RWTH AACHEN UNIVERSITY

AI-based simulation of battery system combined with advanced spectroscopy

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Limei Jin, M.Sc.

aus Jilin, China

Berichter: Prof. Dr. rer. nat. Josef Granwehr
Prof. Dr.-Ing. Ferdinanda Ponci
Prof. Dr. rer. nat. Arne Lüchow

Tag der mündlichen Prüfung: 31.10.2024

Acknowledgements

This doctoral thesis was conducted within the project "Intelligent, individual battery management using spectroscopy and machine learning" (i2Batman), which has been financially supported by the Helmholtz AI Cooperation Unit (HAICU). The project involves collaboration with an interdisciplinary team of natural scientists, engineers and mathematicians from the Institute of Energy and Climate research (IEK-9) at Forschungszentrum Jülich, the Steinbruch Centre for Computing at Karlsruhe Institute of Technology (KIT) and the Theory Department at the Fritz-Haber-Institute (FHI) of the Max Planck Society in Berlin.

First of all, I would like to express my gratitude to my first supervisor, Prof. Dr. Josef Granwehr, for providing me with the opportunity to work in spectroscopy group at IEK-9, and his kind support both in scientific endeavors and administrative matters.

I would also like to extend my sincere thanks to my daily supervisor, Dr. Christoph Scheurer, for guiding me through the research and for providing me with the chance to work in the theory group at FHI. His scientific insights have played a crucial role in shaping this work.

Additionally, I would like to thank Franz Bereck for our enjoyable collaboration on the experiment and theoretical analysis.

Along with them, I would like to thank Dr. Ivan Kondov and Hasan Kalkan from KIT for interesting discussions and collaborative efforts.

Special thanks go to the IEK-9 director, Prof. Dr. Rüdiger-A. Eichel, and director of theory group at FHI, Prof. Dr. Karsten Reuter, for supporting this collaborative project.

And of course I would like to express my appreciation to all my colleagues at IEK-9 and at FHI for their assistance in daily life.

Finally, my deepest thanks go to my parents, whose unwavering support and constant encouragement have been a driving force in my career development.

Eidesstattliche Erklärung

I, Limei Jin, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.
I do solemnely swear that:

1. This work was done wholly or mainly while in candidature for the doctoral degree at this faculty and university;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others or myself, this is always clearly attributed;

4. Where I have quoted from the work of others or myself, the source is always given. This thesis is entirely my own work, with the exception of such quotations;

5. I have acknowledged all major sources of assistance;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Parts of this work will be published as:

   I. **Extending equivalent circuit models for state of charge and lifetime estimation**

      Limei Jin, Franz Bereck, Josef Granwehr, and Christoph Scheurer

      Submitted to the journal and under review

   II. **Lifetime monitoring for lithium-ion battery in the low dimensional latent space**

      Limei Jin, Franz Bereck, Rüdiger-A. Eichel, Josef Granwehr, and Christoph Scheurer

Submitted to the journal and under review

III. **Online battery state analysis using stochastic pulse excitation**

Limei Jin, Franz Bereck, Rüdiger-A. Eichel, Josef Granwehr, and Christoph Scheurer

to be submitted soon

IV. **Weighted distribution of relaxation time analysis of battery impedance spectra using Gaussian process regression for noise estimation**

Christian Bartsch, Limei Jin, Franz Bereck, Andreas Mertens, Rüdiger-A. Eichel, Christoph Scheurer, and Josef Granwehr

Submitted to the journal and under review

Place, Date **Limei Jin**

Abstract

The surge in electric vehicle adoption necessitates the development of high-performance and reliable battery systems. To optimize battery performance and ensure their safe and efficient operation, the Battery Management System (BMS) plays a crucial role. While BMS is commonly utilized at the pack level to oversee the overall health and performance of a battery pack, the recognition of each individual cell with unique operational characteristics within the pack needs more granular and precise management and monitoring at the cell level. Therefore, battery digital twin, which is a real-time, multi-scale virtual representation of the physical battery cell, comes into view. We aim to develop a simplest possible equivalent circuit model-based digital twin. It is designed to accurately reflect the behavior of the actual battery cell under various load conditions. The virtual battery model serves to simulate these conditions and to obtain a wealth of simulated data accordingly. These large amount of data can then be employed by machine learning algorithms to predict the battery's state, providing optimization feedback to enhance the real battery's operation.

Equivalent Circuit Modeling (ECM) of Electrochemical Impedance Spectroscopy (EIS) data is a common technique to describe the current state of batteries and can be used as a virtual battery. To characterize ECM parameters in a data supported way, the Gaussian Process Regression (GPR) based Distribution of Relaxation Times (DRT) technique, which simplify EIS data by deconvolution with a suitable kernel, is used to provide the number of distinguishable features based solely on an individual EIS data set. Here a weighted DRT is employed, where weights are tailored based on the frequency-dependent sensitivity of the data, and GPR is utilized to accurately determine these weights. GPR is capable of estimating an suitable weighting matrix from a single data set, potentially enabling automatized DRT inversion without user intervention. The obtained DRT spectrum is then used for the selection of an equivalent circuit model, its initial parametrization, and setting of constraints. Thereby, the ECM parameters that are fitted to experimental EIS data for a single cycle, vary by the State of Charge (SOC). Eventually, by means of the investigated battery it is discussed that using a combination of DRT and ECM, a more physically relevant description of

processes in an electrochemical system can be achieved.

Unlike SOC, developing a robust and continuous State of Health (SOH) estimation using EIS to build battery digital twins poses a formidable challenge. To bridge the gap between the experiment with data driven techniques that do not rely on fitting of experimental data using *a priori* models, ECM parameters over a single cycle are expanded in a high-dimensional Chebyshev space. It facilitates not only a mapping of the SOC dependence with robust boundary conditions, but also an extension towards a more abstract SOH description is possible. Due to the long impedance measurement time, the Quasi Monte Carlo (QMC) method can be employed to generate differently aged battery models with limited experimental impedance data. As data becomes available, the space spanning the possible states of a battery can be gradually refined. The developed framework, therefore, allows for the training of big data models starting with very little experimental information and assuming random fluctuations of the model parameters consistent with available data.

Battery data, whether obtained through direct measurement or simulation, is inherently complex and high-dimensional. Therefore, detecting anomalies or deciphering the aging pathways from raw data poses significant challenges. To harness effective battery state management, we propose to extract characteristic information from autoencoder latent spaces. First of all, we undertake a comparative analysis of the latent spaces derived from frequency-based impedance data, and from time-based voltage and current data. This comparison provides confidence in the applicability of methods traditionally used for EIS in laboratory environments to the analysis of raw time-series data. Furthermore, in order to estimate realistic battery usage simulation results by incorporating random noise and variations into the data, which is essential for developing robust battery management systems, we compare the latent space trained from ideal sinusoidal data and that from permuted noisy data. The extracted battery features in the latent space were analyzed using Support Vector Machine (SVM) classifiers with both linear and non-linear kernels to segment batteries into three distinct age groups: *fresh*, *aged* and *damaged* cells. The consistent and observable robust aging motion, as depicted by the distribution of classes in the latent space, remains unaffected by the choice of using time-based or frequency-based data, as well as whether the data is ideal or noisy. It underscores the potential and reliability for the development of pseudo-random complex pulse current excitation that can be used for estimating the battery's state, which can then potentially be utilized to inform the development of optimized cell-based load profiles in the future.

While EIS offers valuable insights into a battery's state, real-world battery operation during driving scenarios involves dynamic state changes, where current and voltage

signals are far from ideally sinusoidal. To bridge the gap between EIS and real-world driving cycle analysis, we introduce the concept of a stochastic pulse design compatible with the load profile. This approach starts with frequency-based impedance data as a reference and transitions into stochastic pulse signals that act as a weak perturbation in the time domain, thereby providing optimal contrast regrading the battery's state. The systematic generation of stochastic pulse signals is inspired by fractal curves and is limited by constraints of experimental devices. These signals are designed to mimic the variability and unpredictability of real-world battery usage more closely. The current pulse signals are then simulated through the battery models at different SOC/SOH combinations, yielding the corresponding voltage responses. The analysis of this pair-wise current/voltage data is also conducted in the latent space of an Autoencoder, which comprises essential features extracted from the input data. Here, traditional EIS data is computed as a reference standard to compare the effects of different pulse sequences against a well-established benchmark in battery analysis. For a visual and quantitative comparison between two latent spaces, Quantile-Quantile (Q-Q) plots are employed to estimate performance among different pulse sequences when replicating the distribution of the EIS-derived latent space. Additionally, their performance can also be visual by latent space segmentation. By correlating the segmentation of latent space with battery aging indicators, we validate the representativeness of our identified best and worst performed pulse signals against conventional EIS data. Furthermore, at certain stages of a battery's aging process, altering the pulse sequence can lead to enhanced performance in terms of age group separation. This insight underscores the potential for adaptive strategies in battery management, where pulse sequences can be dynamically adjusted based on the battery's stage of aging to optimize performance and possibly extend its lifespan.

Zusammenfassung

Der Anstieg bei der Adoption von Elektrofahrzeugen macht die Entwicklung leistungsfähiger und zuverlässiger Batteriesysteme notwendig. Um die Leistung der Batterie zu optimieren und ihren sicheren und effizienten Betrieb zu gewährleisten, spielt das Batteriemanagementsystem (BMS) eine entscheidende Rolle. Während das BMS üblicherweise auf der Ebene des Batteriepakets eingesetzt wird, um den Gesamtzustand und die Leistung eines Batteriepakets zu überwachen, erfordert die Erkennung jeder einzelnen Zelle mit einzigartigen Betriebseigenschaften innerhalb des Pakets eine detailliertere und präzisere Verwaltung und Überwachung auf Zellebene. Daher kommt der digitale Zwilling der Batterie ins Spiel, eine Echtzeit-, mehrskalige virtuelle Darstellung der physischen Batteriezelle. Unser Ziel ist es, einen möglichst einfachen digitalen Zwilling auf der Grundlage eines Ersatzschaltbildes zu entwickeln. Es ist darauf ausgelegt, das Verhalten der tatsächlichen Batteriezelle unter verschiedenen Lastbedingungen genau zu reflektieren. Das virtuelle Batteriemodell dient dazu, diese Bedingungen zu simulieren und entsprechend eine Fülle von Simulationsdaten zu gewinnen. Diese großen Datenmengen können dann von Algorithmen des maschinellen Lernens verwendet werden, um den Zustand der Batterie vorherzusagen und Optimierungsrückmeldungen zu geben, um den Betrieb der realen Batterie zu verbessern.

Die Ersatzschaltkreismodellierung (ECM) von Daten der elektrochemischen Impedanzspektroskopie (EIS) ist eine gängige Technik zur Beschreibung des aktuellen Zustands von Batterien und kann als virtuelle Batterie verwendet werden. Um ECM-Parameter auf eine datengestützte Weise zu charakterisieren, wird die auf der Gaußschen Prozessregression (GPR) basierende Verteilung der Relaxationszeiten (DRT) verwendet, welche EIS-Daten durch Deconvolution mit einem geeigneten Kernel vereinfacht und die Anzahl der unterscheidbaren Merkmale basierend allein auf einem individuellen EIS-Datensatz bereitstellt. Hier wird eine gewichtete DRT eingesetzt, bei der die Gewichte auf der Grundlage der frequenzabhängigen Empfindlichkeit der Daten zugeschnitten sind, und GPR wird zur genauen Bestimmung dieser Gewichte verwendet. GPR ist in der Lage, eine geeignete Gewichtungsmatrix aus einem einzigen Datensatz zu schätzen und ermöglicht potenziell eine automatisierte DRT-Inversion

ohne Benutzereingriff. Das erhaltene DRT-Spektrum wird dann für die Auswahl eines Ersatzschaltkreismodells, dessen initiale Parametrisierung und das Setzen von Einschränkungen verwendet. Dadurch variieren die an experimentelle EIS-Daten für einen einzigen Zyklus angepassten ECM-Parameter je nach Ladezustand (SOC). Anhand der untersuchten Batterie wird schließlich erörtert, dass mit einer Kombination aus DRT und ECM eine physikalisch relevantere Beschreibung der Prozesse in einem elektrochemischen System erreicht werden kann.

Im Gegensatz zum SOC stellt die Entwicklung einer robusten und kontinuierlichen Schätzung des Gesundheitszustands (SOH) der Batterie unter Verwendung von EIS zur Erstellung digitaler Batteriezwillingsmodelle eine erhebliche Herausforderung dar. Um die Lücke zwischen dem Experiment und datengesteuerten Techniken zu schließen, die sich nicht auf die Anpassung experimenteller Daten unter Verwendung von *a priori* Modellen stützen, werden die ECM-Parameter über einen einzelnen Zyklus in einem hochdimensionalen Chebyscheff-Raum erweitert. Dies erleichtert nicht nur eine Abbildung der SOC-Abhängigkeit mit robusten Randbedingungen, sondern ermöglicht auch eine Erweiterung hin zu einer abstrakteren SOH-Beschreibung. Aufgrund der langen Impedanzmesszeit kann die quasi Monte Carlo (QMC) Methode eingesetzt werden, um unterschiedlich gealterte Batteriemodelle mit begrenzten experimentellen Impedanzdaten zu generieren. Sobald Daten verfügbar sind, kann der Raum, der die möglichen Zustände einer Batterie umspannt, schrittweise verfeinert werden. Das entwickelte Framework ermöglicht daher das Training von Big-Data-Modellen, beginnend mit sehr wenigen experimentellen Informationen und unter der Annahme von zufälligen Schwankungen der Modellparameter, die mit den verfügbaren Daten übereinstimmen.

Batteriedaten, ob durch direkte Messung oder Simulation gewonnen, sind von Natur aus komplex und hochdimensional. Daher stellt die Erkennung von Anomalien oder die Entschlüsselung der Alterungspfade aus Rohdaten eine große Herausforderung dar. Um ein effektives Batteriezustandsmanagement zu ermöglichen, schlagen wir vor, charakteristische Informationen aus den Latent Space von Autoencodern zu extrahieren. Zunächst führen wir eine vergleichende Analyse der aus frequenzbasierten Impedanzdaten und aus zeitbasierten Spannungs- und Stromdaten abgeleiteten Latent Spaces durch. Dieser Vergleich stärkt das Vertrauen in die Anwendbarkeit von Methoden, die traditionell für EIS in Laborumgebungen verwendet werden, auf die Analyse von rohen Zeitreihendaten. Um realistische Simulationsergebnisse für die Batterienutzung zu ermitteln, indem wir zufälliges Rauschen und Schwankungen in die Daten einbeziehen, was für die Entwicklung robuster Batteriemanagementsysteme unerlässlich ist, vergleichen wir den Latent Spaces, der anhand idealer sinusförmiger Daten trainiert wurde, mit dem aus permutierten verrauschten Daten. Die extrahierten Batteriemerkmale

im Latent Space wurden mit Support-Vektor-Maschinen (SVM)-Klassifikatoren mit sowohl linearen als auch nicht-linearen Kernels analysiert, um Batterien in drei unterschiedliche Altersgruppen zu segmentieren: *frische*, *gealterte* und *beschädigte* Zellen. Die konsistente und beobachtbare robuste Alterungsbewegung, wie durch die Verteilung der Klassen im Latent Space dargestellt, bleibt unbeeinflusst von der Wahl der Nutzung zeitbasierter oder frequenzbasierter Daten, sowie davon, ob die Daten ideal oder verrauscht sind. Dies unterstreicht das Potenzial und die Zuverlässigkeit für die Entwicklung von pseudo-zufälligen komplexen Pulsstromanregungen, die zur Schätzung des Batteriezustands verwendet werden können, welcher dann potenziell zur Information der Entwicklung von optimierten zellbasierten Lastprofilen in der Zukunft genutzt werden könnte.

Während EIS wertvolle Einblicke in den Zustand einer Batterie bietet, beinhaltet der reale Batteriebetrieb in Fahrszenarien dynamische Zustandsänderungen, bei denen Strom- und Spannungssignale bei weitem nicht ideal sinusförmig sind. Um die Lücke zwischen EIS und realer Fahrzyklusanalyse zu schließen, führen wir das Konzept eines stochastischen Pulsdesigns ein, das mit dem Lastprofil kompatibel ist. Dieser Ansatz beginnt mit frequenzbasierten Impedanzdaten als Referenz und geht über zu stochastischen Pulssignalen, die als schwache Störung im Zeitbereich wirken und somit optimalen Kontrast bezüglich des Batteriezustands bieten. Die systematische Erzeugung von stochastischen Pulssignalen ist von fraktalen Kurven inspiriert und wird durch die Beschränkungen der experimentellen Geräte limitiert. Diese Signale sind so konzipiert, dass sie die Variabilität und Unvorhersehbarkeit des realen Batteriegebrauchs enger nachahmen. Die aktuellen Pulssignale werden dann durch die Batteriemodelle bei verschiedenen SOC/SOH-Kombinationen simuliert, was die entsprechenden Spannungsantworten liefert. Die Analyse dieser paarweisen Strom-/Spannungsdaten wird auch im Latent Space eines Autoencoders durchgeführt, der wesentliche aus den Eingabedaten extrahierte Merkmale umfasst. Hier wird traditionelle EIS-Daten als Referenzstandard berechnet, um die Effekte verschiedener Pulssequenzen gegenüber einem gut etablierten Benchmark in der Batterieanalyse zu vergleichen. Für einen visuellen und quantitativen Vergleich zwischen zwei Latent Spaces werden Quantile-Quantile (Q-Q) Diagramme verwendet, um die Leistung unter verschiedenen Pulssequenzen bei der Replikation der Verteilung des aus EIS abgeleiteten Latent Space zu schätzen. Zusätzlich kann ihre Leistung auch durch die Segmentierung des Latent Spaces visualisiert werden. Indem wir die Segmentierung des Latent Spaces mit Batteriealterungsindikatoren korrelieren, validieren wir die Repräsentativität unserer identifizierten besten und schlechtesten Pulssignale gegenüber konventionellen EIS-Daten. Weiterhin kann in bestimmten Phasen des Alterungsprozesses einer Batterie

die Änderung der Pulssequenz zu einer verbesserten Leistung hinsichtlich der Trennung von Altersgruppen führen. Diese Erkenntnis unterstreicht das Potenzial für adaptive Strategien im Batteriemanagement, bei denen Pulssequenzen dynamisch angepasst werden können, basierend auf dem Alterungsstadium der Batterie, um die Leistung zu optimieren und möglicherweise ihre Lebensdauer zu verlängern.

Contents

List of Figures

List of Tables

Abbreviations

ECM	Equivalent Circuit Modeling
EIS	Electrochemical Impedance Spectroscopy
GPR	Gaussian Process Regression
DRT	Distribution of Relaxation Times
SOC	State of Charge
SOH	State of Health
QMC	Quasi Monte Carlo
SVM	Support Vector Machine
BMS	Battery Management System
DT	Digital Twin
ML	Machine Learning
Q-Q	Quantile-Quantile
AC	Alternating Current
CPE	Constant Phase Element
FLW	Finite Length Warburg
FSW	Finite Space Warburg
OCV	Open Circuit Voltage
RC	Resistor Capacitor
XOR	Exclusive OR
HIL	Hardware In the Loop
LS	Latent Space
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
CAE	Convolutional Autoencoder
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
RBF	Radial Basis Function
RC	Resistor-Capacitor
RUL	Remaining Useful Life

RMSE	Root Mean Squared Error
SEI	Solid Electrolyte Interface
CV	Cross-Validation

Chapter 1

Introduction

In recent years, the rapid market growth of electric vehicles as environmentally friendly mobility solution has accentuated the importance of high-performance battery systems [1]. These systems typically comprise complex battery packs composed of numerous cells working in tandem [2]. To optimize battery longevity and ensure their safe and efficient operation, the Battery Management System (BMS) is responsible for monitoring and updating the state of the entire battery pack during operation. However, the BMS primarily works on the overall battery pack, it faces challenges in recognizing and addressing the unique operational characteristics of each individual cell within the pack [3]. This limitation underscores the necessity for a more granular level of management to ensure optimal performance across all cells. To bridge this gap, the introduction of a Digital Twin (DT) enhanced with Machine Learning (ML) at the cell level is proposed. This advanced approach involves creating a virtual model for each cell that mirrors its real-world counterpart in every performance-critical aspect. Thereby, the DT can receive real-time data from each cell, enabling the BMS to distinguish between the unique characteristics and performance of each cell. This detailed level of insight allows for precise monitoring and management, optimizing the health and efficiency of each cell within the pack [4].

Since disassembling a battery or even bringing it to a well-defined thermodynamical state while driving is not feasible, the Electrochemical Impedance Spectroscopy (EIS) technique serves as a valuable tool, providing a high density of information on a cell's state within a comparatively short analysis time [5]. This information is crucial for constructing a robust battery DT as a virtual representation of the physical asset [6]. By leveraging *in operando* impedance data, we establish equivalent circuit models that act as virtual batteries, effectively mapping to the real battery [4].

To detect early signs of battery aging, optimize the actions of the real battery, and provide timely feedback to the user (e.g. a car driver), our model must be predictive.

To achieve this, we incorporate ML techniques in conjunction with a virtual battery model that enables pre-training from simulation data, predict future outcomes and give feedback to control units within the real battery. The extensive initial training data for the neural network uses a comprehensive synthetic dataset, which is compiled from simulations based on theoretical models and a limited set of experimental data for gauging. The simulation process involves the use of detailed equivalent circuit models that represent the intricate processes occurring within a battery, focusing on parameters that allow for estimating State of Charge (SOC) and State of Health (SOH) of a battery. SOC is a measure of the current charge level of the battery relative to its capacity, while SOH reflects the overall condition of the battery and its capacity for storing energy compared to a new battery. For the limited experimental data, we conduct carefully designed experiments that measure the battery's response under controlled conditions. These experiments are crucial for validating the accuracy of our simulations and ensuring that the model reflects all relevant aspects of real battery behavior. Furthermore, the synthetic data training approach facilitates the identification of best-performing, cell adapted load profiles for lifetime improvement through systematic stochastic pulse sequence optimization on the population of virtual batteries, taking into account physical constraints derived from experiments. This gray-box modeling strategy, which integrates physics-based equivalent circuit modeling and data-driven surrogate modeling, enhances the model's efficiency and predictability through iterative refinement and calibration [7]. Importantly, it retains key physical insights into the battery system, ensuring meaningful interpretation within the context of the overall model. The vision of a battery digital twin with cell level granularity that is established in this dissertation is shown in Figure 1.1.

In **Chapter 2** the theoretical understanding of battery modeling techniques addressing directly the macro-scale electric behavior will be discussed. This includes not only the experimental EIS-data based Equivalent Circuit Modeling (ECM) combined with the Distribution of Relaxation Times (DRT) analysis, but also the computational numerical modeling methods involving Chebyshev polynomials and Quasi-Monte Carlo sampling. This comprehensive overview will cover two critical descriptors for the evolution of the battery state on separate timescales: SOC and SOH. Additionally, the real-time simulation framework for battery models used in this work, named ISEA Framework, will be introduced. Furthermore, the chapter will present an exploration of machine learning algorithms, specifically focusing on the autoencoder architecture and Support Vector Machines (SVM). Alongside these machine learning approaches, statistical methods for analysing their results in the latent space, such as Delaunay triangulation and the Quantile-Quantile (Q-Q) plot, will be presented.

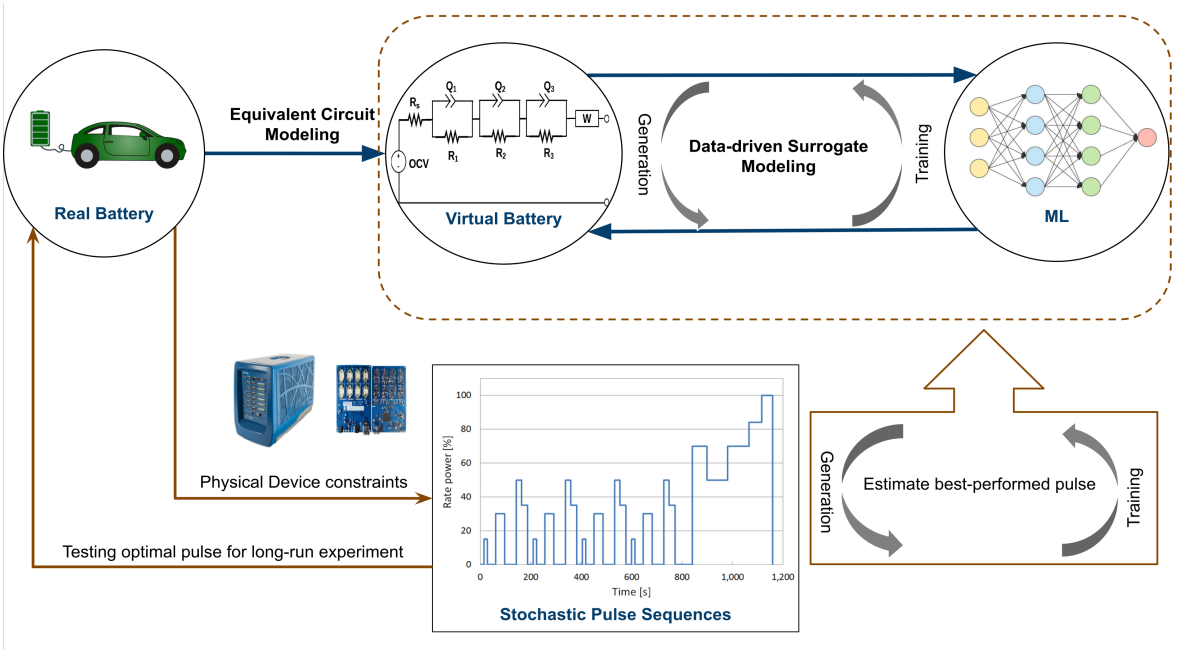


Figure 1.1: The vision of machine learning integrated battery digital twins

In **Chapter 3**, a simplified ECM for robust digital twins and its parameter identification based on experimental EIS data at different SOCs will be discussed. To parametrize the model to EIS data in a physically comprehensive way without over-fitting or under-fitting, the initial assumption and boundaries of ECM fitting are derived by using the DRT method, which transforms an EIS spectrum to the relaxation time domain for extracting separable peak features, particularly for ZARC elements. ZARC is an electrical element occuring in the battery ECM, consisting of a Constant Phase Element (CPE) and an ohmic resistor connected in parallel to represent double-layer capacitance and charge transfer at the electrode-electrolyte interface. The ECM parameters identified in this process are dependent on SOC and are represented in the form of look-up tables for ensuing simulations. However, for more effective lifetime monitoring of the battery, these SOC-dependent parameters should be refined. The goal is to transition from static look-up tables to dynamic functions that flexibly reflect the changing relationship between ECM parameters and a battery's state. SOC changes occur on a relatively short time scale, fluctuating within a single charge-discharge cycle based on immediate power demands and charging protocols. In contrast, SOH evolves over a much longer time scale, reflecting the gradual degradation and loss of capacity of the battery due to repeated charging cycles, aging, and operational conditions. A predictive model aiming to optimize battery performance and longevity, therefore, is required to operate effectively across these disparate time scales. It needs to provide real-time insights for managing SOC to meet immediate energy needs, while also

predicting long-term SOH trends to inform maintenance, replacement schedules, and overall battery management strategies. By dynamically adjusting ECM parameters as a function of SOC, and incorporating a nuanced understanding of SOH progression, the model can offer a comprehensive and accurate tool for battery state monitoring and decision-making.

Chapter 4 aims to employ Chebyshev polynomial approximation to parameterized ECM mapped to experimental cells over a single charging/discharging cycle. Additionally, the chapter delves into the use of Quasi-Monte Carlo sampling techniques to generate large numbers of Chebyshev coefficient vectors to build a space for expressing a battery's health within. This statistical method is employed to systematically investigate and distinguish between three categorically different stages of battery cell aging. These stages are represented as distinct *clusters* or *groups* in the data. Each group corresponds to a specific phase in the battery lifespan, ranging from its fresh state when newly manufactured, through its operational phase where performance might gradually degrade, to its end-of-life stage when the battery's capacity and efficiency significantly diminish. The central point in each cloud uses experimental data in its corresponding aging stage as an achoring point. The term *clouds* is metaphorically used to describe these clusters of data points that represent regions in the simulation space with cells at similar aging stages, providing a visual representation of the spread and overlap of aging characteristics within the dataset. The use of this high-dimensional "Chebyshev space" ensures an optimal robustness and a minimization of divergence in regions of conceivable states of a battery that are not sufficiently sampled experimentally.

In **Chapter 5**, an advanced analysis of battery behavior by leveraging computational simulations and machine learning techniques is introduced to extend the boundaries of traditional EIS analysis. The synthetic input data for the training comes from simulations of battery models that are sampled following the procedures laid out in the previous chapter. The machine learning technique discussed in this chapter involves the use of autoencoders, which are an architecture specifically designed for feature extraction and dimensionality reduction. By alternatively feeding time-based or frequency-based battery data into autoencoders, we can compare their topologies and aging pathways in a lower-dimensional latent space. The similarity of the two resulting latent spaces allows us to abandon traditional impedance measurements and gradually employ raw time series data for further analysis. The time-based battery data is from simulations and is ideally noiseless. Therefore, the chapter addresses the challenge of data scarcity and the robustness of the model by introducing noisy, permuted, and partial signals into the training process. The noisy permuted signals are employed to test model's ability to generalize and explore how little data is re-

quired to still achieve meaningful insights, when obtaining extensive datasets might be challenging or resource-prohibitive.

Chapter 6 describes the systematic generation of stochastic pulse sequences that allows not only for the optimization of excitation schemes specifically designed for battery characterization, but also for stochastic spectroscopic experiments that comply to real-world scenarios during charging or discharging processes where full control of the battery for the measurement process is typically not achievable. The evaluation of how different pulse profiles influence battery diagnostics is conducted in a latent space, using EIS data as a reference. For an effective comparison of two latent spaces, the probability distribution of points within them is analysed by comparing their quantile locations. Furthermore, the segmented latent space resulting from pulse signals, aligned with the division strategy applied to the latent space derived from EIS data, is also visualized. The identified optimal pulse sequence by evaluating how well it aligns with the reference impedance data in terms of distribution and aging-related segmentation within the latent space, could be used for battery lifetime optimization.

In the final **Chapter 7**, the essential contributions of this dissertation are summarized and an outlook is presented to emphasize future research and application perspectives.

Chapter 2

Theory and Background

2.1 Electrochemical Impedance Spectroscopy

Electrochemical impedance spectroscopy is used to characterize the electrical behavior of electrochemical systems, such as batteries. EIS involves measuring the complex impedance as a function of frequency. The analysis of impedance over a broad spectrum of frequencies yields a high density of information about the cell's state within a comparatively short analysis time [8]. Impedance, Z, is a generalization of Ohm's law for Alternating Current (AC) currents $I(t)$, which states that the impedance of a system can be determined by the ratio of the voltage response $U(t)$ to the applied current. Mathematically, impedance $Z(f)$ at a specific frequency f can be calculated from $U(t)$ and $I(t)$ by analyzing the amplitude and phase shift between the voltage and current signals. This relationship allows for the assessment of the cell's impedance across a range of frequencies. The impedance is therefore defined as:

$$Z(\omega) = \frac{|U|e^{j(\omega t+\psi)}}{|I|e^{j\omega t}} = |Z|e^{j\psi} \tag{2.1}$$

where $|U|, |I|$ are, respectively, voltage and current amplitudes of sinusoidal signals, ω is the angular frequency which is equal to $2\pi f$, and ψ is a phase shift. Expanding expression using Euler's relationship $e^{j\theta} = cos\theta + jsin\theta$, in complex notation, the impedance can be represented as [9]:

$$Z(\omega) = |Z|(\cos\psi + j\sin\psi) = \Re(Z) + \Im(Z) \tag{2.2}$$

where $\Re(Z)$ is the real part, $\Im(Z)$ is the imaginary part, and $j = \sqrt{-1}$ is the imaginary unit.

EIS is most effective when the investigated electrochemical system is linear and

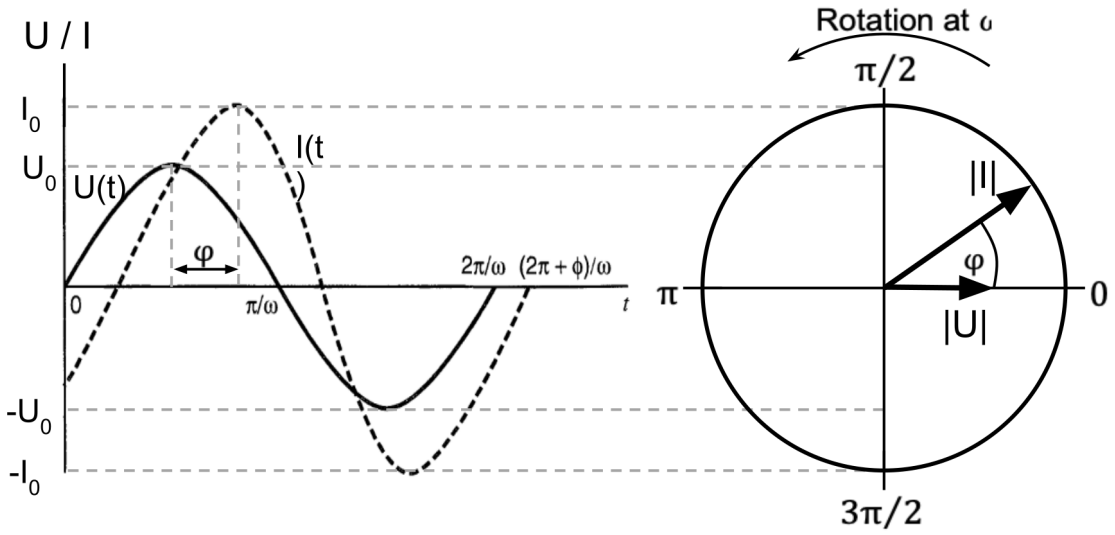


Figure 2.1: The relationship between the voltage and current, and the ratio of them results in impedance

time-independent, as the technique assumes the system's response to an applied sinusoidal signal is proportional and invariant over time. If the system is non-linear or time-dependent, the EIS data may not accurately reflect the system's properties. However, some assumptions of EIS may still be valid for slightly non-linear or quasi-stationary systems if the perturbation signal is sufficiently small, ensuring that the system's response approximates linear behavior within a limited operational range.

Subsequently, the impedance data collected across various frequencies can be plotted on a Nyquist plot or a Bode plot. The Bode plot shows the magnitude $|Z|$ and phase ψ of the impedance as functions of frequency (Equation 2.2). In the Bode plot, the x-axis represents the logarithm of the frequency, and y-axis represents the magnitude of the impedance in decibels (dB) and the phase angle in degrees. Thereby, one can easily observe trends, identify characteristic frequencies, and distinguish between resistive and capacitive behavior [10]. In a Nyquist plot, such as the one shown in Figure 2.2, the real part $Re(Z)$ is usually plotted against the imaginary part $Im(Z)$ for visualizing the system's response to different frequencies, each point on the Nyquist plot corresponds to a specific frequency. The plot often takes the shape of a semicircle or an arc, and its characteristics provide insights into the electrochemical processes occurring within the system [11].

2.2 Equivalent Circuit Modeling

One of the common ways to represent the behaviour of a battery cell is with a simplified electrical circuit that mimics its impedance response, using the ECM technique [12]. The EC model uses a circuit, where the elements are arranged in series and parallel configurations. The choice of circuit elements depends on the electrochemical processes occurring in the system.

Serial Resistor

A serial resistor R_s is a fundamental component often incorporated to represent real or ohmic resistance within a system. According to Ohm's law, theoretically the voltage U_0 at the ohmic resistance is directly proportional to the battery current I_0,

$$Z(\omega) = R_s = \frac{U_0}{I_0}\,. \tag{2.3}$$

Ohmic resistance accounts for voltage decline and power loss as heat. In EIS spectra, the value of the serial resistor is observed through the intersection of the Nyquist curve with the real axis in the high frequency range [13].

ZARC

The ZARC element, which consist of a Constant Phase Element (CPE) and parallel connected resistor, is a common component to describe the charge transfer and the double layer capacitance at an electrode–electrolyte interface. The CPE is an capacitive element to represent the frequency-independent behavior. Mathematically, the impedance of a ZARC element is represented as [14]:

$$Z(\omega) = \frac{R}{1 + (j\omega\tau)^{\phi}} \tag{2.4}$$

where τ is the mean time constant, ϕ is the depression factor, and R is the charge transfer resistance of ZARC element. If $\phi = 1$, the ZARC behaves as an parallel RC circuit element and if $\phi = 0$, it behaves as a ideal resistor. The ZARC element typically manifests as a depressed semicircle in the EIS spectrum [13]. The determination of the number of ZARC elements used in the circuit, along with their characteristic frequencies and parameters, is carried out through techniques such as the DRT [15], which will be discussed in the following section.

Warburg

Diffusion processes are generally represented by Warburg elements and there are two types: the Finite Length Warburg (FLW) and the Finite Space Warburg (FSW), sometimes called the *short* and *open* Warburg elements respectively [16]. FLW considers a specific length or distance over which the Warburg impedance operates. In the EIS spectrum, the shape of the FLW looks like a line transitioning into a semi-circle. It can be thought of as a Warburg being *shorted* by a resistor. FSW implies a limitation on the space through which ions or charge carriers can diffuse. In the Nyquist plot, the classic FSW appears as a straight line with a 45-degree slope at low frequency. However, there is another type of Warburg impedance observed under conditions of finite lengths or specific boundary constraints. In this scenario, the impedance transitions to a predominantly capacitive behavior at very low frequencies, leading to an alignment that is nearly vertical in the Nyquist plot. In a coin cell, it is used to model slow diffusion process of ions within an electrolyte, and its mathematical expression with approximation is [17]:

$$Z(\omega) = \frac{A_W}{\sqrt{\omega}}(1-j)\coth(C_{\text{lim}}A_W\sqrt{2j\omega}) \tag{2.5}$$

where A_W is the Warburg coefficient and C_{lim} is the limit of capacitance.

Open Circuit Voltage

Open Circuit Voltage (OCV) is the potential difference across a cell when there is no current flow. In an ECM, it is often represented as a voltage source in series with other circuit elements. OCV is a crucial parameter for understanding the SOC of a battery and serves as a reference point for the voltage characteristics of the cell [18].

Graphical Representation

A sample EIS spectrum of a commercial coin cell, which is measured from 10 mHz to 300 kHz at SOC 100%, is shown in Figure 2.2. The intersection with the real axis, semi-circles and straight line can be represented each by serial resistor for ohmic behaviour, ZARC elements for charge transfer and double layer capacitance, and Warburg element for diffusion process.

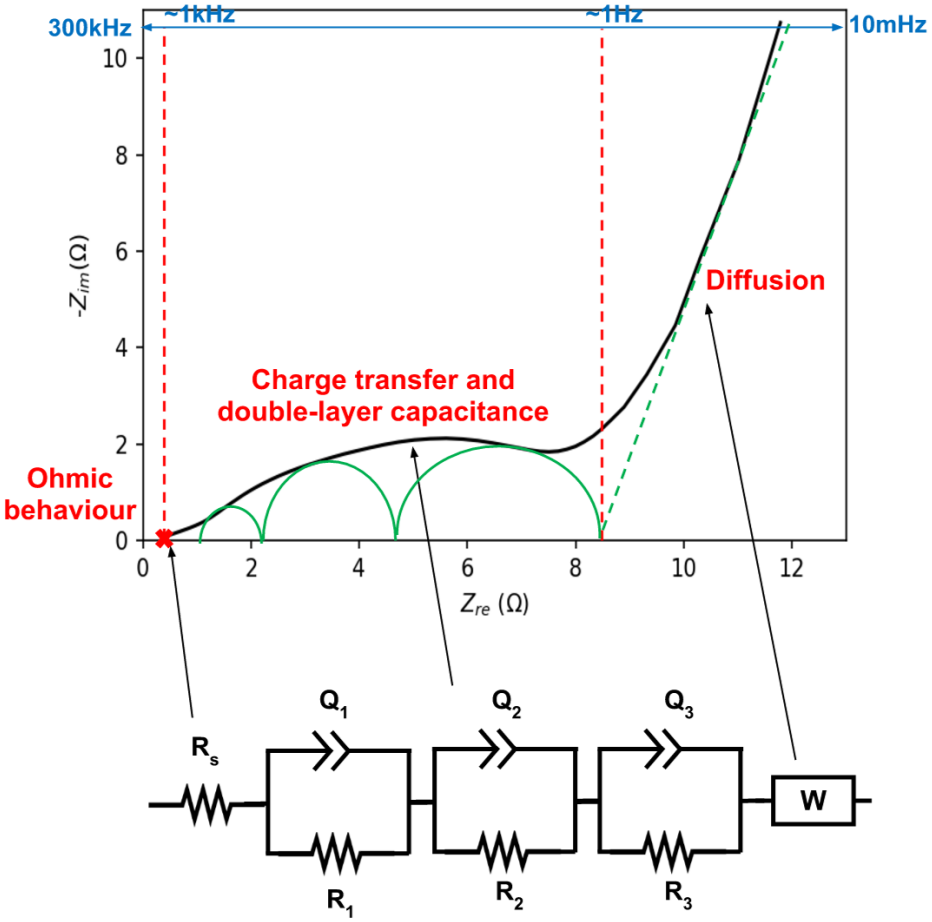


Figure 2.2: ECM based on the EIS response of a lithium-ion battery at 37 mAh capacity and 100% SOC.

2.3 Distribution of Relaxation Times

Identifying the number of ZARC elements in a semicircle of an EIS spectrum without overfitting and determining their physicochemically reasonable parameter values is challenging for pure impedance fitting, because it is highly dependent on initial guesses and boundaries. Therefore, impedance should be pre-processed using the DRT method.

The DRT represents a technique to simplify EIS data $Z(\omega)$ by deconvolution with an Resistor—Capacitor (RC) kernel [19]:

$$Z(\omega) = R_0 + R_p \int_0^\infty \frac{\gamma(\tau)}{1 + j\omega\tau} d\tau \tag{2.6}$$

Here, R_0 is the ohmic resistance, R_p the polarization resistance, and $\tau = RC$ is the time constant of a parallel RC circuit element. $\gamma(\tau)$ with $\int_0^\infty \gamma(\tau)d\tau = 1$ is the distribution function of the RC circuits used to fit the impedance data.

DRT transforms impedance data from frequency domain to a time constant domain,

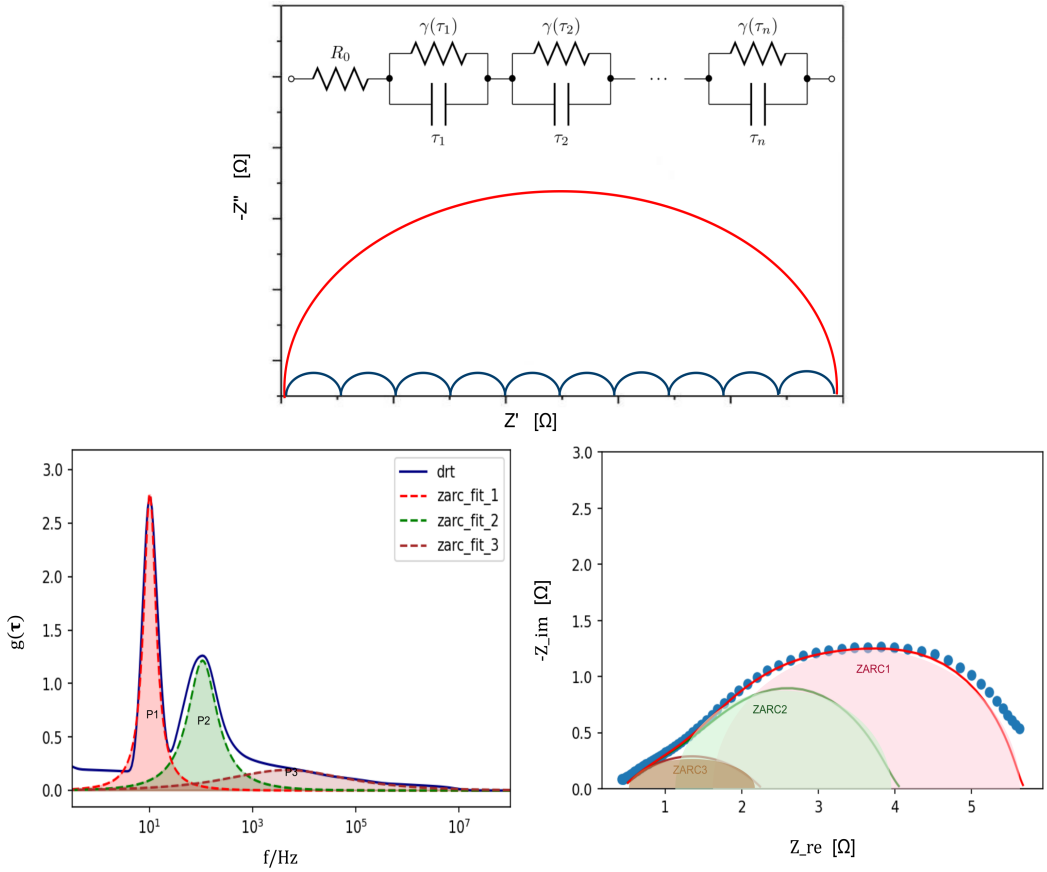


Figure 2.3: DRT transformation of EIS data. (Upper) Principle of deconvolution with RC kernel, (Left) DRT transformation in the time constant domain, (Right) Separation of semi-circle corresponding to three peaks in the EIS spectrum

which represents a distribution of parallel RC circuits by assuming the system to be passive and non-oscillating. Thereby, in the relaxation time domain, inner processes become visible as peak at their respective time scale (Figure 2.3). These peaks give important hints about the number of distinguishable features and their distribution within the sensitivity limit of the EIS data, without *a priori* knowledge of the response of the investigated system. Especially, Gaussian Process Regression (GPR) based DRT method estimate weights based on single dataset, which facilitates the use of DRT as an automatic EIS data pre-processing procedure for ECM parametrization with no user intervention [15]. However, DRT is unstable for low frequency domain features of Warburg elements, which often cannot be meaningfully deconvoluted using a kernel of an RC element [20].

2.4 Battery State

As a battery undergoes charging and discharging cycles, its electrochemical reactions, interfaces, and structural components can experience degradation and changes in performance. For example, the charge transfer resistance and diffusion elements may increase as the battery ages or undergoes cycling, leading to higher internal resistance and reduced battery performance [21]. Here two crucial parameters, the state of charge and the SOH, are indicative of the battery's dynamic state.

2.4.1 State of Charge

SOC refers to the current level of charge stored in a battery relative to its maximum capacity. It provides information about the available energy in the battery and is crucial for estimating the remaining runtime or capacity of the battery. SOC determination can be performed using various methods, including coulomb counting and voltage-based estimation [22].

Coulomb counting estimates SOC by integrating the current flowing in and out of the battery over time. It involves measuring the battery's current and keeping track of the cumulative charge transferred. However, accurate coulomb counting requires precise measurement of current and accounting for efficiency losses [23]. Voltage-based estimation relies on the voltage response of the battery during charging and discharging. A battery's OCV can provide a rough estimation of SOC.

2.4.2 State of Health

SOH represents the overall condition and degradation of a battery relative to its initial capacity or performance. It provides information about the battery's aging, capacity loss, and internal resistance increase over time. SOH determination can vary depending on the specific battery chemistry and available data.

Capacity-based estimation compares the current capacity of the battery to its original capacity when it was new. By measuring the available capacity and comparing it to the expected capacity, SOH can be estimated. This can be done through periodic capacity testing or through algorithms that track capacity degradation over time [24]. However, capacity fade alone may not provide a comprehensive picture of battery health, as other degradation mechanisms may affect performance, and capacity testing requires dedicated equipment, which is not practical for continuous monitoring. Besides that, the internal resistance of a battery increases as it degrades [25]. By measuring the internal resistance through EIS, the changes in resistance over time

can indicate the battery's SOH. EIS measurements over battery full lifetime are time-consuming techniques and interpretation of multiple impedance spectra can be complex and requires expertise. Alternatively, pure data-driven ML models are employed as a promising alternative to predict and estimate the battery's degradation. They utilize various health indicators and large amount of sensor data to effectively capture the non-linear complex relationship among battery parameters and degradation levels [26, 27].

Unlike SOC determination, prediction and estimation of battery age is an ongoing field of research, and accurate and precise determination of SOH can be challenging due to the complexity of battery aging processes and various influencing factors. Besides, measuring different types of batteries at their individual state takes much time.

2.5 Chebyshev Polynomials

Chebyshev polynomials are a family of orthogonal polynomials that arise in various areas of mathematics and physics. There are two kinds of Chebyshev polynomials, known as Chebyshev polynomials of the first kind T_n and the second kind U_n. These polynomials are defined over the interval $[-1, 1]$ and are expressed recursively as [28]:

$$\begin{array}{cc} T_n(cos\theta) = cos(n\theta), & U_n(cos\theta)sin\theta = sin((n+1)\theta) \\ T_0(x) = 1, & U_0(x) = 1, \\ T_1(x) = x, & U_1(x) = 2x, \\ T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), & U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x) \end{array} \tag{2.7}$$

T_n and U_n are the nth degree polynomials, and T_n is orthonormal with respect to the weight $\frac{1}{\sqrt{1-x^2}}$, while U_n is to the weight $\sqrt{1-x^2}$.

Orthogonality in Chebyshev polynomials allows for effective smoothing of noisy data. When fitting a function to data points, the orthogonal property ensures that the influence of each polynomial component on the others is minimized. This leads to a more stable and reliable fit. Besides, orthogonal polynomials can represent complex functions with fewer terms compared to non-orthogonal polynomials. The min-max property of Chebyshev polynomials ensures that the maximum deviation between the polynomial and the actual function is minimized over the interval, achieving a more uniform error distribution compared to other polynomial fits. For instance, in scenarios with unequal sampling density, where some regions of the interval are densely sampled while others are sparse, traditional polynomial fits might exhibit small errors in well-sampled areas and large errors in poorly sampled ones. In contrast, Chebyshev polynomials mitigate this issue by distributing the error more uniformly across the

entire range, regardless of the sampling density variations.

These remarkable orthogonality and min-max properties of the Chebyshev polynomials, particularly T_n, offer a robust framework for exploration of non-linear relationships between multiple battery parameters and SOC without succumbing to overfitting.

2.6 Quasi Monte Carlo

Quasi Monte Carlo methods represent a numerical technique to evaluate multi-dimensional integrals. While traditional Monte Carlo methods use random sampling to approximate integrals, QMC uses low-discrepancy sequences, often referred to as quasi-random sequences [29]. These sequences are designed to fill the integration space more uniformly than uncorrelated random points, leading to potentially faster convergence rates.

The most widely used low-discrepancy sequence is Sobol sequence, which is a digitally constructed sequence that uses a base-2 numeral system [30]. Sobol sequences are particularly effective in high dimensional spaces due to their construction involving bitwise exclusive OR (XOR) operations, which ensure an even distribution across dimensions.

Mathematically, it generates $n = 2^m$ points in unit d-dimensional hypercube $[0, 1)^d$ and m refers to the number of bits of the generator. For each dimension i of the hypercube, the n-th element of the sequence is generated by [31]:

$$x_{n,i} = \oplus_{j=1}^{m}(n_j, v_{i,j}) \tag{2.8}$$

where $\oplus$ denotes the bitwise XOR operation, n_j are the binary digits of n, and $v_{i,j}$ is the direction numbers for the i-th dimension, which are derived from primitive polynomials over the field of two elements ($\mathbb{F}_2$) [32].

Their properties of deterministic and uniform distribution of points in the unit hypercube in higher dimensions helps in achieving better convergence rates in numerical integration. To use these points for analyses based on the Gaussian distribution, they need to be adjusted from their original unit hypercube-shaped spread to align with the Gaussian distribution's specific range and shape. An example of QMC sampling is shown in Figure 2.4.

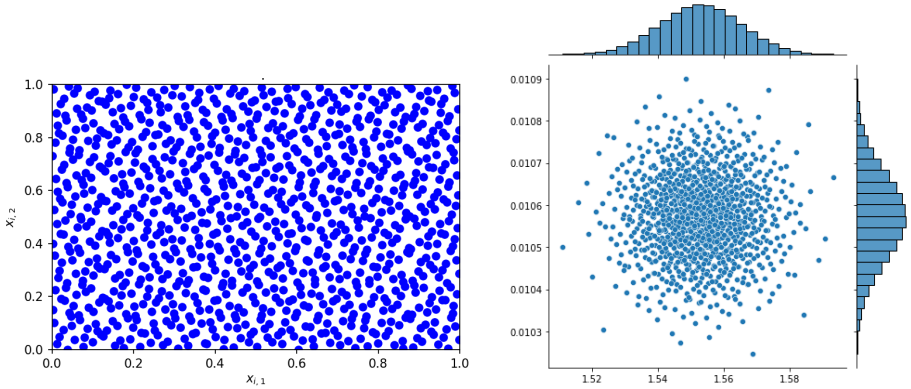


Figure 2.4: QMC sampling in the two-dimensional space using (left) Sobol sequences with 10 Bits for generating 2^{10} points and (right) transform into Gaussian distribution.

2.7 ISEA Framework

The real time capable ISEA Framework written in C++, which is developed at the Institute for Power Electronics and Electrical Drives (ISEA), allows us to simulate EC-models with load profiles and gather the voltage response. As the framework is real-time capable it can be implemented on a Hardware in the Loop (HIL) to estimate crucial states of storage system packs used in electric vehicles [17]. For our research work, we focused on the framework's ability to allow electrical simulations of the battery storage system consisting of arbitrary but reasonable serial connections between its electrical components.

Figure 2.5 provides an overview of the framework's operations. The determined ECM parameters and OCV at various SOCs are stored as Lookup table and an initial SOC and temperature of the battery system are given. For each simulation, the current through the electrical model is given as a function of time. At each time step, it calculates voltage response based on differential and algebraic equations and updates the battery's state, such as SOC or temperature, as well as corresponding ECM parameter values by evaluating the respective lookup tables [33]. These steps are repeated for the specified duration or until the current profile is fully simulated. The calculated voltage at every time step builds a voltage profile that corresponds to the current input. Additionally, a thermal model can be simulated by coupling with the electrical model, so that the thermal heat is directly fed into the thermal simulation. However, currently we only use the electrical simulation, keeping the temperature constant.

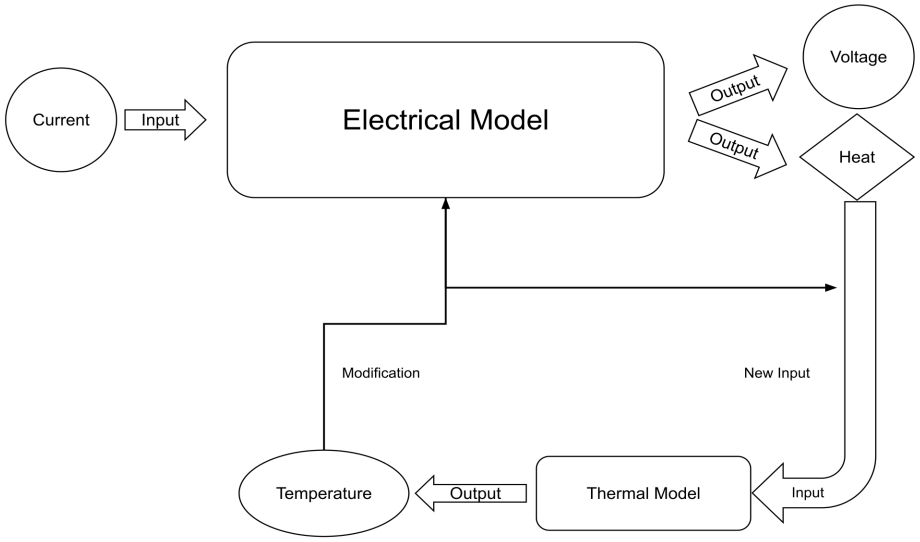


Figure 2.5: Principle of ISEA simulation framework

2.8 Machine Learning

2.8.1 Autoencoder

An autoencoder is a type of artificial neural network used in unsupervised learning to perform dimensionality reduction and feature extraction from input data [34]. It consists of an encoder and a decoder.

The encoder maps an input vector $\mathbf{x} \in \mathbb{R}^n$, which could also be the vectorization of a higher-dimensional input matrix, to a lower-dimensional latent representation $\mathbf{y} \in \mathbb{R}^m$, named Latent Space (LS). This process applies similarly to the output vector $\mathbf{y}$, indicating that both input and output can originate from or be utilized in the form of higher-dimensional structures, efficiently compressed into vector forms for processing. The input vector undergoes a linear transformation in the form of matrix multiplication with learned weights specific to the encoder's layers. The result of the linear transformation is passed through a non-linear activation function, allowing the network to learn complex patterns and relationships. The transformation can be represented as [35]:

$$\mathbf{y} = \mathrm{ReLU}(\mathbf{W} \cdot \mathbf{x} + \mathbf{b}) \tag{2.9}$$

where $\mathbf{W}$ is the weight matrix, $\mathbf{b}$ is the bias vector, and *ReLU* is Rectified Linear Unit (ReLU) activation function, which keeps positive values and sets negative input values equal to zero. The final layer $\mathbf{y}$ represents the compressed, lower-dimensional latent space representation of the original input vector. The decoder then aims to

reconstruct data $\mathbf{z} \in \mathbb{R}^n$ from the latent space representation. The objective of training the autoencoder is to minimize a reconstruction loss between the input $\mathbf{x}$ and its reconstruction $\mathbf{z}$, typically using a loss function, such as Mean Squared Error (MSE): $L(\mathbf{x}, \mathbf{z}) = ||\mathbf{x} - \mathbf{z}||^2$ [36].

Utilizing convolutional layers in both encoding and decoding processes is called Convolutional Autoencoder (CAE), which is designed for handling grid-like data, such as images. CAE extends the basic autoencoder architecture by convolution operation with a set of learnable filters and following max pooling for reduction of spatial dimensions [37]. Therefore, the encoding process can be represented as:

$$\mathbf{y} = \mathrm{ReLU}(\mathbf{W} \cdot (\mathrm{P}(\mathrm{C}(\mathbf{X}))) + \mathbf{b}) \tag{2.10}$$

where C represents convolution, P represents pooling.

Suppose input matrix $\mathbf{X}$ has a size of $L \times H$, and apply a filter $\mathbf{F}$, which is a smaller matrix of size $l \times h$, then the convolution operation is defined as [38]

$$(\mathbf{X} * \mathbf{F})(i, j) = \sum_{u=0}^{l-1} \sum_{v=0}^{h-1} \mathbf{X}(i + u, j + v) \cdot \mathbf{F}(u, v) \tag{2.11}$$

where (i, j) denotes the position on the input image where the filter is applied, and the dot product is summed over the filter dimensions. The stride defines how many pixels the filter moves across the image. A stride of 1 moves the filter one pixel at a time. Padding involves adding extra pixels around the edge of the input image. This is often done to control the size of the output feature map. In Max-pooling, the maximum element is selected from the region of the feature map covered by the filter. Pooling helps the Convolutional Neural Network (CNN) to achieve spatial invariance, allowing the network to detect features regardless of their spatial location in the input [39]. For example, max-pooling with a 2×2 filter and stride 2, the operation is defined as

$$P(i, j) = \mathbf{max}(\mathbf{X}[2i : 2i + 2, 2j : 2j + 2]) \tag{2.12}$$

where $P(i, j)$ represents the element of the pooled output at position (i, j), and the max operation is taken over the 2×2 region of $\mathbf{X}$.

Similarly, decoder mirrors the architecture of the encoder but in reverse. Upsampling layers increase the spatial dimensions of the latent representation. They can be implemented as deconvolutions or convolutions followed by upsampling. Convolutional layers are then used to refine the upsampled features and reconstruct the original input. The aim of training a CAE is to reduce the loss function, similar to that of a basic autoencoder. The basic structure of CAE is shown as Figure 2.6.

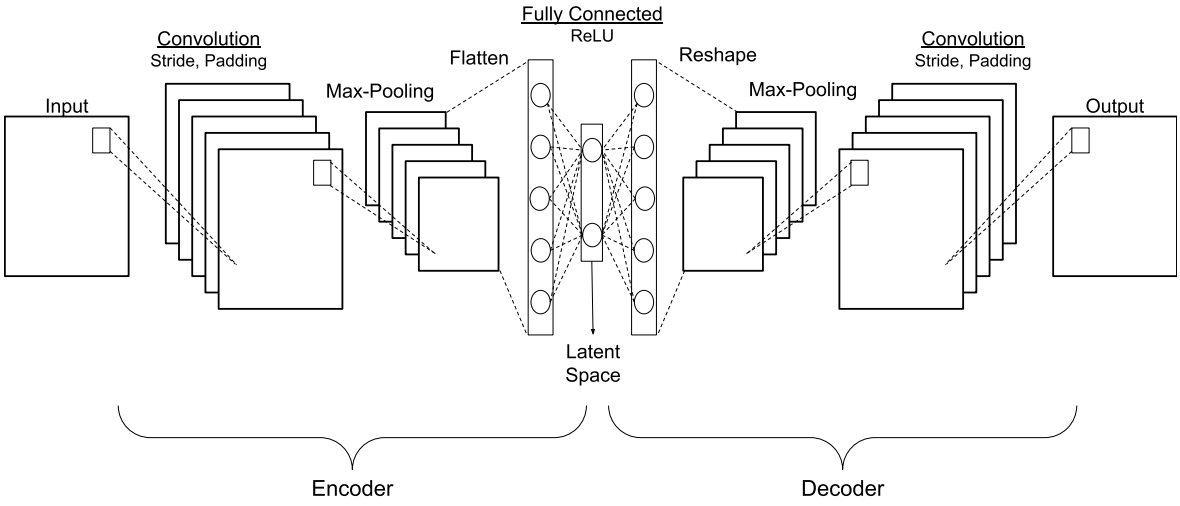


Figure 2.6: Structure of a convolutional autoencoder

CAEs leverage the power of convolutional neural networks, making them adept at capturing spatial hierarchies in data. Convolutional layers share weights, making CAEs more parameter-efficient compared to fully connected layers in basic autoencoders [40]. Both convolution and pooling operations reduce the dimensionality of the input, leading to a more compact representation and reducing the likelihood of over-fitting. Autoencoders for dimensionality reduction is similar to Principal Component Analysis (PCA) but more powerful due to their non-linear capabilities [41]. Additionally, they can learn useful features automatically from the input data. When transforming battery data into the latent space using an autoencoder, the data is typically more amenable to clustering.

However, the individual values of vectors in the latent space might not have an explicit or interpretable meaning on their own. Instead, the crucial information lies in the relative positions and relationships between these vectors, especially in the context of neighboring vectors. The structure and distribution of vectors in the latent space, as well as their proximity or separation, play a vital role in how the model captures and represents patterns and information from the data.

2.8.2 Support Vector Machine

Support vector machine is a supervised machine learning algorithm for classification in a lower-dimensional representation. SVM aims to find an optimal hyperplane that maximizes the margin between different clusters while minimizing classification error

[42]. The convex optimization problem in SVM guaranteed the global optimum, which is not sensitive to the initial guess or starting point.

As a non-parametric model, it makes fewer explicit assumptions about the underlying data distribution or the number of features compared to other models. In other words, it does not assume a specific functional form for the data, nor does it assume particular cluster shapes or densities. Besides, SVMs are effective in handling both linearly and non-linearly separable data. Through the use of appropriate kernel functions, they can learn complex decision boundaries without assuming linearity. These kernel choice allows implicitly mapping of data into a higher-dimensional space, making it easier to find a separating hyperplane. SVM prioritizes a subset of data points, known as support vectors that lie closest to the decision boundary and influence the position and orientation of the hyperplane, therefore it allows a robust decision-making even with a small dataset. Additionally, SVM is less sensitive to outliers due to the margin-maximizing objective. Outliers have minimal impact on the position of the hyperplane, and they represent unusual or faulty behavior in the context of battery clustering for anomaly detection.

In a two-dimensional latent space, an SVM using a linear kernel, the dataset should be $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ with feature x_i and the class label y_i, and their decision boundary is:

$$f(x) = \mathbf{w}^T \cdot \mathbf{x} - b \tag{2.13}$$

where $\mathbf{w}$ is the weight vector perpendicular to the hyperplane, $\mathbf{x}$ is the feature vector, and b is the bias term. The predicted class for a given feature vector $\mathbf{x}$ is determined by the sign of $f(x)$. In binary classification for example, where the class label y_i has values +1 or -1, the objective is to find $\mathbf{w}$ and b that maximize the margin $\frac{2}{\|\mathbf{w}\|}$ [43] while ensuring correct classification $y_i(\mathbf{w}^T \cdot x_i - b) \geq 1$ for all $i = 1, 2, ...n$. That means, the optimization goal is to minimize[44]:

$$\lambda\|\mathbf{w}\|^2 + [\frac{1}{n}\sum_{i=1}^{n}\mathbf{max}(0, 1 - y_i(\mathbf{w}^T x_i - b))] \tag{2.14}$$

where the parameter $\lambda > 0$ determines the trade-off between increasing the margin size and ensuring that the x_i lie on the correct side of the margin. Here a regularization parameter C is introduced, which is inversely related to λ, thus the optimization problem can be messaged into [42]:

$$\underset{\mathbf{w},b,\zeta}{\text{minimize}}(\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\zeta_i) \tag{2.15}$$

subject to

$$y_i(\mathbf{w}^T x_i - b) \geq 1 - \zeta_i,\ \zeta_i \geq 0, \forall \in_i 1, 2, ...n \tag{2.16}$$

where $\zeta_i = \mathbf{max}(0, 1 - y_i(\mathbf{w}^T x_i - b))$. The hyperparameter C helps to control the error of misclassifying. The high values of C means higher penalty and better accuracy, the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly [45]. But it also leads to overfitting because of a low bias and high variance. Conversely, a low value of C will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies some points. It results in underfitting, where the model has high bias but low variance, indicating it is not complex enough to capture the underlying pattern of the data effectively. Tuning C correctly is a vital step in the use of SVM with linear kernel. This is usually achieved through cross-validation, where different values of C are tested, and the one that results in the best performance on a validation set is chosen.

Besides linear classification, SVM can handle non-linear relationships in the data through the kernel function, like polynomial, Radial Basis Function (RBF) and so on [46]. One of the most widely used non-linear boundary function is RBF, and it can be mathematically expressed as

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \tag{2.17}$$

where σ is the variance and $\|x_1 - x_2\|$ is distance between two points x_1 and x_2. Finding right σ for a given dataset is critical and a new hyperparamter γ is introduced to control the distance of the impact of a single training point [47]. γ is inversely proportional to σ, and a high value of γ leads to overfitting. In SVM, the γ parameter determines the scope of influence for a single training example, affecting how closely the model fits to the training data. A high γ value means that a data point must be very close to a support vector to be considered in the same category. Therefore, with a high γ, the influence of each support vector is limited to points that are very close to it, which can result in a more complex decision boundary that fits the training data very closely, and potentially leading to overfitting where the model may not generalize well to new, unseen data. Instead, a low γ value means a broad similarity radius which results in more points being clustered together and cannot capture the complexity or *shape* of the data.

For evaluating the performance and generalization capabilities of the SVM, the training scores and cross-validation scores are essential metrics. The simple validation

process used in machine learning is setting another test set for the trained model to evaluate its accuracy. The training score is defined as the ratio of correctly predicted instances to the total number of instances in the dataset. However, by splitting the data into multiple subsets, the learning results can depend on a particular random choice for the pair of (train, test) sets. The average accuracy is often reported as the cross-validation score. This helps prevent overfitting and increases the generalization ability of the model.

2.9 Statistical analysis for latent space

2.9.1 Delaunay Triangulation

Delaunay triangulation is a mathematical technique used for partitioning a space, particularly when applied to the division of a latent space. The triangulation provides a straightforward visual representation of complex data, making it easier to interpret and understand [48].

The algorithm constructs triangles connecting points in the latent space in such a way that no point lies inside the circumcircle of any triangle [49]. This creates a mesh or network of triangles covering the entire set of points. The triangulation maximizes the minimum angle of all the angles of the triangles, avoiding skinny triangles as much as possible. It provides an efficient way to divide the space, capturing the natural *neighborhood* relationships between points. Areas with a high concentration of points will have smaller triangles, while sparse areas will be represented by larger triangles. By adjusting the criteria for forming triangles, such as limiting the size or area of triangles, it is possible to control the number of groups. Unlike predetermined or rigid clustering algorithms, Delaunay triangulation allows the data itself to dictate the group boundaries. The visual layout of triangles can help in identifying hidden patterns, relationships, and even anomalies within the data.

2.9.2 Quantile Quantile plots

Quantiles are a fundamental concept in statistics, used to compare and analyze distributions. They are particularly important in the context of Q-Q plots, which are graphical tools for assessing whether two data sets come from populations with a common distribution [50].

Quantiles are cut points dividing a dataset into equal-sized, contiguous intervals, each containing a specific proportion of the data. Specific names are given to common quantiles based on how they divide the data: quartiles split the data into four equal

parts, deciles into ten, and percentiles into a hundred distinct groups [51].

Mathematically, a quantile Q for a given proportion p in a dataset X is defined such that a certain percentage p of data in X is less than or equal to Q. For a sorted dataset $X = \{x_1, x_2, ..., x_n\}$ with a finite sample of size n, $x_1 \leq x_2 \leq ... \leq x_n$, the p-quantile is located in the index k of sorted dataset [52]:

$$k = p \cdot (n + 1) \tag{2.18}$$

If k is an integer, the p-th quantile is the value at the k-th position in the sorted dataset. If k is not an integer, interpolate between the closest ranks to find the quantile value. Therefore, the median is the 0.5 quantile, dividing the dataset into two equal halves. For probability distribution with a mean and variance, it is the case that [53]:

$$\mu - \sigma\sqrt{\frac{1-p}{p}} \leq Q_p \leq \mu + \sigma\sqrt{\frac{p}{1-p}} \tag{2.19}$$

where Q_p is the value of the p-quantile for $0 < p < 1$, μ is the mean, and σ is the standard deviation.

To effectively visualize how closely two latent spaces align in terms of their distributional properties, one can construct quantile against quantile plots, where each point (x, y) corresponds to a specific quantile [54]. The y-coordinate of a point represents a quantile from the second distribution, while the x-coordinate corresponds to the same quantile from the first distribution. This arrangement forms a parametric curve on the plot, where the parameter is the index of the quantile interval.

If the distributions are similar, the points on the Q–Q plot will generally fall close to the identity line $y = x$. However, if the distributions have a linear relationship but are not identical, the points will still align along a line, but this may not coincide with the identity line. The similarities or disparities in key distributional properties, such as location, scale, and skewness between the two distributions, can be illustrated. The slope of the data points relative to the identity line offers important insights. If the overall slope of the plot is shallower than the $y = x$ line, it indicates that the distribution represented on the horizontal axis is more spread out compared to the one on the vertical axis. Conversely, a slope that is steeper than the $y = x$ line suggests that the distribution on the vertical axis exhibits greater dispersion than the one on the horizontal axis [55]. Besides, if Q-Q plots display an arced or S-shaped pattern, it signifies that one distribution exhibits greater skewness compared to the other, or that one of the distributions possesses heavier tails than its counterpart. Additionally, the linear regression line drawn through the quantiles can provide valuable information about the relative position and scale of the data samples being compared. The intercept

and slope of this regression line are indicative of these aspects. When the median of the distribution plotted along the horizontal axis is zero, the regression line's intercept offers insight into the relative location of the data, while its slope reflects the relative scale of the distributions.

Chapter 3

ECM Parametrization based on experimental EIS data

This chapter is adapted from the publication that has been submitted to the journal and is now under review: **Weighted distribution of relaxation time analysis of battery impedance spectra using Gaussian process regression for noise estimation**, by C. Bartsch, L. Jin, F. Bereck, A. Mertens, R.-A. Eichel, C. Scheurer, and J. Granwehr. It has been also published in *ChemRxiv*, DOI: 10.26434/chemrxiv-2024-1gxgq (2024).

The GPR-based DRT algorithms and noise characterization for estimation of initial conditions of ECM parameters was developed by Christian Bartsch and Andreas Mertens. The ECM fitting and parametrization at various SOCs were conducted by the author of this thesis, Limei Jin. The experimental setup and measured EIS data of coin cell was provided by Franz Bereck.

3.1 Introduction

With the growing demand for electric vehicles and an increase in decentralized energy supplies, the need for reliable and long-lasting batteries is ever increasing [56]. To reduce the degradation of individual battery cells and to ensure a longer lifespan, advanced characterization methods, which can be applied *in situ* or *in operando*, in combination with adapted analysis and modeling techniques to retrieve a maximum of information, need to be developed [57].

EIS has proven to be a versatile non-invasive method for characterizing lithium-ion batteries [58, 8]. In its most common implementation, a frequency-dependent impedance is determined from an oscillating voltage perturbation and the measured current response of the system.

For an in-depth understanding and prediction of battery characteristics, different modelling strategies are available. The most general and widely used models are electrochemical models and empirical models. Additionally, stochastic models and analytical models are widespread [59]. Electrochemical models are based on the chemical kinetics and reaction dynamics of processes in the battery. These models have a high accuracy, but exhibit disadvantages in the complexity of equations and correspondingly long simulation times. In contrast, empirical models such as ECM are computationally efficient, but deliver limited predictive power regarding ageing and a high stiffness of the model [60]. Among the other two model types, stochastic models are less descriptive but more intuitive, and analytical models use heuristic techniques to model certain specific characteristics of batteries, such as capacity or runtime [59].

Equivalent circuit modelling is the most common technique to analyze EIS data [61]. Thereby, the battery is modelled as an electrical circuit consisting of different classical and non-classical passive circuit elements. Equivalent circuit elements may be related to particular processes in a battery. Resistors and capacitors are the main conventional elements used to characterize electrochemical systems. Resistors are suitable for characterizing not only electronic, but also ionic conductivites, while capacitors describe charge retention or double layers. Diffusive processes behave differently, with a response typically at low frequencies down to µHz [62]. These processes can often be modelled by so-called Warburg elements [63]. A fundamental limitation of ECM fitting is the non-uniqueness of the model to reproduce a certain impedance spectrum, even if data is available at arbitrary sensitivity [64].

ECM parameters need to be adjusted as a function of the battery state for an accurate control and lifetime monitoring [65]. The most important state information is the SOC. To simulate virtual batteries over full cycles, look-up tables of ECM parameters at different SOC are commonly employed [66]. Parametrization relies on fitting of experimental data, which itself relies on the selection of a suitable model and robust data analysis protocols.

In general, two different ECM fitting strategies can be followed. An effective model can be devised, where the dominant processes for the impedance response are macroscopically represented by suitable equivalent circuit elements. This approach facilitates direct fitting of the data, and the obtained ECM may be used as the basis for a digital twin of a battery. Residuals are reduced by increasing the number of circuit elements, but this does not necessarily improve the physical meaning of the fit and may lead to over-fitting. Resistor–capacitor pairs, *i.e.* a resistor in parallel with a capacitor, represent a basic circuit element for such an ECM. To account for distributions of parameters while keeping the number of elements low, the empirical ZARC circuit ele-

ment, where the capacitor in an RC element is replaced by a constant phase element, has been proposed [67].

Nonetheless, such an empirical ECM still contains only a subset of the parameters necessary for obtaining accurate physical properties of the different materials and interfaces. A different approach is followed in physics-based modelling, where materials and devices are accurately modelled based on physical processes, and an electrical circuit is devised based on impedance response of the different processes [68]. In this case, a direct relation between model and physical properties is possible, but the model contains too many parameters for direct data fitting. Therefore, additional assumptions or constraints as well as parameters determined using other analytical techniques or numerical simulations are necessary for a complete parametrization.

Alternatively, a more data-driven approach can be followed for an improved ECM parametrization. It has been shown that the electrochemical circuit elements commonly used to describe batteries in ECMs can be modelled accurately as distributions of RC elements [69]. Thereby the impedance response of an RC element is used as a kernel for the deconvolution of EIS spectra to obtain a DRT. Commonly, DRT is used as a substitution of ECM. The supplemental use of DRT and ECM is rather uncommon despite the evident complementarity of the two techniques [70].

The features of a DRT spectrum are considerably narrower than the discernible features in EIS raw data. By transforming the EIS spectrum of a battery to the relaxation time domain, inner processes become visible as peaks at their respective time constant τ or, equivalently, frequency $f = (2\pi\tau)^{-1}$. While these peaks, or modes, cannot always be clearly separated, they still give important hints about the number of contributing processes, their time scales, and their relative impact on the impedance of the battery. Shifts of these peaks can indicate changes of the internal battery chemistry.

The residuals of a correctly parametrized DRT fit should not contain any systematic, non-random features [71]. Furthermore, no assumptions are necessary regarding the number of underlying processes and their distribution. A DRT analysis provides the number of distinguishable processes and their distribution within the sensitivity limit of the data. If the signal-to-noise ratio is not sufficient to distinguish two processes, they are represented as one broadened mode whose width represents the uncertainty of the data. One of the main challenges with this technique is the heteroscedasticity (non-uniformity) of noise in techniques such as EIS, where data is recorded over a frequency range spanning several orders of magnitude [72, 73, 74, 75]. To extract the maximum amount of available information from an EIS measurement by DRT, the frequency dependence of the noise needs to be considered [76, 70, 77]. Past approaches to DRT often used uniform data weighting [19], which causes local over- or under-regularization,

potentially even both in the same data set. Thereby, through over-regularization, information contained in the data is partly suppressed, producing an overly smooth DRT curve and non-random, systematic residuals, while under-regularization produces a DRT curve with physically implausible features such as narrow spikes or oscillations that often represent an overfitting of noise. A more coarse-grained alternative is to conduct a piece-wise DRT calculation, where regions of similar noise values are individually inverted [71]. This approach is, however, rather laborious, and a robust implementation that does not require user intervention is not straightforward.

Batteries show non-linear dynamic behavior that depends on different internal and external conditions such as temperature, SOC, SOH, calendaric age, or runtime [78, 7]. Hence the characterization and accurate modeling of batteries is quite demanding. More drastically, a non-linear response may even challenge the suitability of spectroscopic characterization techniques altogether, since causality may be violated [79]. To evade this argument, the common assumption is made that a system is quasi-causal on the time scale of the experiment or the amplitude of the perturbation. However, when experiments are repeated, the validity of this assumption must be assessed for the comparability of subsequent experiments, in particular if they are used to characterize noise processes [80].

Here, we present DRT inversion of EIS data of commercial Li-ion button cells using uniform penalty regularization with weighted data [81]. To obtain a noise estimation that is necessary for data weighting, a statistical data analysis of multiple EIS measurement repetitions at identical SOC is compared with an estimate of measurement uncertainty based on GPR using a single data set. Residuals of these fits are assessed for systematic features, and the validity of the assumption of quasi-causality at a particular SOC is critically tested. For data interpretation, DRT spectra are then used for the parametrization of ECM fits, particularly for the choice of the number of ZARC elements, and to obtain starting values and constraints for their fit. As an alternative approach, ZARC elements are fitted directly to the DRT.

3.2 Methodology

3.2.1 DRT of electrochemical systems using Debye kernel

The DRT is an estimation of a distribution function obtained by deconvolution of the frequency-dependent impedance response $Z(\omega)$, with ω an angular frequency, of an electrochemical system using an RC kernel, also called Debye kernel [82]. Thereby the impedance is represented as a distribution of RC parallel circuit elements, each characterized by its relaxation time $\tau = RC$, where R is the resistance and C the

capacitance of the respective element. Then, [19]

$$Z(\omega) = R_0 + R_\mathrm{p} \int_0^\infty \frac{\gamma(\tau)}{1 + \mathrm{j}\omega\tau} \,\mathrm{d}\tau \,, \tag{3.1}$$

where R_0 is the Ohmic resistance, R_p the polarization resistance, $\gamma(\tau)$ with $\int_0^\infty \gamma(\tau) d\tau = 1$ is the distribution function used to fit the impedance data, and $\mathrm{j} = \sqrt{-1}$ is the imaginary unit. Such an inversion has been theoretically shown to be mathematically possible for data of any non-oscillating electrical system [69]. However, a direct physical interpretation is primarily viable for an impedance response of RC or ZARC features.

Since EIS spectra typically span a frequency range of several orders of magnitude, it is common to use a logarithmic frequency variation for the measured data and also represent the DRT logarithmically scaled [82]. Thereby, Equation 3.1 is replaced by

$$Z(\omega) = R_0 + R_\mathrm{p} \int_{-\infty}^\infty \frac{\Gamma(y)}{1 + \mathrm{j}\omega\tau_\mathrm{r} \exp(y)} \,\mathrm{d}y \,, \tag{3.2}$$

substituting $y \equiv \ln(\tau/\tau_\mathrm{r})$ as the independent variable, and $\Gamma(y) = \tau\gamma(\tau)$. Reference frequency τ_r, which is usually selected as $\tau_\mathrm{r} = 1\,\mathrm{s}$, is used to obtain a unitless argument for the logarithm.

Experimentally, Equation 3.2 is discretized as [83]

$$Z(\omega_k) = \sum_{l=0}^{N} \frac{g_l}{1 + \mathrm{j}\omega_k\tau_l} \,. \tag{3.3}$$

Here, ω_k are the experimental frequency values, g_0 represents the Ohmic resistance ($\tau_0 = 0$), g_l for $l \geq 1$ is a discretization of the continuous distribution $R_\mathrm{p}\tau\gamma(\tau)$ with $\sum_{l=1}^{N} g_l = R_\mathrm{p}$, and the corresponding τ_l are user-selectable, logarithmically spaced time constant values τ_l that cover at least the frequency range of the experiment. Equation 3.3 can be expressed in matrix notation using vector $\boldsymbol{\omega}$ containing all experimental frequency values ω_k and vector $\boldsymbol{\tau}$ containing the τ_l values. τ_0 can also be included in $\boldsymbol{\tau}$, albeit g_0 is calculated by only using Tikhonov regularization in standard form (vide infra)[81]. Then, a discretized kernel

$$\mathbf{K} = J \oslash \left(J + \mathrm{j}\,\boldsymbol{\omega}\boldsymbol{\tau}^\top\right) \tag{3.4}$$

in matrix form can be constructed, where J is a matrix of ones and $\oslash$ represents an element-wise division, and [83]

$$\mathbf{Z} = \mathbf{K}\mathbf{g} + \mathbf{e} \,. \tag{3.5}$$

$\mathbf{Z}$ is the vector of experimental impedance values, $\mathbf{e}$ is a vector containing all noise contributions to $\mathbf{Z}$, and $\mathbf{g}$ is a real vector representing the true underlying distribution of relaxation times that describes the data in the frame defined by the choice of $\boldsymbol{\omega}$, $\boldsymbol{\tau}$ and the kernel equation. $\mathbf{g}$ has the same number of elements as $\boldsymbol{\tau}$ and $\mathbf{Z}$ the sames as $\boldsymbol{\omega}$.

3.2.2 Tikhonov regularization

Since the inversion of Equation 3.5 is ill-posed, the equation cannot be directly solved for $\mathbf{g}$ and it is necessary to stabilize the problem numerically [84]. One possibility is a regularization approach, which uses a justified bias as an additional selection criterion among the many possible solutions according to their physical plausibility. This bias is chosen in a way that extreme or unphysical features of the DRT such as spikes, steep slopes, or extreme oscillations of the distribution are penalized. The penalty or constraint may be chosen such that certain features can still occur, but only if justified by the data as an increase in fit quality that is not just overfitting of noise. The regularization algorithm should favor the solution with the broadest distribution of lines that is still consistent with the data, hence line broadening beyond the underlying distribution of time constants corresponds to the uncertainty of the fit due to experimental noise.

The inversion of Equation 3.5 can be formulated as a minimization problem where $\mathbf{g}$ is estimated through the numerical calculation of the Tikhonov regularization in generalized form,

$$\hat{\mathbf{g}} = \min_{\mathbf{g}\in\mathbb{R}} \left[\|\mathbf{K}\mathbf{g} - \mathbf{Z}\|_2^2 + \lambda^2 \|\boldsymbol{\Lambda}\mathbf{g}\|_2^2 \right]. \tag{3.6}$$

Here, $\boldsymbol{\Lambda}$ is the Tikhonov regularization matrix, and λ is a scaling factor that determines the regularization strength. For practical DRT calculations, usually either the squared norm of the real part of the residual matrix, $\left\|\Re\left(\mathbf{K}\mathbf{g} - \mathbf{Z}\right)\right\|_2^2$, or its imaginary part, $\left\|\Im\left(\mathbf{K}\mathbf{g} - \mathbf{Z}\right)\right\|_2^2$, is used, although calculations using complex impedances as written in Equation 3.6 have also been suggested [85, 70]. The letter results in a sensitivity enhancement of about $\sqrt{2}$. Alternatively, an independent DRT calculation with the real and the imaginary part of $\mathbf{Z}$ can be used as consistency check, since the same real-valued $\mathbf{g}$ should be obtained in both cases. Such a check is analogous to a Kramers–Kronig test that is commonly conducted in standard ECM analysis to confirm causality, linearity, and stability of EIS data [86].

For Tikhonov regularization in standard form, $\boldsymbol{\Lambda}$ is the identity matrix. For the generalized form of the regularization, various approaches for the choice of $\boldsymbol{\Lambda}$ have been suggested, and it may be created by merging different contribution matrices designed

to represent specific prior assumptions or constraints.[87] The complete regularization matrix is calculated as

$$\mathbf{\Lambda}^\top \mathbf{\Lambda} = \sum_X \mathbf{\Lambda}_X^\top \mathbf{\Lambda}_X \tag{3.7}$$

by combining its different components $\mathbf{\Lambda}_X$, where X is a placeholder for the different contributions: a uniform penalty term to ensure smoothness of the result, a zero-crossing penalty to prevent unnecessary oscillations, and a boundary penalty to avoid non-physical outliers at the low- and high-frequency edge of the distribution.[88]

For DRT calculation of EIS data, the use of a second derivative operator has been suggested as a suitable choice for $\mathbf{\Lambda}$. A variation of this approach had been demonstrated using an inverse weighting of the individual components with the curvature of the distribution $\mathbf{g}$, effectively leading to a Uniform Penalty (UP) [81]. Thereby the goal is to perform smoothing of the DRT that is uniform in regions with non-negligible spectral density, in order to represent sharp as well as broad features in $\mathbf{g}$ as faithfully as possible.

The commonly employed non-negativity constraint on the DRT is avoided and replaced with a zero-crossing penalty [88] to prevent unnecessary oscillations of the distribution while allowing for inductive contributions that appear with a negative sign when a DRT is calculated using the Debye kernel. In some experimental samples, inductive contributions to the impedance form justified negative contributions to the DRT spectrum [89, 90]. Not suppressing these contributions removes a source of systematic error in the transformation and ensures that the respective information is retained during the DRT step.

Another penalty term regularizes the boundaries to prevent artificial runaway effects at the edges of the distribution. All penalty terms X are collected in $\mathbf{\Lambda}$, which is adjusted iteratively until convergence is achieved or a maximum number of iterations is reached.

Parametrization of $\mathbf{\Lambda}$ is generally done using predetermined weight factors for the different contributions [83], but λ is treated as a user-adjustable parameter. To allow for a fully automatized parametrization without adjustable parameter, one option would be an algorithmic parameter adjustment [91]. Alternatively, it has been suggested to set λ such that the regularization term in Equation 3.6 equals the norm of the experimental noise [87]. Since the UP regularization is rather insensitive to the exact parameter selection [81], the latter option represents a simpler and still robust choice. It has been successfully implemented not only for DRT [71], but also for applications with exponential kernel, which is more problematic in terms of noise enhancement in case of underregularization than the Debye kernel [88]. As a test criterion for a correct

parametrization, it must be ensured, for example by testing with simulated data, that the inverted distribution does not show smaller line widths than the ideal result.

Since noise is not uniform across the whole EIS spectrum [92], Equation 3.6 is extended with a weight matrix **B** to account for variations of the noise of the impedance spectrum. In the simplest case without cross-correlation of noise between different frequencies, **B** is a diagonal matrix containing the inverse of the noise standard deviation of the data vector **Z**. It is used to weight the data and, accordingly, the kernel [81]. Then the residual term in Equation 3.6 becomes

$$\left\|\mathbf{B}(\mathbf{Kg}-\mathbf{Z})\right\|_2^2 . \tag{3.8}$$

To apply UP regularization with weighted data, it is necessary to determine the noise level individually for each experimental frequency.

A DRT performed without frequency-based weighting will either underestimate the high noise that is typically found at low frequencies, which leads to under-regularization and fitting of noise that causes spurious features in the DRT spectrum, or regions with low noise are over-regularized. Thereby, information available in the data may be suppressed, leading to non-random residual differences between data and fit.

The deconvolution is successful if the resulting residuals show no systematic patterns beyond statistical noise.[71] If the residuals show systematic elements, they contain information about the data that has been hidden from the DRT. A DRT, on the other hand, will not always be able to separate all features present in the data. This is influenced by feature distance and shape as well as by the noise amplitude [93]. The main goal is to extract all features that are justified by the data, with uncertainty due to experimental noise represented by the width of a feature in the DRT. In other words, the aim is to find a DRT with modes as broad as possible that still causes no residuals with systematic, non-random contributions.

3.2.3 Noise characterization

Noise model

Based on work by Macdonald *et al.* [92], Agarwal *et al.* suggested an empirical noise model that will serve as a reference to assess the plausibility of the estimated noise values [94]. The model assumes an identical standard deviation σ for both real and imaginary part of the noise,

$$\sigma_\mathrm{r}(\omega)=\sigma_\mathrm{i}(\omega)=\sigma(\omega)=\beta\left|Z_\mathrm{r}(\omega)\right|+\alpha\left|Z_\mathrm{i}(\omega)\right|+\gamma\frac{\left|Z(\omega)\right|^2}{R_\mathrm{m}} . \tag{3.9}$$

Z is the impedance with its real and imaginary part Z_r and Z_i, respectively. R_m is the value of the current-measuring resistor. Parameters α, β and γ are empirically determined based on experimental data. It was suggested that noise can be reproduced well in the mid-to-high frequency range above about 1 Hz, which is the region typically described by ZARC elements [94]. The impedance and its standard deviation are frequency dependent while the other parameters are scalars.

Statistical noise characterization

Repeating an experiment multiple times is the conceptually simplest way to determine its frequency dependent noise amplitude. However, the impedance of a battery depends sensitively on external parameters such as temperature, but also on potentially slowly relaxing kinetic effects in the active material or its surface if fast charge or discharge has been employed. Such factors may lead to fluctuations or drifts of the impedance that could easily surpass the intrinsic noise at a particular frequency. Consequences are twofold: the noise obtained when an experiment is repeated and analyzed independently at each frequency could be exaggerated, and causality of the impedance response could be violated. The latter effect may void the applicability of the DRT concept altogether, but at this point it is assumed that the system can be considered as quasi-stationary for each individual acquisition. The former effect leads to noise components at very low frequencies on the order of the repetition rate of the experiment. Under the assumption of independent and identically distributed (iid) Gaussian white noise for each data point, and the assumption of only a small noise variation between neighbouring frequencies in the EIS spectrum, the problem of very low frequency noise contributions and drifts can be mitigated by noticing that the noise variance of individual data points is a factor 2 smaller than the noise variance of the difference between two data points. The noise variance can then be estimated independently for the real and the imaginary part of the impedance using

$$\sigma^2(f_k) = \frac{\sum_{n=1}^{N_\mathrm{EIS}/2} \Big(\big(Z_{2n}\left(f_k\right) - Z_{2n}\left(f_{k+1}\right) \big) - \big(Z_{2n-1}\left(f_k\right) - Z_{2n-1}\left(f_{k+1}\right) \big) \Big)^2}{2N_\mathrm{EIS}} \,. \tag{3.10}$$

Here, N_EIS is the total number of measurements and $Z_n(f_k)$ is the impedance value in data set number n at frequency f_k of the k^th data point in each measurement. N_EIS is assumed to be even, and $\sigma(f_{k_\mathrm{max}}) = \sigma(f_{k_\mathrm{max}-1})$ is used at the maximum frequency value f_{k_max}. The difference is calculated both along the frequency dimension as well as between individual repetitions of the experiment. The noise level estimations obtained via this method were sufficient to provide the weights for stable DRT inversions.

Gaussian process regression

Repeating each measurement multiple times is not practical at scale. Alternatively, by applying a GPR in a sliding window across the impedance vector, an estimate for the noise level could be achieved from just one measurement. GPR can be defined as an ordinary Bayesian regression with an infinite dimensional parameter space of unknown nonlinear regression functions [95]. It is based on a Gaussian process, which is a stochastic process parameterized by two functions. Firstly, its mean function

$$m(x) = \mathbb{E}\left[g(x)\right], \tag{3.11}$$

where $g(x)$ is the defined, but unknown function of the underlying process, $\mathbb{E}$ denotes the expected value operator and x is variable. The second defining function of the process is its covariance function

$$c(x, x') = \mathbb{E}\left[(g(x) - m(x))(g(x') - m(x'))\right]. \tag{3.12}$$

Here, $m(x)$ is the mean function from the previous step and x' is another variable. A Gaussian process that can be characterized by these two functions is written as

$$g(x) \sim \mathcal{GP}(m(x), c(x, x')). \tag{3.13}$$

This Gaussian process is suitable for specifying the probability distribution of a random function [96]. The goal of GPR is to reconstruct the underlying signal by eliminating the contaminating noise. This is done through computing a weighted average of the noisy observations $\mathbf{y}$ as

$$\bar{g}(x) = \tilde{\mathbf{c}}(x)^\top (\mathbf{C} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}\,, \tag{3.14}$$

Here, $\bar{g}(x)$ is the predicted mean of $g(x)$. $\tilde{\mathbf{c}}(x)$ denotes the vector of covariances between the test point x and the other points in the data set, $\mathbf{I}$ is the identity matrix, $\mathbf{C}$ the covariance matrix of the data vector, and σ^2 the noise variance. Both $\mathbf{C}$ and σ^2 are calculated and optimized in the GPR algorithm implementation provided by the python package `sklearn`. [96]

The algorithm was applied to all possible subsets of adjacent data points in a moving frequency range of two decades inside one EIS measurement. Each frequency data point ω was subsequently assigned a noise level that was calculated as the average of the noise level estimation results $\sigma_{\mathrm{GPR}}(S)$ of all subsets S that this data point was part of,

$$\sigma(\omega) = \sum_{S \ni \omega} \sigma_{\mathrm{GPR}}(S) / \#(S \ni \omega)\,. \tag{3.15}$$

3.2.4 Equivalent circuit model fitting

Fitting of a suitable ECM is a common method for interpreting EIS data. An ECM comprises electrical circuit elements that are connected in series or in parallel to represent electrochemical processes. Each element is characterized by its frequency-dependent complex impedance function with several element-specific parameters. The experimentally probed frequency range of the impedance spectrum can often be subdivided into three sections. The intersection between EIS curve and impedance real axis at the high frequency end can be interpreted as serial Ohmic resistance R_0, which represents the voltage drop and power loss to heat when an outside voltage is applied. The semicircles in the mid-frequency range reflect charge transfer and double layer capacitance at the electrode and electrolyte interface [97], and it can be expressed as serially connected ZARC elements. The diffusion process inside the electrodes, which is visible as a straight line with a characteristic phase angle in the low frequency region, can be modeled by Warburg elements.

The complex impedance function of a ZARC element, which is functionally a parallel circuit of a constant phase element Q and a resistor R, is

$$Z_{\mathrm{ZARC}}(\omega) = \frac{R_{\mathrm{ct}}}{1 + (\mathrm{j}\omega\tau_{\mathrm{ZARC}})^{\phi}} \,. \tag{3.16}$$

Here, R_{ct} is the charge transfer resistance, τ_{ZARC} is the mean time constant, and ϕ is the depression factor. If $\phi = 1$, the ZARC behaves as an RC parallel circuit element. If $\phi = 0$, it behaves as a resistor.

For finite space Warburg elements, the complex impedance is evaluated as [98]

$$Z_{\omega} = \frac{A_W}{\sqrt{\omega}}(1 - \mathrm{j}) \coth\left(C_{\mathrm{lim}} A_W \sqrt{2\mathrm{j}\omega}\right) \,, \tag{3.17}$$

where A_W is the Warburg coefficient and C_{lim} is the limit of capacitance. A direct interpretation in the DRT domain is not straightforward. While fitting is possible [69], the result depends on the signal-to-noise level of the data and the choice of data points used to evaluate the DRT. The latter effect offers a straightforward possibility to identify FSW contributions by calculating a DRT with a varying number of data points. While ZARC elements should be affected only marginally, the distribution peaks related to Warburg elements move and increase in amplitude when additional data points are added at the low-frequency end of the EIS data set.

Finite length Warburg elements, on the other hand, become resistive at low frequencies, and an analytic DRT expression exists that is independent of the selected

data points or the sensitivity of the measurement [99].

Traditionally, ECM fits are used to extract parameters based on a priori knowledge about features of the EIS spectrum of an investigted system, yet this method has certain drawbacks. In particular, finding the correct number and type of elements, such that each element has a physical meaning or is justified as an individual component by the data, is non-trivial. Furthermore, a numerical ECM fit of even an ideal model may converge in a local minimum that makes little physical sense. Fit quality alone is a poor indicator for the scientific utility of the fit. Choosing good initial values and constraints for the fitted parameters is required to obtain physically meaningful results. This may be possible by noticing that different ECM elements depict features at different frequencies in the DRT. A DRT can thus be a useful preparation step for ECM fitting. For one, no prior assumption regarding the number and type of circuit elements is necessary. On the contrary, a DRT provides information regarding the number of elements that can be separated purely based on the data, which can be especially helpful to define the number and values of ZARC elements in a model. Additionally, the DRT technique can also give information regarding the noise bandwidth and a soft boundary of its parameters. For diffusion processes and the parametrization of Warburg elements, DRT with a Debye kernel may not be most suitable [100], but an empirical fit is possible. Therefore the statistics of the residuals does not get degraded by the presence of processes giving rise to Warburg behavior [69].

DRT with a Debye kernel provides direct information regarding ZARC elements. Since a ZARC element can be represented analytically in the DRT domain as [101, 76, 102]

$$\gamma_{\mathrm{ZARC}}(\tau) = \frac{R_{\mathrm{ct}}}{2\pi} \frac{\sin(\phi\pi)}{\cosh\left(\phi \ln \frac{\tau}{\tau_{\mathrm{ZARC}}}\right) + \cos(\phi\pi)}, \tag{3.18}$$

it is possible to directly fit ZARC elements to DRT modes. Thereby, advantage can be taken of the fact that the DRT provides direct information regarding the type and number of distinguishable ECM features, alleviating the drawbacks of traditional ECM fits while retaining their full utility, and requiring only a slightly increased computation time.

3.3 Methods and Materials

3.3.1 Experimental setup

A $LiCoO_2$ LIR2032 button cell battery with a nominal capacity of 45 mAh was investigated. A total of 21 EIS measurements per SOC point were conducted on a

Biologic SP-200 potentiostat with minimal state-of-charge variation for subsequent DRT processing and noise characterization. All experiments were carried out at room temperature in an uncontrolled environment using the same battery.

In a first measurement, the capacity was determined by discharging the fully charged cell at the expected nominal 1C rate of 45 mA until reaching the End of Discharge Voltage (EODV) of 3 V. The resulting sum of the drawn current defined the prevailing capacity of this particular cell to 24.57 mAh.

The button cell was then charged using the CC–CV method. First, the cell was charged with a constant current of 0.5C until reaching the upper voltage limit of 4.2 V. The voltage was then held until the charging current reached the lower threshold of C/20. After resting for 2 h to ensure that the battery had reached the rest potential, the first EIS measurements at 100% SOC were conducted. Afterwards, the cell was discharged at C/10 for 1 h, followed by another rest period of 2 h and the next EIS measurements. This process was repeated until the potential reached the EODV, followed by the last EIS measurement at 0% SOC.

The frequency vector $\mathbf{f} = \boldsymbol{\omega}/2\pi$ used for galvanostatic EIS measurements ranged from 1×10^{-2} Hz to 3×10^{5} Hz. The current amplitude was automatically varied to reach a minimum voltage excitation of 10 mV. Each measurement was averaged over two periods and contained 84 logarithmically spaced data points with 12 points per decade, except for the 6 lowest frequency points that were spaced at 6 points per decade. This allowed faster measurements, as the impact of low frequencies on the duration of EIS experiments was dominant.

3.3.2 DRT analysis

DRT inversion was calculated using a home-written Python script [83], using the imaginary part of the impedance data. The DRT relaxation time vector $\boldsymbol{\tau}$ was chosen from 10^{-10} s to 10^{10} s and contained 500 logarithmically spaced data points. Before calculating the final DRT, the algorithm was performed twice to obtain additional parameters. The first time, an impedance data scaling factor was chosen to normalize the variance of the residuals between the experimental data and the fit, determined by transforming the calculated DRT back into the frequency domain, to a value of one. This stabilized subsequent numerical calculation. The second time, the residuals were used to generate a frequency dependent noise level vector estimation via GPR using the numerical noise estimation algorithm written by Metzen and Lemaitre [103] in scikit [104], which was applied in a logarithmically scaled sliding window with a length of two decades. The third and final DRT was calculated by using the scaling factor and weights obtained in the previous steps.

In all the figures, the shown DRT was unscaled with respect to the definition in Equation 3.5 and labeled as arbitrary units (a.u.). In other words, by multiplying the specified numerical values with the kernel, the fit of the measured impedance was obtained directly.

3.3.3 ECM Fitting

The button cell was modeled by one serial resistance, three ZARC elements [67], and one Warburg element in series. Simulations were conducted using the ISEA Framework [17]. After parametrization, the ECM was also used to simulate the real-time current and voltage response, effectively enabling a fully numerical validation of the proposed data analysis protocol.

3.4 Results and Discussion

3.4.1 EIS experiments

Figure 3.1 shows repeated EIS measurements of a button cell battery at a state of charge of 37.5%. The spectra exhibit variations beyond random noise, particularly at low frequencies. This suggest systematic influences on the experimental data that are not considered in the theoretical noise model [94, 92]. It shows that, due to drifts or very slow fluctuations, a direct use of a set of reference measurements is not sufficient for a selective characterization of frequency dependent random noise, which is necessary for the determination of weights for data fitting. Similar results have also been reported in a recent study with a carefully implemented and optimized experimental setup [105], which illustrates that the presented results are representative for EIS investigations of batteries. Considering difference values between neighbouring data points in the EIS spectrum, systematic contributions get reduced considerably and remain immediately apparent only in the low-frequency region of the real part of Z (Figure 3.1, bottom row). Using the difference method also along the dimension of the experiment repetition, as described by Equation 3.10, further reduces systematic contributions, facilitating a statistical analysis of random noise as required for a weighted DRT analysis.

The noise standard deviation for the real and imaginary part of the data, calculated using Equation 3.10, is shown in Figure 3.2.

The noise level is compared with a fit using the Agarwal model (Equation 3.9). A good correspondence is observed over the whole frequency range for the imaginary part of the noise. The real part shows some deviations in the region below 0.1 Hz. As observed in the low frequency region of the real part of the impedance in Figure 3.1,

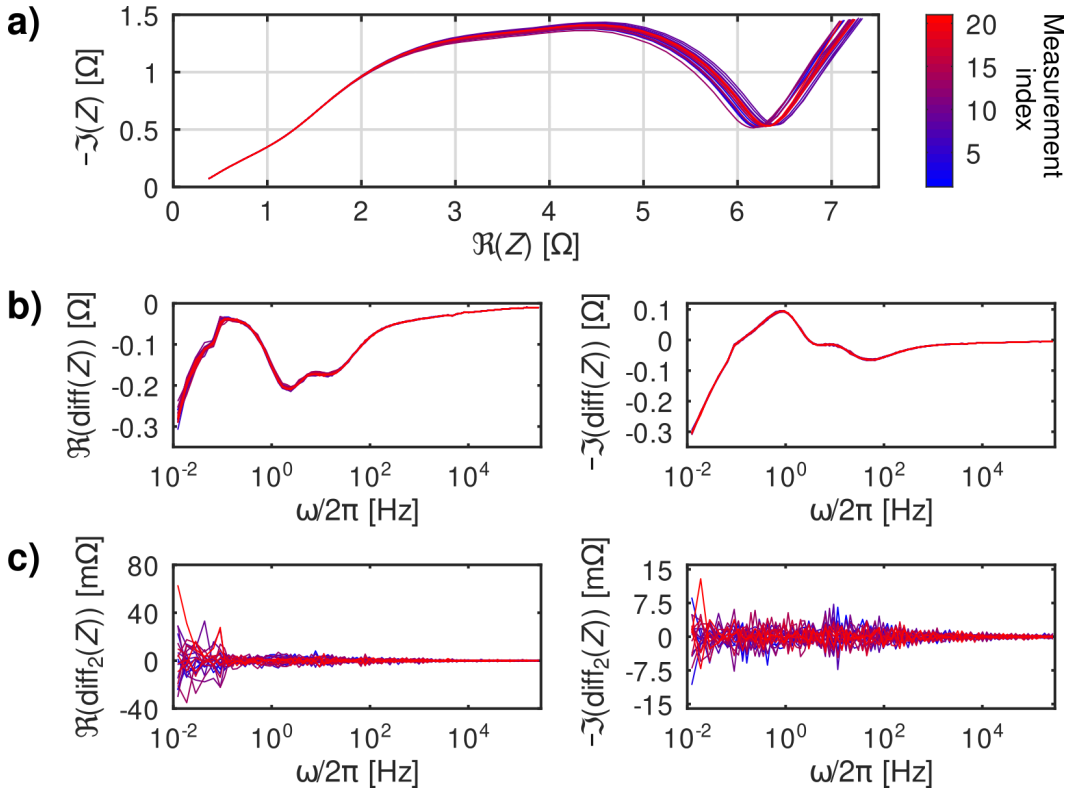


Figure 3.1: *a*) Nyquist plots of a series of 21 EIS measurements of a Li-ion button cell battery at 37.5% charge. A drift of the different measurements is observed, but not fully investigated in this work. *b*) Bode plots of the differences between neighbouring data points along the frequency axis for the real part (left) and for the negative imaginary part (right) of the impedance. *c*) Bode plots of the differences between neighbouring data points along the frequency axis and along the measurement index dimension for the real part (left) and for the negative imaginary part (right) of the impedance, as used in Equation 3.10. These plots illustrate how the measurements are mostly congruent, with the exception of the real part of the impedance at low frequencies. *(These figures are produced by Josef Granwehr)*

there are some systematic drift or very low frequency residuals even when noise is estimated using the difference method. Apart from this, trends correspond well when comparing a statistical noise analysis and the Agarwal model. The values for α, β and γ are comparable when independently determined for the real and the imaginary part of the noise, hence the hypothesis in [94] that $\sigma_r(\omega) \approx \sigma_j(\omega)$ is justified. Overall, this comparison shows that the noise estimated experimentally is consistent with what had been reported in literature for such experiments, hence the determined noise values are a suitable basis for the estimation of weights for a weighted DRT inversion.

The results from noise estimation via GPR and the statistical method are compared in Figure 3.3.

The two methods show a good agreement in the intermediate frequency range be-

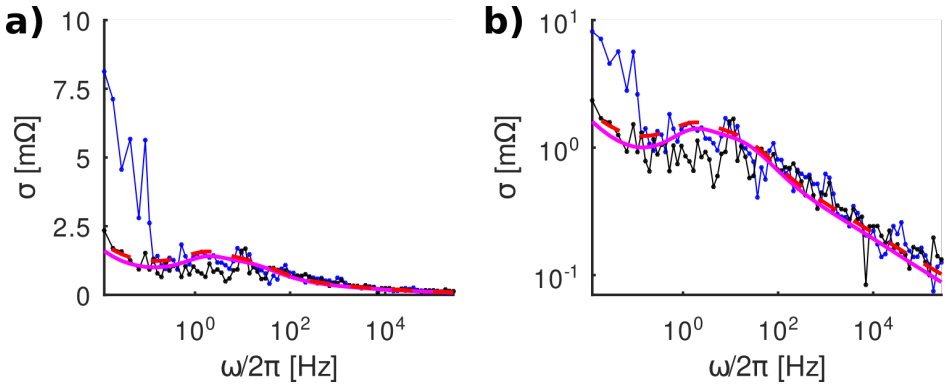


Figure 3.2: Statistically determined noise of real (blue) and imaginary (black) impedance data shown in Figure 3.1. The Agarwal model (Equation 3.9) was fit to provide $\alpha = 6.49 \cdot 10^{-4}$, $\beta = 1.42 \cdot 10^{-4}$, $\gamma/R_m = 0$ for the real part (dashed red line), and $\alpha = 6.61 \cdot 10^{-4}$, $\beta = 1.04 \cdot 10^{-4}$, $\gamma/R_m = 0$ for the imaginary part of the data (solid magenta line). For the fit of σ_r, only data points above $\omega/2\pi = 0.1$ Hz were used. *a*) Linearly scaled *y*-axis. *b*) Logarithmically scaled *y*-axis. *(These figures are produced by Josef Granwehr)*

tween about 0.1 Hz and 100 kHz. At higher frequencies, GPR shows a larger noise level than the statistical analysis. The reason is the presence of a systematic feature, which resembles Gaussian noise at the chosen sampling density but is present in all the data sets. If noise is estimated from a single measurement only, such systematic features cannot be distinguished. If the GPR noise level is determined from repetitions of the experiment, then the similarity of the noise becomes apparent as well, leading to a reduced noise estimate (Figure 3.3c). In the low frequency region, the statistical method produces a higher noise estimate. One reason for this observation is the tendency for overfitting by GPR when insufficient data is available, as may be the case at the boundaries of a data set [106]. Furthermore, the statistical method may overestimate the random noise level due to the above mentioned non-random drift contribution that is not reproduced for different repetitions of the experiment. In this case, GPR may even provide a more faithful set of weights for DRT inversion. Overall, DRT results with GPR noise estimation are close to statistical analysis results while requiring only a single EIS measurement.

3.4.2 DRT calculation

The influence of weighting is demonstrated in Figure 3.4, which compares three DRTs from the same data set, one calculated with uniform weighting based on the mean noise amplitude across all frequencies, one using weights determined statistically using all data sets, and one with weights from GPR.

When comparing the DRTs, they show very similar features. Since the DRT mode

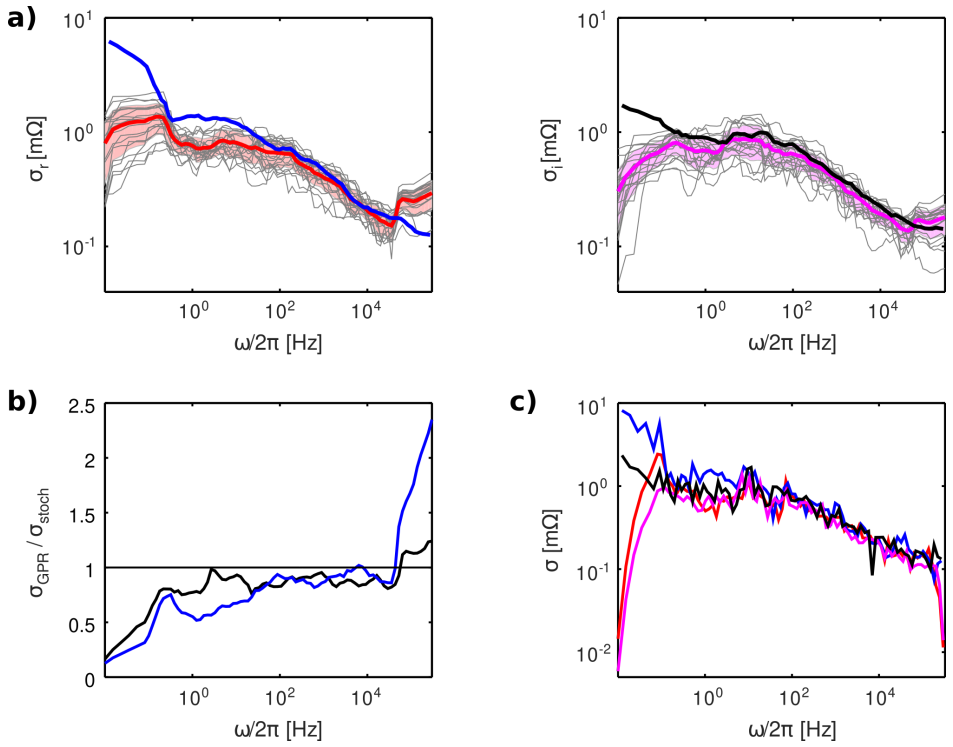


Figure 3.3: Comparison of noise estimate for statistical and GPR analysis of a series of 21 EIS measurements of a Li-ion button cell battery at 37.5% charge. The results for the two methods are similar enough to be used interchangeably for weight determination of DRT transformations. *a*) Noise standard deviation for the real (right) and the imaginary part of the impedance (left), determined using a moving average over two frequency decades. (blue, black) Statistically determined noise. (red, magenta) Mean GPR determined noise level of all the data sets, which were individually averaged over two frequency decades. The shaded regions represent the standard deviation. The gray curves show the GPR noise estimate after moving average for individual data sets. *b*) Ratio of noise standard deviation between statistically determined and GPR noise levels for the real (blue) and the imaginary part of the impedance (black). averaged over two frequency decades. *c*) Comparison of statistically determined noise standard deviation (blue: real; black: imaginary) and standard deviation of residuals between GPR estimate and experimental data (red: real; magenta: imaginary), without moving average. *(These figures are produced by Josef Granwehr)*

is broad at high frequency, where the noise level is small, the distributions appear similar.

As already noticed by Borgia *et al.*, UP regularization is fairly robust with respect to the exact value of the regularization parameter or, in this case, the weights, which play a similar role on a local scale [81]. In this particular case, this is amplified by the broad DRT modes without any particularly narrow features, especially in the high frequency region where the noise level is lower and a higher resolution could be achieved. However, while visually the differences between the three DRTs may appear marginal,

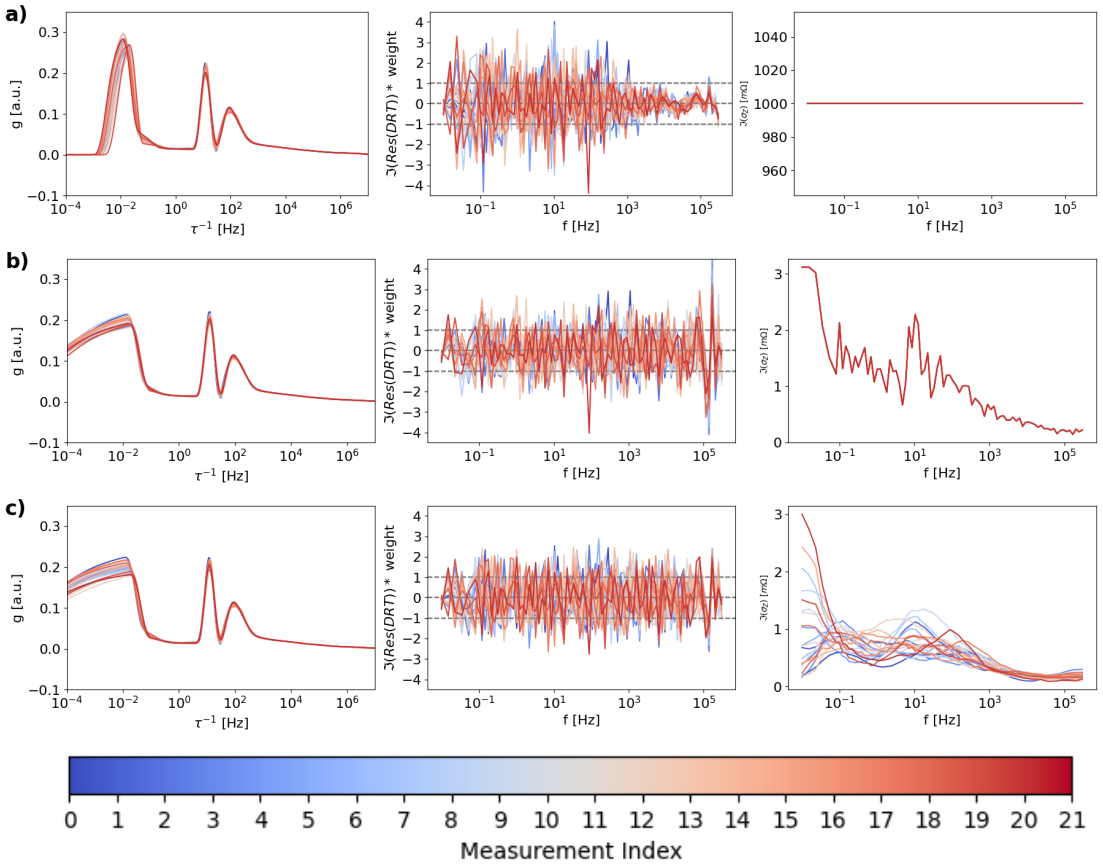


Figure 3.4: DRT of 21 EIS measurements of a Li-ion button cell at an SOC of 37.5% performed using different weighting methods. *Left Column:* Calculated DRT. *Middle Column:* Weighted residuals. *Right Column:* Frequency dependent uncertainty used for weighting. a) DRT performed with uniform weighting. b) Weighted DRT performed using statistically determined weights. c) Weighted DRT performed with algorithmically determined weights using GPR. *(These figures are produced by Christian Bartsch)*

the residuals are more discriminating. With unweighted DRT (Figure 3.4a), systematic residuals above the random noise level are clearly apparent at frequencies higher than 1 kHz. Such residuals are indicative of lost information that is contained in the data. Particularly when looking for small changes during cycling of a battery, such a loss of information in the frequency range that is sampled most sensitively by EIS may be significant. Furthermore, an automatized, data-driven analysis of DRT may be able to discern features that are suppressed by unweighted data processing. And finally, batteries or battery components that with narrower high frequency modes are expected to show a higher resolution with weighted DRT analysis.

A comparison of weighted DRT using weights derived from statistical analysis (Figure 3.4b) with weights generated via GPR (Figure 3.4c) using the same data produces very similar results. This does not only apply to the DRT itself, but also to the residu-

als. In both cases, no systematic features are apparent any more. Since the DRT works similarly well with both weight determination methods, the GPR method is preferred, as it works on single measurements rather than larger sets of measurements. This approach was capable of automatically generating an estimated noise vector suitable for a stable DRT from a single EIS measurement of a battery.

The low-frequency mode of the DRT varies strongly between measurements and changes considerably when multiple DRT inversions of the same measurement are calculated while omitting between zero and five of the lowest-frequency data points. This is consistent with the expected behavior of a Warburg element. Therefore this frequency range is excluded from the ZARC fits below, and no quantitative interpretation is attempted based on the DRT data.

3.4.3 Numerical Verification of Weighted DRT

The fidelity of the DRT deconvolution algorithm was verified by analyzing synthetic data using a simulated noisy data set of a serial circuit of four ZARC elements (Figure 3.5). For this, an impedance spectrum and an analytical DRT spectrum were calculated and compared with the DRT of the impedance spectrum with added noise. The noise was generated according to the empirical frequency dependent noise level determined through the statistical method (Figure 3.2). The numerical result was compared to the expected analytical DRT determined from the source parameters of the simulated ZARC elements. The simulation was created using the parameters in Table 3.1, Equation 3.16 for ZARC impedance, and Equation 3.18 for the analytical DRT of a ZARC element. These equations were evaluated for the same 84 logarithmically spaced frequencies between 1×10^{-2} Hz and 3×10^{5} Hz that were also used in the experimental EIS measurements.

Name	$R_{\mathrm{ct}}[\Omega]$	ϕ	$\tau_{\mathrm{ZARC}}[\mathrm{s}]$
ZARC 1	2.5	0.9	1
ZARC 2	5	0.85	10^{-2}
ZARC 3	2.5	0.3	10^{-4}
ZARC 4	10	0.75	10^{-6}

Table 3.1: Simulation parameters for the ZARC elements in Figure 3.5.*(It is produced by Christian Bartsch)*

The result of the DRT calculation is presented in Figure 3.5, along with the an-

alytically determined DRT of the circuit. The contributions of the individual ZARC elements, calculated using Equation 3.18, are also shown.

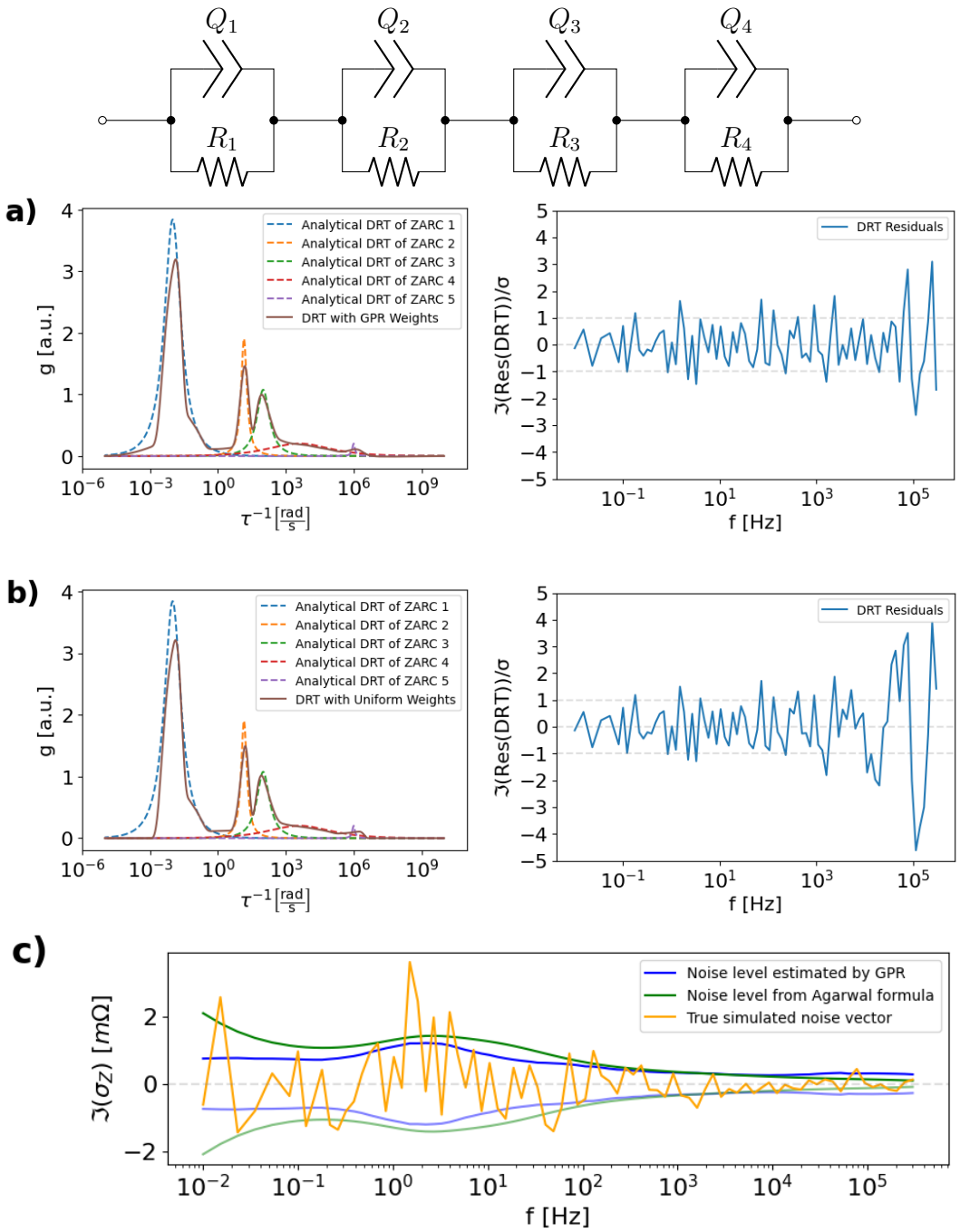


Figure 3.5: *Top:* ECM used for the simulation. It consists of four ZARC elements, numbered 1 to 4, each of which is modeled as a parallel circuit of a resistor R and a constant phase element Q. a,b) Comparison of analytical DRT of simulated circuit, determined from the simulation source parameters in Table 3.1, with different DRT fits for demonstration. a) DRT fit of the simulated impedance using weights determined via GPR. (DRT on the left, residuals on the right side). b) DRT fit of the simulated impedance using uniform weights. (DRT on the left, residuals on the right side). The Z-dependent noise for the simulation was created using the Agarwal model with the parameters determined for the imaginary part by the fit in Figure 3.2. c) Noise level estimated via GPR alongside true noise vector of the simulation. *(These figures are produced by Christian Bartsch)*

Generally, the numerical DRT using GPR determined weights (Figure 3.5a) matches the analytical DRT. The numerical DRT shows a slight broadening of the sharper peaks, which can be attributed the artificial noise added before the DRT calculation. This is not undesired, since a fit of a distribution based on a noisy data set should never give a narrower distribution than the original, noise-free analytic distribution. At the bottom edges, the numerical DRT strives toward zero faster than the analytical DRT. This could be explained by the preference of the regularization method for a different lineshape. The bottom of each line may also be fitted with a slightly different shape, as long as the contribution to the signal is within the noise level. The shape of the high frequency mode somewhat deviates from the analytical shape, since it lies partly outside the measured frequency range, and the respective wing of the line is obtained by extrapolation. Nevertheless, it is important to notice that, based on the weighted residuals, the calculated numerical distribution is consistent with the analytical distribution, both in terms of position and integral of the lines. Overall, the spectrum was represented faithfully, with all features fed into the simulation recreated in the numerical DRT as far as they are supported by the data. Notice also that the GPR determined noise level matches the frequency dependent noise level used to determine a synthetic noise vector well even at the high and the low frequency end of the data set. This may be caused by the exact Gaussian character of the noise, which contrasts the experimental data, where noise is only approximately Gaussian.

The situation is different for the DRT obtained with uniform weighting. At low frequencies, an extrapolated additional feature is obtained that is caused by noise overfitting due to underregularization. This can be seen in the residuals, which are considerably smaller than the synthetic noise added to the ideal data at low frequencies. At high frequencies, the situation is reversed. There the uniform weight obtained by the mean of the noise is penalizing the respective data points too much, leading to additional line broadening and systematic residuals.

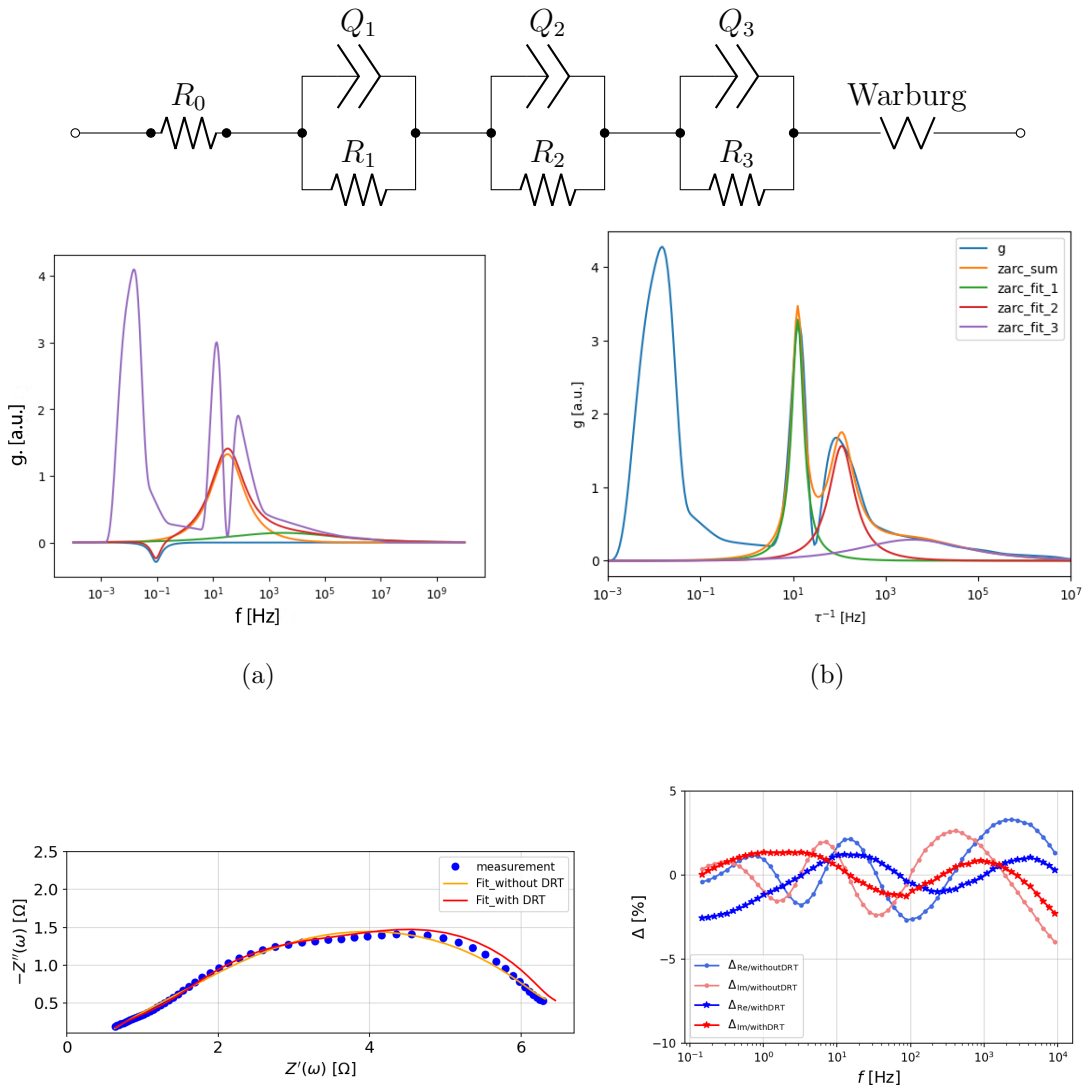
3.4.4 ECM Fitting of Real Battery

Direct ECM fit

Considering the DRT in Figure 3.4, the ECM shown in Figure 3.6b consisting of three ZARC elements, which cover features at frequencies above about 1 Hz, in series with a Warburg element for low-frequency features, appears a suitable model for the investigated Li-ion button cell. Nonetheless, at least one of the modes modeled by a ZARC element is showing an asymmetric shape, which either indicates a more complicated ECM with overlapping features, or a distribution of parameters that is not accurately

represented by a ZARC element.

In Figure 3.6a, the ZARC elements of a direct ECM fit obtained with random parameter starting values and without parameter constraints, transformed into the DRT domain, is compared with the obtained DRT.



(a) (b)

(c) Nyquist plot (d) Residuals

Figure 3.6: ECM fit to an experimental EIS spectrum of a Li-ion button cell at 37.5% charge, based on the battery model shown at the top. (a) Direct ECM fit, shown in the relaxation time domain. The three ZARC elements (yellow, cyan and green) of a direct ECM fit without parameter constraints and their sum (red) are overlaid with a DRT spectrum (purple). The Warburg element of the direct fit is not shown. (b) Three ZARC elements fitted directly to the DRT in an effort to provide a preprocessing step for a better ZARC fit during the electrochemical characterization of the system. (c) Nyquist plot of experimental EIS spectrum (blue), a direct ECM fit (orange), and a fit of the DRT (red). (d) Residuals for the different fits. *(Figures a and b are produced by Christian Bartsch)*

In this particular example, the ZARC element shown in yellow represents two DRT modes, while the cyan ZARC element forms a negative contribution in the DRT that corrects for a discrepancy of low frequency impedances and their fit with a Warburg element, which is not shown since an analytic representation in the DRT domain does not exist. Although the direct ECM fit without constraints is rather heavy-handed, some typical problems faced when fitting an ECM directly to EIS data can be identified. Dominant, asymmetric features tend to get fitted by multiple components at the expense of less prominent features. Furthermore, deviations of a mode from the shape of a fitting function may get compensated with unphysical or unrealistic features, such as the negative ZARC element that implies an inductive contribution as a correction for the Warburg element.

ECM fit using DRT data

Fitting the intermediate frequency range directly in the DRT domain using Equation 3.18 is shown in Figure 3.6b. It shows that three ZARC elements do offer a good description of the sample behavior above 1 Hz. Since the equation used to fit ZARC elements in the DRT domain is symmetric in log space, asymmetric modes in the distribution cannot be accurately reproduced. Nonetheless, the distinguishable features in the DRT could be individually fitted with this simple ECM.

As suggested in subsection 3.2.4, the DRT method using a Debye kernel is most useful for characterizing ZARC element contributions, therefore the focus here lies on the middle and upper frequency range of the EIS spectrum. In Figure 3.6c, Figure 3.6d, fit and residuals of a representative EIS spectrum are shown. The DRT-informed EIS fit and direct numerical EIS fit have a similar quality in the Nyquist plot, with comparable amplitudes of the residuals in the frequency domain despite considerably different DRT spectra and ECM parameters. Additional information from a DRT does not have a strong influence on the root mean square error between experimental and fitted data, consistent with results presented by Boukamp [19]. However, DRT-based information helps to build a more physically reliable battery simulation model. Notice that while DRT residuals can be reduced to essentially heteroscedastic white noise, ECM fitting shows systematic residuals that are several times higher [107]. Therefore, DRT is potentially suitable as a preprocessing step that contains the same information as the raw data, while an ECM fit is an effective, reduced – or compressed – representation.

ECM fit at different SOC

EIS data was first transformed via DRT to obtain a guess of the number and type of features contained in the data in order to get an appropriate circuit model repre-

sentation of the system. Once a suitable ECM was identified, ZARC contributions to the impedance spectra were algorithmically fitted to extract starting values and boundaries for an ECM fit that was performed on the raw data in the frequency domain. These steps were performed sequentially without user interaction or case-specific configuration, and are thereby suitable for large data sets.

Figure 3.7 shows the fitted parameters R_{ZARC}, τ_{ZARC} and ϕ_{ZARC} of the three ZARC elements for different SOC values of the Li-ion button cell with and without use of DRT.

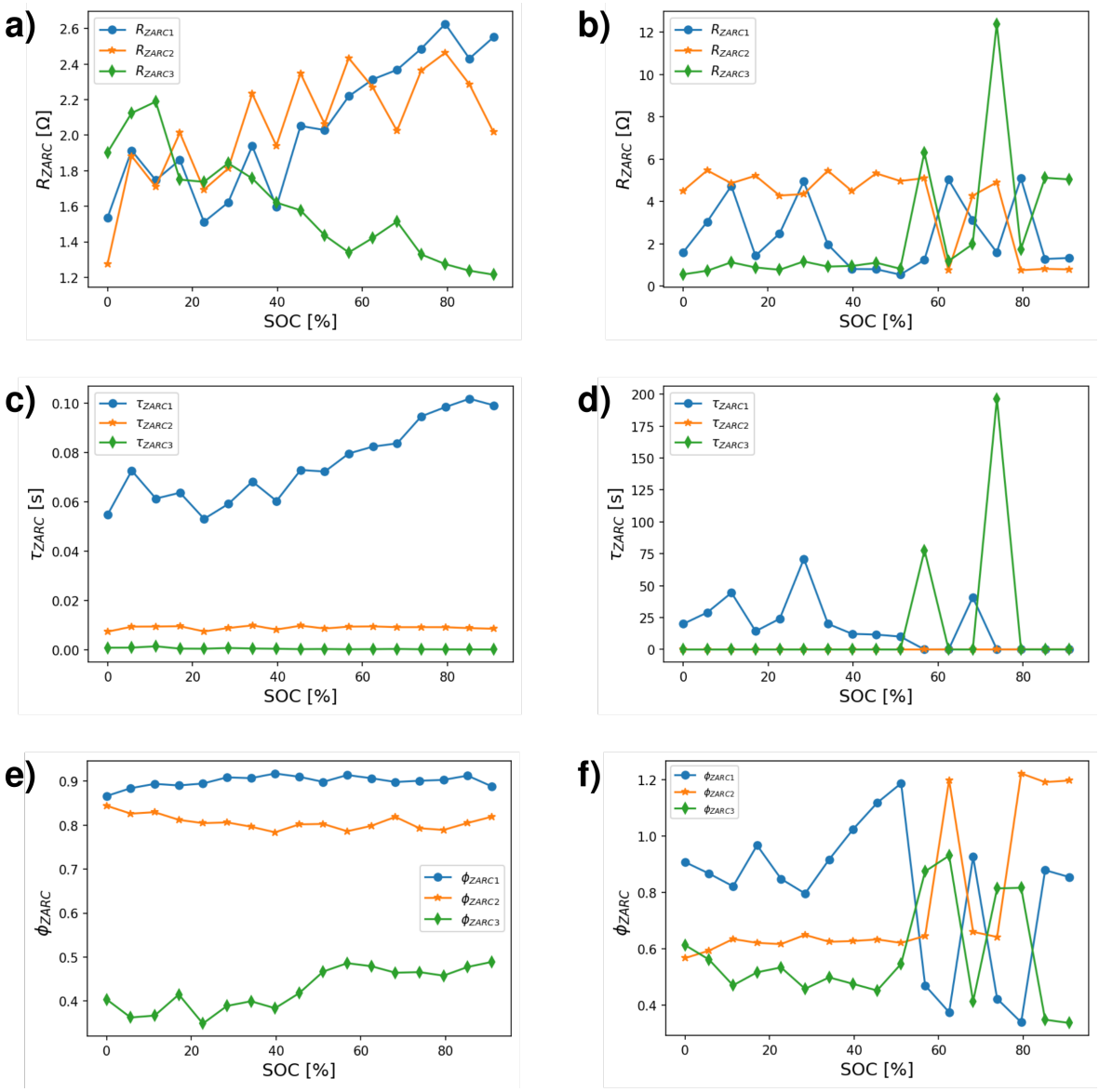


Figure 3.7: ZARC parameters of Li-ion button cell model shown in Figure 3.6, consisting of a resistor, a Warburg element and three ZARC elements in series. (top) Estimation of parameter values R_{ZARC} with (a) and without DRT (b). (middle) Estimation of parameter values τ_{ZARC} with (c) and without DRT (d). (bottom) Estimation of parameter values ϕ_{ZARC} with (e) and without DRT (f).

The resistance of ZARC elements shows less extreme values, fluctuations and outliers in the fits performed with DRT information. The parameter τ_{ZARC} represents the inverse frequency of each ZARC semicircle in the EIS spectrum. According to the DRT spectrum, the three ZARC elements have distinct positions with little overlap, as is observed in Figure 3.7c. Purely numerical parameter estimation without DRT information, on the other hand, exhibits overlap of fitted ZARC elements at higher SOC (Figure 3.7d). Fitting of ϕ_{ZARC} using DRT information shows fairly small, smooth variations as a function of SOC. By numerical fitting, the resulting values of ϕ_{ZARC} strongly fluctuate, which appears difficult to justify for a fresh battery. Furthermore, values of ϕ_{ZARC} are occasionally higher than 1 if no constraints for the fitted parameter are applied.

In the electrochemical data recorded during cycling, a continuous potential variation was observed that would not suggest the strong fluctuations observed in the direct ECM fits. This implies that a DRT supported ECM fit is more robust and allows for a better control regarding outliers. Since DRT and EIS contain the same information if DRT inversion was parametrized correctly, a physically more justified model can be obtained when incorporating DRT in the ECM fit.

3.5 Conclusions

Data weighting was demonstrated as an adequate approach to determine the DRT of a battery that shows typical heteroscedasticity of noise. Stochastic residuals without any non-random systematic features were obtained after a DRT fit, indicating that no information was lost during the DRT step. It was demonstrated that GPR can be used for model-free estimation of weights based on a single data set only. The obtained distributions were consistent with DRT obtained using a statistical noise analysis of multiple repetitions of the same experiment.

The GPR-based method produced reproducible and robust weights, which facilitates the use of DRT for the ECM parametrization with minimal or no user intervention. Thereby, advantage is being taken of the increased resolution of the DRT compared to the standard EIS spectrum without loss of information. Initially, the DRT was used to select a suitable ECM for the investigated button cell battery. Once established, the algorithm was demonstrated for the parametrization of this battery without manual adjustments. Robust ZARC parameters were obtained across the full range of charge states without obvious outliers while previously, ECM fit results of similar quality required manual adjustment of fit parameters and assumptions about the system based on sources other than the measured data. It has been shown that the stability

and quality of ECM fits were improved by including this additional knowledge about the physical state of the measured system in the fitting process without introducing arbitrary assumptions.

Such an automated parametrization based only on the available data opens the door for data-centered applications that use large sets of data spanning the whole lifetime of a battery.

Chapter 4

Sampling battery models at various SOC/SOH combinations

This chapter is adapted from the publication that has been submitted to the journal and is now under review: **Extending equivalent circuit models for state of charge and lifetime estimation**, by L. Jin, F. Bereck, J. Granwehr, and C. Scheurer.

The theoretical analysis about Chebyshev Polynomials for ECM parameters as a function of SOC, and model variation for covering over SOH, as well as modeling and simulation were conducted by the author of this thesis, Limei Jin. The experimental EIS data was provided by Franz Bereck.

4.1 Introduction

In recent years, the rapid growth of electric vehicles and grid energy storage has underscored the importance of high-performance and reliable battery systems [108]. Their efficiency and longevity are crucial for achieving sustainable energy solutions. To optimize battery performance and ensure their safe and efficient operation, the BMS is required to capture charge and health states from an operating battery [109]. While BMS is commonly utilized at the pack level to oversee the overall health and performance of a battery pack, the recognition of each individual cell with unique operational characteristics within the pack needs more granular and precise management and monitoring at the cell level.

These variations in cell behavior necessitate the implementation of a DT at the cell level. A DT is a real-time connected multi-scale virtual representation of physical asset. Its application in the field of battery systems combines multi-disciplinary modelling approaches with different types of on-board data and analysis techniques [4].

Black-box modeling and white-box modeling are two common approaches for bat-

tery modeling [110]. Black-box modeling treats the battery as a complex system with inputs and outputs, focusing solely on capturing its behavior without considering the underlying physical processes [7]. White-box modeling aims to represent the battery's behavior based on the fundamental physics and chemistry governing its operation [7]. However, both approaches have their limitations. Black-box models lack interpretability and do not provide insights into the underlying mechanisms, while white-box models can be computationally expensive and may struggle to capture all intricacies of complex battery systems [111]. To bridge this gap, gray-box battery modeling emerges as an alternative [112]. It incorporates the physics-based understanding of the battery system while leveraging data to enhance model accuracy and efficiency. The model structure captures the key electrochemical processes within the battery for meaningful interpretation of the results. At the same time, by iteratively refining the model using input data to calibrate or predict specific observed parameters, the gray-box model can better capture the nuances and complexities of real-world battery behavior. In most cases, they integrate with machine learning algorithms, to enhance their predictive capabilities while maintaining interpretability [113].

ECM approximate batteries as an electrical circuit with discrete components, such as resistors, capacitors, voltage sources, and non-classical circuit elements representing diffusion processes or chemical transformations. Their parametrization and evolution over time is extracted from measured EIS data at different operation states of a battery. When modeling a battery using an equivalent circuit, different components are considered to represent the various electrochemical processes occurring within the battery and the complexity of the equivalent circuit can vary depending on the specific battery chemistry and the level of detail required for analysis. A simplified ECM offers robustness in the context of creating a DT for a battery system. Although a more complex ECM structure has the potential to closely mimic real battery behavior, a simplified ECM offers a well-balanced compromise. It balances the computational efficiency needed for real-time or large-scale simulations while still providing valuable and interpretable insights into the battery's operation and performance. A snapshot of a simplified equivalent circuit model for a coin cell may include a serial resistance for voltage drop and power loss, ZARC elements for charge transfer and double-layer capacitance at the electrode-electrolyte interfaces, and Warburg elements for diffusion process [61]. Parameterized models can effectively capture the dynamic behavior of batteries while significantly reducing computational requirements compared to physics-based models. However, it is not only time-consuming to conduct EIS measurements at all possible battery states, but also robustly estimating and accurately monitoring internal parameters during aging is challenging. The main difference between an ECM

and a DT is that the former represents the current state of a system, while the latter is predictive. Therefore, an implementation of a DT may be possible based on an ECM that is suitable for a particular electrochemical system. The DT is based on a set of all parametrizations that represent a physically possible state of a battery. They form a hyperspace within which the battery evolves.

In an ideal battery, charging and discharging should be infinitely reproducible under a wide range of reasonable operating conditions. However, in reality, batteries degrade as they undergo multiple charging and discharging cycles [114]. To understand and characterize their behavior over time, the concept of separating time scales is useful. Thereby, *short time scale* refers to the behavior and changes that occur relatively quickly and often within one charging/discharging cycle. At a short time scale, SOC represents the current level of charge within the battery at any given moment. It reflects the immediate effects of charging/discharging cycles and can fluctuate frequently as the battery is used. It can be expressed as a percentage of the battery's full capacity [22]. Conversely, *long time scale* refers to slower changes that occur over the battery's lifetime. Over a long time scale, SOH designates its overall condition and performance compared to its initial, pristine state [115]. While SOC is relatively straightforward to define and measure, albeit not completely unambiguous, SOH is a more complex and nuanced concept that is not solely determined by the slow aging of a battery. For example, SOH can be affected by sudden and unexpected events, such as overcharging, overheating, or other adverse conditions, which can have a rapid and negative impact on the battery's health [116]. A battery can also experience partial recovery from specific types of damage or degradation, particularly when promptly addressing the adverse conditions [21]. Therefore, it is oversimplifying to express SOH as, for example, the number of charging/discharging cycles or the capacity ratio.

In the process of quantifying the dependencies of ECM parameters on SOC, conventional polynomial functions can be employed. However, they come with inherent limitations when dealing with high-order approximations, making them susceptible to issues like overfitting. To address these challenges and enhance the robustness of the parameter quantification process, orthogonal basis fitting methods, such as Chebyshev polynomials with their min-max properties, emerge as a potential approach. Details will be elaborated in the methodology section. The challenge of manifesting SOH within ECM is addressed by the shift of distinct regions, which are generated and delineated based on measurements taken from differently aged cells. The generation of battery models along with aging face limitations in the high-dimensional spaces. As the dimensionality increases, traditional random sampling methods struggle to provide representative and evenly distributed samples, resulting in inadequate coverage of the

space. Quasi-random sequences, such as QMC methods with low-discrepancy properties, emerges as a valuable strategy to overcome these challenges. Details will be followed in the methodology section.

In this chapter, we will demonstrate the application of Chebyshev polynomial fitting to parameterized simplified EC models mapped to experimental cells. Additionally, we will utilize quasi-Monte Carlo sampling to explore and differentiate three distinct aging cell clouds, providing valuable insights into the battery's lifespan dynamics, spanning from fresh to end-of-life conditions.

4.2 Methodology

4.2.1 Parametrization of ECM elements as a function of SOC

EIS is used to measure the batteries' responses at multiple points along charging/discharging cycles. These impedance data are fitted to ECM for a single cycle. This pointwise approach along the SOC dimension yields very noisy parameter values due to inherent measurement noise and the non-linear fitting procedure of equivalent models known to contain massive aliasing between model parameters. It has thus to be repeated without aging the battery noticeably, so that the assumption of a fixed state on the slow "SOH" timescale holds. The simplified ECM for a coin cell includes three elements: a serial resistance, ZARC elements, and a Warburg element, to represent different aspects of its behavior [17]:

$$Z(\omega) = R_s + \frac{\sigma}{\sqrt{\omega}}(1-j)\coth(C_{\text{lim}}\sigma) + \sum_{x=1}^{x=3}\frac{R_x}{1+(j\omega\tau_x)^{\phi_x}} \tag{4.1}$$

where R_s is the serial resistance. ZARC element is functionally a parallel circuit of a constant phase element representing a lossy capacitor and a resistor. Here τ_x is the mean time constant, ϕ_x is the depression factor, and R_x is the charge transfer resistance of three ZARC elements. If $\phi_x = 1$, the ZARC behaves as an parallel RC circuit element and if $\phi_x = 0$, it behaves as a ideal resistor. For FSW element, σ is the Warburg coefficient and C_{lim} is the limit of capacitance. The initial and boundary conditions of ECM fitting has recently been extracted by a transform based DRT analysis which allows one to reduce the ambiguity in the construction of ECM and thus over-fitting [15]. The fitted ECM parameters, such as R_s, R_x, τ_x, ϕ_x, σ and C_{lim},

vary depending on SOC, so that Equation 4.1 should be expanded as

$$Z(\omega, \mathrm{SOC}) = R_s(\mathrm{SOC}) + \frac{\sigma(\mathrm{SOC})}{\sqrt{\omega}}(1-j)\coth(C_{\mathrm{lim}}(\mathrm{SOC})\sigma(\mathrm{SOC})) + \sum_{x=1}^{x=3} \frac{R_x(\mathrm{SOC})}{1+(j\omega\tau_x(\mathrm{SOC}))^{\phi_x(\mathrm{SOC})}} \tag{4.2}$$

These dependencies can be quantified for each parameter as a function of SOC. The relationship between ECM parameters and SOC is often nonlinear, making it difficult to find a simple mathematical model that accurately captures the dependencies. Additionally, double fitting introduces additional sources of error and uncertainty into the parameter estimation process and increases the risk of overfitting. Overfitting occurs when a model is too complex and captures noise or random variations in the data rather than the true underlying relationships. This can lead to poor generalization and inaccurate predictions when the model is applied to new data.

This problem can be mitigated by using an orthogonal function basis, such as Chebyshev polynomials, to project data on the optimal solution. Furthermore, Chebyshev polynomials have the property of minimizing the maximum absolute value over a given interval [117]. This helps preventing physically unrealistically large values of the fitted polynomials in regions with poor sampling to facilitate closer fits in densely sampled regions. It leads to a more robust and accurate representation of the dependencies of ECM parameters on SOC across the entire range. It can ensure that the parametrization is equally well-determined across the SOC range, reducing the risk of overfitting or underfitting in specific regions.

Unlike regular polynomial regression with monomial basis functions x, it uses the orthogonal polynomials in a least-squares functional approximation,

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x) \tag{4.3}$$

where a_n is the coefficient associated with the Chebyshev polynomial $T_n(x)$ of order n, and $T_n(x)$ is defined as,

$$T_n(x) = T_n(\cos\theta) = \cos(n\theta), \quad x \in [-1, 1] \tag{4.4}$$

with weight function $w(x) = \frac{1}{\sqrt{(1-x^2)}}$ [118]. Chebyshev polynomials of the first kind have the advantage of being well-behaved and oscillating within the interval [-1, 1]. This is beneficial because the boundaries of the SOC range are less often realized in practice, and using Chebyshev polynomials can help mitigate any potential issues

related to extrapolation or fitting near the boundaries. For example, the relationship between parameter R_s and SOC could be expressed as

$$\tilde{R}_s(\text{SOC}) = a_1 T_1(\text{SOC}) + a_0 T_0(\text{SOC}), \quad \text{SOC} \in [-1, 1] \tag{4.5}$$

The goal is to find optimal coefficients a_n to minimize the error between functional estimated values $\tilde{R}_s(\text{SOC})$ and experimentally parameterized values $R_s(\text{SOC})$. Using the discrete orthogonality property, coefficients can be approximated as:

$$a_n = \frac{2}{N} \sum_{i=0}^{N-1} f(x_i) T_n(x_i) \tag{4.6}$$

where N is the total number of data points, $f(x_i)$ is the value of the function being fitted at the data point x_i, and $T_n(x_i)$ is the value of the n^{th} Chebyshev polynomial evaluated at x_i. This property simplifies the calculations and makes the convergence of the regression analysis faster, resulting in more efficient parameter estimation.

As a result, the parameters from Equation 4.2 depending on SOC are expressed as 32 dimensional Chebyshev coefficients. These Chebyshev coefficients are used to create a compact representation of the battery's history over single charging/discharging cycle. For representing multiple charging/discharging cycles or an aging procedure, these vectors of coefficients should be concatenated to create a holistic comprehensive representation of the battery's complete history, which allows for an thorough analysis of the battery's behavior, performance degradation, and aging effects.

4.2.2 Model variation for covering over lifetime

Representing SOH changes over multiple charging/discharging cycles through its effects on ECM parameters is challenging. SOH is a specific indicator that quantifies the current condition or health of a system relative to its original or ideal state. It provides a snapshot of the battery's health at a given point in time and it cannot be directly measured but is often inferred based on charging/discharging cycle numbers.

However, relying solely on cycle numbers to infer SOH has limitations, especially when sudden issues and subsequent recovery are involved. Sudden issues, such as rapid voltage drops or battery overheating, might temporarily accelerate degradation in batteries, but the recovery could be partial or nonlinear. Cycle numbers alone may not capture the extent of this recovery. Besides, SOH is a multifaceted and interdependent concept encompassing capacity loss, impedance changes, aging rates, and more. Capturing all these dimensions accurately in a single set of parameters is challenging.

Alternatively, a data-driven approach may be chosen. Machine learning models help to parameterize and represent SOH as multidimensional clusters for quantifying and tracking the health of a battery relative to its initial state. It enables to monitor aging pathway and predict Remaining Useful Life (RUL) of battery based on various scenarios. Nevertheless, obtaining reliable and comprehensive experimental data over a long period for machine learning makes it time-consuming and resource-intensive. Parametrizing a Chebyshev space provides a means to replace experimental data with computationally generated battery models. As more data becomes available during the lifetime of the battery, the numerical data can be replaced by experimental data, and the statistical models for the computational batteries can be consecutively improved.

One Chebyshev coefficients vectors represents a battery's SOC dependent impedance behavior over a single cycle. And SOH is effectively depicted as distinct regions or *clouds* within a broader Chebyshev space. These clouds dynamically shift and evolve as the battery undergoes the aging process. Each cloud within this space signifies a particular state of health the battery can manifest. Visualizing SOH in this manner helps in grasping the continuous and non-linear nature of the aging process. It allows for a more comprehensive understanding of how the battery's health progresses over time, capturing the variability and nuances that come with different aging patterns. Our objective is to derive slight extensions of these coefficient vectors, effectively broadening their representation to cover specific regions of interest throughout the battery's operational lifespan. We aim to encapsulate and encompass crucial phases, behaviors, or characteristics exhibited by the battery during its entire life cycle.

QMC is a method that can be used to generate high-dimensional vectors within a specified interval that is consistent with uncertainty estimates of model parameters by utilizing quasi-random sequences to sample the parameter space in a more uniform and efficient manner compared to standard random sampling. Since Chebyshev coefficients for a single cycle are high dimensional as described above, QMC is particularly suitable.

To generate points in the parameter space, one selects a quasi-random sequence, such as Sobol sequence [119], which have good low-discrepancy properties. The Sobol sequence provides 2^m points in a unit hypercube with d dimensionality $[0,1)^d$, where each point is obtained by bitwise XOR operations on a set of direction numbers. Then transform the generated Sobol points to mimic the normal distribution $N(\mu, \Sigma)$. Therefore, generating a sufficient number of points n through QMC is described as:

$$n = 2^m \ \in \ [0,1)^d \sim N(\mu, \Sigma) \tag{4.7}$$

The foundational set of Chebyshev coefficients is established through a parametrization process, primarily driven by experimental impedance data as elaborated in sub-

section 4.2.1. This process allows for a robust representation of the battery's electrical behavior, encapsulating its impedance characteristics. These d-dimensional n vectors are strategically employed to delineate distinct regions of interest, effectively capturing critical phases within the battery's entire lifespan. By considering the basic vectors corresponding to a *fresh battery*, an *aged battery* and a *dead battery*, we can systematically sample and construct three discernible *clouds*. The *fresh battery* cloud encapsulates the pristine state of the battery, representing its behavior at the initial stage of its life cycle. It provides insights into the battery's optimal performance and behavior when it's in its best condition. In contrast, the *dead battery* cloud represents the end-of-life stage, providing an understanding of the battery's behavior when it nears the end of its operational life. In between, The *aged battery* cloud signifies the battery's behavior as it undergoes the aging process. Sampling and analyzing these three distinct clouds provides a comprehensive understanding of the battery's behavior from its inception to its eventual end-of-life.

4.3 Materials and Methods

4.3.1 Electrical simulation

The generated Chebyshev coefficients vectors at differently aged battery *clouds* are used to construct new ECMs at variety SOCs. These ECMs serve as numerical models that simulate the battery's behavior. The ISEA Framework [17], a real-time capable electrical simulation platform, was used to combine a particular current profile with an ECM to gather a voltage response. The input time-based load profiles are sinusoidal waveforms spanning a wide range of frequencies to mimic the conditions encountered during EIS. These current profile starts the ECM at initialized state, and for each time step, calculate voltage response based on differential and algebraic equations [33] and update the battery's state variables within ECM. These steps are repeated for the specified duration or until the current profile is fully simulated. The calculated voltage at every time step builds a voltage profile that corresponds to the current input. The impedance is then calculated as the ratio of voltage to current in the frequency domain.

4.3.2 EIS measurements for simulation

To fully parametrize a cell for simulation with the ISEA Framework, EIS measurements at multiple SOCs of each cell were acquired. Therefore, the cells were charged with 2C using the CC-CV method until a residual charging current of 0.05C is reached. After a resting period of 10 minutes, an EIS spectra in galvanostatic mode between 200kHz

and 10mHz is measured, followed by a discharge current of 0.1C for 30 minutes to ensure a 5% SOC discharge. Before the next set of EIS measurements, the cell rests for 1h until a steady state is reached. This procedure is repeated until the cell is fully discharged, resulting in 21 SOC points.

To exactly determine the SOC values of each subsequent measurement step, the individual discharge capacities were calculated by integrating the discharge current over time. The sum of all capacity values results in the 0.1C discharge capacity of the investigated cell.

Three cells were taken into consideration:

- a *fresh cell* that has seen no use before the parametrization other than calendaric ageing since production (37mAh).
- an *aged cell* that has been used regularly in real life application and therefore has a lower capacity compared to the fresh cell (25mAh).
- a *dead cell* that has gone through a lot of cyclic and calendaric ageing and shows a significantly lower capacity compared to the fresh cell (16mAh).

4.4 Results and Discussion

Approximation of ECM parameters *vs.* SOC by Chebyshev polynomials

Regression of Chebyshev polynomials to the ECM parameters from fits of experimental impedance data at different SOC values is shown in Figure 4.1. The selection of their best order of polynomials involves Root Mean Squared Error (RMSE) and coefficient of determination, expressed as R^2.

As pointwise parametrized results show, the serial resistance R_s decreases as SOC increases, it implies battery is discharging [120]. This phenomenon is consistent with experimental results [121]. The evaluation results quantifying their relationship with different Chebyshev polynomial degrees illustrate that RMSE decreases and the R^2 increases as the model complexity increases. The order that results in the lowest RMSE and the highest R^2 is considered the best choice. However, our goal is to minimize the dimensionality, in other words, to represent the battery models with as few coefficients as possible. At the point of polynomial with one degree RMSE and R^2 starts to stabilize, therefore, their relationship can be assumed and simplified to a linear one.

R_1, τ_1 and ϕ_1 represent the low frequency ZARC element, since τ represents the inverse frequency and τ_1 is the highest one among three elements. As observed the

increasing parameter values R_1 and τ_1 with increasing SOC exhibit a approximately linear behavior, where both RMSE and R^2 slightly reach an optimal value. The time constant $\tau = RC$ associated with resistance builds up charge during charging and has increased resistance, which may intensify the formation and growth of the Solid Electrolyte Interface (SEI) layer.

In contract to it, R_3 and τ_3, which make up the high frequency ZARC element, their values decrease as SOC increase. It builds up discharge during charging, which implies cathode. Their RMSE and R^2 plot against the order of polynomial suggest a linear regression function.

The another distinguishable ZARC element in the middle frequency has more fluctuating values along with SOC and they needs higher degree polynomial functions for mimic the dependencies. Obviously R_2 can be described by 2-polynomial function, but RMSE and R^2 plot of τ_2 don't show clear stabilization within expected polynomial degree. Here we choose the point where both lines meet, 3 degree of polynomials, while it represents a balance between fitting the data and model complexity.

ϕ values strongly fluctuate and their polynomial function normally has also higher degree. Based on the strategy of choosing either stabilizing or intersecting points, ϕ_1 selects 2 polynomials, ϕ_2 3 polynomials, and ϕ_2 1 polynomial respectively. The Warburg coefficient σ and limit of capacitance C_{lim} for diffusion process represent the straight line in the low frequency in the Nyquist plot. σ decreases as SOC increases, but conversely, $Clim$ increases as it increases, and both can be expressed by second-degree polynomial functions.

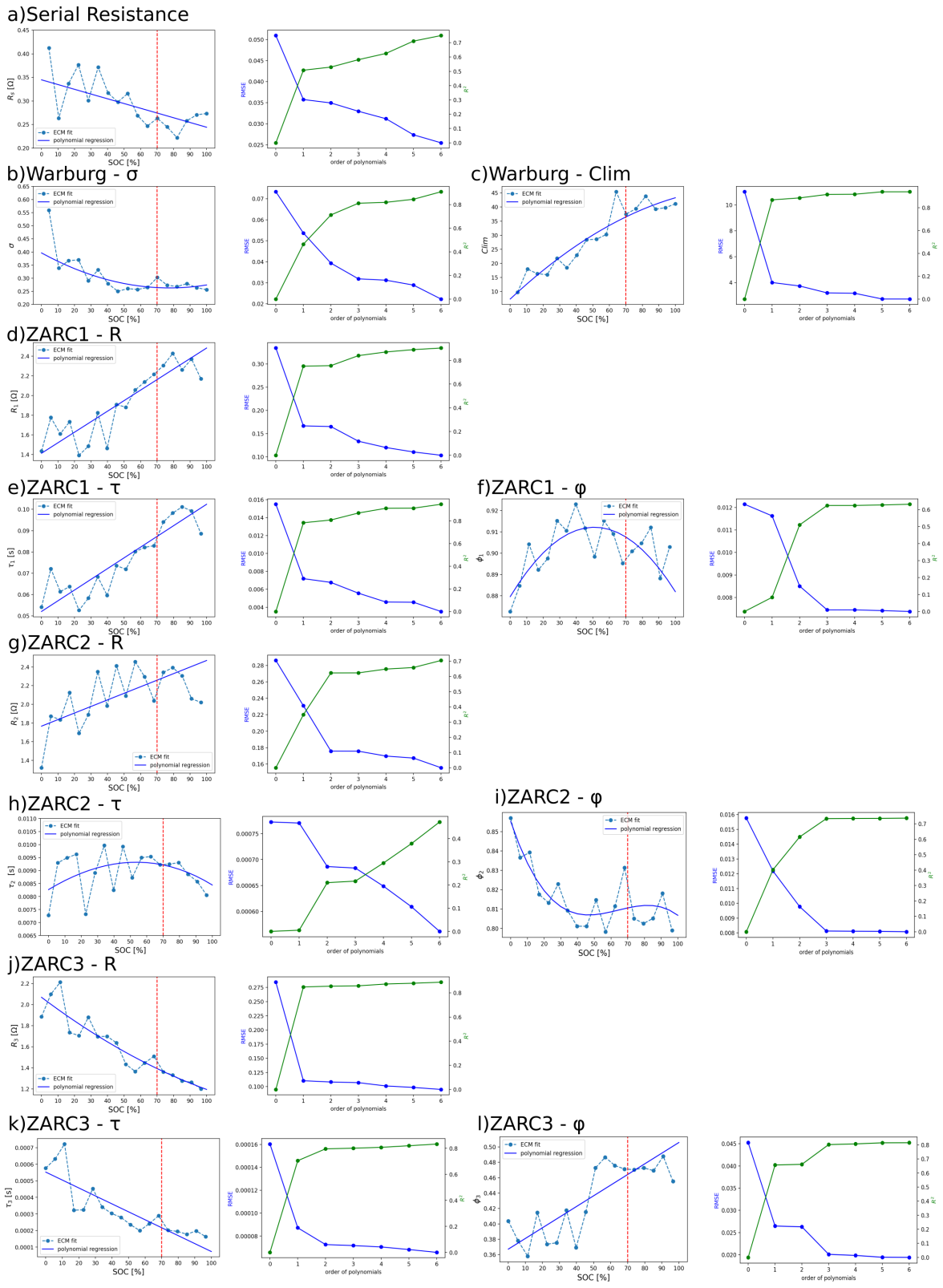


Figure 4.1: Chebyshev polynomial regression of ECM parameter values along with SOC (left) and their RMSE and R^2 of different degrees of Chebyshev polynomial regression (right)

These Chebyshev polynomials can replace lookup tables to model the dependencies of ECM parameters and SOC in a more compact and mathematically efficient manner. To evaluate the polynomials, plug the specific SOC value at 70% into the defined Chebyshev polynomials and regular polynomial function to obtain the corresponding parameter values and calculate new impedance according to Equation 4.2. As a representative middle point of SOC range, the charging and discharging effects on the battery are relatively balanced and 70% SOC is relevant to practical usage for optimal performance. In Figure 4.2, the mathematically calculated and experimentally observed EIS spectrum, and their residuals are shown.

The chosen order of the polynomial for a parameter is limited to maximum of 4, signifying a highly regularized model. Consequently, the effects and differences between two polynomials are not substantial. Although both polynomial fitting methods provided good visual fits on the Nyquist plot, the Chebyshev polynomial regression demonstrated better performance in terms of residual distribution and fit accuracy. The more homogeneous distribution of residuals and the smaller RMSE across the entire frequency range indicate a higher level of precision and accuracy in capturing the impedance behavior of the system. Thereby Chebyshev polynomial coefficients are capable and reliable to replace traditional ECM parameters representation along with SOC. However, the extremely high and low frequencies may not have practical significance in real-world applications. Using RMSE to average values across the entire frequency spectrum might not be a suitable measure, as it could obscure the meaningful insights in the more relevant frequency ranges. This analysis is selected from individual spectrum at a certain SOC, and the subsequent section will delve into a comparative analysis of multiple spectra.

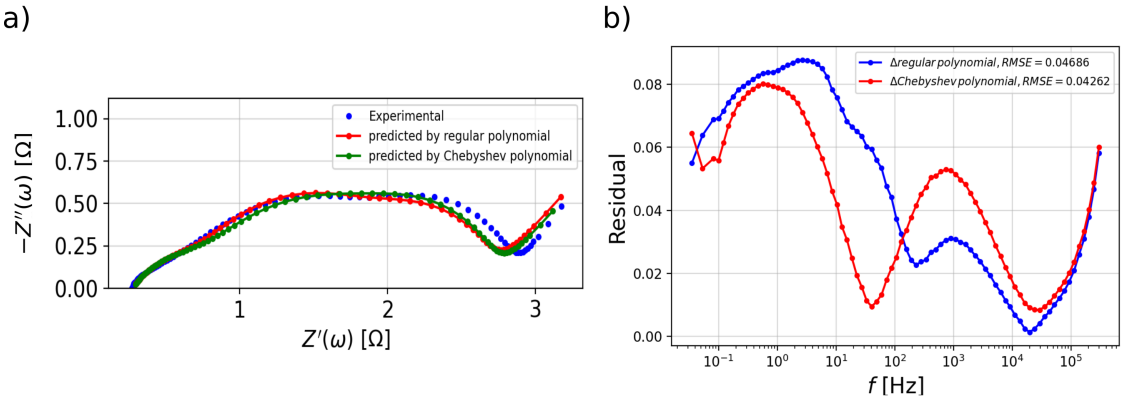


Figure 4.2: The differences between the observed impedance from EIS measurement and predicted impedance from regular and Chebyshev polynomial regression model at SOC=70% (a) and their residuals (b)

When aiming for a global EIS fit of the experimental data from a single cell at fixed age for all measured SOCs, instead of conducting individual fits for each SOC separately, securing an appropriate starting set of Chebyshev coefficients for this non-linear fitting might be crucial. For this, the Chebyshev coefficients taken from pointwise approach that describes in this context can provide a solid foundation for initial conditions and reasonable boundaries. The optimized Chebyshev coefficients directly fitted from experimental EIS data are detailed in the Supporting Information (section 4.6).

Model variation over lifetime

The determined Chebyshev coefficients above are based on ECM fitting on measured impedance at a single cycle. The total number of coefficients to represent an ECM is 32, which is hard to visualize for such high dimensional vectors. Here we take the parameter ϕ_1 as an example, which is quantified by second order polynomial function with three coefficients and is able to visualize in Figure 4.3.

In this study, three fundamental sets of coefficients, derived from experimental data analysis and denoted as fresh cell, aged cell, and dead cell, serve as key reference points in a 3-dimensional space, which are plotted as star points. The generation of coefficients based on these fundamental points is carried out using QMC Equation 4.7 and should take care of boundaries. A simple concept is to create circles around each fundamental point, the radius is set to half the distance between any two neighboring fundamental points. Thousands of points are generated within respective circles and this process is applied independently in each dimension.

The aged cell group and fresh cell group are located nearly, while dead cell group is far away from them. It indicates that the healthy batteries and unhealthy batteries are significant noticeable in the Chebyshev space. The generated coefficients serve as the basis for defining parameter values. These coefficients are utilized to construct polynomial functions along the SOC. The new polynomial functions should align with the fundamental Chebyshev polynomial function's behavior. The fundamental Chebyshev polynomial function of each group serves as a reference for the desired tendencies in the generated polynomial functions. If the generated polynomial functions do not exhibit the desired tendencies akin to the fundamental Chebyshev polynomial function, the corresponding points are discarded. Additionally, ϕ_1 of dead cells increases dramatically with increasing SOC, while its change in healthy cells is not significant. At higher SOC, the three distinct clouds become heavily intertwined, while at lower SOC, they remain well separated. This observation suggests that, even when examining a single parameter, batteries can be effectively grouped based on their lifetime characteristics. Notably, within the fresh and aged cell clouds, there is a outlier present even at low

SOC levels, thus it becomes essential to observe the speed at which outliers are generated as the standard deviation of QMC generation increases. This sensitivity to the standard deviation highlights the need for robust analysis and management of battery behavior, especially in cases where outliers play a significant role.

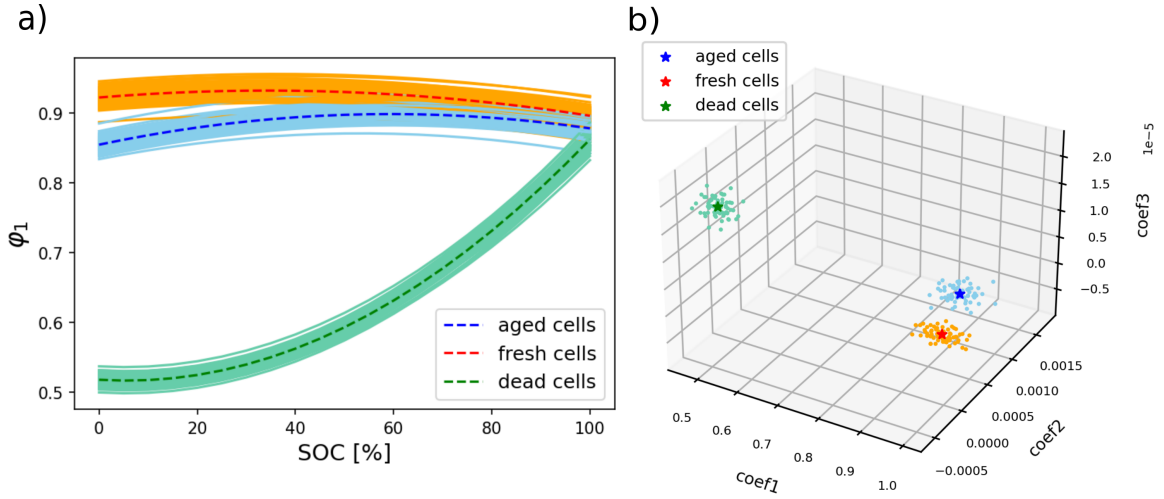


Figure 4.3: Generated parameter ϕ_1 along with SOC (a) and its coefficients for aged, fresh and dead cells (b)

Similarly, applying QMC generation methods to the rest of parameters allows one to build high dimensional vectors to represent differently aged cells and store big data for further machine learning process (Figure 4.4). Notably, the impedance generated using Chebyshev polynomials exhibits a distribution that closely aligns with the experimental data, outperforming regular polynomials. As a result, the overall mean value of Chebyshev polynomials is smaller than that of regular polynomials. Similarly, when comparing both polynomial types at different SOC values, Chebyshev polynomials exhibit a smoother transition in response to changes in SOC. The sensitivity of impedance variation to the standard deviation of QMC generation highlights the speed of cloud expansion. As the variance of Gaussian distribution increases, the variance of impedance with standard polynomials accelerates more rapidly than with Chebyshev polynomials. In practical terms, this means that Chebyshev polynomials introduce outliers more slowly within cloud boundaries, thereby causing clouds to expand gradually and ensuring that the distinct clouds do not mix together. This behavior contributes to ensuring that the scaling up of the clouds remains under control and occurs gradually, with a measured pace.

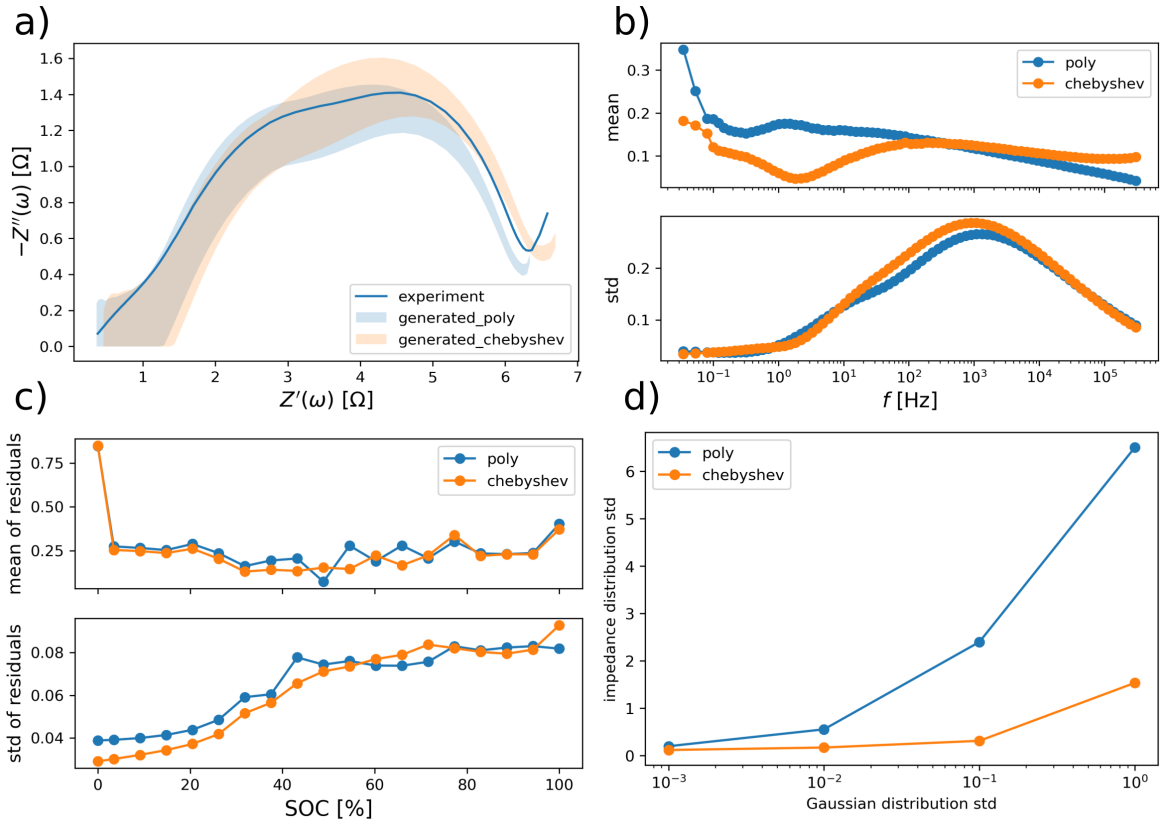


Figure 4.4: The generated aged cell models based on Chebyshev polynomials and regular polynomials. (a) shows the impedance calculated from the generated parameters of the aged cell cloud at a single SOC; (b) illustrates the mean and standard deviation of residuals when compared to the experimental data; (c) shows residuals over multiple SOCs; (d) is impedance variance along with standard deviation of QMC generation over multiple SOCs

4.5 Conclusions

In order to reduce the dimensionality of model representation at a single charging/discharging cycle, Chebyshev polynomials quantify the SOC dependencies on ECM parameters in a robust way, which allows to reconstruct the ECM at any desired SOC value in a concise manner. Each charging/discharging cycle can be represented by a certain set of Chebyshev coefficients that correspond to the fitted parameters at different SOC values during that cycle. These coefficients form the battery's history over that specific cycle.

Since parametrizing SOH is a complex task due to the multifaceted nature of battery health, it makes difficult to use multiple charging/discharging cycle numbers to represent the aging process comprehensively. Aging behavior can only be tracked by segmentation of different battery health categories. QMC method helps to generate

sufficient sets of high dimensional Chebyshev coefficients vectors based on experimental data to build certain region of battery health without massive experiments. These sampled large amount of numerical models work as foundation for further machine learning, which allows for tailored approaches to mitigate or manage specific degradation patterns. The data delivered in this article includes Chebyshev coefficients vectors of different battery health categories, and their corresponding simulated time-based current and voltage data, as well as their analyzed impedance and DRT data.

4.6 Supporting Information

The process initiates with a set of 32 Chebyshev coefficients derived from a pointwise approach as described in this context, which serve as the initial variables for optimization. The object function employs this set of coefficients to determine the ECM parameters across full SOCs. For each ECM parameter set at each SOC, impedance is calculated over a wide range of frequencies. These calculated impedances are then compared against the experimental impedance data, with the aim of minimizing the residuals between them. Following this, an iterative Bayesian optimization process takes the initial guess for the Chebyshev coefficients and the fitting boundaries, applying the predefined objective function to find the minimal residuals. It intelligently selects the next set of parameters for evaluation, informed by prior iterations. The goal is to identify the optimal set of Chebyshev coefficients that leads to a more accurate fit to the experimental EIS data. Table 4.1 presents an example of the optimized Chebyshev coefficients obtained with a good initial set side by side with those obtained without such initialization. The parameter $Rser_0$ and $Rser_1$ in the table are the coefficients of serial resistor R_s to represent their SOC dependencies by 1-polynomial regression. Similarly, the parameters are related to polynomials of element in the table.

element	parameter	initial	optimized	without good initialization
R_s	$Rser_0$	0.295	0.310	13.979
	$Rser_1$	-9.363e-5	-8.895e-5	-1.724
$Clim$	WC_0	8.053	8.456	-19254.8
	WC_1	0.077	0.081	-233.2
	WC_2	9.425e-5	9.897e-5	-0.703
σ	WS_0	0.049	0.047	-0.609
	WS_1	8.551e-4	8.978e-4	0.034
	WS_2	1.243e-5	1.181e-5	4.852e-4
R_1	$Z1R_0$	1.414	1.484	-1.657
	$Z1R_1$	0.011	0.010	0.142
τ_1	$Z1T_0$	0.052	0.055	-0.113
	$Z1T_1$	5.029e-4	4.778e-4	0.0329
ϕ_1	$Z1P_0$	0.879	0.835	0.93
	$Z1P_1$	0.00128	0.00122	0.0173
	$Z1P_2$	-6.285e-6	-6.599e-6	1.251e-3
R_2	$Z2R_0$	1.766	1.854	-7.786
	$Z2R_1$	0.00704	0.00739	-0.244
τ_2	$Z2T_0$	0.00827	0.00867	-0.12
	$Z2T_1$	3.801e-5	3.611e-5	0.011
	$Z2T_2$	-1.596e-4	-1.676e-4	-1.7e-4
	$Z2T_3$	-1.078e-10	-1.049e-10	7.645e-7
ϕ_2	$Z2P_0$	0.855	0.898	-0.388
	$Z2P_1$	-0.00255	-0.00243	0.0289
	$Z2P_2$	2.152e-5	2.259e-5	2.158e-3
	$Z2P_3$	-5.576e-8	-5.297e-8	-1.121e-5
R_3	$Z3R_0$	2.069	2.141	-5.257
	$Z3R_1$	-0.012	-0.011	0.755
	$Z3R_2$	1.532e-5	1.061e-5	4.818e-3
τ_3	$Z3T_0$	5.526e-4	5.802e-4	7.551e-4
	$Z3T_1$	-4.794e-6	-4.554e-6	-7.112e-6
ϕ_2	$Z3P_0$	0.367	0.349	0.575
	$Z3P_1$	-0.00138	-0.00145	-0.0559

Table 4.1: Table to 32 Chebyshev coefficients of 12 ECM parameters fitted to experimental impedance data by initial pointwise approach, iterative global optimization, and fitting without good initialization. The SOC is sampled at 19 discrete points within the range of 0-100%. For each SOC, the impedance is measured across a frequency spectrum consisting of 84 points, spanning from 10 mHz to 300 kHz.

Chapter 5

Observation of battery aging process in latent space

This chapter is adapted from the publication that has been submitted to the journal and is now under review: **Lifetime monitoring for lithium-ion battery in the low dimensional latent space**, by L. Jin, F. Bereck, R.-A. Eichel, J. Granwehr, and C. Scheurer.

The theoretical analysis about design of autoencoder, classification and validation of battery aging state in latent space, as well as simulation and training of battery data were conducted by the author of this thesis, Limei Jin. The experimental EIS data was provided by Franz Bereck.

5.1 Introduction

Batteries in electric vehicle contain intricate battery packs composed of numerous cells working in tandem. The BMS is responsible for effective management of these cells, ensuring optimal performance and overall lifetime [4]. However, achieving precise control over the complex electrochemical processes occurring within each cell remains a challenge. Therefore, the integration of a DT into cell management practices empowers the BMS with advanced capabilities for monitoring, analysis, and control [113].

A DT is a virtual representation of the physical battery system, and the data obtained from EIS can be integrated into this digital model, ensuring that the virtual model stays synchronized with the actual behavior of the battery cell and updated in real time [8]. The EIS stands out as a spectroscopic technique capable of providing insights beyond conventional measures of potential and current. However, EIS assumes a linear and time-invariant system, which is often not true in real-world applications, and EIS provides a snapshot of the system's behavior at a specific point in time,

usually represented as SOC [58]. It cannot easily capture the system's non-periodic dynamics or changes over time, such as SOH, which can be crucial for understanding the behavior of a system in real-world, time-varying conditions (chapter 4). Besides, obtaining high-quality, comprehensive experimental impedance data from batteries is time-consuming.

Addressing these challenges, the frequency domain analysis will be transitioned to time domain. Initially, the time-based data analysis is applied on ideal sinusoidal current and voltage signals for better comparison with traditional EIS analysis. However, in real-world scenarios, battery behavior can vary significantly due to factors such as temperature fluctuations [122], electrode degradation [21] and other environmental conditions. It results in an idealized sinusoidal waveform that is no longer realistic. Hence, introducing controlled noise [77], such as Gaussian noise, into simulation signals enhances realism, making them more representative of the complexities in real-world battery applications. Additionally, real-world data collection, especially with comprehensive measurements, can be time-consuming and resource-intensive [123]. Hence, testing the model's ability to generalize and provide reliable outcomes with limited input information is necessary. Partitioning each noisy sinusoidal signal into subsets and scrambling them among various frequencies contributes to reduce the amount of data required for analysis, as long as their resulting signals, that are independent on frequency order, give the same results as idealized signals. It allows researchers to investigate how little data with how many frequency points is sufficient for objective analysis while preserving essential characteristics. It also helps avoid bias introduced by analyzing only idealized sinusoidal signals and introduces variability for capturing a broader range of potential scenarios, such as stochastic pulse signals or other fluctuated signals. Thereby, model's robustness and adaptability can be tested with fewer data points.

However, time-based battery data analysis involves large amount of high-dimensional data, thus, ML emerges as a transformative approach, offering potentials to leverage large datasets and extract meaningful features from these complex data in a non-linear way, improving the precision and efficiency of battery monitoring [7]. Especially, dimensionality reduction techniques like autoencoders are used to identify the most relevant features and focus on the most informative aspects of the data.

Autoencoders are a deep neural network architecture commonly used in unsupervised learning to perform dimensionality reduction and feature extraction from input data [34]. They typically consist of an encoder and a decoder. The encoder takes the normalized data as input and compresses it into a lower-dimensional latent representation, named Latent Space. The decoder then aims to reconstruct the original data

from the latent space representation. The objective of training the autoencoder is to minimize a reconstruction loss between the input data and its reconstruction.

Analyzing battery data in the LS provides the chance to observe aging process more easily. By mapping different battery ages onto the latent space, it becomes possible to visualize the progression from fresh to aged and potentially end-of-life states. The position of data points in the latent space reflects the underlying dynamics of aging, providing insights into how batteries evolve over time. Furthermore, the low-dimensional LS enables efficient classification of battery states. Through techniques like SVM [42], distinct regions in the latent space can be associated with different battery ages.

In this article, we will explore the transition from linear to non-linear regimes, aiming to extend the limits of traditional EIS analysis for battery system. By mapping different types of battery data into LS, derived through autoencoders, we will delve into capturing the evolving characteristics of batteries over time, paving the way for live aging tracking.

5.2 Methodology

5.2.1 Structure of autoencoder

In our approach, we exclusively employ the encoder portion of the autoencoder while omitting the decoder, as our primary focus lies on the observation of battery aging within the latent space. The activation function chosen for our neural network is ReLU for its simplicity and effectiveness, and CNN layers are utilized due to our input data comprising multiple channels [124]. Opting for a latent space dimensionality of two is sufficient for categorization and convenient for visualization. Although alternative scenarios might necessitate higher dimensions, our specific case for tracking battery aging benefits from the simplicity and interpretability offered by a two-dimensional space. The foundational library of the autoencoder provides the backbone for implementing the selected algorithms in our analysis.

5.2.2 Mapping data to autoencoder input/output layers

In this study, we work with time-based and frequency-based input data over a single cycle, featuring 18 SOC points, at each 5% over discharging cycle. The frequency-based input comprises real and imaginary parts of impedance, each with 78 values at every SOC point, mirroring the frequency points taken in the EIS measurement. Additionally, an SOC sequence with 18 measuring points is included, as illustrated in

Figure 5.2a. Similarly, the time-based data consists of a current sequence with 1000 time steps, a corresponding voltage sequence with the same time steps, and an SOC sequence aligning with the frequency-based analysis, showcased in Figure 5.2b and Figure 5.2c.

Given the sensitivity of neural networks to input scale variations, we employ min-max scaling normalization for each sequence [125]. This process transforms each sequence into a standardized range, typically between 0 and 1. It ensures uniformity across different input scales, preventing any single sequence from dominating the network and promoting stability during the training process.

The normalized three-channel input sequence undergoes a convolutional transformation in the encoder, designed with multiple layers. The kernel size, stride, and padding parameters control the convolutional operations and downsampling. The initial convolutional layer applies operation with a kernel size of 3x3x3, maintaining the spatial dimensions through padding and a stride of 1, and transforms into an output tensor with 16 channels. Then, a ReLU activation function follows to introduce non-linearity to the transformation. The subsequent layer with 16 channels employs a similar convolutional operation with a 3x3x3 kernel but introduces spatial downsampling through a stride of 2, effectively reducing the spatial dimensions. Again, a ReLU activation function is followed afterwards. The last convolutional layer refines the representation to 2 channels with another 3x3x3 kernel and a ReLU activation. The resulting tensor is then flattened, transforming it from a 3D tensor to a 1D tensor. Three linear layers follows to further reduce the dimensionality of the representation through ReLU activations. The final linear layer outputs the compressed representation in the latent space with a size two. Thereby, each three-sequence input data set over a single cycle compressed as one point in the 2D latent space. When applying the same structure of input data over multiple cycles, multiple three-sequence input data sets are condensed independently, leading to the emergence of different points in the latent space (Figure 5.1).

5.2.3 Classification

Each point in the latent space serves as a representation of cells with varying aging conditions. However, the individual values of vectors in the LS might not have an explicit or interpretable meaning on their own. Instead, the crucial information lies in the relative positions and relationships between them, especially in the context of neighboring points. The structure and distribution of vectors in the LS, as well as their proximity or separation, play a vital role in how the model captures and represents patterns and information from the data. Hence, a classification is attempted regarding

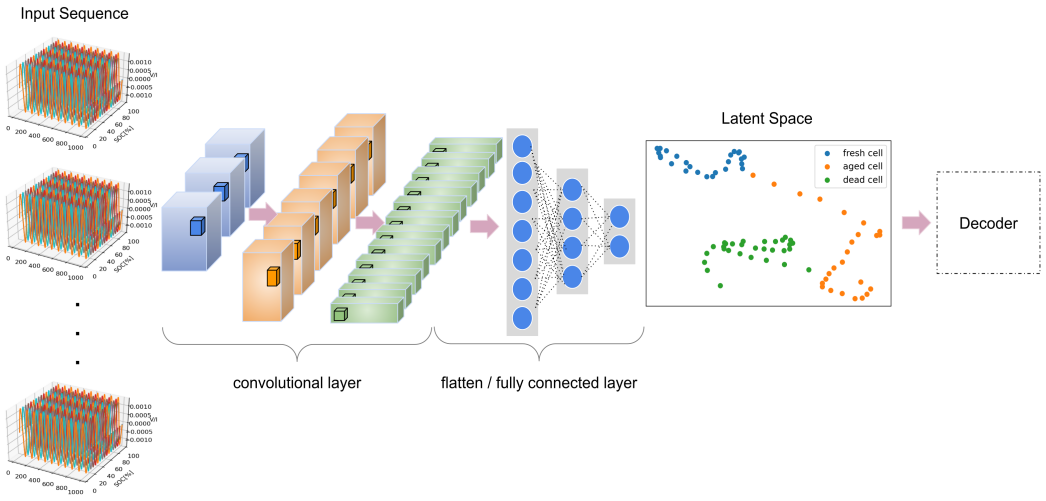


Figure 5.1: Structure of an Autoencoder for training ideal sinusoidal time series data for differently aged cells

to the SOH of battery cells.

SVM is a supervised machine learning algorithm, which aims to find an optimal hyperplane that maximizes the margin between different clusters while minimizing classification error [42]. The convex optimization problem in SVM guaranteed the global optimum, which is not sensitive to the initial guess or starting point.

As a non-parametric model, it makes fewer explicit assumptions about the underlying data distribution or the number of features compared to other models. In other words, unlike some clustering methods that might create circular decision boundaries around clusters, SVM seeks the hyperplane, that is not constrained by a predefined shape. Therefore, SVM is particularly suitable for tracking aging pathways in the latent space. It segregates entire space into three age regions, if necessary can be more. By observing the distribution of these regions, one can gain a comprehensive understanding of how cells age. Additionally, SVM prioritizes a subset of data points, known as support vectors that lie closest to the decision boundary and influence the position and orientation of the hyperplane, therefore it allows a robust decision-making even with a small dataset.

SVM handels both linear and non-linear kernels. The parameter C helps to control the error of misclassifying in SVM with linear kernel function, while γ parameter determines the classification results of non-linear kernel function, polynomial and RBF [46].

The high values of C and γ mean higher penalty and better accuracy, the optimizer will choose a more complex decision boundary that fits all the training data very closely, and potentially leading to over-fitting because of a low bias and high variance [45].

Conversely, a low value of C and γ will cause robust boundaries with mis-classification, which results in a high bias and low variance for under-fitting problem.

5.2.4 Cross-validation

To find optimal hyperparamter of a classifier, Cross-Validation (CV) technique is involved to estimate the accuracy of a model along with hyperparamter variation [126].

The simple validation process used in machine learning is setting another test set for the trained model to evaluate its accuracy. The training score reflects how well the model predicts the labels of the data it was trained on. However, this alone does not guarantee the model's ability to generalize to new, unseen data.

In a 3-fold CV, the training set is partitioned into three non-overlapping subsets, and the model is trained and validated three times. Each time using a different subset as the test set and the remaining two subsets as train sets. The average accuracy over the three folds is reported as the CV-score. It provides an overall evaluation and an indication of how well the model performs on data it hasn't seen during training.

Combination of training scores and cross-validation scores ensure that SVM model fits without over-fitting nor under-fitting.

5.2.5 Scrambling

We take experimental EIS data during a discharging cycle on three distinct battery cells with varying ages: a fresh cell with a capacity of 37 mAh, an aged cell with 25 mAh, and a damaged cell with 16 mAh. ECM fitting combined with DRT techniques allows us to parameterize virtual models that represent three real battery cells. Then the use of Chebyshev polynomials enables to quantify of SOC dependencies on ECM parameters, and these Chebyshev coefficients form the battery's history over a specific cycle. Subsequently, QMC method is applied to generate sufficient sets of high dimensional Chebyshev coefficients vectors. These sets, anchored around the three experimental data points, define specific regions of battery health (chapter 4). These sampled large amount of numerical models are simulated with ideal sinusoidal current profile and their corresponding voltage output is stored for further ML.

We start with these prepared ideal sinusoidal voltage and current data (Figure 5.2b) ranging from high- to low-frequencies. The frequency range precisely aligns with the frequency components encountered in EIS measurements, to ensure that the simulated data maintains relevance and comparability to experimental impedance data. For more representative real-world scenarios, we introduce Gaussian noise to each of the sinusoidal voltage data with different frequencies (chapter 3).

To investigate how many individual oscillations are needed for the same LS topology as traditional periodic signals, scrambling signals independently of frequency order is a necessary step. It reduces data requirements, which is valuable for testing the model's ability to generalize with limited input. Each noisy sinusoidal signal is meticulously partitioned into several subsets or segments. The number of subsets precisely corresponds to the number of frequency points, ensuring that each recombined signal covers all frequency range. The subsequent step involves taking the subsets from different frequency signals and rearranging or scrambling them among various frequency components. For example, subsets from a high-frequency signal may be placed within low-frequency signals, and vice versa. As a result, the new signals now contain components from randomly different frequencies (Figure 5.2c).

5.3 Materials and Methods

The open-source machine learning library *PyTorch* is used to construct convolutional autoencoder architecture [127]. An real-time capable electrical simulation platform, *ISEA Framework* is utilized to simulate EC models with current input and gather voltage output [33].

5.4 Results and Discussion

Latent Space: time-based data vs. frequency-based data

Battery data has been subjected to training using an autoencoder, and the subsequent analysis is conducted in a two-dimensional latent space for simplicity. Higher dimensional latent space may be required for more complex system responses. Each point within this latent space represents a distinct battery model corresponding to a specific cycle number, in other words a certain battery age.

While traditional impedance analysis is well-established and widely used, comparing it with time domain data analysis ensures that the latter method is reliable and capable of capturing dynamic battery performance. Comparing the LS results obtained from ideal periodic signals with it derived from EIS data provides a means to validate the linearity assumption. To account for non-linear battery's behaviour, the LS representation from scrambled data is validated, aiming to achieve results consistent with a linear system, thus providing a secure foundation for extending the linear limits.

Figure 5.3 shows the latent space representation gained from training by ideal sinusoidal current and voltage data (Figure 5.2b), by scrambled noisy time series data

(Figure 5.2c), and by their impedance data (Figure 5.2a). The blue points have been generated using QMC methods for understanding aging behavior, and the three star points are actual data obtained from experiments as a basis for comparison and validation. The measurement on a battery with capacity 37 mAh serves as fresh cell, with 25 mAh as an aged cell model, and with 16 mAh as defective cell.

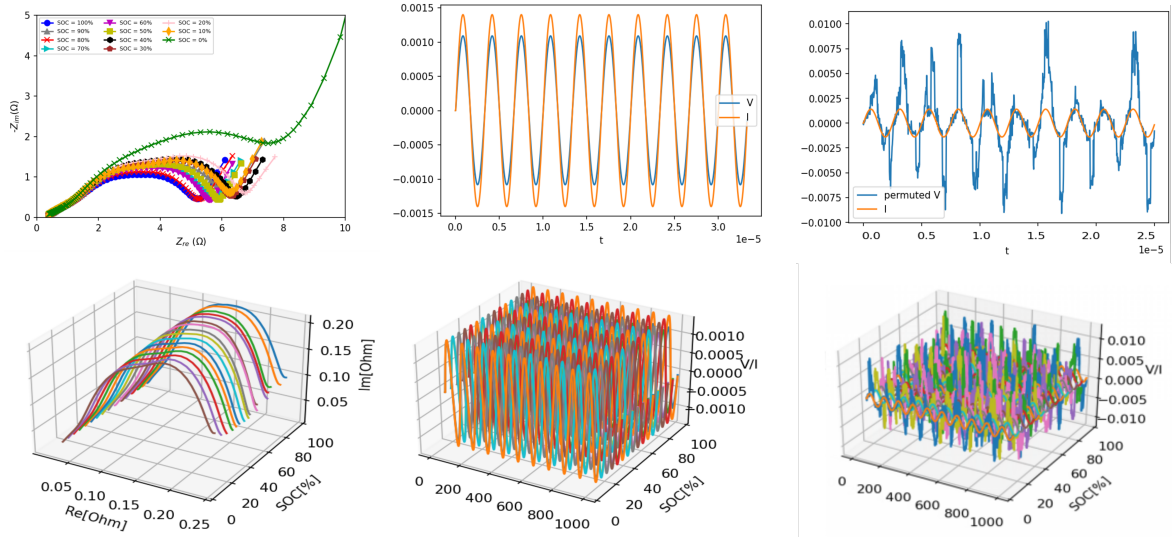
As observed, the topology of the latent space appears similar in shape for all training data sets, from blue star point through green point to red point can form a directed line. The similarity between Figure 5.3a and Figure 5.3b implies that analysis of ideal time-based data can capture key information and features similar to the frequency-based impedance data analysis under linear assumption and the same conditions. Comparing Figure 5.3c with the others indicate that frequency independent noisy signals still give us the similar results as linear system. Therefore, it allows a reliable dynamic battery performance analysis by using traditional impedance analysis as reference or benchmark.

However, the latent space obtained from time-series data appears to be rotated in relation to the latent space from impedance data. Such a rotation in latent space does not indicate a problem especially in the context of unsupervised learning, rather is a characteristic of the dimensionality reduction process. The non-linear transformation of the autoencoder can result in a change in the reorientation and redistribution of the data points while the essential features and patterns in the data are preserved.

The positions of the three experimental datasets in the latent space, representing distinct battery ages, effectively imprint the motion of aging onto the latent space topology. The color progression from blue to green to red corresponds to the transition from a fresh battery to an end-of-life battery, and these aging pathways consistently manifest across the three different types of training data (Figure 5.3). Visually, the blue point representing the fresh cell, the green point denoting the aged cell, and the red point indicating the dead cell are aligned along a continuum, symbolizing the chronological progression of battery life. Despite the minimal number of experimental samples, the concept of associating the SOH with the latent space topology is effectively realized, showcasing the its commitment to data economy and resource efficiency. This imprinting process establishes a clear correlation between the observed aging patterns and the latent space, affirming the autoencoder's capability to capture essential features even with a sparse dataset.

Classification: permuted noisy data vs. ideal periodic data

The classification of battery age in the latent space is implemented using SVM with a linear kernel and a non-linear RBF kernel. The optimization of SVM performance

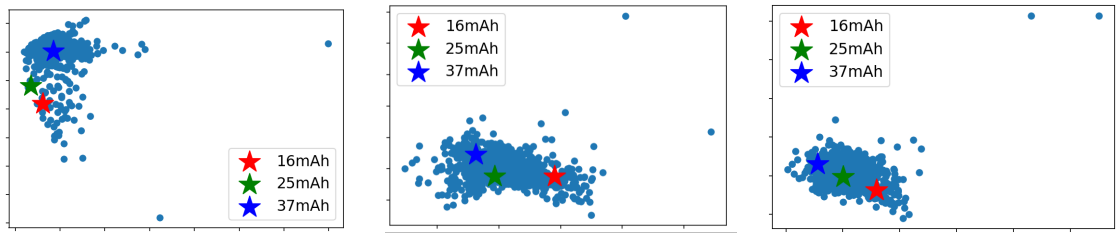


(a) frequency-based impedance data

(b) ideal sinusoidal current and voltage data

(c) scrambled noisy voltage and current data

Figure 5.2: Different types of battery training data over a single cycle



(a) training by impedance data

(b) training by periodic time series data

(c) training by scrambled time series data

Figure 5.3: Latent space of generated uncategorized battery cells

involves the careful selection of the kernel type and their associated coefficients, a crucial step in ensuring the accuracy and reliability of the classification results.

When evaluating the linear kernel, the left validation curves in Figure 5.4a and Figure 5.4b are presented to illustrate training scores and cross-validation scores as a function of the linear kernel parameter C. As C values increase, both training and validation scores rise and eventually stabilize. The optimal balance between model bias and variance is often achieved when both scores reach a plateau. This plateau signifies a point where further increasing the complexity of the model does not lead to substantial improvements in performance and may even result in overfitting. Accordingly, C values of 10^6 and 10^7 are chosen to define the coefficient for the linear kernel.

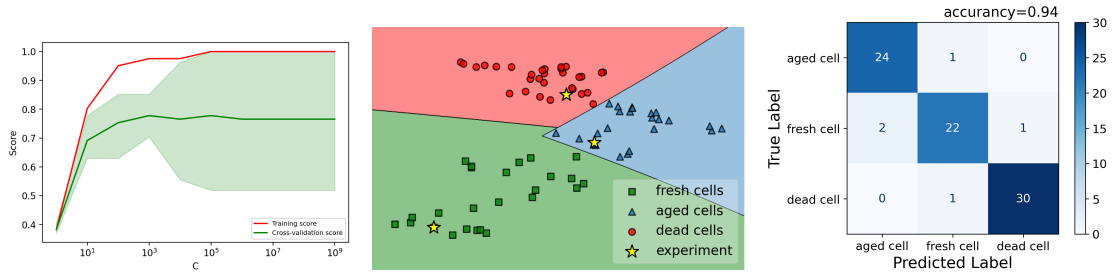
Furthermore, left validation curves in Figure 5.4c and Figure 5.4d delve into coefficient γ for RBF kernel. The analysis reveals that very low γ values result in both training and validation scores being low, termed underfitting, whereas medium γ val-

ues lead to high scores, signifying effective classifier performance. On the other hand, excessively high γ values indicate overfitting, with strong training scores but poor validation scores. Consequently, γ values of 10^2 and 10^8 are selected to define γ for the RBF kernel function for ideal and permuted signals, respectively. This rigorous parameter selection process ensures that the SVM classifier is finely tuned and proficient in accurately classifying battery age within the latent space.

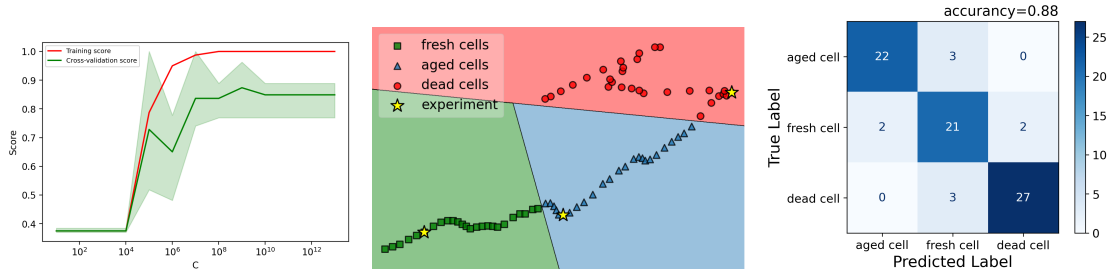
As observed, there are substantial differences in the selected γ values for the two datasets, while two C values selected are similar. This discrepancy arises from the scaling differences in the two latent spaces derived from distinct input data. It's worth noting that a linear boundary is infinite and remains unaffected by scaling, while nonlinear boundaries are indeed influenced by scaling factors.

In the middle classifier plots in Figure 5.4, the classifiers effectively define classification boundaries in the two-dimensional latent space. It signifies that the two-dimensional features derived from the latent space are sufficiently informative for the segmentation of fresh, aged, and damaged battery cells. Notably, the latent spaces resulting from the training of ideal, noiseless time-series signals (Figure 5.4a and Figure 5.4c) and those obtained from training with permuted noisy signals (Figure 5.4b and Figure 5.4d) yield similar classification results. In both cases, the trajectory of battery aging pathways is observed to transition from the bottom-left blue region, through the center-right orange region, and eventually to the upper-center green region. It suggests that the reduction in frequency range, as employed in the further analysis of stochastic pulse signals, does not have a detrimental effect on the classification results. In essence, the classifiers exhibit consistent and reliable performance, validating the utility of the chosen latent space and data processing techniques in battery age classification.

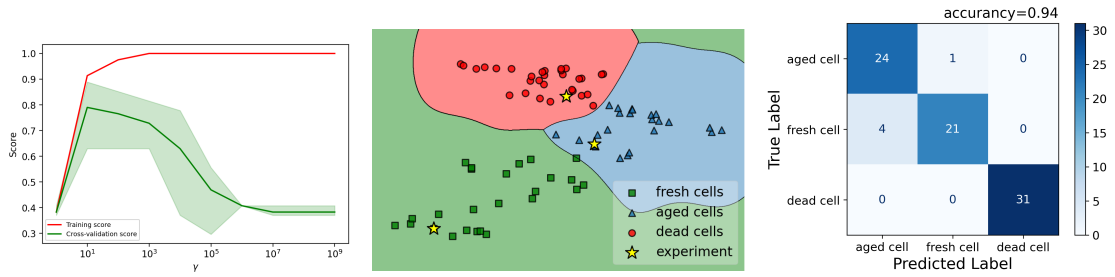
The trained classifier is put to the test using a new battery dataset to assess its accuracy, and the results are presented in the left confusion matrix in Figure 5.4. In this matrix, the diagonal elements represent the number of data points for which the predicted label matches the true label, and they exhibit the highest values. These diagonal elements reflect correct predictions by the classifier. The off-diagonal elements indicate misclassified points, and they have the maximum value of 4 in this case. A higher concentration of diagonal values within the confusion matrix is indicative of more accurate predictions. The accuracy, which quantifies the overall performance of the classifier, reveals that the classifier employing the RBF kernel function performs slightly better than the one using a linear kernel. Nevertheless, the differences in accuracy between the two classifiers are relatively small, underlining their overall effectiveness in accurately categorizing battery age.



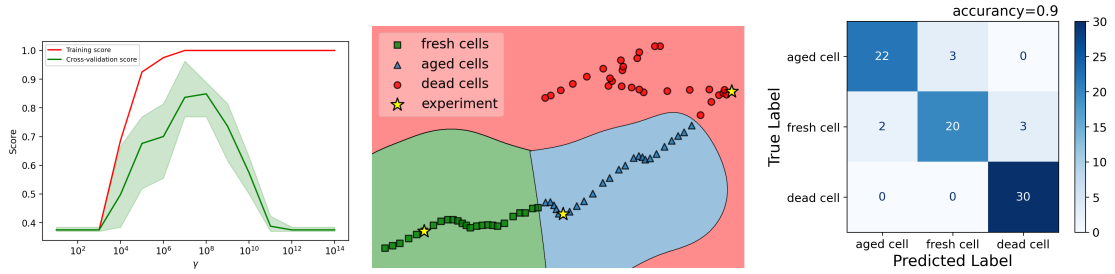
(a) linear classification for ideal signal



(b) linear classification for permuted signal



(c) RBF classification for ideal signal



(d) RBF classification for permuted signal

Figure 5.4: Classification of batteries into three aging groups within the latent space.(Left) SVM validation curve along with hyperparameter variation; (Middle) Segmentation visualization of three different battery's ages; (Right) Test results of trained classifiers

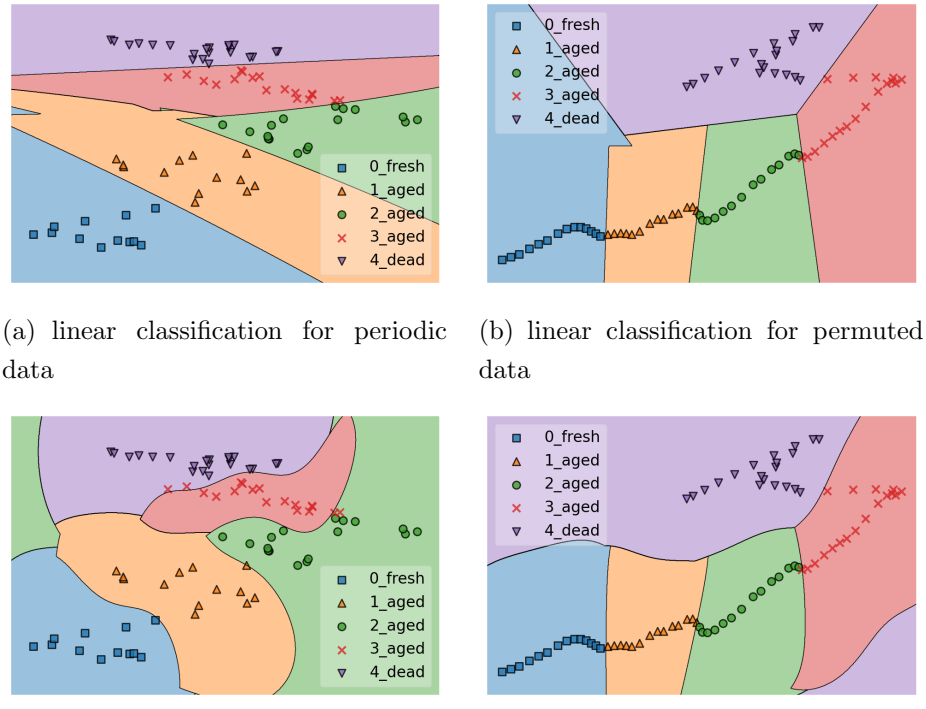
5.5 Conclusions

This paper has led to the development of a robust classifier capable of effectively segmenting three distinct battery cell groups: fresh cell, aged cell, and dead cell. The

classification process, as well as the observation of battery aging pathways, is conducted within a two-dimensional latent space obtained through the machine learning algorithm autoencoder. To transition from linear to non-linear regimes and to mimic real-world operational conditions, ideal periodic and noisy permuted time-based battery data are utilized to be mapped into the latent space. To validate their robustness and efficiency, traditional impedance is employed as a reference. The consistent and observable robust aging motion, as depicted by the distribution of classes in the latent space, remains unaffected by the choice of using time-based or frequency-based data, as well as whether the data is ideal or noisy. This finding underscores the potential and reliability for the development of stochastic pulse sequences that can optimize charging strategy to extend the battery's lifetime.

5.6 Supplementary Information

The segmentation of battery data into 5 classes in the latent space still provides aging motion, from blue fresh cells through orange, green, and red aged cells into purple damaged cells (Figure 5.5). The aging pathway stay consistent in periodic and permuted data.



(a) linear classification for periodic data

(b) linear classification for permuted data

(c) non-linear classification for periodic data

(d) non-linear classification for permuted data

Figure 5.5: Classification of batteries into five aging groups within the latent space.

Chapter 6

Stochastic pulse excitation

This chapter is adapted from the pre-final version of publication that will be submitted soon: **Online battery state analysis using stochastic pulse excitation**, by L. Jin, F. Bereck, R.-A. Eichel, J. Granwehr, and C. Scheurer.

The theoretical design of stochastic pulse excitation, simulation of battery models, machine learning and feature analysis in latent space were conducted by the author of this thesis, Limei Jin. The experimental limitation was provided by Franz Bereck.

6.1 Introduction

EIS is a standard spectroscopic method that provides a more detailed understanding of a battery's state compared to simple current-voltage curves. However, traditional EIS methods often face challenges, such as the assumption of operation in the linear response regime, and the time required to acquire a spectrum. These limitations make it difficult to record a full EIS spectrum during driving, as the transient state changes of the battery may occur on time scales shorter than the acquisition time. Therefore, a more realistic and comprehensive assessment of battery load profiles in diverse and dynamic real-world scenarios would be desirable.

In battery system assessment, a variety of excitation signals are utilized to probe and model the dynamics of a system based on its response to these inputs. One widely used technique is the single-frequency method [128], which encompasses excitation signals that are characterized by their for practical purposes monochromatic nature, presenting one frequency at a time to the system. This category includes techniques such as the sine sweep and square sweep, allowing for an in-depth analysis of the system's response across the frequency spectrum. The EIS method also uses the sine sweep excitation signals. However, the single-frequency sweep requires a longer measurement time and a higher Signal-to-Noise ratio [129]. Because each frequency is tested sequentially,

extended periods of time are required to capture a stable system response and ensure clear signal detection amidst background noise.

Broadband methods [128], on the other hand, are distinguished by their multi-frequency content, capable of stimulating a system across a wide frequency range simultaneously. Within these methods, the step signal, which is a non-periodic signal that abruptly transitions from one level to another, provides a basic excitation scheme. This abruptness makes it a potent tool for evaluating a system's transient response, especially for high frequency response [130]. Multi-sine signals are another option, combining several sine waves of different frequencies. This method is advantageous as it allows for observing a system's behavior efficiently, with the benefit of requiring simple setups and minimal computational analysis, thus reducing overall testing complexity [131]. For the specific focus of our study, we employ pseudo-random sequences as a flexible scheme within the space of conceivable implementations of the broadband strategy. These sequences, although seemingly random, are precisely defined and reproducible, offering the advantage of a broad frequency range stimulation while ensuring consistency and reliability in experimentation. Pseudo-random sequences mimic the stochastic nature of real-world signals, providing a robust assessment of a system's performance under conditions that closely resemble actual operational scenarios.

To study the characteristics of voltage and current responses in batteries under dynamic conditions, we apply stochastic pulse current signals as pseudo-random excitation to the battery and analyze the resulting voltage responses. Varying the amplitude and duration of the pulse signals allows exploring how the battery responds to different levels of excitation and sudden changes in input conditions. Then, current pulse input signals and their response voltage signals for different states of various batteries are used to train a convolutional autoencoder to produce low-dimensional latent spaces. The resulting latent spaces derived from different excitation signals are systematically evaluated by employing traditional impedance spectroscopy as reference. The comparative analysis involves probability distribution comparison through quantiles and segmentation analysis of the latent space, focusing on regions that can be assigned to aging. The ultimate goal is to identify an optimal pulse sequence by evaluating how well it aligns with the reference impedance data in terms of distribution and aging-related segmentation within the latent space.

6.2 Methodology

6.2.1 Pulse excitation concept

The construction principle for the systematic generation of stochastic current signals as pseudo-random excitation to the battery is inspired by the Koch curve [132]. Their geometric shape displays self-similarity across different scales. It is constructed by starting with a straight line segment, dividing it into thirds, replacing the middle segment with two segments that form an equilateral triangle, and then removing the base of the triangle. This process is repeated indefinitely, creating a complex and detailed fractal pattern. In our approach, we utilize a related iterative and self-similar construction to design excitation signals that can test the battery's response under varied yet precisely defined conditions, offering insights into its imitated real-world performance. As Figure 6.1 shows, the first hierarchy level is a simple rectangular pulse

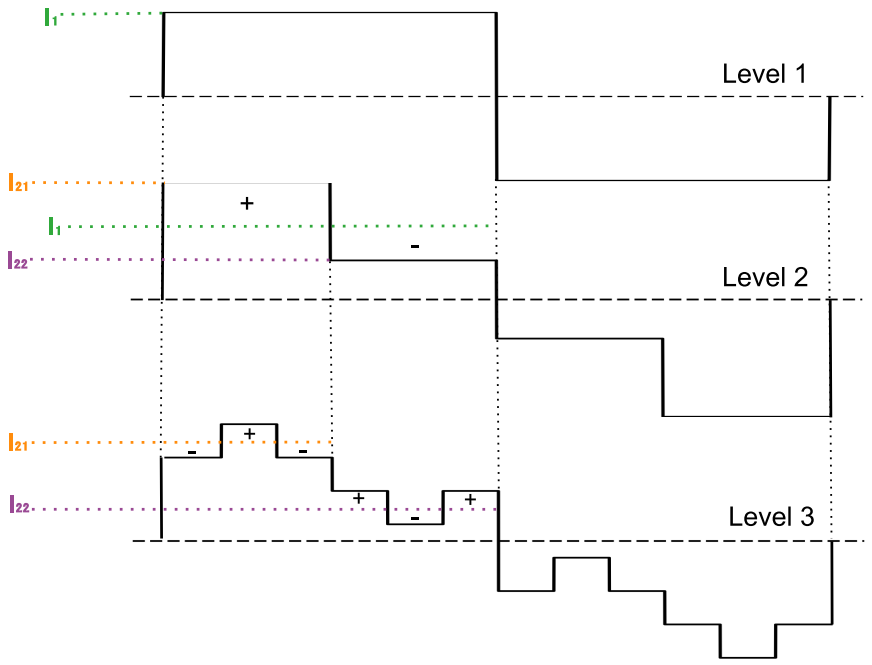


Figure 6.1: The concept of stochastic pulse generation from first level to higher levels

waveform and total excitation with equal positive and negative amplitudes asserts that the overall SOC remains essentially unchanged. The subsequent level is created by splitting level 1 into two parts. The amplitude of a subpart is selected based on the previous level's amplitude, with a small variation added and subtracted to keep the total values consistent with the previous level. Similarly, higher levels continue this pattern. Pulse signal at level 3 is generated by splitting sub-pulse at level 2 into 3 parts, and the integrated amplitude values remain identical to level 2. By using iterations of this kind, one can generate more complex pulse sequences from simple wavelets.

The actual protocol for these hierarchical levels is described by amplitude and time duration variation, which is systematically controlled as follows:

- the starting *Level 1* has a pulse amplitude I_1, which is limited by the experimental physical system to avoid significant SOC fluctuations on the slowest timescale. The duration is set as $t_1 = 1$s, which means a formal frequency scale given by $f_1 = 1$ Hz, not accounting for the high frequency components due to the step edges.

- *Level 2* has amplitude I_2 derived from I_1, with small variation on it, $I_2 = |I_1 \pm \alpha|$. The random selection of sub-waveform is reflected in taking plus or minus. And α represents a small variation on the amplitude. Their duration is multiplied by previous level $t_2 = \frac{1}{2}t_1$, which is equal to $f_2 = 2f_1$.

- Similarly, *Level 3* follows with amplitude $I_3 = |I_2 \pm \alpha|$ and duration $t_3 = \frac{1}{3}t_2$.

- Continue this pattern, at *Level l*, the amplitude would be $|I_l = I_{l-1} \pm \alpha|$, and duration $t_l = \frac{1}{l} \cdot t_{l-1}$, which means frequency $f_l = l \cdot f_{l-1}$.

The number of excitation levels uses prime numbers to minimize duplication and enhancing the uniqueness. In signal processing and system excitation, avoiding harmonic overlap is crucial for clear and distinct results. Prime numbers, by their nature, do not share common factors with each other, reducing the likelihood of harmonic interference or resonance effects that could skew the results or make them less distinct. Therefore, excitation levels $l = \{1, 2, 3, 5, 7, 11, 13...\}$ ensure that each level is distinct from the others and minimizes the risk of overlapping. Additionally, the prime number sequence grows in a way that inherently spaces the numbers apart at increasing intervals, which contributes to an efficient and broad coverage of the excitation spectrum.

When designing experiments that involve current pulse excitation, the number of excitation levels is closely tied to the constraints and capabilities of the experimental hardware. Critical parameters include the maximum sampling rate of the system and the total number of steps per excitation profile that can be effectively managed on memory-limited, cheap commodity hardware. For instance, if 3 levels, *Level 1, 2, 3*, are chosen, the highest frequency for the pulse excitation would be 6Hz. To accurately capture this, the system's achievable time resolution must be at least 0.167s, ensuring each waveform is properly recorded. Additionally, the profile comprises a minimum of 3 charging steps and 3 discharging steps, totally 6 steps for a complete operational cycle. These requirements are crucial to ensure that the hardware can handle rapid transitions between charging and discharging states without any significant delay or error.

Combining these generated levels into a single current pulse sequence introduces randomness. At each level, a sub-pulse is randomly selected. For example, at level 2, there are two sub-pulses with different amplitudes. One of these can be chosen to represent level 2. Through the combination and permutation of different levels with their respective sub-pulses, various sets of pulse current profiles are generated. Figure 6.2 shows one such generated pulse current profile, which includes seven different levels arranged in a random order. The negative pulse excitation is designed similarly

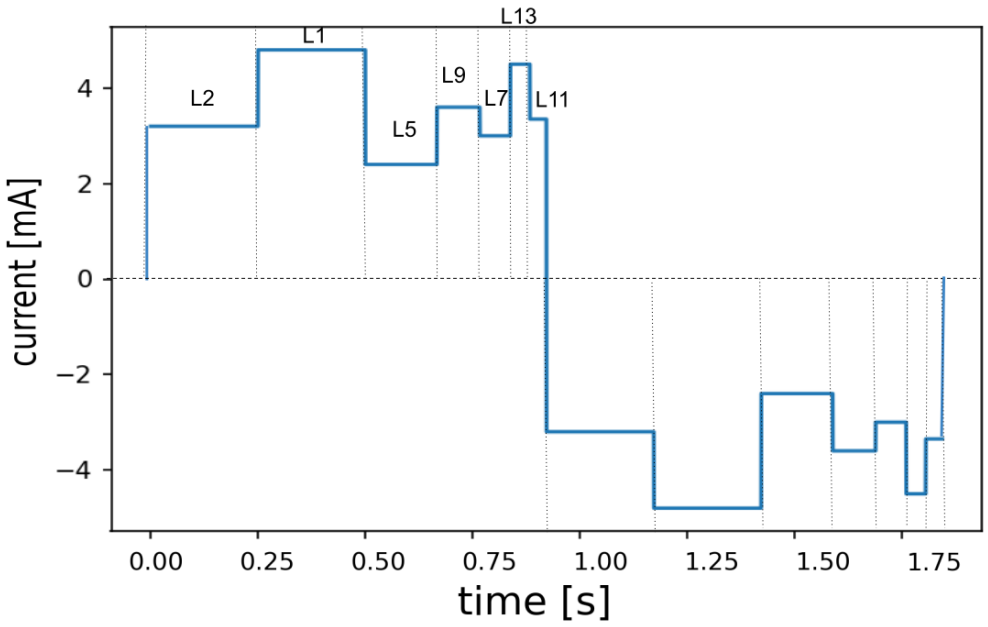


Figure 6.2: Example of generated pseudo-random current pulse sequence

to the positive pulse excitation, ensuring that the total integrated current amplitude of the pulse profile remains zero to maintain SOC consistency. Consequently, it allows not only for the excitation specifically designed for battery characterization, but also for the excitation that simulate real-world scenarios to be taken as input for charging or discharging process without external control. Furthermore, this scheme includes both low and high-frequency components, which is beneficial for battery testing, as it allows for a comprehensive evaluation of the system's response across an extensive spectrum of frequencies. Unlike Pseudo-Random Binary Sequence (PRBS) excitation, where the bandwidth is constrained by the sample period at high frequencies and by the sequence's period at low frequencies [133], our approach ensures the inclusion of frequencies as high and as low as practically feasible. This broader frequency capability is crucial, especially since PRBS excitation often encounters system identification errors near the edges of its bandwidth and is typically limited to load-free time periods [134].

6.2.2 Latent space segmentation

The generated pulse excitation sequences are used in simulations of the numerical battery models at different states (chapter 4) to gather a wide variety of response voltage signals. To extract meaningful features from extensive simulation datasets of paired pulse current and voltage, an unsupervised ML technique based on a convolutional autoencoder, is employed. This reduces the high dimensionality of the space of battery models in a non-linear way and enables battery aging pattern recognition in a low-dimensional Latent Space (chapter 5). Thereby, each pulse profile leads to a projection of time-series battery response data into a distinct and independent latent space composed of multiple points, where each point represents a battery cell at a different stage of aging. For evaluating the effect of different pulse signals on battery performance, their derived latent space population should be compared with reference space produced by traditional EIS data. This comparative analysis provides insights into how various pulse profiles influence the representation of differently aged battery cells in the reduced-dimensional latent space.

Geometrically, neighboring points in the latent space can be organized into simple triangular regions, to simply assess the relations within the SOH landscape of battery models. Monitoring the movement across these regions enables tracking of battery aging dynamics effectively. As further experimental data becomes available to substitute the synthetica points, more refined approaches for topological analysis can be employed. Delaunay triangulation is a powerful method for creating a triangulation mesh, connecting points in a way that maximizes the minimum angle of all the angles of the triangles [135]. This approach provides a structured visualization of how these points or states are related to each other spatially. The edge lengths of all edges in the triangulated mesh are combined in a histogram. This histogram serves as a tool to identify a suitable edge length as a hyperparameter for segmenting the latent space into reasonably resolved regions. The aim is to avoid segments that are too small, which may represent insignificant variations, or too large, which might cloak important yet subtle differences between states.

Once suitable segmentation criteria are established in the latent space derived from standard EIS data, this segmentation framework can be transfered to the latent space derived from various realizations of pulse data. This allows for a comparison between two latent spaces, aiding in the assessment which pulse signal produces a latent space segmentation most similar to the one derived from traditional EIS.

Comparing two latent spaces derived from pulse and EIS data through segmentation presents certain limitations due to the presence of misclassified points within the groups. This issue arises because the segmentation process is based purely on geometri-

cal separation, and when this segmentation is mapped onto another space, outliers can occur at the edges of the separation. Consequently, a point-to-point comparison between the two latent spaces is considered a more accurate and computationally efficient method. This approach mitigates the challenges posed by geometrical segmentation and provides a more precise alignment between the latent spaces derived from pulse and EIS data.

6.2.3 Point-to-point comparison

Assessing the correlation between segments in the LS faces limitations in detailed understanding of the variations and relationships between individual points within a specific segment. Our goal is to scrutinize whether each point in the LS from EIS data precisely aligns with its counterpart in the LS derived from pulse signals. To facilitate a direct point-to-point comparison of two latent spaces, a more explicit approach, Quantile-Quantile plot, is adopted.

A Q-Q plot is a graphical tool used to estimate the similarity between two probability distributions. If the observed distribution is similar to theoretical distribution, the points in the Q-Q plot will approximately lie along an identity line [136]. In the process of comparing latent spaces derived from EIS data and pulse signals, the feature vectors from both spaces are collected and used to compute the quantiles. To facilitate a meaningful comparison, both datasets undergo a standardization procedure, ensuring that they share a common scale with a mean of 0 and a unit standard deviation. Then voltage responses for 21 evenly spaced SOC percentage points ranging from 0% to 100%, at each 5%, are generated. At each specified percentage point, the actual quantile values for the standardized EIS data and pulse data are calculated. They form a set that is subsequently utilized to construct a Q(EIS)-Q(pulse) plot. The inherent continuity of the distributions being compared, such as normal distributions, result in smooth Q-Q plots. This smoothness arises from the gradual, uninterrupted change in quantile values found in continuous data, as opposed to the discrete jumps that might be seen in more granular or categorical data. With finer gradations between quantiles and larger number of data points, the transition from one quantile to the next becomes less abrupt, leading to a visually smoother curve or line on the graph.

The distance of each point on the Q-Q plot to the identity line serves as a score matrix, enabling the evaluation of different pulse signal performances in comparison to the EIS data. The smaller the distance, the better the alignment between the latent space derived from the two sources, indicating better performance in replicating the

traditional EIS analysis. The score function is defined as:

$$\text{Score} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - y_i)^2} \tag{6.1}$$

where N is the total number of points on the Q-Q plot, (x_i, y_i) are the coordinates of the i-th point in the plot. x_i refers to the quantile value from the pulse sequences data, while y_i is corresponding quantile value from EIS data.

6.3 Methods and Materials

6.3.1 Experimental limitation

Experimental device limitation of the sampling rate is up to $42\mu s$, approximately 23kHz, with 40 steps per profile. Hence, the levels are limited up to *Level 13*, in other words, *Level 1, 2, 3, 5, 7, 11, 13* are taken. The designed current profile encompasses 7 charging steps and 7 discharging steps, totally 14 steps. The hierarchical starting level takes frequency 1Hz with duration 1s, aiming to capture significant characteristics of the battery load that are most relevant and informative under typical driving conditions in EVs [137]. Additionally, assume that using up to 0.035% of the battery's total capacity in Coulombs during this short pulse time is acceptable to ensure that the SOC remains unchanged. When working with a fresh coin cell battery with maximum 40mAh capacity and the pulse duration is restricted to no more than 5 seconds, consequently, the amplitude should be capped at 10mA to avoid significant SOC fluctuations [138]. Therefore, the starting level's amplitude is set at 3.2mA as basis to ensure minimal SOC changes with small excitation for coin cell with capacity from 30mAh to 16mAh.

6.3.2 Simulation

The combination and permutation of chosen 7 levels can generate a diverse set of pulse current profiles, estimated to be around 5000 profiles. Besides, the battery models at different SOC/SOH combinations are computationally generated (chapter 4) and stored, around 1500 cell models. A real-time capable electrical simulation platform, *ISEA Framework*, is utilized to simulate these models with current profiles and gather voltage output [33]. This means the electrical simulation runs 7.5 million times and correspondingly millions of voltage/current data were gathered. These large amount of pairwise current and voltage data with their SOC information are fed into constructed

convolutional autoencoder (chapter 5). Each pairwise pulse voltage/current data produces a latent space, which is constructed from 1500 models. Consequently, we obtain 5000 diverse latent spaces and each space has 1500 points distributed differently, which represents aging states.

6.4 Results and Discussion

Pulse sequence estimation by quantiles

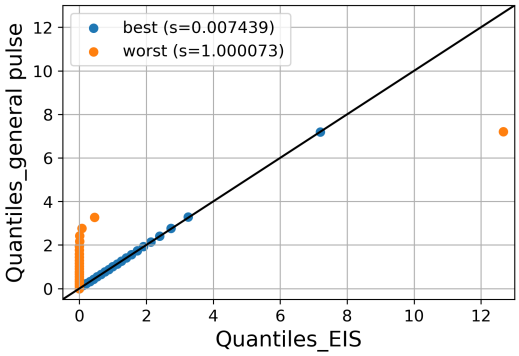
To estimate how various pulse current profiles affect battery's aging process, the response voltage data with its input current data are employed to analyzed in the latent space, which is a reduced representation of complex data sets. Every unique pulse sequence generates its own distinctive latent space, comprising 2D features extracted by the encoder. For example, blue pulse sequence in Figure 6.3c produces its corresponding LS illustrated in Figure 6.5c. The points in the latent space represent the battery cell at its different aging states. Additionally, traditional EIS data is taken as a reference standard hence it also yields its own latent space (Figure 6.5a). It allows us to compare the effects of different pulse sequences against a well-established benchmark in battery analysis.

For a visual and quantitative comparison between two latent spaces, we constructed Q-Q plots. These plots were generated by computing the quantiles from the points distributed on the LS derived from both pulse sequence and EIS data. The mean of distances between the points on Q-Q plot and the identity line is employed as a score (Equation 6.1). The scores over all generated pulse sequences are shown in Figure 6.3b, the higher score implies the LS derives from pulse is more similar to it from EIS, therefore, this pulse performs better than the others. Out of all the 5000 pulse sequences generated, we identified the best and the worst one.

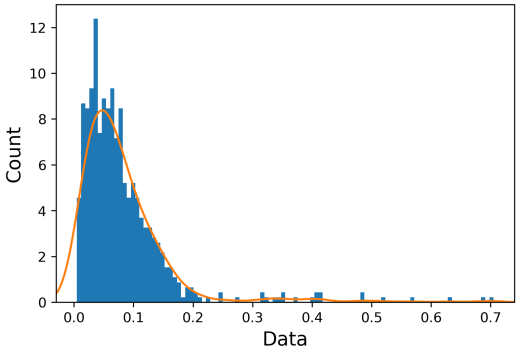
Figure Figure 6.3a presents the Q-Q plots for the pulse sequences with the highest and the lowest scores, showcasing the extremes in performance among different pulse sequences when replicating the distribution of the EIS-derived latent space. The best pulse sequence (shown in blue), indicated by a score value of 0.005 exactly aligning with the identity line, suggests that the distribution of features in its latent space closely resembles that of the EIS-derived latent space. In contrast, the worst pulse sequence (orange), with a score value of 0.7, deviates significantly from the identity line. The fast vertical layout and a region that is sampled very poorly in the Q-Q plot indicate substantial disparities between the latent space produced by this pulse sequence and the one obtained from EIS data. These discrepancies might signify a poor replication of the electrochemical characteristics in the latent space, potentially hindering the

effectiveness of the pulse sequence in capturing the underlying electrochemical behavior of the battery.

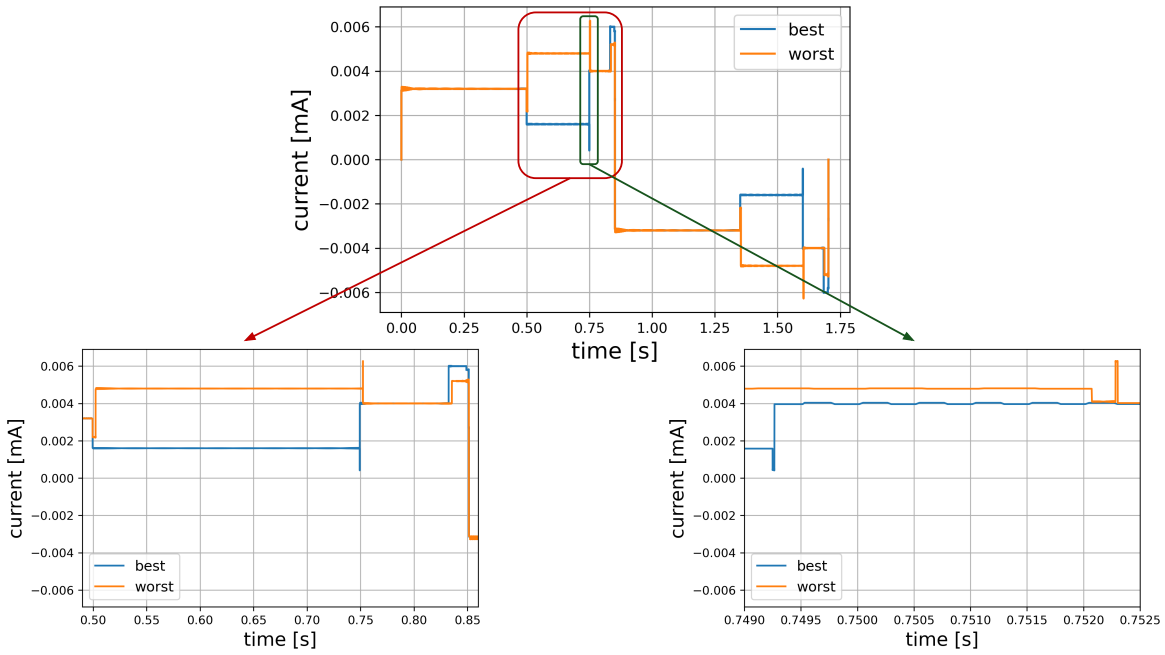
In Figure Figure 6.3c, the specific instances of the best and worst pulse sequences are visually represented over time. Due to the presence of high-frequency values in certain segments of the pulse sequences, the shapes of these pulses are challenging to discern within small time intervals.



(a) Q-Q plot



(b) histogram of scores

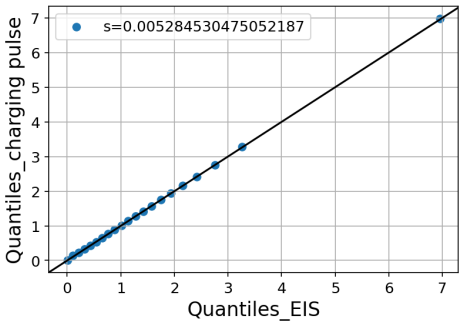


(c) pulse sequences over time

Figure 6.3: Selected best and worst pulse sequences using quantiles to estimate similarity to latent space derived from EIS data

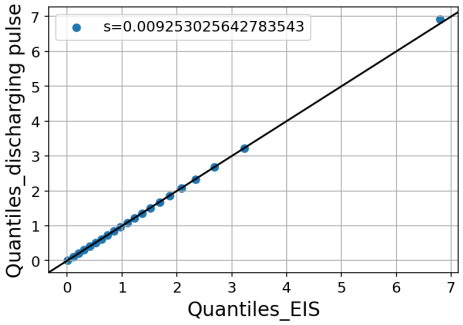
Additionally, we analyze pure charging sequences and pure discharging sequences, which are subdivided from the generated general sequences. These are then fed into an autoencoder for training. Utilizing the same estimation method to the distribution of the latent space through quantiles, we have identified the optimal charging

pulse sequence (Figure 6.4a) and the optimal discharging pulse sequences (Figure 6.4c), comparing to EIS data as reference. While the estimated optimal charging sequence performs similarly to the best general sequence that previously selected (Figure 6.4b), the optimal discharging sequence demonstrates a significant discrepancy, as illustrated in Figure 6.4d. It underscores the sensitivity of battery performance to specific charging and discharging patterns. Even minor variations in pulse sequences can have significant impacts on battery health and efficiency. This sensitivity suggests a potential avenue for further optimization in designing charging and discharging protocols to maximize performance and longevity. Figure 6.4e displays the selected optimal general sequence, charging sequence, and discharging sequence over time, which patterns of these sequences are distinctly different from each other.

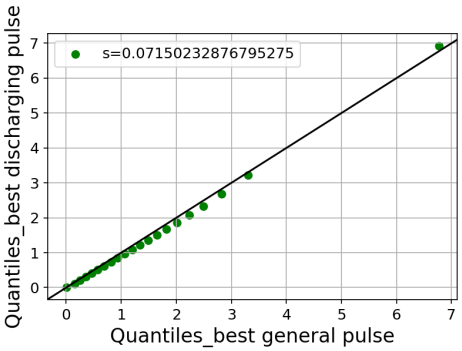


(a) optimal charging pulse against to EIS data

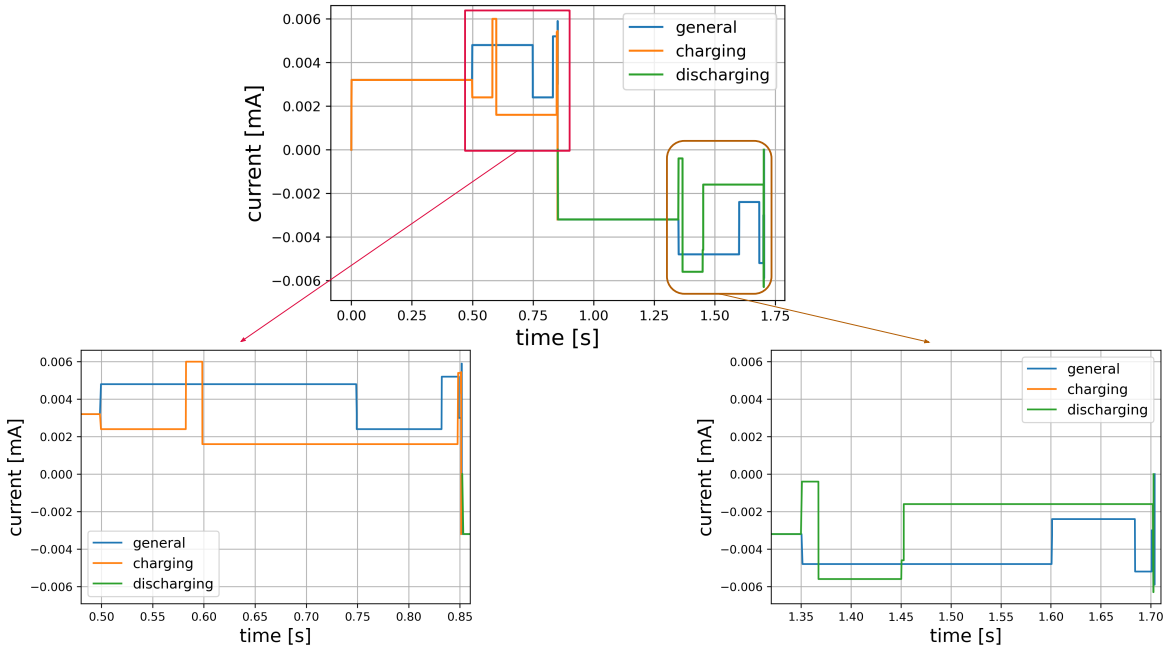
(b) optimal charging pulse against to the best general pulse



(c) optimal discharging pulse against to EIS data



(d) optimal discharging pulse against to the best general pulse



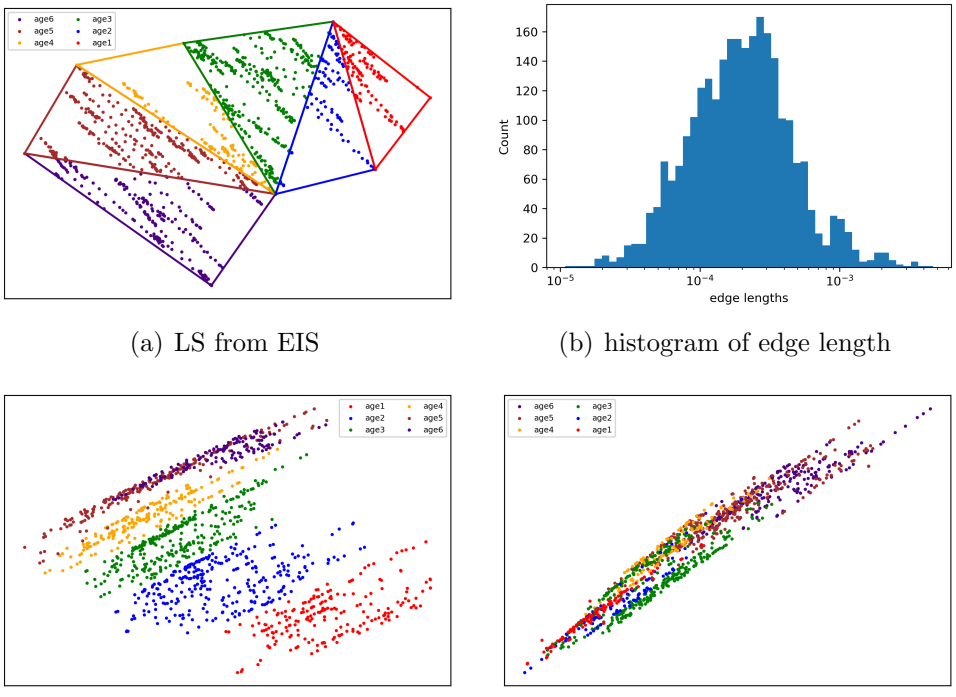
(e) optimal general, charging, discharging pulse sequence over time

Figure 6.4: Selected best charging and discharging pulse sequences using quantiles to estimate similarity to latent space derived from EIS data, and their correlation to the best general sequences

Pulse sequence estimation by LS segmentation

The performance of identified general best and worst pulse sequences can be also visual in the latent space segmentation (Figure 6.5). The division of space uses Delaunay triangulation to create a mesh of triangles and find reasonable edge length for partition into roughly similar areas (Figure 6.5b). As Figure 6.5a shows, the reference latent space can be divided into six regions, setting battery aging stages from 1 to 6, indicated by colors transitioning from red to purple. These classified points are mapped to the points in the latent spaces derived from general best and worst pulses. Since the battery models with different health states simulated for pulse current and sinusoidal current are identical, mapping model points from EIS LS to the best pulse LS (Figure 6.5c) and to the worst pulse LS (Figure 6.5d) is straightforward. Their coloring is according to the EIS segments.

Overall, the mapped segmentation in the LS from the best pulse remains similar to the EIS data, while the mapped groups in the LS from the worst pulse are significantly mixed up. It indicates that the latent space derived from the worst pulse sequence fails to consistently represent the distinct battery aging stages that are clearly observed in the reference EIS data. The mixed-up groups in the worst pulse latent space suggest an incapability in capturing the underlying patterns of battery aging.



(a) LS from EIS

(b) histogram of edge length

(c) LS from best pulse

(d) LS from worst pulse

Figure 6.5: Mapping latent space segmentation from EIS data into latent spaces from the best and worst pulse sequences

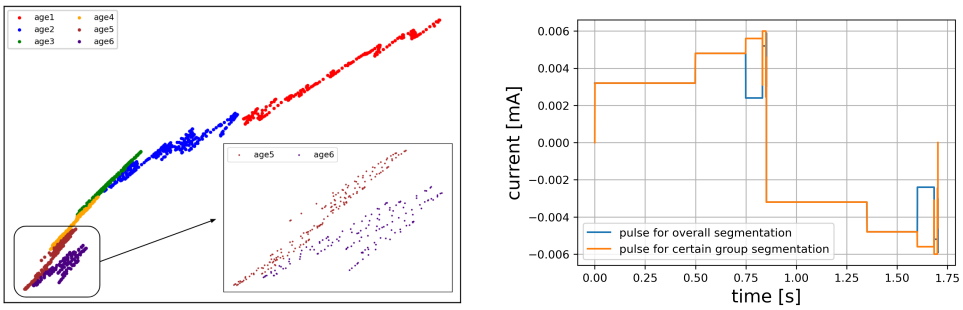


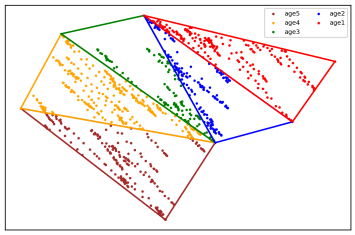
Figure 6.6: alternative pulse sequence for better segmentation of age group 5 and age group 6

Although, the overlap between the brown and purple regions in Figure Figure 6.5c, where some purple group members appear in the brown region, can be attributed to the geometric division of the latent space from EIS data. The division creates triangular regions, and at the edges of these triangles, points are situated very closely to each other. As a result, points near the edges might be assigned to the same group despite belonging to different geometric regions.

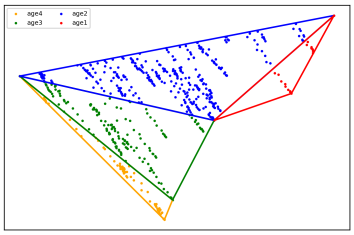
Additionally, the separation of age 1 and 2 is even more broadly distributed in the best pulse case than in the EIS, while age 5 and 6 are poorly separated. Consequently, we have excluded the features of ages 5 and 6 and reapplied the quantile criteria to identify an alternative pulse excitation. It aims to treat the age 5 and 6 categories similarly to how ages 1 and 2 are treated. As illustrated in the Figure 6.6, we successfully identified an alternative pulse sequence that more effectively separates the age 5 and 6 groups. While the total latent space of this new sequence does not closely resemble that of the EIS latent space and there is some overlap with other age groups, specifically ages 2 and 3, it nonetheless offers improved clarity in distinguishing between certain groups. It highlights that at certain stages of a battery's aging process, altering the pulse sequence can lead to enhanced performance in terms of age group separation. This insight underscores the potential for adaptive strategies in battery management, where pulse sequences can be dynamically adjusted based on the battery's stage of aging to optimize performance and possibly extend its lifespan.

Therefore, we take different number of groups to assess the clarity of segmentation, as Figure 6.7 shows. Regardless of which number of categorization is chosen, overlapping between neighboring categories always persists. The possible reason is the algorithm behind for selecting edge length of triangles among a mesh of Delaunay triangles, such that no point lies inside the circumcircle of any triangles. Here, the mean of the edge lengths of each triangle is computed. The criteria is to extract the top number of triangles with the largest mean edge lengths for cutting into corresponding number

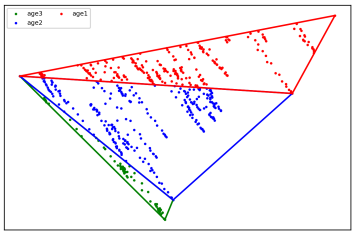
of groups. Hence, selecting top 5, top 4, or top 3 largest triangles, result in absolutely different geometric division. Mapping this separation into latent space derived from best pulse sequence makes the categorization not really explicit. But it does not significantly impact the overall segmentation. The movement of aging patterns in the LS from the best pulse sequence remains consistent with the segmentation observed in the EIS data. The key features and trends in the battery aging process are still accurately captured, underscoring the robustness and reliability of the latent space representation derived from the best pulse sequence in mirroring the behavior seen in EIS data.



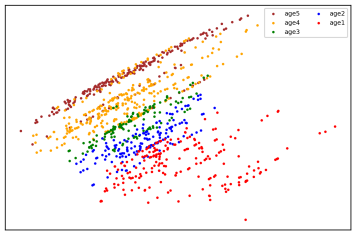
(a) 5-groups segmentation of EIS space



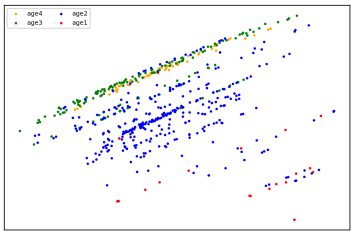
(b) 4-groups segmentation of EIS space



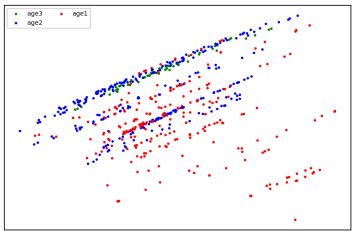
(c) 3-groups segmentation of EIS space



(d) Mapping 5-groups segmentation to pulse space



(e) Mapping 4-groups segmentation to pulse space



(f) Mapping 3-groups segmentation to pulse space

Figure 6.7: Latent space segmentation into 3, 4 or 5 groups derived from EIS data and mapping to optimal pulse space

6.5 Conclusions

Through the systematic generation of stochastic pulse sequences without SOC changes, and using them to simulate battery models at various states, we collect a substantial amount of pulse current and voltage data. They are mapped into a latent space by encoder, which facilitates a straightforward comparison of the distribution of models with a reference space obtained from EIS data. By conducting a point-to-point comparison by Q-Q plot between these two spaces, we can identify the optimal and worst-performing pulse sequences. The performance of the optimal pulse in grouping and tracking in terms of aging in the latent space demonstrates a similar pattern

to EIS analysis, showcasing its effectiveness in capturing crucial characteristics of the battery's state. Furthermore, testing the optimal pulse in a real battery experiment verifies its adaptability and serves as a foundation for future investigations, which may involve running specific charging or discharging profiles of pulse sequences for extended durations, aiming to explore their optimization for battery lifetime and their impact on fast-charging scenarios.

Chapter 7

Summary and Outlook

The thesis exploits the application of battery digital twins with machine learning algorithms to estimate its aging behaviour and optimize charging profiles. The EIS measurement on the coin cell provides real impedance data for construction of equivalent virtual battery models with relatively short analysis time. The GPR-based DRT method is facilitated for a physically justified ZARC series parametrization from an EIS measurement with minimal or no user intervention. Thereby, selected suitable ECM for the investigated battery and its robust ZARC parameters were obtained across the full range of charge states without obvious outliers. The other meaningful elements for an ECM is extracted by curve fitting. It has been shown that the stability and quality of ECM fits were improved by including this additional knowledge about the physical state of the measured system in the fitting process without introducing arbitrary assumptions.

The fitted ECM parameters at various SOCs are stored as look-up table. In order to reduce the dimensionality of model representation, Chebyshev polynomials quantify the SOC dependencies in a robust way, which allows to reconstruct the ECM at any desired SOC value in a concise manner. Thereby, each charging/discharging cycle can be represented by a certain set of Chebyshev coefficients that correspond to the fitted parameters at different SOC values during that cycle. These coefficients form the battery's history over that specific cycle.

To extend this coefficients to broader cycles for tracking aging performance, theoretically, one should take EIS measurements at multiple charging/discharging cycles and repeat the parametrization and SOC-quantification procedures, which is extremely time-consuming. However, since aging is a multifaceted process, encompassing a diverse range of aspects and dimensions of battery's behaviour, cycle numbers cannot be used to represent the aging process comprehensively. Therefore, the QMC method helps to generate sufficient sets of high dimensional Chebyshev coefficient vectors based

on a limited set of experimental data to build certain regions of battery health without massive experiments. The limited experiments are taken at distinct states of health: fresh cell, aged cell, and damaged cell. Thereby, the aging behavior can be tracked by segmentation of different battery health categories. These sampled large amounts of numerical models work as foundation for further machine learning, which allows for tailored approaches to mitigate or manage specific degradation patterns.

These numerical EC-models were simulated under different current waveforms. Ideal periodic sinusoidal current with its voltage response data is utilized to compute impedance, serving as a reference. The noisy permuted current and paired voltage data is used to mimic real-world operational conditions. Furthermore, stochastic pulse current with its voltage response data is used to estimate best charging profiles for optimization of battery performance. The electrical simulation, which entails applying a current input profile to an EC-model at a single SOC to obtain a corresponding voltage profile, is carried out using the ISEA Framework.

To deal with a large number of pairwise current and voltage data with its SOCs, an unsupervised machine learning algorithm, convolutional autoencoder is employed to reduce dimensionality and extract 2D-features into latent space. In the latent space, each EC-model with its various SOCs over a single cycle is represented as a single point. The aggregation of these points, corresponding to different cycles, and their spatial distribution collectively delineate an aging pathway. Thereby, applying a single current profile across all numerical models results in the creation of one latent space, comprising multiple data points.

To transition from linear to non-linear regimes and to mimic real-world operational conditions, latent spaces from impedance data, from ideal periodic data, and from noisy permuted time-based data, are compared. This comparison of latent spaces involves utilizing a robust classifier by SVM that is adept at effectively segmenting the data into three distinct groups representing different states of battery cells: fresh, aged, and damaged. Consequently, the consistent and observable robust aging motion, as depicted by the distribution of classes in the latent space, remains unaffected by the choice of using time-based or frequency-based data, as well as whether the data is ideal or noisy. This finding underscores the potential and reliability for the development of stochastic pulse sequences that can optimize a charging strategy.

Considering experimental limitation of device, we systematically generate thousands of stochastic pulse sequences, and simulate them with our large number of numerical models one by one. Subsequently, their pairwise current and voltage data are processed through a convolutional autoencoder. As a result, the thousands of pulse profiles used in the analysis generate thousands of distinct latent spaces. Addition-

ally, the latent space derived from traditional impedance is employed as a reference for validation of their robustness and efficiency. To efficiently compare the distribution of models among a large number of latent spaces, the Q-Q plot is used to identify the optimal and worst-performing pulse sequence compared to reference space. The performance of identified optimal pulse in grouping and tracking in terms of aging in latent space demonstrates a similar pattern to EIS analysis, showcasing its effectiveness in capturing crucial characteristics of the battery's state. Furthermore, testing the optimal pulse in a real battery experiment verifies its adaptability and serves as a foundation for future investigations. Further longer optimal charging pulse sequences for operating batteries over cycles can be targeted in the future.

In summary, this study demonstrates that using a battery digital twin effectively bridges the gap between data acquisition and data analysis. It showcases the feasibility of developing these processes separately while utilizing the digital twin as a pivotal connecting element. Moreover, it enables the management of the battery at the cell level, offering a granular and precise approach to battery maintenance and operation. Furthermore, it indicates that the presented parametrization strategy has the potential to evolve the digital twin into a predictive model, rather than limiting it to a mere downstream EC model based on measured or simulated data. Additionally, this approach aids in identifying the optimal load profile for operating the battery, while considering the physical constraints of the device.

The current methods still face certain limitations that could impact their efficacy and applicability. One such limitation is the reliance on limited experimental datasets. In this research, a limited set of experimental battery data was used for the development and validation of the framework. Further, an area for potential exploration could be the extensive comparison of the performance of numerical models in latent space against a larger dataset of experimentally obtained battery aging data. This future endeavor would not only validate and potentially refine the current models but also enhance their generalizability in the real-world battery aging processes.

Another limitation is the exclusion of temperature variations, which are critical to battery performance and aging. In this study, we have focused on SOC and SOH as the two crucial parameters for maintaining battery operation, with the assumption that the temperature is set to room temperature. Looking ahead, integrating temperature as an additional parameter in the virtual battery models and within the machine learning networks represents a promising avenue for future exploration. This consideration would enhance the accuracy and applicability of the models in real-world scenarios.

Lastly, further refining the parametrization strategy to include more sophisticated or additional parameters could improve the predictive capabilities of the digital

twin. By continuously updating and validating the model against larger, more varied datasets, it's possible to achieve a digital twin that not only accurately reflects current battery states but also reliably forecasts future conditions and performance.

References

[1] Yaosen Tian et al. "Promises and Challenges of Next-Generation "Beyond Li-ion" Batteries for Electric Vehicles and Grid Decarbonization". In: *Chemical Reviews* 121 (2021), pp. 1623–1669. ISSN: 0009-2665. DOI: 10.1021/acs.chemrev.0c00767. URL: https://doi.org/10.1021/acs.chemrev.0c00767.

[2] Xianzhi Gong, Rui Xiong, and Chunting Chris Mi. "Study of the Characteristics of Battery Packs in Electric Vehicles With Parallel-Connected Lithium-Ion Battery Cells". In: *IEEE Transactions on Industry Applications* 51.2 (2015), pp. 1872–1879. DOI: 10.1109/TIA.2014.2345951.

[3] Hossam A. Gabbar, Ahmed M. Othman, and Muhammad R. Abdussami. "Review of Battery Management Systems (BMS) Development and Industrial Standards". In: *Technologies* 9.2 (2021). ISSN: 2227-7080. DOI: 10.3390/technologies9020028. URL: https://www.mdpi.com/2227-7080/9/2/28.

[4] Billy Wu et al. "Battery digital twins: Perspectives on the fusion of models, data and artificial intelligence for smart battery management systems". In: *Energy and AI* 1 (2020), p. 100016. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2020.100016. URL: https://www.sciencedirect.com/science/article/pii/S2666546820300161.

[5] Allen J. Bard and Larry R. Faulkner. "Electrochemical Methods: Fundamentals and Applications". In: *Russian Journal of Electrochemistry* 38 (2002), pp. 1364–1365. ISSN: 1608-3342. DOI: 10.1023/A:1021637209564.

[6] F. Naseri et al. "Digital twin of electric vehicle battery systems: Comprehensive review of the use cases, requirements, and platforms". In: *Renewable and Sustainable Energy Reviews* 179 (2023), p. 113280. ISSN: 1364-0321. DOI: 10.1016/j.rser.2023.113280. URL: https://www.sciencedirect.com/science/article/pii/S1364032123001363.

[7] Ulrike Krewer et al. "Review—Dynamic Models of Li-Ion Batteries for Diagnosis and Operation: A Review and Perspective". In: *Journal of The Electrochemical*

Society 165.16 (Nov. 2018), A3656. DOI: 10.1149/2.1061814jes. URL: https://dx.doi.org/10.1149/2.1061814jes.

[8] Pooja Vadhva et al. "Electrochemical Impedance Spectroscopy for All-Solid-State Batteries: Theory, Methods and Future Outlook". In: *ChemElectroChem* 8.11 (2021), pp. 1930–1947. DOI: 10.1002/celc.202100108. URL: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/celc.202100108.

[9] Carl R. Nave. *Complex impedance*. http://hyperphysics.phy-astr.gsu.edu/hbase/electric/impcom.html#c1. Accessed: 01.2022.

[10] Yicheng Liao and Xiongfei Wang. "General Rules of Using Bode Plots for Impedance-Based Stability Analysis". In: *2018 IEEE 19th Workshop on Control and Modeling for Power Electronics (COMPEL)*. 2018, pp. 1–6. DOI: 10.1109/COMPEL.2018.8460168.

[11] Hend S. Magar, Rabeay Y. A. Hassan, and Ashok Mulchandani. "Electrochemical Impedance Spectroscopy (EIS): Principles, Construction, and Biosensing Applications". In: *Sensors* 21.19 (2021). ISSN: 1424-8220. DOI: 10.3390/s21196578. URL: https://www.mdpi.com/1424-8220/21/19/6578.

[12] Lijun Zhang et al. "Comparative Research on RC Equivalent Circuit Models for Lithium-Ion Batteries of Electric Vehicles". In: *Applied Sciences* 7.10 (2017). ISSN: 2076-3417. DOI: 10.3390/app7101002. URL: https://www.mdpi.com/2076-3417/7/10/1002.

[13] D. Andre et al. "Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. II: Modelling". In: *Journal of Power Sources* 196.12 (2011), pp. 5349–5356. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2010.07.071. URL: https://www.sciencedirect.com/science/article/pii/S0378775310012942.

[14] Juan A. López-Villanueva et al. "A compact model of the ZARC for circuit simulators in the frequency and time domains". In: *AEU - International Journal of Electronics and Communications* 153 (2022), p. 154293. ISSN: 1434-8411. DOI: 10.1016/j.aeue.2022.154293. URL: https://www.sciencedirect.com/science/article/pii/S1434841122001820.

[15] Christian Bartsch. "Scalable Data Aggregation and Interpretation of Electrochemical Impedance Spectroscopy for Battery Characterization". MA thesis. RWTH Aachen University, 2022.

[16] Hui-Chia Yu et al. "Simulation of the diffusional impedance and application to the characterization of electrodes with complex microstructures". In: *Electrochimica Acta* 354 (2020), p. 136534. ISSN: 0013-4686. DOI: 10.1016/j.electacta.2020.136534. URL: https://www.sciencedirect.com/science/article/pii/S0013468620309270.

[17] Heiko Witzenhausen. "Elektrische Batteriespeichermodelle: Modellbildung, Parameteridentifikation und Modellreduktion". PhD thesis. RWTH Aachen University, 2017. DOI: 10.18154/RWTH-2017-03437. URL: https://publications.rwth-aachen.de/record/687819.

[18] C. R. Birkl et al. "A Parametric Open Circuit Voltage Model for Lithium Ion Batteries". In: *Journal of The Electrochemical Society* 162.12 (2015), A2271. DOI: 10.1149/2.0331512jes. URL: https://dx.doi.org/10.1149/2.0331512jes.

[19] Bernard A Boukamp. "Distribution (function) of relaxation times, successor to complex nonlinear least squares analysis of electrochemical impedance spectroscopy?" In: *Journal of Physics: Energy* 2.4 (2020), p. 042001. DOI: 10.1088/2515-7655/aba9e0. URL: https://dx.doi.org/10.1088/2515-7655/aba9e0.

[20] E. Ivers-Tiffée and A. Weber. "Evaluation of electrochemical impedance spectra by the distribution of relaxation times". In: *Journal of the Ceramic Society of Japan* 125.4 (2017), pp. 193–201. DOI: 10.2109/jcersj2.16267.

[21] Jacqueline S. Edge et al. "Lithium ion battery degradation: what you need to know". In: *Phys. Chem. Chem. Phys.* 23 (14 2021), pp. 8200–8221. DOI: 10.1039/D1CP00359C. URL: http://dx.doi.org/10.1039/D1CP00359C.

[22] Sabine Piller, Marion Perrin, and Andreas Jossen. "Methods for state-of-charge determination and their applications". In: *Journal of Power Sources* 96.1 (2001). Proceedings of the 22nd International Power Sources Symposium, pp. 113–120. ISSN: 0378-7753. DOI: 10.1016/S0378-7753(01)00560-2. URL: https://www.sciencedirect.com/science/article/pii/S0378775301005602.

[23] Kiarash Movassagh et al. "A Critical Look at Coulomb Counting Approach for State of Charge Estimation in Batteries". In: *Energies* 14.14 (2021). ISSN: 1996-1073. DOI: 10.3390/en14144074. URL: https://www.mdpi.com/1996-1073/14/14/4074.

[24] Amelie Krupp et al. "Incremental Capacity Analysis as a State of Health Estimation Method for Lithium-Ion Battery Modules with Series-Connected Cells". In: *Batteries* 7.1 (2021). ISSN: 2313-0105. DOI: 10.3390/batteries7010002. URL: https://www.mdpi.com/2313-0105/7/1/2.

[25] Ziyong Xu et al. "Aging Characteristics and State-of-Health Estimation of Retired Batteries: An Electrochemical Impedance Spectroscopy Perspective". In: *Electronics* 11.23 (2022). ISSN: 2079-9292. DOI: 10.3390/electronics11233863. URL: https://www.mdpi.com/2079-9292/11/23/3863.

[26] Yunwei Zhang et al. "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning". In: *Nature Communications* 11.1 (2020), p. 1706. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15235-7. URL: https://doi.org/10.1038/s41467-020-15235-7.

[27] Darius Roman et al. "Machine learning pipeline for battery state-of-health estimation". In: *Nature Machine Intelligence* 3.5 (May 2021), pp. 447–456. ISSN: 2522-5839. DOI: 10.1038/s42256-021-00312-3. URL: https://doi.org/10.1038/s42256-021-00312-3.

[28] J.C. Mason and D.C. Handscomb. *Chebyshev Polynomials*. CRC Press, 2002. ISBN: 9781420036114. URL: https://books.google.de/books?id=g1DMBQAAQBAJ.

[29] Pierre L'Ecuyer and Christiane Lemieux. "Recent Advances in Randomized Quasi-Monte Carlo Methods". In: *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*. New York, NY: Springer US, 2002, pp. 419–474. ISBN: 978-0-306-48102-4. DOI: 10.1007/0-306-48102-2_20.

[30] Art B. Owen. *On dropping the first Sobol' point*. 2021. arXiv: 2008.08051 [math.NA].

[31] Emanouil Atanassov and Sofiya Ivanovska. "On the Use of Sobol' Sequence for High Dimensional Simulation". In: *Computational Science – ICCS 2022*. Cham: Springer International Publishing, 2022, pp. 646–652. ISBN: 978-3-031-08760-8.

[32] Multi-Output GP Emulator's documentation. *Procedure: Generating a Sobol's Sequence*. https://mogp-emulator.readthedocs.io/en/latest/methods/proc/ProcSobolSequence.html. Accessed: 01.2023.

[33] ISEA-team. *ISEA Framework Documentation*. https://isea.pages.rwth-aachen.de/framework/index.html. Accessed: 01.2022.

[34] Mark A. Kramer. "Nonlinear principal component analysis using autoassociative neural networks". In: *AIChE Journal* 37.2 (1991), pp. 233–243. DOI: 10.1002/aic.690370209. URL: https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690370209.

[35] Yasi Wang, Hongxun Yao, and Sicheng Zhao. "Auto-encoder based dimensionality reduction". In: *Neurocomputing* 184 (2016). RoLoD: Robust Local Descriptors for Computer Vision 2014, pp. 232–242. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2015.08.104. URL: https://www.sciencedirect.com/science/article/pii/S0925231215017671.

[36] Ganggang Dong et al. "A Review of the Autoencoder and Its Variants: A Comparative Perspective from Target Recognition in Synthetic-Aperture Radar Images". In: *IEEE Geoscience and Remote Sensing Magazine* 6.3 (2018), pp. 44–68. DOI: 10.1109/MGRS.2018.2853555.

[37] Zhengxue Cheng et al. "Deep Convolutional AutoEncoder-based Lossy Image Compression". In: *2018 Picture Coding Symposium (PCS)*. 2018, pp. 253–257. DOI: 10.1109/PCS.2018.8456308.

[38] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. 2017, pp. 1–6. DOI: 10.1109/ICEngTechnol.2017.8308186.

[39] Sumit Saha. *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way*. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. Accessed: 01.2023.

[40] Diego Unzueta. *Fully Connected Layer vs. Convolutional Layer: Explained*. https://builtin.com/machine-learning/fully-connected-layer. Accessed: 01.2024.

[41] Seifallah Fetni et al. "Capabilities of Auto-encoders and Principal Component Analysis of the reduction of microstructural images; Application on the acceleration of Phase-Field simulations". In: *Computational Materials Science* 216 (2023), p. 111820. ISSN: 0927-0256. DOI: 10.1016/j.commatsci.2022.111820. URL: https://www.sciencedirect.com/science/article/pii/S0927025622005316.

[42] Cortes Corinna and Vapnik Vladimir. "Support-vector networks". In: *Machine Learning* 20 (1995), pp. 273–297. ISSN: 1573-0565. DOI: 10.1007/BF00994018.

[43] Stack Exchange. *Why is the SVM margin equal to* $\frac{2}{\|\mathbf{w}\|}$. https://math.stackexchange.com/questions/1305925/why-is-the-svm-margin-equal-to-frac2-mathbfw. Accessed: 01.2024.

[44] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: A Library for Support Vector Machines*. https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf. 2001.

[45] Alaa Tharwat. "Parameter investigation of support vector machine classifier with kernel functions". In: *Knowledge and Information Systems* 61 (2019), pp. 1269–1302. ISSN: 0219-3116. DOI: 10.1007/s10115-019-01335-4.

[46] Jakub Nalepa and Michal Kawulok. "Selecting training sets for support vector machines: a review". In: *Artificial Intelligence Review* 52 (2019), pp. 857–900. ISSN: 1573-7462. DOI: 10.1007/s10462-017-9611-1.

[47] F. Budiman. "SVM-RBF Parameters Testing Optimization Using Cross Validation and Grid Search to Improve Multiclass Classification". In: *Scientific Visualization* 11 (2019), pp. 80–90. DOI: 10.26583/sv.11.1.07.

[48] Jean-Philippe Pons and Jean-Daniel Boissonnat. "Delaunay Deformable Models: Topology-Adaptive Meshes Based on the Restricted Delaunay Triangulation". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007, pp. 1–8. DOI: 10.1109/CVPR.2007.383019.

[49] Leonidas J. Guibas, Donald E. Knuth, and Micha Sharir. "Randomized incremental construction of Delaunay and Voronoi diagrams". In: *Algorithmica* 7 (1992), pp. 381–413. ISSN: 1432-0541. DOI: 10.1007/BF01758770.

[50] Adam Loy, Lendie Follett, and Heike Hofmann. "Variations of Q–Q Plots: The Power of Our Eyes!" In: *The American Statistician* 70.2 (2016), pp. 202–214. DOI: 10.1080/00031305.2015.1077728.

[51] Anuj shah. *Quantile, Percentile (one tail and two tail distribution), Confidence Interval, Box Plot*. https://medium.com/@anuj_shah/quantile-percentile-one-tail-and-two-tail-distribution-confidence-interval-box-plot-ce80b1b6db37. Accessed: 01.2024.

[52] Rob J. Hyndman and Yanan Fan. "Sample Quantiles in Statistical Packages". In: *The American Statistician* 50.4 (1996), pp. 361–365. DOI: 10.1080/00031305.1996.10473566. URL: https://www.tandfonline.com/doi/abs/10.1080/00031305.1996.10473566.

[53] Emery N. Brown. "Course 14 - Theory of Point Processes for Neural Systems". In: *Methods and Models in Neurophysics*. Ed. by C.C. Chow et al. Vol. 80. Les Houches. Elsevier, 2005, pp. 691–727. DOI: 10.1016/S0924-8099(05)80020-4. URL: https://www.sciencedirect.com/science/article/pii/S0924809905800204.

[54] Daniel B. Carr. "Graphics in the Physical Sciences". In: *Encyclopedia of Physical Science and Technology (Third Edition)*. Ed. by Robert A. Meyers. Third Edition. New York: Academic Press, 2003, pp. 1–14. ISBN: 978-0-12-227410-7. DOI: 10.1016/B0-12-227410-5/00297-0. URL: https://www.sciencedirect.com/science/article/pii/B0122274105002970.

[55] Brian S. Cade and Barry R. Noon. "A gentle introduction to quantile regression for ecologists". In: *Frontiers in Ecology and the Environment* 1.8 (2003), pp. 412–420. DOI: 10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/1540-9295%282003%29001%5B0412%3AAGITQR%5D2.0.CO%3B2.

[56] Yanyan Zhao et al. "A Review on Battery Market Trends, Second-Life Reuse, and Recycling". In: *Sustainable Chemistry* 2.1 (2021), pp. 167–205. ISSN: 2673-4079. URL: https://www.mdpi.com/2673-4079/2/1/11.

[57] Monika Kwiecien et al. "Correct processing of impedance spectra for lead-acid batteries to parameterize the charge-transfer process". In: *J. Appl. Electrochem.* 48 (2018), pp. 885–900. DOI: 10.1007/s10800-018-1217-z.

[58] M. Itagki et al. "*In-situ* EIS to determine impedance spectra of lithium-ion rechargeable batteries during charge and discharge cycle". In: *J. Electroanal. Chem.* 737 (2015), pp. 78–84. DOI: 10.1016/j.jelechem.2014.06.004.

[59] Arash Shafiei, Ahmadreza Momeni, and Sheldon S. Williamson. "Battery modeling approaches and management techniques for Plug-in Hybrid Electric Vehicles". In: *2011 IEEE Vehicle Power and Propulsion Conference*. 2011, pp. 1–5. DOI: 10.1109/VPPC.2011.6043191.

[60] R.R. Gaddam et al. "Review on physical impedance models in modern battery research". In: *Phys. Chem. Chem. Phys.* 23 (2021), pp. 12926–12944. DOI: 10.1039/d1cp00673h.

[61] Christian Fleischer et al. "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". In: *Journal of Power Sources* 260 (2014), pp. 276–291. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2014.01.129. URL: https://www.sciencedirect.com/science/article/pii/S0378775314002249.

[62] Andreas Mertens et al. "Quantitative Analysis of Time-Domain Supported Electrochemical Impedance Spectroscopy Data of Li-Ion Batteries: Reliable Activation Energy Determination at Low Frequencies". In: *Journal of The Electro-*

chemicl Society 163.7 (2016), H521–H527. DOI: 10.1149/2.0511607jes. URL: https://doi.org/10.1149/2.0511607jes.

[63] M. Schönleber et al. "A Consistent Derivation of the Impedance of a Lithium-Ion Battery Electrode and its Dependency on the State-of-Charge". In: *Electrochim. Acta* 243 (2017), pp. 250–259. DOI: 10.1016/j.electacta.2017.05.009.

[64] Andrzej Lasia. *Electrochemical Impedance Spectroscopy and its Applications*. New York: Springer Science+Business Media, 2014.

[65] Xueyuan Wang et al. "A review of modeling, acquisition, and application of lithium-ion battery impedance for onboard battery management". In: *eTransportation* 7 (2021), p. 100093. DOI: 10.1016/j.etran.2020.100093. URL: https://doi.org/10.1016/j.etran.2020.100093.

[66] Alexander Farmann and Dirk Uwe Sauer. "A comprehensive review of on-board State-of-Available-Power prediction techniques for lithium-ion batteries in electric vehicles". In: *Journal of Power Sources* 329 (2016), pp. 123–137. ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2016.08.031. URL: https://www.sciencedirect.com/science/article/pii/S0378775316310308.

[67] Alexander Farmann and Dirk Uwe Sauer. "Comparative study of reduced order equivalent circuit models for on-board state-of-available-power prediction of lithium-ion batteries in electric vehicles". In: *Applied Energy* 225 (2018), pp. 1102–1122. DOI: 10.1016/j.apenergy.2018.05.066. URL: https://doi.org/10.1016/j.apenergy.2018.05.066.

[68] Yang Li et al. "A physics-based distributed-parameter equivalent circuit model for lithium-ion batteries". In: *Electrochimica Acta* 299 (2019), pp. 451–469. ISSN: 0013-4686. DOI: https://doi.org/10.1016/j.electacta.2018.12.167. URL: https://www.sciencedirect.com/science/article/pii/S0013468618328883.

[69] M. Schönleber and E. Ivers-Tiffée. "Approximability of impedance spectra by RC elements and implications for impedance analysis". In: *Electrochemistry Communications* 5 (June 2015), pp. 15–19. DOI: 10.1016/j.elecom.2015.05.018. URL: http://dx.doi.org/10.1016/j.elecom.2015.05.018.

[70] Leo Wildfeuer, Philipp Gieler, and Alexander Karger. "Combining the Distribution of Relaxation Times from EIS and Time-Domain Data for Parameterizing Equivalent Circuit Models of Lithium-Ion Batteries". In: *Batteries* 7 (2021), p. 52. DOI: 10.3390/batteries7030052. URL: https://doi.org/10.3390/batteries7030052.

[71] Andreas Mertens et al. "Superionic bulk conductivity in $Li_{1.3}Al_{0.3}Ti_{1.7}(PO_4)_3$ solid electrolyte". In: *Solid State Ionics* 309 (2017), pp. 180–186. DOI: 10.1016/j.ssi.2017.07.023. URL: https://doi.org/10.1016/j.ssi.2017.07.023.

[72] J.R. Macdonald, J. Schoonman, and A.P. Lehnen. "The applicability and power of complex nonlinear least squares for the analysis of impedance and admittance data". In: *J. Electroanal. Chem.* 131 (1982), pp. 77–95. DOI: 10.1016/0022-0728(82)87062-9. URL: ttps://doi.org/10.1016/0022-0728(82)87062-9.

[73] P. Zoltowski. "The error function for fitting of models to immittance data". In: *J. Electroanal. Chem.* 178 (1984), pp. 11–19.

[74] J. Ross Macdonald and William J. Thompson. "Strongly heteroscedastic nonlinear-regression". In: *Commun. Statist. Simula.* 20.4 (1991), pp. 843–886. DOI: 10.1080/03610919108812987. URL: https://doi.org/10.1080/03610919108812987.

[75] Mark E. Orazem, Pavan Shukla, and Michael A. Membrino. "Extension of the measurement model approach for deconvolution of underlying distributions for impedance measurements". In: *Electrochimica Acta* 47 (2002), pp. 2027–2034.

[76] Ting Hei Wan et al. "Influence of the Discretization Methods on the Distribution of Relaxation Times Deconvolution: Implementing Radial Basis Functions with DRTtools". In: *Electrochimica Acta* 184 (2015), pp. 483–499. ISSN: 0013-4686. DOI: 10.1016/j.electacta.2015.09.097. URL: https://www.sciencedirect.com/science/article/pii/S0013468615305090.

[77] Amin Moradpour et al. "Measurement Uncertainty in Battery Electrochemical Impedance Spectroscopy". In: *IEEE Trans. Instrum. Meas.* 71 (2022), p. 1006209.

[78] P. Keil et al. "Calendar Aging of Lithium-Ion Batteries". In: *J. Electrochem. Soc.* 163 (2016), A1872–A1880. DOI: 10.1149/2.0411609jes.

[79] Digby D. Macdonald. "Reflections on the history of electrochemical impedance spectroscopy". In: *Electrochimica Acta* 51.8 (2006). Electrochemical Impedance Spectroscopy, pp. 1376–1388. ISSN: 0013-4686. DOI: 10.1016/j.electacta.2005.02.107. URL: https://www.sciencedirect.com/science/article/pii/S0013468605007991.

[80] Alessio De Angelis et al. "Uncertainty Characterization of a Practical System for Broadband Measurement of Battery EIS". In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–9. DOI: 10.1109/TIM.2022.3156994.

[81] Giulio C. Borgia, Robert J.S. Brown, and Paola Fantazzini. "Uniform-Penalty Inversion of Multiexponential Decay Data". In: *Journal of Magnetic Resonance* 132 (May 1998), pp. 65–77. DOI: 10.1006/jmre.1998.1387. URL: https://doi.org/10.1006/jmre.1998.1387.

[82] J. Ross Macdonald. "Generalizations of "universal dielectric response" and a general distribution-of-activation-energies model for dielectric and conducting systems". In: *J. Appl. Phys.* 58.5 (1985), pp. 1971–1978. DOI: 10.1063/1.336004. URL: https://doi.org/10.1063/1.336004.

[83] Andreas Mertens and Josef Granwehr. "Two-dimensional impedance data analysis by the distribution of relaxation times". In: *Journal of Energy Storage* 13 (Oct. 2017), pp. 401–408. URL: http://dx.doi.org/10.1016/j.est.2017.07.029.

[84] Bernard A. Boukamp and Aurélie Rolle. "Analysis and Application of Distribution of Relaxation Times in Solid State Ionics". In: *Solid State Ionics* 302 (2017), pp. 12–18. DOI: 10.1016/j.ssi.2016.10.009.

[85] U. Tröltzsch and O. Kanoun. "Implementierung der Impedanzspektroskopie in vollautomatischen Messsystemen am Beispiel der Batteriezustandsdiagnose (Battery Diagnosis as an Example for Implementing Impedance Spectroscopy in Automated Measurement Systems)". In: *Tech. Mess.* 73 (2006), pp. 393–403. DOI: 10.1524/teme.2006.73.7-8.393.

[86] Pankaj Agarwal, Mark E. Orazem, and Luis H. Garcia-Rubio. "Application of Measurement Models to Impedance Spectroscopy: III . Evaluation of Consistency with the Kramers-Kronig Relations". In: *J. Electrochem. Soc.* 142 (1995), pp. 4159–4168. URL: https://doi.org/10.1149/1.2048479.

[87] Per Christian Hansen. *Rank-Deficit and Discrete Ill-Posed Problems*. Philadelphia: SIAM Society for Industrial and Applied Mathematics, 1998.

[88] Josef Granwehr and Peter J. Roberts. "Inverse Laplace Transform of Multidimensional Relaxation Data Without Non-Negativity Constraint". In: *Journal of Chemical Theory and Computation* 8.10 (2012). PMID: 26592997, pp. 3473–3482. DOI: 10.1021/ct3001393. eprint: https://doi.org/10.1021/ct3001393. URL: https://doi.org/10.1021/ct3001393.

[89] Luc H. J. Raijmakers et al. "Non-Zero Intercept Frequency: An Accurate Method to Determine the Integral Temperature of Li-Ion Batteries". In: *IEEE Transactions on Inductrial Electronics* 63 (2016), pp. 3168–3178. DOI: 10.1109/TIE.2016.2516961.

[90] A. Schiefer, Marcel Heinzmann, and André Weber. "Inductive Low-Frequency Processes in PEMFC-Impedance Spectra". In: *Fuel Cells* 20 (Mar. 2020). DOI: 10.1002/fuce.201900212.

[91] M. D. Huerlimann and L. Venkataramanan. "Quantitative Measurement of Two-Dimensional Distribution Functions of Diffusion and Relaxation in Grossly Inhomogeneous Fields". In: *Journal of Magnetic Resonance* 157 (2002), 31–42. URL: https://doi.org/10.1006/jmre.2002.2567.

[92] James R. Macdonald. "Impedance spectroscopy: old problems and new developments". In: *Electrochim. Acta* 35 (1990), pp. 1483–1492. DOI: 10.1016/0013-4686(90)80002-6. URL: http://dx.doi.org/10.1016/0013-4686(90)80002-6.

[93] R.J.S. Brown. "Information Available and Unavailable from Multiexponential Relaxation Data". In: *J. Magn. Reson.* 82 (1989), pp. 539–561.

[94] P. Agarwal et al. "Application of measurement models to impedance spectroscopy. II. Determination of the stochastic contribution to the error structure". In: *J. Electrochem. Soc.* 142 (1995), pp. 4149–4158. DOI: 10.1149/1.2048478.

[95] Taeryon Choi. "Bayesian Regression with Gaussian Processes : A Tutorial". In: *Proceedings for the Spring Conference*. The Korean Statistical Society, 2011.

[96] Carl E. Rasmussen and Christopher K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. URL: http://www.gaussianprocess.org/gpml/chapters/.

[97] Thomas Heil and Andreas Jossen. "Continuous approximation of the ZARC element with passive components". In: *Measurement Science and Technology* 32.10 (July 2021), p. 104011. DOI: 10.1088/1361-6501/ac0466. URL: https://doi.org/10.1088/1361-6501/ac0466.

[98] Heiko Witzenhausen. "GUI for Fitting Models to measured Impedance Spectra with focus on Lithium Ion Batteries". In: (2022). URL: https://github.com/HWitz/FittingGUI.

[99] Bernard A. Boukamp. "Derivation of a Distribution Function of Relaxation Times for the (fractal) Finite Length Warburg". In: *Electrochim. Acta* 252 (2017), pp. 154–163. DOI: 10.1016/j.electacta.2017.08.154.

[100] Emanuele Quattrocchi et al. "A general model for the impedance of batteries and supercapacitors: The non-linear distribution of diffusion times". In: *Electrochimica Acta* 324 (2019), p. 134853. DOI: 10.1016/j.electacta.2019.134853.

[101] B.A. Boukamp. "Fourier transform distribution function of relaxation times; application and limitations". In: *Electrochim. Acta* 154 (2015), pp. 35–46. DOI: 10.1016/j.electacta.2014.12.059.

[102] Helge Schichlein. "Experimentelle Modellbildung für die Hochtemperatur-Brennstoffzelle SOFC". PhD Thesis. Universität Karlsruhe (TH), Mar. 2003. ISBN: 3-86130-229-2.

[103] Jan Hendrik Metzen and Guillaume Lemaitre. *Gaussian process regression (GPR) with noise-level estimation*. https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_noisy.html. Accessed on Jan 9, 2022.

[104] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[105] Manuel Kasper et al. "Electrochemical impedance spectroscopy error analysis and round robin on dummy cells and lithium-ion-batteries". In: *Journal of Power Sources* 536 (2022), p. 231407. DOI: 10.1016/j.jpowsour.2022.231407.

[106] M. Gulian, A. Frankel, and L. Swiler. "Gaussian process regression constrained by boundary value problems". In: *Computer Methods in Applied Mechanics and Engineering* 388 (2022), p. 114117. DOI: https://doi.org/10.1016/j.cma.2021.114117. URL: https://www.sciencedirect.com/science/article/pii/S0045782521004485.

[107] G. Spinolo et al. "Error analysis and choice of conflicting models in impedance spectroscopy". In: *Solid State Ionics* 28–30 (1988), pp. 1602–1606.

[108] Joshua Thomas Jameson Burd et al. "Improvements in electric vehicle battery technology influence vehicle lightweighting and material substitution decisions". In: *Applied Energy* 283 (2021), p. 116269. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2020.116269. URL: https://www.sciencedirect.com/science/article/pii/S0306261920316597.

[109] Yujie Wang et al. "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems". In: *Renewable and Sustainable Energy Reviews* 131 (2020), p. 110015. ISSN: 1364-0321. DOI: 10.1016/j.rser.2020.110015. URL: https://www.sciencedirect.com/science/article/pii/S1364032120303063.

[110] Malin Andersson et al. "Parametrization of physics-based battery models from input–output data: A review of methodology and current research". In: *Journal of Power Sources* 521 (2022), p. 230859. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2021.230859. URL: https://www.sciencedirect.com/science/article/pii/S0378775321013458.

[111] S. Tamilselvi et al. "A Review on Battery Modelling Techniques". In: *Sustainability* 13.18 (2021). ISSN: 2071-1050. DOI: 10.3390/su131810042. URL: https://www.mdpi.com/2071-1050/13/18/10042.

[112] Wendi Guo et al. "Review of "grey box" lifetime modeling for lithium-ion battery: Combining physics and data-driven methods". In: *Journal of Energy Storage* 56 (2022), p. 105992. ISSN: 2352-152X. DOI: 10.1016/j.est.2022.105992. URL: https://www.sciencedirect.com/science/article/pii/S2352152X22019806.

[113] Wenwen Wang et al. "Application of Digital Twin in Smart Battery Management Systems". In: *Chinese Journal of Mechanical Engineering* 34.1 (June 2021), p. 57. ISSN: 2192-8258. DOI: 10.1186/s10033-021-00577-0. URL: https://doi.org/10.1186/s10033-021-00577-0.

[114] Yibo Guo et al. "Insight into fast charging/discharging aging mechanism and degradation-safety analytics of 18650 lithium-ion batteries". In: *Journal of Energy Storage* 72 (2023), p. 108331. ISSN: 2352-152X. DOI: 10.1016/j.est.2023.108331. URL: https://www.sciencedirect.com/science/article/pii/S2352152X23017280.

[115] Zuolu Wang et al. "A review on online state of charge and state of health estimation for lithium-ion batteries in electric vehicles". In: *Energy Reports* 7 (2021), pp. 5141–5161. ISSN: 2352-4847. DOI: 10.1016/j.egyr.2021.08.113. URL: https://www.sciencedirect.com/science/article/pii/S2352484721007150.

[116] Yu Yang et al. "Towards a safer lithium-ion batteries: A critical review on cause, characteristics, warning and disposal strategy for thermal runaway". In: *Advances in Applied Energy* 11 (2023), p. 100146. ISSN: 2666-7924. DOI: 10.1016/j.adapen.2023.100146. URL: https://www.sciencedirect.com/science/article/pii/S2666792423000252.

[117] N. Karjanto. *Properties of Chebyshev polynomials*. 2020. arXiv: 2002.01342 [math.HO].

[118] Arthur T. Benjamin et al. "Combinatorial trigonometry with Chebyshev polynomials". In: *Journal of Statistical Planning and Inference* 140.8 (2010). Lattice Path Combinatorics and Applications, pp. 2157–2160. ISSN: 0378-3758. DOI: 10.1016/j.jspi.2010.01.011. URL: https://www.sciencedirect.com/science/article/pii/S0378375810000224.

[119] I.M Sobol'. "On the distribution of points in a cube and the approximate evaluation of integrals". In: *USSR Computational Mathematics and Mathematical Physics* 7.4 (1967), pp. 86–112. ISSN: 0041-5553. DOI: 10.1016/0041-5553(67)90144-9. URL: https://www.sciencedirect.com/science/article/pii/0041555367901449.

[120] Taufiq Alif Kurniawan et al. "Real time monitoring state-of-charge battery using internal resistance measurements for remote applications". In: *Journal of Physics: Conference Series* 1528.1 (Apr. 2020), p. 012034. DOI: 10.1088/1742-6596/1528/1/012034. URL: https://dx.doi.org/10.1088/1742-6596/1528/1/012034.

[121] Dian Wang, Yun Bao, and Jianjun Shi. "Online Lithium-Ion Battery Internal Resistance Measurement Application in State-of-Charge Estimation Using the Extended Kalman Filter". In: *Energies* 10.9 (2017). ISSN: 1996-1073. DOI: 10.3390/en10091284. URL: https://www.mdpi.com/1996-1073/10/9/1284.

[122] Sagar Vashisht et al. "Thermal behaviour of Li-ion battery: An improved electrothermal model considering the effects of depth of discharge and temperature". In: *Journal of Energy Storage* 70 (2023), p. 107797. ISSN: 2352-152X. DOI: 10.1016/j.est.2023.107797. URL: https://www.sciencedirect.com/science/article/pii/S2352152X23011945.

[123] Valentin Sulzer et al. "The challenge and opportunity of battery lifetime prediction from field data". In: *Joule* 5.8 (2021), pp. 1934–1955. ISSN: 2542-4351. DOI: 10.1016/j.joule.2021.06.005. URL: https://www.sciencedirect.com/science/article/pii/S2542435121002932.

[124] Hidenori Ide and Takio Kurita. "Improvement of learning for CNN with ReLU activation by sparse regularization". In: *2017 International Joint Conference on Neural Networks (IJCNN)*. 2017, pp. 2684–2691. DOI: 10.1109/IJCNN.2017.7966185.

[125] Weijun li and Zhenyu Liu. "A method of SVM with Normalization in Intrusion Detection". In: *Procedia Environmental Sciences* 11 (2011). 2011 2nd International Conference on Challenges in Environmental Science and Computer Engineering (CESCE 2011), pp. 256–262. ISSN: 1878-0296. DOI: 10.1016/j.proenv.

2011.12.040. URL: https://www.sciencedirect.com/science/article/pii/S1878029611008632.

[126] Muhammad Asrol, Petir Papilo, and Fergyanto E Gunawan. "Support Vector Machine with K-fold Validation to Improve the Industry's Sustainability Performance Classification". In: *Procedia Computer Science* 179 (2021). 5th International Conference on Computer Science and Computational Intelligence 2020, pp. 854–862. ISSN: 1877-0509. DOI: 10.1016/j.procs.2021.01.074. URL: https://www.sciencedirect.com/science/article/pii/S1877050921001046.

[127] *Pytorch Documentation*. https://pytorch.org/docs/stable/index.html. accessed on Nov., 2023.

[128] Jussi Sihvo et al. "Fast Approach for Battery Impedance Identification Using Pseudo-Random Sequence Signals". In: *IEEE Transactions on Power Electronics* 35.3 (2020), pp. 2548–2557. DOI: 10.1109/TPEL.2019.2924286.

[129] Keith Godfrey. *Perturbation signals for system identification*. Prentice Hall International (UK) Ltd., 1993.

[130] Jaber A. Abu Qahouq and Zhiyong Xia. "Single-Perturbation-Cycle Online Battery Impedance Spectrum Measurement Method With Closed-Loop Control of Power Converter". In: *IEEE Transactions on Industrial Electronics* 64.9 (2017), pp. 7019–7029. DOI: 10.1109/TIE.2017.2686324.

[131] David A. Howey et al. "Online Measurement of Battery Impedance Using Motor Controller Excitation". In: *IEEE Transactions on Vehicular Technology* 63.6 (2014), pp. 2557–2566. DOI: 10.1109/TVT.2013.2293597.

[132] Mamta Rani, Riaz ul Haq, and Avinash Diwedi. "Variants of Koch curve: A Review". In: 2012.

[133] Z Ren and G G Zhu. "Pseudo-random binary sequence closed-loop system identification error with integration control". In: *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 223.6 (2009), pp. 877–884. DOI: 10.1243/09596518JSCE794. eprint: https://doi.org/10.1243/09596518JSCE794.

[134] Dominik Schneider et al. "Enhanced State and Parameter Estimation within Reconfigurable Battery Systems for Electric Vehicles". In: *2022 12th International Conference on Power, Energy and Electrical Engineering (CPEEE)*. 2022, pp. 71–77. DOI: 10.1109/CPEEE54404.2022.9738700.

[135] Pierre Legendre and Louis Legendre. "Chapter 13 - Spatial analysis". In: *Numerical Ecology*. Ed. by Pierre Legendre and Louis Legendre. Vol. 24. Developments in Environmental Modelling. Elsevier, 2012, pp. 785–858. DOI: 10.1016/B978-0-444-53868-0.50013-7. URL: https://www.sciencedirect.com/science/article/pii/B9780444538680500137.

[136] M. B. WILK and R. GNANADESIKAN. "Probability plotting methods for the analysis for the analysis of data". In: *Biometrika* 55.1 (Mar. 1968), pp. 1–17. ISSN: 0006-3444. DOI: 10.1093/biomet/55.1.1. eprint: https://academic.oup.com/biomet/article-pdf/55/1/1/730568/55-1-1.pdf.

[137] Nina Meddings et al. "Application of electrochemical impedance spectroscopy to commercial Li-ion cells: A review". In: *Journal of Power Sources* 480 (2020), p. 228742. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2020.228742. URL: https://www.sciencedirect.com/science/article/pii/S0378775320310466.

[138] Christof Backhaus. *Smart Battery Management: Optimization of electrochemical impedance spectroscopy for machine learning based battery management*. 2023.