vary, and therefore, a sufficiently informative drive-cycle dataset is needed for a successful battery model identification attempt. Consequently, if a real drive-cycle dataset needs to be used for the model identification (and update) procedure, then the informative segments of the dataset should be selected instead of using the whole dataset *as is*. Alternatively, a synthetic signal design, such as the one proposed in this paper, that excites all the relevant dynamics can be used instead of a real drive-cycle dataset to obtain suitable battery models.

Finally, as hinted previously in Section 3.4, the translation of the identified LPV model into the ECM components requires diagonalization of the parameter-varying matrix $\mathcal{A}(p_k)$. Fortunately, contrary to the remaining ECM components, the feedthrough resistance (subsequently denoted with R_0) can be directly related to the scheduling variable p using an analytical relation, that is, $R_0(p_k) = \bar{D}(p_k) = D(p_k) = b_0(p_k)$. For instance, the functional relation for R_0 corresponding to the model $\mathcal{M}_{4,3}$ has been identified as follows

$$R_0(p_k) \approx 0.104 + 0.019s_k^{-1} + 0.058e^{\frac{1}{20}\sqrt{|u_k|}} + 0.019e^{\frac{1}{10}\sqrt{|u_k|}} - 0.012e^{\frac{3}{20}\sqrt{|u_k|}} - 0.035e^{\frac{1}{5}\sqrt{|u_k|}} + 0.001s_k^{-2} \ln(s_k)e^{\frac{1}{20}\sqrt{|u_k|}} \quad (19)$$

where R_0 does not exhibit dependence on the current-direction δ for the set of dynamics excited during the identification experiment. All in all, it can be remarked that a successful identification attempt has been demonstrated in this paper aided with suitable physical insights to yield sparse battery models. It must be emphasized again that the sparsity of an identified model may not always lead to a better approximation of the true system, and therefore should be accompanied with appropriate prediction/simulation capabilities to enhance the trust of the user on the identified model.

8. Conclusions

In this paper, a battery model structure has been proposed by extending the *archetypal* equivalent-circuit model (ECM) structure in the linear time-invariant (LTI) framework to the linear parameter-varying (LPV) framework. Subsequently, the need for a sufficiently informative experimental dataset has been emphasized in light of the widely accepted notion that the identification experiments should resemble the model application as much as possible, thereby rendering the model application as interpolation and not extrapolation. Furthermore, the principle of parsimony has been embraced to obtain the simplest possible battery model for a given set of dynamics using a model identification strategy that exploits the sparsity-inducing properties of the l_1 -norm penalization (LASSO). It has been shown that the increase in the model order generally enhances the simulation capabilities of the identified battery models — a result that intuitively makes sense but often remains contested in the literature. Finally, the proposed framework can be extended to investigate several other factors affecting the battery behaviour, namely temperature and various ageing phenomena, and will be considered for future studies.

CRediT authorship contribution statement

A.M.A. Sheikh: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M.C.F. Donkers:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **H.J. Bergveld:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The research data and the software code associated with this work can be found online in the form of an open-source Python package at <https://github.com/tue-battery/PyBatteryID>.

Acknowledgements

This work has received financial support from the Ministry of Economic Affairs and Climate, under the grant 'R&D Mobility Sectors' carried out by the Netherlands Enterprise Agency.

References

- [1] M. Doyle, T.F. Fuller, J. Newman, Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell, *J. Electrochem. Soc.* 140 (6) (1993) 1526, <http://dx.doi.org/10.1149/1.2221597>.
- [2] G. Ning, B.N. Popov, Cycle life modeling of lithium-ion batteries, *J. Electrochem. Soc.* 151 (10) (2004) A1584, <http://dx.doi.org/10.1149/1.1787631>.
- [3] A. Berlinsky, W. Unruh, W. McKinnon, R. Haering, Theory of lithium ordering in Li_xTi₂, *Solid State Commun.* 31 (3) (1979) 135–138, [http://dx.doi.org/10.1016/0038-1098\(79\)90421-6](http://dx.doi.org/10.1016/0038-1098(79)90421-6).
- [4] M.P. Mercer, S. Finnigan, D. Kramer, D. Richards, H.E. Hoster, The influence of point defects on the entropy profiles of lithium ion battery cathodes: a lattice-gas Monte Carlo study, *Electrochim. Acta* 241 (2017) 141–152, <http://dx.doi.org/10.1016/j.electacta.2017.04.115>.
- [5] T.R. Ferguson, M.Z. Bazant, Nonequilibrium thermodynamics of porous electrodes, *J. Electrochem. Soc.* 159 (12) (2012) A1967, <http://dx.doi.org/10.1149/2.048212jes>.
- [6] R.B. Smith, M.Z. Bazant, Multiphase porous electrode theory, *J. Electrochem. Soc.* 164 (11) (2017) E3291, <http://dx.doi.org/10.1149/2.0171711jes>.
- [7] D. Jiani, L. Zhitao, W. Youyi, W. Changyun, A fuzzy logic-based model for Li-ion battery with SOC and temperature effect, in: 11th IEEE International Conference on Control & Automation, ICCA, 2014, pp. 1333–1338, <http://dx.doi.org/10.1109/ICCA.2014.6871117>.
- [8] Q.-K. Wang, Y.-J. He, J.-N. Shen, Z.-F. Ma, G.-B. Zhong, A unified modeling framework for lithium-ion batteries: An artificial neural network based thermal coupled equivalent circuit model approach, *Energy* 138 (2017) 118–132, <http://dx.doi.org/10.1016/j.energy.2017.07.035>.
- [9] R.R. Richardson, C.R. Birk, M.A. Osborne, D.A. Howey, Gaussian process regression for in situ capacity estimation of lithium-ion batteries, *IEEE Trans. Ind. Inform.* 15 (1) (2019) 127–138, <http://dx.doi.org/10.1109/TII.2018.2794997>.
- [10] A. Lasia, *Electrochemical Impedance Spectroscopy and Its Applications*, Springer, 2002, <http://dx.doi.org/10.1007/978-1-4614-8933-7>.
- [11] J.P. Schmidt, P. Berg, M. Schönleber, A. Weber, E. Ivers-Tiffée, The distribution of relaxation times as basis for generalized time-domain models for Li-ion batteries, *J. Power Sources* 221 (2013) 70–77, <http://dx.doi.org/10.1016/j.jpowsour.2012.07.100>.
- [12] F. Hoekstra, M. Donkers, H. Bergveld, Rapid empirical battery electromotive-force and overpotential modelling using input–output linear parameter-varying methods, *J. Energy Storage* 65 (2023) 107185, <http://dx.doi.org/10.1016/j.est.2023.107185>.
- [13] B. Yann Liaw, G. Nagasubramanian, R.G. Jungst, D.H. Doughty, Modeling of lithium ion cells—A simple equivalent-circuit model approach, *Solid State Ion.* 175 (1) (2004) 835–839, <http://dx.doi.org/10.1016/j.ssi.2004.09.049>, Fourteenth International Conference on Solid State Ionics.
- [14] D. Andre, M. Meiler, K. Steiner, H. Walz, T. Soczka-Guth, D. Sauer, Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. II: Modelling, *J. Power Sources* 196 (12) (2011) 5349–5356, <http://dx.doi.org/10.1016/j.jpowsour.2010.07.071>.
- [15] P. Iurilli, C. Brivio, R.E. Carrillo, V. Wood, EIS2MOD: A DRT-based modeling framework for Li-ion cells, *IEEE Trans. Ind. Appl.* 58 (2) (2022) 1429–1439, <http://dx.doi.org/10.1109/TIA.2021.3134946>.
- [16] H. Beelen, H. Bergveld, M. Donkers, On experiment design for parameter estimation of equivalent-circuit battery models, in: 2018 IEEE Conference on Control Technology and Applications, CCTA, 2018, pp. 1526–1531, <http://dx.doi.org/10.1109/CCTA.2018.8511529>.
- [17] R. Tóth, Modeling and identification of linear parameter-varying systems, Vol. 403, Springer, 2010, <http://dx.doi.org/10.1007/978-3-642-13812-6>.
- [18] A. Fotouhi, K. Propp, D.J. Auger, Electric vehicle battery model identification and state of charge estimation in real world driving cycles, in: 2015 7th Computer Science and Electronic Engineering Conference, CEEC, 2015, pp. 243–248, <http://dx.doi.org/10.1109/CEEC.2015.7332732>.
- [19] Y. Hu, S. Yurkovich, Linear parameter varying battery model identification using subspace methods, *J. Power Sources* 196 (5) (2011) 2913–2923, <http://dx.doi.org/10.1016/j.jpowsour.2010.10.072>.
- [20] A. Nikolian, J. Jaguemont, J. de Hoog, S. Goutam, N. Omar, P. Van Den Bossche, J. Van Mierlo, Complete cell-level lithium-ion electrical ECM model for different chemistries (NMC, LFP, LTO) and temperatures (−5 °C to 45 °C) – Optimized modelling techniques, *Int. J. Electr. Power Energy Syst.* 98 (2018) 133–146, <http://dx.doi.org/10.1016/j.ijepes.2017.11.031>.
- [21] J. Mairal, F. Bach, J. Ponce, Sparse modeling for image and vision processing, *Found. Trends® Comp. Graph. Vis.* 8 (2–3) (2014) 85–283, <http://dx.doi.org/10.1561/0600000058>.
- [22] D.W. D.Sc., H.J.M. D.Sc., XLII. On certain fundamental principles of scientific inquiry, *The London, Edinb., and Dublin Philos. Mag. and J. Sci.* 42 (249) (1921) 369–390, <http://dx.doi.org/10.1080/14786442108633773>.
- [23] F. Hoekstra, L. Rajmakers, M. Donkers, H. Bergveld, Comparison of battery electromotive-force measurement and modelling approaches, *J. Energy Storage* 56 (2022) 105910, <http://dx.doi.org/10.1016/j.est.2022.105910>.
- [24] X. Lin, H.E. Perez, S. Mohan, J.B. Siegel, A.G. Stefanopoulou, Y. Ding, M.P. Castanier, A lumped-parameter electro-thermal model for cylindrical batteries, *J. Power Sources* 257 (2014) 1–11, <http://dx.doi.org/10.1016/j.jpowsour.2014.01.097>.
- [25] J.P. Christoffersen, G.L. Hunt, C.D. Ho, D. Howell, Pulse resistance effects due to charging or discharging of high-power lithium-ion cells: A path dependence study, *J. Power Sources* 173 (2) (2007) 998–1005, <http://dx.doi.org/10.1016/j.jpowsour.2007.08.025>, X Polish Conference on Systems with Fast Ionic Transport.
- [26] F. Hoekstra, Y. Heuts, H. Bergveld, M. Donkers, Empirical battery modelling for high currents: The effect of nonlinear overpotential and inevitable self-heating, *IFAC-PapersOnLine* 53 (2) (2020) 12440–12445, <http://dx.doi.org/10.1016/j.ifacol.2020.12.1323>, 21st IFAC World Congress.
- [27] A. Sheikh, M. Donkers, H. Bergveld, Investigating identification input designs for modelling lithium-ion batteries with hysteresis using LPV framework, in: 2024 American Control Conference, ACC, IEEE, 2024.
- [28] T.-T. Nguyen, A.B. Khan, Y. Ko, W. Choi, An accurate state of charge estimation method for lithium iron phosphate battery using a combination of an unscented Kalman filter and a particle filter, *Energies* 13 (17) (2020) <http://dx.doi.org/10.3390/en13174536>.
- [29] R. Toth, J.C. Willem, P.S.C. Heuberger, P.M.J. Van den Hof, The behavioral approach to linear parameter-varying systems, *IEEE Trans. Autom. Control* 56 (11) (2011) 2499–2514, <http://dx.doi.org/10.1109/TAC.2011.2109439>.
- [30] Y. Hu, S. Yurkovich, Battery cell state-of-charge estimation using linear parameter varying system techniques, *J. Power Sources* 198 (2012) 338–350, <http://dx.doi.org/10.1016/j.jpowsour.2011.09.058>.
- [31] R. Toth, H.S. Abbas, H. Werner, On the state-space realization of LPV input-output models: Practical approaches, *IEEE Trans. Control Syst. Technol.* 20 (1) (2012) 139–153, <http://dx.doi.org/10.1109/TCST.2011.2109040>.
- [32] P.V.D. Kloet, F.L. Neerhoff, Diagonalization algorithms for linear time-varying dynamic systems, *Int. J. Syst. Sci.* 31 (8) (2000) 1053–1057, <http://dx.doi.org/10.1080/002077200412195>.
- [33] L. Ljung, Perspectives on system identification, *Annu. Rev. Control* 34 (1) (2010) 1–12, <http://dx.doi.org/10.1016/j.arcontrol.2009.12.001>.
- [34] P. Royston, D.G. Altman, Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 43 (3) (1994) 429–467, <http://dx.doi.org/10.2307/2986270>.
- [35] Z.-D. Yuan, L. Ljung, Unprejudiced optimal open loop input design for identification of transfer functions, *Automatica* 21 (6) (1985) 697–708, [http://dx.doi.org/10.1016/0005-1098\(85\)90043-3](http://dx.doi.org/10.1016/0005-1098(85)90043-3).
- [36] M. Gevers, L. Ljung, Optimal experiment designs with respect to the intended model application, *Automatica* 22 (5) (1986) 543–554, [http://dx.doi.org/10.1016/0005-1098\(86\)90064-6](http://dx.doi.org/10.1016/0005-1098(86)90064-6).
- [37] W. Widanage, A. Barai, G. Chouchelamane, K. Uddin, A. McGordon, J. Marco, P. Jennings, Design and use of multisine signals for Li-ion battery equivalent circuit modelling. Part 1: Signal design, *J. Power Sources* 324 (2016) 70–78, <http://dx.doi.org/10.1016/j.jpowsour.2016.05.015>.
- [38] J. Ding, V. Tarokh, Y. Yang, Model selection techniques: An overview, *IEEE Signal Process. Mag.* 35 (6) (2018) 16–34, <http://dx.doi.org/10.1109/MSP.2018.2867638>.
- [39] T. Hastie, R. Tibshirani, R. Tibshirani, Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons, *Statist. Sci.* 35 (4) (2020) 579–592, <http://dx.doi.org/10.1214/19-STS733>.
- [40] J.P. Dahlgren, Alternative regression methods are not considered in Murtough (2009) or by ecologists in general, *Ecol. Lett.* 13 (5) (2010) E7–E9, <http://dx.doi.org/10.1111/j.1461-0248.2010.01460.x>.
- [41] J. Ranstam, J.A. Cook, LASSO regression, *Br J Surg* 105 (10) (2018) 1348, <http://dx.doi.org/10.1002/bjs.10895>.
- [42] N. Meinshausen, Relaxed lasso, *Comput. Statist. Data Anal.* 52 (1) (2007) 374–393, <http://dx.doi.org/10.1016/j.csda.2006.12.019>.
- [43] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.* 96 (456) (2001) 1348–1360, <http://dx.doi.org/10.1198/016214501753382273>.
- [44] H. Zou, The adaptive lasso and its oracle properties, *J. Amer. Statist. Assoc.* 101 (476) (2006) 1418–1429, <http://dx.doi.org/10.1198/016214506000000735>.
- [45] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Statist.* 32 (2) (2004) 407–499, <http://dx.doi.org/10.1214/009053604000000067>.

- [46] H. Liu, B. Yu, Asymptotic properties of Lasso+mLS and Lasso+Ridge in sparse high-dimensional linear regression, *Electron. J. Stat.* 7 (2013) 3124–3169, <http://dx.doi.org/10.1214/14-EJS875>.
- [47] M. Andersson, M. Johansson, V.L. Klass, A continuous-time LPV model for battery state-of-health estimation using real vehicle data, in: 2020 IEEE Conference on Control Technology and Applications, CCTA, 2020, pp. 692–698, <http://dx.doi.org/10.1109/CCTA41146.2020.9206257>.