Journal of The Electrochemical Society, 2023 **170** 120537

A Grey-box Model with Neural Ordinary Differential Equations for the Slow Voltage Dynamics of Lithium-ion Batteries: Model Development and Training

Jennifer Brucker,[z] Wolfgang G. Bessler,[*] and Rainer Gasper

Institute of Sustainable Energy Systems (INES), Offenburg University of Applied Sciences, 77652 Offenburg, Germany

Lithium-ion batteries exhibit slow voltage dynamics on the minute time scale that are usually associated with transport processes. We present a novel modelling approach toward these dynamics by combining physical and data-driven models into a Grey-box model. We use neural networks, in particular neural ordinary differential equations. The physical structure of the Grey-box model is borrowed from the Fickian diffusion law, where the transport domain is discretized using finite volumes. Within this physical structure, unknown parameters (diffusion coefficient, diffusion length, discretization) and dependencies (state of charge, lithium concentration) are replaced by neural networks and learnable parameters. We perform model-to-model comparisons, using as training data (a) a Fickian diffusion process, (b) a Warburg element, and (c) a resistor-capacitor circuit. Voltage dynamics during constant-current operation and pulse tests as well as electrochemical impedance spectra are simulated. The slow dynamics of all three physical models in the order of ten to 30 min are well captured by the Grey-box model, demonstrating the flexibility of the present approach.
 [DOI: 10.1149/1945-7111/ad14cd]

Manuscript submitted August 2, 2023; revised manuscript received November 22, 2023. Published December 28, 2023.

List of Symbol

Symbol	Unit	Meaning
a	V	proportionality factor
A_e	m^2	active surface area
b	$A^{-1}\,s^{-1}$	proportionality factor
c	mol m^{-3}	(lithium) concentration
$\bar{c}$	mol m^{-3}	average (lithium) concentration
C	—	nondimensionalized (lithium) concentration
C_1	F	conductance
C_{bat}	A h	battery capacity
$\bar{c}_i$	mol m^{-3}	average (lithium) concentration in discretization element i
$\bar{C}_i$	—	average nondimensionalized (lithium) concentration in discretization element i
$\bar{c}_{i\,0}$	mol m^{-3}	initial average (lithium) concentration in discretization element i
$\bar{C}_{i\,0}$	—	initial nondimensionalized average (lithium) concentration in discretization element i
c_{init}	mol m^{-3}	initial (lithium) concentration
C_{init}	—	initial nondimensionalized (lithium) concentration
$c_{Li\ act}$	mol m^{-3}	actual lithium concentration
$c_{Li\ eq}$	mol m^{-3}	equilibrium lithium concentration at the actual operation point
c_{max}	mol m^{-3}	(lithium) concentration of a fully lithiated particle
c_{min}	mol m^{-3}	(lithium) concentration of a fully delithiated particle
C_N	A h	nominal battery capacity
c_S	mol m^{-3}	(lithium) surface concentration
C_S	—	nondimensionalized (lithium) surface concentration
$C_{S\ GB}$	—	predicted nondimensionalized (lithium) surface concentration
$C_{S\ true}$	—	true nondimensionalized (lithium) surface concentration
$c_{V\ act}$	mol m^{-3}	actual concentration of vacancies
$c_{V\ eq}$	mol m^{-3}	equilibrium concentration of vacancies at the actual operation point
D	$m^2\,s^{-1}$	diffusion coefficient
D^*	s^{-1}	partially nondimensionalized diffusion coefficient
f	—	neural network
F	C mol^{-1}	Faraday constant
f^*	—	scaled neural network
g_D	—	nonlinear function
i	—	index of discretization elements
I_{bat}	A	(battery) current
j	mol $m^{-2}\,s^{-1}$	molar flux
j_i	mol $m^{-2}\,s^{-1}$	molar flux across the right boundary of discretization element i
j_{influx}	mol $m^{-2}\,s^{-1}$	molar flux across the particle surface
k	—	number of inputs to a neural network layer

*Electrochemical Society Member.
[z]E-mail: jennifer.brucker@hs-offenburg.de

(Continued).

l	m	length of the cuboid particle
L	—	loss function
MSE	—	mean squared error
n	—	charge number
N	—	number of discretization elements
r	m	length along the radius
R	m	radius of the spherical particle
R_1	Ω	ohmic resistance
R_g	J mol^{-1} K^{-1}	universal gas constant
r_i	m	position of the right boundary of discretization element i
SOC	—	state of charge
t	s	time
t	—	index of layers of a neural network
T	K	temperature
T	—	number of layers of a neural network
$\boldsymbol{u}$	*	vector of external variables
$\mathcal{U}$	—	uniform distribution
V	m^3	volume
V_{diff}	V	voltage resulting from diffusion
$V_{diff\ GB}$	V	predicted diffusion voltage
$V_{diff\ true}$	V	true diffusion voltage
V_{final}	V	final voltage drop of the RC circuit following a current step
V_i	m^3	(scaled) volume of discretization element i
V_{RC}	V	voltage drop across the RC circuit
x	m	length along the diffusion path
x_i	m	position of the right boundary of discretization element i
z	—	nondimensionalized length
$\boldsymbol{z}_0$	*	vector of input states
Z_{diff}	Ω	diffusion impedance
z_i	—	nondimensionalized position of the right boundary of discretization element i
$\boldsymbol{z}_t$	*	vector of states at layer t of a neural network
α_i	A^{-1} s^{-1} or—	learnable parameter
α_i^*	A^{-1} s^{-1} or—	scaled learnable parameter
δ	s^{-1}	partially nondimensionalized molar flux
δ_i	s^{-1}	partially nondimensionalized molar flux across the right boundary of discretization element i
γ_i	s^{-1}	grey-box model parameter
ε	—	volume fraction
η	—	learning rate
$\boldsymbol{\theta}$	*	vector of weights and biases of a neural network
$\boldsymbol{\theta}_f$	—	vector of weights and biases of the neural network $\boldsymbol{f}$
$\boldsymbol{\theta}_{f^*}$	—	vector of weights and biases of the neural network $\boldsymbol{f}^*$
$\boldsymbol{\theta}_t$	*	vector of weights and biases of layer t of a neural network
τ	s	time constant of the RC circuit
ω	s^{-1}	angular frequency
ω_1	V	learnable parameter
ω_1^*	V	scaled learnable parameter

* depending on the quantity considered

Lithium-ion batteries are complex dynamic systems. They exhibit current-voltage dynamics on multiple time scales, from milliseconds to hours.[1] This can be observed both in the time domain (e.g., during a pulse test[2]) and in the frequency domain (i.e., in electrochemical impedance spectra).[3] While "fast" dynamics (up to several seconds) are well-understood today,[4] "slow" dynamics on long timescales (typically one minute and beyond) have only recently become subject of interest. For example, slow relaxation times become problematic when trying to measure the true open-circuit voltage (OCV).[5–8] Understanding the slow dynamics is challenging because they have multiple potential causes, including transport processes at the microscale (solid-state diffusion within active material particles[9]), mesoscale (liquid-phase diffusion of lithium ions in the electrolyte[10]), and macroscale (thermal gradients,[11] lithium concentration gradients in the anode overhang[12]).

Modelling and simulation have proven powerful tools to support understanding of lithium-ion batteries.[13] Different model categories can be distinguished. White-box (WB) modelling uses prior physical, chemical or engineering knowledge in the form of mathematical equations to describe the behavior of the system under consideration. It is therefore limited to an understanding of the underlying processes. In contrast, black-box (BB) models learn relationships between inputs and outputs of systems without any physical knowledge. They therefore require a large amount of training data. Neural networks are an important representative of BB models. So-called Grey-box (GB) models combine WB and BB modelling techniques to take advantage of their respective strengths. They use prior physical knowledge in combination with parametric parts for unknown or imprecisely-known parameters and relationships. This results in a reduction in the amount of data required for training.[14–17]

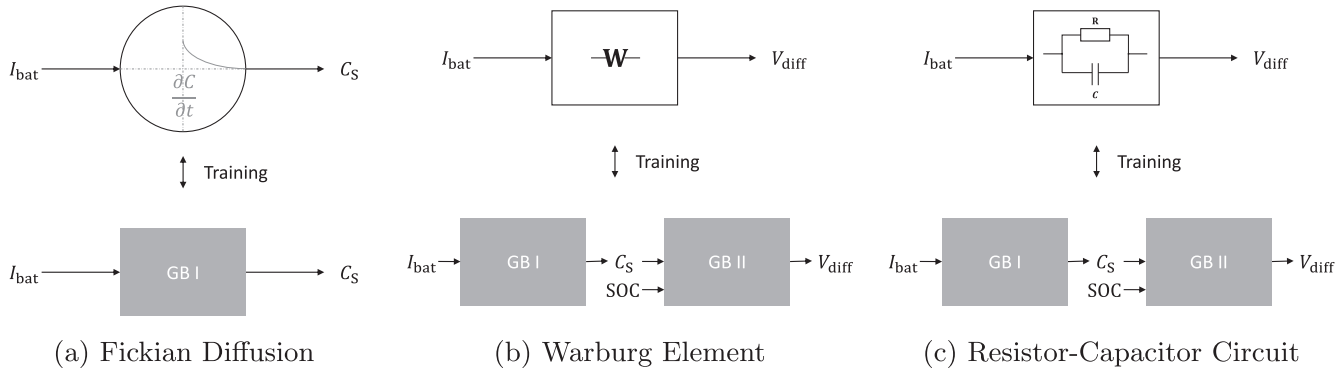

Figure 1. Simulation and training approach. The present study consists of three parts: (a) concentration dynamics based on Fickian diffusion model, (b) voltage dynamics resulting from a Warburg element, (c) voltage dynamics resulting from a resistor-capacitor circuit. Goal is the prediction of the slow dynamics of the diffusion voltage V_{diff}.

A typical WB model used for slow dynamics in lithium-ion batteries is Fickian diffusion, which is used in physicochemical models of transport in the solid or liquid phases.[18,19] Numerical time-domain simulation of a Fickian diffusion process requires spatial discretization techniques. A different setting, still WB but using a more global viewpoint, are equivalent circuit model (ECMs). Translated to equivalent circuits, Fickian diffusion leads to Warburg type elements.[20] These belong to the class of constant phase elements (CPEs). Although easily calculated in the frequency domain, time-domain simulation of Warburg elements gives rise to a differential equation of fractional order,[21] which is difficult to solve numerically. Therefore, instead of explicitly using Warburg elements in ECMs, various workarounds have been proposed. The authors of Ref. 22 used multiple resistor-capacitor (RC) elements to model a CPE. The parallel connection of $m \in \mathbb{N}$ series RC branches with recursive parameters served as the basis for the model. They complemented this basic RC model with an additional conductance for phase correction at low frequencies and an additional capacitance for phase correction at high frequencies. In Ref. 21 an ECM was built with an ohmic resistor, two CPEs and a Warburg element. The authors used the Grunwald-Letnikov[23] definition to discretize the resulting fractional equations. To keep the computation time short, they only considered the states from the previous time step instead of all previous data in the discretized equations. The authors of Ref. 8 found a linear relationship between the open-circuit time and the time constant of a second-order RC model. Therefore, they introduced a new voltage relaxation model with a time-varying relaxation time. This could reduce the waiting time during OCV measurements. However, this approach is only applicable to rest phases when no current is being drawn. In Ref. 24 an ECM with multiple RC circuits was supplemented with a diffusion element. Fick's law was used as a starting point for the derivation of the diffusion element. The authors considered the diffusion overpotential of a single current step with a pulse duration significantly shorter than the diffusion time constant. The simplification of the diffusion equation from Ref. 25 was used to calculate the diffusion overpotential. In contrast, the authors of Ref. 26 extended an ECM with a simple diffusion model for solid-state diffusion in the active material particles. The particle consists of two regimes with individual capacity in this core-shell model. The chosen approach is a simplified version of a single particle model[27–29] with only two discrete volumes. In conclusion, there are many different approaches to model Fickian diffusion and the resulting Warburg-type behavior of lithium-ion batteries. The matter gets further complicated by the fact that the experimentally-observed dynamics are often not pure Warburg.[30]

Because of the multiple possible causes and often ill-defined nature of the slow voltage dynamics, as well as difficulties in numerical simulation of Warburg-type elements, we introduce here a novel GB modelling approach. In particular, we use neural ordinary differential equations (NODEs)[31] for a GB model embedded in an ECM. The physical structure of the GB model is based on the Fickian diffusion law, where the transport domain is discretized using finite volumes. Within this physical structure, unknown parameters (e.g., diffusion coefficient, diffusion length, discretization) and dependencies (e.g., state of charge (SOC), lithium concentration) are replaced by learnable parameters and neural networks. This combination has a physical basis but at the same time provides sufficient flexibility for describing non-ideal behavior.

An overview of our modelling and training approach is shown in Fig. 1. Our study consists of three parts. We first build a GB model for concentration dynamics due to Fickian diffusion, which we train with numerical simulation results of a WB diffusion model (Fig. 1a). We then extend this model by a description of a concentration-voltage relationship, allowing to predict the diffusion voltage. This extended model is first trained to a Warburg element simulated in the time domain (Fig. 1b). Finally, in order to demonstrate the flexibility of the approach, the same GB model is trained to an RC circuit (Fig. 1c). All three WB models are parameterized such that they show dynamics in the range of ten to 30 min.

The paper is organized as follows. In the following section we give a short introduction to WB modelling of Fickian diffusion. Furthermore, we show how to model a Warburg element and an RC circuit. Afterwards, we introduce the GB modelling approach for Fickian diffusion. First, we show how to model the concentration dynamics of lithium in a spherical active material particle. Second, we extend the approach to model the resulting voltage drop. We perform model-to-model training and tests. Finally, the results are presented and discussed. At the end of the paper we summarize the results and give an outlook.

Physical Basis—White-Box Modelling

In this section we present the physics-based models (WB models) used as the structural basis of the GB model, and also used for simulating training and test data. We first present the Fickian diffusion model. Next, we give a brief overview of the Warburg element and the modelling of the resulting voltage dynamics. Finally, we show how to model the voltage dynamics of an RC circuit.

Fickian diffusion.— *Model.*—There are various diffusive transport mechanisms taking place in a lithium-ion battery, including solid-state diffusion in the active material particles of the electrodes, liquid-phase diffusion of lithium ions in the electrolyte, or solid-state diffusion of lithium into (or out of) the anode overhang. The spatiotemporal change of concentrations due to diffusion is generally

described by Fick's second law:[32]

$$\frac{\partial c}{\partial t} = \nabla(D\nabla c) \quad [1]$$

where, c is the concentration (e.g., lithium concentration inside the active materials), D is the diffusion coefficient, and ∇ is the Nabla operator. Fick's first law describes the species flux as

$$j = -D\nabla c \quad [2]$$

where, j is the molar flux.

In the following, we assume solid-state diffusion within active material particles as main diffusive transport mechanism (cf. Figure 1a). For simplicity, the active material particles of lithium-ion batteries are often assumed to be spherical. We use here a so-called single-particle model, where the electrodes are assumed to consist of a single spherical particle with radius R[27,28] and the externally-applied battery current I_{bat} is scaled to this single particle.

In spherical coordinates, the diffusion Eq. 1 simplifies to

$$\frac{\partial c}{\partial t} = \frac{1}{r^2}\cdot\frac{\partial}{\partial r}\left(Dr^2\cdot\frac{\partial c}{\partial r}\right), \quad 0 \leqslant r \leqslant R \quad [3]$$

where, r is the radial position.

Electrochemical charge-transfer reactions take place at the surface of the particle. A (de-)intercalation process with a given flux through the particle surface results in the following initial and boundary conditions:[33–36]

$$c = c_{\text{init}}, \quad 0 \leqslant r \leqslant R,\ t = 0 \quad [4a]$$

$$\frac{\partial c}{\partial r} = 0, \quad r = 0,\ t \geqslant 0 \quad [4b]$$

$$-D\frac{\partial c}{\partial r} = j_{\text{influx}}, \quad r = R,\ t \geqslant 0 \quad [4c]$$

where, c_{init} denotes the initial concentration, and[37]

$$j_{\text{influx}} = \frac{R}{3\varepsilon FV}\cdot I_{\text{bat}} \quad [5]$$

is the molar flux across the surface of the sphere. Here, ε is the volume fraction of the active material, F is the Faraday constant, and V is the volume of the electrode. Note that the particle center is at $r = 0$ and the particle surface at $r = R$.

Discretization and nondimensionalization.—Spatial discretization of the diffusion Eq. 3 in spherical coordinates with the finite-volume method leads to the following expression:[38]

$$\frac{\partial \bar{c}_i}{\partial t}\cdot V_i = D\left(\frac{\bar{c}_{i+1}+\bar{c}_i}{2}\right)\cdot r_{i+1}^2\cdot\frac{\bar{c}_{i+1}-\bar{c}_i}{(r_{i+2}-r_i)/2} - D\left(\frac{\bar{c}_i+\bar{c}_{i-1}}{2}\right)\cdot r_i^2\cdot\frac{\bar{c}_i-\bar{c}_{i-1}}{(r_{i+1}-r_{i-1})/2} \quad \forall\, i = 1,\ldots,N-2 \quad [6]$$

where, $N \in \mathbb{N}$ is the number of spherical shells chosen, $V_i = \frac{1}{3}(r_{i+1}^3 - r_i^3)$ the scaled volume of the shell between $[r_i, r_{i+1}]$, and $\bar{c}_i$ the average concentration within its volume. Note that the factor 4π in the usual volume term $\frac{4}{3}\pi(r_{i+1}^3 - r_i^3)$ is cancelled by the same factor in the surface.[38] Taking into account the initial and boundary conditions according to Eqs. 4 and 5, we can reformulate Eq. 6 as:

$$\frac{\partial \bar{c}_i}{\partial t} = 3\frac{j_{i+1}r_{i+1}^2 - j_i r_i^2}{r_{i+1}^3 - r_i^3}, \quad 0 \leqslant i \leqslant N-1 \quad [7a]$$

$$j_0 = 0 \quad [7b]$$

$$j_i = 2\frac{D\left(\frac{\bar{c}_i+\bar{c}_{i-1}}{2}\right)\cdot(\bar{c}_i - \bar{c}_{i-1})}{r_{i+1}-r_{i-1}}, \quad 1 \leqslant i \leqslant N-1 \quad [7c]$$

$$j_N = -\frac{R}{3\varepsilon FV}\cdot I_{\text{bat}} \quad [7d]$$

$$\bar{c}_{i\,0} = c_{\text{init}}, \quad 0 \leqslant i \leqslant N-1,\ t = 0 \quad [7e]$$

where, $\bar{c}_{i\,0}$ is the initial concentration in the shell i.

The finite-volume method gives the average concentration in each discretization volume. To approximate the surface concentration c_S, a linear extrapolation to the particle surface is recommended:[38]

$$c_S = \frac{3\bar{c}_{N-1} - \bar{c}_{N-2}}{2}. \quad [8]$$

To minimize the influence of rounding errors, we (partially) nondimensionalize the equations.[34–36] For this purpose, the following (partially) nondimensionalized quantities and functions are introduced:

$$C = \frac{c - c_{\min}}{c_{\max} - c_{\min}};\ z = \frac{r}{R};\ \delta = \frac{j}{R\cdot(c_{\max} - c_{\min})};\quad D^*(C) = \frac{D(c)}{R^2} \quad [9]$$

where, $c_{\max}$ and $c_{\min}$ are the (average) concentrations when the particle is fully lithiated or delithiated, respectively. In the single-particle model this can be interpreted as the battery being fully charged or discharged, respectively.

The following relationships result from Eqs. 7 and 8 with the partially nondimensionalized quantities according to Eq. 9:

$$\frac{\partial \bar{C}_i}{\partial t} = 3\frac{\delta_{i+1}z_{i+1}^2 - \delta_i z_i^2}{z_{i+1}^3 - z_i^3}, \quad 0 \leqslant i \leqslant N-1 \quad [10a]$$

$$\delta_0 = 0 \quad [10b]$$

$$\delta_i = 2\frac{D^*\left(\frac{\bar{C}_i+\bar{C}_{i-1}}{2}\right)\cdot(\bar{C}_i - \bar{C}_{i-1})}{z_{i+1}-z_{i-1}}, \quad 1 \leqslant i \leqslant N-1 \quad [10c]$$

$$\delta_N = -\frac{1}{3\varepsilon FV\cdot(c_{\max} - c_{\min})}\cdot I_{\text{bat}} \quad [10d]$$

$$\bar{C}_{i\,0} = C_{\text{init}}, \quad 0 \leqslant i \leqslant N-1 \quad [10e]$$

$$C_S = \frac{3\bar{C}_{N-1} - \bar{C}_{N-2}}{2}. \quad [10f]$$

In the single-particle setting used here, the average nondimensionalized concentration $\bar{C}$ represents the SOC of the electrode.

Equations 10 are the final result of this section. We used these equations to generate training data for GB modelling and also as a basis for the GB model of Fickian diffusion that will be presented below.

Parameterization.—We chose to use a representative graphite electrode particle as basis for this study. We used parameters representing the lithium iron phosphate (LFP) cell from the Chinese manufacturer CALB, model CA180FI, which is described

in detail in Ref. 39. According to the manufacturer, the cell has a nominal capacity of $C_N = 180$ A h. The graphite electrode has a volume of $V = 7.203 \cdot 10^{-4}$ m^3 and a volume fraction of $\varepsilon = 0.554$ at 50% SOC.[39] The diffusion coefficient of lithium in graphite has been observed to strongly depend on the lithium concentration, spanning more than an order of magnitude.[40] For the present study, we decided to use a value of $D = 3.9 \cdot 10^{-14}$ m^2 s^{-1} taken from Ref.,[41] and multiplied this value with a concentration-dependent parabolic function $g_D = -3.6 \cdot (C - 0.5)^2 + 1$ ranging between values of 0.1 (for $C = 0$ and 1) and 1 (for $C = 0.5$). This generic function allows to assess the ability of the GB model to represent nonlinear diffusion coefficients. For the graphite electrode of the cell in Ref. 42, the maximum lithium concentration is $3.1833 \cdot 10^4$ mol m^{-3}. For simplicity we assumed that $c_{max} - c_{min} = 3 \cdot 10^4$ mol m^{-3}. The particle radius is $R = 1.25 \cdot 10^{-5}$ m.[42]

Simulation.—Equations 10 were implemented in Python (version 3.7.6) to generate training and test data. We used the odeint method from the torchdiffeq library (version 0.2.3)[43] to solve the differential equations with the Dopri5 method. The radius of the spherical particle was discretized in 100, 20 and 10 equidistant sections for comparison. Note that for the GB model (see below) we only chose 5 discretization elements. Extrapolation of the concentration according to Eq. 10f gave the surface concentration.

We simulated the concentration dynamics for different battery currents and different initial concentrations. We used eight different time series as input for the training data, representing eight different virtual experiments with the cell. Specifically, we simulated constant-current charge (i.e., lithiation, starting from a fully delithiated particle) and discharge (i.e., delithiation, starting from a fully lithiated partice) with absolute external currents I_{bat} of 18 A, 50 A and 180 A. Additionally, we simulated pulsed charge and discharge as follows. An absolute current of 50 A was applied to the cell. After each 300 s the current was reduced to an absolute value of 25 A for 100 s. At the end of each time series we included a rest period. The rest phase following the constant current (CC) phase with an absolute current of 180 A was chosen to be sufficiently long to achieve (almost) complete relaxation, at least following one lithiation and one delithiation process in the training data set. This is especially helpful when GB modelling the battery voltage dynamics as shown below. For the (dis-)charging with an absolute current of 18 A we simulated 18000 s current load and then 2000 s rest phase. For an absolute current of 50 A the current load lasted 12000 s, the subsequent rest phase lasted 8000 s. The CC phase for an absolute current of 180 A was kept up for 3600 s followed by a rest phase of 16400 s duration. For the pulsed currents we chose a total duration of 7500 s. The subsequent rest phase took 2500 s.

As test data sets for the final GB model, we simulated two additional dynamic protocols. One time series started with a fully delithiated particle, the other with a half-lithiated particle. We used current steps with different lengths, absolute values and sign, followed by rest phases of different lengths. The current profiles will be shown below in the Results section.

Warburg element.—In an electrochemical system such as a lithium-ion battery, diffusion-induced changes of the internal concentrations result in changes of the externally observable voltage. In the following we refer to this as diffusion voltage V_{diff}. In this section we describe the Warburg element commonly used as equivalent circuit component (cf. Fig. 1b).

Background.—The intercalation reaction of lithium is given by following equation:[3]

$$Li^+ + e^- + V_{Li} \rightleftarrows Li^I \qquad [11]$$

where, Li^+ denotes a lithium ion in the electrolyte, e^- an electron, V_{Li} a lithium vacancy in the host lattice of the electrode and Li^I intercalated lithium. Based on the Nernst equation, assuming small perturbations around the equilibrium, the relationship between the change in battery voltage and the concentration is given as (cf. Refs. 3, 20)

$$V_{diff} = \frac{R_g T}{F}\left(\ln\frac{c_{V\,act}}{c_{V\,eq}} - \ln\frac{c_{Li\,act}}{c_{Li\,eq}}\right) \qquad [12]$$

with the universal gas constant R_g, the temperature T, the Faraday constant F, the concentrations of intercalated lithium $c_{Li\,act}$ and $c_{Li\,eq}$, and the concentrations of vacancies $c_{V\,act}$ and $c_{V\,eq}$, where the subscript act means the actual and the subscript eq means the equilibrium concentration at the actual operation point. Using this relationship together with a Fickian diffusion process, the dynamic voltage response can be derived in the Laplace domain, resulting in the so-called Warburg element.[3,20]

The complex Warburg impedance Z_{diff} depends on the boundary conditions on either side of the diffusion domain. We have chosen Cartesian coordinates to generally describe Fickian diffusion in particles with a one-dimensional diffusion path, in contrast to the previous representation of Fickian diffusion in spherical particles.

In general, the current flux into the electrode defines the first boundary condition at $x = 0$:

$$j_{influx} = -D \cdot \frac{\partial c}{\partial x}, \quad x = 0,\ t \geqslant 0. \qquad [13]$$

The second boundary depends on the type of diffusion.[44] For semi-infinite diffusion, the second boundary condition is given by[44]

$$\frac{\partial c}{\partial x} = 0, \quad x = \infty\ ,\ t \geqslant 0. \qquad [14]$$

The resulting complex impedance, referred to as Warburg impedance, is inversely proportional to the square root of the frequency ω:[45]

$$\begin{aligned} Z_{diff} &= \frac{R_g T}{cn^2 F^2 A_e} \cdot \left(\frac{1}{\sqrt{\omega D}} - j \cdot \frac{1}{\sqrt{\omega D}}\right) \\ &= \frac{R_g T}{cn^2 F^2 A_e} \cdot \frac{1}{\sqrt{j\omega D}} \end{aligned} \qquad [15]$$

where, R_g represents the universal gas constant, c the molar concentration, n the charge number, A_e the active surface area, and ω the angular frequency. In the Nyquist plot it is represented by a straight line inclined at an angle of $-45°$ to the real axis.[9,20,44,45]

Model.—For the present study we assume semi-infinite diffusion. The dynamic voltage behavior of the corresponding Warburg impedance is given by Eq. 15. Due to the inverse proportionality to the square root of the frequency, the semi-infinite diffusion leads to a time-domain differential equation of fractional order according to[45,46]

$$\frac{d^{0.5} V_{diff}}{dt^{0.5}} = \frac{R_g T}{cn^2 F^2 A_e \sqrt{D}} \cdot I_{bat} \qquad [16]$$

where, $d^{0.5}/dt^{0.5}$ represents a fractional differential with order 0.5.

The battery current serves as the model input, and the voltage drop across the Warburg element is the model output.

Parameterization and simulation.—We used Matlab/SIMULINK (Matlab: version R2020a, SIMULINK: version 10.1) to implement a white-box model of the Warburg impedance according to Eq. 16. The fractional integrator was taken from the FOMCON toolbox (version 1.50.3).[47] The integrator is implemented as an Oustaloup filter.[48] We have used the following settings: Frequency range [10^{-4} Hz, 1000 Hz] and an approximation order of 5. The selected integrator can reproduce the desired integration of fractional order with good accuracy in the defined frequency range. For very low frequencies and very high

frequencies the behavior becomes resistive (cf. Ref. 49). We used the automatic solver selection with variable step size and a maximum step size of 0.5 s.

The proportionality factor $R_g T/cn^2 F^2 A_e \sqrt{D}$ in Eq. 16 is unknown. We arbitrarily set the factor such that, in a virtual experiment where a current step from 0 A to 180 A is applied, the voltage drop due to the Warburg impedance reaches a value of 200 mV after 1000 s. These are reasonable values for the chosen battery with a capacity of $C_{bat} = 180$ A h. The proportionality factor determined in this way was $3.210 \cdot 10^{-5}\ \Omega\ s^{-0.5}$. Different to the Fickian diffusion case, this value was kept constant (independent of SOC), mainly because we want to directly compare the behavior of the Warburg element to that of the RC circuit (cf. below).

To generate training data, we used similar current signals as for the Fickian diffusion case. However, the durations of current application were chosen such that the final lithiation was approximately 0.5 in each case when simulating the diffusion voltage with the specified battery capacity of $C_{bat} = 180$ A h. For the (dis-)charging with an absolute current of 18 A we simulated 18000 s current load and then 2000 s rest phase. For an absolute current of 50 A the current load lasted 6480 s, the subsequent rest phase lasted 3520 s. The CC phase for an absolute current of 180 A was kept up for 1800 s followed by a rest phase of 18200 s duration. For the pulsed currents we chose a total duration of 7500 s. The subsequent rest phase took 2500 s.

For testing, we used the same current signals as for the Fickian diffusion case.

Resistor-capacitor circuit.—One of the goals of the present study is to investigate whether a GB model based on finite-volume discretization of Fickian diffusion equation can approximate time-dependent behavior other than Fickian diffusion. We therefore replaced the Warburg element with a simple RC circuit (cf. Fig. 1c).

The RC circuit is modelled with an ordinary differential equation describing the variation of the voltage drop across the element with time, according to

$$\frac{dV_{RC}}{dt} = \frac{1}{C_1} \cdot I_{bat} - \frac{1}{R_1 \cdot C_1} \cdot V_{RC} \qquad [17]$$

with the voltage drop V_{RC}, the ohmic resistance R_1 and the capacitance C_1. The time constant $\tau = R_1 \cdot C_1$ is a characteristic property of the RC circuit. It is defined as the time it takes to discharge the capacitor across the resistor to $e^{-1} \approx 36.8\%$ of its initial voltage.

For direct comparison between RC circuit and Warburg element, we wanted the RC circuit voltage drop to reach 200 mV after 1000 s with a current load of $I_{bat} = 180$ A, as for the Warburg element. We chose the time constant to be $\tau = 1000$ s. We first calculated the final voltage drop V_{final} resulting from a current load of 180 A as

$$V_{final} = \frac{200.0\ \text{mV}}{0.632} = 316.5\ \text{mV}. \qquad [18]$$

The ohmic resistance of the RC circuit can then be calculated from the relation between the final voltage and the current step with Ohm's law as

$$R_1 = \frac{V_{final}}{I_{bat}} = \frac{316.5\ \text{mV}}{180\ \text{A}} = 1.758\ \text{m}\Omega.$$

Finally, we used this value together with the time constant to calculate the capacitance according to

$$C_1 = \frac{\tau}{R_1} = \frac{1000\ \text{s}}{1.758\ \text{m}\Omega} = 568.8\ \text{kF}.$$

These parameters were kept constant (independent of SOC).

The RC circuit model was implemented in SIMULINK. We generated another set of training and test data with the same current signals as for modelling the Warburg element (cf. above).

Grey-box Modelling

In this section we give a brief introduction to NODEs. We then introduce a GB model for modelling the lithium concentration profile in a spherical active material particle. Finally, we expand the GB modelling approach to model the voltage drop resulting from diffusion.

Background: neural ordinary differential equations.—NODEs[31] are a special type of neural networks. For a general overview of neural networks, the interested reader is referred to Ref. 50. One particular type of neural network is the residual neural network (ResNet), introduced by He et al.[51] Unlike simple feedforward networks, ResNets have additional short-cut connections. These shortcuts add a neuron's input directly to its output. ResNets overcome problems of training loss degradation with increasing number of hidden layers in deep neural networks. They can be used for time series prediction.[51]

The state transformation from layer t to layer $t+1$ in a ResNet can be described by the following recursive formula:[51]

$$z_{t+1} = z_t + f(z_t, \theta_t), \quad t = 0, ..., T-1 \qquad [19]$$

with the vector $z_t \in \mathbb{R}^d$ of states at layer t, the learned parameters θ_t of layer t, and the learnable function f: $\mathbb{R}^d \to \mathbb{R}^d$. The parameter vector θ_t summarizes the learned weights and biases. Reducing the time step Δt toward zero and thus increasing the number of observed points per time series toward infinity, while sharing the parameters across the layers ($\theta_t = \theta$ for $t = 0,...,T-1$), leads to the explicit Euler discretization of the initial value problem[31,52–57]

$$\frac{dz(t)}{dt} = f(z(t), t, \theta), \quad z(0) = z_0. \qquad [20]$$

The neural network f appears on the right-hand side of the differential equation. Therefore, the differential equation is called NODE. Given the initial state $z(0)$, a differential equation solver returns the output state $z(T)$.[31,54,55,57] The differential equation according to Eq. 20 is generalized to consider external variables $u(t)$:[58]

$$\frac{dz(t)}{dt} = f(z(t), u(t), t, \theta), \quad z(0) = z_0. \qquad [21]$$

The authors of Ref. 59 call the consideration of prior knowledge in a NODE *universal differential equation*. We want to emphasize the combination of WB and BB modelling techniques and therefore use the term GB modelling with NODEs.

Simulation and optimization methodology.—We implemented our GB models in Python (version 3.7.6). The open-source machine learning framework PyTorch (version 1.11.0)[60] was used. It provides two main features: Tensor computing and automatic differentiation for neural networks. We also used the torchdiffeq library (version 0.2.3).[43] It is based on PyTorch and allows solving differential equations and backpropagating through their solutions.

Unless otherwise stated, we used the Dopri5 method with an absolute tolerance of 10^{-9} and a relative tolerance of 10^{-7} to solve the differential equations in our models. As we used different quantities in the definition of the loss function, we used the numerical values of the respective quantities in SI units to calculate the training loss. The backpropagation of the training loss through the differential equations was done using the standard method from torchdiffeq. The standard method backpropagates through the solutions by applying the chain rule. For algebraic equations, the

automatic differentiation of PyTorch was used. Finally, the loss was minimized using an Adam optimizer.

After the introduction of the necessary background information and the description of the simulation and optimization methodology we now focus on GB modelling.

Concentration dynamics: Fickian diffusion.— *Approach.*—For simplicity, the active material particles of lithium-ion batteries are often assumed to be spherical. For example, in the single-particle model, the electrodes are each assumed to consist of one single spherical particle.[27,28] However, in real electrodes, microstructures tend to be much more complex, and truly spherical particles are rather uncommon; this is particularly true for graphite, as used in the experimental cell studied here.[39] We therefore further simplified the diffusion model by assuming a simple cartesian one-dimensional domain. This might be interpreted as cuboid particles of length $l = 2R$, where diffusion takes place in one direction only. Real particle shapes are likely neither spherical nor cuboid, nor consistent throughout the electrode. In the present approach, we assume that the particle shape and its distribution can be implicitly captured by training the free parameters of the neural network (diffusion coefficient, length, and discretization) within the GB model to the effective (macroscopic) diffusion response.

Based on these considerations, we used the ordinary differential equation system resulting from finite-volume discretization of Fickian diffusion in a one-dimensional cartesian domain as a basis for our GB model (called "GB I" in Fig. 1a). The corresponding equations are derived in the Appendix (Eqs. A·5). As some of the parameters in these equations are not known, we introduced learnable parameters and neural networks. As mentioned above, the active material particles in the electrodes are irregularly shaped and have different sizes. It is unclear which value to assume for the diffusion coefficient, the diffusion length, and the proportionality factor between the battery current and the molar flux across the particle surface (which is related to the active surface area).

We chose five discretization volumes in order to keep the computation time low. Using equidistant spatial discretization would be the simplest. However, as mentioned in Ref. 38, a non-uniform distribution could be advantageous. We therefore decided to include the individual lengths of the discretization volumes as learnable parameters during the training process.

A neural network approximates the diffusion coefficient. A direct dependence on the nondimensionalized concentration in the discretization volumes was considered. As the diffusion coefficient is positive, we decided to use the absolute value of the output of the neural network to simplify the training. Otherwise, the diffusion coefficient might become (partly) negative due to the initialization of the weights and biases of the neural network or due to the optimization steps during training. Furthermore, the proportionality factor in the Eq. A·5d is unknown. Here we introduced another learnable parameter that directly includes the length of the outermost discretization element.

Overall, these considerations lead to the following GB model:

$$\frac{\partial \bar{C}_i}{\partial t} = \gamma_{i+1} - \gamma_i, \quad 0 \leqslant i \leqslant 4 \qquad [22a]$$

$$\gamma_0 = 0 \qquad [22b]$$

$$\gamma_i = \left| f\left(\frac{\overline{C}_i + \overline{C}_{i-1}}{2}, \boldsymbol{\theta}_\mathrm{f}\right) \right| \cdot \alpha_i \cdot (\overline{C}_i - \overline{C}_{i-1}), \quad 1 \leqslant i \leqslant 4 \qquad [22c]$$

$$\gamma_5 = -\alpha_5 \cdot I_\mathrm{bat} \qquad [22d]$$

$$C_\mathrm{S} = \frac{3\overline{C}_4 - \overline{C}_3}{2} \qquad [22e]$$

with the learnable parameters α_i and the neural network f. For f, we chose a feedforward network with one hidden layer, ten hidden neurons and ReLU activation. The neural network takes a non-dimensionalized concentration as input and returns the partially nondimensionalized diffusion coefficient given in s^{-1}. The parameters α_i, $1 \leqslant i \leqslant 4$ represent dimensionless quantities and the parameter α_5 is given in $\mathrm{A}^{-1}\,\mathrm{s}^{-1}$. The battery current I_bat is the input to the model and the normalized surface concentration C_S is the output. The current for delithiation of the particle is defined as positive and the current for lithiation is defined as negative.

Note that we could have combined the neural network f and the learnable parameters α_i, $1 \leqslant i \leqslant 4$ in Eq. 22c. However, we decided to split the neural network and the parameters. This allows an evaluation of the diffusion coefficient represented by the neural network f. In addition, the concentration dependence of the diffusion coefficient is difficult to learn. When considering the learnable parameters within the neural network, there would be a risk of incorrectly learning different concentration dependencies for the different i, $1 \leqslant i \leqslant 4$. It is worthwhile noting that, due to its physical basis and the choice of finite volume discretization, the model is mass-conservative: independently of the values of the parameters, the spatially-averaged concentration only depends on the influx of lithium ions over the particle boundary.

Equation 22c involving a neural network is part of the right-hand side of the differential Eq. 22a. Therefore, this is a NODE with prior knowledge built in, which we call a GB model.

Normalization and initialization.—The initialization and scaling of the parameters to be learned are crucial for good training.[61] The authors of Ref. 61 recommend scaling the inputs of neural networks: The average of the input variables over the training set should be close to zero (note that this condition is fulfilled for the battery current of a rechargeable battery, since negative currents for charging and positive currents for discharging integrate to zero; the mean SOC is about 0.5). In addition, their covariance should be about the same.

Since the SOC is in the range 0 to 1, we decided to scale all inputs to values between 0 and 1. We also normalized the output value of the neural network to a similar range of values. We scaled the learnable parameters according to the range of values and the expected deviation from the chosen initial value.

Exemplary, we discuss how to scale and initialize the neural network f and the parameters α_i in Eq. 22c. Assuming equidistant discretization volumes, the output of the neural network times the respective learnable parameter α_i in Eq. 22c substitutes the factor $\frac{D^*}{(z_{i+1}-z_i)^2} = \frac{D}{\left(\frac{l}{2}\right)^2 \cdot (z_{i+1}-z_i)^2}$, $1 \leqslant i \leqslant N-1$ (cf. Eqs. A·5).

Using the physical parameters from the WB model representing a graphite electrode particle (cf. above), and using $l/2 = R$ (cartesian coordinates), the value for the α_i times the output of the neural network can be roughly estimated as $\frac{3.9 \cdot 10^{-14}\,\mathrm{m^2\,s^{-1}}}{(12.5 \cdot 10^{-6}\,\mathrm{m})^2 \cdot 0.02^2} = 1.872 \cdot 10^{-2}\,\mathrm{s^{-1}}$ for $N = 5$.

To keep the output of the neural network and the learnable parameters in the value range from 0 to 1, we introduced a scaling factor of 10^{-1}. The output of the scaled neural network times the scaled learnable parameter is then given by $f^*\left(\frac{\overline{C}_i + \overline{C}_{i-1}}{2}, \boldsymbol{\theta}_{\mathrm{f}^*}\right) \cdot \alpha_i^* = 10 \cdot f\left(\frac{\overline{C}_i + \overline{C}_{i-1}}{2}, \boldsymbol{\theta}_\mathrm{f}\right) \cdot \alpha_i$. We initialized the weights and biases of the scaled neural network from the uniform distribution $\mathcal{U}\left(-\sqrt{\frac{1}{k}}, \sqrt{\frac{1}{k}}\right)$, with $k \in \mathbb{N}$ the number of inputs to the respective layer (cf. Ref. 61). Note that the input of the neural network is dimensionless and that the output is given in s^{-1}. The learnable parameters α_i^*, $1 \leqslant i \leqslant 4$ were initialized with 1. The scaling and initialization of the other learnable parameter is described in the Appendix.

The adapted scaled model is given by the following equations:

$$\frac{\partial \bar{C}_i}{\partial t} = \gamma_{i+1} - \gamma_i, \quad 0 \leqslant i \leqslant 4 \tag{23a}$$

$$\gamma_0 = 0 \tag{23b}$$

$$\gamma_i = 0.1 \cdot \left| f^*\left(\frac{\bar{C}_i + \bar{C}_{i-1}}{2}, \boldsymbol{\theta}_{f^*}\right) \right| \cdot \alpha_i^* \cdot (\bar{C}_i - \bar{C}_{i-1}), \quad 1 \leqslant i \leqslant 4 \tag{23c}$$

$$\gamma_5 = -10^{-5} \cdot \alpha_5^* \cdot I_{\text{bat}} \tag{23d}$$

$$C_S = \frac{3\bar{C}_4 - \bar{C}_3}{2} \tag{23e}$$

with the learnable parameters α_i^* and the neural network f^* with one hidden layer, ten hidden neurons and ReLU activation. Since the scaling factors are dimensionless, the parameters have the same units as the unscaled ones. Equations 23 represent the GB model "GB I" in Fig. 1.

Training and test.—We implemented the final GB model for Fickian diffusion according to Eqs. 23 in Python. We used the time-dependent surface concentration simulated with the WB model which uses a spherical diffusion domain with 100 spherical shell discretization volumes for training and testing (cf. above).

After setting the initial concentration, the battery current $I_{\text{bat}}(t)$ served as an external variable. We solved the differential equations using the Dopri5 method and performed backpropagation through their solutions using the standard odeint method from torchdiffeq. The backpropagation of the loss through the algebraic equations was done with the automatic differentiation of PyTorch. We used the mean squared error (MSE) of the true and approximated surface concentration to define the loss function L as

$$L = 100 \cdot \text{MSE}(100 \cdot C_{S\,\text{GB}},\ 100 \cdot C_{S\,\text{true}}). \tag{24}$$

Additionally, we penalized negative values as output of the neural network f^*. We calculated the output of the neural network for input values in the range between -1 and 2 in steps of 0.1. We chose this wide range of values because due to the definition of the nondimensionalized concentration it is also possible to get negative values or values greater than 1. If an output value of the neural network was smaller than zero, 10000 times this value was subtracted from the error.

An Adam optimizer with a decaying learning rate between $\eta = 10^{-2}$ and $\eta = 10^{-3}$ minimized the loss. The parameters were saved as the training loss decreased. We decided to exclude some data at the beginning of the training to speed up the training process. During the first 99 epochs we only considered the first 10% of the data points of the time series. The number of data points was increased until we considered the entire time series from epoch 300 on. In total, we ran 750 training epochs with stochastic gradient descent which means that the backpropagation of the training loss was carried out separately for each time series. The simulation and optimization methodology is described above in detail. The training results are presented and discussed below.

The aim of the present study is to develop a GB modelling approach for Fickian diffusion as well as the resulting voltage drop across a Warburg element and an RC circuit. We want to show that the modelling approach works in principal. Finding the best possible model for this synthetic case is not our intention. Therefore, we did not perform hyperparameter tuning and we did not split off a validation data set.

Still, we tested the resulting GB model against the two additional test data sets. After providing the initial concentration, the current served as external input. Again, we solved the differential equations using the Dopri5 method.

Voltage dynamics: Warburg element and resistor-capacitor circuit.—In the previous section a GB model for the concentration dynamics due to Fickian diffusion ("GB I" in Fig. 1) was derived. In the present section, we extend this model to approximate the resulting voltage dynamics ("GB II" in Fig. 1). The proposed modelling approach is trained not only to the dynamics of a Warburg element (which is based on Fickian diffusion), but also to the dynamics of an RC circuit.

Approach.—The voltage drop due to concentration gradients is given by the Nernst Eq. 12. The concentration dynamics are given by Eqs. A·5. The combination of both equations serves as a basis for the full GB model. To this goal, we reformulated Eq. 12 by neglecting the vacancies:

$$V_{\text{diff}} = -a \ln \frac{c_S}{\bar{c}} \tag{25}$$

with a proportionality factor a, the surface concentration c_S of lithium, and the mean lithium concentration $\bar{c}$ in the particle. Applying the logarithmic laws $\ln \frac{c_S}{\bar{c}} = \ln c_S - \ln \bar{c}$, using the non-dimensionalized quantities C_S and $\bar{C}$ instead of c_S and $\bar{c}$, recognizing that $\bar{C} = \text{SOC}$, and linearizing the logarithmic terms leads to the following simplified expression:

$$V_{\text{diff}} = \omega_1 \cdot (\text{SOC} - C_S) \tag{26}$$

with the learnable parameter ω_1. This equation is based on the nondimensionalized quantity C_S which results from the GB transport model Eqs. 22.

We add an additional equation for the SOC as we need to know the average lithium concentration in the particle. The SOC of the battery is given by current integration ("Coulomb Counter"):

$$\frac{\text{dSOC}}{\text{d}t} = -\frac{1}{C_{\text{bat}}} \cdot I_{\text{bat}} \tag{27}$$

with the battery capacity C_{bat}.

According to the manufacturer, our representative cell (cf. above) has a nominal capacity of $C_N = 180\,\text{A h}$. Therefore, for the Coulomb Counter, we set the battery capacity to $C_{\text{bat}} = 180\,\text{A h} \cdot 3600\,\text{s h}^{-1} = 648000\,\text{A s}$.

Taken together, these assumptions result in the following GB model:

$$\frac{\partial \bar{C}_i}{\partial t} = \gamma_{i+1} - \gamma_i, \quad 0 \leqslant i \leqslant 4 \tag{28a}$$

$$\gamma_0 = 0 \tag{28b}$$

$$\gamma_i = \left| f\left(\frac{\bar{C}_i + \bar{C}_{i-1}}{2}, \boldsymbol{\theta}_f\right) \right| \cdot \alpha_i \cdot (\bar{C}_i - \bar{C}_{i-1}), \quad 1 \leqslant i \leqslant 4 \tag{28c}$$

$$\gamma_5 = -\alpha_5 \cdot I_{\text{bat}} \tag{28d}$$

$$C_S = \frac{3\bar{C}_4 - \bar{C}_3}{2} \tag{28e}$$

$$\frac{\text{dSOC}}{\text{d}t} = -\frac{1}{C_{\text{bat}}} \cdot I_{\text{bat}} \tag{28f}$$

$$V_{\text{diff}} = \omega_1 \cdot (\text{SOC} - C_S) \tag{28g}$$

with the learnable parameters α_i and ω_1 and the neural network f. The parameter ω_1 represents a voltage and is given in volts. The parameters α_i represent the same quantities, and the neural network f has the same configuration, as before when GB modelling Fickian diffusion. Overall, the equation system yields the diffusion voltage $V_{\text{diff}}(t)$ as output for a given current $I_{\text{bat}}(t)$ as input.

Normalization and initialization.—Again, we had to scale and initialize the parameters and the neural networks. Equations 28a to 28e are equal to the Eqs. 22. Therefore we used the same scaling factors as for the Eqs. 23 and initialized the parameters and neural networks as before.

There is an additional learnable parameter ω_1 in Eq. 28g which represents the factor a in Eq. 25. The estimation of a resulting from comparison with Eq. 12 is given in the Appendix.

These assumptions result in the following scaled GB model:

$$\frac{\partial \bar{C}_i}{\partial t} = \gamma_{i+1} - \gamma_i, \quad 0 \leqslant i \leqslant 4 \tag{29a}$$

$$\gamma_0 = 0 \tag{29b}$$

$$\gamma_i = 0.1 \cdot \left| f^*\left(\frac{\bar{C}_i + \bar{C}_{i-1}}{2}, \boldsymbol{\theta}_{f^*}\right)\right| \cdot \alpha_i^* \cdot (\bar{C}_i - \bar{C}_{i-1}), \quad 1 \leqslant i \leqslant 4 \tag{29c}$$

$$\gamma_5 = -10^{-5} \cdot \alpha_5^* \cdot I_{\text{bat}} \tag{29d}$$

$$C_S = \frac{3\bar{C}_4 - \bar{C}_3}{2} \tag{29e}$$

$$\frac{\text{dSOC}}{\text{d}t} = -\frac{1}{C_{\text{bat}}} \cdot I_{\text{bat}} \tag{29f}$$

$$V_{\text{diff}} = 10 \cdot \omega_1^* \cdot (\text{SOC} - C_S) \tag{29g}$$

with the scaled learnable parameters α_i^* and ω_1^* and the scaled neural network f^* with one hidden layer, ten hidden neurons and ReLU activation. Since the scaling factors are dimensionless, the parameters have the same units as the unscaled ones. The neural network takes a dimensionless quantity as input, the output is given in s^{-1}. Equations 29a to 29e represent the model "GB I" in Fig. 1, whereas Eq. 29g represents model "GB II".

We applied this GB model to the voltage dynamics of both, a Warburg element (Fig. 1b) and an RC circuit (Fig. 1c).

Training and test.—We implemented the final GB model for slow voltage dynamics according to Eqs. 29 in Python. The differential Eq. 29f and the NODE given by Eqs. 29a to 29d were summarized in one differential equation system and solved together. The SOC and the lithium concentrations in the outer discretization elements were then used to calculate the surface concentration and the voltage drop across the diffusion element. We trained the final GB model against the voltage drop across either the Warburg element or the RC circuit.

For the Warburg element, simulation results from the SIMULINK model with fractional integrator (as shown above) were used as training and test data. The training is described in the following. The initial concentrations and SOC were set to 0 or 1 depending on the time series. The battery capacity was given as $C_{\text{bat}} = 180$ A h and the current I_{bat} served as the external variable. We solved the differential equations using the Dopri5 method and performed backpropagation through their solutions using the standard odeint method from torchdiffeq. The loss function was defined as

$$L = 100 \cdot \text{MSE}(100 \cdot V_{\text{diff GB}}/\text{V},\ 100 \cdot V_{\text{diff true}}/\text{V}). \tag{30}$$

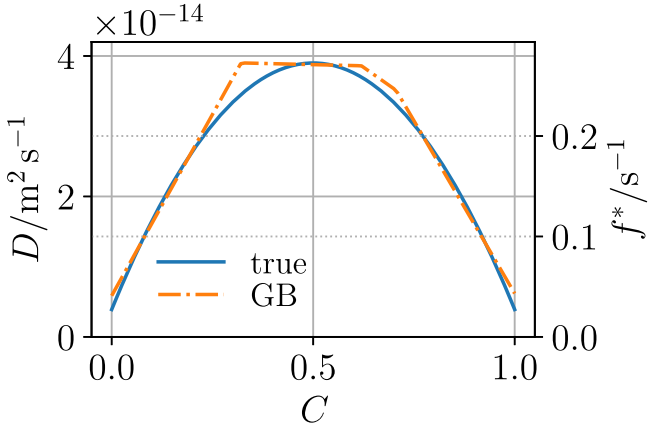

Figure 2. Output of the scaled neural network f^* ("GB") after completing the training in comparison to the true diffusion coefficient ("true") for Fickian diffusion.

Additionally, we penalized negative values as output of the neural network f^*. We calculated the output of the neural network for input values in the range between -1 and 2 in steps of 0.1. If an output value of the neural network was smaller than zero, 10000 times this value was subtracted from the error.

An Adam optimizer with a decaying learning rate between $\eta = 10^{-2}$ and $\eta = 10^{-4}$ minimized the loss. The parameters were saved as the training loss decreased. We decided to exclude some data at the beginning of the training to speed up the training process. During the first 299 training epochs, we considered only the training data for pulsed current charging and discharging. Afterwards, we also took the other training data sets into account. During the first 19 epochs we froze the parameter ω_1^*. We only considered the first 10% of the data points of the time series during the first 99 epochs. Note that we used an adaptive solver to generate the training data. Therefore, the data points were irregularly sampled. The number of data points was increased until we considered the entire time series from epoch 300 on. This was done to facilitate the approximation of the voltage dynamics and to speed up the training. In total, we ran 500 training epochs with stochastic gradient descent.

As before, we used the two remaining data sets for model testing. The Dopri5 method was used to solve the differential equations. The initial concentrations and the initial SOC as well as the current signals were provided.

For the RC circuit, we used simulation results of the SIMULINK model with the RC circuit (as shown above) as training and test data. The training and test procedure was the same as for the Warburg element.

Results and Discussion

This section presents the results of the GB models derived in the previous section, starting with the concentration dynamics due to Fickian diffusion (Fig. 1a) and followed by the voltage dynamics from a Warburg element (Fig. 1b) and an RC circuit (Fig. 1c).

Fickian diffusion.—A total of eight time series of $I_{\text{bat}}(t)$ (cf. above) with the resulting single-particle surface concentration dynamics (using parameters inspired by a graphite electrode, cf. above) were used as training data. The absolute value of the output of the scaled neural network f^*, representing the scaled diffusion coefficient, at the end of training is shown in Fig. 2 as function of C on the right axis. On the left axis, the figure shows the chosen concentration-dependent diffusion coefficient used to generate the training and test data. For a direct comparison, both axes were individually scaled such that the maxima of the curves

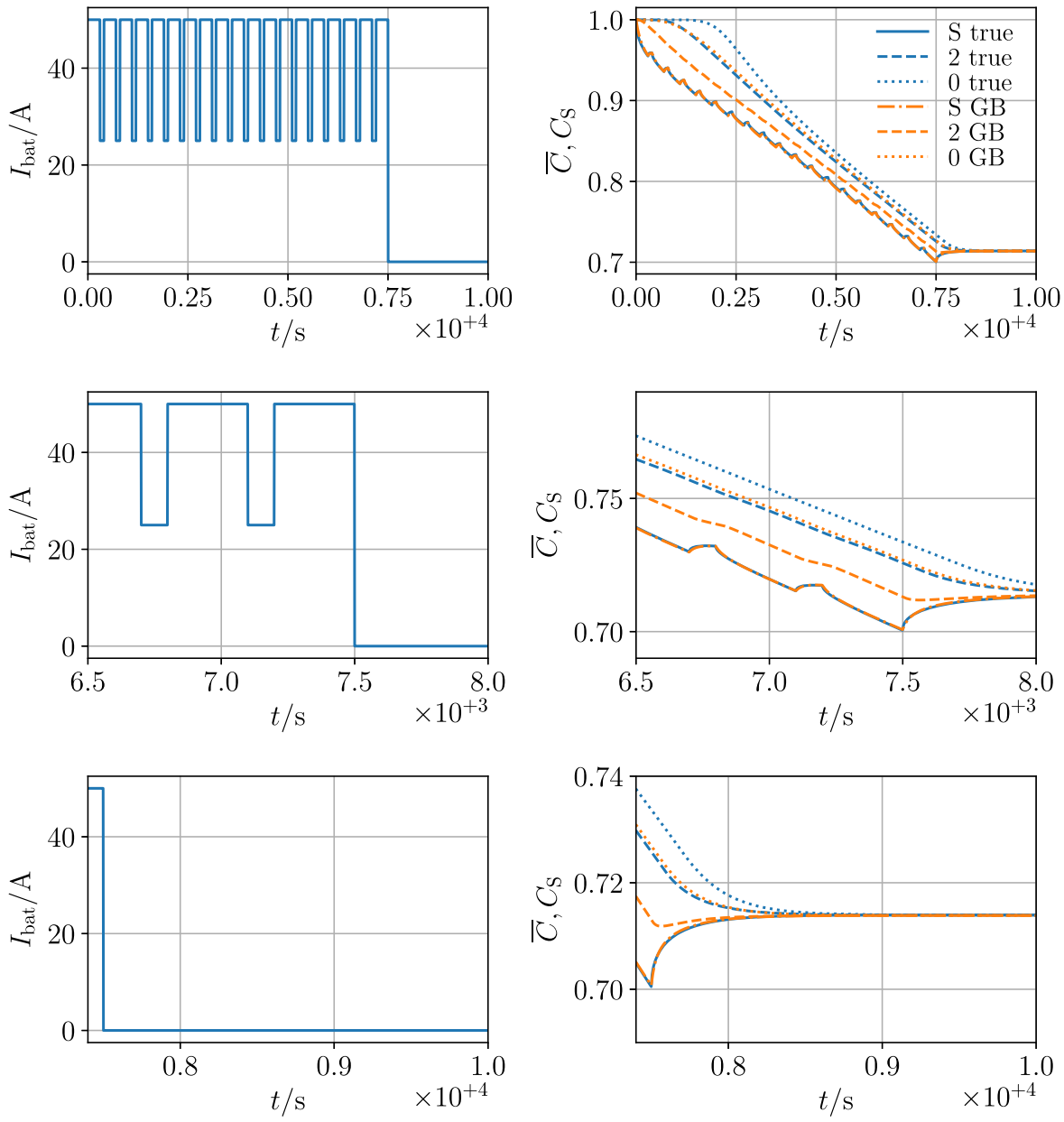

Figure 3. Training results for delithiation of a spherical particle with a pulsed current. The upper panels show the complete time series, the middle panels show the last two pulses and the lower panels show the rest phase; left: battery current, right: concentrations. The right panels compare results from the Fickian diffusion model with 100 discretization elements ("true") and the GB model ("GB"). The legend indicates the position within the diffusion length, where S is the surface, 2 is the middle discretization element and 0 is the innermost element.

Table I. Results for the scaled learnable parameters after completing the training and the mean squared error (MSE) between the true and predicted nondimensionalized surface concentration (Fickian diffusion) or diffusion voltage (Warburg element and resistor-capacitor circuit) for delithiation with a pulsed current.

	α_1^*	α_2^*	α_3^*	α_4^*	α_5^*	ω_1^*	MSE
Fickian Diffusion	0.1869	0.7905	1.1810	1.3498	0.4330 $\mathrm{A}^{-1}\,\mathrm{s}^{-1}$	—	$9.6959 \cdot 10^{-7}$
Warburg Element	0.0132	0.0951	0.4458	1.3048	0.7842 $\mathrm{A}^{-1}\,\mathrm{s}^{-1}$	0.0899 V	$3.0955 \cdot 10^{-7}\ \mathrm{V}^2$
Resistor-Capacitor Circuit	0.6145	0.4965	0.2672	0.1166	0.7651 $\mathrm{A}^{-1}\,\mathrm{s}^{-1}$	0.0204 V	$2.8343 \cdot 10^{-7}\ \mathrm{V}^2$

coincide. The learned diffusion coefficient shows good qualitative agreement with the true one over a wide range of concentration. Only the areas around the boundary points and the extreme point show greater deviation. Here one has to keep in mind that we chose a small neural network with only 10 hidden neurons and ReLU activation.

The trained parameters are given in Table I. The values α_1^* to α_4^* represent the different lengths of the discretization volumes. They increase with increasing i, that is, the volumes become smaller toward the surface. According to Ref. 38, a nonuniform distribution with smaller discretization elements toward the surface allows a dramatic reduction of the number of discretization elements

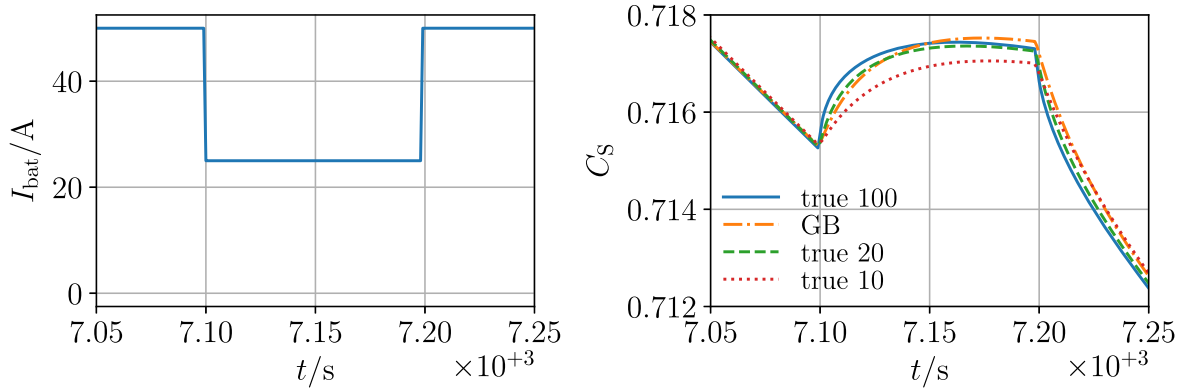

Figure 4. Training results for delithiation of a spherical particle with a pulsed current. The focus is on one current pulse; left: current, right: surface concentration. The figure compares results from Fickian diffusion models ("true") with 100, 20 and 10 discretization volumes along the particle radius to the results of the GB model ("GB") trained to the Fickian diffusion model with 100 discretization volumes.

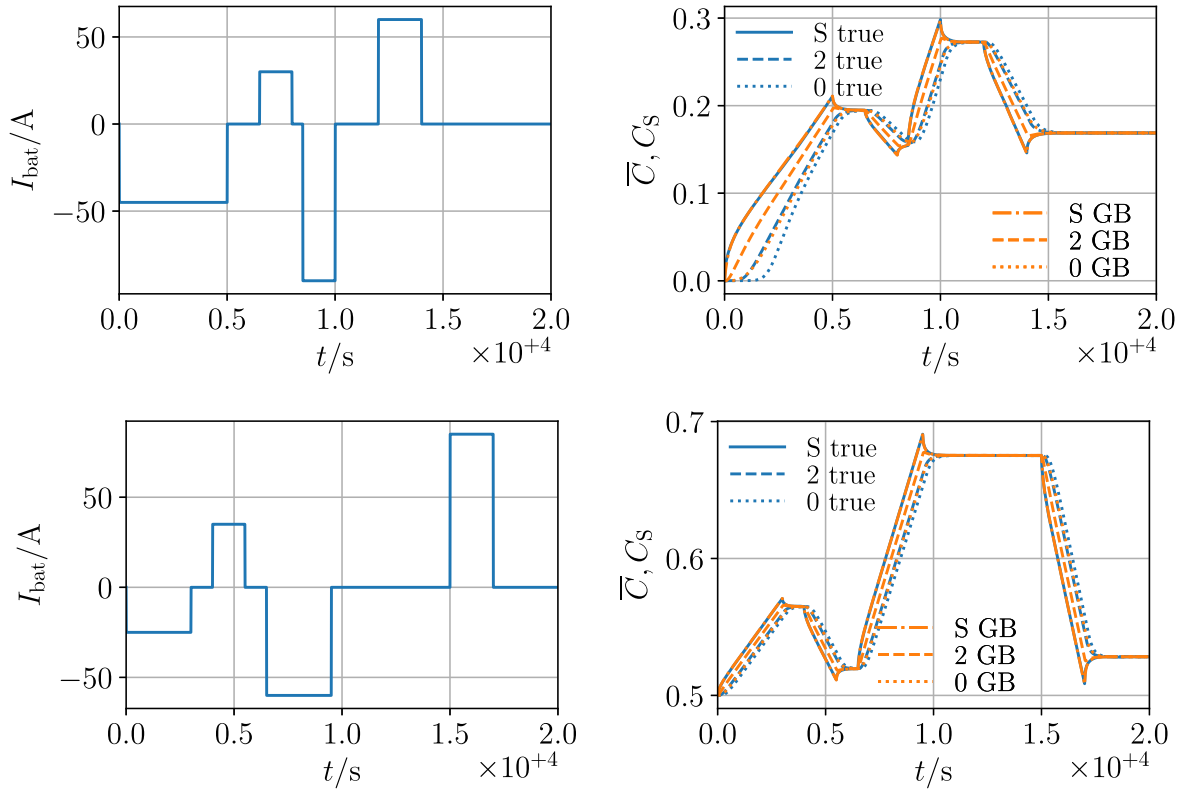

Figure 5. Test results for Fickian diffusion. The upper panels show the results for test data set one, the lower panels show the results for test data set two; left: battery current, right: concentrations. The right panels compare results from the Fickian diffusion model with 100 discretization elements ("true") and the GB model ("GB"). The legend indicates the position within the diffusion length, where S is the surface, 2 is the middle discretization element and 0 is the innermost element.

compared to a uniform distribution while maintaining a high accuracy. The training process has "automatically" recognized this requirement and has adapted the discretization lengths accordingly. The parameter α_5^* represents the proportionality factor between the battery current and the molar flux and additionally takes into account the particle surface area and the length of the discretization volume closest to the particle surface.

Figure 3 shows the training results of delithiation of the spherical particle using a pulsed current as an example. In addition to the concentration at the surface, which is the value used in the loss function, the average concentrations in the innermost discretization volume (indicated with 0 in the legend) and the middle discretization volume (indicated with 2 in the legend) are shown. The upper panels show the entire time series, the middle panels focus on the last two current pulses, and the lower panels show the rest phase at end of delithiation. The surface concentration dynamics predicted by the GB model are in very good agreement with the training data. However, the concentrations in the middle and the center deviate from the true values. They are both lower than for the WB model. We believe that this difference results from both, the significantly lower number of only five discretization volumes in the GB model compared to 100 in the WB model, as well as the non-equidistant distribution of these volumes in the GB model (cf. previous paragraph) compared to the equidistant distribution in the WB model. During the rest phase the approximated concentrations approach each other until they are almost the same in the end, as required from mass conservation.

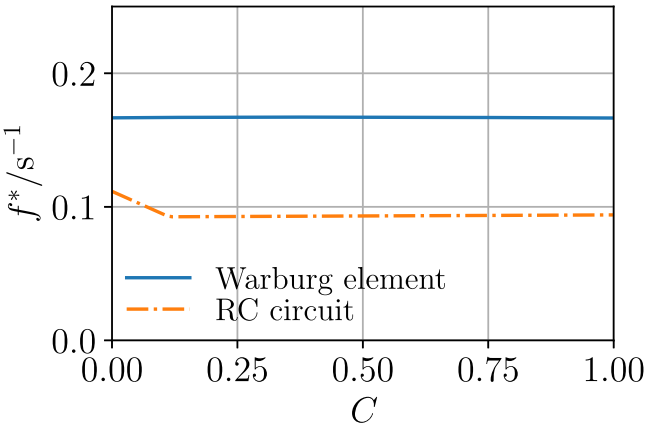

Figure 6. Output of the scaled neural network f^* for GB modelling the Warburg element and the RC circuit after completing the training.

For a more detailed analysis, Fig. 4 focuses on the value of interest, the surface concentration, shown here for the last pulse. Overall, the estimated concentration dynamics are consistent with the true values. However, in the first seconds after the current step, there is a clear deviation—the estimated concentration changes slower than the true concentration. It should be noted that we only considered 5 discretization volumes in our GB model, while the training data was calculated with our WB model (Eqs. 10) using a much finer discretization of 100 volumes. In order to test the influence of discretization on the results, we carried out additional simulations with the WB model using only 10 and 20 volume elements. The results are also shown in Fig. 4. These simulations show a significant deviation compared to the WB model with 100 volumes. The GB model, based on only 5 volumes, shows a behavior somewhat in between the WB models with 10 and 20 volumes in the first seconds after the current step. At the end of the current pulse, the concentration is slightly overestimated by the GB model.

In addition to the learned parameters, Table I gives the MSE between the true and the approximated surface concentration for the delithiation with a pulsed current.

After completing the training process, we wanted to test the model on data that was not included in the training. Figure 5 shows the results for the two test data sets. The long rest periods between the pulses with alternation between positive and negative current steps of different duration characterize the two test profiles and distinguish them from the training data set with CC training data and regular current steps. Test data set two is also characterized by starting with a half-lithiated particle. For both data sets, the surface concentration predicted by the GB model shows very good agreement to the results of the WB model.

The results show that our GB model can successfully predict the slow concentration dynamics resulting from Fickian diffusion. It is able to at least partially compensate for the difference in the assumed transport domain (cartesian vs. spherical) and in the number of discretization volumes (5 vs. 100). This demonstrates the validity and the flexibility of the chosen approach for modelling the surface concentration.

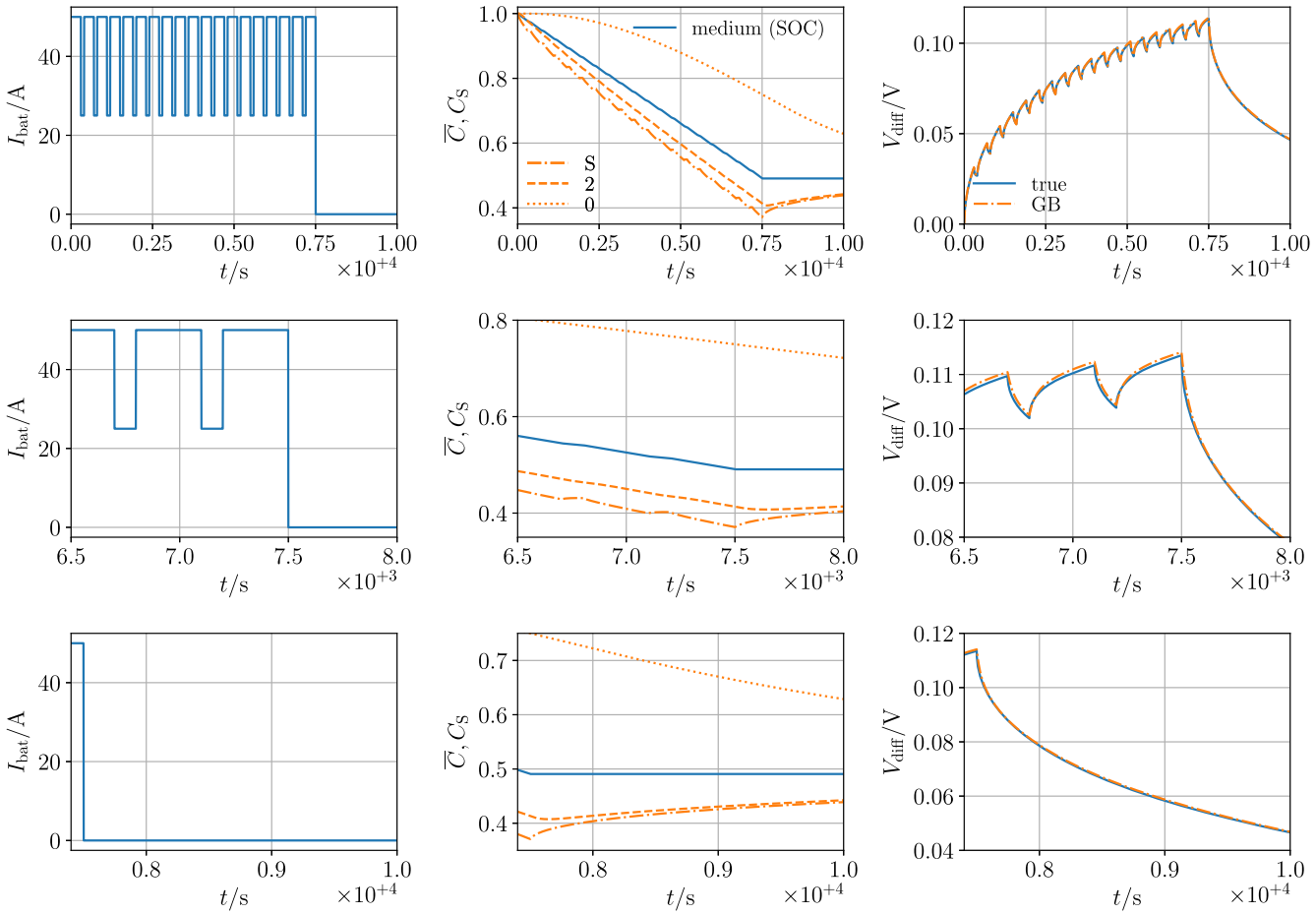

Figure 7. Training results for discharge of a Warburg element with a pulsed current. The upper panels show the complete time series, the middle panels show the last two pulses and the lower panels show the rest phase; left: battery current, middle: concentrations, right: voltage drop over the Warburg element. The figure compares results from a Warburg element ("true") to results of the GB model ("GB"). The legend indicates the position within the diffusion length, where S is the surface, 2 is the middle discretization element and 0 is the innermost element.

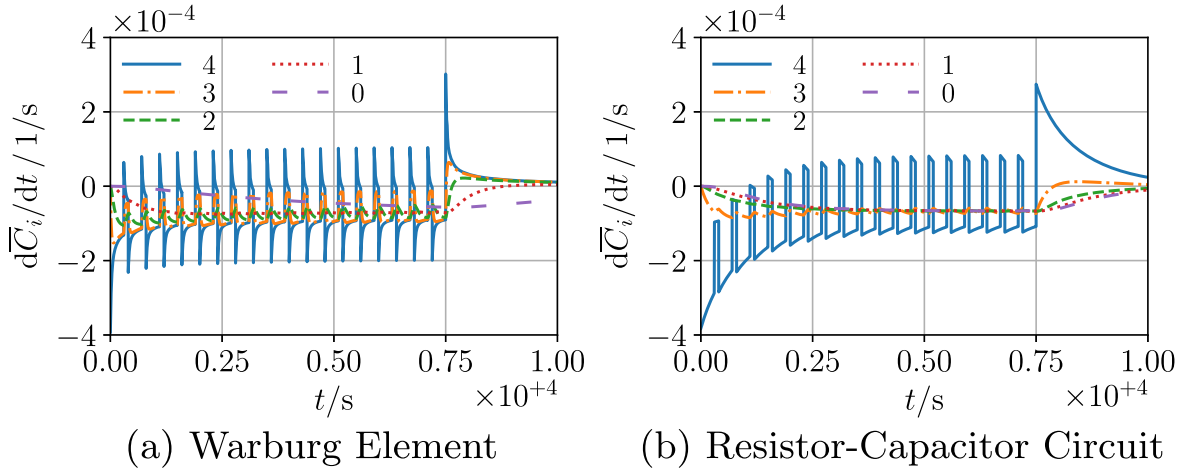

Figure 8. Change of the GB model concentrations with time for battery discharge with a pulsed current; a) GB model trained to a Warburg element, b) GB model trained to an RC circuit. The legend indicates the position within the diffusion length, where 4 is the discretization element closest to the surface and 0 is the innermost element.

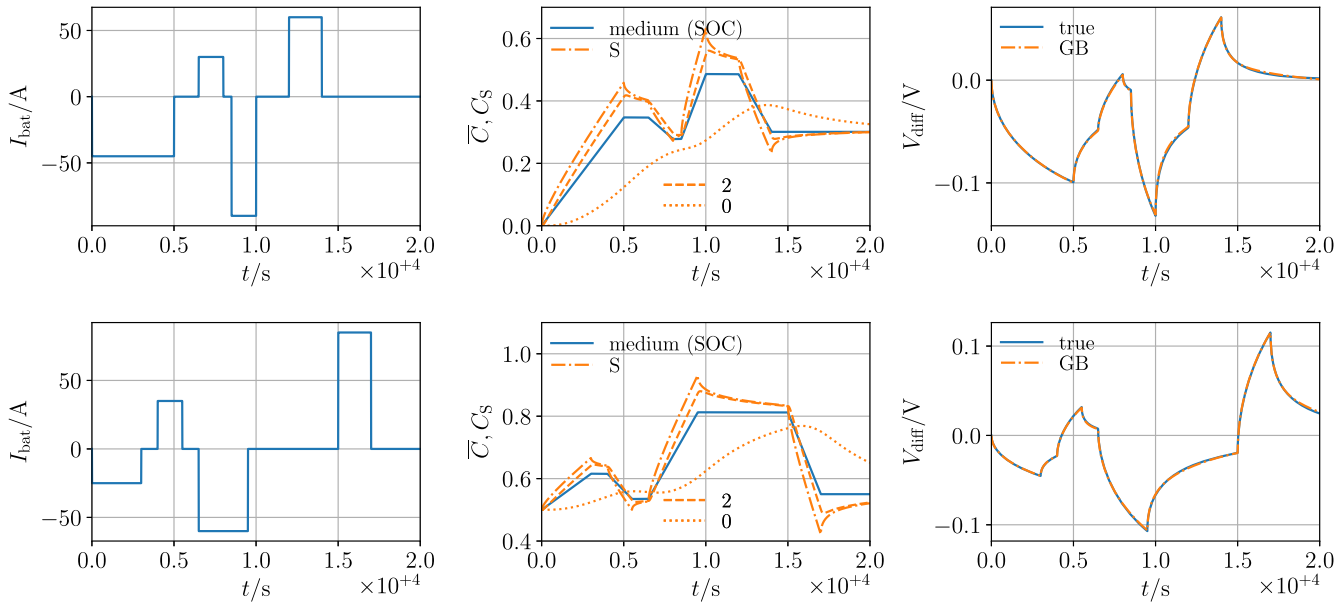

Figure 9. Test results for the simulation of a Warburg element. The upper panels show the results for test data set one, the lower panels show the results for test data set two; left: battery current, middle: concentrations, right: voltage drop over the Warburg element. The figure compares results from a Warburg element ("true") to results of the GB model ("GB"). The legend indicates the position within the diffusion length, where S is the surface, 2 is the middle discretization element and 0 is the innermost element.

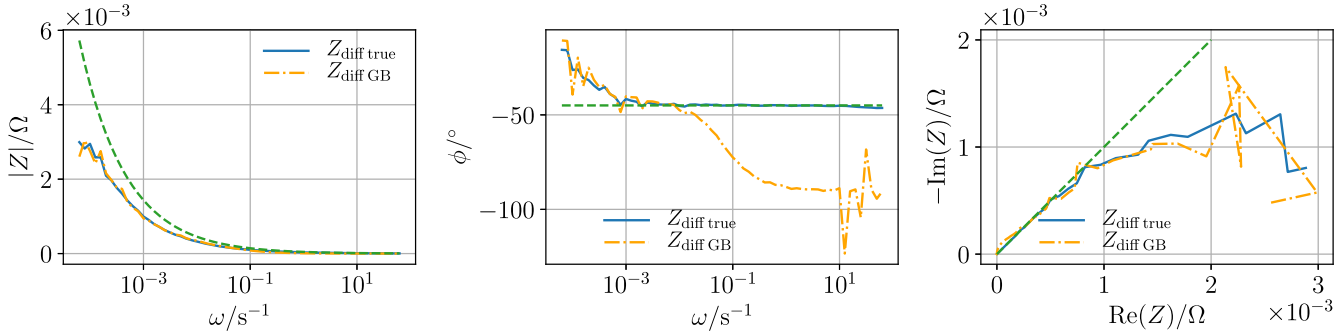

Figure 10. Simulated electrochemical impedance spectrum of the GB model trained to the Warburg element. Impedance was simulated using a current step voltage response method; left: the amplitude response, middle: the phase response, right: a Nyquist plot from the Warburg model ("true") and the GB model ("GB"). Angular frequency scale is logarithmic. The dashed green line indicates a −45° angle representing ideal Warburg behavior resulting from semi-infinite diffusion.

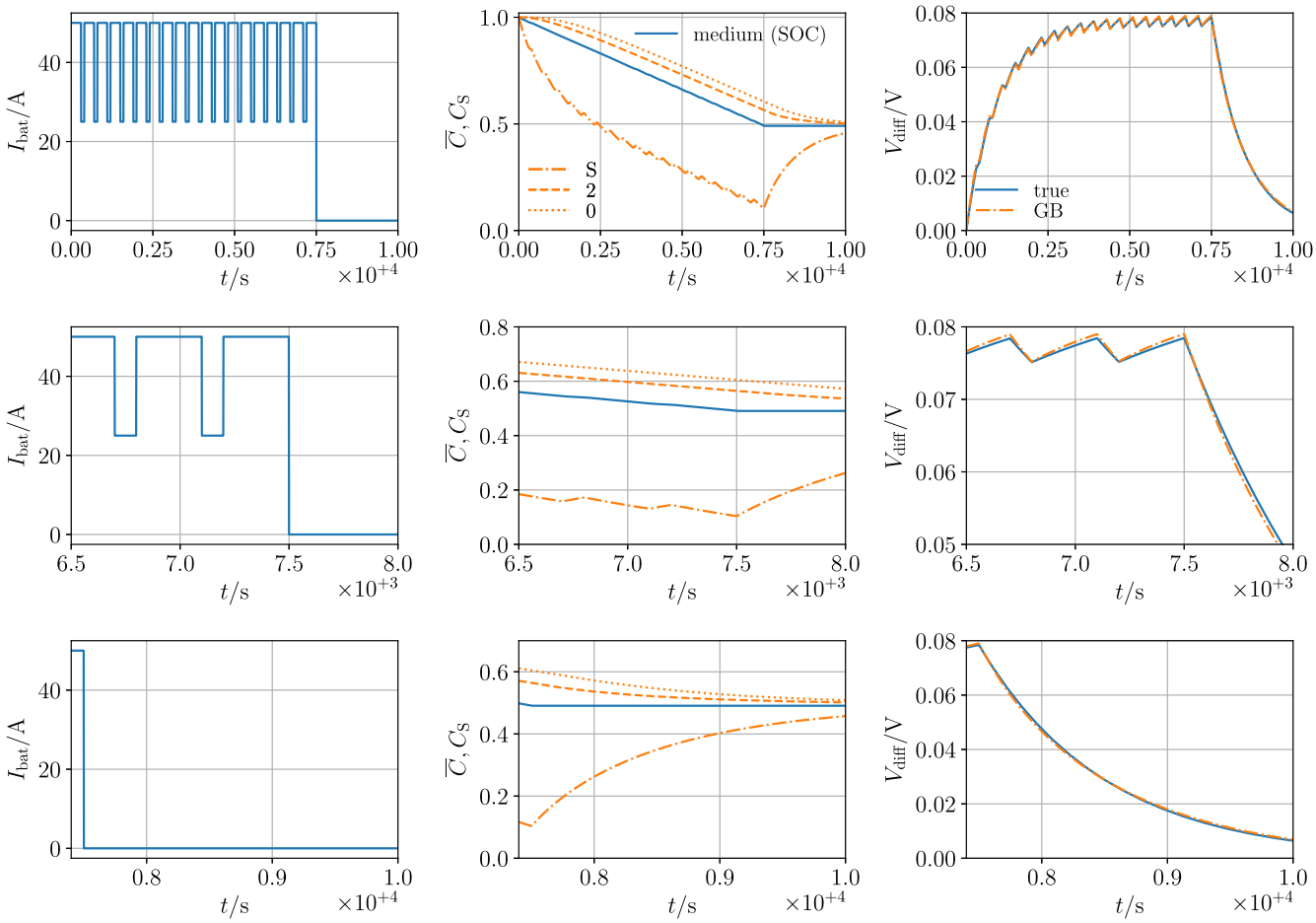

Figure 11. Training results for discharge of an RC circuit with a pulsed current. The upper panels show the complete time series, the middle panels show the last two pulses and the lower panels show the rest phase; left: battery current, middle: concentrations, right: voltage drop over the RC circuit. The figure compares results from an RC circuit ("true") to results of the GB model ("GB"). The legend indicates the position within the diffusion length, where S is the surface, 2 is the middle discretization element and 0 is the innermost element.

Warburg element.—In the following we discuss the results for GB modelling the voltage drop across a Warburg element. Again, we used six times series covering CC lithiation and delithiation and two times series with pulsed currents for training. The training and test data were generated with the SIMULINK model featuring a fractional order integrator to replicate the Warburg element (cf. above).

Figure 6 shows the output of the scaled neural network f^* representing the partially nondimensionalized diffusion coefficient at the end of training. It is almost independent of the concentration. As we only used constant SIMULINK model parameters, this was expected. Still, after having demonstrated that the GB model is able to capture nonlinear dependencies when modelling Fickian diffusion (see above), these results demonstrate the flexibility of the GB model by showing that it can also capture constant diffusive behavior. The scaled learnable parameters α_i^* and ω_1^* can be found in Table I. As before, the parameters α_1^* to α_4^* increase with increasing i which means that the discretization volumes become smaller toward the surface.

Figure 7 shows the results for exposure to the pulsed discharge current. The left column shows the current, which is the model input. The middle column shows the SOC resulting from the integration of the current over time. In addition to the SOC, the GB model concentrations at surface, and the average concentration in the middle discretization element (indicated with 2 in the legend) and in the innermost discretization element (indicated with 0 in the legend) are shown. The right column shows the voltage drop across the Warburg element, which is the main value of interest as it is used in the loss function. The top row shows the entire time series, while the middle row focuses on the last two current pulses. The lower row shows the rest phase. We have defined the nondimensionalized concentrations in such a way that the average concentration is equal to the SOC. In the middle columns we can see that the SOC is always between the concentration in the innermost volume element and the surface concentration. During the rest phase the concentrations approach each other. The approximated voltage drop across the Warburg element agrees very well with the training data. During the pulses, the approximation deviates slightly from the true voltage. Notably, the model captures the slow voltage dynamics during the rest phase (here around 0.7 h) very well. The rest phase was too short to allow complete relaxation.

The MSE between the true and the approximated diffusion voltage is given in Table I for the pulsed delithiation.

For a more detailed analysis, Fig. 8a shows the change of the concentrations within the individual GB finite volumes with time. The concentration change is largest at the surface, from where it clearly propagates into the inner discretization volumes.

We tested the final GB model against the two test data sets that were not included in the training process. We used the same current profiles as for modelling Fickian diffusion. Figure 9 shows the results. Even during the long rest phases, the voltage predicted by the GB model corresponds very well to the output of the WB model. This is the case for both test data sets.

We further investigated the behavior in the frequency domain. We used the computational approach of Bessler[62] to simulate the

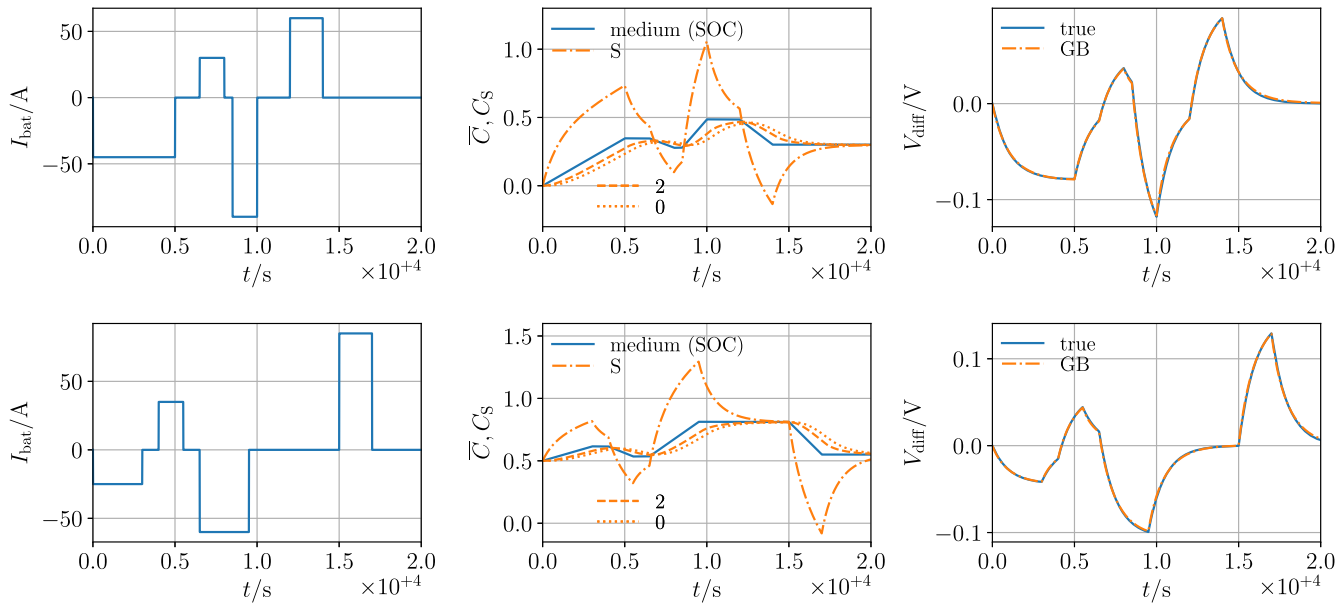

Figure 12. Test results for the voltage drop across an RC circuit. The upper panels show the results for test data set one, the lower panels show the results for test data set two; left: battery current, middle: concentrations, right: voltage drop over the RC circuit. The figure compares results from an RC circuit ("true") to results of the GB model ("GB"). The legend indicates the position within the diffusion length, where S is the surface, 2 is the middle discretization element and 0 is the innermost element.

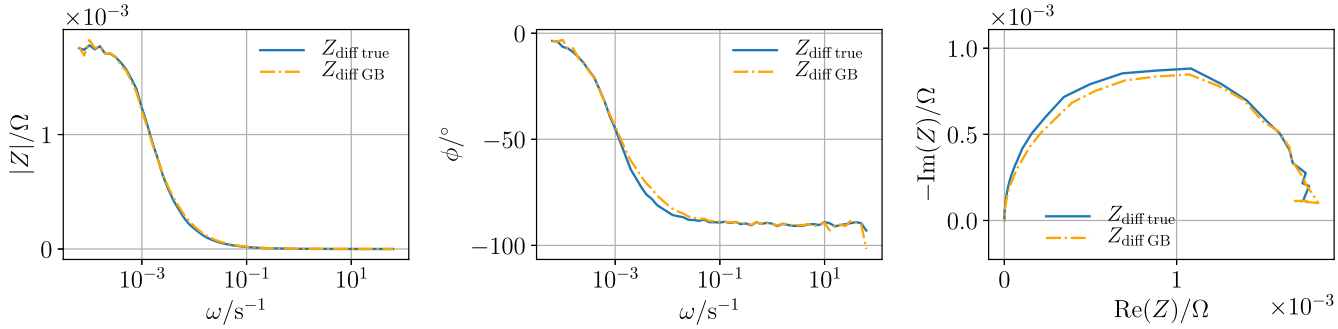

Figure 13. Simulated electrochemical impedance spectrum of the GB model trained to the RC element. Impedance was simulated using a current step voltage response method; left: the amplitude response, middle: the phase response, right: a Nyquist plot from the RC model ("true") and the GB model ("GB"). Angular frequency scale is logarithmic.

electrochemical impedance spectrum using our time domain models. We applied a current step excitation and evaluated the voltage relaxation using Fourier transforms. We set the initial nondimensionalized concentrations in the discretization elements to 1. Starting with zero current, we simulated a fast linear current step of $\Delta I_{bat} = 0.1$ A during 10^{-5} s. The simulation was carried out over a time of 10^5 s. We used the current signal and the voltage response of the model to calculate the impedance according to the findings in Ref. 62 for a frequency band between 10^{-5} Hz and 10 Hz.

A proper choice of solvers and tolerance turned out to be important for the impedance simulations. Using the Dopri5 method with an absolute tolerance of 10^{-9} and a relative tolerance of 10^{-7} resulted in very noisy impedance spectra. With lower absolute and relative tolerances of 10^{-20} the results improved significantly. These settings are used in the following. We also tested the SciPy[38] solver LSODA, which in turn wraps a Fortran solver from ODEPACK.[63] For the same tolerances, the results were only slightly affected.

The results are shown in Fig. 10. The left panel shows the amplitude response, the middle panel shows the phase response of the trained GB model, and the right panel shows the resulting Nyquist plot. A pure Warburg element would give a straight line with an angle of $-45°$ starting at an impedance of $Z = 0\,\Omega$ in the Nyquist plot (cf. dashed lines). It should be mentioned that neither the WB nor the GB model shows this ideal behavior. Due to the use of the Oustaloup filter in SIMULINK, the training data deviates from the straight line for low angular frequencies ($\omega < 10^{-3}\,s^{-1}$, as the lower frequency range of the Ostaloup filter was set to $f = 10^{-4}$ Hz, or $\omega = 2\pi f = 6.283 \cdot 10^{-4}\,s^{-1}$ respectively). The amplitude of the impedance as well as the phase angle are smaller than for the pure Warburg element. This is reproduced by the GB model. At angular frequencies above ca. $20 \cdot 10^{-3}\,s^{-1}$ there are significant deviations of the GB model from the training data: the phase angle does not stay at $-45°$ but decreases toward $-90°$. The limited discretization (only 5 volumes) used in the GB model is insufficient to capture fast dynamics (which was, however, not the goal of the present study). Towards low and high frequencies, numerical noise is visible in the impedance spectrum for both the SIMULINK model and the GB model. However, the GB model has more numerical noise.

Overall, the GB model shows a very good quantitative agreement with the time-domain training and test data, and good agreement with the frequency-domain behavior below an angular frequency of ca. $20 \cdot 10^{-3}\,s^{-1}$. We can conclude that the proposed GB modelling approach is suitable for describing the voltage drop across the Warburg element at long time scales.

Resistor-capacitor circuit.—Finally, we used the same GB modelling framework to approximate the voltage drop across an RC circuit. The GB model was trained and tested against a

SIMULINK model with a single RC circuit, using the same simulation protocols as for the Warburg element.

The output of the scaled neural network f^* after completion of training is shown in Fig. 6. It represents the partially nondimensionalized diffusion coefficient. Again, as expected and consistent with the training data, it does not show a strong dependence on concentration. Note that the value is significantly lower than for the GB model trained to the Warburg element. This reflects the difference in voltage dynamics between the Warburg element and the RC circuit. The scaled learnable parameters α_1^* to α_4^*, representing the lengths of the discretization volumes, are given in Table I. In contrast to the values for the other cases, the values here decrease with increasing i. This means that the elements become larger toward the surface. This difference will be further discussed below.

Figure 11 shows training results for a pulsed discharge. The figure has the same layout as in the case of the Warburg element (Fig. 7). The middle column shows SOC and concentrations. As expected and physically required, the SOC is always between the center concentration and the surface concentration. During the rest phase the concentrations approach each other until they are almost equal. The approximated voltage dynamics are plotted in the right column. The voltage shows good agreement with the training data. A slight deviation is visible right after the current step to high currents. During the rest phase, the voltage drop across the RC circuit approaches zero volts.

Again, the MSE between the true and the approximated diffusion voltage is given in Table I for the delithiation with a pulsed current.

Compared to the results of the GB model trained to the Warburg element (Fig. 7), the surface concentration is significantly lower here during the current load. This is consistent with the lower value for the diffusion coefficient with a similar range of values for α_i^*, $1 \leqslant i \leqslant 4$ and almost the same proportionality factor α_5^*. Furthermore, the values of the parameters α_i^*, $1 \leqslant i \leqslant 4$ suggest a different distribution of the discretization elements, which is reflected in the results.

Figure 8 shows the change of the concentrations with time for modelling the RC circuit in comparison to the Warburg element. In the WB model used for training, the voltage dynamics are given by one single differential equation (see Eq. 17). The GB model used in the Eqs. 29 consists of five differential equations. After training, the outer discretization volume is the largest (physically, this means that it contains most of the electrode capacity). Due to this particular distribution of the discretization volumes in combination with the low diffusion coefficient, the concentrations in the inner discretization volumes are only slightly affected by the external current. Therefore, the behavior of the GB model is dominated by a single differential equation representing concentration change at the surface of the diffusion domain. This means that the training process has reproduced the mathematical structure of the RC circuit model.

Again, we tested the final GB model against the two test data sets. The results are shown in Fig. 12. The voltage approximated by the GB model shows excellent agreement with the true voltage given by the WB model for both test data sets. In the middle panels it can be noted that the surface concentration can reach values lower than 0 and larger than 1. Here, a physical interpretation of this internal state of the GB model is not possible.

Using the same approach as described in the previous section, we simulated the electrochemical impedance spectrum of the GB model. The results are shown in Fig. 13. An ideal RC circuit appears as semicircle in the Nyquist plot with the ohmic resistance as the diameter. The right panel shows good approximations of a semicircle for both the WB model and the GB model. We used an ohmic resistance of $R_1 = 1.758\ \mathrm{m\Omega}$ in our WB model to generate the training data. This resistance value is well estimated by the final GB model. Numerical noise is present at low and high frequencies for both the SIMULINK model and the GB model.

Overall, the results show that the proposed GB modelling framework can successfully reproduce the slow voltage dynamics of an RC circuit. One has to keep in mind that the GB model is based on Fickian diffusion. However, by training the parameters, it can also reproduce dynamic behavior other than pure Warburg (here: RC circuit behavior). This shows the flexibility of the approach, which is required when training to experimental data that may not show ideal Warburg type behavior.

Summary and Conclusions

In this article we have presented the development and validation of a GB modelling framework for the voltage dynamics of lithium-ion batteries, focusing on long timescales. The finite-volume discretization of solid-phase diffusion in an active material particle in combination with an expression for the diffusion voltage served as the physical base structure of the GB model. We performed model-to-model comparisons. The main results can be summarized as follows.

In the first part of this study, we have shown how to derive a GB model for a Fickian diffusion process, in particular for lithium diffusion inside an active material particle, using a finite-volume discretization of Fick's second law with an appropriate choice of learnable functions (formulated as NODEs) and parameters. We trained and tested the GB model with simulation data obtained from numerical simulations of Fick's law for a spherical particle. As training data we used two time series with pulsed current lithiation and delithiation, as well as six more time series applying a constant current with a subsequent rest phase. The test data set consisted of two more time series covering current pulses of different amplitude and length followed by long rest phases. For generating the training and test data, we chose a spatial discretization of 100 equidistantly distributed finite volume elements along the particle radius. The GB model only consisted of five finite volume elements. However, the spatial distribution of the elements was used as a degree of freedom. Despite the small number of discretization volumes, the GB model was able to approximate the dynamics of the surface concentration with good accuracy. The trained GB model showed a decreasing size of the discretization volumes toward the particle surface. Such a non-equidistant discretization scheme is known to lead to a higher accuracy of the simulations.[38] We conclude that the training has successfully resulted in an adaptive spatial discretization scheme.

In the second part of this study, we extended the GB model by adding a physically motivated expression for the voltage drop resulting from a difference between the surface concentration and the average concentration over the diffusion length. The GB model was trained and tested with two different equivalent circuit model elements typically used to represent long-timescale dynamics. The first training model was a Warburg element, implemented in the time domain using an Oustaloup filter to represent the required fractional integrator of order 0.5. The voltage drop over the Warburg element was simulated for similar current profiles as before. After training, the GB modelling results were in good agreement with the training data. Again, the finite-volume elements were observed to get smaller toward the surface. The second training model was an RC circuit, which shows qualitatively different voltage dynamics than the Warburg element. Again, the GB model was able to reproduce the dynamic behavior of the RC circuit. Here, the finite-volume element closest to the surface was observed to be the largest. Both models showed a very good prediction accuracy of the test data.

As a key result, the developed GB model was able, through appropriate training of the learnable parameters, to approximate three different types of dynamics in lithium-ion batteries: concentration dynamics of Fickian diffusion, voltage dynamics of a Warburg element, and voltage dynamics of an RC circuit. Noteworthy, the characteristic dynamics toward long timescales, here in the order of ten to 30 min, could be well reproduced, both in the time domain and in the frequency domain. The GB model should therefore be flexibly applicable to experimental training data, even to

those that do not show ideal Warburg-type or RC-type behavior. In future work we will integrate the GB model into a full ECM of a lithium-ion battery, including open-circuit voltage and ohmic resistances.

In conclusion, we have shown that the use of NODEs can be a powerful methodology for GB modelling of the long-timescale voltage dynamics of batteries.

Acknowledgments

J. B. and R. G. acknowledge funding from the Carl Zeiss Foundation. J. B. is an associated member of the Research Training Group GRK 2218 SiMET—Simulation of mechano-electro-thermal processes in lithium-ion batteries, project number: 281041241. She thanks the German Research Foundation (DFG) for the cooperative support. W. B. would like to thank Marcel Franke and Julia Kowal (TU Berlin, Germany) for providing an initial SIMULINK implementation of the Ostaloup filter for time-domain simulation of the Warburg element.

Appendix

A.1. Finite Volume Discretization of the Fickian Diffusion Equation for a Cuboid Particle.—In this section we derive the equation system for one-dimensional Fickian diffusion in a cartesian coordinate system using finite-volume discretization. This diffusion domain represents cuboid particles of length $l = 2R$. Lithium ions enter the diffusional domain on both sides of the domain with equal flux. Due to this symmetry, we only have to consider half of the particle. The diffusion Eq. 1 is given as

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D \cdot \frac{\partial c}{\partial x}\right), \quad 0 \leqslant x \leqslant \frac{l}{2}. \qquad [A\cdot1]$$

The electrochemical reaction takes place at the particle surface. Considering a (de)intercalation process with a constant flux over the particle surface, the initial and boundary conditions are as follows:[33–36]

$$c = c_{\text{init}}, \quad 0 \leqslant x \leqslant \frac{l}{2}, t = 0 \qquad [A\cdot2a]$$

$$\frac{\partial c}{\partial x} = 0, \quad x = 0, t \geqslant 0 \qquad [A\cdot2b]$$

$$-D\frac{\partial c}{\partial x} = j_{\text{influx}}, \quad x = \frac{l}{2}, t \geqslant 0 \qquad [A\cdot2c]$$

where, c_{init} is the initial concentration and j_{influx} is the molar flux across the particle surface. Note that the particle center is at $x = 0$ and the particle surface is at $x = l/2$.

The numerical discretization of the diffusion Eq. A·1 with the finite volume method leads to the following expression:

$$\frac{\partial \overline{c}_i}{\partial t} \cdot V_i = D\left(\frac{\overline{c}_{i+1} + \overline{c}_i}{2}\right) \cdot A \cdot \frac{\overline{c}_{i+1} - \overline{c}_i}{(x_{i+2} - x_i)/2} - D\left(\frac{\overline{c}_i + \overline{c}_{i-1}}{2}\right) \cdot A \cdot \frac{\overline{c}_i - \overline{c}_{i-1}}{(x_{i+1} - x_{i-1})/2} \quad \forall\, i = 1, \ldots, N-2 \qquad [A\cdot3]$$

where, $N \in \mathbb{N}$ is the number of discretization volumes chosen, V_i the volume of the discretization volume between $[x_i, x_{i+1}]$, $\overline{c}_i$ the average concentration within its volume, and A the area. The volume V_i is given by $V_i = A \cdot (x_{i+1} - x_i)$. Taking into account the initial and boundary conditions according to the Eqs. A·2, we can reformulate Eq. A·3 to

$$\frac{\partial \overline{c}_i}{\partial t} = \frac{j_{i+1} - j_i}{x_{i+1} - x_i}, \quad 0 \leqslant i \leqslant N-1 \qquad [A\cdot4a]$$

$$j_0 = 0 \qquad [A\cdot4b]$$

$$j_i = 2\frac{D\left(\frac{\overline{c}_i + \overline{c}_{i-1}}{2}\right) \cdot (\overline{c}_i - \overline{c}_{i-1})}{x_{i+1} - x_{i-1}}, \quad 1 \leqslant i \leqslant N-1 \qquad [A\cdot4c]$$

$$j_N = j_{\text{influx}} \qquad [A\cdot4d]$$

$$\overline{c}_{i\,0} = c_{\text{init}}, \quad 0 \leqslant i \leqslant N-1, t = 0 \qquad [A\cdot4e]$$

where, $\overline{c}_{i\,0}$ is the initial average concentration in the discretization volume i. Additionally, we used the extrapolation according to the Eq. 8 to calculate the surface concentration.

In order to minimize the influence of rounding errors, the equations were (partially) nondimensionalized.[34–36] For this purpose, we used the nondimensionalized quantities according to the Eq. 9. Using Eqs. A·4 and 8 under the assumption of a proportional relationship between the molar flux across the particle surface and the battery current, this results in the following final equation system:

$$\frac{\partial \overline{C}_i}{\partial t} = \frac{\delta_{i+1} - \delta_i}{z_{i+1} - z_i}, \quad 0 \leqslant i \leqslant N-1 \qquad [A\cdot5a]$$

$$\delta_0 = 0 \qquad [A\cdot5b]$$

$$\delta_i = 2\frac{D^*\left(\frac{\overline{C}_i + \overline{C}_{i-1}}{2}\right) \cdot (\overline{C}_i - \overline{C}_{i-1})}{z_{i+1} - z_{i-1}}, \quad 1 \leqslant i \leqslant N-1 \qquad [A\cdot5c]$$

$$\delta_N = b \cdot I_{\text{bat}} \qquad [A\cdot5d]$$

$$\overline{C}_{i\,0} = C_{\text{init}}, \quad 0 \leqslant i \leqslant N-1 \qquad [A\cdot5e]$$

$$C_{\text{S}} = \frac{3\overline{C}_{N-1} - \overline{C}_{N-2}}{2} \qquad [A\cdot5f]$$

with the proportionality factor b including the nondimensionalization of the quantities.

A.2. Estimation of initial values of the learnable parameters.—This section shows how to estimate the initial values of the learnable parameters.

*A.2.1. Fickian diffusion.—*The normalization and initialization of the learnable parameters for GB modelling Fickian diffusion are described above. Here, we show how to estimate the learnable parameter α_5 in the Eqs. 28 to complement the information.

The proportionality factor between the battery current and the partially nondimensionalized molar flux δ_N across the particle surface in Eq. A·5d is unknown. However, we used the proportionality factor between the two quantities that results for a spherical particle. It is given by $-\frac{1}{3\varepsilon FV(c_{\max} - c_{\min})} = -\frac{1}{3 \cdot 0.554 \cdot 96485\ \text{A s mol}^{-1} \cdot 720.3\ \text{cm}^3 \cdot 30000\ \text{mol m}^{-3}} = -2.886 \cdot 10^{-7}\ \text{A}^{-1}\,\text{s}^{-1}$ according to Eq. 10d. We had to divide this value by the length of the fifth discretization volume $z_5 - z_4 \approx 0.2$ to get an estimation for the learnable parameter α_5 (see Eqs. 22a and A·5a). Therefore, in the optimal case, the learnable parameter has the value $\frac{-2.886 \cdot 10^{-7}\ \text{A}^{-1}\,\text{s}^{-1}}{0.2} = -1.443 \cdot 10^{-6}\ \text{A}^{-1}\,\text{s}^{-1}$. For an irregular distribution of discretization volumes, we would expect the discretization volumes near the surface to be smaller than the estimated value. The absolute value of the parameter α_5 would be higher. Therefore we introduced the scaling factor 10^{-5} and initialized the normalized parameter $\alpha_5^* = 10^5 \cdot \alpha_5$ with $0.5\ \text{A}^{-1}\,\text{s}^{-1}$. Note that the proportionality factor between the current and the

flux is negative (we define positive current for delithiation). The minus sign was directly implemented in the equation system.

A.2.2. Warburg element and resistor-capacitor circuit.—The normalization and initialization of the learnable parameters for GB modelling long-timescale battery dynamics is described above. Here we show how to estimate the learnable parameters to complement the information.

As the other model parameters are the same as for modelling Fickian diffusion, we only had to provide an estimation for the learnable parameter ω_1 in Eq. 28g. This parameter represents among other things the factor a in Eq. 25. The absolute value of the factor a is given by $\frac{R_g T}{F} = \frac{8.3145\ \text{J mol}^{-1}\ \text{K}^{-1}\cdot 298.15\ \text{K}}{96485\ \text{A s mol}^{-1}} = 2.569\cdot 10^{-2}\ \text{V}$.

The learnable parameter ω_1 is affected by the conversion factors for nondimensionalization of the lithium concentrations. Additionally, one has to keep in mind that we used a linearization of the logarithmic function. Therefore, the parameter is difficult to estimate. Due to the shape of the voltage response for pulsed charging and discharging (see true voltages in Figs. 7 and 11), we expected a higher value for the Warburg element than for the RC element. We introduced the scaled parameter $\omega_1^* = \omega_1 \cdot 10$ and initialized it with 0.3 V for the Warburg element and with 0.02 V for the RC element.

ORCID

Jennifer Brucker https://orcid.org/0000-0003-3334-2819
Wolfgang G. Bessler https://orcid.org/0000-0001-8037-9046

References

1. U. Krewer, F. Röder, E. Harinath, R. D. Braatz, B. Bedürftig, and R. Findeisen, "Review–dynamic models of li-ion batteries for diagnosis and operation: a review and perspective." *J. Electrochem. Soc.*, **165**, A3656 (2018).
2. E. Goldammer and J. Kowal, "Determination of the distribution of relaxation times by means of pulse evaluation for offline and online diagnosis of lithium-ion batteries." *Batteries*, **7**, 36 (2021).
3. M. Schönleber, C. Uhlmann, P. Braun, A. Weber, and E. Ivers-Tiffée, "A consistent derivation of the impedance of a lithium-ion battery electrode and its dependency on the state-of-charge." *Electrochimica Acta*, **243**, 250 (2017).
4. M. Quarti and W. G. Bessler, "Model-based overpotential deconvolution, partial impedance spectroscopy, and sensitivity analysis of a lithium-ion cell with blend cathode." *Energy Technology*, **9**, 2001122 (2021).
5. M. Petzl and M. A. Danzer, "Advancements in ocv measurement and analysis for lithium-ion batteries." *IEEE Transactions on Energy Conversion*, **28**, 675 (2013).
6. M. S. Hosen, R. Gopalakrishnan, T. Kalogiannis, J. Jaguemont, J. van Mierlo, and M. Berecibar, "Impact of relaxation time on electrochemical impedance spectroscopy characterization of the most common lithium battery technologies–experimental study and chemistry-neutral modeling." *World Electric Vehicle Journal*, **12**, 77 (2021).
7. L. Pei, R. Lu, and C. Zhu, "Relaxation model of the open-circuit voltage for state-of-charge estimation in lithium-ion batteries." *IET Electrical Systems in Transportation*, **3**, 112 (2013).
8. L. Pei, T. Wang, R. Lu, and C. Zhu, "Development of a voltage relaxation model for rapid open-circuit voltage prediction in lithium-ion batteries." *Journal of Power Sources*, **253**, 412 (2014).
9. A. Jossen, "Fundamentals of battery dynamics." *Journal of Power Sources*, **154**, 530 (2006).
10. L. O. Valøen and J. N. Reimers, "Transport properties of lipf6-based li-ion battery electrolytes." *J. Electrochem. Soc.*, **152**, A882 (2005).
11. S. Paarmann, L. Cloos, J. Technau, and T. Wetzel, "Measurement of the temperature influence on the current distribution in lithium-ion batteries." *Energy Technology*, **9**, 2000862 (2021).
12. J. P. Fath, L. Alsheimer, M. Storch, J. Stadler, J. Bandlow, S. Hahn, R. Riedel, and T. Wetzel, "The influence of the anode overhang effect on the capacity of lithium-ion cells–a 0d-modeling approach." *Journal of Energy Storage*, **29**, 101344 (2020).
13. A. A. Franco et al. (ed.), *Physical Multiscale Modeling and Numerical Simulation of Electrochemical Devices for Energy Conversion and Storage: From Theory to Engineering to Practice* (Springer, London) 1st ed. (2016), Green Energy and Technology (10.1007/978-1-4471-5677-2).
14. S. Estrada-Flores, I. Merts, B. de Ketelaere, and J. Lammertyn, "Development and validation of "Grey-box" models for refrigeration applications: a review of key concepts." *International Journal of Refrigeration*, **29**, 931 (2006).
15. Y. Oussar and G. Dreyfus, "How to be a gray box: dynamic semi-physical modeling." *Neural Networks*, **14**, 1161 (2001).
16. B. Duarte, P. M. Saraiva, and C. C. Pantelides, "Combined mechanistic and empirical modelling." *International Journal of Chemical Reactor Engineering*, **2**, A3 (2004).
17. F. Hamilton, A. L. Lloyd, and K. B. Flores, "Hybrid modeling and prediction of dynamical systems." *PLoS Computational Biology*, **13**, e1005655 (2017).
18. M. Doyle, T. F. Fuller, and J. Newman, "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell." *J. Electrochem. Soc.*, **140**, 1526 (1993).
19. C. Kupper and W. G. Bessler, "Multi-scale thermo-electrochemical modeling of performance and aging of a $LiFePO_4$/graphite lithium-ion cell." *J. Electrochem. Soc.*, **164**, A304 (2017).
20. E. Barsoukov and J. R. Macdonald, *Impedance Spectroscopy: Theory, Experiment, and Applications* (Wiley, Hoboken, New Jersey) 2nd ed. (2005).
21. M. Hu, Y. Li, S. Li, C. Fu, D. Qin, and Z. Li, "Lithium-ion battery modeling and parameter identification based on fractional theory." *Energy*, **165**, 153 (2018).
22. J. Valsa and J. Vlach, "Rc models of a constant phase element." *International Journal of Circuit Theory and Applications*, **41**, 59 (2011).
23. I. Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of their Solution and Some of their Applications, volume 198 of Mathematics in Science and Engineering* (Academic Press, San Diego) (1999).
24. A. G. Li, K. Mayilvahanan, A. C. West, and M. Preindl, "Discrete-time modeling of li-ion batteries with electrochemical overpotentials including diffusion." *Journal of Power Sources*, **500**, 229991 (2021).
25. W. Weppner and R. A. Huggins, "Determination of the kinetic parameters of mixed-conducting electrodes and application to the system Li_3Sb." *J. Electrochem. Soc.*, **124**, 1569 (1977).
26. J. A. Braun, R. Behmann, D. Schmider, and W. G. Bessler, "State of charge and state of health diagnosis of batteries with voltage-controlled models." *Journal of Power Sources*, **544**, 231828 (2022).
27. M. Mayur, S. C. DeCaluwe, B. L. Kee, and W. G. Bessler, "Modeling and simulation of the thermodynamics of lithium-ion battery intercalation materials in the open-source software cantera." *Electrochimica Acta*, **323**, 134797 (2019).
28. S. Santhanagopalan, Q. Guo, P. Ramadass, and R. E. White, "Review of models for predicting the cycling performance of lithium ion batteries." *Journal of Power Sources*, **156**, 620 (2006).
29. A. Jokar, B. Rajabloo, M. Désilets, and M. Lacroix, "Review of simplified pseudo-two-dimensional models of lithium-ion batteries." *Journal of Power Sources*, **327**, 44 (2016).
30. F. Hall, J. Touzri, S. Wußler, H. Buqa, and W. G. Bessler, "Experimental investigation of the thermal and cycling behavior of a lithium titanate-based lithium-ion pouch cell." *Journal of Energy Storage*, **17**, 109 (2018).
31. R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations." (2018), arXiv:1806.07366 [cs.LG].
32. R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena* (Wiley, New York and Chichester and Weinheim and Brisbane and Singapore and Toronto) 2nd ed. (2007).
33. T. F. Fuller, M. Doyle, and J. Newman, "Simulation and optimization of the dual lithium ion insertion cell." *J. Electrochem. Soc.*, **141**, 1 (1994).
34. V. R. Subramanian, J. A. Ritter, and R. E. White, "Approximate solutions for galvanostatic discharge of spherical particles: I. Constant diffusion coefficient." *J. Electrochem. Soc.*, **148**, E444 (2001).
35. M. Guo and R. E. White, "An approximate solution for solid-phase diffusion in a spherical particle in physics-based li-ion cell models." *Journal of Power Sources*, **198**, 322 (2012).
36. S. Liu, "An analytical solution to Li/Li^+ insertion into a porous electrode." *Solid State Ionics*, **177**, 53 (2006).
37. A. M. Bizeray, J.-H. Kim, S. R. Duncan, and D. A. Howey, "Identifiability and parameter estimation of the single particle lithium-ion battery model." *IEEE Transactions on Control Systems Technology*, **27**, 1862 (2019).
38. Y. Zeng, P. Albertus, R. Klein, N. Chaturvedi, A. Kojic, M. Z. Bazant, and J. Christensen, "Efficient conservative numerical schemes for 1d nonlinear spherical diffusion equations with applications in battery modeling." *J. Electrochem. Soc.*, **160**, A1565 (2013).
39. M. C. Yagci, R. Behmann, V. Daubert, J. A. Braun, D. Velten, and W. G. Bessler, "Electrical and structural characterization of large-format lithium iron phosphate cells used in home-storage systems." *Energy Technology*, **9**, 202000911 (2021).
40. J. H. Park, H. Yoon, Y. Cho, and C.-Y. Yoo, "Investigation of lithium ion diffusion of graphite anode by the galvanostatic intermittent titration technique." *Materials*, **14**, 4683 (2021).
41. P. Ramadass, B. Haran, P. M. Gomadam, R. White, and B. N. Popov, "Development of first principles capacity fade model for li-ion cells." *J. Electrochem. Soc.*, **151**, A196 (2004).
42. M. Guo, G. Sikha, and R. E. White, "Single-particle model for a lithium-ion cell: thermal behavior." *J. Electrochem. Soc.*, **158**, A122 (2011).
43. R. T. Q. Chen, "torchdiffeq." (2021), https://github.com/rtqichen/torchdiffeq, accessed on 23.02.2022.
44. J. Illig, *Physically based Impedance Modelling of Lithium-Ion Cells* (KIT Scientific Publishing, Karlsruhe) Schriften des Instituts für Werkstoffe der Elektrotechnik, 27 (2014).
45. C. Fleischer, W. Waag, H.-M. Heyn, and D. U. Sauer, "On-line adaptive battery impedance parameter and state estimation considering physical principles in reduced order equivalent circuit battery models: part 1. Requirements, critical review of methods and modeling." *Journal of Power Sources*, **260**, 276 (2014).
46. M. E. Fouda, A. S. Elwakil, A. G. Radwan, and A. Allagui, "Power and energy analysis of fractional-order electrical energy storage devices." *Energy*, **111**, 785 (2016).
47. A. Tepljakov, E. Petlenkov, and J. Belikov, "Fomcom: a matlab toolbox for fractional-order system identification and control." *International Journal of Microelectronics and Computer Science*, **2**, 51 (2011).

48. A. Oustaloup, F. Levron, B. Mathieu, and F. M. Nanot, "Frequency-band complex noninteger differentiator: characterization and synthesis." *IEEE Trans. Circuits Syst. I*, **47**, 25 (2000).
49. S Das, I Pan, S Saha, A Kumar, S Das, and A Gupta, "Revisiting oustaloup's recursive filter for analog realization of fractional order differintegrators." *2011 International Conference on Energy, Automation and Signal, IEEE*, Piscataway, NJ, 1 (2011).
50. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, Massachusetts) (2016), Adaptive computation and machine learning.
51. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." *XXIX IEEE Conference on Computer Vision and Pattern Recognition. IEEE*, Piscataway, NJ 770 (2016).
52. E. Haber and L. Ruthotto, "Stable architectures for deep neural networks." *Inverse Problems*, **34**, 014004 (2018).
53. L. Ruthotto and E. Haber, "Deep neural networks motivated by partial differential equations." *Journal of Mathematical Imaging and Vision*, **62**, 352 (2020).
54. E. Dupont, A. Doucet, and Y. W. Teh, "Augmented neural odes." *Advances in Neural Information Processing Systems 32, Curran Associates, Inc.*, Red Hook, NY, USA3140 (2019).
55. T. Zhang, Z. Yao, A. Gholami, K. Keutzer, J. Gonzalez, G. Biros, and M. W. Mahoney, "Anodev2: a coupled neural ode evolution framework." (2019), arXiv:1906.04596 [cs.LG].
56. E. Haber, L. Ruthotto, E. Holtham, and S.-H. Jun, "Learning across scales—multiscale methods for convolution neural networks.",Proceedings of the AAAI Conference on Artificial Intelligence, 32 **1** (2018).
57. A. Gholami, K. Keutzer, and G. Biros, "Anode: unconditionally accurate memory-efficient gradients for neural odes." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* 730 (2019).
58. J. Brucker, W. G. Bessler, and R. Gasper, "Grey-box modelling of lithium-ion batteries using neural ordinary differential equations." *Energy Informatics*, **4**, 15 (2021).
59. C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman, (2020)"Universal differential equations for scientific machine learning." arXiv:2001.04385 [cs.LG].
60. A. Paszke et al., "Pytorch: an imperative style, high-performance deep learning library." *Advances in Neural Information Processing Systems 32, Curran Associates, Inc.*, Red Hook, NY, USA8024 (2019).
61. Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop." *Neural Networks: Tricks of the Trade*, ed. G. B Orr and K-R Müller (Springer, Berlin, Heidelberg) (1998).
62. W. G. Bessler, "Rapid impedance modeling via potential step and current relaxation simulations." *J. Electrochem. Soc.*, **154**, B1186 (2007).
63. A. C. Hindmarsh, "Odepack, a systematized collection of ode solvers." *Scientific Computing*, ed. R. S Stepleman et al. (North-Holland, Amsterdam) IMACS Transactions on Scientific Computation,155 (1983).