UNIVERSITÀ DEGLI STUDI DI CASSINO E DEL LAZIO MERIDIONALE

DIPARTIMENTO DI INGEGNERIA ELETTRICA E DELL'INFORMAZIONE
"MAURIZIO SCARANO"

Dottorato di Ricerca in Metodi, Modelli e Tecnologie per l'Ingegneria

Optimized data-driven measurement methods and devices for modeling diagnosis and prediction of batteries behavior

Tutors:

Prof: Luigi Ferrigno

Prof: Domenico Capriglione

Candidate:

Carmine Bourelly

XXXVI cycle

Contents

List of Figures

Introduction

Batteries are playing an increasingly central role in today's society, driven by national and supranational entities as a means to facilitate the energy transition from fossil fuels to renewable sources, reducing emissions and pollution while aiming to enhance the quality of life in cities. An example of this push is the European Union's mandate that, by 2035, all vehicles sold in the European common market must be battery-powered. Simultaneously, the transition to an economy no longer reliant on fossil fuels is occurring, particularly in large-scale energy production, where renewable resources are being incentivized due to their intermittent energy supply nature, making energy storage devices crucial.

Redox reactions-based batteries currently stand as the most efficient system for energy storage. They provide a higher energy density compared to any other available storage method. However, electrochemical cells are not without flaws, and improper management can lead to an increased risk of accidents with potentially tragic consequences. In this context, it is vital to develop methods that ensure proper battery operation, effective diagnostic tools, and predictive modeling. The topics addressed by predictive diagnostics align perfectly with the needs of electric energy storage systems. Notably, predictive diagnostics have gained significant traction in recent years, becoming a part of Industry 4.0, representing the latest industrial revolution focused on incorporating new production technologies to improve working conditions, increase the interconnection of production systems, and enhance plant resilience. Predictive diagnostics play a pivotal role in each of these aspects and are, therefore, fundamental within the framework of Industry 4.0.

Predictive diagnostics, also known as predictive maintenance, is defined by two fundamental components: "data acquisition" and "condition monitoring." The "data acquisition" process encompasses all the tools and systems used to acquire knowledge about the system. Without adequate measurement of the phenomenon that accurately reflects reality, any subsequent action is unlikely to succeed and can even be detrimental.

The thesis work conducted spans both areas characterizing predictive maintenance.

In the realm of measurements on electrochemical cells, data acquisition involves sensor-based systems, measuring intrinsic parameters (such as voltage, current, temperature, and internal impedance) and derived parameters (such as State of Charge and State of Health). This thesis addresses the issue of data quality, seeking methods that not only provide an understanding of the measured data but also its associated quality. Specifically, a measurement system for battery impedance was created, capable of stimulating by injecting an alternating current and measuring both the current and voltage to derive impedance. The system's flexibility allowed for the study of various signal forms (stepped sine and multisine). An analysis of processing was also carried out, introducing sinusoidal fitting as a method to overcome the time constraints dictated by Fast Fourier Transform (FFT). The effectiveness of the created measurement system led to the development of a distributed embedded measurement system designed specifically for UPS (Uninterruptible Power Supply).

"Condition monitoring" can be divided into two significant sectors: Diagnosis, which involves understanding the current state of the battery through data acquisition systems, and

Prognosis, which entails forecasting the future state of the battery. In this thesis, diagnosis is addressed through the development of a methodology for optimizing measurement times in estimating the state of charge of lithium iron phosphate batteries.

Knowledge of the state of charge is crucial not only for providing users with information about system autonomy but also for supplying the Battery Management System (BMS) with insights into the state of individual cells and enabling actions to protect battery modules. Currently, state of charge estimation relies on techniques such as Coulomb Counting, which are based on prior knowledge of the state of charge and battery capacity. Impedance measurement has excellent potential to surpass the limits imposed by current methodologies but has the significant drawback of requiring extended measurement times, especially when low frequencies are used (typically, these systems have stimulus frequencies ranging from 10 mHz to 10 kHz).

A Support Vector Machine (SVM) is trained to estimate the state of charge based on impedance spectroscopy measurements within a genetic algorithm framework designed to reduce the number of stimulus frequencies for measurement time optimization. The results obtained enable a reduction in measurement time while preserving the accuracy of state of charge estimation. In the context of diagnosis, a preliminary study of non-destructive testing techniques for batteries has been conducted. In particular, a correlation between state of charge and the inductance of an inductor, printed on a PCB, attached to the side of a VRLA lead-acid battery, has been established. The proposed system is intriguing as it allows state of charge estimation without any galvanic connection to the battery.

Once the current battery state is determined, the challenge is predicting the future state, i.e., the so-called prognosis. Prognosis cannot be separated from system modeling. Initially, the focus shifted away from batteries to model electric loads using state-based algorithms based on Markov Chain. Each device was characterized metrologically and then modeled in a series of states that vary through coefficients listed in a probability matrix. The result is a synthetic dataset capable of emulating the simultaneous use of multiple loads and allowing the training of algorithms for recognizing devices connected to an electrical network by measuring the electrical signature of the system. In the case of batteries, the modeling concentrated on the parametric identification of equivalent electrical models (ECMs) with the aim of predicting terminal voltage knowing the state of charge. The problem was approached in two ways. The first, already known in the literature, involves pulsed characterization tests in which batteries are discharged with current pulses, and the voltage response provides information about the ECM. The second method uses a genetic algorithm to provide a closed form of a function that relates voltage to state of charge. The pulsed approach has the major disadvantage of requiring specific characterization with precision instruments capable of discharging the battery while measuring voltage and current at a high sampling frequency. The approach based on Genetic Programming (GP) uses a simpler characterization to train the algorithm and also provides a closed form output that can be easily integrated into a BMS. The system can be constrained to provide an invertible solution, meaning it can provide the battery's state of charge based on terminal voltage as input. In conclusion, this thesis work has covered all the distinctive aspects of predictive maintenance by providing methodologies, defining and refining hardware tools, and creating datasets useful for the development of software tools.

Chapter 1

Batteries

1.1 What is a secondary battery

Today, batteries are one of the most popular ways of storing electrical energy by converting it into chemical energy. In fact, all batteries fall into the category of electrochemical cells. They consist of two electrodes, anode and cathode, an electrolyte and sometimes an ion separator [133]. Primary batteries are those in which the conversion of chemical energy into electrical energy is not reversible, i.e. once all the material of which the battery is composed has participated in the reaction, there is no way of returning it to its initial state by adding energy from outside. On the other hand, secondary batteries are those that use chemical elements that allow the elements that make up the battery to ideally return to their initial state by adding energy from outside. This feature allows secondary batteries to have more than one discharge cycle, making them extremely versatile and environmentally friendly. In fact, the extreme popularity of secondary batteries has limited primary batteries to those applications where energy demand is extremely low.

1.2 Working Principle

Batteries are a particular subgroup of electrochemical cells. An electrochemical cell is a device capable of converting chemical energy into electricity or vice versa if the reactions allow it. To do this, they use reactions called oxidation-reduction reactions in which electrons naturally pass from an element that is oxidized to one that is reduced. Equation 1.1 and Equation 1.2 show the two half-reactions of oxidation Equation 1.1 and reduction Equation 1.2 of a Zinc-Copper battery, known as a Daniell battery. Zinc is oxidized by losing two electrons, while copper is reduced by gaining two electrons. In the overall reaction, Equation 1.3, the number of electrons lost by the Zinc balances the number of electrons gained by the copper.

$$Zn(s) \Rightarrow Zn^{2+}(aq) + 2e^{-} \tag{1.1}$$

$$Cu^{2+} + 2e^{-} \Rightarrow Cu(s) \tag{1.2}$$

$$Zn(s) + Cu^{2+}(aq) \Rightarrow Cu(s) + Zn^{2+}(aq) \tag{1.3}$$

In batteries, the two half-reactions are kept separate and the exchanging electrons are forced to pass outside the battery, thus providing work in the form of electrical energy. The battery voltage is dependent on the elements used. The elements used as reactants are not

chosen at random but are derived from the scale of electrochemical reduction potentials, in which, each element has a specific voltage which is nothing other than the voltage of a battery in which one of the two electrodes is the element in question while the other electrode is a reference electrode. Typically, the hydrogen electrode is chosen as the reference electrode to construct the scale. This particular electrode consists of a platinum foil immersed in a 1Molar solution of hydrogen ions in which hydrogen gas is bubbled at 1 atmosphere.

Figure 1.1 [107] shows the voltage values for each element. Given two different elements, the battery voltage will be the difference between the voltage of the element chosen as the cathode and the voltage of the element chosen as the anode. There are many pairs of elements capable of generating an oxidation-reduction reaction, and the greater the distance between the elements in the greater the difference in battery potential will be. For example, Nickel-Cadmium batteries have a terminal voltage of 1.2 V while LFPs have 3.7 V. Sometimes individual cells can be connected in series and/or in parallel to increase the output voltage and/or current of the battery.

Table I.2.1 Standard potentials of electrode reactions[a]

Half cell	Electrode reaction	$E^{\ominus}$ (V vs. SHE)
$Li \mid Li^+$	$Li^+ + e^- \rightleftarrows Li$	−3.040
$Rb \mid Rb^+$	$Rb^+ + e^- \rightleftarrows Rb$	−2.924
$K \mid K^+$	$K^+ + e^- \rightleftarrows K$	−2.924
$Cs \mid Cs^+$	$Cs^+ + e^- \rightleftarrows Cs$	−2.923
$Ca \mid Ca^{2+}$	$Ca^{2+} + 2\,e^- \rightleftarrows Ca$	−2.76
$Na \mid Na^+$	$Na^+ + e^- \rightleftarrows Na$	−2.713
$Mg \mid Mg^{2+}$	$Mg^{2+} + 2\,e^- \rightleftarrows Mg$	−2.375
$Al \mid Al^{3+}$	$Al^{3+} + 3\,e^- \rightleftarrows Al$	−1.706
$Zn \mid Zn^{2+}$	$Zn^{2+} + 2\,e^- \rightleftarrows Zn$	−0.7628
$Pt \mid Cr^{2+}, Cr^{3+}$	$Cr^{3+} + e^- \rightleftarrows Cr^{2+}$	−0.41
$Fe \mid Fe^{2+}$	$Fe^{2+} + 2\,e^- \rightleftarrows Fe$	−0.409
$Cd \mid Cd^{2+}$	$Cd^{2+} + 2\,e^- \rightleftarrows Ni$	−0.4026
$Ni \mid Ni^{2+}$	$Ni^{2+} + 2\,e^- \rightleftarrows Ni$	−0.23
$Pb \mid Pb^{2+}$	$Pb^{2+} + 2\,e^- \rightleftarrows Pb$	−0.1263
$Pt \mid H_2,\ H^+_{aq}$	$2H^+ + 2\,e^- \rightleftarrows H_2$	0.0000
$Pt \mid Cu^{2+},\ Cu^+$	$Cu^{2+} + e^- \rightleftarrows Cu^+$	+0.167
$Cu^{2+} \mid Cu$	$Cu^{2+} + 2\,e^- \rightleftarrows Cu$	+0.3402
$Pt \mid [Fe(CN)_6]^{3-}, [Fe(CN)_6]^{4-}$	$[Fe(CN)_6]^{3-} + e^- \rightleftarrows [Fe(CN)_6]^{4-}$	+0.356
$Pt \mid [W(CN)_8]^{3-}, [W(CN)_8]^{4-}$	$[W(CN)_8]^{3-} + e^- \rightleftarrows [W(CN)_8]^{4-}$	+0.457
$Pt \mid [Mo(CN)_8]^{3-}, [Mo(CN)_8]^{4-}$	$[Mo(CN)_8]^{3-} + e^- \rightleftarrows [Mo(CN)_8]^{4-}$	+0.725
$Ag \mid Ag^+$	$Ag^+ + e^- \rightleftarrows Ag$	+0.7996
$2\,Hg \mid Hg_2^{2+}$	$Hg_2^{2+} + 2\,e^- \rightleftarrows 2\,Hg$	+0.7961
$Pt \mid Cr_2O_7^{2-}, Cr^{3+}, H^+$	$Cr_2O_7^{2-} + 14\,H^+ + 6\,e^- \rightleftarrows 2\,Cr^{3+} + 7H_2O$	+1.36
$Pt \mid O_2,\ H_2O, H^+$	$1/2\,O_2 + 2H^+ + 2\,e^- \rightleftarrows H_2O$	+1.229
$Au^+ \mid Au$	$Au^+ + e^+ \rightleftarrows Au$	+1.83
$Pt \mid MnO_4^-,\ Mn^{2+},\ H^+$	$MnO_4^- + 8H^+ + 5\,e^- \rightleftarrows Mn^{2+} + 4H_2O$	+1.491
$Pt \mid H_4XeO_6,\ XeO_3$	$H_4XeO_6 + 2\,H^+ + 2\,e^- \rightleftarrows XeO_3 + 3\,H_2O$	+2.42
$Pt \mid F_2,\ F^-$	$F_2 + 2\,e^- \rightleftarrows 2\,F$	+2.866

Figure 1.1: Standard potentials of electrode reactions [107]

1.3 Battery typologies

Many electronic devices produced in recent times integrate a battery inside them; the choice of the type of battery to be adopted depends largely on design constraints such as cost, power output, energy capacity, device weight, output voltage, usage environment, etc. Figure 1.2 extracted from [42] shows some of the most widely adopted battery types in the industry. In recent years, lithium-based batteries have gradually replaced other technologies. Lead-acid-based batteries still resist, as they remain economical to produce compared with lithium batteries of equivalent capacity and discharge power.

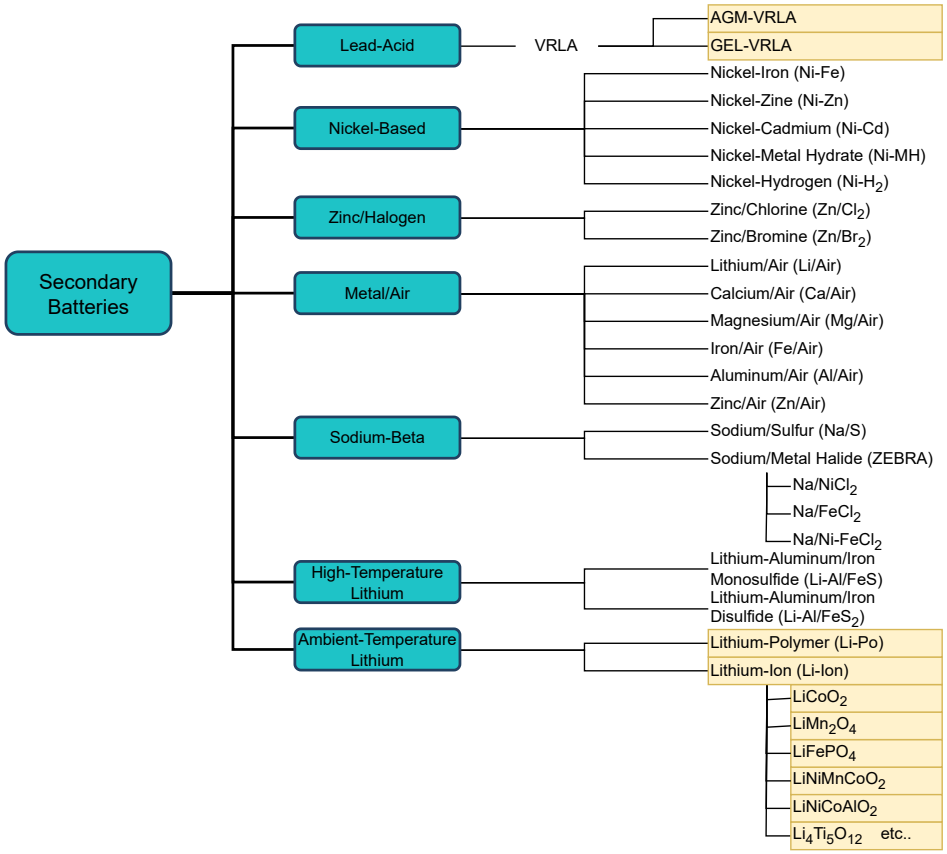


Figure 1.2: Secondary Batteries

1.3.1 Lead Acid Batteries

Since World War II, lead-acid batteries have been the only or at least the most widely used technology for storing electrical energy using a reversible reaction. The development of this technology is linked to the development of the automobile, transportation in general, and telecommunications. The two electrodes of the lead acid battery are composed of Lead (negative electrode) and Lead Dioxide (Positive electrode)

$$Pb + PbO_2 + 2H_2SO_4 \rightleftharpoons 2PbSO_4 + 2H_2O \tag{1.4}$$

Unlike other types of batteries, lead acid batteries have the special feature of having the electrolyte actively participate in the oxidation-reduction reaction. Equation 1.4 reports the overall reaction that takes place inside the cell. On the negative electrode the oxidation of lead

($Pb- > Pb^{2+} + 2e^-$) takes place, releasing two electrons. On the positive electrode there is reduction of lead dioxide ($Pb^{4+} + 2e^- - > Pb^{2+}$). If the two electrodes are immersed in a solution of H_2SO_4 lead sulfate is generated from both of the two sub-reactions on both electrodes. Also on the positive electrode, during charging, there is as sub reaction the production of hydrogen that remains mostly dissolved in aqueous solution but can be released as a gas. The equilibrium potential of the $Pb/PbSO_4$ electrode, compared with the reference potential of the hydrogen electrode, is very negative. In contrast, the $PbO_2/PbSO_4$ electrode has a positive equilibrium potential, again compared with the hydrogen electrode. By composing the two electrodes, a relatively high cell voltage is achieved, around 2V per cell. However, the cell voltage is dependent on the concentration of $H2SO_4$ in the electrolyte [96]. Since sulfuric acid participates in the reaction its concentration of the electrolyte depends on how much the battery has been discharged, this implies that the voltage at the cell terminals varies as the remaining charge of the battery varies. When the voltage at the cell terminals reaches $1.7 - 1.8V$ the cell is called discharged.

Batteries exploiting the Equation 1.4 reaction are mainly of two types:

- **Flooded batteries**, typically adopted in internal combustion engines as SLI (starting, lighting, ignition) starters, however, require regular maintenance.
- **Maintenance-free batteries** with a large electrolyte excess, containing calcium or antimony in addition to lead, which help to have less water loss. In addition to Evaporation Loss (Temperature-dependent) most of the water loss is through electrolysis during charging, over-charging and self-discharge of the battery VRLABs (Valve Regulated Lead Acid Batteries) are also sealed batteries, but they have a valve capable of expelling gases that may form inside the battery itself.
  - **AGM-VRLA**: these are batteries in which there is a separator, usually made of fiberglass, between the two electrodes that absorbs the electrolyte. This expedient allows the electrolyte to appear solid and avoid leakage. The concentration of sulfuric acid in the electrolyte is higher than in flooded batteries.
  - **GEL-VRLA**: are batteries that have a similar principle to AGMs but instead of absorbing the electrolyte in a mat, in GELs the electrolyte is dispersed in an extremely viscous gel. Compared with AGMs they have lower performance and generally lower durability

Defining "Coulomb efficiency" as the ratio of the amount of charge that can be extracted from a charged battery to the amount of charge fed into the battery, lead-acid batteries achieve an efficiency of around 85%, the efficiency, however, is closely related to the output current of the battery. In Figure 1.3 you can see how the energy delivered by the battery decreases as the output current increases.

The main causes of deterioration of lead-acid batteries are [77]:

- **Positive-Plate Expansion**: as shown above, on the positive electrode there is a lead dioxide reduction reaction that has lead sulfate as a product. This occupies a larger volume than the lead dioxide causing the electrode volume to grow. The continuous growth and shrinkage of the electrode leads to mechanical deterioration of the electrode and loss of active material
- **Water loss**: Oxygen and hydrogen gas can be generated during battery operation, which, after being expelled, reduce the amount of water in the battery leaving the electrodes partially unimmersed and therefore unusable.

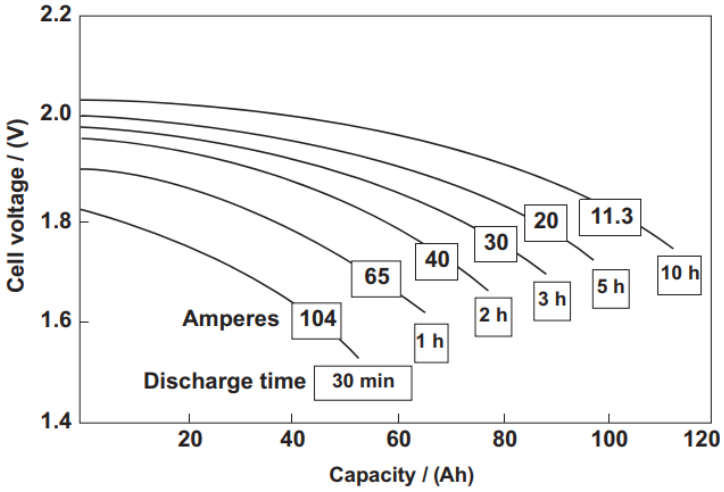


Figure 1.3: Typical discharge curves for lead-acid batteries [98].

- **Acid Stratification**: sulfuric acid has a higher density than water and therefore tends to stratify in the lower part of the battery causing uneven utilization of the active materials

Lead-acid batteries perform better in terms of time life when their discharge rate is very small, in fact, their use is relegated to those situations where the battery is used as a buffer battery that is, where discharge is limited and sporadic. Under these conditions, lead-acid batteries can reach up to 2,000 life cycles. However, unfavorable environmental conditions can bring down battery performance leading to premature death or loss of capacity. The main parameter affecting capacity is definitely temperature. As shown in Figure 1.4, fixed the discharge current (expressed in C-rate) the change in temperature affects the battery's ability to store energy.

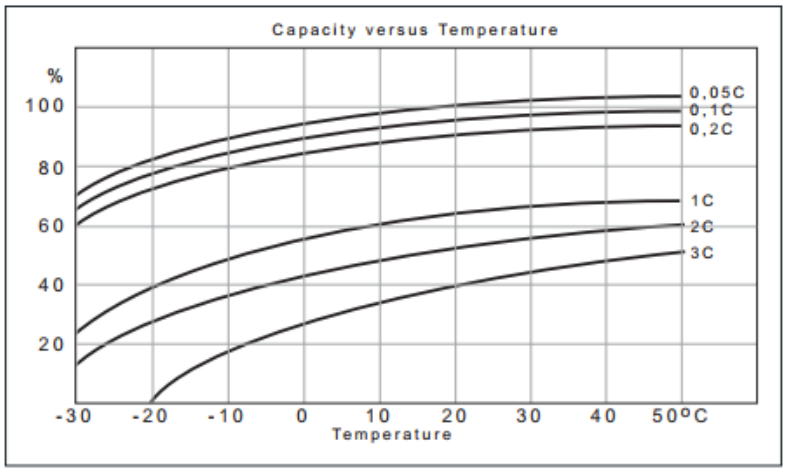


Figure 1.4: Lead Acid Batteries typical Capacity dependence of Temperature and Current [1]

The recyclability of lead-acid batteries is extremely high, to date the percentage of recycled batteries is close to 100% [118]. The main reason lies in the profitability of recycling and that in more developed nations it is illegal to discard lead-acid batteries into the environment. Also favoring recyclability is the standardization that batteries have achieved, all are formed from the same elements (ABS case, lead electrodes, Electrolyte as a mixture of sulfuric acid in water,

etc..) which simplifies the recyclability process. Finally, the elements obtained downstream from the recycling process have a high quality that allows them to be used for the construction of new batteries.

1.3.2 Lithium batteries

To date, lithium batteries represent the most popular technology among electrical energy storage systems in the form of chemical energy. The reason behind this success is the performance that lithium batteries are able to offer. Lithium is the lightest existing alkali metal and is also the least electronegative element existing in nature. These characteristics make it an excellent electron donor (and therefore an element with a low reduction potential E° = - 3.045 V), combined with a very electronegative material within an electrochemical cell it is capable of developing a very high potential difference. Another point in favor of lithium is its low atomic weight i.e., a low density that allows extremely high energy densities to be achieved

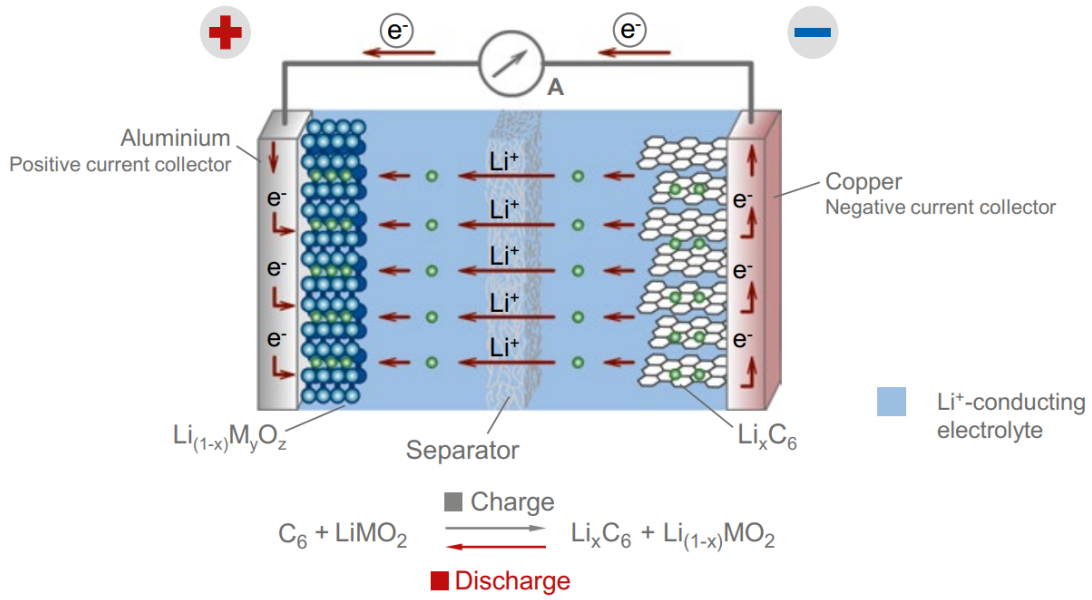


Figure 1.5: Lithium-Ion Battery overview [64]

Figure 1.5 shows a simplified diagram of the internal structure of the battery where the four main components can be identified: cathode, anode, electrolyte, and separator. The cathode consists of lithium metal oxide, which has a tunneled or layered structure on an aluminum electrode. The anode, on the other hand, is generally made of lithiated graphite also with a layered structure but on a copper electrode. The electrolyte is an ionic conductor interposed between the two electrodes. The separator is a porous polymer membrane capable of passing ions from one side of the interface to the other but blocking the passage of electrons that are then forced outside the battery providing useful work. The lithium ions then move between the two electrodes by intercalating in the porous structures of the anode and cathode. During the discharge process, the lithium ions are deintercalated from the negative electrode, releasing electrons, to migrate to the positive electrode.

There are many types of lithium-based batteries, most differ in the metal oxides used for the cathode, the Figure 1.6 shows the most common technologies also indicating the battery voltage and energy density. One of the parameters that most differentiates lithium batteries

is the cost of the basic elements, for example, Nickel and Cobalt have generally higher costs than other metals, so batteries using these elements have a higher final cost.

Acronym	Cathode	Anode	Cell voltage (V)	Energy density (Wh kg^{-1})
LCO	$LiCoO_2$	Graphite	3.7–3.9	140
LNO	$LiNiO_2$	Graphite	3.6	150
NCA	$LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$	Graphite	3.65	130
NMC	$LiNi_xMn_yCo_{1-x-y}O_2$	Graphite	3.8–4.0	170
LMO	$LiMn_2O_4$	Graphite	4.0	120
LNM	$LiNi_{1/2}Mn_{3/2}O_4$	Graphite	4.8	140
LFP	$LiFePO_4$	$Li_4Ti_5O_{12}$	2.3–2.5	100

Figure 1.6: Energy density and cell voltage for varius Lithium Ion Batteries

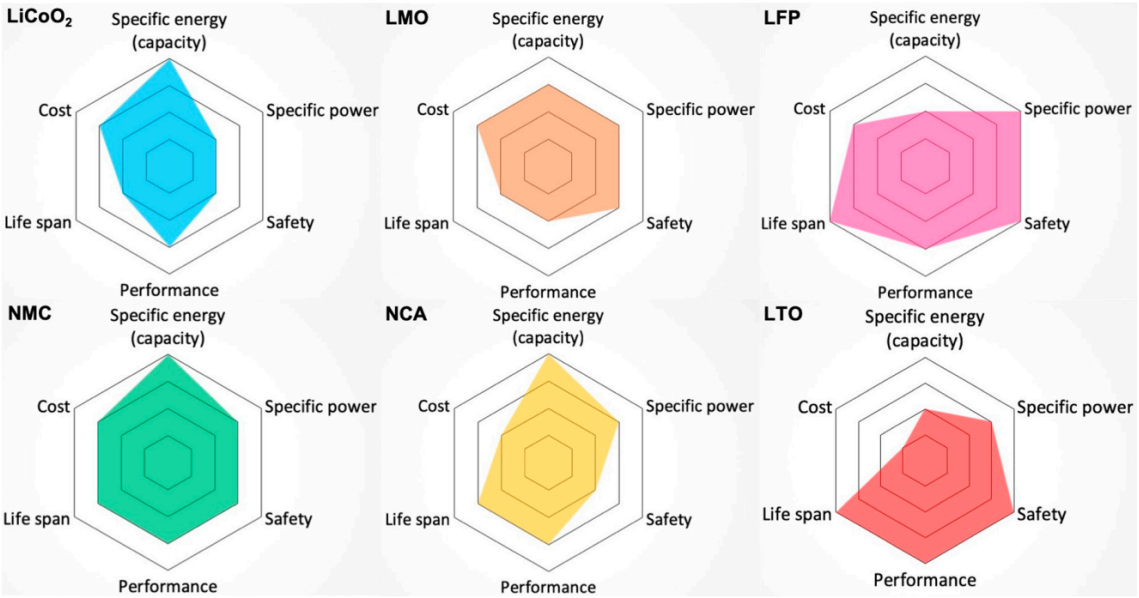


Figure 1.7: Comparisons of different types of Li-ion batteries used in EVs from the following perspectives: specific energy (capacity), specific power, safety, performance, life span, and cost (the outer hexagon is most desirable).

Figure 1.7 graphically shows the performance of individual lithium-based construction technologies, the larger the colored area the greater the battery performance. The parameters examined are cost, specific energy, specific power, safety, and life span. Where specific energy is the ratio of the amount of energy the battery is capable of storing to the weight of the battery, specific power is the ratio of the current the battery is capable of delivering to capacity.

Talking of the safety concerns, the 2022 Lucid Air Dream edition Performance has a battery pack of 118 kWh this energy is equivalent to 101 kg of equivalent TNT [1]. Furthermore,

[1]Equivalent TNT is used to express the relative explosive power or energy output of an explosive material in comparison to TNT (trinitrotoluene), which is a standard reference explosive with well-established properties. Equivalent TNT is often used in the context of measuring or quantifying the destructive potential of various explosive materials.

lithium batteries are not well managed and can take on fire. The problem is due to two reasons:

- The lithium salt added to the electrolyte: This salt has many advantages inside the battery, but it decomposes at low temperatures (50 °C)In case of fire the products of the decomposition of the salt interact with the oxygen released by the positive electrode. At the a 50 kWh battery produces 6 kg of HF (Hydrogen Fluoride) and a toxic phosphoryl fluoride gas [69] which can provoke irreversible neuronal damage [86].
- The positive electrode oxygen release: Each electrode composition has its own characteristics, but here some general information is reported. Cobalt used inside the positive electrode tends to lose oxygen on deep lithium extraction so in the time the concentration of cobalt is lowered or it is doped with aluminum (NCA batteries). The substitution of Cobalt with another transition metal (like Nickel or Manganese) has been abandoned.

LFPs are significantly cheaper than other Lithium-based technologies while lying a similar capacity [73]. In addition, LFPs are significantly safer than layered metal oxide. The strong P-O bond does not allow oxygen to be released as is the case with LCOs, also tolerating abuse well during their operation [112] [73]. Experiments have shown that LFPs release a small amount of heat (147 J g-1) if the temperature exceeds 250 °C. [73, 135].

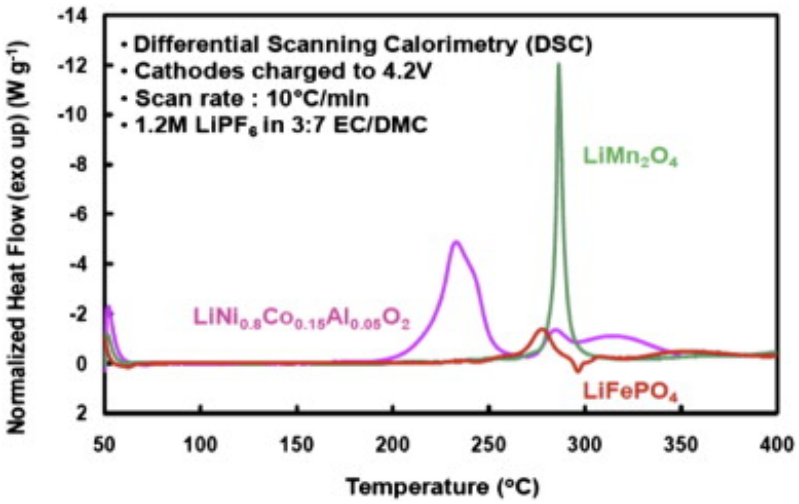


Figure 1.8: DSC spectra of over charged layered, spinel and olivine cathodes with traces of 1.2 mol L-1 LiPF6 in EC-EMC (3:7) electrolyte at 10 $°Cmin^{-1}$

In Figure 1.8 form [138] there is a differential scanning calorimetry (DSC) spectra of three different cathode chemistry during the charging process. When the curve goes up means that there is an exothermic reaction. The figure shows that the NCA electrode undergoes in exothermic reaction between 200 °C and 250 °C, much earlier than the LMO or LFP cathodes. Thus for NCA batteries the self-heating limit could be easily reached when the battery is mechanically stressed. Recycling when talking of LIBs is still a current problem, actually, the trend for electric vehicles (EVs) is, despite trashing, to reuse the batteries in another application like stationary storage systems. This is called secondary use, and it's possible because the EVs batteries are considered at the end of their life when the capacity drops under the 80% of the original capacity [84, 143]. Li-ion battery recycling is not as easy as lead acid, mainly because lead acid batteries are composed of few elements, instead, Li-ions have a wider variety of materials and elements. As a consequence, the LIBs recycling is not easy and barely

sustainable from an economic point of view. There are three main methods for recycling the LIBs [30, 37]:

- **Pyrometallurgical recycling**: batteries are melted in a furnace, plastic, and electrolyte burn and the result is an alloy of Cobalt Copper nickel, and Iron and a slug with Lithium Aluminum, and others. Recovering lithium or Aluminum from the slag is not profitable, instead by leaching is possible to recover the metals from the alloy.
- **Intermediate recycling process**: The batteries are reduced in small particles by hammering. Then the parts are filtered by an aqueous stream. This technique is commercially used in Canada but, as Pyrometallurgical recycling is profitable only the batteries have nickel or cobalt.
- **Direct recycling**: In this method, the cells are placed in a chamber with CO2 at high temperature and pressure to separate the electrolyte and the lithium. The rest are pulverized, and the materials are filtered by their different properties of density or conductivity.

The recycling research is still ongoing, and we can expect in future new methods for profitable recycling the Lithium based batteries.

1.4 Batteries equivalent circuits models

The development of precise and dependable models holds a fundamental significance across all phases of a battery's lifecycle. These phases encompass (i) guiding and optimizing its design, (ii) efficiently managing its standard operational conditions through the Battery Management System (BMS) to enhance charging/discharging techniques, (iii) averting hazardous operational states like overcharging or over-discharging, which may result in significant damage or hazards, and (iv) reliably forecasting its long-term behavior for applications such as predictive maintenance, secondary utilization, or end-of-lifecycle estimation [109, 121]. Of these models, the electro-thermal model assumes a unique role as it possesses the capability to elucidate the battery's voltage response concerning State of Charge (SoC), C-rate, and temperature (T). The intricate interconnection of various battery parameters, including voltage, current, temperature, and SoC, necessitates the development of complex, multi-physical models. Due to its critical importance, battery modeling has received substantial and escalating attention from the scientific community over recent decades. Consequently, a plethora of battery models, varying in complexity and accuracy, are now available in the scientific literature, classifiable into three primary categories based on their level of physical description. On one hand, electrochemical models offer a comprehensive depiction of batteries, delving into the intricacies of their constituent materials (conductors, electrolytes, etc.) and structural elements (electrodes, package shapes, etc.). These models simulate the intricate chemical and electrical processes occurring within the battery, often considering external factors like temperature distribution [5, 36]. Every parameter in such models carries a physical significance, directly tied to the geometry, physical characteristics of materials, and similar attributes. However, these models entail certain limitations, demanding a profound understanding of the device for parameterization and incurring a substantial computational burden, especially when analyzing intricate systems like battery packs. Conversely, behavioral models are constructed based on empirical data obtained through experimental characterization, defining input-output relationships inherent to the battery. In this context, the battery is treated as a 'black box,' resulting in model parameters devoid of physical meaning, as they are determined through the chosen identification strategy, similar to coefficients in a fitting function. Behavioral models can take either

analytical [83] or stochastic [83] forms. Analytical models often rely on a limited set of equations to describe overall battery characteristics, while stochastic models, particularly Markov processes, enable future predictions based on present states without necessitating knowledge of the entire historical record. Notably, these models suffer from the complexity of the methods employed in their derivation.

In the middle ground between these two extremes, a third category of models, known as hybrid models, comes into play, with the popular Equivalent Circuit Models (ECMs) being a prime example. These hybrid models take a unique approach: they combine elements from both electrochemical and behavioral modeling paradigms.

In the case of ECMs, these models adhere to a predefined and fixed circuit structure. For instance, an ECM might utilize a cascade of resistors and capacitors along with appropriate voltage sources in its equivalent circuit. While these circuit elements can be ascribed physical meanings, akin to the electrochemical models, their values are determined through empirical input-output characterizations of the battery, similar to behavioral models. These hybrid models have gained widespread acceptance due to their ability to effectively capture the electro-thermal behaviors of batteries with a commendable level of accuracy, all without introducing unnecessary complexity [44, 54, 99, 126].

It's worth noting that electrical models are available for various types of batteries, ranging from traditional lead-acid batteries to modern Li-ion ones [26]. These models can be classified into three main classes based on their circuit structures: Thévenin-based models, impedance-based models, and runtime-based models, each of which has its own set of advantages and disadvantages.

The Thévenin-based model, representing the most straightforward approach of the three, relies on a series resistor and an RC parallel network to anticipate a battery's response to abrupt load changes at a specific State of Charge (SoC). It simplifies the modeling by assuming a constant open-circuit voltage [46, 105]. Nevertheless, this model encounters limitations in accurately predicting the battery's behavior during steady-state and continuous runtime operations [26].

Conversely, impedance-based models craft an AC-equivalent impedance profile of the battery, usually in the frequency domain, through techniques like Electrochemical Impedance Spectroscopy (EIS) [6, 15, 66]. The parameters derived from EIS measurements remain valid only for particular battery SoC and temperature conditions, necessitating periodic recalibration during battery use. However, EIS measurements can be time-intensive, and recalibration might not always be a practical solution, particularly for complex ECMs with many components. Furthermore, impedance-based models are unsuitable for forecasting the battery's DC response or runtime characteristics [26].

Lastly, runtime-based models adopt a sophisticated circuit network to emulate the battery's runtime and DC voltage response under constant discharge current conditions, typically within SPICE-compatible simulation platforms [40]. However, this approach sacrifices the capability to conduct dependable AC or transient analyses. Regardless of the chosen model structure, calibrating an ECM for transient battery analysis is a laborious and time-consuming process. It involves a substantial number of experimental trials conducted under varying operational conditions, including alterations in charging/discharging current amplitudes and temperature settings.

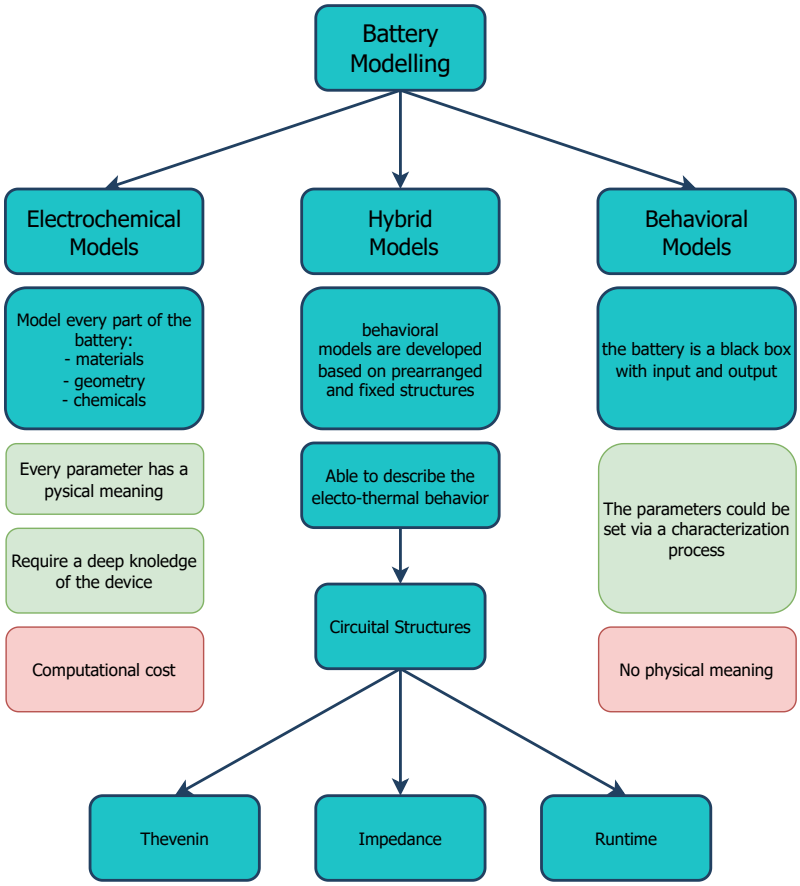


Figure 1.9: Battery modelling

1.5 Batteries status indicies

1.5.1 What and Which they are

Batteries represent intricate systems in which chemical and physical phenomena interplay, rendering it a challenging endeavor to succinctly depict their operational status. This complexity stems from the amalgamation of chemical reactions and physical processes that happen inside the battery cell. Consequently, producing a precise state of a battery is not always straightforward. To help the users and designers in battery status estimation, a series of indices are created.

The state of Charge (SoC) is an indicator of the remaining energy in the battery. It is expressed in percentage between 0% to 100% where the first one means that the battery is empty then the second one means that the battery is fully charged. Equation 1.5 shows how the SoC is computed [52] where "C_{remain}" is the remaining capacity of the battery while "$C_{present}$" is the present maximum capacity of the battery in that specific moment.

$$SoC = \frac{C_{remain}}{C_{present}} \cdot 100 \tag{1.5}$$

Together with SoC another widely used parameter is the State of Health, but while SoC gives the actual energy present in the battery SoH gives a metric to evaluate the aging level of the batteries. The SoH, as reported in Equation 1.6, is the ratio between the present maximum capacity, "$C_{present}$", and the original battery capacity, "$C_{initial}$".

$$SoH = \frac{C_{present}}{C_{initial}} \cdot 100 \tag{1.6}$$

The State of Health (SoH), akin to the State of Charge (SoC), is quantified in percentage terms, with its range spanning from 0% denoting a battery incapable of storing energy to 100% representing a battery in a pristine, brand-new condition. Other parameters used to describe the state of a battery are the State of Power (SoP) and State of Energy (SoE). Before explaining SoP is useful to determine the battery Safe Operating Area (SOA) which is determined by the operation range regarding temperature, voltage, and current. The SoP can be represented as the maximum electrical power that the battery is able to absorb or provide while remaining within the prescribed boundaries of its SOA [114]. SoP is expressed in units of Electrical power. The State of Energy is intrinsically linked to the State of Charge. SoC is a relative metric expressed as a percentage, serving as an indicator of the battery's charge level relative to its present maximum capacity. In contrast, SoE serves as an absolute metric, delineating the precise energy content residing within the battery. SoE is expressed in a unit of energy [75]. Remain Useful Life (RUL) relates to a metric used to estimate how much more operational time or cycles a battery can provide before it reaches the end of its serviceable life. The RUL is always a prediction and not a measured value.

1.5.2 How they are estimated

SoC and the SoH represent the predominant and widely employed metrics for delineating the batteries' status. Consequently, there exists a dynamic and ongoing research effort within the scientific community to advance the methods to establish these metrics. Over time, this concerted research endeavor has yielded a plethora of diverse methodologies and approaches aimed at increasing the precision in SoC and SoH estimation. In this Thesis, I'm focusing on the SoC methods because, on one hand, the methods used for SoC and SoH are most of the time the same, and on the other hand, SoC is easier to replicate and allows the researchers to have more data in less time. In [100] the Authors classify the methods used for SoC estimation, a condensed graph is reported in Figure 1.10.

In general, because the methods are not precise, the SoC estimation is done by union of two or more methods together. Direct methods are based on the physical battery proprieties i.e. current voltage or temperature to determine the SoC.

Coulomb Counting

The most used method is the Coulomb Counting where the only parameter measured is the battery output current. The integration of the current in time, as shown in Equation 1.7, is physical and energy and in particular, is the Battery capacity as shown in Equation 1.5.

$$Capacity(t) = Capacity(t0) + \int_{t0}^{t} I_{bat} dt \tag{1.7}$$

Where $Capacity(t0)$ represents the energy stored within the battery at time $t = t0$; I_{bat} denotes the measured electric current. Typically, the electric current is sampled at regular intervals of τ, referred to as the "sampling interval" or the inverse "sampling frequency". Both the "$C_{present}$" and "C_{remain}" of Equation 1.5 could be determined with this method. This algorithm, despite being widely used, has some issues [91]:

- The measured current has an intrinsic measurement error

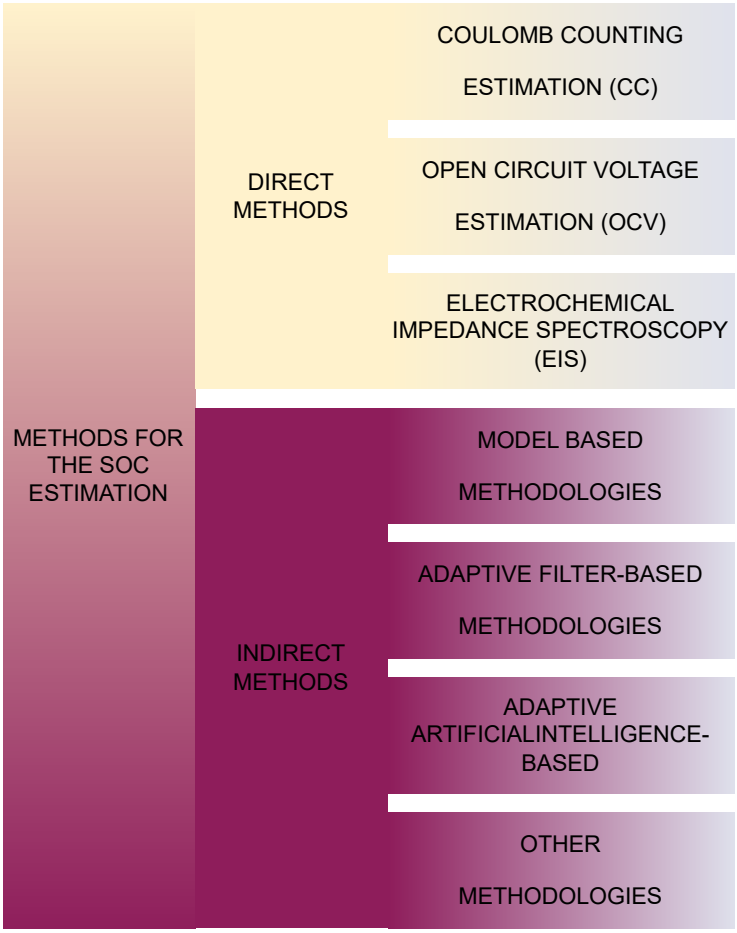


Figure 1.10: Most used methods for SoC estimation in batteries

- Also the time measurement has an intrinsic measurement error
- The integration operation with discrete data gives an approximation of the real result which will result in capacity estimation errors.

All these sources of uncertainty bring the capacity estimation to be inaccurate. Furthermore, both the "$C_{present}$" and "C_{remain}" of Equation 1.5 could be determined with the Coulomb Counting which results in an increase of the SoC estimation error.

Open Circuit Voltage

In Li-ion batteries, the Lithium Ions are intercalated into the crystalline lattice structures of the battery electrodes. This insertion leads to a differential accumulation of ions between the positive and negative electrodes, consequently generating an electromotive force (EMF). EMF exhibits a strict correlation with the SoC of the battery. Open Circuit Voltage (OCV) is defined as the electrical potential difference observed between the terminals of the battery when no current is flowing. Under specific operational conditions, OCV closely approximates the EMF. As a result, it becomes feasible to estimate the SoC of the battery by measuring the voltage across its terminals and applying a straightforward mathematical transformation. The OCV methodology is frequently employed due to its inherent advantage of not requiring complex calculations, as battery voltage can be readily monitored during operation.

The strict closeness between OCV and EMF is true only when the battery is in a steady state condition; after a charge or a discharge the time to reach the steady state condition depends on the battery but it can vary from several minutes to hours because the process is mainly dominated by the diffusion. If the steady state is not achieved, the measurements are not reliable [101]. Furthermore the State of Charge-Open Circuit Voltage (SOC-OCV) curve depends also on the battery temperature and battery SoH.

LiBs employing cathode materials such as Lithium Iron Phosphate (LFP) or similar variants exhibit a nearly flat SOC-OCV and a notable OCV hysteresis phenomenon. Consequently, for LFP, the utilization of OCV-based SOC estimation methods within the SOC range associated with the flat SOC-OCV curve becomes unreliable.

Electrochemical Impedance Spectroscopy

Electrochemical impedance spectroscopy (EIS) emerges as a promising alternative for replacing Coulomb/OCV counting methods in State of Charge (SoC) estimation, as indicated by prior studies [10, 87]. Notably, EIS circumvents the challenges previously outlined and offers the potential advantage of real-time applicability in actual operating conditions [65]. EIS could be performed in galvanostatic or potentiostatic method. In the first case the current is controlled by the measurement equipment while in the second case the voltage is controlled. Whatever the measurement method, both current and voltage are measured. The generated stimulus signal, often sinusoidal, typically has span frequencies ranging from a few millihertz (mHz) to a few kilohertz (kHz) [68, 89]. Equation 1.8 shows the complex value of the measured impedance ($\dot{Z}(f_k)$) at different frequencies of investigation (f_k); $\dot{V}(f_k)$ is the voltage phasor, and $\dot{I}(f_k)$ is the current phasor. It's anyway important to note that the Equation 1.8 is valid only for small signals because batteries are complex non linear systems, so thay can be considered linear only when the stimulus is small enough. To define the amplitude of a small signal, the total harmonic distortion (THD) could be a good metric, higher is the THD and more the cell could not be considered a linear load.

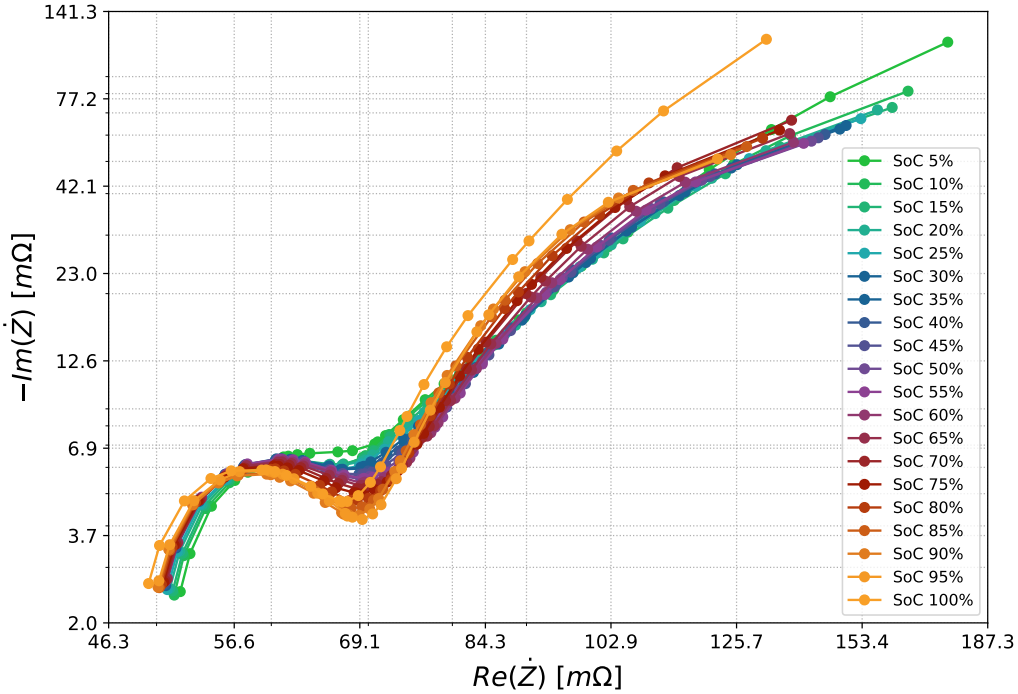


Figure 1.11: Nyquist plot of a Lithium Iron Phosphate battery impedance at different State of Charge

$$\dot{Z}(f_k) = \frac{\dot{V}(f_k)}{\dot{I}(f_k)} = Re\left(\dot{Z}(f_k)\right) + jIm\left(\dot{Z}(f_k)\right). \tag{1.8}$$

Being the impedance a complex number, it has a Real part and an Imaginary part. The Nyquist diagram reports on the horizontal axis the real part and on the vertical axis the imaginary part of the impedance. An example of it is reported in Figure 1.11 at different SoC for the same battery. Tipically, the electrochemical cell impedances have a capacitive contribution rather then inductive, so the vertical axis is negative. Low-frequency impedances are on the right of the curve while the high-frequency impedances are on the left. EIS is interesting because the battery impedances could be linked to the physical behavior of the elements inside the battery. The battery's diffusion processes are indicated by the low-frequency tail [2], the double-layer capacitance effect is indicated by the mid-frequency semicircle [3], and the battery bulk's Ohmic resistance is indicated by the high-frequency region's intercept of the EIS curve with the real axis. For this reason, EIS could be also considered as a basic measurement for a data-driven SoC estimation.

Indirect Methods

The indirect methods are methods developed to connect direct methods to SoC estimation.

[2]The diffusion layer is a virtual layer in the bulk solution and next to the electrodes where the concentration of ions is different from the bulk ions concentration

[3]The double layer is a region at the interface between an electrode and an electrolyte solution where significant electrical charge separation occurs.

In other words, they use the direct measurement (current, voltage, temperature, impedance) and correlate them with the battery SoC.

Model based Methodologies

Model-based methods are widely used for SoC estimation. In general, there are many applications based on Electrochemical, Hybrid, and behavioral Models. Electrochemical models uses a combination of electrical parameters (Resistance, Capacitance, etc..) and non linear models to take in account the real behavior of the system. The obtained differential equations from the model [90] could be used to estimate the state of charge. Despite the good performances, Electrochemical models are computationally heavy. Thevenin models are composed of a non-linear voltage source, and one or more RC branches. The number of RC branches depends on the tradeoff between computational costs and model accuracy. More branches are used and higher are the performances. In general, in literature, one or two branches [9]. Electrical models have the advantages of being simple and easy to compute.

Adaptive filter based Methodologies One of the most used methodology for Soc and SoH

estimation is based on the Kalman Filters (KF). The key idea behind the Kalman Filter is to maintain a probabilistic representation of the state of the system and update it recursively as new measurements become available. KF could be divided in three branches:

- *Linear Kalman Filters (KF)*: Incorporates measurement models and linear dynamics. When dealing with nonlinear systems, it can produce poor results, but it works well when the system and sensor models are actually linear.
- *Extended Kalman Filters (EKF)*: Extends the Kalman Filter to handle nonlinear models by linearizing them at each iteration. Even though EKF is capable of handling nonlinear systems, accuracy problems could arise if the linearization is not a good approximation.
- *Unscented Kalman Filters (UKF)* employs the Unscented Transform, a deterministic sampling method, to approximate the distribution of the state variables in order to overcome the limitations of the EKF. Because it doesn't rely on linearization, the UKF is better suited for highly nonlinear systems.

KFs are suitable and widely used for the battery status estimation [120].

Adaptive Aritficial Intelligent based Methodologies

Due to their powerful computational intelligence capabilities, machine learning techniques like artificial neural networks (ANN) [78], fuzzy logic (FL) [53], and support vector machines (SVM) [18] have attracted a lot of attention in SOC estimation in recent years. A well-liked machine learning technique called an artificial neural network (ANN) can accurately analyze SOC while taking into account non-linear characteristics of lithium-ion batteries, aging, noises, and temperature effects. However, ANN requires a significant amount of data, a storage device, and a lengthy training period while the maximum error in the SoC estimation is around 5%. SVM has quick and accurate estimation with a good result in terms of classification accuracy, [18] report the best case at 0.98%, furthermore the computational cost of SVM is small if compared with other Machine Learning algorithm

Other Methodologies

All the methods examined above take as input the current or the voltage from the battery terminals, but there are several methods that take into account other physical changes in the battery. In [47], one of many authors who have investigated magnetic methods for charge state estimation, a coil wound around a ferromagnetic core was mounted on the side of a lead-acid battery. When the state of charge changes, the magnetic field is impacted due to electrochemical phenomena that occur inside the battery. At an excitation frequency of $10kHz$ and a coil base value of $168mH$, a variation of $1mH$ is obtained. A similar approach has been tried in [59] but instead of using a single coil, in this case, two coils are placed on either side of the battery, one of the two coils is used to generate the magnetic field at a frequency of 70 kHz while the voltage induced on the second battery turns out to be proportional to the SoC of the battery. In [128] the author places the single excitation coil inside the battery to minimize the distance to the negative electrode. In fact, in lead-acid batteries, during discharge, the negative electrode becomes coated with lead sulfate, which has a different permeability than metallic lead.

Chapter 2

Predictive Maintenance

2.1 Predictive Maintenance on Battery powered devices

The following analysis has the aim of analyzing the status of the art of battery diagnosis by finding the most valuable contributions to the community knowledge. Furthermore, the most prolific researcher and istitutions are outlined.

2.1.1 Data collection

The data used in this investigation were obtained from the Web of Science Core Collection (WoS CC) which is an enormus citations database, where is possibile selecting all the indexed papers that satisfy the query conditions. This study is focused on the reasearches pubblished in the last ten years, between the *01/09/2013* and *01/09/2023*. The query is structured in two parts connected in logic "AND". The first block select all the papers that cointains one or more of the following terms *(TS = "lithium-ion batter*" OR TS = "lead acid batter*" OR TS = "Lithium ion batter*" OR TS = "Li ion batter*" OR TS = "batter*")*. The symbol "*" is works as wildcard. The second block instead is composed by *(TS = "fault diagnosis" OR TS = "fault prediction" OR TS = "fault detection" OR TS = "fault monitoring" OR TS = "failure*" OR TS = "diagnos*" OR TS = "early diagnos*" OR TS = "prognostic*" OR TS = "fault identification" OR TS = "condition monitoring" OR TS = "prognostics and healthment" OR TS = "condition-based maintenance" OR TS = "PHM" OR TS = "prognostics and health management")*. With the built query, 16′093 pubblications are selected. Refinig the search by including the article and review contribution the number of pubblications are 14′117, where 13786 are in english. The last refinment is applied on the Research Areas, in order to exclude the articles focused on the material science which is not the focus of this thesis. The selected areas are: *Engineering, Energy Fuels, Science Technology Other Topics, Instruments Instrumentation, Meteorology Atmospheric Sciences, Spectroscopy, Computer Science*. With all the applied filters, 5′784 pubblications are selected.

2.1.2 Topic Engaging

The volume of publications within a specific subject over a defined timeframe serves as a significant indicator of the research focus it garners. A comprehensive comprehension of the temporal patterns and evolving trends in publications is made more accessible through the systematic examination of publication trends related to the diagnosis of Lithium-Ion Batteries (LIB).

Figure Figure 2.1 vividly illustrates the escalating interest surrounding battery diagnostic studies. In the year 2014, the literature contained a mere 190 articles devoted to this subject. Remarkably, this number surged to a substantial 1156 articles by the year 2022. Over the span

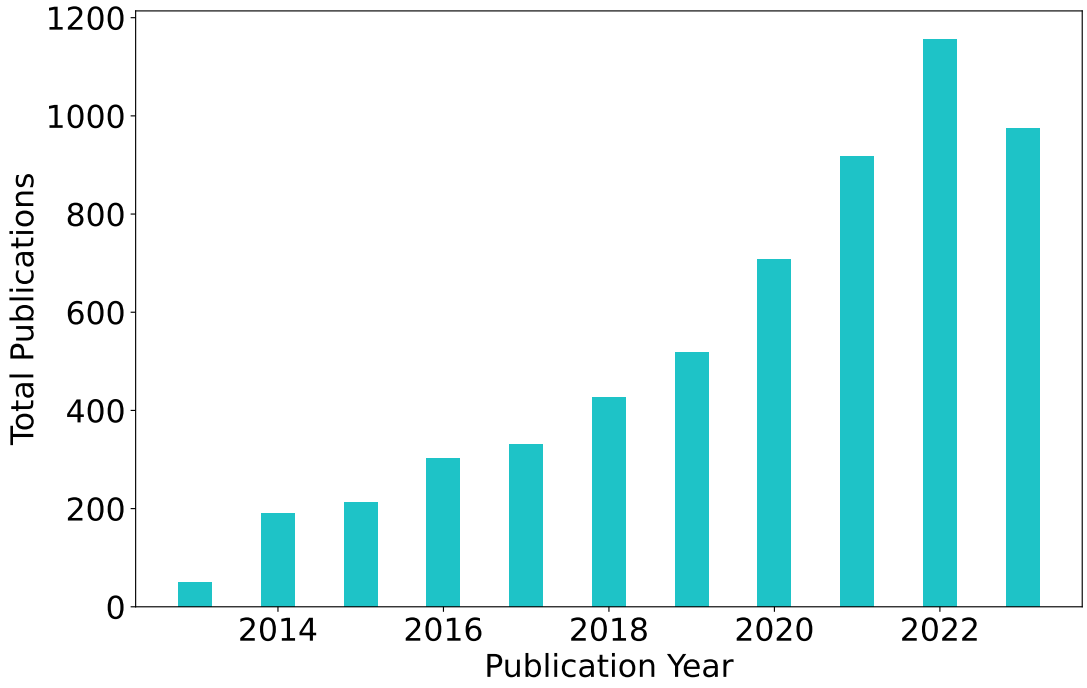


Figure 2.1: Publication rate in the last 10 years, the analysis is made from September 2013 to September 2023

of just 8 years, this growth amounts to an impressive 600% increase in the volume of scholarly works dedicated to this field.

Based on this growth pattern, it is possible to predict that publications on related research will continue to rise in the years to come, with 2023 not marking the end of exponential growth in the number of publications.

2.1.3 Geographical Distribution

The rankings of the top ten most productive countries are reported in Figure 2.2. In this context, both China and the United States stand out as the foremost prolific nations, with South Korea following behind. It's worth noting that when considering the European Union as a single entity, the cumulative contributions of Germany, France, and Italy are substantial enough to position the EU as the third most productive region.

An analogous discovery is echoed in Table 2.1 as presented in [67]. China emerges as the highest contributor in terms of sheer volume, yet interestingly, it occupies only the seventh position in the average citation ranking. This finding suggests that, despite its prolific output, articles originating from China are comparatively less compelling to the scientific community.

Conversely, research originating from France receives notable recognition within the scientific community. The average citation per document for French contributions stands at a remarkable sixfold higher than that of China.

2.1.4 Journal Distribution

The pivotal journals in the field of battery fault detection are highlighted in Figure 2.3. Remarkably, the top two journals, namely, the *Journal of Power Sources* and the *Journal of Energy Storage*, are both under the esteemed banner of Elsevier. In addition, the third influential

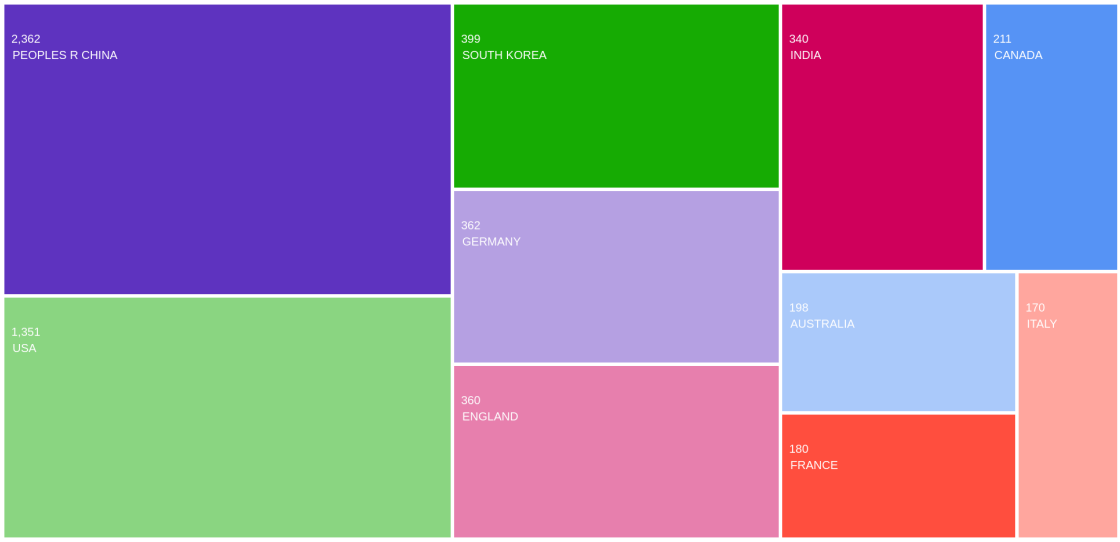


Figure 2.2: Article or review published in the last 10 years from the most prolific countries

journal, emphEnergies, is published by *MDPI*. Table 2.3 shows the keywords chosen by the Authors. Lithium-ion batteries are the most studied battery beacause their diffusion in the market but also because of the safety issues related to them.

Table 2.1: 10 productive countries/regions in LIB fault diagnosis

Rank	Label	Continent	Publications	AC	APY	TLS
1	China	Asia	1139	29	2018.9	470
2	USA	North America	810	44.3	2017	426
3	South	Korea Asia	185	30.2	2017.9	63
4	Germany	Europe	167	27.3	2018.3	100
5	UK	Europe	159	40.2	2018.6	132
6	Canada	North America	105	179.5	2018.1	96
7	France	Europe	105	200.4	2016.8	95
8	Australia	Oceania	61	36.5	2018.9	70
9	India	Asia	56	20.8	2019.4	40
10	Japan	Asia	53	22	2016.5	30

Table 2.2: Note: TC = Total citations, AC = Average citations per document, APY = Average publication year, TLS = Total link strength.

Figure 2.3: Ten most used Journal on battery fault detection, the area of each rectangle is proportional to the journal publications on the selected topic

Table 2.3: Most frequent used keywords

	Authors Keywords	Number of Articles
1	lithium-ion battery	1404
2	batteries	255
3	state-of-health	255
4	electric vehicles	254
5	thermal runaway	219
6	remaining useful life	178
7	fault diagnosis	173
8	degradation	131
9	reliability	128
10	machine learning	126
11	battery	121
12	state of charge	119
13	energy storage	118
14	safety	106
15	battery safety	93
16	energy harvesting	83
17	deep learning	81
18	battery management system	73
19	internal short circuit	65
20	estimation	64
21	wireless sensor networks	64
22	predictive models	62
23	aging	62
24	remaining useful life prediction	61
25	prognostics and health management	60
26	remaining useful life (rul)	59
27	fault detection	58
28	electrochemical impedance spectroscopy	56
29	particle filter	55

2.2 State of Art

From the analysis carried out in the previous paragraph, a series of works were selected to make up the state of the art illustrated below. Maintenance is a fundamental operation within industry, especially given the growing interest in automation and mechanisation of processes. Depending on the chosen approach, maintenance can be divided into three different types: "Failure-driven", "Time-based" and, "Condition-Based" as shown in the Figure 2.4. "Failure-driven maintenace (FDM), also known as 'run to failure' is that maintenance based on intervention after failure or breakdown. There are no maintenance plans, but simply repairs are carried out should an incident occur. The advantages of such an approach are obvious, there is no need for planning and since there are no ready-made spare parts, there is no inventory cost. The disadvantages are also obvious, at any time a breakage can render the instrument unusable for an indefinite period of time, and the repair is an operation carried out in an emergency situation which tends to be more costly [134]. A possible aid to the 'failure-driven' approach can be the 'time-based' approach. In this case, the process has a series of planned maintenance operations. Reducing the maintenance plan is not a trivial task, one must carefully identify the elements that need periodic interventions and decide how often to carry out the interventions. Too frequent interventions lead to increased costs due to maintenance costs and downtime. Intervention times are decided on the basis of statistical considerations, operator experience and instrument datasheets [134]. In recent years, thanks in part to the drive of Industry 4.0, a new type of maintenance, Condition Based Maintenance (CBM), has begun to be considered. CBM is based on the need to determine the state of a system (diagnosis) and identify a set of parameters able to indicate a tendency of the device or process towards a fault condition before it actually occurs (prognosis) [134]. Prognostic parameters give an indication of potential issues and developing faults that could cause a component or piece of equipment to perform below an acceptable standard. Equipment aging and deterioration are frequent issues in maintenance. Before equipment fails or is damaged catastrophically, these equipment conditions serve as useful warning signs of potential faults and issues.

All the above-defined principles could be applied also to batteries and battery-powered devices. Extensive efforts in the academic literature have been dedicated to developing effective methods for enhancing the reliability and availability of batteries. CBM methods enable the automated and real-time monitoring of concealed indicators related to battery degradation, such as state-of-health (SOH), state-of-charge (SOC), and the prediction of their remaining useful lifetime (RUL). These predictions are valuable for supporting maintenance strategies based on the condition or even predictions. CBM techniques are used to optimize battery maintenance and ensure that the battery can satisfy the required performance over its lifetime. CBM techniques are normally comprised of three primary components [103]:

- Data acquisition: Data acquisition involves collecting data from the battery's sensors and other monitoring equipment. Data acquisition represents a crucial element within battery prognostic and health management methodologies, encompassing a variety of approaches for collecting data. Sensor-based data acquisition involves using sensors to collect data on the battery's temperature, voltage, current, and other parameters. The data collected from these sensors can be used to monitor the battery's condition and to develop models for predicting the battery's remaining useful life (RUL) [57] [139]. External feature-based data acquisition involves collecting data on external features such as the battery's charge and discharge cycles, usage patterns, and environmental conditions. This data can be used to develop empirical models for predicting the battery's RUL [103]. Between these two extreme edges, there are several combinations of Sensor-based and External feature that depends on the specific application and the related costs.

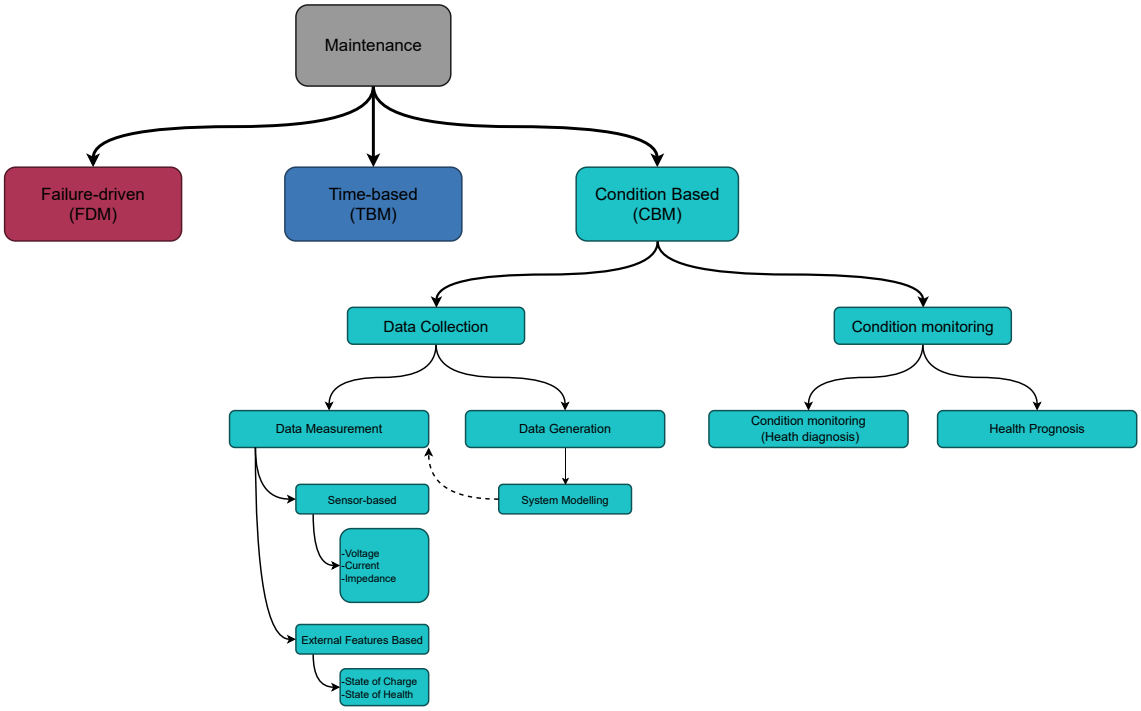


Figure 2.4: Maintenance topologies

- Condition monitoring and health diagnosis: Battery condition monitoring is a crucial component of battery prognostic and health management (PHM) techniques. Monitoring means having the knowledge at any time of the battery status indices.
- Health management and health prognosis: Health management involves monitoring the battery's condition and taking corrective action to prevent failure. This can include actions such as adjusting the charging rate or replacing a faulty cell [88].

2.2.1 Data Acquisition

Current: The generation of Joule heat is a consequence of the flow of electric current [142]. The heat generated in a battery is proportional to the square of the current flowing through it [27]. If the heat generated is not dissipated more quickly than it is produced within a battery, it can potentially lead to thermal runaway. Another risk associated with high current charging on the negative electrode is the potential for lithium plating, which can pose a safety hazard [79]

Voltage: Excessive voltage, as seen in overcharging, can trigger decomposition in both the electrolyte and the positive electrical conductor. This leads to the production of gas and heat. Conversely, insufficient voltage, as experienced in under-voltage situations such as over-discharge or deep discharge, can cause lithium plating on the negative electrode. This can also result in the deterioration of the copper current collector, potentially leading to the formation of copper dendrites and an increased risk of internal short circuits [3].

Internal Impedance: The internal impedance of batteries can be an indicator to describe the state of a battery. The battery generally has a behaviour that is definitely resistive but also more or less capacitive depending on the stimulus frequency. Currently, there are commercial systems that allow impedance measurements in the range from a few mHz to tens of kHz. Impedance depends not only on the state of health and charge of the battery, but

also on its internal temperature and the state of operation, online (measurement at the same time as the current supply) or offline (measurement with the battery electrically disconnected from the load). In recent years, research has spread that aims to perform online impedance measurements directly via DC-DC converters in battery modules [2, 119].

Temperature: Battery temperature is an extremely important parameter that must be measured and kept under strict control both to keep the battery in the ideal thermal operating regime and to detect any malfunctions. To date, battery modules have a series of temperature sensors inside them, but these increase the cost of the module and do not allow a real knowledge of the internal temperature of the cells, but only an external knowledge that delays the intervention of heating or cooling of the battery pack. Research is therefore moving towards sensorless systems based on impedance spectroscopy [71] or innovative materials combined with neural networks [51].

State of Battery: The most commonly used external parameters are state of charge, state of health, but power status, energy status and others are also defined as required. These indicators bring together in a single piece of information that is of interest to the designer or end user. They are dealt with extensively in the section 1.5.

2.2.2 Diagnosis

Battery diagnosis encompasses the comprehensive evaluation of a battery's present condition and operational performance, encompassing critical parameters such as State of Health (SOH) and State of Safety (SOS). In recent years, there has been notable progress in the field of battery diagnosis through the utilization of data–driven methodologies, which primarily rely on empirical data. One prominent avenue in battery diagnosis involves the application of machine learning algorithms. These sophisticated algorithms are adept at scrutinizing battery performance data, enabling them to make precise estimations of critical metrics, such as capacity and cycle life. Through the training of these models with extensive datasets, machine learning techniques have demonstrated their potential in achieving highly accurate SOH estimations for batteries [142]. Another facet of battery diagnosis centers on the creation of diagnostic tools and techniques. These tools are designed to continuously monitor battery behavior and promptly identify any irregularities or signs of deterioration. Researchers have gained a more profound comprehension of battery degradation mechanisms by scrutinizing various electrochemical side reactions within batteries.

2.2.3 Prognosis

The trends in battery prognosis are [142]:

- Cloud-Edge Interaction: The future trend in battery prognosis involves the use of cloud-based AI-powered diagnosis and prognosis. This interconnected framework enables intelligent data processing and analysis for accurate battery diagnosis and prognosis. It allows for the uploading of battery operating data to the cloud, providing valuable insights for battery health management.
- Full-Scale Diagnosis: Another future trend is the development of comprehensive diagnostic approaches using in-vehicle field data. This approach aims to assess the state of health (SOH) and state of safety (SOS) of the battery. By analyzing real-world operating conditions and aging mechanisms, it provides a more accurate understanding of battery performance and reliability.
- Artificial Intelligence: The integration of artificial intelligence (AI) techniques is a key trend in battery prognosis. Data-driven, machine learning-based approaches have shown

promise in accurately predicting battery SOH, cycle life, remaining useful life (RUL), and identifying cells with a high risk of failure. AI models trained on laboratory and field data can provide valuable insights for EV owners.

Overall, the future trends in battery prognosis involve leveraging cloud-based AI-powered approaches, comprehensive diagnostic methods using field data, integration of artificial intelligence techniques, and the development of electronic health reports. These advancements aim to improve battery performance, reliability, and safety in real-world EV applications.

Chapter 3

Methods for the Battery Status Diagnosis

3.1 Accuracy and Measurement Time Optimization in Impedance Spectroscopy

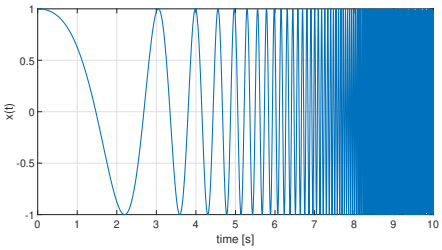
The importance of impedance measurement extends significantly across a wide range of applications, each with specific and critical needs in terms of time required and desired accuracy on the impedance measurement. This quantity, which reflects the resistance to current flow in an electrical system, is a fundamental parameter for understanding and characterizing systems and devices in many engineering contexts.

In recent years, impedance measurement has also become increasingly important in battery research. Batteries are key components in numerous devices and applications, and their reliability is crucial. Impedance measurement has been used to estimate parameters relating to the operating state of batteries, such as the State of Charge (SoC) and the State of Health (SoH), but also to identify behavioral models based on equivalent electrical circuits, which allow prediction of battery behavior under different operating conditions.

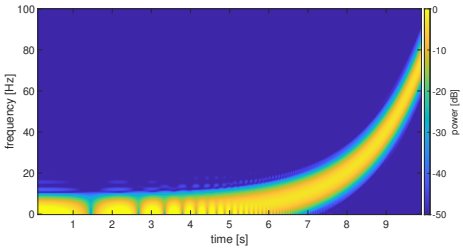
Impedance measurement on batteries is commonly performed through the Electrochemical Impedance Spectroscopy (EIS) technique. This technique provides detailed information on the electrochemical properties of batteries and can be applied either during battery operation (online mode) or by interrupting the battery from its normal operation to perform the measurement (offline mode). Through EIS, a frequency–varying electrical stimulus signal is applied to the battery, and then the electrical response of the battery is measured. In this way, battery impedance can be measured at different frequencies. This technique offers numerous advantages, including the ability to characterize the battery in a non–destructive way and the possibility of obtaining detailed information on various electrochemical processes occurring within the battery by analysis due to the wide range of analyzed frequencies.

However, the main disadvantages of EIS concerns the time required to implement the technique, that is the time to cover all frequencies of interest. Especially in online mode, it is crucial to minimize the measurement time to avoid significant variations in battery behavior during the measurement process. Even in offline mode, it is important to minimize the measurement time to limit the disservice caused by the interruption of normal battery operation.

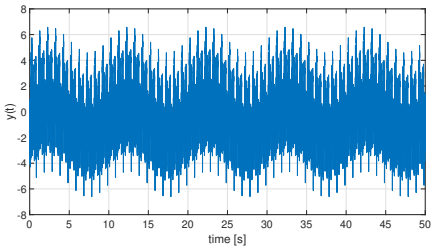
Therefore, the purpose of this section is to analyze some stimulus signal and processing techniques approaches in order to identify optimal combinations that allow EIS to be used to efficiently estimate battery impedance, achieving good trade–offs between measurement time and measurement uncertainty associated with impedance estimation itself.



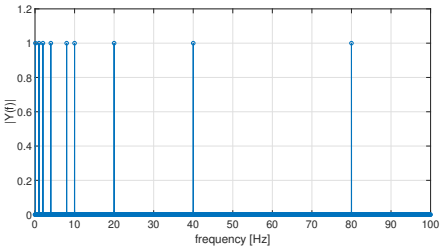
(a) *Stepped sine signal in the time domain.*



(b) *Spectrogram of the stepped sine signal.*



(c) *Multisine signal in the time domain.*



(d) *Multisine signal in the frequenc domain.*

Figure 3.1: Considered stimulus signals.

The analyzed stimulus signals

Two different stimulus signals were considered: stepped sine and multisine. The stepped sine represents the most adopted stimulus signal for implementing EIS on batteries. It involves applying a pure sinusoidal signal with constant amplitude and phase, varying the frequency of the sinusoidal over time. 3.1a and 3.1b show an example of a stepped sine stimulus signal in which the frequency varies during the stimulus time with a logarithmic trend. In detail, 3.1a shows the trend of the signal over time while 3.1b shows the spectrogram of the signal in which the logarithmic change in frequency during the stimulus execution time can be seen.

Instead, the multisine signal represents a sum of several sinusoidal signals at different frequencies with adjustable amplitude and phase. Therefore, a multisine signal simultaneously encloses within it all the frequencies to be analyzed. The general form of a multisine signal is as follows

$$y\left(t\right) = \sum_{k=1}^{N_k} A_k \cos\left(2\pi f_k t + \phi_k\right), \tag{3.1}$$

in which the multisine signal $y\left(t\right)$ is sum of N_k sinusoids having amplitudes A_k, frequencies f_k and phase ϕ_k. Furthermore, the phases are generally optimized to obtain a minimum crest factor, therefore the best signal–to–noise ratio. A simple choice of phases is based on Schröder phases [?], as shown below

$$\phi_k = \frac{-k\left(k-1\right)\pi}{N_k}. \tag{3.2}$$

3.1c and 3.1d show an example of a multisine signal in the time and frequency domains, respectively.

The analyzed processing techniques

Two different processing techniques were considered: frequency domain and time domain analysis. In all cases, the aim is to analyze the voltage and current signals measured on the battery in order to estimate the impedance $\dot{Z}(f_k)$ at different frequencies of investigation.

Regarding the evaluation technique based on frequency analysis, the procedure is to perform the ratio between the voltage phasor $\dot{V}(f_k)$ and the current phasor $\dot{I}(f_k)$ to measure the real and imaginary part of the impedance, as shown below

$$\dot{Z}(f_k) = \frac{\dot{V}(f_k)}{\dot{I}(f_k)} = Re\left(\dot{Z}(f_k)\right) + jIm\left(\dot{Z}(f_k)\right). \tag{3.3}$$

The voltage and current phasors are estimated with the FFT algorithm. It is worth noting that processing by FFT can be applied to both stepped sine and multisine stimulus signals.

Regarding processing by time–domain analysis, the procedure involves making the ratio between the voltage rms value $V_{rms}(f_k)$ and the current rms value $I_{rms}(f_k)$ in order to estimate the amplitude of impedance $\left|\dot{Z}(f_k)\right|$ and the difference between the voltage phase $\phi_V(f_k)$ and the current phase $\phi_I(f_k)$ allows the impedance phase $\phi_{\dot{Z}}(f_k)$ to be measured, as shown below

$$\dot{Z}(f_k) = \frac{V_{rms}(f_k)\, e^{j2\pi f_k t + \phi_V(f_k)}}{I_{rms}(f_k)\, e^{j2\pi f_k t + \phi_I(f_k)}} = \left|\dot{Z}(f_k)\right| e^{j\phi_{\dot{Z}}(f_k)}. \tag{3.4}$$

The voltage and current parameters needed to evaluate (3.4) ($V_{rms}, I_{rms}, \phi_V, \phi_I$) were calculated by sinusoidal fitting (sinefit). In fact, a sinusoidal signal can be described as follows

$$s(t) = A_0 + A\sin(2\pi f t + \phi), \tag{3.5}$$

where A_0 is the DC component, A is the amplitude, f is the frequency and ϕ is the phase. In the analyzed case, for voltage and current signals from the battery, the frequency f is known, as it is set by the stimulus signal, and therefore only three unknown parameters are needed for the reconstruction of the signals (A_0, A, ϕ). Therefore, from a set of M samples $\{s_1, s_2, \ldots, s_M\}$ acquired at time instants $\{t_1, t_2, \ldots, t_M\}$ a least–square approach can be used to estimate the parameters of the sinusoidal signal that minimize the sum of squares of the errors, as shown below

$$\{\hat{A}_0, \hat{A}, \hat{\phi}\} = \underset{\{A_0, A, \phi\}}{\arg\min} \sum_{i=1}^{M} \left(s_i - (A_0 + A\sin(2\pi f t_i + \phi))\right)^2, \tag{3.6}$$

where $\{\hat{A}_0, \hat{A}, \hat{\phi}\}$ represent the estimated unknown parameters. This procedure is applied to both voltage and current signals in order to evaluate (3.4). It is worth noting that, because of the signal definition in (3.5), processing by sinefit is only applicable to stepped sine stimulus signals.

Finally, it is important to emphasize that the performance of all processing techniques depends on the acquisition settings of the stimulus signals. For this reason, an analysis of the performance of the techniques is carried out by varying some signal acquisition conditions such as the duration and power of the acquired signals.

Considered Measurement Methods

Figure 3.2 summarizes all the measurement methods considered in this section. In particular, all the analyzed conditions in terms of types of stimulus signals, processing techniques, power and length of stimulus signals used are shown.

Regarding the stimulus signals, 17 frequencies from 50 mHz to 1 kHz with logarithmic spacing were analyzed for both stepped sine and multisine.

Next, two different configurations of signals acquisition settings for processing based on frequency analysis were adopted, FFT–A and FFT–B. In detail, for each frequency to be analyzed, FFT–A has a fixed sampling frequency of 10 kHz and a fixed observation time of 100 s, i.e. 5 periods of the lowest analysis frequency (50 mHz). The FFT–A allows both stepped sine and multisine signals to be acquired with the same acquisition settings. About FFT–B, for each frequency to be analyzed, the number of signal periods was set equal to 10 and the number of points per period equal to 500. These settings make FFT–B a processing oriented to optimize acquisition time, compared with FFT–A, and applicable only to the stepped sine stimulus signal because the sampling frequency is variable as the frequency of the stimulus signal varies (this condition is not applicable to a multisine stimulus signal). Concerning the acquisition settings used for processing based on time analysis (sinefit), for each frequency to be analyzed, the number of signal periods was set equal to 3 and the number of points per period equal to 500. Sinefit, like FFT–B, is also a processing oriented toward optimizing acquisition time.

Furthermore, since the measurement performance is sensitive to the power of the signals, an additional consideration should be given to this aspect. First, the signals are current driven and this aspect will be detailed after. In order to compare the two analyzed stimulus signals it was chosen to set the amplitude of the stepped sine at 35 mA rms (i.e. the amplitude of each sinewave of the stepped sine), Power–A in Figure 3.2, while for the multisine two different power levels have been adopted. Power–A means that each sinewave of the multisine has an amplitude of 35 mA rms resulting in an overall amplitude of the multisine equal to 142 mA rms ($35 \cdot \sqrt{17}$, where 17 is the number of adopted frequencies). Power–B, on the other hand, intends to fix the overall amplitude of the multisine equal to 35 mA rms, resulting in 8.5 mA rms the amplitude of the individual sinewaves ($35/\sqrt{17}$).

Finally, a further analysis was conducted by varying the length of the acquired stimulus signals. As shown in Figure 3.2, starting from a record of the acquired signal (for example 5 periods for the stepped sine with FFT–A) signal reductions were conducted (2 periods and 1 period for the considered example) in order to compare the measurement performances to vary the length of the acquired signal. This procedure was carried out for all combinations of stimulus signals, processing techniques and signal power shown in Figure 3.2.

Experimental Setup

In order to carry out the comparison among the different impedance measurement methods described into the previews Section, an ad–hoc experimental set–up was developed. It is important to emphasize the need to design and implement a flexible and versatile experimental set–up. In fact, this set–up must be able to handle any type of signal to be forced on the battery and must allow the processing of signals starting from the measured raw data. Consequently, we developed a custom experimental set–up capable of performing impedance spectroscopy on batteries offering flexibility in the processes of stimulus signal generation and data acquisition.

Figure 3.19 shows the schematic diagram of the experimental set–up. The designed experimental set–up involves stimulating the battery by an impressed current signal; there are no constraints on the shape of the current to allow testing of different stimulus signals. This operation is performed by a current pump circuit that receives as input a voltage signal from an arbitrary waveform generator and provides the battery stimulus current as output. The voltage signal is supplied by a generation channel of a TiePie Handyscope HS5 usb oscilloscope. In addition, the same TiePie provides measurement of the battery current and voltage. Due to the low impedance of the batteries, a four wire connection is adopted to measure impedance by reducing the contribution of wires. A software in the MATLAB environment was developed and installed on a PC; through a usb communication protocol between PC and the TiePie it

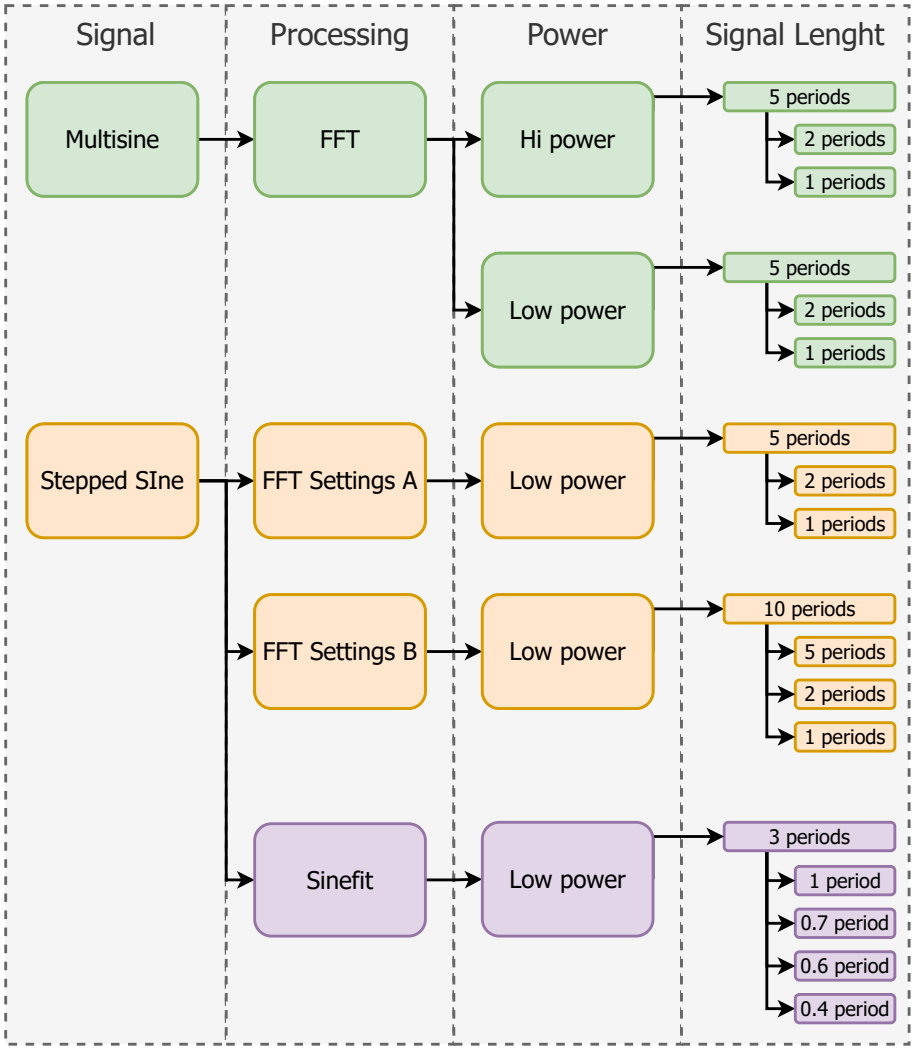
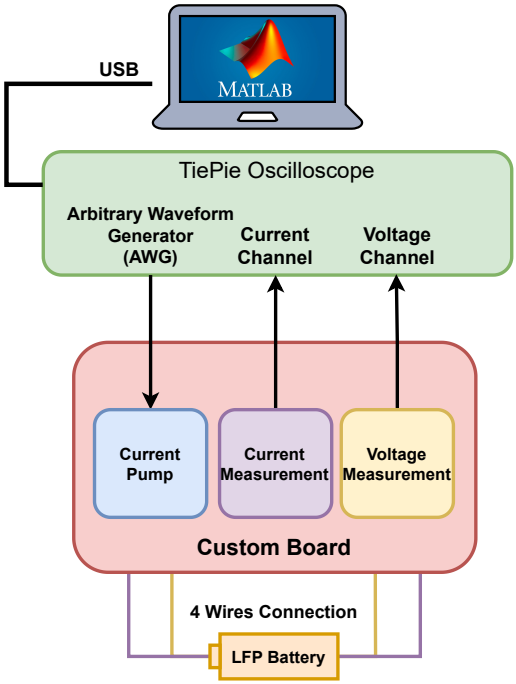


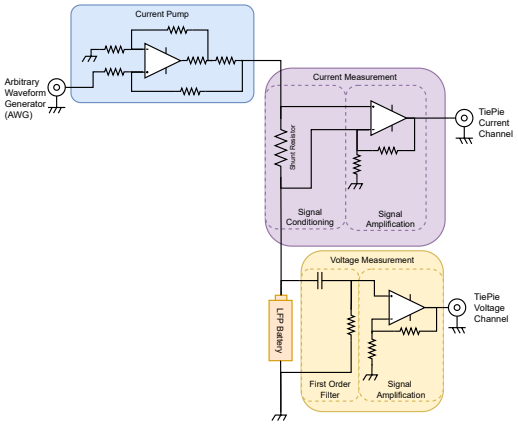
Figure 3.2: Considered measurement methods for EIS–based impedance estimation.

is possible to manage the whole experimental set–up.

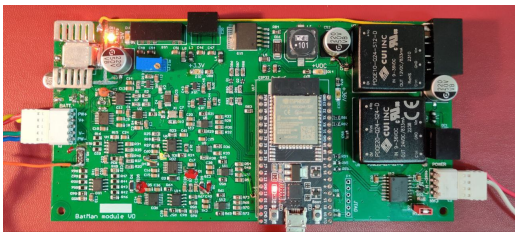
3.6f shows the circuit detail of the realized experimental set–up. Specifically, the wiring diagram of the current pump that performs the transformation from an input impressed voltage signal to an output impressed current signal is highlighted. At the output of the current pump, a shunt resistor and the battery under test are connected in series. The battery current is measured using a shunt resistor by measuring the voltage across the shunt resistor using an instrumentation amplifier in buffer configuration (i.e. gain equal to 1). The output signal is acquired by the TiePie. Regarding the battery voltage measurement, it is initially filtered using a first–order low–pass filter with a cutoff frequency equal to 5 mHz (an order of magnitude smaller than the minimum test frequency used for the battery stimulus current). The filtering operation is used to remove the DC component of the battery voltage in order to optimize signal acquisition. Downstream of filtering, due to the extremly low battery impedance, the voltage signal is amplified through an instrumentation amplifier with a gain equal to 25.7 and finally acquired by the TiePie. The adopted TiePie to acquire the battery current and voltage uses a 14–bit analog–to–digital converter. 3.3c shows the engineering of the custom board made on pcb circuit. Finally, a lithium iron phosphate ($LiFePO_4$ – LFP) battery from "NX" with a nominal capacity of 2300 mAh and a nominal voltage of 3.2 V was used to perform the



(a) *Schematic diagram of the custom experimental set–up.*



(b) *Detail of the electrical circuit of the realized set–up.*



(c) *Realized custom board.*

Figure 3.3: Implemented experimental set–up.

tests.

Validation of the custom experimental set–up

Before using the custom board for the comparison of different measurement methods regarding impedance measurement, discussed in the preview Section, it is necessary to verify the proper operation of the realized board. For this purpose, a performance comparison was carried out between the board and a reference instrument for impedance measurement on batteries. In particular, the Hioki BT4560 was used as the reference instrument, which uses sinusoidal stimulus voltage signals in a frequency range of 10 mHz to 1 kHz. The output of this instrument is directly the complex impedance of the battery.

In order to carry out the performance comparison tests, the same LFP battery was used for the tests with both the custom board and the Hioki. In both cases, before carrying out the tests the battery was charged until a SoC level of 100% was reached. In both cases, the same 17 stimulus frequencies within the frequency range of the Hioki (10 mHz – 1 kHz) were used. As for the custom board tests, a stepped sine was chosen to be used as the stimulus signal and FFT-A as the processing type. This choice respects the classical approach widely used in the literature for performing an EIS test. Each EIS test was repeated 10 times for both the

custom board and Hioki. Figure 3.4 shows the obtained results for both the custom board and the Hioki on the Nyquist plane. In this representation, the x–axis shows the real part of the impedance $Re(\dot{Z})$ while the y–axis shows the imaginary part of the impedance $Im(\dot{Z})$ and each point on the curves represents a different stimulus frequency. Performance comparisons between the custom board and the Hioki were carried out by estimating the mean values and the associated measurement uncertainties for each analyzed stimulus frequency. Relative to the estimation of measurement uncertainties, a Type A evaluation was followed for the custom board based on the performed repeated measurements, while for the Hioki the uncertainty was estimated using both the repeated measurements (Type A evaluation) and the performance information on the instrument datasheet (Type B evaluation). The estimation of uncertainty components by Type A and Type B evaluations follows what is stated in the ISO ENV 13005 "Guide to the expression of uncertainty in measurement (GUM)". The estimated uncertainties for both the custom board and the Hioki were added to the average trends Figure 3.4 as horizontal bars (for the real part of the impedance) and vertical bars (for the imaginary part of the impedance). Finally, it can be seen that the obtained trends are perfectly compatible with each other. This result validates the performance of the custom board, allowing us to use it to implement the measurement methodologies described before.

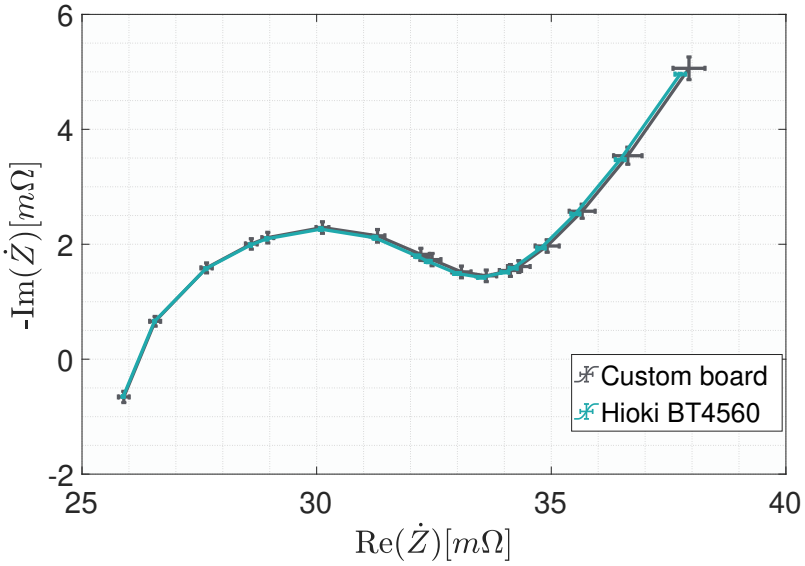


Figure 3.4: Performance comparison between the custom board and the Hioki BT4560 instrument.

Measurement Time

After the setup has been constructed, it is possible to assess the performance in terms of measurement time for all combinations of signal, acquisition settings, and processing. As can be seen from Figure 3.5, the stepped sine with the Fourier transform with type A settings requires more time, especially when considering acquisitions with 5 periods for the lowest stimulus frequency. This is not an unexpected result, as the stepped sine signal requires stimulating the system with all frequencies one after the other, naturally taking more time compared to other signal types. Performance improves significantly when using type B acquisition settings. Both type A and type B multisine signals exhibit the same measurement time, as the difference between the two signals lies solely in the stimulus signal's power, which does not

affect the measurement time. With the same number of acquired periods, the multisine signal allows for considerable time savings compared to the stepped sine signal, as it stimulates the system with all frequencies simultaneously. All processing signals, except for the stepped sine processed with the "sinefit" method, cannot be sampled with less than one period of the stimulus sinusoid for each frequency due to the using of FFT algorithm. In contrast, the "sinefit" method, utilizing a sinusoidal fitting algorithm, has a lower limit of K data points to be well conditioned. K is three for processing when the frequency is known and four when the frequency is unknown. This allows for acquiring less than one period of the stimulus sinusoid, resulting in significant time savings compared to the classic stepped sine signal processed with the Fast Fourier Transform.

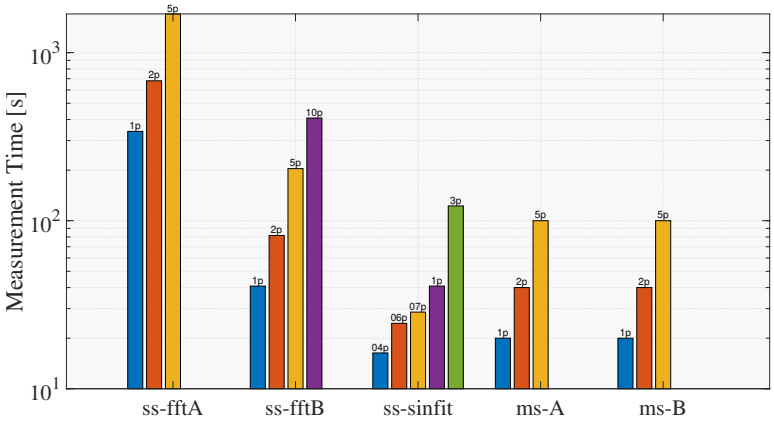


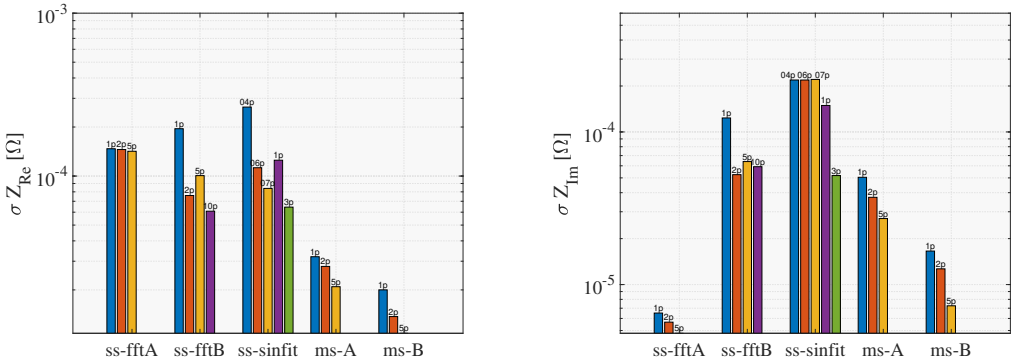
Figure 3.5: Measurement time comparison with all the different combinations of signal, acquisition settings, and processing.

Measurement Repeatability

To assess measurement performance, 12 tests were conducted for each combination of signal acquisition settings and processing. These 12 tests were carried out while keeping the sample's temperature and state of charge constant. Therefore, variations among the 12 tests can be attributed solely to the measurement system's performance.Figure 3.6 illustrates the impedance variations at three different frequencies (100 mHz, 8 Hz, 500 Hz), taken as examples. Since impedance is a complex quantity, it was preferable to separate the real and imaginary parts.

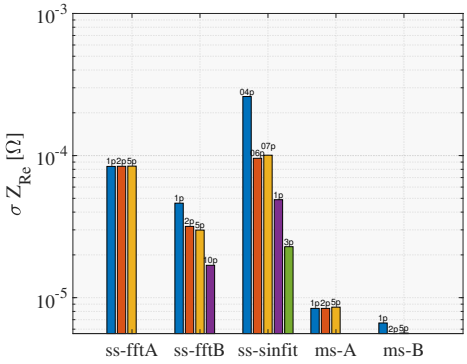
In general, it is evident that the best repeatability can be attributed to the multisine signal with type B settings. Interestingly, despite this signal having lower power compared to the type A multisine, it still exhibits superior repeatability performance. Repeatability appears to be significantly influenced by the number of acquired periods, i.e., the observation time of the signal.

the choice of a signal or processing over the others could be done only define the applicant scenario, because the application gives constrains in terms of measurement time, measurement accuracy, and repeatability. This work is currently ongoing and his aim is to understand the better combination of signals, acquisition settings, and processing to fit the parameters of equivalent circuits and how the measurement performance impacts on the parameters' value.

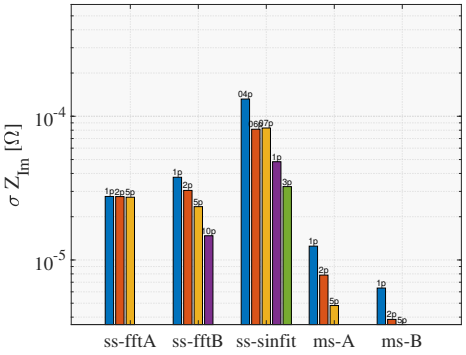


(a) $\sigma Re(\dot{Z}(100\ mHz))$

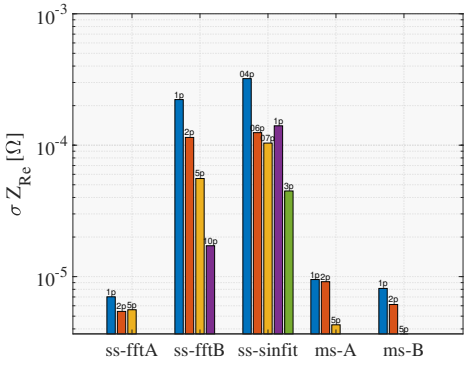
(b) $\sigma Im(\dot{Z}(100\ mHz))$



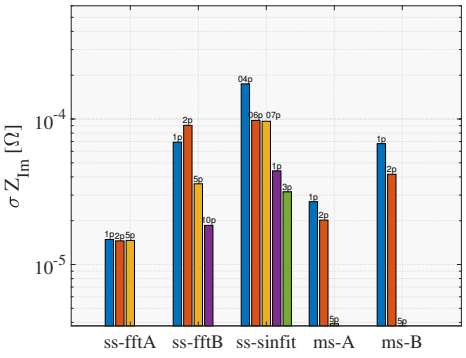
(c) $\sigma Re(\dot{Z}(8\ Hz))$



(d) $\sigma Im(\dot{Z}(8\ Hz))$



(e) $\sigma Re(\dot{Z}(500\ Hz))$



(f) $\sigma Im(\dot{Z}(500\ Hz))$

Figure 3.6: Repeatability in terms of standard deviation through 12 repetitions of the different signals, acquisition settings, and processing.

3.2 Stimulus Frequencies Selection

The primary objective of a Battery Management System (BMS) is to ensure battery security and maximize its lifespan. This goal is achieved through the continuous and precise monitoring and control of the battery's State of Charge (SoC) and State of Health (SoH). Presently, the predominant techniques for estimating SoC rely on Coulomb counting and Open Circuit Voltage (OCV), as detailed in chapter 1. Notably, the application of OCV to Lithium Iron Phosphate (LFP) batteries presents challenges due to their voltage-SoC relationship, which exhibits limited sensitivity. This means that even a minor voltage measurement error can result in a substantial SoC estimation error. Electrochemical impedance spectroscopy (EIS) emerges as a promising alternative to Coulomb and OCV counting methods for SoC estimation, as highlighted in references [10, 87]. EIS offers the advantage of not being susceptible to the aforementioned issues and can potentially be implemented in real-world operating conditions, as discussed in [65]. EIS techniques involve generating stimulus signals, typically sinusoidal waves, and measuring both battery voltage and current. These stimulus signals typically span frequencies ranging from a few millihertz to a few kilohertz, as indicated in references [68, 89]. The primary challenge associated with EIS pertains to extended measurement times due to the utilization of lower-frequency stimulus signals, which considerably hinders the feasibility of real-time SoC estimation.

The aim of this section is to outline methodologies for optimizing the use of Electrochemical Impedance Spectroscopy (EIS) in battery State of Charge (SoC) estimation. The proposed approach is versatile and can be applied in any scenario where the challenge of designing a real-time SoC estimation system using EIS arises.

Depending on the specific requirements of different applications, including desired accuracy levels and the acceptable measurement time for SoC estimation, the proposed method facilitates the selection of optimal EIS frequencies to meet imposed constraints. Essentially, this method has the potential to optimize two out of three key parameters in SoC estimation: the number of SoC classes, target accuracy, and measurement time. For instance, when the number of SoC classes is predetermined, the method allows for the maximization of classification accuracy without constraints on measurement time or the minimization of measurement time without compromising classification accuracy. Alternatively, it can find a balanced trade-off between classification accuracy and measurement time when both parameters are subject to specific application constraints.

The method is structured around two main stages:

- **Experimental Characterization**: This stage involves the empirical study of the actual behavior of the batteries under consideration.

- **Time-Accuracy Optimization**: This stage relies on suitable feature selection and Machine Learning (ML) techniques. In the context of EIS, a range of stimulus frequencies spanning from millihertz to kilohertz is typically employed, with the operator having control over the number of frequencies used. A greater number of frequencies yields a more comprehensive spectrum knowledge but prolongs measurement time. The critical aspect of this optimization is selecting the minimum number of measurements at key frequencies. By adopting an ML approach, each measurement at a different frequency can be treated as a feature for the algorithm. Consequently, this problem can be effectively addressed through a feature selection (FS) approach. FS serves to reduce the number of input variables when constructing a predictive model. The advantages of FS encompass enhanced model performance, mitigation of overfitting, increased interpretability, and decreased computational complexity, as discussed in [72].

The FS approach was also employed in a study referenced in [8], where a dataset comprising 54 different frequencies measured on Li-ion batteries across five distinct temperatures was utilized. In this study, the authors employed a method based on Pearson's Correlation Coefficient to identify frequencies with the highest correlation to State of Charge (SoC). These selected frequencies were subsequently used as inputs for a Machine Learning (ML) model and linear regression analysis. However, the specific frequencies chosen and the extent of their impact on reducing measurement time remain somewhat unclear in their approach. Furthermore, as highlighted in this section, the correlation-based selection method demonstrated inferior performance compared to other proposed fitness functions. Another notable application of feature selection (FS) pertains to assessing the health status of batteries, as demonstrated in [76]. In this study, the authors conducted FS using voltage, temperature, and current measurements to identify indicators most strongly correlated with the health status of Lithium Iron Phosphate (LFP) batteries. These batteries were subjected to aging through 121 equivalent charge and discharge cycles. The features used in this analysis included the time the battery spent between two voltage thresholds, temperature data, current measurements, and energy-related metrics, resulting in a total of 69 features. Following a thorough analysis, the features were ranked based on their correlation to State of Health (SoH). By establishing a correlation threshold, it was possible to select the most significant features for SoH estimation. The proposed measurement method extend the research presented in [16], which aimed to reduce the time required for Electrochemical Impedance Spectroscopy (EIS) by employing a feature selection technique grounded in Genetic Algorithms (GA). This research effort extends in multiple directions. To enhance the validity of the findings, the experimental dataset was expanded to include a larger number of batteries than those considered in [16]. Additionally, the study explored the effectiveness of the Particle Swarm Optimization algorithm, as an alternative search strategy to GA-based techniques, as introduced in [111]. Furthermore, two distinct fitness functions were implemented. The first function served to assess the performance achieved when training a supervised learning model, employing a wrapper approach. In this context, various Machine Learning (ML) models were tested to determine the most suitable one. Conversely, the second fitness function was based on a correlation analysis employing Pearson's Correlation Coefficient.

The Proposed Method

Utilizing impedance spectroscopy as the method for State of Charge (SoC) assessment within Battery Management Systems (BMS) offers an alternative to relying on Coulomb counting or Open Circuit Voltage methods. These conventional methods have inherent limitations, often exhibiting divergence over time or reduced reliability during battery charging and discharging phases. However, impedance spectroscopy presents a significant drawback in the form of extended measurement times, primarily stemming from the low inspection frequencies, resulting in measurement durations on the order of minutes. Such prolonged measurement times may not align with the requirements of certain applications. To mitigate measurement time while preserving the efficacy of SoC estimation, a data-driven methodology for SoC assessment is introduced. This method is specifically tailored to address the challenge of designing a real-time SoC estimation system utilizing the Electrochemical Impedance Spectroscopy (EIS) technique. Depending on the unique demands of diverse applications, including the desired level of accuracy and the acceptable measurement duration for SoC estimation, the proposed approach facilitates the optimization of frequency selection within EIS to meet predefined constraints effectively. Within this methodology, the selection of optimal frequencies is achieved through feature selection techniques employing search algorithms and

various fitness functions.

The method is summed up in Figure 3.7 reported in [17]

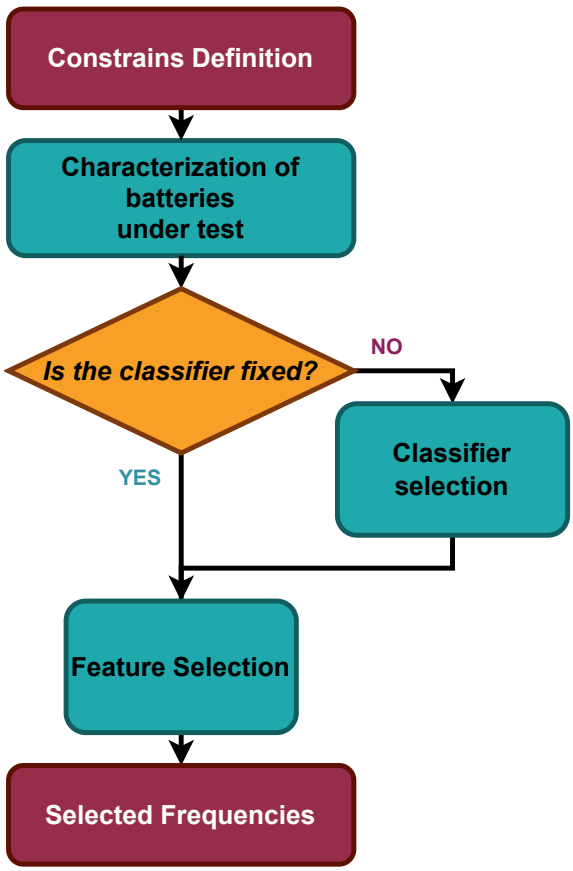


Figure 3.7: The proposed method workflow [17].

The first operation is to define the required constraints and the parameters, below all the requirements are listed:

- Resolution of SoC output estimation: the higher the resolution the longer the characterization step and in general the greater the number of classes and the greater the classification error.
- Target measurement time: the target measurement time is the maximum acceptable time to obtain the measurement.
- Accuracy target: is the minimum acceptable Accuracy of the classifier for SoC estimation.
- The type of batteries: the characteristics of the batteries you want to use (chemistry, nominal capacity, nominal voltage etc.).
- Classifier: the proposed method uses a classifier built into the search algorithms to estimate the population. The choice of classifier is critical because the next optimization step will be stitched around the classifier, in other words, the features selected also depend on the classifier adopted.

The classifier is sometimes defined by design parameters related to hardware requirements, while at other times the choice of the classifier is related solely to classification performance.

The proposed method caters to both of these aspects: the classifier can either be specified as an input parameter or chosen through a preliminary step to identify the most suitable classifier for the dataset. This selected classifier will serve as the data-driven mechanism employed by the Battery Management System (BMS) to estimate State of Charge (SoC).

The subsequent step involves the characterization of the devices under examination. During this characterization phase, stringent parameters are employed to ensure the highest possible SoC resolution and an extensive array of measured features. The selection of batteries for inclusion in the database should align with the anticipated application of the BMS. Employing diverse batteries of the same model imparts generality to accommodate various battery types, just as incorporating batteries at varying states of Health (SoH) ensures adaptability to different SoH scenarios.

In cases where the initially chosen classifier does not conform to the design constraints, an evaluation is performed to determine the most suitable classifier for the dataset obtained in the preceding section. Among a set of pre-selected classifiers, the one that delivers superior accuracy is subsequently integrated into the feature selection algorithm.

The final stage of the proposed methodology is feature selection, where search algorithms are employed to select features. This process serves dual objectives: reducing measurement time to a value equal to or less than the target time while simultaneously attaining accuracy surpassing the prescribed target level.

The proposed method: Characterization of batteries under test A study aimed at the selection

of Features cannot disregard an initial phase of collecting an adequate dataset. There are two possible ways: (i) the use of public datasets, i.e. datasets produced by researchers and then made available to the community for further analysis or (ii) the creation of a measurement setup to collect the measurement data independently. A number of public datasets of impedance measurements on batteries were explored but the characteristics required for this study in terms of number of measurements, stimulus frequencies and temperature did not allow for the use of a public dataset. It was therefore decided to create a setup and acquire the data independently. The dataset collection and the related experimental setup will be illustrated in detail in the next paragraph and Figure 3.19.

At the end of the measurement collection the obtained dataset, as shown in Figure 3.8 [17], has 20 levels of SoC, and 28 complex impedances for each SoC for a total of 7 different LFP batteries.

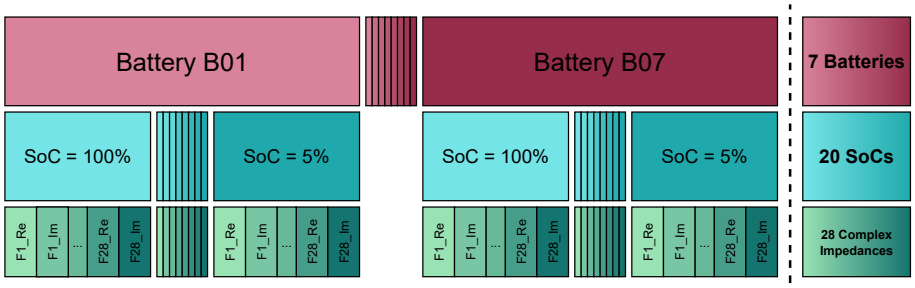


Figure 3.8: Structure of the entire obtained experimental dataset [17].

The proposed method: Classifier Selection

As shown in Figure 3.7 the adopting classifier could be a parameter or could be fixed by

external limits (i.e. computational limit). For this reason, the feature selection algorithm is built to properly work with all types of classifiers The initial experiments consisted of two distinct phases. In the first phase, multiple ML models were examined to identify a subset of algorithms that best represented the problem at hand, taking into consideration both their achieved performance and computational complexity. The second phase involved a more in-depth analysis of the selected models to assess their suitability for real-world applications.

For the SoC estimation task, a 10-class classification approach was adopted, where each class represented a 10% SoC interval, resulting in estimation intervals such as [0-10], (10-20], and so on up to (90-100]. All available features were utilized, totaling 28 impedances for each SoC (comprising Real and Imaginary components) acquired at various frequencies, resulting in a total of 56 features used in the analysis. Notably, no normalization techniques or general pre-processing methods were applied because conventional techniques did not yield significant performance improvements across various models.

During these experiments, key metrics commonly associated with multiclass classification problems, as highlighted by Grandini et al. in [41], were employed for performance assessment. These metrics included Accuracy (Acc), Matthews Correlation Coefficient (MCC), Precision (P), Recall (R), and F1 Score (F1). It's worth noting that these metrics, originally designed for binary classification, were adapted for use in a multiclass context by averaging the unweighted mean per label.

A comparison of different classifiers is performed in this phase, the aim is to identify the most suitable classifier for this problem. A classical division 80-20 is applied to split the dataset into two parts where the bigger part is used to train the models while the second part is used to test them. This approach is well established in the literature because the algorithms are tested on data that are not used for the training. The classifier selection is conducted in two phases, first, the most promising classifiers are selected from a series of classifiers, these are then tuned and adjusted in order to increase the classifiers' performances. In **??** [17] all the first phase tested classifiers are reported. Based on this analysis, three models were selected for more in-depth performance evaluation. The selected models are as follows:

- K-Nearest Neighbors: Known for its low computational complexity during training but not particularly high performance.
- Random Forest: Chosen for its favorable balance between performance and complexity.
- Support Vector Machine: Selected with the aim of achieving the highest performance possible.

The previously chosen models were employed in subsequent testing phases. Specifically, the primary objective was to ensure controlled separation of data for training and testing, ensuring that measurement data from the batteries did not overlap between the training and testing sets. These experiments consistently followed the k-fold method, employing six folds in which measurements from two batteries were allocated for testing, while the remaining data was used for training. This approach was designed to maximize dataset utilization, resulting in six trained models for each classification algorithm. The results obtained are presented in Table 3.2 . The metrics considered in the evaluation confirm that the Support Vector Machine (SVM)-based model performed the best, achieving a mean accuracy of 0.83 with a standard deviation of 0.04. The resulting confusion matrix is depicted in Figure 3.9.

As a result of these preliminary classification tests, the SVM emerges as the most effective ML model among those under consideration. Consequently, in subsequent Feature Selection (FS) methodologies that involve the use of a classification model, the SVM model will be the one employed.

Table 3.1: Average values of the considered evaluation metrics and main best obtained hyper-parameters for the ML models.

Model	Acc	MCC	P	R	F1	Hyper–parameter	Value
AdaBoost (DecisionTree)	0.34	0.31	0.32	0.36	0.27	Base estimator	Decision Tree
						Number of estimators	50
						Learning rate	0.1
Logistic Regression	0.45	0.39	0.43	0.45	0.43	Inverse of regularization strength	10000
						Norm of the penalty	l2
Multi–Layer Perceptron	0.55	0.52	0.52	0.57	0.51	Activation function	ReLu
						Hidden layer sizes	100, 200, 100
						Learning rate	Constant
						Solver	Adam
K–Nearest Neighbor	0.57	0.53	0.59	0.57	0.57	Number of neighbors	3
						Minkowski metrics	Euclidean
						Weight function	Distance
Gaussian Naive Bayes	0.68	0.65	0.72	0.68	0.67	Variance smoothing	1e-9
Bagging (DecisionTree)	0.75	0.72	0.77	0.74	0.74	Base estimator	Decision Tree
						Number of estimators	50
Random Forest	0.75	0.73	0.78	0.74	0.74	Criterion	log_loss
						Number of estimators	100
Support Vector Machine	0.89	0.88	0.92	0.89	0.89	Regularization parameter	10000
						Kernel type	RBF
						Kernel coefficient(gamma)	1
						Decision function	one-vs-one

The proposed method: Fatures Selection

Feature Selection (FS) is used to reduce the number of utilized features by eliminating those that are extraneous or contain noise. Let's consider a collection of data samples for classification denoted as $\mathcal{Z}$, which is described by a set Y consisting of N features. The FS task involves the identification of the optimal subset $X^* \in Y$ composed of M $(M < N)$ features. This selection is guided by an objective function denoted as J. The function $J(X)$ assesses the discriminative capability of the feature subspace represented by the subset X when the samples in $\mathcal{Z}$ are projected onto it. To achieve this goal, a well-defined search strategy, an evaluation function, and a stopping criterion are essential components. Nevertheless, owing to the exponential expansion of the solution space, which encompasses all 2^N conceivable subsets of Y,

Table 3.2: Average values and standard deviation of the considered evaluation metrics examined, where Train and Test datasets are divided by batteries [17].

	k-NN	Random Forest	Support Vector Machine
Acc	0.62 ± 0.12	0.79 ± 0.05	0.83 ± 0.04
MCC	0.58 ± 0.13	0.77 ± 0.05	0.82 ± 0.04
P	0.65 ± 0.11	0.81 ± 0.05	0.85 ± 0.04
R	0.63 ± 0.12	0.79 ± 0.05	0.84 ± 0.04
F1	0.62 ± 0.13	0.79 ± 0.05	0.84 ± 0.04

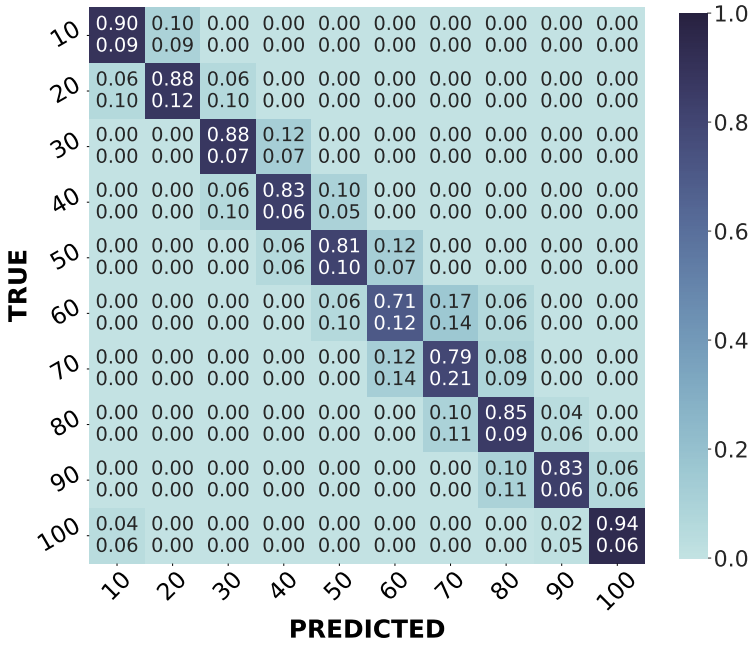


Figure 3.9: Obtained confusion matrix of the Support Vector Machine model, with mean value (top) and standard deviation (bottom) for each class [17].

conducting an exhaustive search to identify the optimal subset X^* becomes infeasible, particularly when dealing with a large number of features (i.e., $N > 50$). In the realm of Feature Selection (FS), algorithms typically need to specify both an evaluation function and a search procedure. Evaluation functions can be categorized into two general classes: filter and wrapper methods. Filter methods gauge the relevance of features based on statistical measures or scores. They evaluate a feature subset independently of the Machine Learning (ML) algorithm being employed. Common examples of filter methods encompass correlation-based FS, mutual information, and chi-square tests. In contrast, wrapper methods assess the performance of an ML model with various feature subsets. They entail training and evaluating the model using diverse combinations of features and ultimately selecting the subset that yields the best performance. It's worth noting that wrapper methods can be computationally intensive but generally offer more precise results. Evolutionary Computation (EC) techniques have gained widespread utilization in this domain due to their renowned capacity for global exploration. Furthermore, EC techniques operate without requiring domain-specific knowledge and make no presumptions regarding the nature of the evaluation function, such as linearity or differentiability.

In this study, we tackled the challenge of determining the optimal set of frequencies for impedance measurements via Electrochemical Impedance Spectroscopy (EIS) in the context of battery State of Charge (SoC) estimation. To attain this objective, we employed two optimization algorithms as search strategies. The first is rooted in the Genetic Algorithm (GA), while the second is founded on Particle Swarm Optimization (PSO). Additionally, for both of these search algorithms, we incorporated two distinct fitness functions: the first is based on a supervised learning model, and the second relies on a correlation analysis. We will elaborate on

the proposed approaches, including the selection of search algorithms and the implementation of fitness functions, in the subsequent sections.

Figure 3.25 reports both the implemented searching algorithm.

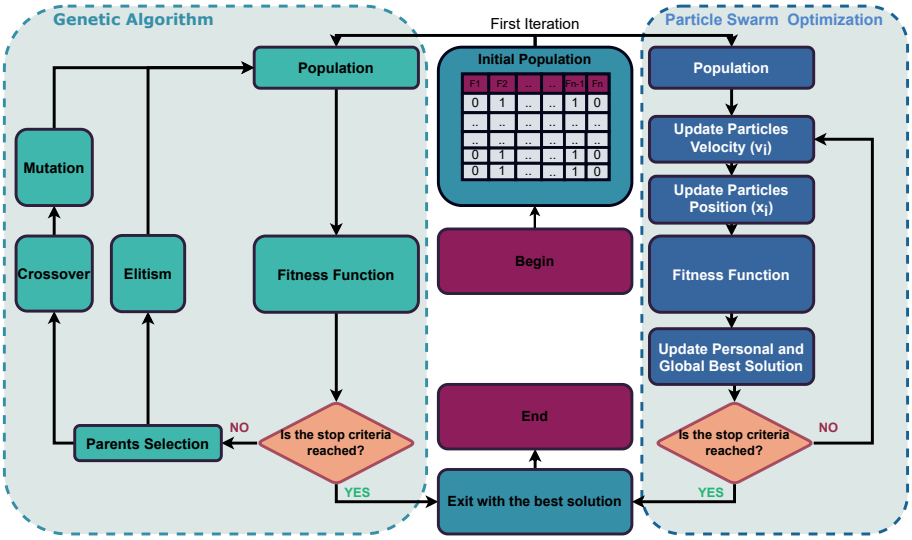


Figure 3.10: Genetic Algorithm (left) and Particle Swarm Optimization (right) flow charts [17].

A Genetic Algorithm (GA) mimics the natural process of evolution by creating a population of potential solutions, each represented as binary strings. These solutions are then subjected to crossover and mutation operations to explore the search space effectively [116]. In our implemented approach, we employed a generational GA, where individuals were depicted as binary vectors that encoded specific feature subsets. Each individual took the form of a binary vector with a fixed length equal to N, with feature inclusion and exclusion represented by 1 and 0, respectively. The algorithm's flowchart can be seen on the left side of Figure 3.25.

To initiate the algorithm, we generated a population of individuals denoted as P, with each individual's values randomly set as either 1 or 0. Subsequently, we applied a fitness function to assess each individual's fitness, taking into account the frequencies encoded by that particular individual. After the evaluation phase, we created a new population. Initially, we implemented an elitist strategy, copying the best individuals from the current population to retain the top elements. The remaining $(P - e)/2$ pairs of individuals, referred to as parents, were chosen through a tournament selection method with a size of t. This selection method involved picking an individual from the population by selecting the best among t randomly chosen individuals. We applied the uniform crossover operator to each selected pair, followed by the mutation operator, with both operations having probabilities equal to pc and pm, respectively. The resulting offspring were then evaluated and added to the new population. This entire process was reiterated for Ng generations. You can find the parameter values for pc, pm, and Ng in Table 3.3.

The Particle Swarm Optimization (PSO) algorithm [56] stands out as a global optimization technique particularly well-suited for addressing problems where the optimal solution resides within a multidimensional space. PSO can be applied to optimize both real-valued and discrete-data scenarios. These swarm-based algorithms belong to a family of nature-inspired, population-based methodologies known for their ability to generate robust solutions for a variety of intricate problems. A swarm comprises a population of simple, homogeneous agents that execute elementary tasks through interactions with one another and their environment.

Table 3.3: Genetic Algorithm parameters.

Parameter	Symbol	Value
population size	P	100
selection type	-	tournament
tournament size	t	2
elitism	e	2
crossover probability	p_c	0.6
mutation type	-	random
mutation number of genes	m_{size}	1 bit
generation number	N_g	250

From a mathematical standpoint, as expressed in Equation 3.7, assuming we are tackling a maximization problem and denoting the objective function as f, our objective is to locate the coordinate point x_{best} within the search space $\mathbb{R}^D$. The value of $f(x)$ is commonly referred to as the "fitness" of the point. In the context of PSO, each potential solution (x_i) is termed a "particle" and represents a point within the search space.

$$\begin{aligned} &f : \mathbb{R}^D \longrightarrow \mathbb{R} \\ &x_{best} | f(x_{best}) \geq f(x) \quad \forall x \in \mathbb{R}^D \end{aligned} \tag{3.7}$$

The ensemble of these elements constitutes the swarm. In its quest to find the problem's solution, the swarm continually updates its positions through an equation of motion, as illustrated in (3.8). Here, v_i is a D-dimensional vector that represents the velocity components of the i-th particle. The time instants t and $t+1$ correspond to two successive iterations of the algorithm (Marini et al., 2015). The algorithmic progression of this iterative process is depicted in Figure 3.25.

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{3.8}$$

To address the problem at hand, we applied the optimization algorithm to discrete data [58]. In this context, each particle is represented by a binary vector that characterizes a feature subset. The parameters employed in the subsequent experiments are detailed in Table 3.4.

Table 3.4: Particle Swarm Optimization parameters [17].

Parameter	Symbol	Value
optimizer	-	Binary PSO
cognitive parameter	c1	2
social parameter	c2	2
inertia parameter	w	0.8
iterations	i	250
number of particles	n	100

As previously mentioned, two fitness functions were implemented for each of the considered search algorithms. These fitness functions play a pivotal role in determining which features (and consequently, which frequencies) should be utilized for State of Charge (SoC) estimation. The underlying concept behind the definition of these fitness functions encompasses two primary objectives: maximizing the information content of the features and minimizing the overall measurement time.The measurement time is computed as the sum of all the

selected stimulus periods. Given that the measured impedances consist of complex numbers, comprising both real and imaginary components, two features are generated for each stimulus frequency. It's important to note that the impedance measurement instrument outputs both components simultaneously for each measurement. Consequently, the measurement time is calculated by considering a measurement period whenever one or both of the features at the same frequency are taken into account.

A fitness function was developed and evaluated based on an SVM (Support Vector Machine) classifier because, as previously detailed, it demonstrated superior performance in classifying the State of Charge based on the acquired data. The structure of the fitness function is presented in (3.9). The SVM operates as a supervised classifier that partitions the feature space using hyperplanes. The optimal hyperplane is chosen based on its effectiveness in separating different classes within the test dataset. In the context of Feature Selection (FS), the feature space undergoes changes in each iteration. For instance, let Ch represent one of the chromosomes containing M features. The algorithm utilizes the M-dimensional feature space from the training sub-dataset to identify the most suitable hyperplane. Subsequently, a test is conducted on the test sub-dataset using the same M-dimensional feature space, and the resulting SoC estimation accuracy is assessed as described in (3.10), where the parameter A denotes the ratio of correct predictions (CP) to total predictions (TP).

The primary objective of FS in this application is to reduce measurement time while simultaneously preserving or achieving a satisfactory compromise with SoC estimation accuracy. To this end, measurement time is considered through the B parameter, as detailed in (3.11). This parameter exhibits an inverse relationship with the duration of the measurement process. In this context, T_{meas} represents the measurement time, calculated as the sum of the durations of the selected features (frequencies). Conversely, the maximum measurement time, T_{max}, is determined by aggregating the durations of all features. In our specific case, T_{max} is approximately 4 minutes. Utilizing measurement periods instead of the number of features enables the algorithm to minimize the feature count while prioritizing higher frequencies over lower frequencies, as the impact of lower frequencies on measurement time is more pronounced. Moreover, both A and B terms are confined within the range of $[0, 1]$. The weight coefficient α offers a mechanism to adjust the balance between the contributions of SoC estimation accuracy and measurement time considerations.

$$S = \alpha \cdot A + (1 - \alpha) \cdot B \tag{3.9}$$

$$A = \frac{CP}{TP} \tag{3.10}$$

$$B = 1 - \frac{T_{meas}}{T_{max}} \tag{3.11}$$

The fitness function, based on Pearson's Correlation Coefficient (PCC) as defined in (3.12) for paired data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, serves the purpose of enabling the identification of a feature subset with minimal redundancy in the dataset. For every solution generated by the algorithms, we compute the correlation matrix among the selected features. Subsequently, an index associated with the individual is derived by calculating the arithmetic mean of the unique elements within the matrix. Additionally, the optimization problem takes measurement time into consideration. The resultant function is formulated as a maximization problem, as shown in (3.13), and consists of two components weighted by an α weight coefficient. The first component, denoted as C and presented in (3.14), represents the correlation index of the solution under examination. Meanwhile, the second component, denoted as B, exhibits an inverse relationship with measurement time, as explained previously in (3.11).

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.12}$$

$$S = \alpha \cdot C + (1 - \alpha) \cdot B \tag{3.13}$$

$$C = 1 - \frac{1}{n}\sum_{i=1}^{n} |r_{xy}(i)| \tag{3.14}$$

Performance Evaluation

In this section, the results of the Feature Selection (FS) process, which involves the application of the previously discussed search algorithms and fitness functions to the experimental dataset as previously described. Specifically, both SVM and PCC-based fitness functions were implemented on the GA and PSO search algorithms, yielding four potential combinations: GA+SVM, PSO+SVM, GA+PCC, and PSO+PCC. The experimental parameters for GA and PSO, used consistently throughout the subsequent experiments, were determined following preliminary trials, as outlined in Table 3.3 and Table 3.4.

To assess the performance of the tested systems, a total of 50 independent runs were conducted, each initiated with a distinct randomly generated initial population. At the conclusion of each run, the best individual discovered in that particular run was selected as the output solution. The results presented herein represent the best outcome in terms of accuracy among the 50 runs conducted. The hyperparameters employed for SVM remained consistent with those detailed in Table 3.1. In each run, the model construction followed the structure described before, with one of the available folds being randomly chosen. Furthermore, the achieved performances were evaluated while considering various α weight coefficients within the implemented fitness functions, as outlined in (3.9) and (3.13). Two distinct application scenarios were considered. In the first scenario, the objective was to select all solutions that exceeded a predefined target classification accuracy, without regard to the time involved. In contrast, the second scenario introduced multiple objectives: selecting the solution with the shortest measurement time while ensuring that there was no degradation in classification performance compared to the scenario using all features. To address these objectives, Figures 3.11-3.14 illustrates the trends of the maximum accuracy and measurement time achieved for each α value, encompassing all combinations of fitness functions and search algorithms.

In this practical case study, we established a target accuracy of 0.95, irrespective of the measurement time, as outlined in the requirements. Figures 3.11 3.12 3.13 3.14 presents the maximum accuracy achieved over 50 runs along with the corresponding measurement time as a function of the α weight coefficient. The blue star-marked point signifies the solution with the highest accuracy. Additionally, the shaded band illustrates the SoC estimation accuracy obtained when using all features with the SVM classifier, along with its associated uncertainty, as presented in Table 3.2.

For accuracy, higher values represent superior solutions, whereas for measurement time, lower values indicate better solutions. Regardless of the combination of fitness function and search algorithms, accuracy increases as the alpha coefficient increases. This trend emerges because, with lower alpha values, the search algorithms tend to generate solutions that prioritize minimizing measurement time over maximizing classifier accuracy. Conversely, higher alpha values prompt the search algorithms to prioritize accuracy at the expense of minimizing measurement time.

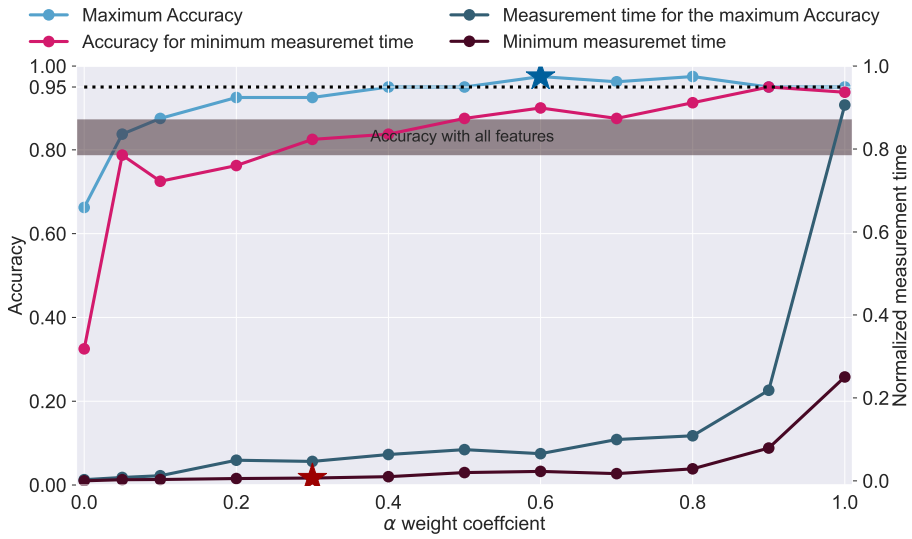


Figure 3.11: Accuracy and measurement time as a function of α weight coefficient for GA+SVM combination [17].

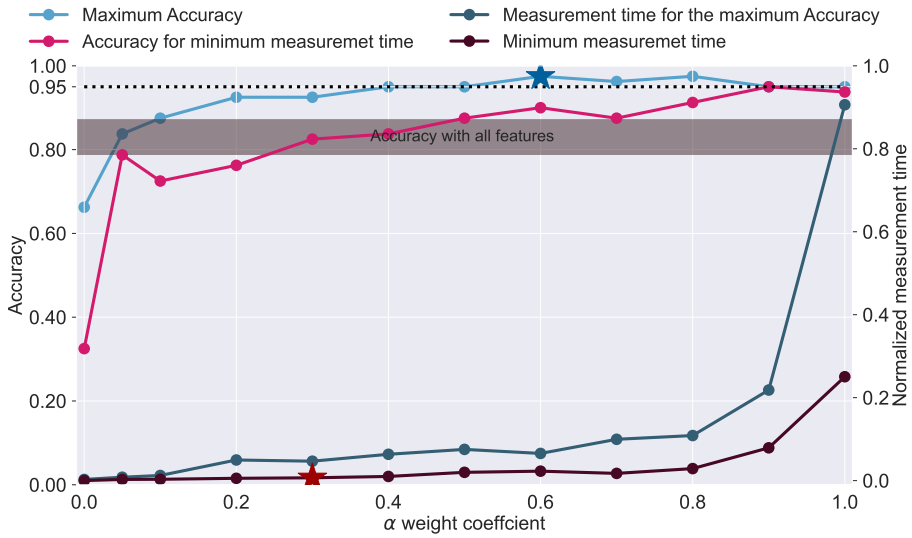


Figure 3.12: Accuracy and measurement time as a function of α weight coefficient for PSO+SVM combination [17].

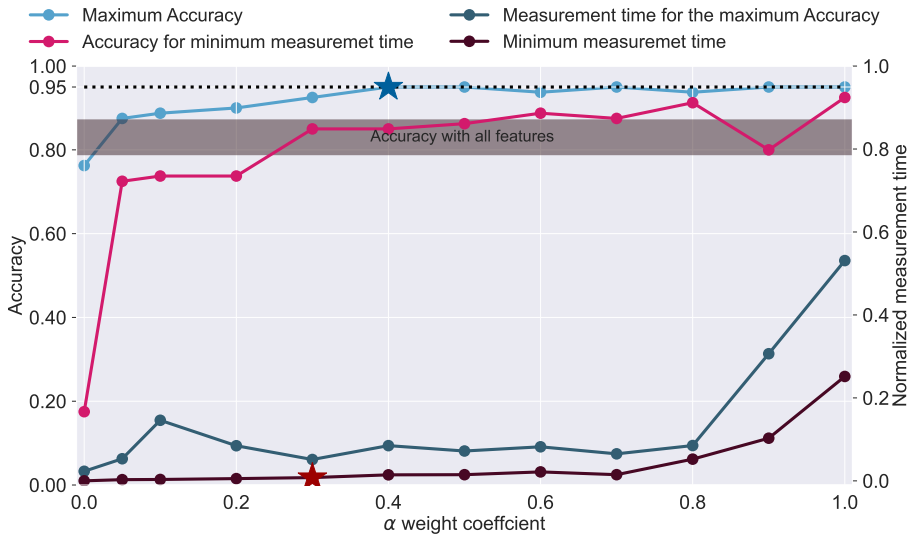


Figure 3.13: Accuracy and measurement time as a function of α weight coefficient for GA+PCC combination [17].

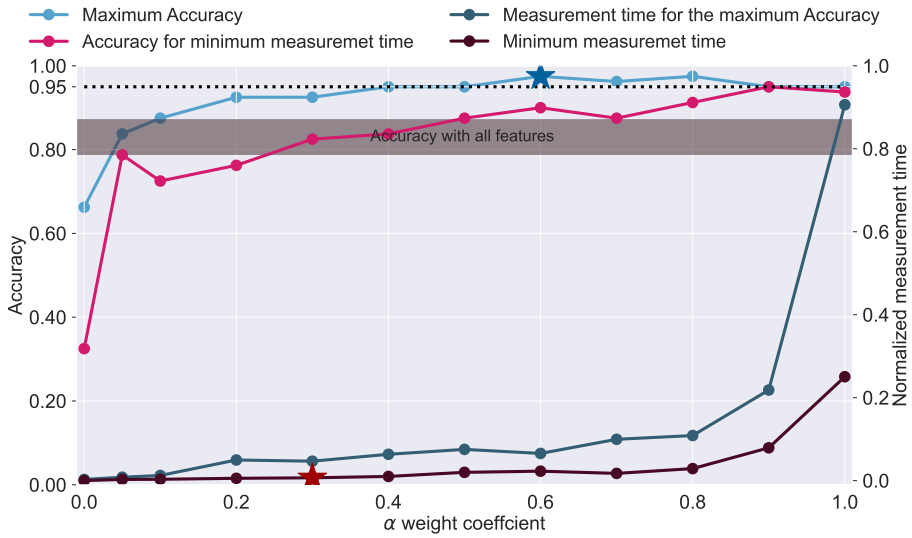


Figure 3.14: Accuracy and measurement time as a function of α weight coefficient for PSO+PCC combination [17].

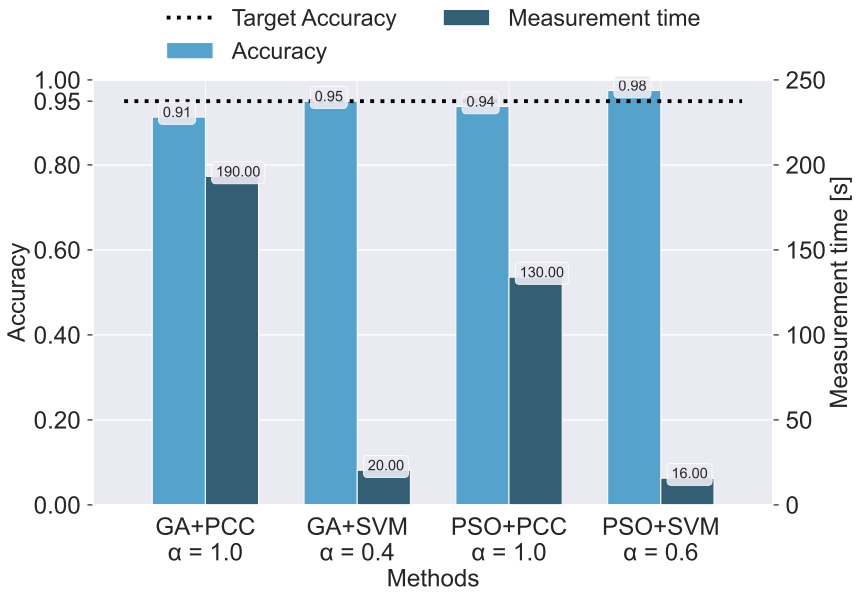


Figure 3.15: Accuracy and time performances of the solutions with the higher SoC classification Accuracy [17].

When assessing the comparative solution-searching performances of GA and PSO, there are no significant disparities in accuracy between the obtained solutions, as observed in Figures 3.11, 3.12, 3.13, and 3.14. This suggests that their performances are comparable.

Regarding the fitness functions, it is evident that the one based on the SVM classifier consistently outperforms the PCC algorithm in terms of accuracy. In fact, the target accuracy is achieved only when the fitness function incorporates the SVM classifier. An interesting finding is that feature selection can even enhance SoC evaluation performance. There are several cases where the accuracy surpasses that achieved with all features. This phenomenon can be attributed to the presence of ambiguous features that may mislead the classification algorithm. Thus, even when there is no imperative need to reduce measurement time, this method can exclude suboptimal features and improve overall classification accuracy.

Figure ?? provides a detailed view of the star-marked points from Figures 3.11, 3.12, 3.13, and 3.14. The left axis presents accuracy, while the right axis depicts measurement time for all combinations of search algorithms and fitness functions. It becomes evident that the most effective combination for the acquired dataset is PSO+SVM, which achieves an accuracy of 0.98 and a measurement time of 16 seconds. In this case, even without a specific focus on optimizing time, the method saves over 93% of the total measurement time, which was initially set at 240 seconds.

The employed Feature Selection (FS) methodology relies on multidimensional search algorithms, specifically utilizing the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) to determine the most pertinent features within the Electrochemical Impedance Spectroscopy (EIS) signals. Two distinct fitness functions have been taken into account, one grounded in supervised learning, notably the Support Vector Machine (SVM), and the other centered on feature correlation assessed through Pearson's Correlation Coefficient (PCC). The data source employed for driving the FS algorithms comprises an experimental EIS-derived dataset encompassing seven Lithium Iron Phosphate batteries.

In scenario A, the primary objective is to attain a SoC classification Accuracy of 0.95, with

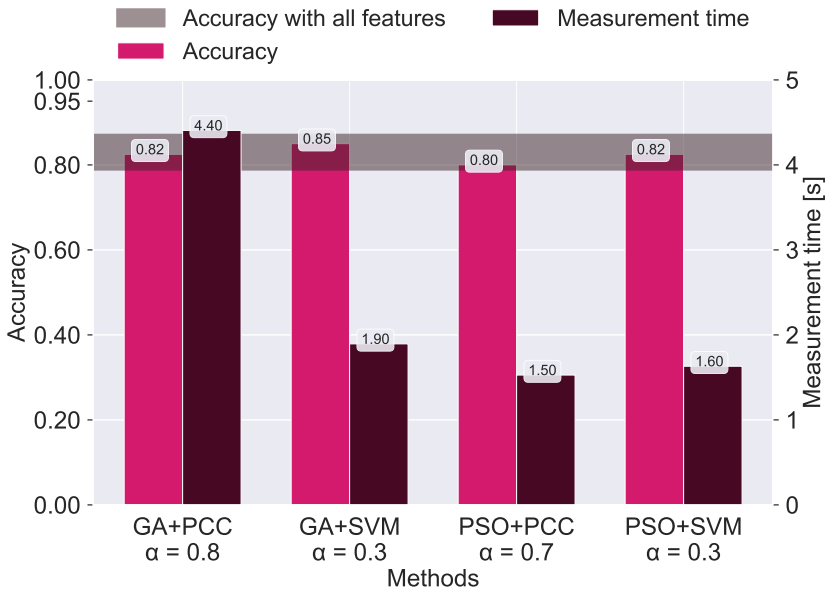


Figure 3.16: Accuracy and time performances of the solutions with two constraints on time and Accuracy [17].

no consideration given to the measurement time. The optimal solution in this scenario achieves a classification Accuracy of 0.98, accompanied by a measurement time of 16 seconds, which is shorter than the measurement time required for all stimulus frequencies.

On the other hand, scenario B introduces two constraints: one on the classification Accuracy and another on the measurement time. In this case, the best solution identified by the search algorithms achieves an Accuracy of 0.82, along with a notably reduced measurement time of 1.5 seconds.

Scenario A solely emphasizes achieving a SoC classification Accuracy of 0.95, without taking the measurement time into account. In this context, the most favorable solution achieves a classification Accuracy of 0.98, with a measurement time of 16 seconds.

In contrast, scenario B imposes dual constraints, targeting both classification Accuracy and measurement time. In this scenario, the search algorithms identify the best solution, characterized by an Accuracy of 0.82 and a significantly reduced measurement time of 1.5 seconds.

The substantial time savings of up to 99% underscore the efficiency gained by employing the selected features rather than utilizing all measured features. Furthermore, it is essential to recognize the impact of misleading features on classification algorithms. The elimination of these misleading features results in an improvement in SoC estimation Accuracy of over 10%, highlighting the significance of Feature Selection (FS) in enhancing SoC evaluation performance, even when time savings are not the primary focus.

3.2.1 Data Collection

Studying the relationship between battery States (SoC, SoH) but also temperature with the impedance requires sometimes expensive instrumentation which is a big obstacle for researchers. Just to give an order of costs, the minimum requirement in terms of instrumentation includes a bidirectional, an impedance meter, and a climate chamber. The first one is used to set the levels of charge in the batteries (SoC) while the second one is used to properly measure the battery impedance and since the impedance is a function of temperature the climate chamber helps to keep the temperature constant. The price of a bidirectional in Europe is between 2

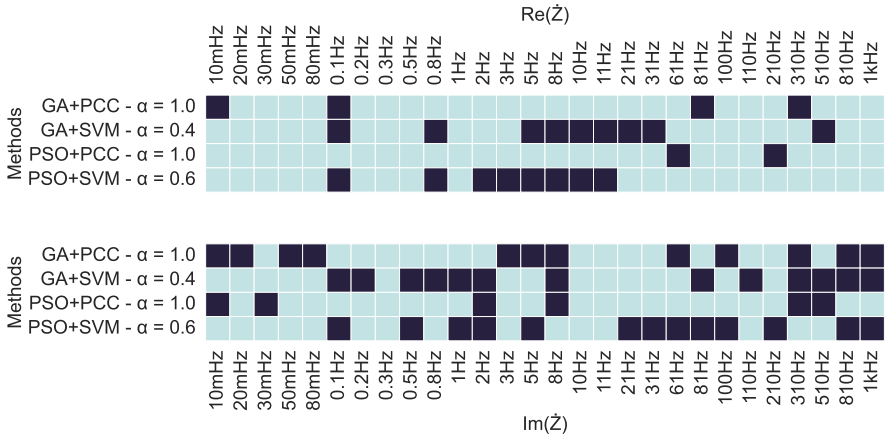


Figure 3.17: Distribution of selected features through the different combinations of searching algorithms and fitness functions for scenario A [17].

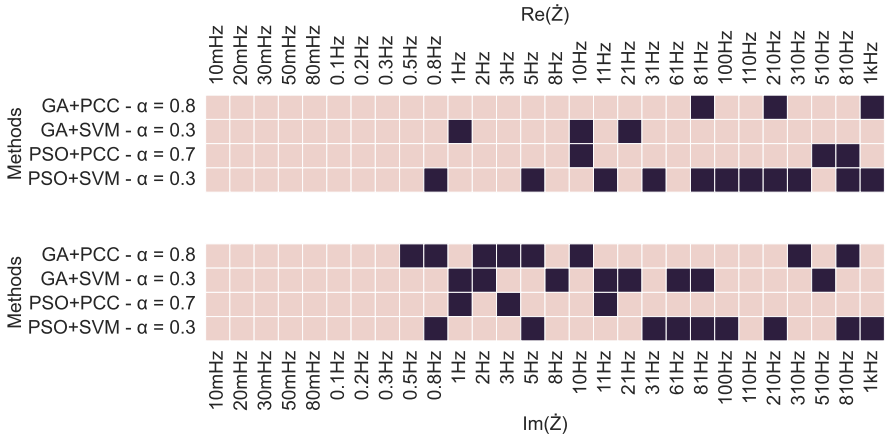


Figure 3.18: Distribution of selected features through the different combinations of searching algorithms and fitness functions for scenario B [17].

and 8 k while the impedance meter is around 10 k and the chamber from 15k. To overcome this issue, researchers with the instrumentations create measurement datasets and then share them with the community.

Background and Summary

Currently, batteries are regarded as a highly efficient means of energy storage when considering the energy-to-volume ratio and electrical power output. Among the various battery technologies available, Li-ion batteries stand out for their exceptional performance in terms of aging, cycle life, and rapid charging capability, as noted in previous studies [74, 144]. Consequently, Li-ion batteries are increasingly being utilized in a wide range of applications, including portable electronic devices, electric vehicles, and energy storage systems, as highlighted by Ding et al. [33]. Despite significant advancements, understanding the electrochemical phenomena taking place within batteries remains a formidable challenge. Electrochemical Impedance Spectroscopy (EIS) has emerged as a powerful tool for evaluating the performance and degradation processes of Li-ion batteries, as emphasized by Islam et al. [50]. EIS mea-

surements involve the application of an alternating signal across various frequencies, with the resulting measurements of currents and voltage enabling the assessment of impedances [50]. These impedances are closely related to the electrochemical processes occurring within the battery, encompassing charge transfer kinetics, ion diffusion, and interfacial reactions, as elucidated by Lazanas et al. [70].

Novelty of Data

However, addressing these relationships is not straightforward, and the literature contains numerous contributions proposing various approaches for the analysis of EIS data, as outlined by Jiang et al. [52] and others [48]. A significant challenge in this field is data acquisition, which necessitates high-end instruments capable of measuring small impedances and a complex measurement setup. Consequently, obtaining reliable data on battery aging is a challenging endeavor. This section seeks to address this challenge by presenting a novel EIS dataset for Li-ion batteries at different States of Charge (SoC) levels.

The data contained within this dataset is entirely original and has not been previously published in any publication or data repository. The dataset primarily comprises Electrochemical Impedance Spectroscopy (EIS) measurements conducted on LFP batteries during the discharge process at various States of Charge (SoC). The selected batteries for this dataset are the 600mAh LFP batteries manufactured by *O'Cell New Energy Technology CO. LTD*. Each battery undergoes three full discharge cycles. The primary contribution of this dataset to the research community lies in the substantial number of stimulus frequencies employed (58) and the wide range of SoC values (20). To the best of our knowledge, there is no existing dataset that combines these two characteristics. The inclusion of such a large number of stimulus frequencies facilitates the exploration of the frequency-domain properties of batteries and their variations throughout the discharge process. Furthermore, these data can be harnessed for developing prediction models and training machine learning algorithms to evaluate methods for the reliable and efficient management of rechargeable battery devices. Researchers in the fields of electrochemical studies, power generation technologies, and energy storage systems can derive significant benefits from this dataset. Moreover, industries involved in the manufacturing of consumer electronics, the advancement of power systems, and the progress of electric transportation can also find these data valuable for their applications.

Existing Li-ion Batteries Datasets

Here are some publicly available Li-ion battery Electrochemical Impedance Spectroscopy (EIS) measurement datasets, along with brief descriptions of each:

- Zhang et al. Dataset [140]:
  - *Dataset Description*: This dataset, provided by Zhang et al., is one of the largest EIS datasets publicly available. It comprises over 20,000 impedance spectra values obtained from 12 Eunicell LR2032 45 mAh LCO/graphite batteries. These batteries underwent cycling at different temperatures, and EIS measurements were conducted at various States of Charge (SoC) levels, including 45°C, 35°C, and 25°C.
  - *Purpose*: The primary purpose of this dataset is to facilitate State of Health (SoH) analysis of the batteries rather than State of Charge (SoC) analysis. As a result, it contains only three impedance spectroscopy measurements during discharge.
  - *Data Format*: The data is provided in ".txt" format and includes both the EIS values and independent capacity measurements.

- Buchicchio et al. Dataset [19]:
  - *Dataset Description*: This dataset, compiled by Emanuele et al., features EIS measurements from commonly used Li-ion batteries. The measurements were conducted using a random-phase multi-sine excitation signal, covering a range of fourteen distinct frequencies from 0.05 Hz to 1000 Hz for six batteries. The temperature was maintained at a constant 25 $°C$.
  - *Purpose*: This dataset includes 10 impedance measurements at various States of Charge (SoC), making it suitable for the development of data-driven methods for SoC evaluation. However, it should be noted that this dataset contains fewer stimulus frequencies compared to others.
  - *Data Format*: All the data is shared in a single ".csv" format for all the batteries.
- Toyota Dataset [110]:
  - *Dataset Description*: In this dataset, the authors focus on the charging phase of a 1100 mAh Lithium Iron Phosphate cylindrical battery. The charging phase is conducted using a four-step fast charging protocol, while the discharge is carried out at a fixed current. The dataset includes tests on 240 cells, divided into 5 batches.
  - *Purpose*: This dataset primarily emphasizes the charging phase of battery operation.
  - *Data Format*: Detailed data format information is not provided in the source.
- NASA Dataset [104]:
  - *Dataset Description*: This dataset consists of 34 2 Ah battery cells tested at three different temperatures. It provides EIS data with 39 frequencies ranging from 0.1 Hz to 5 kHz. However, specific frequency values are not given. The primary focus of this dataset is State of Health (SoH).
  - *Purpose*: The primary objective of this dataset is to facilitate SoH analysis.
  - *Data Format*: Detailed data format information is not provided in the source.
- Panasonic Dataset [97]:
  - *Dataset Description*: This dataset involves testing a 2900 mAh Panasonic 18650PF cell at five controlled environment temperatures, ranging from 25 $°C$ to -20 $°C$. Each impedance spectroscopy test covers a frequency range from 1 mHz to 6 kHz with a ΔSoC of 5%.
  - *Purpose*: The dataset provides extensive EIS measurements across various temperatures and SoC values.
  - *Data Format*: Detailed data format information is not provided in the source.

These datasets offer valuable resources for researchers and practitioners in the field of battery characterization, with each dataset catering to different aspects of battery performance analysis and providing unique insights into Li-ion battery behavior.

Data Acquisition

Figure 3.19 shows the bult data Acquisition Setup. The equipment utilized comprises a specialized battery impedance meter, namely the *Hioki BT4560*. This impedance meter is capable of measuring battery impedance using stimulus frequencies ranging from 10 mHz

to $1kHz$. Due to the inherently minute nature of battery impedances, a four-contact measurement configuration is employed by the meter to minimize contact and cable resistances.

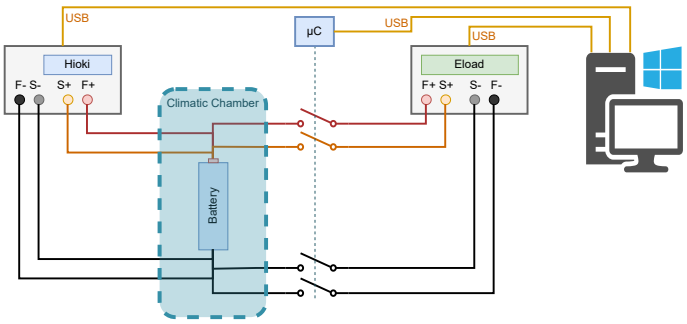


Figure 3.19: Implemented Setup scheme.

For battery discharging, the system relies on the Electronic *Load Zketech EBD-A20H*. This electronic load device enables the preselection of a discharge current, accommodating currents of up to 20 amps with a resolution of $10mA$. It maintains a maximum error rate of $\pm 0.5\%$ in current measurement accuracy. Similar to the impedance meter, the Electronic Load features four terminals: two for voltage measurement and an additional two for setting the discharge current. An intricate switch system is integrated into the setup to physically disconnect the electronic load during measurements. This complexity is necessary to ensure the impedance meter operates in accordance with its datasheet specifications. The management of the switch system is overseen by a dedicated microcontroller, with each switch exhibiting a contact resistance of roughly 100 $m\Omega$. Importantly, this contact resistance does not impact the discharge current since, as explained earlier, the electronic load's four terminals allow for compensation of the switches' resistance.

The entire system is under the control of a Python-based control script. This script requires input parameters, including the desired discharge current, the minimum battery voltage ($V_{cut-off}$), the number of impedance measurements, and the corresponding stimulus frequencies. The system generates output data, including the discharge current, voltage readings, and impedance measurements conducted by the impedance meter.

The experimental setup involves eleven batteries, each with a capacity of 600 mAh and employing LFP technology. Each battery is uniquely identified by an identification code. The key characteristics of these selected batteries are summarized in Table 3.5.

Measurement Protocol

Figure 3.20 illustrates the flowchart delineating the process of conducting a comprehensive test, which is divided into two primary segments: battery charging and battery discharging.

Battery charging adheres to the CC-CV (Constant Current-Constant Voltage) methodology. This approach encompasses two distinct phases:

- In the initial phase, a constant charging current (CC) is applied until the battery voltage attains the maximum voltage threshold (V_{charge}).
- Subsequently, the voltage is sustained at a constant level equal to the maximum voltage (CV), and charging continues.
- Once the charging current reaches the predetermined threshold value (I_{cutoff}), the battery is considered fully charged, signaling the readiness for the ensuing discharge test

Table 3.5: Adopted Batteries Characteristics.

	Battery Characteristics
Manufacturer	O'Cell New Energy Technology CO., LTD
Model	IRF14500EC-600 mAh
Capacity	600 mAh
Nominal Voltage	3.2 V
Charge Voltage	3.65 V
Charge Current	300 mA (0.5C)
Cut-off Current	30 mA (0.05C)
Cut-off Voltage	2.0 V

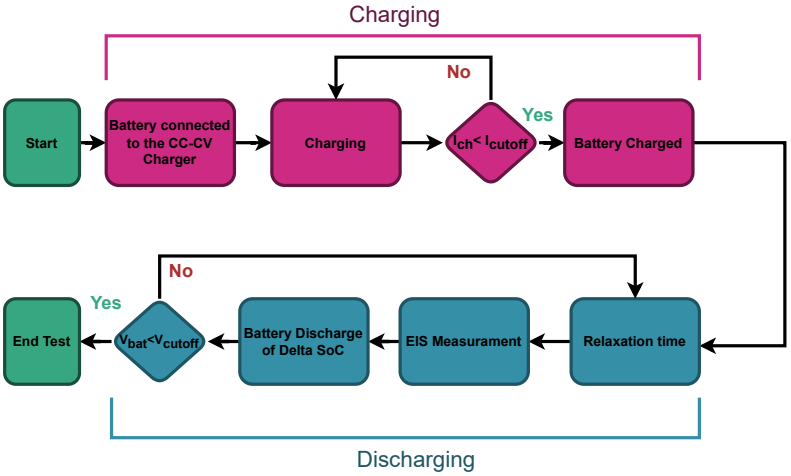


Figure 3.20: Adopted procedure to obtain the dataset.

and impedance measurement.

After absorbing or delivering current, batteries necessitate a designated duration for electrochemical recombination processes to occur, termed the relaxation time. Following this interval, the switches are opened, the electronic load is disconnected, and impedance measurements are executed at all the specified frequencies established during setup. The resulting data is then stored in a file on the control PC.

Upon completion of the measurement phase, the electronic load is reconnected to the circuit, and the battery undergoes discharge, expending the anticipated amount of energy. The battery voltage is continuously monitored throughout this phase, and if it descends below the minimum battery voltage threshold (V_{cutoff}), the test is terminated.

The values of V_{charge}, I_{cutoff}, and V_{cutoff} are provided by the manufacturer and are documented in Table 3.5. The State of Charge (SoC) is defined as the ratio of the remaining energy within the battery to its total energy capacity at a specific state of health. It's worth noting that determining SoC necessitates knowledge of the battery's total capacity. In some instances, nominal capacity is employed for simplicity. However, it's essential to acknowledge that the energy a battery can deliver depends on various factors, including discharge current, cell temperature, health state, and more.

Therefore, within the acquired dataset, a discharge test is conducted at the designated

current and temperature without concurrent measurements. This test subsequently serves to ascertain the total energy the battery can deliver by simply integrating the measured current over time.

Data Validation

The electrochemical impedance spectroscopy is valid under the small signal hypothesis, or, in other words, when the conditions of linearity, causality, stability, and finiteness are verified. In order to prove these conditions, the Kramer–Kronigs relations and their particular application named "Lin–KK" proposed in [14,108] are used. In order to apply this method, it is necessary to have a mathematical model of the battery's impedance behavior as the frequency varies.

The model estimation is performed by fitting the impedance measurements by varying the frequency with an equivalent circuit model composed of a series of RC elements. An important operation is to determine the number of RC elements necessary to accurately define the impedance behavior. To this end, relative errors committed by the fitting operations on the real part $Re(\dot{Z})$, imaginary part $Im(\dot{Z})$, and complex impedance $\dot{Z}$ are used as a figure of merit to fix the correct number of RC elements. In detail, Equation 3.15,Equation 3.16 and, Equation 3.17 shows the mathematical definitions of the above relative errors

$$\varepsilon_{Re(\dot{Z})} = \sqrt{\sum_{i=1}^{N} \left(\Delta_{Re}(f_i)\right)^2} = \sqrt{\sum_{i=1}^{N} \left(\frac{Re(\dot{Z}(f_i)) - Re(\dot{Z}_s(f_i))}{|\dot{Z}(f_i)|} \right)^2}, \tag{3.15}$$

$$\varepsilon_{Im(\dot{Z})} = \sqrt{\sum_{i=1}^{N} \left(\Delta_{Im}(f_i)\right)^2} = \sqrt{\sum_{i=1}^{N} \left(\frac{Im(\dot{Z}(f_i)) - Im(\dot{Z}_s(f_i))}{|\dot{Z}(f_i)|} \right)^2}, \tag{3.16}$$

$$\begin{aligned} \varepsilon_{\dot{Z}} &= \sqrt{\sum_{i=1}^{N} \left(\Delta_{Re}(f_i)\right)^2 + \sum_{i=1}^{N} \left(\Delta_{Im}(f_i)\right)^2} = \\ &= \sqrt{\sum_{i=1}^{N} \left(\frac{Re(\dot{Z}(\omega_i)) - Re(\dot{Z}_s(f_i))}{|\dot{Z}(f_i)|} \right)^2 + \sum_{i=1}^{N} \left(\frac{Im(\dot{Z}(f_i)) - Im(\dot{Z}_s(f_i))}{|\dot{Z}(f_i)|} \right)^2}, \end{aligned} \tag{3.17}$$

where N indicates the number of analyzed frequencies, f_i is the i-th considered frequency, $Re(\dot{Z}(f_i))$ and $Im(\dot{Z}(f_i))$ represents the measured real and imaginary part for the i-th specific frequency, $Re(\dot{Z}_s(f_i))$ and $Im(\dot{Z}_s(f_i))$ represent the estimated real and imaginary part for the i-th specific frequency and $|\dot{Z}|$ is the module of the impedance for the i-th specific frequency. Figure 3.21 shows the obtained relative errors by varying the number of considered RC elements for the model estimation operation.

As an example, the figure reports the analysis to only one impedance spectra, in particular, the spectra of battery B11 at 20% of SoC. In particular, we choose to adopt 10 RC elements since from this point on all the relative errors show a very flat behavior with respect to the number of RC elements.

Once obtained a mathematical model of the battery impedance as the frequency varies, it is possible to apply the "Lin–KK" method for the measurements validation. The Lin–KK method is a specific method based on the Kramer Kronig relations and proposed by [108] for the impedance spectra measurement validation. Figure 3.22 reports the obtained results for the Lin-KK method.

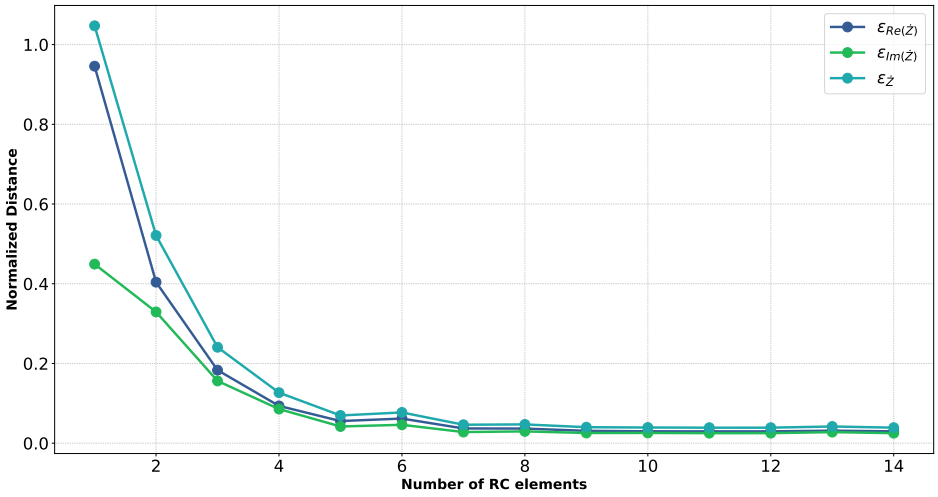


Figure 3.21: Committed relative errors among the measured and the estimated impedances

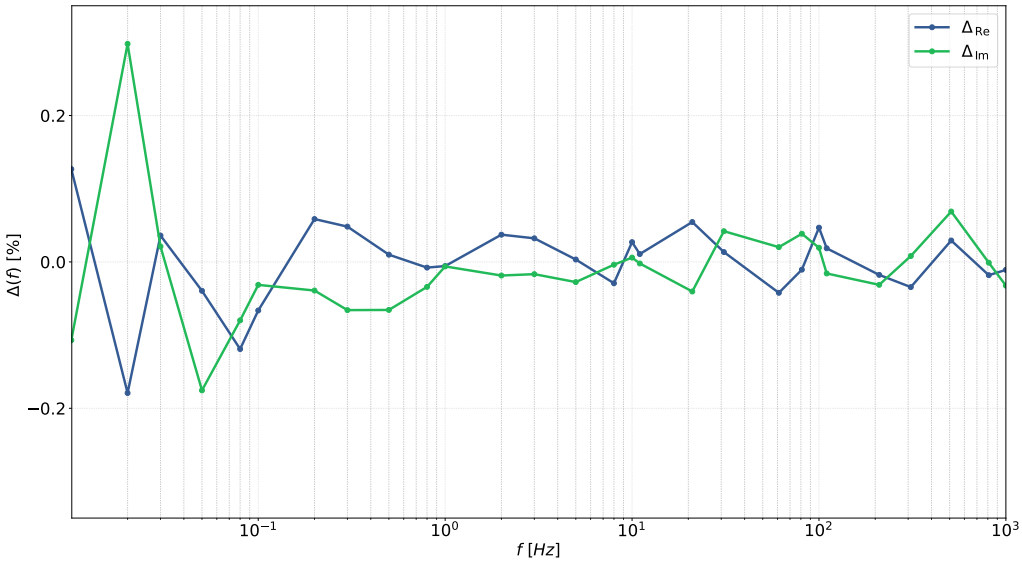


Figure 3.22: Obtained performance with the Lin-KK method validation.

It is possible to note that the estimation performances are always less than 0.3% for the estimation of both the real and imaginary parts. This confirms the reliability of the conducted analysis.

3.3 Innovative Methods: Resonance-based technique

The state of charge is an index that provides how much energy a battery has delivered relative to its presumed maximum capacity (measured in a precedent cycle). The state of charge is critical information for both the Battery Management System, i.e., that system that oversees the battery, and the user. In the case of electric vehicles, for example, the SoC provides macroscopic information about the remaining range but the microscopic information about the individual cells. In fact, SoC is very useful to the BMS for balancing operations. In battery modules, there are several cells connected in series and in parallel, a cell that has a lower SoC than the others in series limits the discharge capacity of all the branch because the cell must not fall below the minimum voltage for safety issues [31] More advanced BMSs use the energy from the other cells to recharge the cell with the lowest SoC and bring it up to the level of the others. The most widely used way to evaluate the SoC of batteries is a combination of battery voltage and the integral of the delivered current from a fully charged situation. The battery voltage depends on the state of charge but also on the current delivered at that time or the current delivered in some previous time frame, which may affect the quality of the remaining charge estimate. In addition, the integration operation on the current results in a natural divergence of the result over time; continuous recalibration is necessary to avoid this. To overcome the limitations mentioned above, several alternative methods have been proposed in the literature, exploiting different principles such as impedance spectroscopy, as extended shown in this thesis, or so-called data-driven methods. Some authors have also investigated magnetic methods for charge state estimation, in particular in [47] a coil wound around a ferromagnetic core placed on the side of a battery in lead-acid technology was used. The change in the state of charge affects the magnetic field for reasons related to the electrochemical phenomena occurring inside the battery. With an excitation frequency of 10 kHz a variation of 1 mH is obtained on a coil base value of 168 mH. A similar approach has been tried in [59] but instead of using a single coil, in this case, two coils are placed on either side of the battery, one of the two coils is used to generate the magnetic field at a frequency of 70 kHz while the voltage induced on the second battery turns out to be proportional to the SoC of the battery. In [128] the author places the single excitation coil inside the battery to minimize the distance to the negative electrode. In fact, in lead-acid batteries, during discharge, the negative electrode becomes coated with lead sulfate, which has a different permeability than metallic lead. In this section, we present a novel approach based on a LC tank attached to the battery side and measured near the resonant frequency with the aim of a *Texas Instrument LDC1101*. The used PCB coil has some clear advantages, it is very cheap to realize and also easy to attach to the battery shell without taking up space.

3.3.1 Background & Overview

Battery modules are made out of dozen or hundred of single cells. For the BMS its not possible identify the status of every single cell because they are connected in series or in parallel and its impossible esaminate one cell without the interference of the others. For this reason, evaluating the SoC without a physical connection to the battery terminals is interesting. The utilization of Inductance to Digital Converter (LDC) technology offers a potential solution for assessing the State of Charge (SoC) of a battery. LDC operates on the principle of the eddy currents effect, wherein an alternating current is supplied to a primary coil, thereby generating an alternating magnetic field adjacent to the coil. When the coil is in close proximity to a conductive object, the magnetic field induces eddy currents within the object, subsequently giving rise to a reactionary magnetic field that interacts with the original field. As a consequence of this phenomenon, the inductance of the primary coil undergoes a change. The

magnitude of the eddy currents and the alteration in inductance depends on several factors such as the object's conductivity, permittivity, and distance from the coil. The transducer is able to measure the impedance and the resonance frequency of a LC tank. On the left side of Figure 3.23, the equivalent circuit of a resonant LC tank is reported, the resistance Rs is a parasitic resistance, this circuit could be transformed into the circuit on the right side which is the model used by the LDC transducer. The smd capacitance C is placed next to the coil. In this

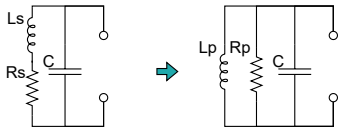


Figure 3.23: Caption

section, the target is a lead acid battery, specifically a Valve Regulated Lead Acid (VRLA). Like all electrochemical cells, lead acid batteries are characterized by three basic elements, the cathode, the anode, and the 'electrolyte. In the case of VRLAs, the anode is composed of lead metal while the cathode is composed of lead dioxide. The electrolyte is a solution of varying concentration of sulfuric acid. Unlike batteries with liquid electrolyte in VRLAs the electrolyte is diffused within an absorbent mat with the advantage of making the battery maintenance-free in that no topping up of electrolyte is required during life. The elements that makeup VRLA batteries have the characteristic of varying electrical and magnetic characteristics depending on the state of charge.

Electrodes

Rechargeable batteries use a reversible oxidation-reduction reaction to image energy during charging and supply it to the load during discharging. as stated in Equation 3.18, at the anode the metallic lead reacts with the sulfuric acid present in the electrolyte generating two electrons a hydrogen ion that remains in aqueous solution and a lead sulfate molecule that is instead a solid that is deposited on the electrode. The magnetic and electrical characteristics of lead sulfate are very different from those of metallic lead as can be seen from Table 3.6. In Equation 3.19 shows the half-reaction that occurs at the cathode, again during discharge, lead dioxide combines with sulfuric acid, hydrogen ions and two electrons to form a molecule of water and lead sulfate. The variation in magnetic susceptibility between lead dioxide and lead sulfate is less pronounced than the variation between metallic lead and lead sulfate. Equation 3.20 then reports the whole oxidation-reduction equation, The lead sulfate produced is a salt that creates a thin distributed layer on the electrodes.

$$Pb + HSO_4^- \rightarrow PbSo_4 + H^+ + 2e^- \tag{3.18}$$

$$PbO_2 + 3H^+HSO_4^- + 2e^- \rightarrow PbSO_4 + 2H_2O \tag{3.19}$$

$$Pb + PbO_2 + 2H_2SO_4 \rightarrow 2PbSO_4 + 2H_2O \tag{3.20}$$

Electrolyte

In lead-acid batteries the electrolyte is composed of a solution of sulfuric acid in water, the acid concentration is dependent on the construction technology but for lead-acid batteries it is about $30 - 40\%$. As can be deduced from the oxidation-reduction Equations 3.18 3.19, 3.20 sulfuric acid contributes to the reaction that takes place during discharge, reacting with the

Table 3.6

Material	Magnetic Susceptibilities	Electrical Conductivity
Lead	-23 [128]	5.30 E5 S/m [38]
Lead Oxide	-42 [128]	1.35 E4 S/m [38]
Lead Sulphate	-69.7 [128]	1 uS/m [38]
Electrolyte	-57E-6 [93]	0.8S/m [7]
Distilled Water	– []	– []

electrode material and generating lead sulfate and water as residual material. Therefore, the concentration of sulfuric acid is inversely proportional to the state of charge of the battery.

3.3.2 Setup

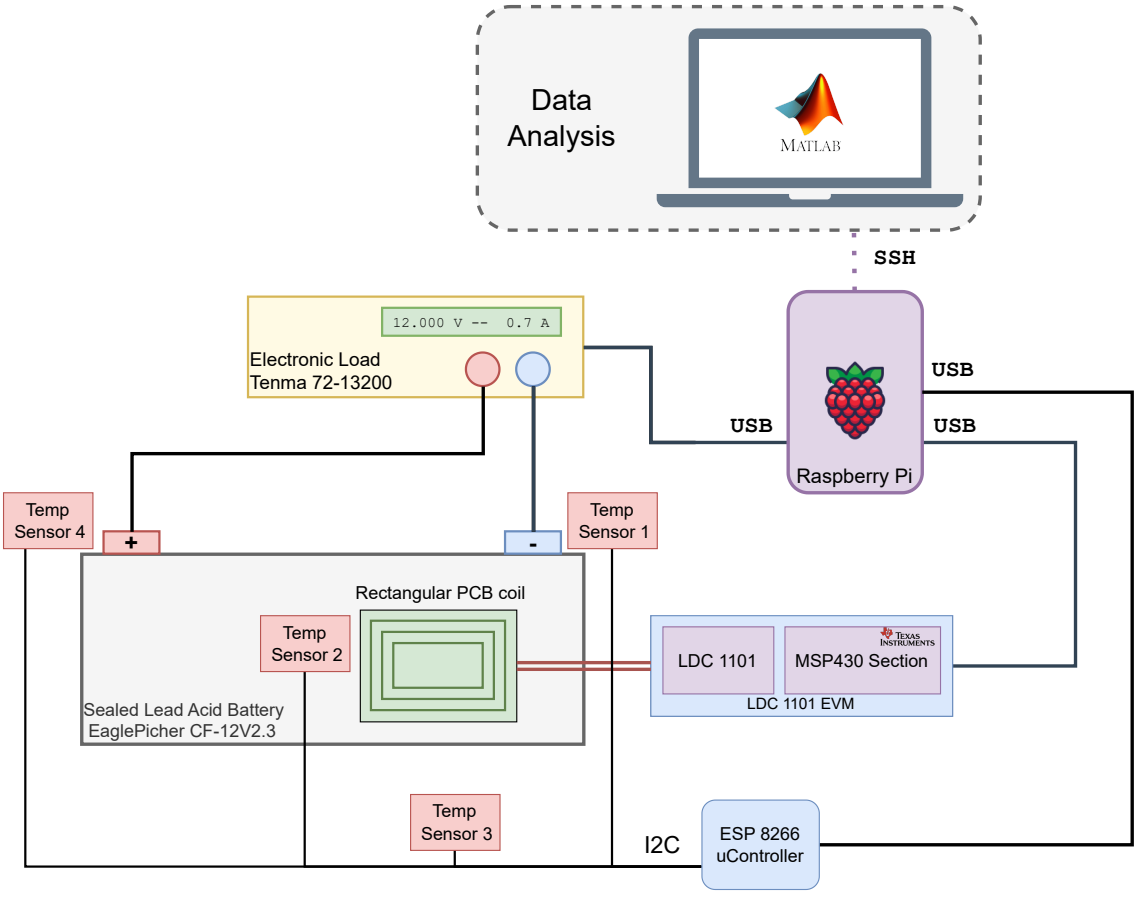


Figure 3.24: Measurament Setup

In this section, the experimental setup is presented. The battery adopted is EaglePicher CF-12V2.3 Sealed Lead Acid Battery whose characteristics are shown in Table 3.7.

The characteristics required of the setup are:

- The ability to bring the battery to known states of charge either starting from a discharged battery or a fully charged beater, in other words, the ability to discharge and charge the battery and measure the energy being delivered or drawn is required.

- Continuous monitoring of temperature, as it is one of the major interferers for measurements on batteries.
- Measurements must be simultaneous in order to ensure time synchronization of results.

All these aspects have been implemented in the setup in Figure 3.24

Table 3.7: Main characteristics of the considered batteries.

Battery features	Values
Battery Type	Valve Regulated Lead Acid
Nominal capacity	2.3 Ah @ 0.2C Discharge
Nominal voltage	12 V
Minimum voltage	10.67 V@ $Id < 1.78C$
Charging voltage	15 V
Maximum operating temperature	-20°C – 50°C (Charging)

The instruments adopted are:

- Tenma 72 13200 Electonic Load: the electronic load allows the battery to be discharged and brought to known states of charge by setting the chosen discharge current; it also outputs a measurement of battery voltage a and current with resolution and accuracy of 0.1 mA and $\pm 0.05\% FS$ for current and 0.1 mV and $\pm 0.05\% FS$ respectively
- Texas Instrument LDC 1101: forms the heart of the measurement system and enables high-speed measurement of the inductance of a coil; the output is digital and sampled through a 24-bit resolution converter. The parameters provided are the coil inductance and the equivalent parallel resonant resistance.
- AZ-DElivery dS18B20 8-bit Temperature sensors: Temperature is one of the major interfering null measurements on batteries for this reason it was chosen to monitor it with a system based on four different sensors located at different points of the battery. two near the terminals one near the coil and the last one measures the temperature in the air. the ESP8266 microcontroller is used as a bridge to convert the I2C protocol to USB.

The coil is a custom spiral PCB coil of rectangular shape and size 20 mm by 30 mm , the substrate is composed of FR-4 and has 48 turns arranged in two layers. The resulting Inductance is 31 μH. According to the manufacturer's datasheet, a 330pF capacitor is soldered in parallel with the coil to have a resonant frequency in the measurement range of the LDC1101. All of the above instruments are managed by a python scrip running on a *Raspberry Pi 3B+*.. Thanks to the *Threading library* all measurements are acquired simultaneously albeit with different sampling rates to account for typical variations in the quantities to be measured or maximum sampling rate. For example, the temperature is monitored with a sampling rate of 4.3 Sa/s, while inductance with 0.9 Sa/s and finally 80 Sa/s for battery current and voltage. Figure 3.25 shows the flowchart of the measurement program, with "Thread 1" being the main thread since it is the only one capable of triggering the "end of measurement" event that forces "Thread 2-3" to stop the acquisition. The trigger is determined by the end of the test or when the minimum battery voltage is reached.

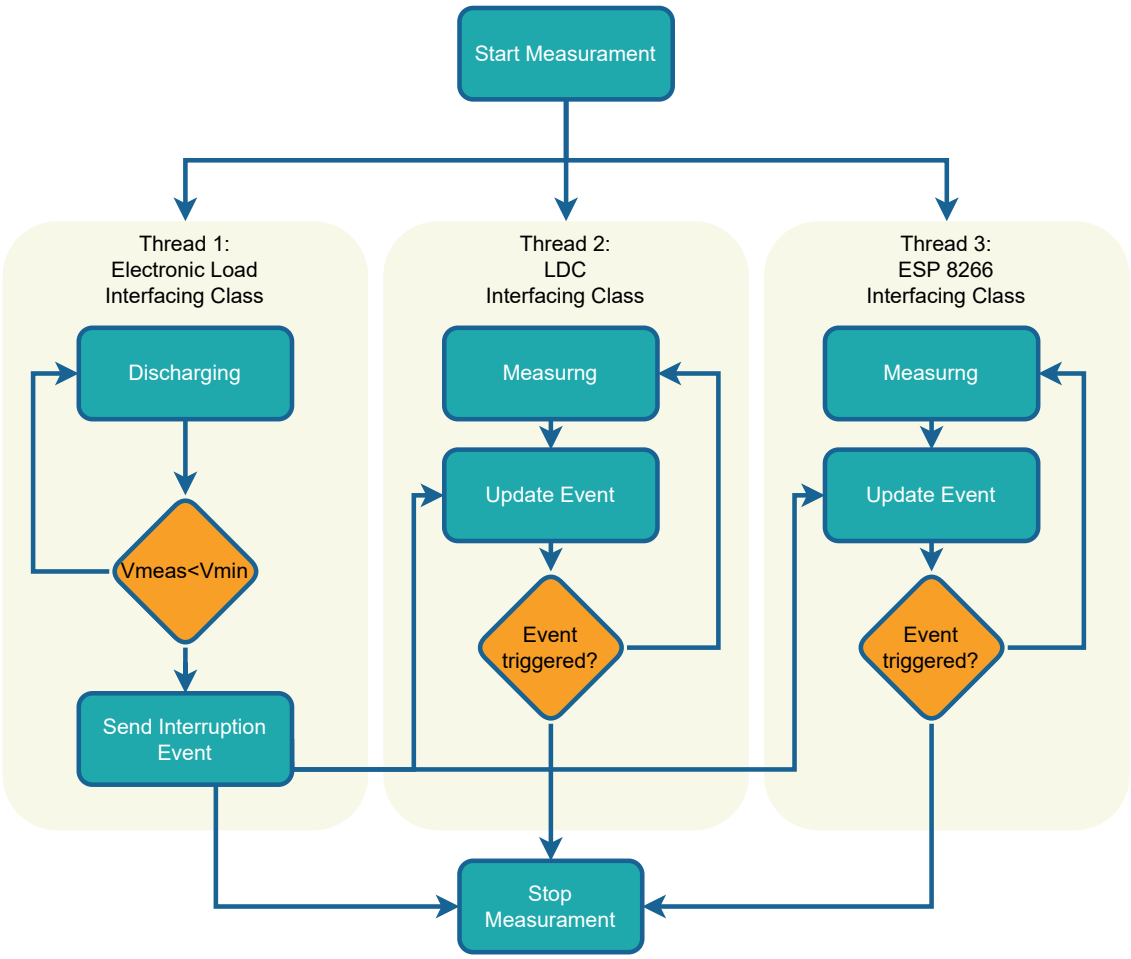


Figure 3.25: Measurement Protocol

3.3.3 Performed Test and Discussion

As verified in the overview section, the chemical and physical transformations of the battery elements (electrolyte and electrodes) during discharge result in a variation of the battery's magnetic characteristics. Since the effect is driven by two different battery elements, it is not possible to define analytically which sensor position maximises sensitivity, i.e., the position that has the greatest variability as the state of charge changes. the greater the sensitivity, the greater the ability to discriminate between two neighbouring SoC values. In addition to sensitivity, another characteristic that is commonly considered when analysing the performance of a transducer is linearity, which is the transducer's ability to have sensitivity over its entire operating range. Transducers with linear behavior are preferable. The discharge test is performed starting with the battery fully charged and with at least one hour of resting time before the discharge test. The battery is considered to be fully discharged when the terminal voltage reaches the minimum voltage given in the datasheet and equal to $10.87/V$. The discharge is a constant current discharge ($0.3C$ $0.71A$). In order to identify the best sensor position, several battery discharge tests were carried out by varying the sensor position. In Figure 3.26 all six positions tested are shown. The lead-acid battery used consists of six cells arranged in series. After a visual inspection carried out on a battery of the same type, it is reported that the electrodes are arranged perpendicular to positions A and B, and U and parallel to positions P and N. Hence, similar sensitivities are to be expected between positions A and B. Position U, for

construction reasons, is further away from the electrodes. Positions P and N are at the positive and negative battery terminals. Finally, in position AR the sensor is concentric to position A but rotated by 90 degrees.

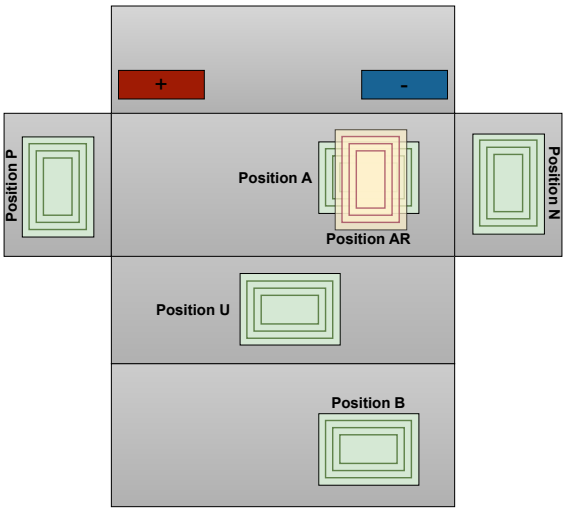


Figure 3.26: Tested positions on battery

The LC tank and the measuring device, in particular, provide two quantities as output, the parallel inductance "L_p" and the parallel resistance "R_p" of the circuit in figure Figure 3.23, the values are provided directly in ADC code, i.e. the numerical value output from the analog to digital converter. Figure Figure 3.27 shows the parallel resistance trend as the battery state of charge changes; the tests were carried out during a discharge. To make the graphs easier to understand, the resistance values have been normalized using the formula given in Equation 3.21. Furthermore, the values are the mean of several tests in order to give more robust results.

$$Rp^{Norm} = Rp(i) - Rp(1); \quad 1 \leq i \leq N \tag{3.21}$$

Position A appears to have a sensitivity compatible with the other positions; For all positions, the behavior appears to be exponential.

A similar behavior occurs for the inductance, the main difference is that the sensitivity is lower for all positions, in fact it begins to be comparable with the resolution of the ADC converter. For "R_p" the variation of the SoC between 100% and 0% is within about 1000 ADC levels while for "L_p" the number of levels used by the converter is about 100 ("L_p" is normalized through the Equation 3.22). The ADC converter used by the instrument to convert the two values is a 16-bit SAR-type converter. In the case of "L_p", the position relative to the positive electrode also has a consistent pattern to the other positions.

$$Lp^{Norm} = Lp(i) - Lp(1); \quad 1 \leq i \leq N \tag{3.22}$$

Figure Figure 3.29 reports a 4-hour measurement without battery discharge, the objective being to rule out that the measured variations are due to a natural drift of the measurement systems rather than to an actual variation in the measurand, which cannot be ruled out a priori since the discharge tests, for the chosen current, last more than 3 hours. The change in "R_p" in more than 4 hours was about 60 ADC levels while the change in "L_p", which is less like a signal drift and more like noise, is at most 7 ADC levels.

another possible influencing factor is the discharge current, this in fact could make the measurement system more or less sensitive to the change in charge state as well as vary the

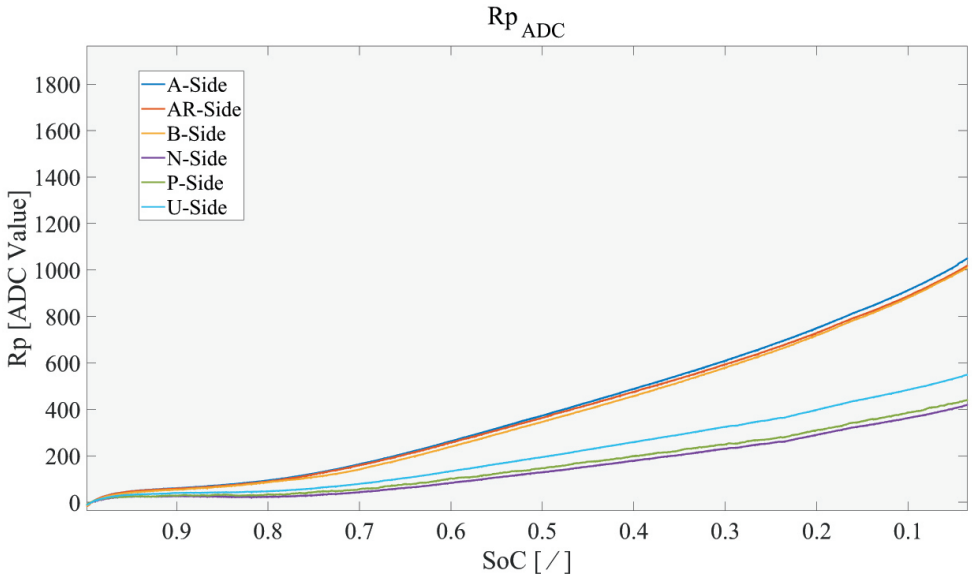


Figure 3.27: Mean Parallel Resistance Measurement with the sensor at all selected positions

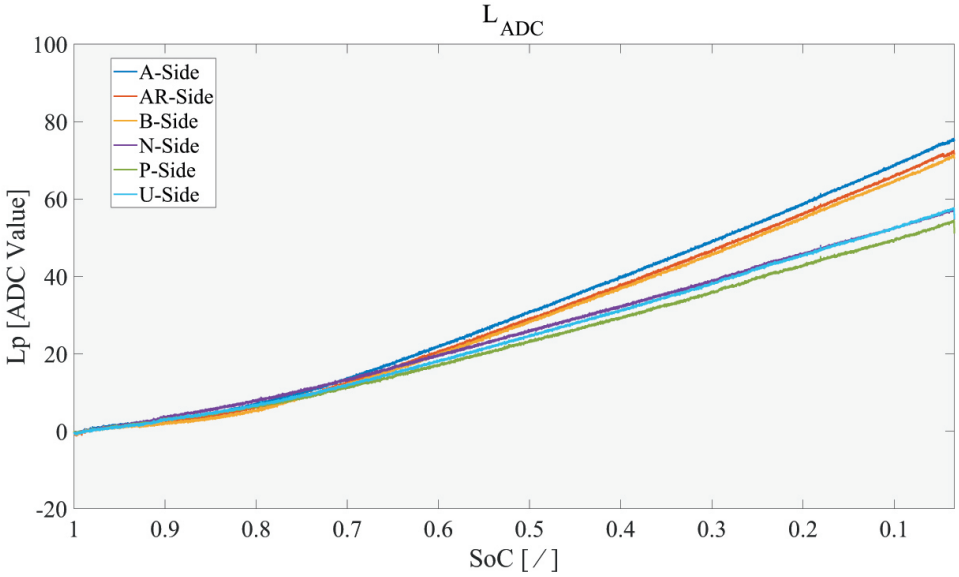


Figure 3.28: Mean Inductance Measurement with the sensor at all selected positions

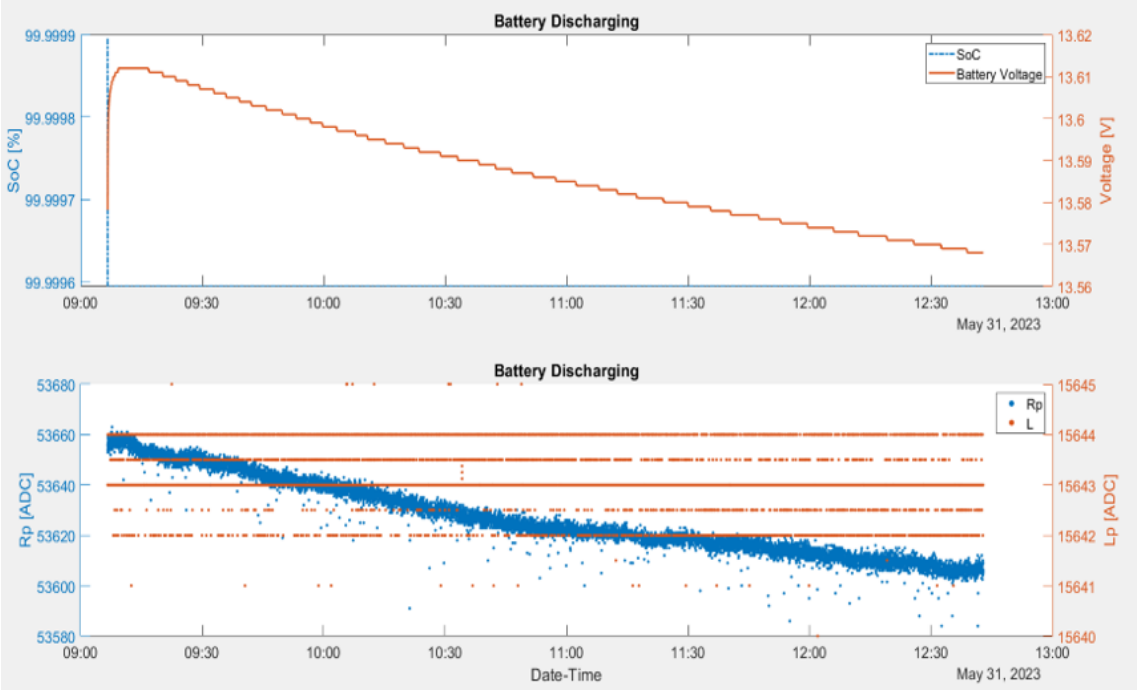


Figure 3.29: Drift analysis - The sensor value is acquired for a long time without changing the battery SoC

transfer function. In Figure 3.30 and Figure 3.31 are shown the trends of "R_p" and "L_p" respectively for four different values of discharge current. It can be seen that as the current increases, the transfer function becomes more linear. We also note a well-known behavior of batteries and particularly evident in lead-acid batteries, the amount of energy the battery is able to deliver is inversely proportional to the discharge current, in other words, the higher the discharge current, the less energy the battery is able to deliver.

Figure 3.30 and Figure 3.31 also shows the repeatability of the sensors through different tests

Measure the inductance of a PCB printed coils at its resonant frequency is successful for non-intrusively monitoring the state-of-charge of sealed lead acid batteries. The main problem is due to an offset of the measurement that forces to apply a normalization procedure to the measurement. Temperature affects the measurements and further studies should be provided to understand the relation with the temperature, in this preliminary test was not possible to fix the temperature due to a lack of instrumentation but it was at least measured. Further work is required in order to confirm the present results and their interpretation, plus to determine the practicality of the technique in the field.

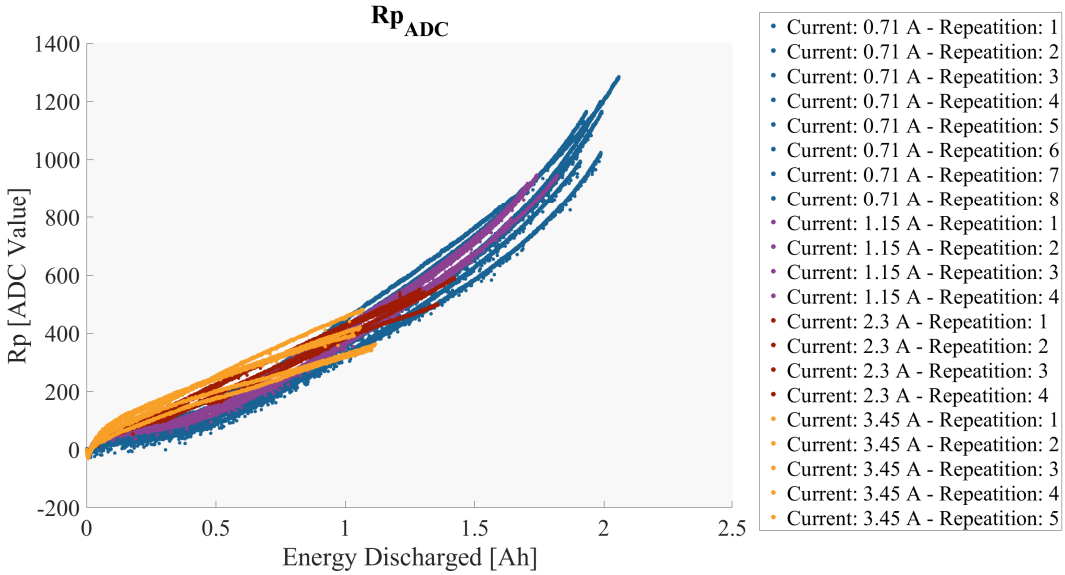


Figure 3.30: Measured parallel resistance for different discharge currents

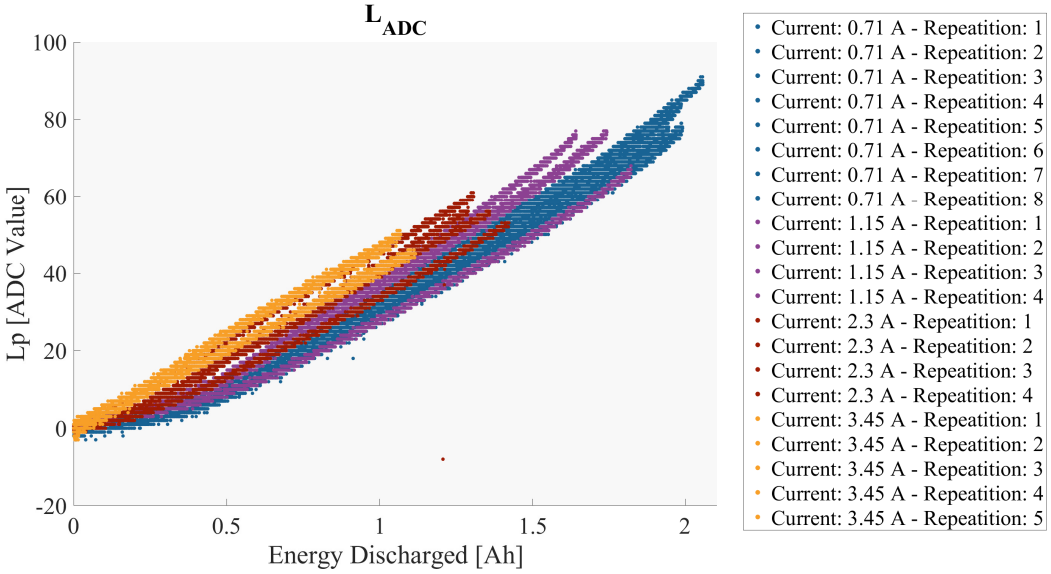


Figure 3.31: Measured parallel inductance for different discharge currents

Chapter 4

Data Driven based Modelling Methods for Batteries Prognosis

Predicting a system behavior is impossible without modelling it. Modelling means creating a mathematical object whose behaviour is similar to that of the system in question. Modelling makes it possible to test different life conditions of the modelled devices and to try to predict their behaviour in the future. In particular, when dealing with batteries, modelling plays a central role in trying to predict a future state, e.g. SoC or SoH, of the battery. For example, SoC or SoH as a result of certain usage conditions. There are many different types of models and modelling techniques, depending on the complexity of the resulting system, the technique used to obtain it and the properties required of the model. Three different modelling techniques are discussed in this chapter: the first is based on statistical analysis, identifying devices as state machines in which the transition from one state to another is linked to a statistical probability defined by the model. In this case, the method is applied to generic electrical loads, but a future application to batteries is being studied. The second method uses a battery-pulsed discharge to fit the values of an equivalent electrical circuit. The third method, on the other hand, exploits the ability of a genetic algorithm to obtain a mathematical model capable of predicting the battery voltage during discharge, given certain boundary conditions.

4.1 Makov chain Modelling for electrical appliances

4.1.1 Introduction

Applications related to the smart energy paradigm are demanding novel devices and techniques for satisfying the increasing issues of more efficient and sustainable use of energy. Nowadays, these applications are attracting more and more interest, especially due to energy and ecological transitions, whose goals are the reduction of CO_2 emissions and the increase of energy efficiency. As the latter increases, international treaties set limits on emissions and global temperature rise that must be met, such as those defined at COP26 [29]. As an example, in Europe, the development and use of intelligent measurement systems and techniques has been boosted by European directives in recent years. In 2006, the Directive 2006/32/EC [21] identified the use of intelligent metering systems as one of the tools to improve energy efficiency through suitable algorithms and techniques. In 2009, the Electricity Directive 2009/72/EC [22] declared the obligation for the Member States to ensure the adoption of smart metering systems. Finally, in 2012 the European Energy Efficiency Directive 2012/27/EU [23] reaffirmed the importance of using smart meters, with strict requirements for the Member States regarding both metering and billing. In this new paradigm, in which energy efficiency and the de-

velopment of new services based on the measurement of energy parameters are demanding data, the smart meter assumes a key role. On one hand, it accomplishes the primary task of monitoring energy consumption. On the other hand, it could provide new quantities useful for modern smart energy applications. Among these quantities we can mention those related to the electrical signature of a device, the power quality, and to the use of energy trends over time, to cite a few, which could enable the implementation of new high-value services for energy and plants management. In other words, the smart meter can be the key-enabling technology to implement modern algorithms and emerging techniques in the field of smart energy applications such as loads defragmentation through the non-intrusive load monitoring (i.e NILM), digital twin of the equipment, clustering of devices and predictive detection and diagnosis of faults on both plants and grids. More in-depth, the above-mentioned techniques are generally based on machine learning algorithms or on optimization techniques exploiting electrical parameters related to the electrical signature of a load and/or Power Quality parameters. For example, considering the development of NILM solutions that is involving many researchers in the scientific community, some algorithms are based on the use of Markovian models (HMM) and their variants [4] [63] [131], while others prefer signal processing techniques using graphs (Graph Signal Processing) [117] [141] or Combinatorial Optimization [13]. In recent years, other Machine Learning techniques have also been applied for non-intrusive monitoring, such as Multilayer Perceptron (MLP) [24], Convolutional Neural Networks (CNN) [132], Deep Learning [60] [115], Recurrent Neural Network (RNN) [85], Extreme Learning Machine [106] and Bayes Classifier [39]. Whatever the followed approach, all these techniques require two fundamental steps to be implemented to reliable train and tune the algorithms on the considered case study: *a)* the accurate measurement of the energy consumption and parameters related to the power signatures and other parameters relying on the involved devices, *b)* the monitoring for a time interval wide enough for capturing an adequate amount of data. As for the first step *a)*, it is crucial since it strictly affects the quality and the quantity of the collected information. It is very important to have energy measurements characterized by low uncertainties and, at the same time, to provide several energy parameters, such as the harmonic and inter-harmonic power, current and voltage values, the total harmonic distortion, the power factor, etc. that can characterize how a device, or a group of devices, is consuming the electrical energy. As for *b)*, the monitoring for wide time intervals is very important not only to provide big data to the smart energy algorithms but also to let the algorithms work on a few operating conditions able to give reliable information on the monitored process or equipment. For example, to optimize the energy consumption of a family it is important to monitor the consumption habits considering several days and the effects of seasonality. As a further example, such as diagnosis or predictive maintenance of devices and grids, to predict faults on equipment, it is important to have a reliable footprint of its energy status considering all the possible operating conditions of the involved device. Since the above-mentioned requirements impose to perform very expensive and time-consuming measurement campaigns, in the last years some datasets have been proposed in the literature with the aim of facilitating the development of new algorithms addressing the cited emerging needs. Many are the contributes to the literature that propose the use of a dataset to train or tune smart energy algorithms. Examples for NILM can be found in [102], load profiling can be found in [130] and predictive maintenance can be found in [136]. In literature two different types of datasets can be found: real [11,12,35,55,62,81,82,92,95,113,129] and simulated datasets [20,25,45,61]. In particular, real datasets are composed of data, collected on the field, while simulated ones are obtained through suitable software simulators. However, these datasets do not meet all the requirements for effective data usage, i.e. the limited amount of saved electrical parameters, data variability, seasonality, closeness to real scenarios, and defini-

tion of the current operating states are just some of the missing or not complete information that makes them poorly suitable for modern Smart Energy Applications. To overcome such limitations, in this study we present an innovative energy dataset. The main advantages of the proposal are: *(i)* data variability, in terms of the operating states of the electrical loads and the adoption of appropriate consumption models; *(ii)* the total number of available electrical parameters (433 in our case) enormously larger than the above-cited datasets, as the full analysis of frequency behaviors and harmonic computations; *(iii)* the seasonality of the data. Furthermore, the presented dataset is a hybrid solution between the simulated datasets and the ones based on real data. Indeed, it is generated through simulation, starting by real measurements, and customized by means of an effective computed variability model. Finally, it has been validated with further *a-posteriori* measurement campaign on real loads. The developed dataset, namely eLAMI, is made publicly available to the whole research community at [34] *(download here)* through a modular structure, allowing customized downloads and analyses.

4.1.2 Related Works

The growing demand of new monitoring devices with smart algorithm and technique has led to the development of several dataset capable to tune the performance. In Tab. 4.1, some examples of public datasets with data collected from the field are shown; they differ for the number of devices, number and type of provided electrical parameters and length. In the following they are briefly described.

Table 4.1: **Main Characteristics of Some Public Dataset Based on Real Data**

Name	Number Appliances	Parameters	Period
AMPds [82]	20	P,Q,S,V,I	1 year
AMPds2 [81]	20	P,Q,S,V,I	2 years
BLUED [35]	50	P,Q,V,I	1 week
Dataport [95]	8598	P,S	4 years
DRED [129]	12	P	6 months
ECO [12]	45	P,V,I	8 months
ENERTALK [113]	75	P,Q	1714 days
iAWE [11]	63	P,Q,S,V,I	73 days
REDD [62]	92	P,V,I	119 days
REFIT [92]	20	P	2 years
UK-DALE [55]	109	P,Q,S,V,I	2247 days

- **AMPds** [82] The Almanac of Minutely Power dataset is a public dataset, published in 2013, containing 1 year of collected data of residential appliances from a single household in Canada. This first version contains measurements of electricity, water and natural gas at one-minute intervals, for a total of 525600 readings per year per meter.
- **AMPds2** [81] This second version of the AMPds dataset differs from the previous one only in the number of total readings, 1051200, corresponding to 2 years of acquisition.
- **BLUED** [35] The Building-level fully labeled dataset for electricity disaggregation was released in 2012 by 1 week electricity data from 1 building in the USA. This dataset contains not only the steady-state, but also the state transition of each appliance.

- **Dataport** [95] The Dataport database was created by Pecan Street Inc and published in 2015. It contains electricity data from 722 houses and commercial buildings across different cities in the USA. As it has a sampling period of 1 min for aggregate and appliance signal, this is considered a low frequency dataset.
- **DRED** [129] The Dutch Residential Energy Dataset was released in 2015 and contains energy consumption data from a household in Netherlands, with a total duration over six months. It includes electricity measurements for the aggregate and submetered signal of each device.
- **ECO** [12] Electricity Consumption and Occupancy dataset was collected in 6 Swiss households over a period of 8 months. It contains data recordings of active power, voltage and current at low frequency sampling rate.
- **ENERTALK** [113] The ENERTALK dataset was created in Korea from 22 houses with a total period of 1714 days and it was published in 2019. It provides active and reactive power measurements (both aggregate and each device), with a sampling frequency of 15 Hz.
- **iAWE** [11] Indian Dataset for Ambient Water and Energy was released in 2013, from recordings electricity, water and ambient data in a house in New Delhi, for a total duration of 73 days. The electrical data were recorded with a sampling period from 1 to 6 seconds over 63 electrical appliances.
- **REDD** [62] Reference Energy Disaggregation Dataset has been published in 2011. It contains 119 days of collected data from 6 households in the USA and includes both high and low frequency recordings.
- **REFIT** [92] The REFIT Electrical Load Measurements dataset includes cleaned electrical consumption data from 20 households in the UK from 2 years of recordings in 2016. It contains electrical data with a sampling period of 8 s and active power as only parameter.
- **UK-DALE** [55] UK-Domestic Appliance Level Electricity was published in 2015 and it contains 2247 days of data by 5 residential buildings in the UK. Just like REDD, it reports high and low-frequency data and all appliances are sub-metered.

Collecting data to build a dataset is a fairly complex process. As long measurement campaigns have to be carried out, the process requires a considerable amount of time, effort, and instrumentation to measure and record data. One of the main limitations of the real datasets in the literature today is the small number of reported electrical parameters (at most P, Q, S, V, I).

This could represent a limitation for algorithms in the field of Smart Energy. In particular, considering methods for load profiling, NILM and fault or predictive diagnosis, the optimal choice of needed parameters is still a research topic: therefore the availability of a large number of electrical quantities, able to define the complete "electrical signature of the load", could help in performing an accurate selection.

Furthermore, real datasets are often characterized by missing data due to several problems that may occur during the measurement campaigns, time mismatching between individual load data and aggregate data or possible errors due to malfunctioning and inaccuracy of the adopted instrumentation.

The absence of information about current operating states of each device can also be a limitation for these datasets, e.g. in the case of supervised artificial intelligence algorithm training or in the definition of consumption quality as well as system efficiency indices. A solution to the aforementioned problems could be the simulation of data.

A simulated dataset does not require lengthy monitoring campaigns, saving time, costs, and instrumentation. All these advantages are provided as long as the simulation process is correctly implemented, which is not trivial in terms of suitable modeling and computational costs.

In literature, some simulated datasets are presented as reported in Tab. 4.2 and discussed below.

Table 4.2: **Main Characteristics of Some Public Dataset Based on Simulated Data**

Name	Number Appliances	Parameters	Period
AMBAL [20]	14	-	1 days
SHED [45]	66	P, S, V, I	14 days
SmartSim [25]	25	P	7 days
SynD [61]	21	P, S, V, I	180 days

- **AMBAL** [20] Automated model builder for appliance load dataset was published in 2017 and it comprises 14 domestic loads at a sampling rate of 1 Hz for a time duration of one day. The AMBAL dataset allows the user to build models using real energy consumption data, based on parameterised signature sequences. The main operational phase of the AMBAL dataset includes preprocessing, extraction of active segments, segmentation, and model fitting.
- **SHED** [45] Simulated high-frequency energy disaggregation dataset, released in 2018. It is a commercial dataset containing the power consumption of 66 buildings at a sampling frequency of 1/30 Hz. The data is generated synthetically and based on modeling the current flowing through an electrical device and is matched with the real model of electrical devices.
- **SynD** [61] SynD is a synthetic dataset that was published in 2020 and simulates readings of electricity consumption for a house for 180 days. The measurements campaign was based on the monitoring of 21 different residential devices from 2 households in Austria. In particular, the consumption patterns were observed, the absorption profiles of each device were then extracted and finally the dataset was generated.
- **SmartSim** [25] A Device Accurate Smart Home Simulator for Energy Analytics was released in 2016 and is a simulated dataset of 1 week total duration. It uses the energy modelling of individual devices to build the final dataset with the aim of generating accurate domestic energy traces that are qualitatively and quantitatively similar to real energy data traces.

Despite the large amount of simulated datasets in the literature, most of them still suffer some of the aforementioned problems, as limited number of saved electrical parameters, absence of harmonic and power quality information of seasonality, monthly, daily variability, as well as variability of operating states; furthermore, they often provide a low likelihood value with respect to real scenario profiles.

4.1.3 Design of Dataset generation

In this section, the design process of the eLAMI (electrical Loads Acquisition for Monitoring Instruments) dataset and its implementation is described. At first, main requirements that a modern energy dataset must have are discussed. Issues related to the the choice of a simulator

and possible solution are then faced. Chosen electrical loads, along with consumption patterns and the simulator description close the section.

Requirements

A modern energy dataset for monitoring and investigating consumers' energy behaviors must have specific characteristics. The main requirements are:

a) High number of saved electrical parameters, both in time and frequency domain;

b) High likelihood to real scenario profiles;

c) Faithful representation of the devices;

d) Considering the metrological performance of commonly used energy smart meters;

e) Presence of Power Quality and harmonic data;

f) Appropriate observation times congruent with the objectives;

g) Current operating state of the monitored device.

As regards *a)* having a large number of electrical parameters allows for a better representation of the electrical signature of the load. In this way, for example, as also proved in Section IV, energy efficiency algorithms can drastically increase their performance. Therefore, in the eLAMI dataset, 433 electrical parameters are calculated at each measurement interval. Furthermore, in our simulated dataset we try to represent an electrical scenario as faithful as possible to reality. Specifically, reported consumptions are related to a 2–person–household.

Having a good match between the simulated scenario and the real one, simply summarized as likelihood, is crucial as it offers the possibility, for instance, of training AI algorithms on consistent data, as mentioned in *b)*. This avoids possible mismatches between performance obtained in simulation and real scenarios.

As said in *c)* it is necessary to take into account both how the electrical energy is consumed, but also how the individual load behaves in terms of electrical operation in reality. In particular, in our case we have chosen to represent equipment as 'state machines', simulating the corresponding nominal operating states. We consider this choice to be valid, having chosen to simulate the assumed scenario under permanent regime. Of course, in reality, between different 'operating states' there are transients that can lead to more or less marked variations in electrical quantities. As regards *d)*, furthermore, these variations can also be related to the natural duty cycle of the equipment or be due to the uncertainty of the monitoring system used [13]. eLAMI reports this variability thanks to the mathematical model for generating absorption profiles implemented.

As said in *e)* today, given the widespread use of electronic equipment, we consider the study of the harmonic behavior of electrical loads to be of great interest. For example, the analysis of the frequency spectra of voltage and current absorption of devices can provide useful information on the health of loads and in general of the entire system, also in terms of the quality of the power supply system (Power Quality). eLAMI in this case reports a considerable amount of harmonic and quality parameters.

With regard to the observation time of the monitored system *f)*, it must be chosen in such a way as to meet the objectives of the applications for which it was hypothesized. Classification, clustering, NILM, Load Profiling, and energy retrofit algorithms in most cases aim at analyzing the system over sufficiently long time horizons. In our case, eLAMI refers to a time

horizon of one year. As regards, *g)*, an operating state of the monitored device can be defined as a steady-state voltage and current joint profile, whose availability allows the dataset user to evaluate the proper load working cycle and extract electrical signature quality indices and detecting possible incoming anomalies.

Indeed, referring to "electrical signature", it not only deals with the typical quantities (P, Q, I_{RMS}, etc.) but with the entire frequency spectrum of voltage and current profiles absorbed by the electrical load. Such information are therefore also state-related, i.e. they can change for each operating state of the load.

Therefore, the proposed simulator must be designed to generate voltage and current profiles under all possible load states tested and for a predefined simulated time, in order to get the spectrum information.

An issue raised at this stage is related to the way the frequency spectrum could be faithfully simulated. The approach here followed is to make measurements on real loads under different tested operating states and adopt acquired information as a basis to generate simulated profiles.

Based on the knowledge of these quantities and the known usage habits of the electrical loads, it is possible to create the dataset by using the acquired information.

Any accidental failures or malfunctions that might occur in the real electrical system are neglected in the current status of the simulator. This is reasonable as these are very rare events in reality that, when compared with the simulation time frame, can be neglected. The simulator modular structure would eventually permit to add such situations in a fairly easy fashion.

Electrical Loads Description

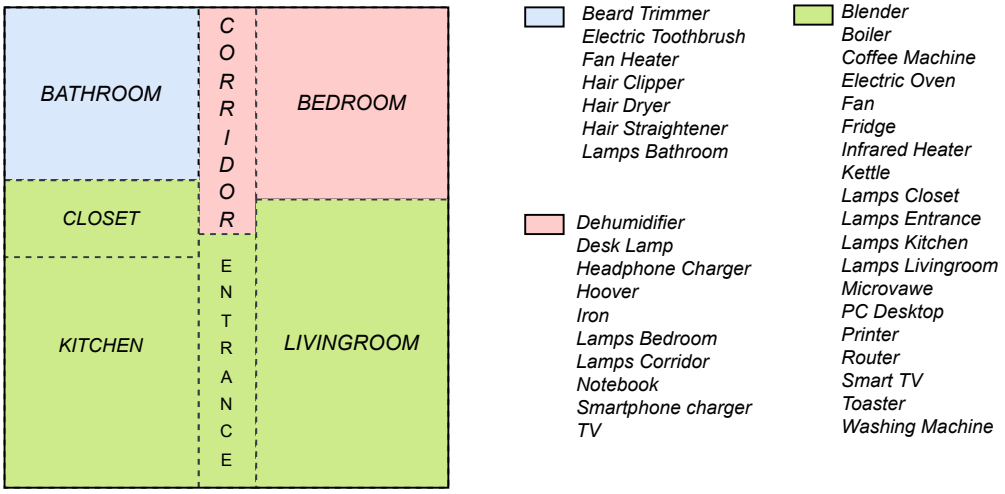


Figure 4.1: **House Scheme and Electrical Loads Distribution. The blue zone represents the *"Partial 1"* with 7 electric loads; the pink zone represents the *"Partial 2"* with 10 electric loads; the green zone represents the *"Partial 3"* with 19 electric loads** [80]

According to the aforementioned requirements of a new simulated energy consumption dataset for the residential appliances with innovative saved electrical parameters, namely eLAMI, has been developed. The choice of simulating a residential building consumption profile has the aim to provide a means with innovative characteristics compared to the datasets currently present in the literature for the evaluation of new techniques and algorithms in the field of Smart Energy, including NILM, Load profiling, Management Systems, and Energy Efficiency algorithms.

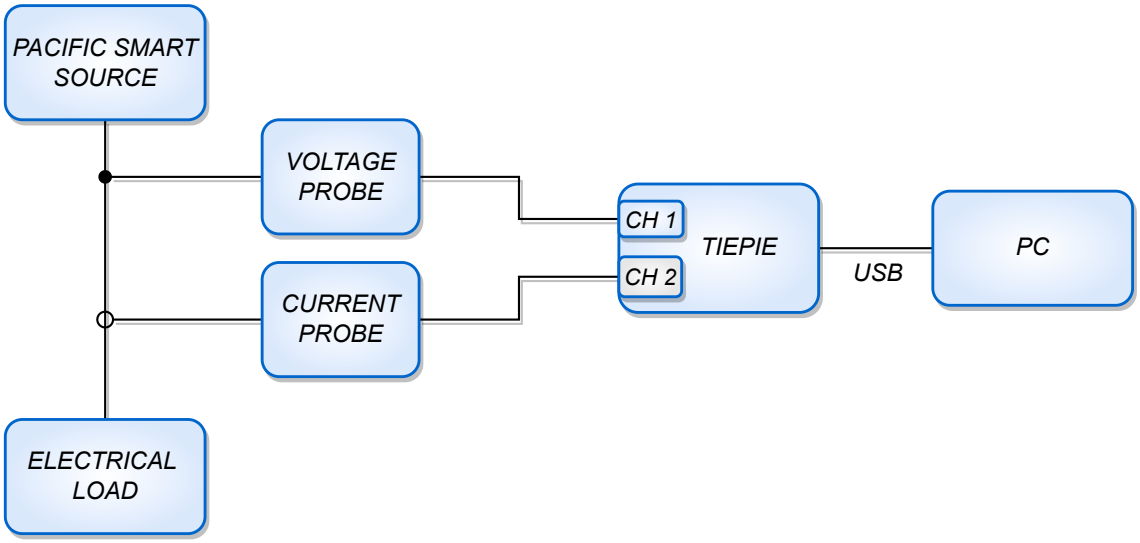


Figure 4.2: **Block Diagram of Data Acquisition Setup** [80]

eLAMI refers to 360 days of simulation of a house having 36 connected appliances, as described in Tab. 4.3. But in the future other years will be added. The "ID" is the appliance identifier, N_S is the overall discrete number of tested operating states (including the OFF state) and the P_{Nom} is the upper bound each appliance can absorb, according to the manufacturer's indications. At each measurement interval, 433 electrical parameters are computed, whose details are reported in Subsection 4.1.3. In addition to the information on the individual load, the same parameters are also calculated for the total aggregate and three partial sub-aggregates consisting of subsets of loads. In addition to the calculated quantities, current operating state is also provided.

The partial aggregates are provided for the three different identified zones of the virtual house. For each zone, a subset of loads was defined. The combination of loads for each sub-group was chosen to obtain aggregates with a progressive number of loads, as shown in Fig. 4.1. This is an important feature for the structure of eLAMI as it offers researchers the possibility to test the algorithms on an increasing number of loads, therefore on an increasing complexity level.

Although all loads are connected in parallel, their supply is carried out by means of radial lines. Therefore, in eLAMI, the voltage at the terminals of each load is not the same but depends on the load conditions.

Moreover, the reported OFF state, for some devices, is an indication of a standby state. Therefore, the absorbed current is slightly different than zero.

Simulation basis: data acquisition

In order to define the frequency spectrum of a load, as mentioned above, it is necessary to experimentally acquire its absorption profile. Through a measurement campaign in Industrial Measurement Laboratory (LAMI) in the University of Cassino and Southern Lazio, 36 residential typical electric loads (Tab. 4.3) have been acquired. Current and voltage waveforms for each devices has been recorded and get the corresponding reference profiles for use in data generation software. The block diagram of the experimental setup for data acquisition is shown in Fig. 4.2.

The adopted power supply system is a Pacific Smart Source, an electrical network emulator which allows reproducing any mains profile both in terms of amplitude and harmonic content

Table 4.3: **Electrical Loads Specifications. Returns the appliance identifier (ID), the name of the loads (Appliance), the number of states tested (N_S) and and the maximum active power that each load can absorb (P_{Nom})**

ID	Appliance	N_S	P_{Nom} [W]
1	Beard Trimmer	2	4
2	Blender	2	60
3	Boiler	2	220
4	Coffee Machine	2	1200
5	Dehumidifier	2	190
6	Desk Lamp	4	10
7	Electrical Oven	3	2400
8	Electric Toothbrush	2	5
9	Fan	4	40
10	Fan Heater	4	1900
11	Fridge	2	140
12	Hair Clipper	2	6
13	Hair Dryer	3	420
14	Hair Straightener	2	65
15	Headphones Charger	2	5
16	Hoover	3	600
17	Infrared Heater	2	1350
18	Iron	2	1300
19	Kettle	2	2000
20	Lamps Bathroom	2	30
21	Lamps Bedroom	2	90
22	Lamps Closet	2	75
23	Lamps Corridor	2	55
24	Lamps Entrance	2	65
25	Lamps Kitchen	2	125
26	Lamps Livingroom	2	145
27	Microwave	2	1300
28	Notebook	2	60
29	PC-Desktop	2	300
30	Printer	2	750
31	Router	1	5
32	Smart TV	2	105
33	Smartphone Charger	2	25
34	Toaster	2	640
35	TV	2	85
36	Washing Machine	4	800

[94]. In particular, it has been used as arbitrary voltage generator to supply the electric loads.

In order to emulate real working conditions, the power grid voltage was first acquired and the corresponding harmonic characteristics were calculated up to the 50th harmonic order. The harmonic coefficients obtained, in terms of amplitude and phase, are used as an input for the Pacific Power Source.

The emulated mains voltage profiles both with and without load, are shown in Figs. 4.3, 4.4

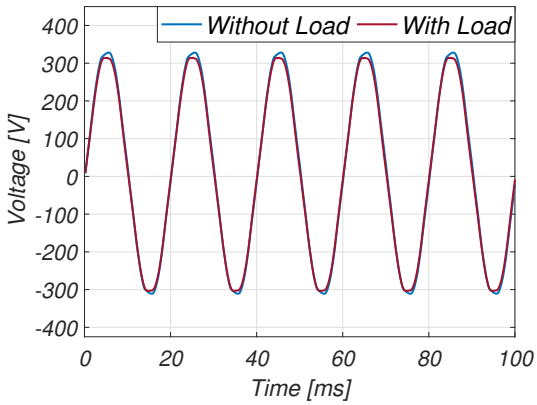


Figure 4.3: **Emulated Voltage profiles with load and without load** [80]

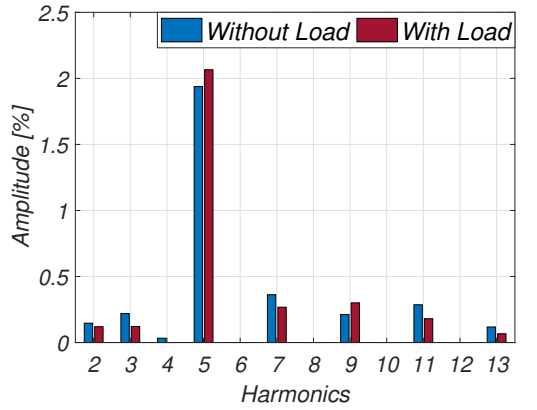


Figure 4.4: **Percentage harmonic coefficients normalized with respect to the fundamental frequency** [80]

along with the harmonic contents. For clarity we report the distribution of the percentages harmonic coefficients from the 2th to the 13th order. The harmonic coefficients shown in Fig. 4.4 were obtained by extrapolating the characteristics of the harmonics of the voltage signals and evaluating the amplitudes in percentage terms with respect to the corresponding fundamental tones.

Furthermore, it can be seen from the figures that there are differences between the voltages in different load conditions. In particular, the voltage under load is lower and has a slightly different distribution of harmonic coefficients. As expected, the voltage profiles of Fig. 4.3 do not reproduce a perfect sinewave, since it contains the harmonics contributions.

In Tab. 4.4, is reported a numerical comparison between the main acquired voltage and the emulated one, both with and without load. The compared values are: the root mean square (RMS) value (V_{RMS}) of both the overall voltage and only the voltage first harmonic ($V_{RMS\ Fundamental}$). Furthermore, the offset component (V_{DC}), the peak value (V_{PK}), and the total harmonic distortion (THD) voltage are also compared.

Through the use of a Tektronix P202A Hall effect probe [125], powered by a Tektronix 1103 power supply [124], with a transformation ratio of 100 mV/A, the electrical current flowing in the circuit was measured.

Using a Tektronix P5200 differential probe [123], with a transformation ratio of 1:500, the voltage supplied by the Pacific to the electrical loads was measured.

A TiePie HS5 [127] was used to acquire the measurements through a customized software

Table 4.4: **Comparison of some characteristics of the voltages: Acquired Main Voltage, Emulated Voltage With Load and Emulated Voltage Without Load**

	V_{RMS} [V]	$V_{RMS\ Fund}$ [V]	V_{DC} [V]	V_{pk} [V]	THD_V [%]
Acquired Main Voltage	231.35	230.82	12.79	326.27	3.94
Emulated Voltage With Load	232.27	231.99	10.06	327.31	4.85
Emulated Voltage Without Load	226.89	226.39	10.07	317.86	4.72

developed in Matlab™ environment.

30 repeated measurements (N_{ACQ}) were performed for each operating state tested. The iterated procedure allowed obtaining the corresponding average absorption profiles and the associated standard deviations for voltage and current waveforms.

The sampling frequency (F_s) used for profile acquisition is 5 kHz. For each acquisition (of each load state) 25000 points (N_P) were acquired.

The amplitude resolution of the acquisition system, through the TiePie Hs5, was set to 14 bits per channel, with a full scale of 0.8 for the voltage channel and 2 V for the current channel. Considering the conversion factors of the probes used, therefore (1:500 and 100 mV/A) this gives 400 V for the voltage channel and 20 A for the current channel. Values are chosen in relation to the electrical systems considered. All the sampling characteristics are summarized in Tab. 4.5.

Table 4.5: **Sampling characteristics used for the acquisition of electrical load profiles**

Specifications	Descriptions
F_s	5 kHz
N_P	25000
T_{OSS}	5 s
N_{ACQ}	30
A/D converter resolution	14 bit
Full scale of the voltage channel	0.8 V
Full scale of the Current channel	2 V
Voltage probe conversion factor	1:500
Current probe conversion factor	100 mV/A

Simulator Description

The purpose of this section is to provide a general description of some of the fundamental parts that make up the innovative simulator created for the realisation of eLAMI, highlighting the procedures followed in the simulation. All processing operations were performed in Matlab™ environment.

The simulator consists of three sections: (1) INPUT, (2) PROCESSING and (3) OUTPUT.

For each section, the chronological flow and the main operations are described.

1) The "INPUT" section consists of 4 macro blocks:

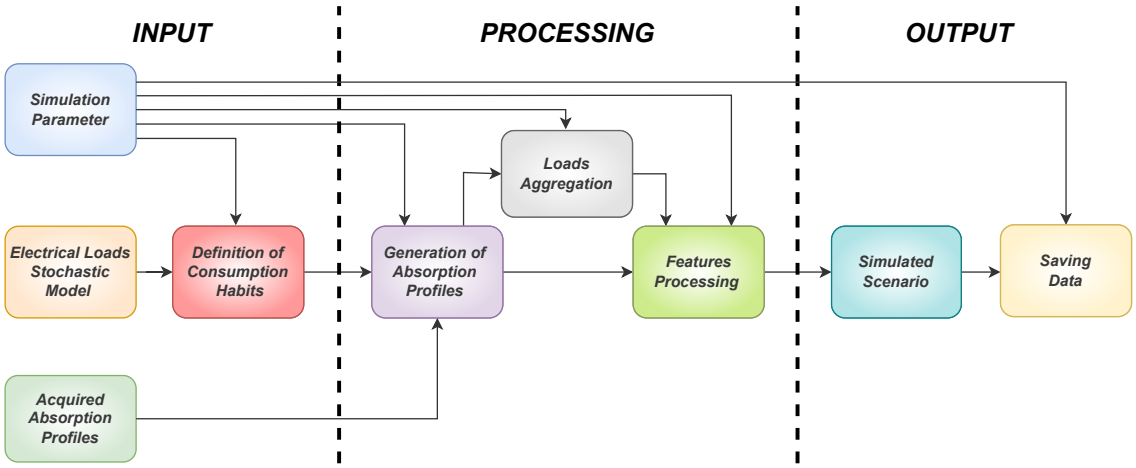


Figure 4.5: **Block diagram of the simulator composed of three sections: "INPUT", "PROCESSING" and "OUTPUT"** [80]

1.1) In "Simulation Parameter" block, specifications for the simulation are defined, e.g. simulation interval (1 year), measurement time (5 s), and aggregation criteria, just to cite a few. The content in round brackets is the assignment of possible parameter values. They are sent to the following blocks, defining the characteristics of the electrical scenario to be simulated.

1.2) In "Electrical Loads Stochastic Model", mathematical relations about operating state changes are defined according to output got by previous surveys performed by the authors in residential buildings.

1.3) In "Definition of Consumption Habits", taking into account the many factors influencing the electric consumption and adopting the mathematical models defined in the previous block, along with the hypothesized simulation interval, typical consumption habits, or patterns, are defined and used as input to generate faithful absorption profiles. The behavioural consumption patterns, defined for each load, take into account daily variability and seasonality. This is made possible by the implementation of a stochastic process, studied ad-hoc for the assumed dynamic system. To explain the suitability of the implemented stochastic model to avoid a mismatch with real scenarios, in our simulation framework it is absolutely unlikely that the blender or hoover will switch on during the night as well as lighting or heating during winter periods for consecutive days is more likely than in summer times. In eLAMI, the defined stochastic model also considers all these factors. In this way, for each load, the days turn out to be dependent.

1.4) In "Acquired Absorption Profiles" contain the reference absorption profiles obtained as described in Section 4.1.3.

2) The PROCESSING section is composed of:

2.1) "Generation of Absorption Profiles" is a block that, taking inputs as described before, generates the voltage and current waveforms related to the conditions to be simulated. Such signals are simultaneously sent to both "Loads Aggregation" and "Features Processing" to perform different operations, as described below. Such waveforms are referred to the i-th iteration for a specific instant of a simulated day and condition of each considered appliance. The compounded values define the "Current Operating State" of each load.

2.2) The "Loads Aggregation" block receives the output of "Generation of Absorption Profiles" and "Simulation Parameter" and aggregates individual loads in a macro-load condition, i.e. considers all appliances belonging to a specific category (e.g. bathroom) as they were one only aggregated load.

2.3) In "Features Processing", all electrical quantities assumed for individual loads and aggregates are calculated. For electrical parameters calculations, our simulator implements the definitions of Power in IEEE-1459, especially referring to single-phase non-sinusoidal case [?]. The calculated electrical parameters refer to: $V_{RMS}{}^{TOT}$ and $I_{RMS}{}^{TOT}$, V_{RMS} and I_{RMS} at the fundamental, of harmonics and DC components alone; active, apparent, non-active and distorted power; power factor and harmonics distorted parameters, for a total of 26 electrical parameters. Furthermore, for better identification of electrical loads signature, the simulator also implements the definitions of measurements in IEC 61000-4-7 Standard [49], regarding harmonics and interharmonics, including the rms values of voltage and current groups up to the 50th harmonic order are calculated, including harmonic sub-groups ($SubG_rV$, G_rV_1...G_rV_{50} ; $SubG_rI$, G_rI_1...G_rI_{50}) and the corresponding phase values of each group calculated for a total of 202 parameters. In addition, the harmonic index of the maximum amplitude tones of voltage/current in each group is also provided, the phase of each harmonic group and the current operating state, to achieve further 205 parameters. Considering the overall processing, the total number of electrical parameters reported in eLAMI is 433.

3) The "OUTPUT" section is composed of:

3.1) In "Simulated scenario", the computed electrical parameters are packed considering the "simulation interval" (e.g. 1 day, 1 month, 1 year) parameter and their size also depends on the measurement time (e.g. 5 s), which is the time resolution over which the 433 parameters are computed. Future simulated years of eLAMI will be added to the main folder.

3.2) In "Saving Data" block, a hierarchical structure is created for saving the dataset, as illustrated in Fig. 4.6. The latter was created to make the data as much usable as possible for the end user. From a hierarchical point of view, eLAMI is divided into a first level "by months", then "by loads" and finally "by calculated electrical parameters".

All information are stored in granular ".csv" files, one for each basic condition (day of the month).

4.1.4 Results

The aim of our work is to provide a dataset with innovative features compared to datasets currently found in the scientific literature, for the evaluation and development of new techniques and algorithms in the field of Smart Energy Applications. Of course, it is of paramount importance that the dataset is physically consistent with real case scenarios.

To this end, we first validate our dataset simulator; then, some dataset peculiarities are highlighted, and finally some examples of eLAMI applications in the field of Smart Energy, in particular Load Profiling, NILM, and Energy Management systems, are proposed.

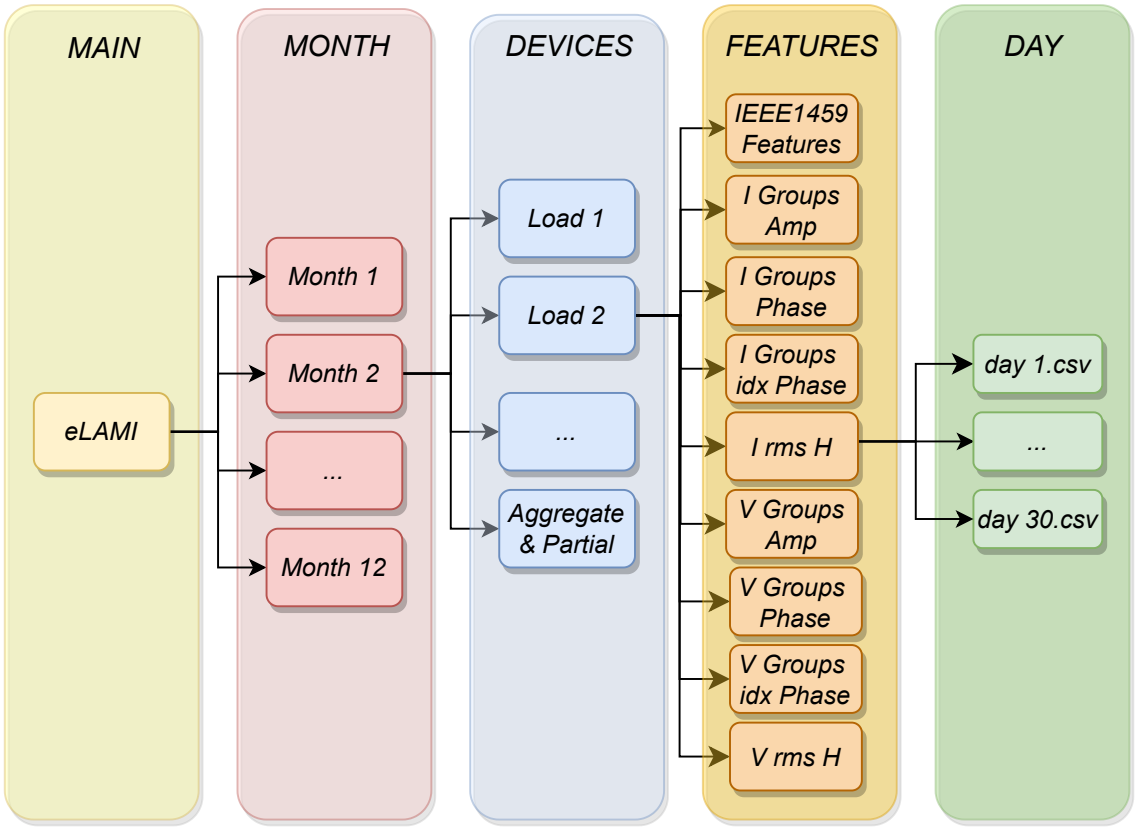


Figure 4.6: **Saving eLAMI Structure** [80]

Validation of Dataset Simulator

For the technical validation of the simulator and the corresponding consistency of the generated data, a comparison between the measurements obtained from a real test and a simulated one has been carried out, by assessing their metrological compatibility. In particular, the electrical scenario assumed for the test, is composed of 3 real loads of eLAMI with different electrical characteristics, namely *"Fan"*, *"Fan Heater"* and *"Smart TV"*. They have been connected to the Pacific network emulator [94], then fed in parallel with the same voltage signal used for the creation of eLAMI reference profiles 4.1.3. At the same time, the absorption profiles, at the ends of every single load and the *"Aggregate"*, were monitored using a laboratory wattmeter, the Precision Power Analyzer WT3000 [137]. The same scenario, without WT3000 measurement instrument, has been replayed in the simulation environment, by starting from the reference profiles previously acquired. Test set-up settings are total test duration (1 hr), measurement time (5 s), and the total number of measurement points (720).

At the end of the test, a comparison was made between the results obtained in the two cases, with the aim of showing: *i)* a comparison between variability ranges in the case of real and simulated data; *ii)* the metrological compatibility of the simulated measurements, and thus of eLAMI, with real acquired values; the combination of *i)* and *ii)* leads to state the validity of the simulator implementation.

Starting from *(i)*, a comparison between the values obtained in the real and simulated cases is shown in Fig. 4.7, in terms of variability ranges. In this figure, the behaviour of V_{RMS} (4.7.a), I_{RMS} (4.7.b), P (4.7.c) and S (4.7.d) are reported, for each individual considered load and its corresponding aggregate.

In this case, the variability ranges are overlapped, although WT3000's related range is

almost always narrower than the simulated one. This is because the WT3000 has a higher accuracy level than what can be obtained by adopting the set-up used in the acquired reference profiles for simulation. In any case, since WT3000 has been chosen as a reference instrument, it is expected that it can exhibit a far better metrological performance. Furthermore, the reference profiles for the generation of eLAMI were constructed to take into account the variability of the data over a time horizon longer than 1 hour (total test duration). Of course, increasing the acquisition time would tend to increase the variation intervals of the WT3000 distributions, due to measurand variability.

An interesting aspect, that can be seen in this figure looking at the WT3000 measured values, is the behaviour of the V_{RMS} (Fig. 4.7.a)). The 3 loads are simultaneously supplied by the same power source, each through its own power line: such a setting can cause a potential voltage drop. As reported in Fig. 4.7.c), the *Fan Heater* is the load with the highest absorption: consequently, the V_{RMS} at its ends is the lowest (Fig. 4.7.a)). Conversely, the WT3000 records the highest voltage at the ends of the *"Fan"*, which is the closest to the aggregate's one, i.e. the power source. Looking at the voltage behavior of eLAMI, the same trend can be observed.

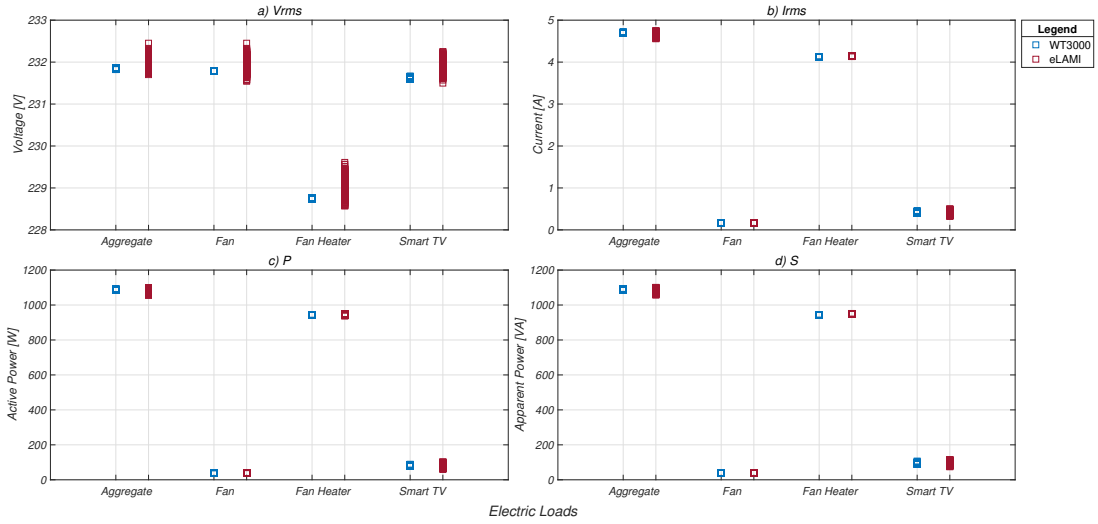


Figure 4.7: **Comparison between WT3000 and eLAMI for devices: *"Fan"*, *"Fan Heater"*, *"Smart TV"* and *"Aggregate"*. Comparison made in terms of features: a) Voltage RMS - b) Current RMS c) Active Power d) Apparent Power** [80]

To demonstrate *(ii)*, only the mean value and standard deviation of a few monitored Features for the *"Fan"* load are reported, for sake of brevity, in Tab. 4.6. In particular, the considered features are: V_{rms}, I_{rms}, P and S.

The standard deviation values of the measurements recorded by the WT3000, as we expected, are much smaller than those related to eLAMI, due to the simulator design parameters, which had the purpose to replicate a typical commercial smart meter less accurate than the adopted reference (WT3000). Nevertheless, from a measurement point of view, the intervals (μ-σ, μ-σ) belonging to WT3000 and eLAMI are generally overlapped, demonstrating the validity of the generated dataset.

The validity of the algorithms implemented in the simulator for the generation of eLAMI is evident when analyzing the values and behaviors obtained from the features analyzed in Fig. 4.7 and Tab. 4.6. Furthermore, the similarity between the behaviors of the *"Aggregate"* highlights the consistency of the process implemented in the simulator.

Table 4.6: **Comparison of mean and standard deviation values obtained from WT3000 and eLAMI for the electrical load *"Fan"*. Considered features: V_{rms}, I_{rms}, P and S**

Feature		WT3000	eLAMI
V_{rms} [V]	μ	231.68	231.94
	σ	0.13	0.59
I_{rms} [mA]	μ	159.90	161.30
	σ	0.58	1.40
P [W]	μ	36.38	37.12
	σ	0.22	0.37
S [VA]	μ	36.82	37.63
	σ	0.31	0.32

Descriptions of the general characteristics of eLAMI

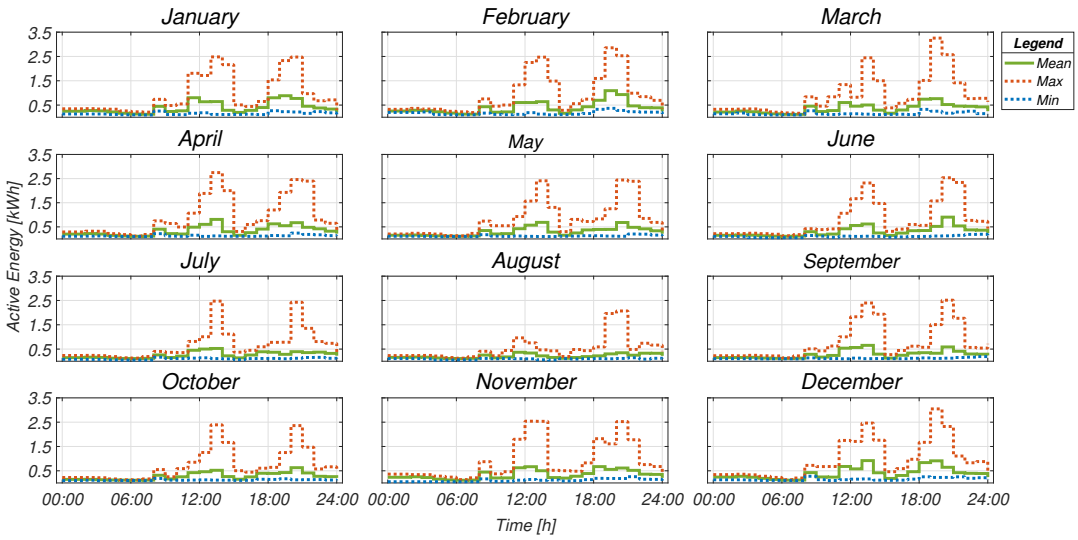


Figure 4.8: **Average Monthly Active Energy Consumption Profiles of the Aggregate Load**

As highlighted above, one of the peculiarities of eLAMI is the variability of the data, both in terms of operating states of the monitored devices and consumption habits. Fig. 4.8 shows the average monthly active energy consumption profiles of the aggregate load, for the 24 hours of the day (x-axis), for each month. In particular, in terms of active energy consumption (y-axis), the curves show: in green, the average trend, in red the maximum reached for each hourly interval of the month considered, and similarly in blue the minimum. First of all, when analyzing the individual month, we can see the variability of the absorption curve during the 24 hours of the day, consistent with what happens in the residential area. In particular, the curve shows an increase in the early morning hours followed by a rapid decrease until midday when the second absorption peak occurs. A further decrease follows this in consumption before arriving at the evening hours characterized by the highest energy absorption.

In addition to the variability during the day, by comparing the different months, we can value the seasonality of consumption and thus the variation in electrical-behavioral habits. In particular, consumption is higher during the winter months than in the summer months, which is particularly evident when comparing August and December. This is because in the

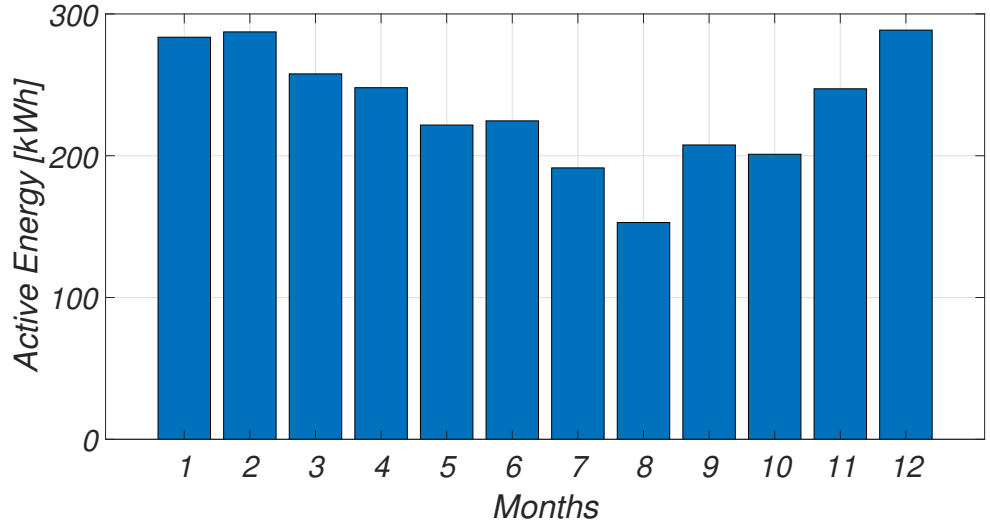


Figure 4.9: **Aggregate Active Energy Load Curve per Months** [80]

winter months, according to defined habits, there is greater use of certain high-consumption devices, such as *Boiler*, *Electric Oven* and *Microwave*. Furthermore, it should be noted that during the winter months lamp utilization is higher than in the summer months, which has an impact on the consumption peaks mentioned above.

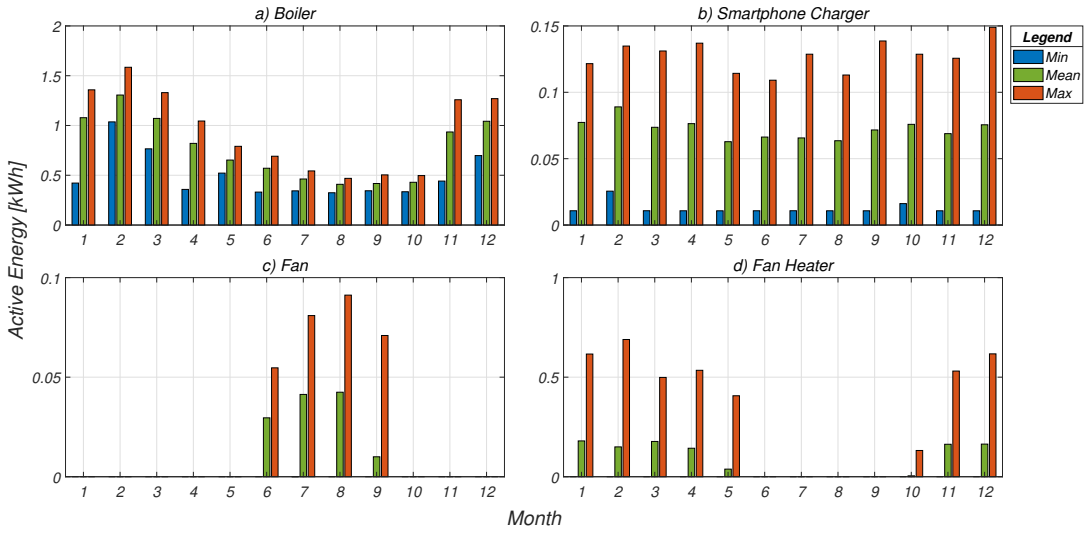


Figure 4.10: **Seasonality Example of Some Different Electrical Loads - a) Boiler - b) Smartphone Charger - c) Fan - d) Fan Heater** [80]

A further interesting aspect is a non-zero consumption during nighttime hours present in all months, due to the presence of some devices in standby mode, characterized by minimal but not zero power consumption. This ceiling does not show much variability in terms of consumption, as the devices on stand-by during the night are almost always the same, so the 3 curves (average, minimum and maximum) almost overlap, with a consumption of less than 0.5 kWh.

In Fig.4.9 the total electricity consumption for each month of the simulated year is shown.

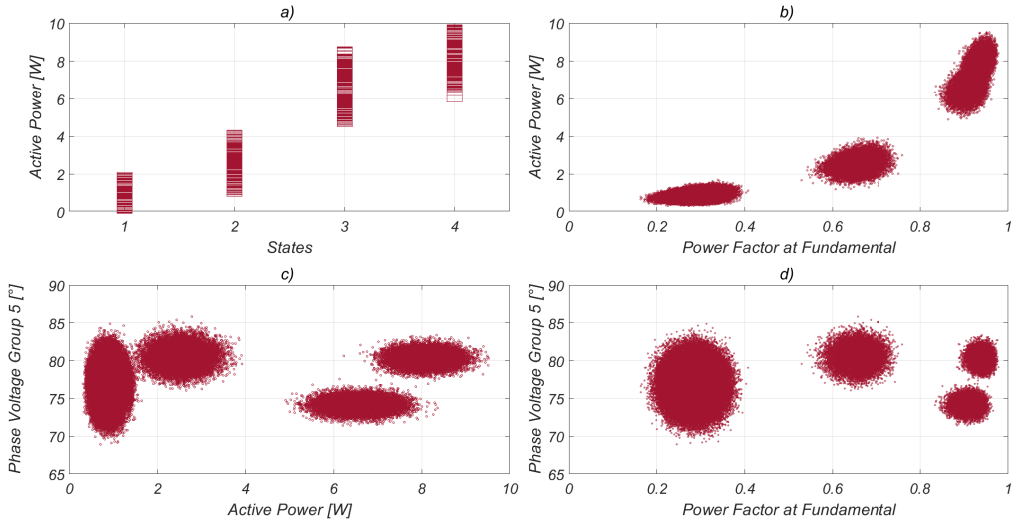


Figure 4.11: **Example of Load Profiling Application for Desk Lamp - a) Power consumption per state b) Active Power as a Function of Power Factor at 50 Hz c) Phase of 5th Voltage Harmonic Group as a Function of Active Power d) Phase of 5th Voltage Harmonic Group as a Function of Power Factor at 50 Hz** [80]

In this case, the seasonality of consumption and thus the variation of the energy absorbed in the different months is particularly evident. Since there is no cooling system, total consumption is higher in autumn and winter than in spring and summer. In particular, starting the year in January, electricity consumption is high and remains more or less constant until March, and then begins to decrease in April with the arrival of spring. The lowest peak is reached in August and then starts to rise again from September forward. The highest electricity consumption is recorded in December, while the lowest is recorded in August in perfect analogy with what is shown in Fig. 4.8.

In Fig. 4.10 we report the average daily consumption of 4 electrical loads for each month of eLAMI (in terms of average, maximum and minimum daily consumption), taken as an example, to show the variation in consumption patterns in eLAMI. Specifically, the devices are: *Boiler* (4.10.a), *Smartphone Charger* (4.10.b), *Fan* (4.10.c), *Fan Heater* (4.10.d). Analyzing *Boiler* consumption (4.10.a), it can be seen that the average consumption in winter is higher than in summer, where it is used only for hot water and not for heating. In contrast, the *Fan Heater* (4.10.d) is only used in winter, from October to April, peaking in February. The opposite behavior is obtained by analyzing the *Fan* (4.10.c), which is only used in the summer months, from June to September, with a peak in August. Unlike the others, the *Smartphone Charger* (4.10.b) does not show substantial variations in consumption between months. This is because it is used on average every day of the year in the same way.

Smart Energy Application Examples

Load Profiling

Machine Learning and Artificial Intelligence techniques, in general, are based on the use of large amounts of input data. However, if these techniques have input data that do not correctly describe the phenomenon to be studied, the output may be far from the desired result. This is why feature selection algorithms are very often used to find the best set of features to build useful and robust models of the phenomena studied [122].

In Fig. 4.11.a the active power absorbed by the *Desk Lamp* in January eLAMI is reported as a function of the corresponding assumed states. As previously verified, state variability is present. Consequently, the active power P alone is not able to discriminate the 4 different load operating state because states 1 and 2, and similarly 3 and 4, overlap in terms of P. Therefore, in terms of load profiling, other features must be found for the correct identification of operating states. For example, the active power P as a function of the power factor at fundamental is reported in Fig. 4.11.b. This feature is able to discriminate states 1 and 2 better than P alone, while states 3 and 4 still remain indistinguishable. Conversely, in Fig. 4.11.c we see the phase of the 6th harmonic voltage group identify states 3 and 4 well but not the first two. Combining the two features identified, power factor at fundamental and phase of the 6th harmonic voltage group, Fig. 4.11.d shows an optimal situation where 4 operating states of the load (cluster) considered are clearly visible. It is therefore very clear how the greater number of electrical parameters in eLAMI results in a better representation and distinction between the different electrical signatures of the individual load.

NILM

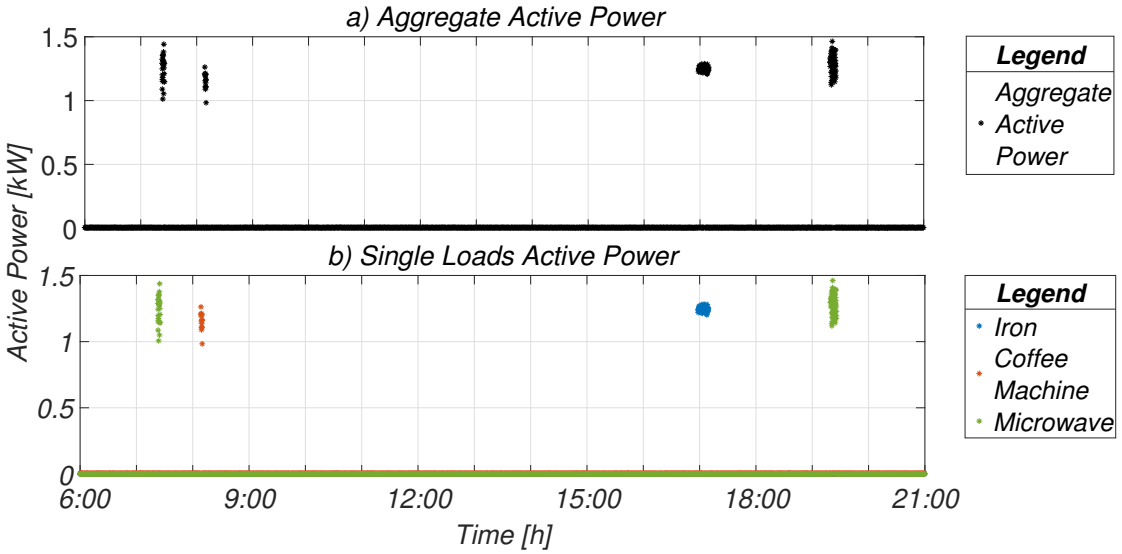


Figure 4.12: **Active Power in time: a) Aggregate - b) Single Loads** [80]

Below is reported an example of application in the NILM (Non-intrusive Load Monitoring) field. NILM is presented as a time series classification problem where the objective is to detect which appliances are active at a given instant and how much each of them contributes to the total percentage of consumption. Due to their advantages, techniques based on the analysis of steady-state features are the most widely used, typically referring to active power only [43]. This feature, however, is not always able to distinguish devices that absorb similar power or have similar operating principles.

In Fig. 4.12.a we report the total active power obtained from an aggregation of 3 loads present in eLAMI and in Fig. 4.12.b the individual active powers absorbed by the 3 considered loads, are reported. In particular, in Fig. 4.12.a and Fig. 4.12.b the active power in time, both for aggregate and some example of single load, are reported. Conversely, Fig. 4.13.a and Fig. 4.13.b show the RMS values of the 6th harmonic current group in time, both for the individual loads considered above and for the corresponding aggregate.

It is evident when analyzing the aggregate active power in Fig. 4.12.a and the single active

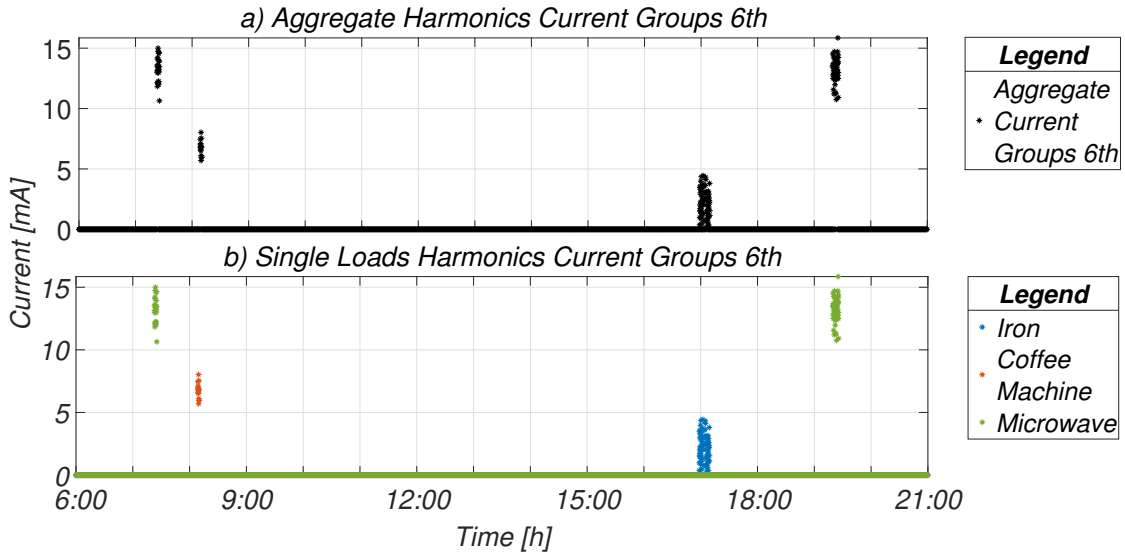


Figure 4.13: **Current Harmonics of 6th group in time: a) Aggregate - b) Single Loads** [80]

power in Fig. 4.13.b, that the 3 loads are indistinguishable due to the problems mentioned above. This represents a critical case for NILM algorithms. Instead, analyzing the current harmonics of 6th group in Fig. 4.13.a for the aggregate and in Fig. 4.13.b for every single load, three different levels of absorption are present, allowing for correct identification of the active loads.

Forecasting

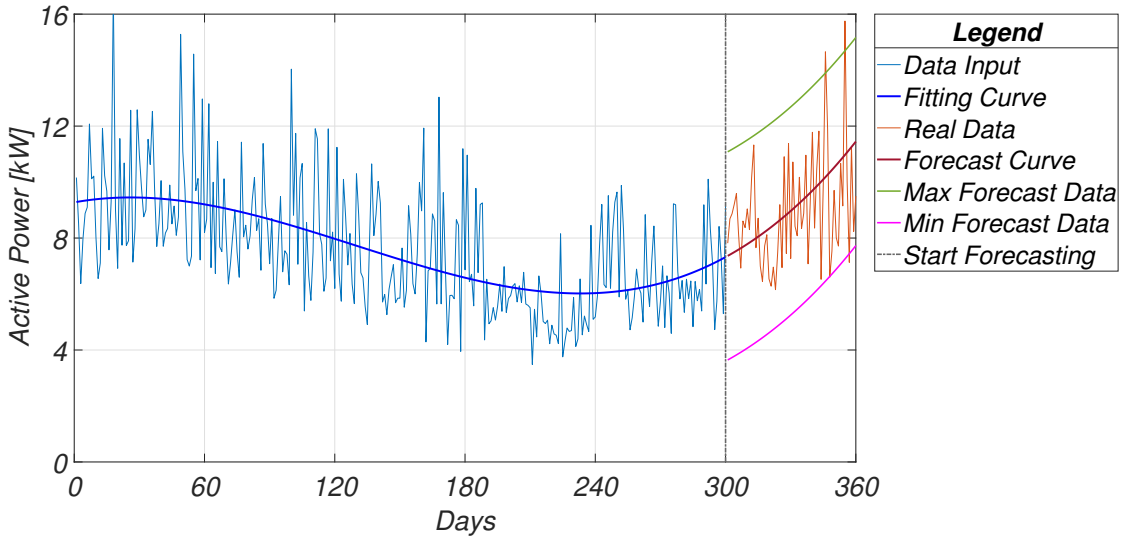


Figure 4.14: **Example of Forecasting for Electric Consumption with eLAMI** [80]

In the world of Smart Energy, statistical and forecasting analyses based on time series are often carried out. In particular, one of the goals of an Energy Management System is to create mathematical models that can simulate trends in electricity consumption as a function of var-

ious factors. Furthermore, through statistical analysis, it is possible to define statistical and/or performance indices of the analyzed system. Some of the benefits of modeling analyzed electricity consumption are: *i)* construction of past seasonal trends and consequently of future ones, *ii)* definition of energy efficiency and optimization plans, *iii)* forecast balancing of electricity networks and performance verification.

To this end, eLAMI also offers the possibility of testing forecasting and modeling algorithms for electrical systems. For example, in Fig. 4.14 based on the knowledge of the first 10 months of the year (300 days) relating to the power consumption trend of the total aggregate, a consumption forecast was made for the last two months of the year. The mathematical model derived is based on a fitting of the input data. Using the mathematical model, a band (*2*DevStd* width) was derived within which the consumption for the 60 forecast days is estimated. As can be seen from this figure, the real data falls within the obtained forecast band with an error of approximately 6.6 %, i.e. 56 days out of 60 estimated correctly.

Furthermore, in diagnostic and predictive terms, thanks to the analysis of time series of data, it is possible to estimate the operating ranges of electrical loads in order to assess and/or predict any decay in their performance and any drift towards fault states.

4.1.5 Conclusion

A novel approach to provide a simulated electrical energy dataset, from the reference signal acquisition to the data validation has been reported in this study. To enforce motivations leading to its building and to prove its suitability for Smart Energy applications, a final section regarding examples of Smart Energy profiling and management is also discussed. The output data, composed of more than 400 electrical parameters and reporting one-year period energy profile simulation of a residential building, are made available for download to enhance research in Smart Energy sector with novel, detailed, validated and wide-applicable data. The acquisition set-up has been chosen according to typical metering capabilities of currently adopted home smart meters. Voltage and Current signals, produced for 36 home appliances, have been processed both in time and frequency domain, to provide a comprehensive set of electrical parameters composing the electrical signature of each considered appliance. To be as close as possible to the real scenario, stochastic models have been also implemented to obtain consumption habits to manage state transitions for each load. In its current status, only nominal operating conditions have been considered, i.e. no failures have been hypothesized during the simulation interval. This is, to the authors' opinion, still reasonable given the very low failure rate of the considered apparatuses in the tested period. The produced data are anyway suitable for most Smart Metering Applications. A second release is intended to be developed, where common failures will be implemented and its aim would be voted to fault location research efforts: it could be seen as an appendix to the current dataset, which results as much complete as possible in terms of detailed representation of typical operating states of considered appliances in a home environment.

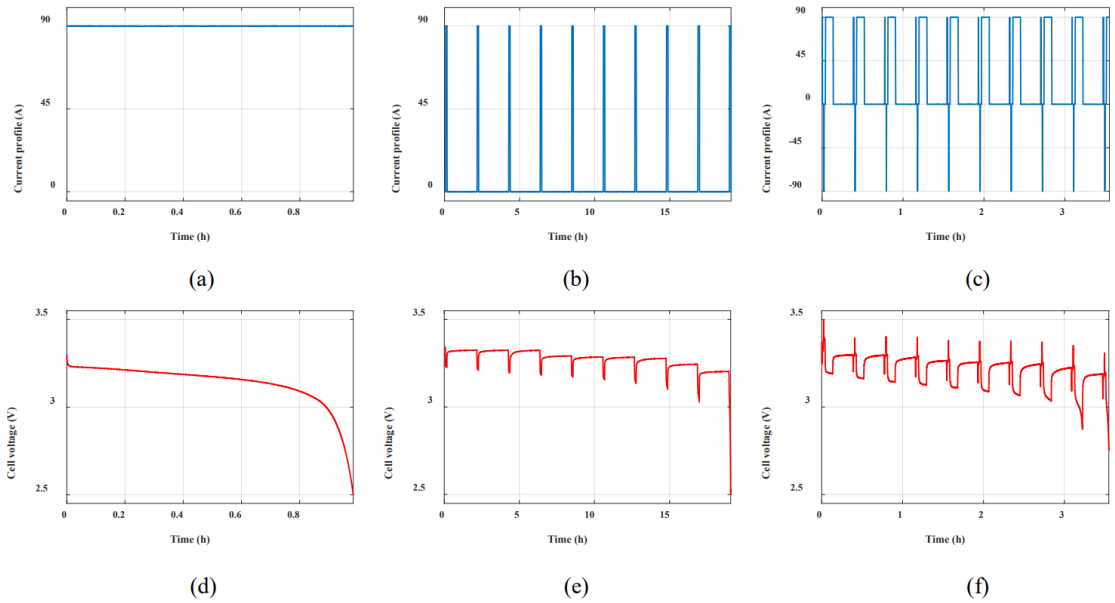


Figure 4.15: Current profiles for (a) static, (b) basic pulsed and (c) HPPC test procedures, with the related voltage responses (d, e, f) [32].

4.2 Characterization procedures

The characterization procedures typically employed can be summarized into three categories: static, basic pulsed, and HPPC (Hybrid Pulse Power Characterization). All procedures are carried out during the battery discharge phase and involve monitoring the voltage while the discharge current is imposed by the electronic load. The operational conditions for characterizations include the discharge current and temperature.

Figure 4.15 (a) and (d) depict a standard discharge, where at t=0, the battery is fully charged and has undergone a relaxation period, essential to bring the battery to an equilibrium state. The discharge is performed without interruptions at a constant current until the minimum voltage at the battery terminals is reached. The static test is employed when assessing the effective capacity of an electrochemical cell.

On the other hand, the pulsed procedure, as shown in Figure 4.15 (b)-(e), involves discharging the battery with a series of current pulses interspersed with relaxation times during which the battery reverts to its electrochemical equilibrium state. Similar to static tests, the test is terminated when the battery reaches the minimum voltage.

The HPPC test, illustrated in Figure 4.15 (c)-(f), was formalized by the "Idaho National Laboratory" in the "Battery Test Manual For Electric Vehicles" [28] and is designed to determine the dynamic capabilities of cells through a test involving discharge, regeneration, and charging phases.

All these tests can be valuable for battery diagnostics or prognosis as will be discussed in the next sections.

4.2.1 Battery models identified through pulsed discharges

Experimental Setup

Characterizing batteries through dynamic discharge methods necessitates a specific setup. Specifically, the setup must be capable of discharging the battery with a constant current

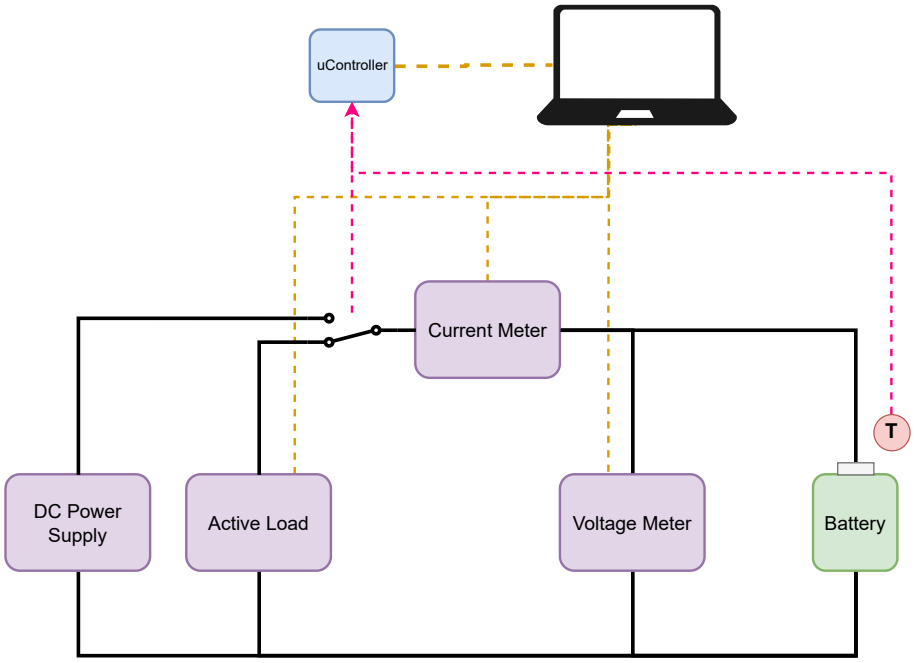


Figure 4.16: Experimental setup for static and dynamic charging and discharging procedures

set by the operator. Furthermore, the setup should be capable of discharging the battery in a pulsed manner, i.e., by interspersing the discharge with relaxation periods for the battery. Throughout the entire test, both the current supplied by the battery and the terminal voltage must be measured and recorded. Finally, the setup should be able to set (via a dedicated climatic chamber) or at least measure the ambient temperature.

The implemented setup is depicted in Figure 4.16, and the instruments within the setup include two AGILENT 34401A multimeters used as voltmeter and ammeter. The ammeter is placed to the valley of the voltmeter because the impedance of the battery is lower than the ammeter's internal resistance. This way allows to reduce to the minimum the measurement error because the voltmeter measures only the battery voltage, while the ammeter measures both the current drawn by the battery and the current drawn by the voltage meter. The voltage meter drawn current is negligible compared with the battery current. There is also a TTi power supply for battery charging, and finally, an active load that, when appropriately controlled, enables discharging the battery with a constant current.

Due to interoperability constraints of the instruments, it was necessary to add relays to change the circuit configuration. The relays and temperature sensors are controlled by a microcontroller. A MATLAB script configures the circuit and the instruments to be used based on the chosen characterization. Additionally, the script retrieves the measured values from the instruments and creates a text file containing the timestamp and the measurements of current, voltage, and temperature.

ECM circuit model selection

In literature, various circuit models are available, and in general, each model aims to strike a balance between circuit complexity (and hence computational complexity) and prediction accuracy based on specific needs. Furthermore, there are models that focus more on predicting battery voltage, while others emphasize "storage capability" prediction. In this case, the focus is primarily on predicting battery voltage. The chosen model is a Thevenin model, which

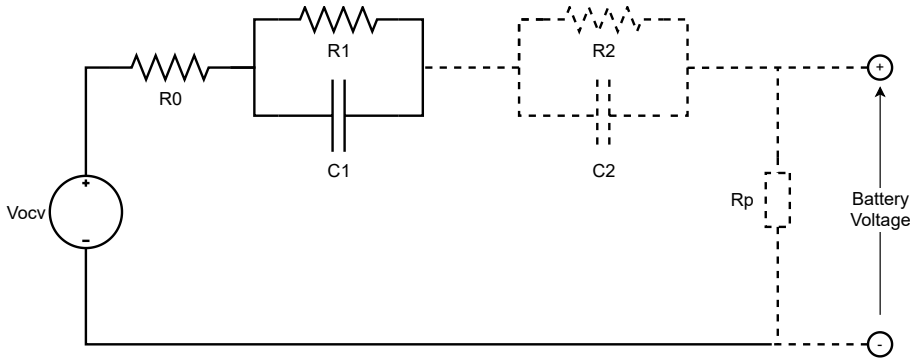


Figure 4.17: Example of the electrical circuit model

employs an ideal voltage source in series with a passive resistor and an RC loop, as shown in the Figure 4.17.

The more loops considered in the circuit model, the higher the computational cost for fitting and parameter determination. Some circuit models also incorporate a parasitic branch. Its inclusion becomes necessary because, during battery charging and discharging processes, there are irreversible reactions that lead to battery efficiency loss and, subsequently, energy loss. The extent of these "side reactions" greatly depends on the battery's construction technology; for example, lead-acid batteries exhibit much lower efficiency compared to lithium cells. Modeling these side reactions is challenging, and various models in the literature use voltage and temperature as inputs, while others include current and State of Charge (SoC). In this initial phase, it was decided to adopt a simple circuit model with a single RC loop to minimize computational requirements and neglect the contribution of the parasitic parameter.

Testing Protocol

The chosen measurement protocol depicted in Figure 4.18 involves completely discharging the battery through a standard discharge, which is a constant current discharge. The purpose of this discharge is to bring the battery back to a known state. The discharge is terminated when the battery voltage reaches the minimum voltage specified by the manufacturer. In this condition, the State of Charge (SoC) is 0%.

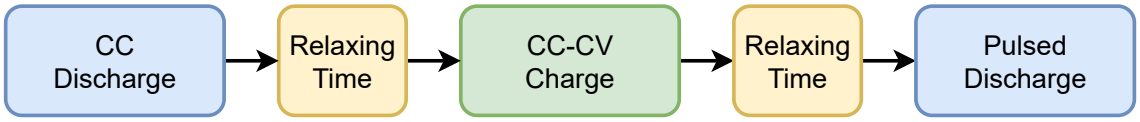


Figure 4.18: Testing protocol for pulsed test

Subsequently, after a relaxation time of 1 hour, the battery is charged following the instructions provided in the datasheet using a "constant current, constant voltage" (CC-CV) method. Initially, the battery is charged with a constant current until it reaches the maximum terminal voltage. Then, the voltage remains fixed at the maximum voltage, and the charging continues until the current falls below a threshold also specified in the battery's datasheet. At this point, a pulsed discharge test is conducted with a constant current. Further details regarding the test discharge are provided in Table 4.7.

Figure 4.19 displays the voltage and current over time, as measured by the two multimeters throughout the entire test. The chart clearly illustrates the initial discharge phase, followed by the second charging phase, and, finally, the pulsed discharge. The relaxation times have been

Table 4.7: Pulsed discharge test configuration values

Discharging Pulsed Test		
$Current_{pulse}$	500	mA
$Current_{off}$	0	mA
$Time_{pulse}$	1800	s
$Time_{off}$	3600	s
$Numberofpulses$	20	-
$Cut_{off}Voltage$	5.25	V

omitted from the presentation.

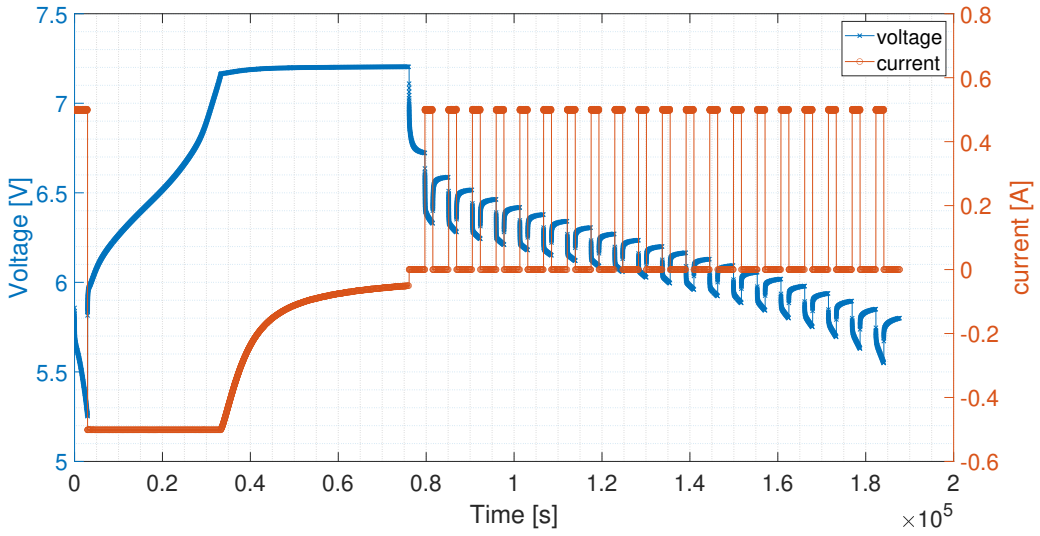


Figure 4.19: Experiment Voltage and Current

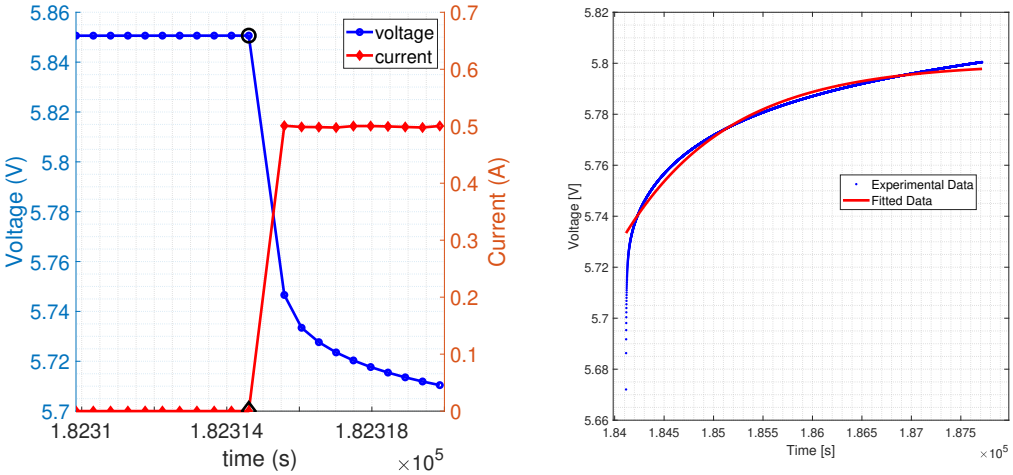
ECM Parameters Fitting

In the context of the circuit depicted in the Figure 4.17, featuring a single RC loop and no parasitic parameters, the following values should be evaluated:

- Open Circuit Voltage (V_{ocv})
- Internal resistance of the battery (R_0)
- Capacitance of the RC loop (C_1)
- Resistance of the RC loop (R_1)

All of these parameters can be determined based on the behavior of voltage and current (which is set by the active load but measured by the ammeter).

The open circuit voltage is measured at the end of each relaxation period, just before the next discharge current is applied. In Figure 2, a specific part of the experiment from Figure 4.19 is shown, and the black circle indicates the moment when the open circuit voltage is extracted. All the extracted open circuit voltages are plotted in Figure 3 as a function of the battery's state of charge.



(a) *Open Circuit Voltage Estimation* (b) *Dynamic branch fitting*

Figure 4.20: Open circuit voltage and dynamic branch estimation

The change in battery voltage is linked to two phenomena modeled through the resistance R_0 and the RC loop $R_1 - C_1$. As the resistance R_0 lacks any dynamic components, it exhibits an immediate response, referred to as "time-invariant." This means that any current variation instantly results in a voltage change. The value of resistance R_0 is evaluated according to Equation 1, particularly as the ratio between the voltage change and the current change as soon as the battery is requested to deliver current. The typical time window used to determine R_0 is denoted by τ and is on the order of tens of seconds.

$$R_0 = \frac{\Delta V}{\Delta I} \tag{4.1}$$

On the other hand, the $R_1 - C_1$ loop is obtained through a curve-fitting process performed using MATLAB's nonlinear least squares curve fit. Equation 2 represents the relationship used within the curve fitting. This relation came from the Kirchoff laws applied to the circuit in Figure 4.17. whereas the initial ($t = 0$) condition for the first order differential equation over the $R_1 - C_1$ loop is the voltage across the R_1 resistance.

$$V_{bat} = V_{ocv} - R_1 I e^{-t/(R_1 C_1))} \tag{4.2}$$

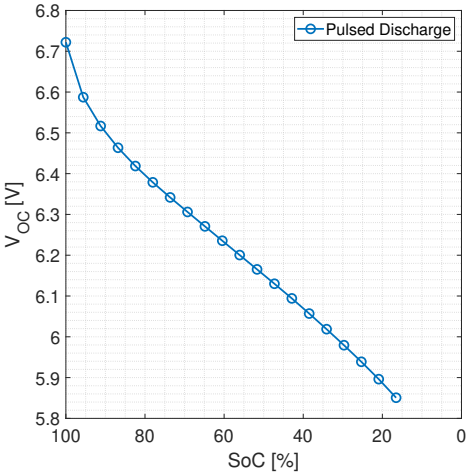
Results and Validation

4.21a displays the open circuit voltage as a function of the battery's state of charge (SoC). The open circuit voltage exhibits a linear relationship with SoC, except in the higher SoC range where it appears to follow an exponential trend. In contrast, the resistance R_0, shown in 4.21b, exhibits a strongly nonlinear and non-biunivocal behavior, much like R_1 and C_1, which are presented in 4.21c and 4.21d, respectively.

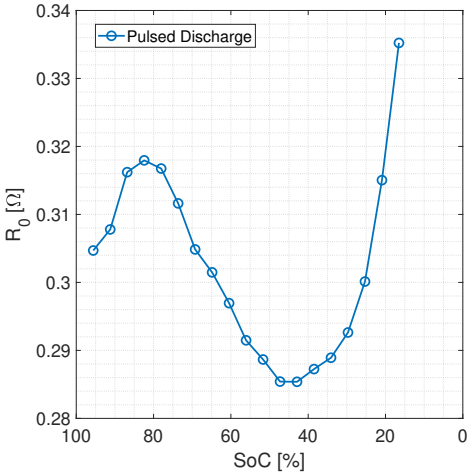
Model validation is performed using the protocol outlined inFigure 4.22.

The measured current serves as a forcing function for both the model, with parameters V_{ocv}, R_0, R_1, and C_1 depending on SoC, and of course, for the battery.Figure 4.23 illustrates the comparison between the experimentally measured battery voltage and the battery voltage simulated by the model.

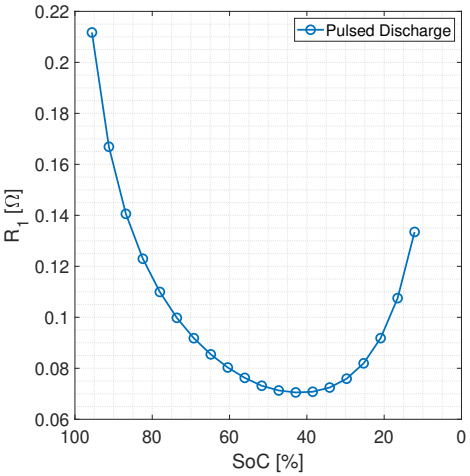
Consequently, the extracted model can be used to predict battery voltage, provided that a characterization is conducted to determine the circuit parameters and the knowledge of the



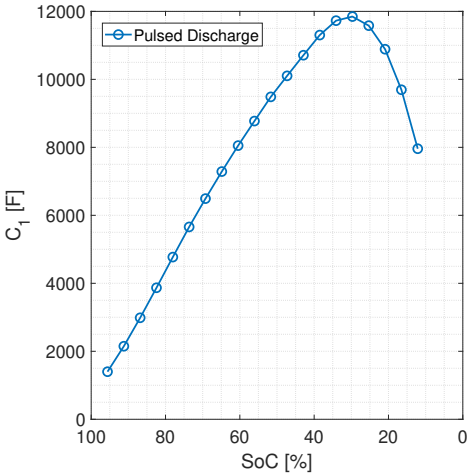
(a) *Open Circuit Voltage as function of SoC*



(b) *Battery internal resistance as function of SoC*



(c) *Dynamic Component Resistance as function of SoC*



(d) *Dynamic Component Capacitance as function of SoC*

Figure 4.21: Fitted Equivalent Circuit Parameters as a function of the SoC

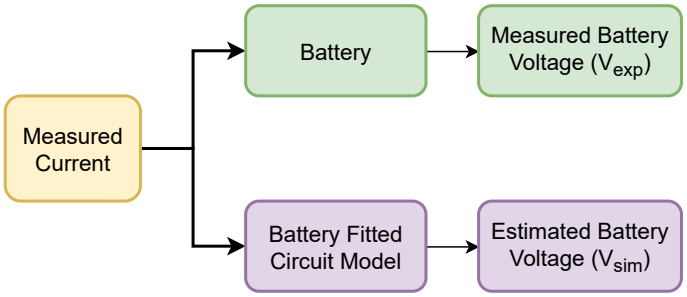


Figure 4.22: Validation Protocol

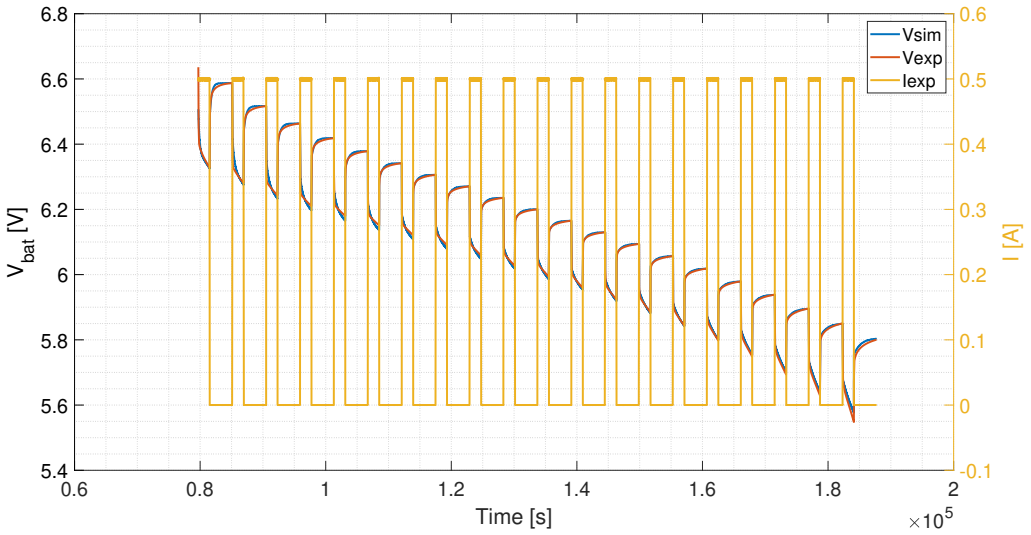


Figure 4.23: Validation results

state of charge. However, a limitation of this approach lies in its inability to reverse the system and use the measured voltage to estimate the state of charge since the relationships of R_0, R_1, and C_1 are non-biunivocal.

4.2.2 Characterization Methods based on Abstract Models with Genetic Programming

This section introduces an innovative methodology that utilizes a Genetic Programming algorithm to establish behavioral models for capturing the transient behavior of a battery's output voltage. These models establish analytical relationships between the battery's voltage and factors such as its state-of-charge, charge/discharge rate, and temperature. In contrast to widely-used equivalent circuit-based models, explained and implemented in the previous section, one of the key advantages of this approach is the significant reduction in the effort required to generate the experimental dataset needed for identifying model parameters. Furthermore, the system could be forced to find a solution which could be inverted in order to determine the SoC.

The Genetic Programming algorithm generates a set of optimal "candidate" analytical models, each associated with relevant metrics that assess performance indicators, including simplicity and accuracy. The methodology is applied here to describe the transient discharge phase of a Lithium Iron Phosphate (LiFePO4 or LFP) battery under realistic operational con-

Table 4.8: Main specifications of the 3.2 V 90 Ah LFP battery.

Battery Features	Value
Chemistry	Lithium Iron Phosphate (LFP)
Nominal capacity	90 Ah
Nominal voltage	3.2 V
Minimum voltage	2.5 V
Maximum voltage	3.65 V
Maximum charging/discharging continuous Crate	1C
Maximum operating temperature	0°C – 55°C (Charging) -20°C – 55°C (Discharging)

ditions. This includes variations in state-of-charge ranging from 20% to 80%, discharge rates spanning from 0.25C to 1C, and temperatures from 5°C to 35°C.

The resulting models are rigorously validated against experimental data, providing diverse solutions that allow for the selection of an appropriate trade-off between accuracy and simplicity. These selected models ensure a relatively low level of relative root mean square error, with a maximum of 0.31% and 0.22% across the entire analysis range.

Battery Characterization

this study employs a precise static multi-rate testing procedure to conduct a thorough analysis of a 90-Ah LFP battery in its discharging mode. The key characteristics of this battery are outlined in Table 4.8. The voltage at the battery terminal (V_B) varies depending on the SoC, the discharge rate (Crate), and the temperature (T). As a result, four distinct Crate values are chosen (0.25C, 0.33C, 0.5C, and 1C) and four temperature settings (5 °C, 15 °C, 25 °C, and 35 °C) to serve as our operating conditions for the static characterization. For each test, we determine the SoC using the Coulomb counting method, which involves integrating the current over time.

The experimental setup is simple, the battery is connected to a bidirectional device ("SM70-CP-450 Delta Elektronika") which is capable of both charging and discharging the battery. Furthermore, this device can measure both voltage and current.

From the extensive data collected during the static characterization, the number of operating conditions for our modeling dataset are defined. The specifics of the experimental settings for our Gaussian Process-based modeling approach are summarized in Table 4.9. In particular, we include all Crate and T values, but we focus on SoC values in increments of 2.5%, within the range of 20% to 80% (resulting in 25 values) for the GP dataset.

Consequently, the resulting dataset comprises a total of 400 distinct operating conditions, encompassing various combinations of SoC, Crate, and T values, for which we have experimentally obtained the battery terminal voltage. To ensure precision, the standard measurement uncertainty based on the specifications of the bidirectional power supply is estimated. Specifically, is calculated the expanded measurement uncertainty for the selected operating conditions with a confidence level of approximately 98%, assuming that the measured voltage adheres to a uniform probability distribution. This allows to represent the measured battery voltage within a coverage interval with a confidence level of approximately 98%.

Table 4.9: Adopted settings for GP-based modeling.

Settings	Conditions
4 Crate values	0.25 C, 0.33 C, 0.5 C, 1 C
4 Temperature values	5 °C, 15 °C, 25 °C, 35 °C
25 SoC values	Minimum 20%, maximum 80%, with step of 2.5%

GP settings and multi-objective optimization

The GP algorithm's objective is to express the battery terminal voltage ($V_{B,bhv}$) as a function that depends on State of Charge (SoC), Crate, and temperature, as stated in Equation 4.3. To assess the "quality" of each model generated by the GP algorithm, three objective functions are employed: the error between the model's outputs and the training data (evaluated by the objective function F_{RMS}), the model's complexity (evaluated by the objective function F_{cmp}), and the monotonic behavior of its coefficients concerning Crate (assessed by the objective function F_{mnt}). This approach results in a multi-objective optimization problem, with the fitness function F_{FIT} expressed as a weighted sum, as shown in Equation 4.4.

$$V_{B,bhv} = f[SoC, C-rate, u_1, ..., u_n] \tag{4.3}$$

$$F_{FIT} = \alpha F_{RMS} + \beta F_{cmp} + \gamma F_{mnt} \tag{4.4}$$

Following there is an explanation of GP settings (e.g., how the GP operators are set) and on the objective functions F_{RMS}), F_{cmp}, and F_{mnt}.

GP Settings

Table 4.10 provides a summary of the GP settings and their associated parameter values. Figure 4.24 outlines the components of the terminal set (external nodes) and the non-terminal set (internal nodes) used in the GP algorithm for generating models. In this study, different complexity factors have been assigned to the external and internal nodes to control the vertical development of the models, favoring simplicity in the terminal set elements.

The complexity of a model (c_f) is determined as follows:

- Each external node (input or coefficient) contributes 0.8 to the overall complexity.
- Each internal node contributes to complexity, and this contribution varies depending on the type of algebraic operator or basic function it represents.
- For each internal node, there is an additional contribution, equal to the product of the complexity of the function and its argument, with different complexities for the function depending on whether the argument is an external or an internal node.

These choices are made to mitigate the "bloat" phenomenon, which might lead to models with excessive layers of nested functions or complex operations on simple functions.

GP objective functions

Table 4.10: GP algorithm settings and parameters.

Description	Value
Number of runs	50
Population size	100
Generation number	50
Maximum tree size	25 nodes
Individuals subjected to elitism	1% of the current population
Selection operator	binary tournament
Crossover operator	subtree crossover
Mutation operator	subtree & node mutation
Crossover probability	0.80
Subtree mutation probability	0.20

Function	Arity	Description	External nodes complexity	Inner node complexity
sum	2	$h+g$	0.6	1
subtraction	2	$h-g$	0.6	1
multiplication	2	$h \cdot g$	0.75	1.1
division	2	h/g	0.85	1.2
reciprocal	1	$1/h$	0.85	1.2
negation	1	$-h$	0.6	1
square root	1	$\sqrt{h}$	1	1.5
square	1	h^2	1	1.5
natural log.	1	$\ln(h)$	1	1.5
natural exp.	1	e^h	1	1.5
sine	1	$\sin(h)$	1	1.5
cosine	1	$\cos(h)$	1	1.5
tangent	1	$\tan(h)$	1	1.5
inverse sine	1	$\sin^{-1}(h)$	1	1.5
inverse cosine	1	$\cos^{-1}(h)$	1	1.5
inverse tangent	1	$\tan^{-1}(h)$	1	1.5

Figure 4.24: GP external and internal node complexities.

- **Model error** F_{RMS}: To derive each coefficient $u_i (i = 1, \ldots, n)$, the m values u_{ij} $(j = 1, \ldots, m)$ are processed using the Levenberg-Marquardt Non-Linear Least Square (NLLS) optimization method. These u_{ij} values are determined based on a data vector of secondary variables with a size of m. Specifically, for each $j = 1, \ldots, m$, the NLLS method identifies the optimal values of uij that minimize the relative root mean square (RMS) error between the output values within the training dataset and the model's predictions. This evaluation is carried out by computing the objective function for the relative RMS error, as described in Equation 4.5:

$$F_{RMS} = \frac{1}{100}\sqrt{\frac{1}{p_{Tr}}\sum_{k=1}^{p_{Tr}}(\frac{V_{B,k} - V_{B,bhv,k}}{V_{B,k}})^2} \tag{4.5}$$

- **Model complexity** F_{cmp}: Numerous methods exist for assessing the complexity of a model. A widely employed technique involves the evaluation of the depth and quantity of nodes constituting the tree structure. In this study, we have introduced varying complexity factors for the external and internal nodes, as outlined in Figure 4.24. Subsequently, the overall complexity (c_f) of each generated GP model is calculated following the procedure explained before. The normalized complexity objective function, denoted as F_{cmp}, is determined as shown in Equation 4.6.

$$F_{cmp} = \frac{c_f - c_{f,min}}{c_{f,max} - c_{f,min}} \tag{4.6}$$

in Equation 4.6 c_f represents the complexity factor of a specific model function, with predefined minimum (c_f, and maximum ($c_{f,max}$ = 80) complexity levels tailored for this particular scenario. It's important to emphasize that the function F_{cmp} exclusively pertains to the complexity of the model function f and is thus linked solely to its reliance on the primary variables SoC and Crate.

- **Model monotonicity** F_{mnt}: The objective function Fmnt expresses a qualitative characteristic of the u coefficients, namely their monotonicity with respect to temperature values for the case under study, and is calculated as in Equations 4.7 - 4.11.

$$F_{mnt} = \frac{1}{k}\sum_{i=1}^{k} 2min\left\{X^{(-)}, X^{(+)}\right\} \tag{4.7}$$

$$X^{(+)} = \sum_{j=1}^{m-1}(C_{ratej+1} - C_{ratej})\frac{\dot{u}_{i,j+1}^{(+)} + \dot{u}_{i,j+1}^{(+)}}{2} \tag{4.8}$$

$$X^{(-)} = \sum_{j=1}^{m-1}(C_{ratej+1} - C_{ratej})\frac{\dot{u}_{i,j+1}^{(-)} + \dot{u}_{i,j+1}^{(-)}}{2} \tag{4.9}$$

$$\dot{u_{i,j}^{(+)}} = max(\frac{u_{i,j} - u_{i,j-1}}{C_{ratej} - C_{ratej-1}}, 0) \tag{4.10}$$

$$\dot{u_{i,j}^{(-)}} = max(\frac{u_{i,j-1} - u_{i,j}}{C_{ratej} - C_{ratej-1}}, 0) \tag{4.11}$$

Table 4.11: Fitness weighting coefficients and bias/scaling coefficients.

α	β	γ	bias coefficient	scaling coefficient
0.9	0.05	0.05	forced	forced
0.85	0.075	0.075	forced	forced
0.8	0.1	0.1	forced	forced
0.9	0.05	0.05	free	free
0.85	0.075	0.075	free	free
0.8	0.1	0.1	free	free
0.33	0.33	0.34	forced	forced
0.75	0.2	0.05	forced	forced
0.5	0.3	0.2	forced	forced

Genetic Programming: dataset, metrics and results

The dataset of the 90-Ah LPF battery serves as the input for executing the GP algorithm. It comprises a total of 400 operating conditions, involving various combinations of SoC, Crate, and temperature values (as outlined in Table 4.9). Like all data-driven algorithms for predictions and decision-making, these input data are typically divided into multiple datasets. Specifically, three datasets play distinct roles in the model development process: training, validation, and test datasets.

The training dataset encompasses the data samples utilized in the identification of the model during the GP evolution. Subsequently, the validation dataset is employed to provide an unbiased evaluation of how well a model aligns with the training dataset, all while fine-tuning the model's hyperparameters. Finally, the test dataset consists of data samples used to provide an impartial assessment of the ultimate model selected from among all models discovered through the training dataset.

In the context of this battery case study, a subset of 40 conditions, equivalent to 10% of the original dataset, is randomly selected to constitute a common test dataset (Tts) for all GP runs. The remaining 360 conditions, accounting for 90% of the original dataset, are randomly partitioned for each run, with 288 conditions forming the training dataset (Ttr) and 72 conditions forming the validation dataset (Tvl).

To explore solutions emphasizing different objective functions, a distinct setup of fitness weighting coefficients (as detailed in Table 4.11) is employed. Among the α, β and γ coefficients, greater weight is assigned to the RMS error objective function F_{RMS}, ensuring the prevalence of individuals (models) with low RMS error. This decision is informed by the initial analysis of experimental data, which reveals a inherent smooth behavior in the battery voltage concerning SoC, Crate, and T, leading to relatively straightforward functions (f) and monotonic coefficients (u). Additionally, various combinations of bias and scaling coefficients have been implemented in the GP algorithm (as described in Table 4.11).

The GP algorithm has been executed through 50 runs for each configuration of weighting coefficients, with a population comprising 100 models evolving over 50 generations in each run. Following each run, all GP models can be preliminarily selected based on their RMS error values computed for both the training and test datasets. In this context, models exhibiting an error lower than 10 mV, which aligns with the level of accuracy required for modeling this specific battery and accounts for the associated measurement uncertainties described in detail later, are considered. Among these solutions, models that are not dominated in the (F_{RMS}), F_{cmp}, and F_{mnt}.) domain are identified, sorted, and assigned numerical labels in ascending order based on their corresponding F_{FIT} values. These models, along with their fitness value

Models	Expressions	F_{FIT}	F_{rms}	F_{cmp}	F_{mnt}	# coeff.
#1	$u_1 + u_2\left[\frac{1}{SoC}\sqrt{C_{rate}} + \sqrt{atan(C_{rate})}\right]$	0.0068	0.0030	0.0560	~0	2
#2	$u_1 + u_2\left[\frac{1}{SoC}\sqrt{C_{rate}} + \sqrt{\sqrt{C_{rate}}}\right]$	0.0082	0.0029	0.0761	~0	2
#3	$u_1 + u_2[u_3 + SoC + atan(C_{rate})]\frac{1}{SoC}$	0.0148	0.0028	0.0342	0.2099	3
#4	$\frac{1}{SoC}(u_1 C_{rate} + u_2) + u_3 C_{rate} + u_4$	0.0203	0.0026	0.1015	0.2583	4
#5	$SoC^2(u_1 SoC + u_2 C_{rate} + u_3) + C_{rate}(u_4 SoC + u_5 C_{rate} + u_6) + u_7 SoC + u_8 C_{rate} + u_9$	0.0207	0.0013	0.3890	0.0008	9
#6	$\frac{1}{SoC}(u_1 C_{rate} + u_2) + C_{rate}(u_3 C_{rate} + u_4) + u_5 C_{rate} + u_6$	0.0259	0.0023	0.2216	0.2559	6
#7	$u_1 + u_2\frac{1}{SoC}[atan(C_{rate}) + u_3]$	0.0304	0.0029	0.0227	0.5339	3
#8	$SoC(u_1 C_{rate} + u_2) + C_{rate}(u_3 C_{rate} + u_4) + \frac{u_5}{SoC} + u_6 C_{rate} + u_7$	0.0308	0.0028	0.2858	~0	7
#9	$\frac{1}{SoC}(-u_1 C_{rate} + u_2) + \frac{1}{C_{rate}}(-u_3 C_{rate} + u_4) - u_5 C_{rate} + u_6$	0.0311	0.0023	0.2768	0.1112	6
#10	$\frac{1}{SoC}(u_1 C_{rate} + u_2) + u_3 C_{rate} + C_{rate} + u_4$	0.0319	0.0025	0.1120	0.2845	4
#11	$\frac{1}{SoC}(u_1 C_{rate} + u_2) + C_{rate}(u_3 C_{rate} + u_4) + u_5 SoC + u_6 C_{rate} + u_7$	0.0331	0.0016	0.2733	0.1501	7
#12	$SoC\ C_{rate}(u_1 SoC + u_2) + (u_3 SoC + u_4)atan(C_{rate}) + C_{rate}(u_5 SoC + u_6) + u_7 SoC + u_8$	0.0376	0.0016	0.3632	~0	8
#13	$C_{rate}[SoC\ (u_1 SoC + u_2) + u_3 SoC + u_4] + SoC(u_5 SoC + u_6) + \frac{1}{SoC}u_7 + SoC + u_8$	0.0386	0.0014	0.3745	0.0003	8
#14	$ln(SoC)(u_1 SoC + u_2 C_{rate} + u_3) + u_4 SoC + u_5 C_{rate} + Soc + u_6$	0.0416	0.0022	0.2236	0.1749	6
#15	$C_{rate}[u_1 SoC + u_2 C_{rate} + u_3] + \frac{1}{SoC}(u_4 C_{rate} + u_5) + u_6 SoC + u_7 C_{rate} + u_8$	0.0428	0.0015	0.3291	0.4998	8
#16	$C_{rate}\left(u_1\frac{1}{SoC} + u_2 C_{rate} + u_3\right) + \frac{1}{SoC}(u_4 + C_{rate}) + u_5 + C_{rate}$	0.0475	0.0025	0.2178	0.3868	5

Figure 4.25: Expressions of best models obtained from GP.

FFIT, as well as the objective function values F_{cmp}, F_{RMS}, and F_{mnt}, are comprehensively listed in Figure 4.25.

It's essential to highlight that, at this stage, all u coefficients are purely numerical and are assessed through a nonlinear least square optimization method. For all GP-based models in question,Equation 4.2.2 presents a summary of statistical measures, including the mean (μ err), standard deviation (σ err), maximum (errmax), and RMS values of the absolute error computed for both the training dataset Ttr and the test dataset Tts.

Equation 4.2.2 employs a color gradient, ranging from blue to red, to emphasize the minimum and maximum values for each parameter. These models exhibit a maximum RMS error of the absolute error less than 9.5 mV for the training dataset and 9.8 mV for the test dataset. Some models are preferred over others within this group. For instance, models such as #1 and #2 exhibit the lowest F_{FIT} value (representing an optimal trade-off among the three objective functions) and have the added advantage of relying on just two numerical coefficients,u_1 and u_2. Both excel on the test dataset, where they achieve the minimum errmax value, an aspect not explicitly considered during the GP's evolutionary process.

Conversely, models with more than five u coefficients demonstrate strong metrics on the training dataset (e.g., models #5, #11, #12, #13, #14, and #16) and offer a reasonable compromise on the test dataset. Model 15, featuring eight numerical coefficients, performs exceptionally well on the training dataset but exhibits the highest values for standard deviation (σ err), maximum error (errmax), and RMS over the entire test dataset. Model 5 strikes a notable balance, as evidenced by the predominance of blue-colored cells inEquation 4.2.2.

Thus, we proceed to compare the following two models: - Model #1, which is the simplest with the fewest coefficients, boasts the best fitness value F_{FIT}, the highest RMS value (F_{RMS} = 0.003), a relatively low complexity value (F_{cmp} = 0.056), and minimal monotonicity factor (F_{mnt} = 0). - Model 5, the most complex with the maximum number of coefficients, achieves a higher fitness value FFIT, the lowest RMS value (F_{RMS} = 0.0013), the highest complexity

Model	Training dataset T_{tr}				Testing dataset T_{ts}			
	μ_{err} (mV)	σ_{err} (mV)	err_{max} (mV)	RMS (mV)	μ_{err} (mV)	σ_{err} (mV)	err_{max} (mV)	RMS (mV)
#1	7.0	6.4	43.5	9.5	6.3	5.3	21.2	8.2
#2	7.0	5.9	35.2	9.2	6.7	5.7	21.8	8.9
#3	6.5	5.9	37.1	8.8	6.1	7.2	36.3	9.5
#4	5.9	5.7	45.4	8.2	6.4	6.2	31.9	8.9
#5	2.8	3.8	38.9	4.7	3.8	5.0	25.4	6.3
#6	5.5	5.0	37.0	7.4	5.1	5.3	29.3	7.3
#7	6.5	5.9	36.1	8.8	6.0	7.3	35.2	9.5
#8	5.9	6.5	58.9	8.8	6.3	6.3	29.1	8.9
#9	5.6	4.9	34.4	7.4	5.2	5.7	32.3	7.7
#10	6.0	5.5	39.7	8.1	6.7	6.5	27.4	9.3
#11	3.9	3.3	19.3	5.1	4.8	7.6	48.0	9.0
#12	3.3	4.3	31.8	5.4	5.6	6.5	26.7	8.6
#13	3.4	3.8	26.4	5.1	5.2	7.6	42.7	9.2
#14	4.7	5.0	27.4	6.8	5.9	6.3	29.5	8.6
#15	3.5	3.0	19.6	4.6	4.9	8.5	54.2	9.8
#16	5.6	4.7	28.7	7.3	5.6	5.8	30.9	8.1
	max relative RMS error = 0.31% min relative RMS error = 0.15%				max relative RMS error = 0.36% min relative RMS error = 0.22%			

Figure 4.26: Metrics values for selected GP models

value (F_{cmp} = 0.389), and a moderately low level of monotonicity factor (F_{mnt} = 0.0008).

It's important to note that a lower monotonicity factor increases the likelihood of expressing the u coefficients as a function of temperature T. For instance, both models #6 and #16 might appear as compelling alternatives, featuring fewer coefficients than model #5 and superior error metrics compared to model 1. However, models #6 and #16 exhibit a relatively high level of monotonicity factor, which could impede the practical ability to represent the u coefficients as temperature-dependent functions.

Figure 4.27 and Figure 4.28 illustrate the absolute error (Ea) denoted as $V_{B,exp}V_{B,bhv}$, which represents the difference between the experimentally observed battery voltage and the corresponding voltage predicted by GP models #1 and #5. In this analysis, all experimental data points from both the training and validation datasets are considered (indicated by circle markers), as well as those from the test dataset (designated by star markers).

The results overall affirm the high level of accuracy achieved by behavioral model #1, particularly evident for intermediate SoC values and lower Crate values. The most significant error values occur at lower SoC levels, such as when model #1 records the highest error of errmax = 43.5 mV at Crate = 0.25, SoC = 20%, and T = 5°C.

Conversely, the reliability and strong accuracy of behavioral model #5 are evident for almost all combinations of SoC, Crate, and temperature. For instance, at Crate = 0.25, SoC = 20%, and T = 5°C, behavioral model #5 records the highest error of errmax = 38.9 mV at Crate = 1, SoC = 20%, and T = 15°C.

Both model #1 and model #5 consistently offer a high level of accuracy, with maximum relative RMS error values of 0.31% and 0.22%, respectively. These error levels align with the required accuracy standards.

Final Validation and discussion

The final section aims to validate the applicability and reliability of the GP-based models across the entire range of investigation for the SoC, Crate, and T parameters. For each specific temperature value (T = 5, 15, 25, 35 °C), the numerical coefficients u have already been determined by the GP algorithm through a nonlinear least square optimization method. This enables us

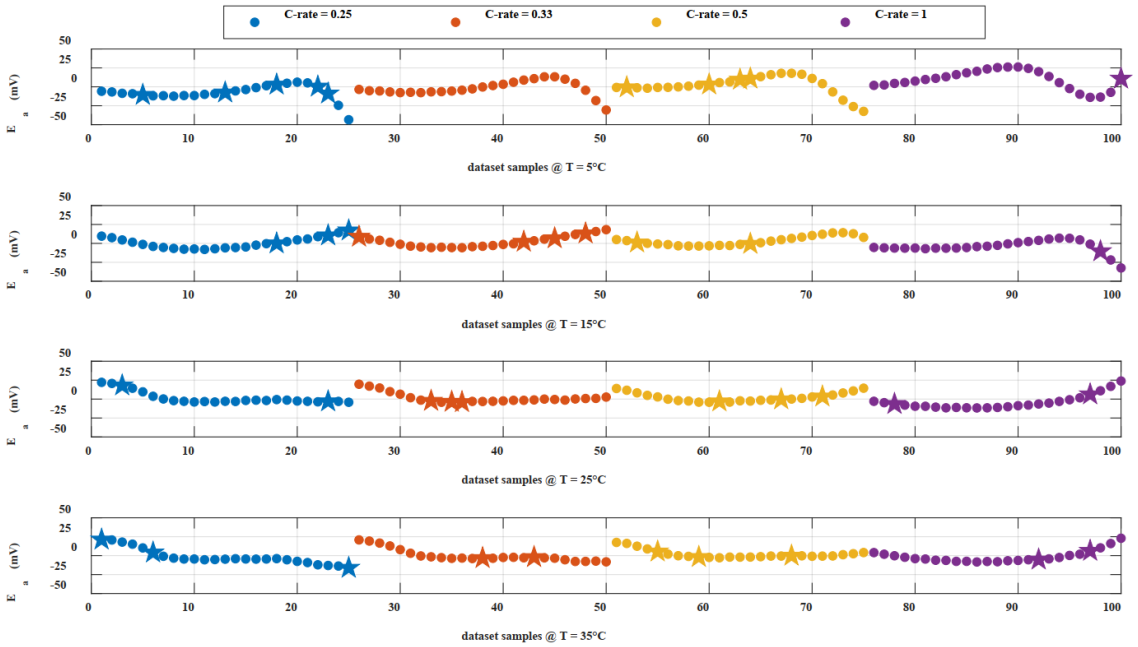


Figure 4.27: Absolute error between the experimental dataset and output battery voltage predicted by using the GP model #1: circle markers = training dataset samples, star markers = test dataset samples. For each temperature value and Crate subset (Crate=0.25, Crate=0.33, Crate=0.50, Crate=1), data samples are sorted from SoC = 80% to SoC = 20%

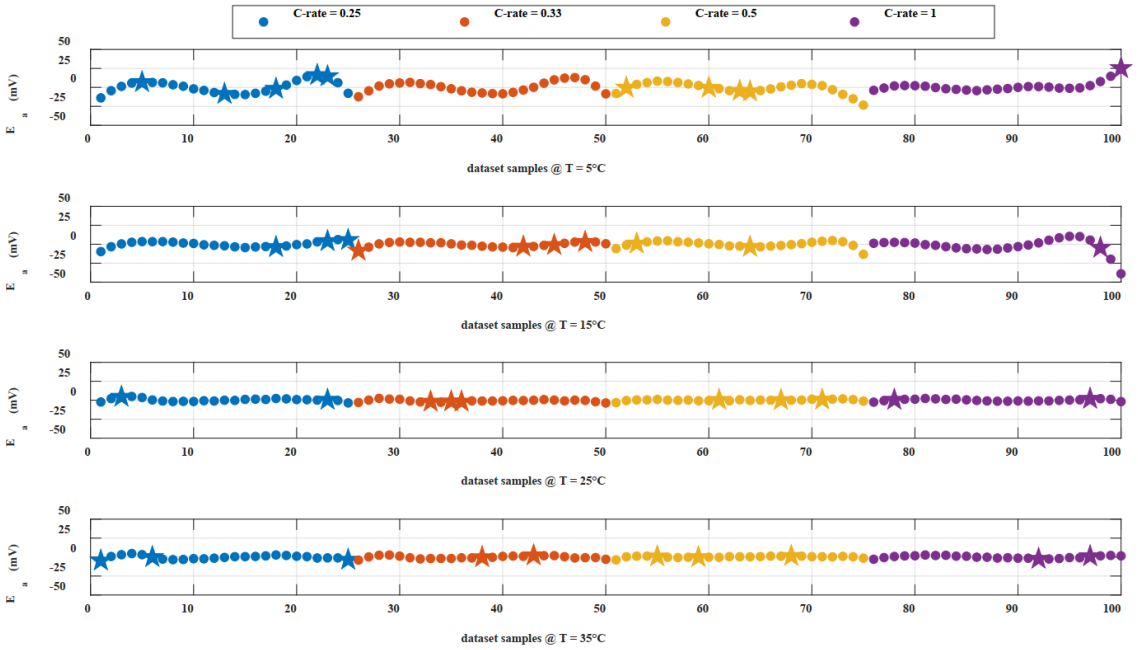


Figure 4.28: Absolute error between the experimental dataset and output battery voltage predicted by using the GP model #5: circle markers = training dataset samples, star markers = test dataset samples. For each temperature value and Crate subset (Crate=0.25, Crate=0.33, Crate=0.50, Crate=1), data samples are sorted from SoC = 80% to SoC = 20%.

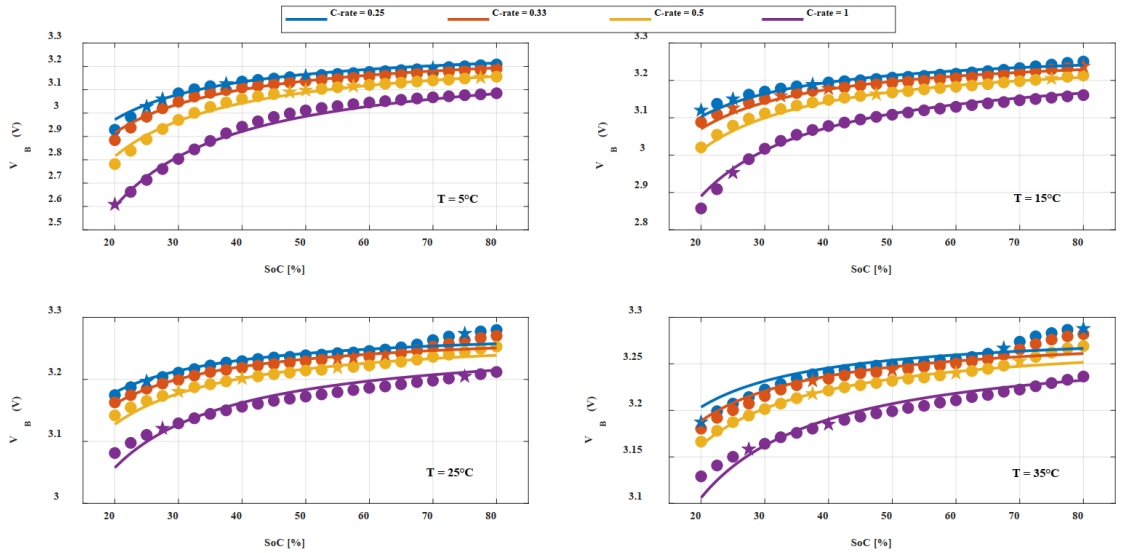


Figure 4.29: Output battery voltage experimental values (circle markers = training dataset samples, star markers = test dataset samples), and results based on model #1 with constant coefficients u (continuous lines).

to obtain the $V_{B,bhv}$ versus SoC characteristics for all possible SoC values ranging from 20% to 80%, in conjunction with the given Crate values of 0.25C, 0.33C, 0.5C, 1.0C.

In the earlier analysis, only 25 SoC values were considered for each Crate value, starting from SoC = 20% and progressing in increments of 2.5%. Figure 4.29 and Figure 4.30 illustrate the fitting of the output battery terminal voltage, represented by continuous lines corresponding to SoC values within the 20% to 80% range, as predicted by models #1 and #5. These predictions are compared against all the experimental data points adopted from the training and validation datasets (indicated by circle markers), as well as from the test dataset (designated by star markers).

Consequently, both models facilitate the prediction of the battery terminal voltage trends for any SoC value within the original dataset. In detail, model #1 accurately predicts the behavior of the battery terminal voltage for intermediate SoC values, with a more significant error observed for the lowest SoC values. Note that the voltage scales in Figure 5 have been adjusted to the curves for enhanced visualization. Conversely, model #5 effectively captures the expected trends of experimental data points for SoC, Crate, and temperature, displaying reliable performance.

In the final validation step of our proposed approach, we have confirmed the reliability of the GP-based models for C-rate and T values within the original definition range but not included in the GP dataset. To achieve this, we determined the trend of coefficients u with respect to temperature (T). The plots in Figure 4.31 display the values of coefficients u1 and u2 for T = 5, 15, 25, 35°C (indicated by red dots), as well as the coefficient trends u1(T) and u2(T) for model #1. These trends can be expressed as shown in Equation 4.12 for i = 1, 2), taking into account the numeric coefficients ci1, ..., ci4 listed in Figure 4.33.

$$U_i(T) = c_{i1}T^3 + C_{i2}T^2 + c_{i3}T + C_{i4} \tag{4.12}$$

Similarly, Figure 4.32 displays the coefficient values $u_1, \ldots, u_9$ for T = [5, 15, 25, 35] Celsius Degree (indicated by red dots), along with the coefficient trends $u_1(T), \ldots, u_9(T)$ for model #5. These coefficient trends can be expressed as in Equation 4.12 for $i = 1, \ldots, 9$, taking into account the numeric coefficients $c_{i1}, \ldots, c_{i4}$ listed in Figure 4.32.

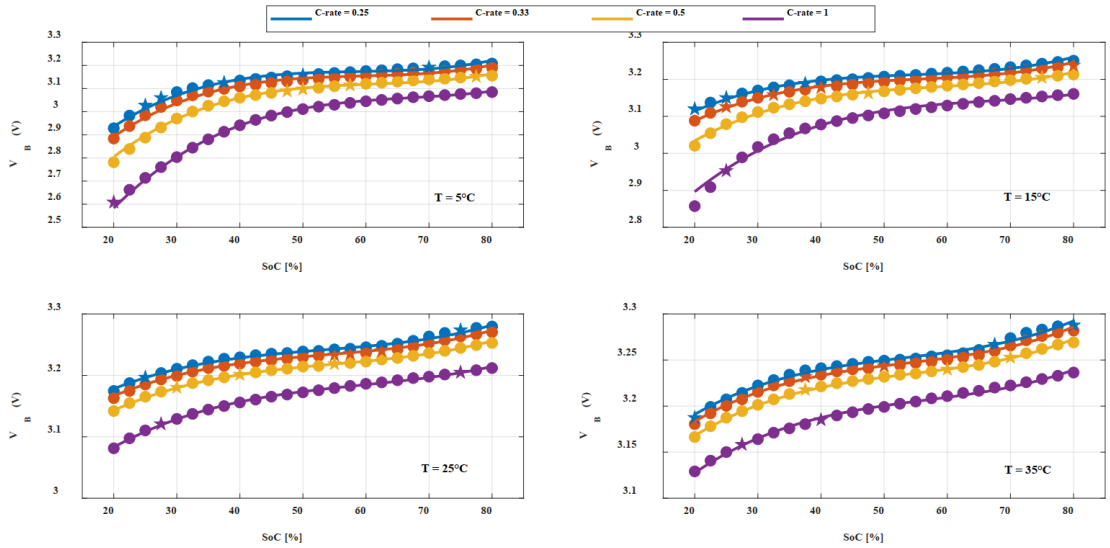


Figure 4.30: Output battery voltage experimental values (circle markers = training dataset samples, star markers = test dataset samples), and results based on model #5 with constant coefficients u (continuous lines).

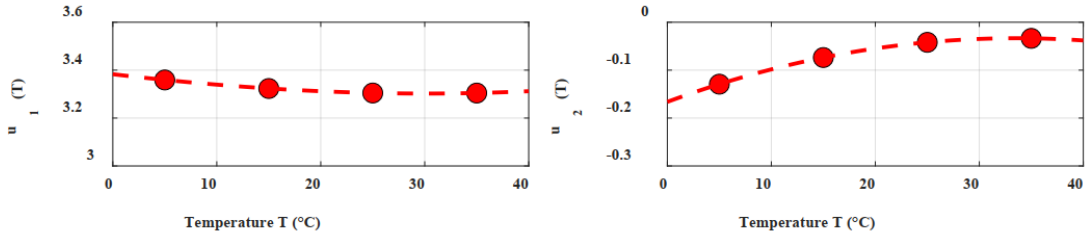


Figure 4.31: . Coefficients values u1 and u2 for T = 5, 15, 25, 35°C (red dots), and coefficients trend u1(T) and u2(T) for the GP model #1.

Consequently, both model #1 and model #5, along with their coefficient trends as described in Equation 4.12, enable the prediction of the battery terminal voltage for any SoC values within the range of 20% to 80%, Crate values spanning from 0.25 to 1, and temperature values between 5°C and 35°C. It's worth noting that certain metrics provided in Figure 4.35 may exhibit variations when model #1 and model #5 are used in conjunction with $u_i(T)$. For the original dataset, the resulting new metrics are presented in Table **??**. In this context, the new metrics closely resemble those obtained with purely numerical coefficients u identified by the GP through a non-linear least square algorithm, as detailed in

. This similarity primarily arises from the monotonic trend of the numerical coefficients u, which can be readily fitted with analytical functions.

As the ultimate step of validation, the same experimental configuration employed for prior datasets was used to conduct a fresh analysis of the 90-Ah LFP battery in discharge mode. Specifically, for a Crate of 0.6C and a temperature of 10°C, a total of 55,500 state of charge (SoC) data points, ranging from 20% to 80%, were gathered. Figure 4.36 illustrates the disparity between this newly acquired experimental dataset of battery voltage output and the corresponding battery voltage predictions made using GP model #1 (depicted by the blue

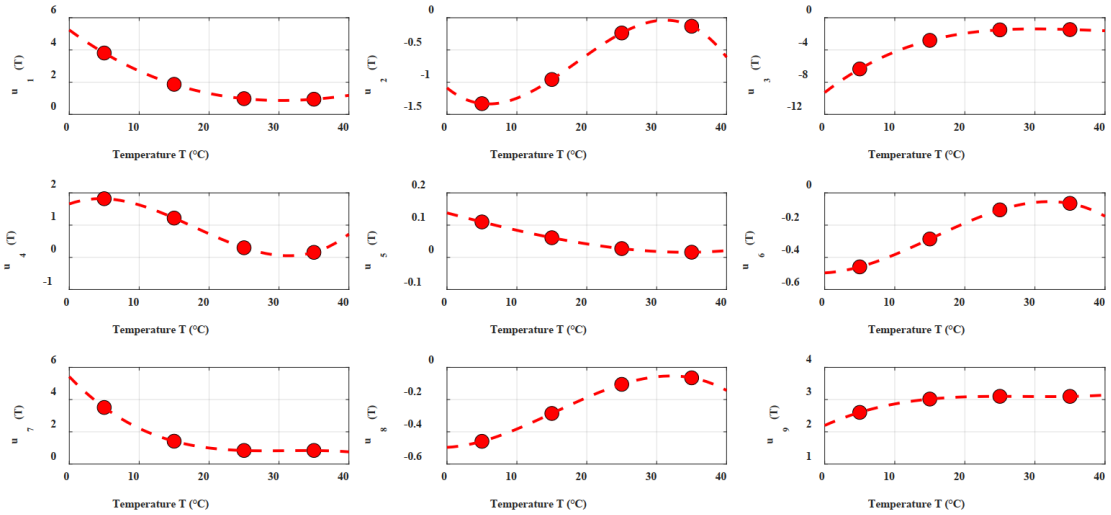


Figure 4.32: Coefficient values u1, …, u9 for T = 5, 15, 25, 35°C (red dots), and coefficient trend u1(T), …, u9(T) for the GP model #5

Table 6. Coefficient values for the behavioral model #1.

Coefficients for model #1	c_{i1}	c_{i2}	c_{i3}	c_{i4}
u_1	2.860e-07	7.169e-05	-0.510e-02	3.383
u_2	8.726e-08	-1.235e-04	8.004e-03	-1.662e-01

Figure 4.33: Coefficient values for the behavioral model #1

Coefficients for model #5	c_{i1}	c_{i2}	c_{i3}	c_{i4}
u_1	-3.360e-05	6.787e-03	-3.192e-01	5.241
u_2	-1.606e-04	8.963e-03	-8.965e-02	-1.088
u_3	1.638e-04	-1.861e-02	6.740e-01	-9.282
u_4	1.829e-04	-9.828e-03	7.706e-02	1.657
u_5	1.196e-06	2.242e-05	-5.736e-03	1.379e-01
u_6	-2.469e-05	1.149e-03	2.335e-03	-4.964e-01
u_7	-1.573e-04	1.467e-02	-4.518e-01	5.423
u_8	-2.466e-05	1.148e-03	2.365e-03	-4.965e-01
u_9	4.038e-05	-3.461e-03	9.741e-02	2.197

Figure 4.34: Coefficient values for the behavioral model #5

Coefficients for model #5	c_{i1}	c_{i2}	c_{i3}	c_{i4}
u_1	-3.360e-05	6.787e-03	-3.192e-01	5.241
u_2	-1.606e-04	8.963e-03	-8.965e-02	-1.088
u_3	1.638e-04	-1.861e-02	6.740e-01	-9.282
u_4	1.829e-04	-9.828e-03	7.706e-02	1.657
u_5	1.196e-06	2.242e-05	-5.736e-03	1.379e-01
u_6	-2.469e-05	1.149e-03	2.335e-03	-4.964e-01
u_7	-1.573e-04	1.467e-02	-4.518e-01	5.423
u_8	-2.466e-05	1.148e-03	2.365e-03	-4.965e-01
u_9	4.038e-05	-3.461e-03	9.741e-02	2.197

Figure 4.35: Coefficient values for the behavioral model #5

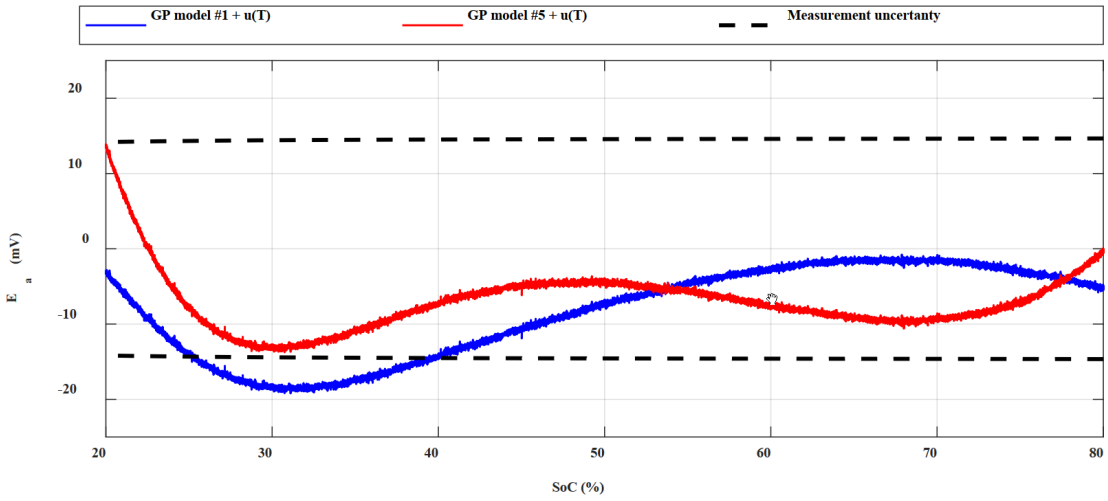


Figure 4.36: Absolute error between the experimental output battery voltage for Crate = 0.6C and T = 10°C and the output battery voltage predicted by using model #1 (blue line) and model #5 (red line) along with coefficients u(T)

line) and GP model #5 (represented by the red line), along with their coefficient trends concerning temperature (denoted as "u(T)"). The dashed black lines indicate the measurement uncertainty calculated with a confidence level of approximately 98%.

In summary, model #5 predicts the battery's output voltage with a maximum absolute error of less than $15mV$, falling well within the specified measurement uncertainty for a Crate of 0.6C and a temperature of 10°C, across all SoC values ranging from 20% to 80%. In contrast, the absolute error of model #1 does not consistently fall within the designated measurement uncertainty but still offers a fairly accurate prediction of the battery's output voltage, with a maximum absolute error of less than $20mV$.

Despite the fact that the two GP models analyzed here have been arranged in ascending order of complexity, it is essential to note that both are analytical models. Therefore, in a direct problem scenario, their cost would be the same, for instance, when implementing them into a Battery Management System (BMS) to forecast output voltage. In direct problems, complexity can be heightened, and constraints on the coefficients can be relaxed, especially if the ultimate goal is to enhance accuracy. However, a significant advantage of an analytical representation lies in its potential for reverse usage in inverse problems, such as optimization

or uncertainty analysis during the design phase.It is evident that, for these alternate purposes, having simpler analytical models with monotonous coefficients holds substantial added value.

Chapter 5

Case Study: BATMAN

5.1 Platform description

The aim of the project is to realize a distributed, cost-effective, and intelligent platform based on IoT solutions for the predictive diagnosis and thermal management of batteries involved in safety-critical and mission-critical applications. In particular, the distributed platform will be used for monitoring Uninterruptible Power Supply (UPS) in HV/MV and MV/LV transformer substations. In order to meet the functional requirements of the project, the platform to be implemented will consist of three basic parts:

- A local sensing system to be applied to the individual batteries being monitored.
- One or more data collection gateways with instrument fault detection capabilities.
- A processing and supervision system with data fusion, predictive diagnostics, and benchmarking capabilities.

The sensing system will enable continuous sensing of quantities of interest for battery monitoring, including electrical, electrochemical, and environmental parameters that can provide detailed information on the state of the batteries. The data collection gateways will consist of devices that will have multiple functions including:

- Acquisition and pre-filtering of data from the sensors.
- Local data processing.
- Reliable data transmission to the subsequent processing and supervision system.

The processing and supervision system with data fusion, predictive diagnostics, and benchmarking capabilities will aim to collect data from the field and process it, using artificial intelligence algorithms, in order to estimate the state of the batteries and identify possible anomalies that might require battery maintenance.

5.2 Sensing Platform

The characteristics of the local measurement system to be applied to individual batteries (or to a limited number of batteries) within UPSs, are defined according to the quantities to be measured. In particular, it is necessary, in order to estimate the status of the batteries in UPSs and identify possible anomalies, to continuously monitor the quantities indicated below.

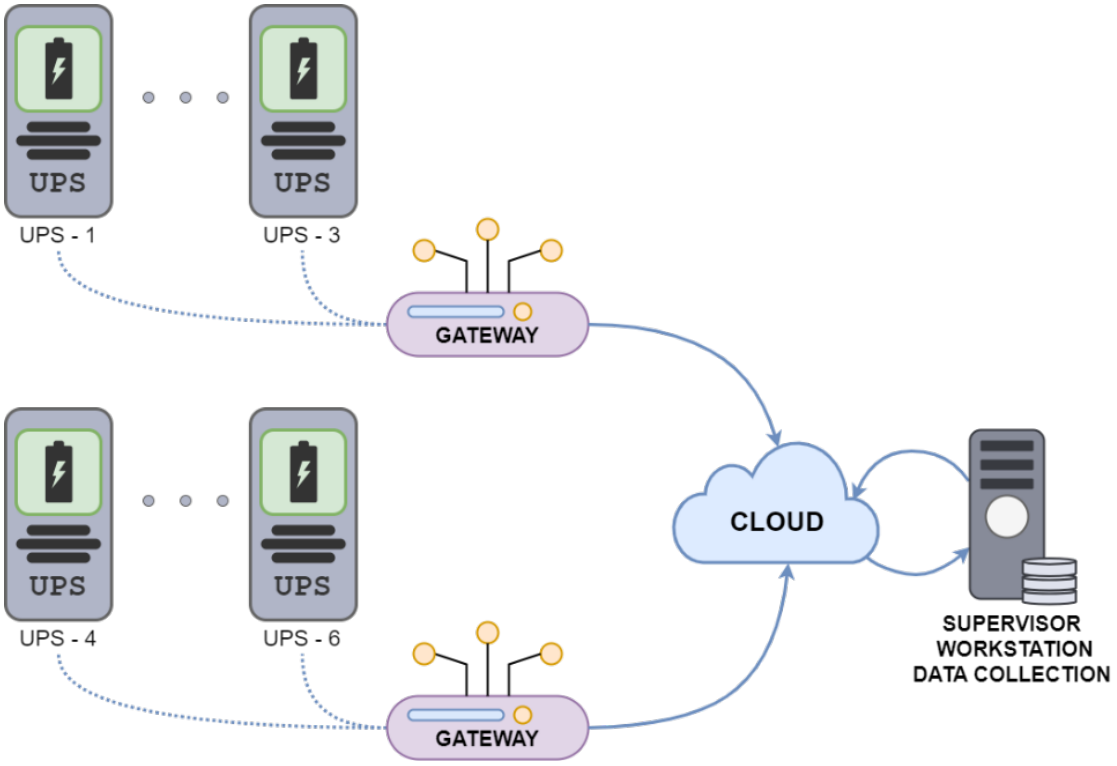


Figure 5.1: Structure of the acquisition, transmission, processing, and archiving system to be implemented.

- The continuous current delivered (absorbed) by the batteries during the discharge (charging) process.
- The continuous voltage of the individual batteries.
- The impedance of the individual batteries, estimated from AC voltage and AC current measurements.
- The temperature of the individual batteries.

Furthermore, in order to provide the system with greater diagnostic capabilities, it is planned to measure the power flows at the input and output of UPSs in order to have a global view of the system status, which is essential for performing planned and predictive maintenance of the system. Figure 5.2 shows the logic structure and all the measured parameters by the Board.

5.2.1 Device Description

Stimulus Signal

The generation and injection of a sinusoidal signal are achieved through a function generator, taking as reference Figure 5.3, U1 (AD9833), which can be remotely programmed via the SPI communication port. The signal is then converted into a current signal using a power amplifier, OPA544, configured as a voltage-driven current source. Its operational principle, using negative feedback, ensures that the desired current is achieved when the voltage drop

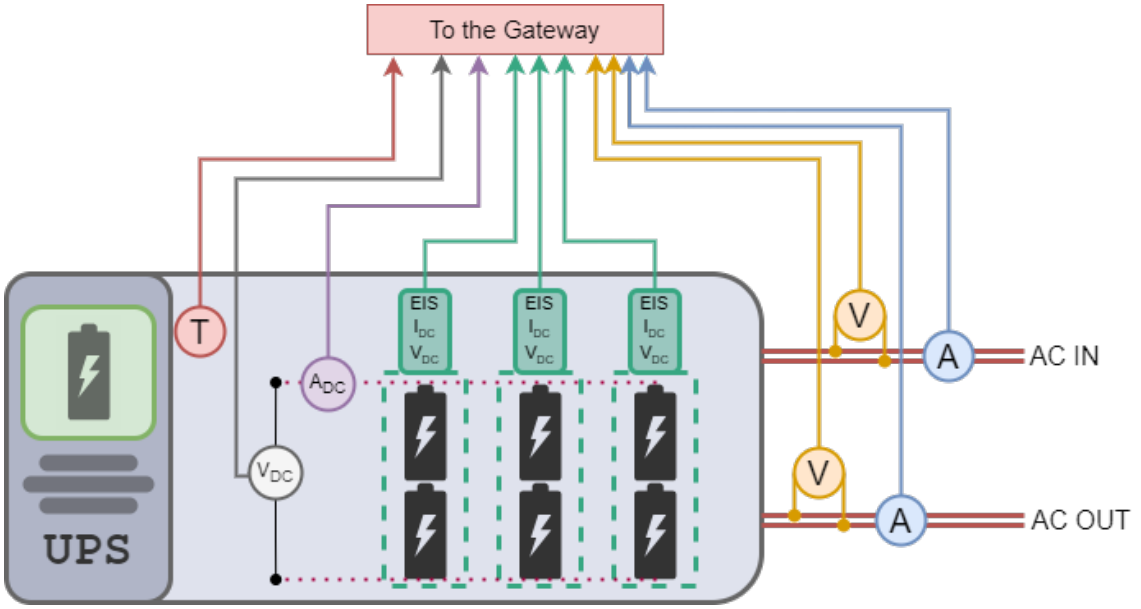


Figure 5.2: Logical structure of the measurement system within the UPS.

across the measurement resistor RS_G1 matches the voltage set at the non-inverting input of the amplifier.

The function generator, U1 (AD9833), receives frequency setting instructions from the ESP32 microcontroller (U17) via the MOSI channel of SPI, while selecting the chip using the CS2- (chip select) pin. The output signal, OUT_G, has a fixed amplitude ranging between $Vmin = 38mV$ and $Vmax = 650mV$. To make this signal have zero DC offset, a precision voltage regulator VR1 (TC1087) and operational amplifier U3B in an inverting configuration are used to subtract an offset voltage, $Voff = 344mV$. U3B subtracts Voff from the function generator signal to achieve a zero DC offset signal.

The device U2 (TMUX7219) serves as an analog signal switch and can be remotely controlled with a digital signal (1 or 0). This switch is employed to connect the input of the next stage to the output of amplifier U3B or to ground through R6. The latter configuration allows for a null signal at the input of amplifier U3A, effectively cancelling the test current in the battery, thereby enabling the measurement of the open-circuit voltage (OCV).

Amplifier U3A is configured as an inverting amplifier. This configuration allows for an adjustable gain via the digital potentiometer RD1B, with gain values both greater than and less than 1 being possible.

AC Voltage

The battery is connected to the measuring circuit with a 4-terminal configuration to limit the influence of contact resistances and the measuring circuit. Taking the region 1 in Figure 5.4 as reference, the Battery DC voltage is measured using operational amplifier U10A in a differential configuration with a unity gain. This configuration is employed to eliminate any potential offsets caused by the voltage drop across the measurement resistor RS_G1 when it carries the test current. The output signal from U10A is then passed through a voltage divider of 1/11, consisting of R46 and R49. This voltage divider is designed to generate a voltage within the range of 0 to 3.3V, corresponding to the operational voltage range of the sampler. The output signal from the voltage divider passes through a circuit with zener diodes for

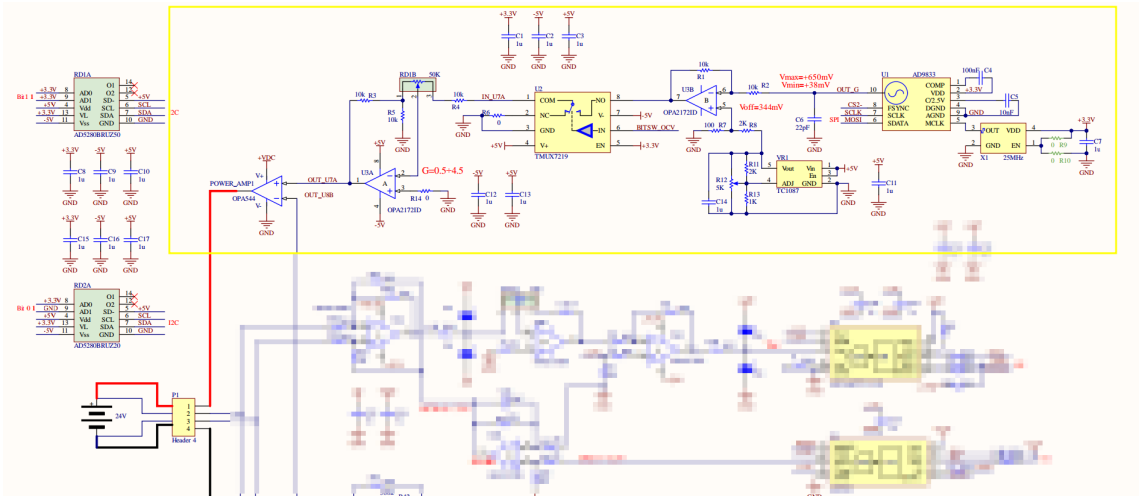


Figure 5.3: Signal generation section.

protection and is subsequently sampled at 16 bits by the device U11, an MCP33131D. This operation prepares the signal to be transmitted via the SPI bus to the microcontroller. The transmission channel used is the SDO, which is connected to the MISO pin of the microprocessor U17, an ESP32. Furthermore, the microcontroller selects the target device using the "chip select" signal CNVST2.

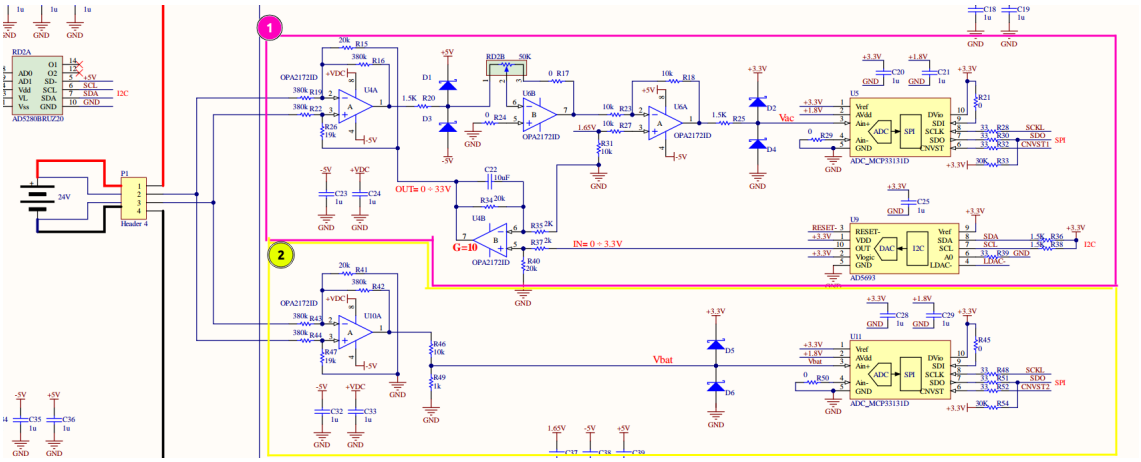


Figure 5.4: DC and AC Voltage measuring section

DC Voltage

The zero mean signal, coming from the output of U4A, as shown into the region 2 of Figure 5.4, undergoes appropriate amplification through the variable-gain amplifier U6B, controlled by the digital potentiometer RD2B. Subsequently, the signal is shifted by $1.65V$ through op-amp U6A and voltage regulator U15. The output from U6A is then sampled by the ADC converter U5 and transferred to the microcontroller via the SPI interface.

AC Current The current is measured via the measuring resistor RS_G1, whose voltage drop

is read by the op-amp U10B. The resulting signal is subsequently amplified using an amplifier

with variable gain, U14B, which is adjusted via digital potentiometer RD3B. In addition, the signal is shifted by 1.65V via the operational U14A. The output signal from the latter op-amp is sampled by U13 and then transmitted to the microcontroller via the SPI interface.

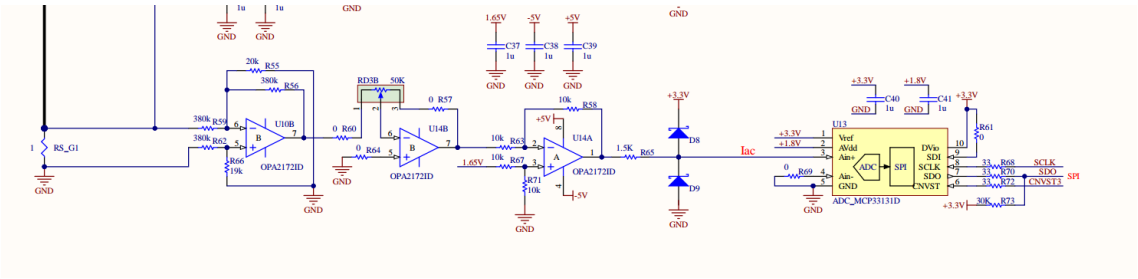


Figure 5.5: Current measure section

5.3 Gateway Platform

The gateway system plays an essential role by acting as an intermediate interface between the sensing device and the supervisor. Its presence is of fundamental importance since, if the entire system consisted only of the sensing device and the supervisor, the latter would have to possess too much computing power to be compatible with current systems. However, thanks to the intermediate gateway, it is possible to shift part of the computational load and the management of certain devices to the sensing side, rather than having everything be delegated to the supervisor system. This approach allows easier scalability of the whole system, as the gateway can be optimized to handle specific functions and workloads, allowing the supervisor to concentrate on high-level tasks. Furthermore, considering the community's orientation towards systems with high repairability and long life cycles, the use of an intermediate gateway simplifies replacement in the event of obsolescence caused by an increase in the number of quantities measured by the sensing system. In summary, the introduction of an intermediate gateway between the sensing device and the supervisor optimizes the overall efficiency of the system, reducing the load on the supervisor and allowing easier adaptability and replacement of components, aligning with the objectives of robustness and long life of today's systems.

The gateway plays a crucial role in the collection and processing of data from UPSs (Uninterruptible Power Supply). Each UPS transmits a series of parameters that are essential for the monitoring and effective operation of the overall system. Among the data collected are:

- V_{Sac} and I_{Sac}: these values represent the stimulus voltages and currents used to calculate the battery impedance at specific frequencies. By monitoring these parameters, the gateway can assess the stability and efficiency of the power system.
- V_{dc}: This value indicates the individual battery voltage, providing valuable information on the charge level and operational status of the batteries.
- I_{1i}: The current drawn by the individual UPS represents the electrical load that the system is handling at the input. Monitoring this current helps to optimize UPS utilization and prevent overloads.
- I_{1o}: This value indicates the current supplied by the individual UPS to the connected appliances and equipment. By carefully monitoring this current, the gateway can verify the appropriate power delivery and identify any abnormal situations.

- I_{bat}: This is the total continuous current delivered by the battery pack, in practice it is the current that powers the subsequent DC/AC converter (inverter).
- V_{1i}: The input voltage of the individual UPS represents the quality of the mains power supply. An accurate control of this parameter allows voltage fluctuations and problems on the mains to be detected.
- V_{1o}: This value indicates the voltage generated by the individual UPS to the connected devices. Accurate monitoring of this voltage is essential to ensure a stable and secure power supply to the devices.
- V_{Bat}: this is the total voltage of the battery module, knowing the connection of the individual cells within the UPS the Vbat value could be calculated as the composition of the individual voltages, the advantage of also having the total measurement however allows a controlled and more robust measurement system to be created even at the fault of a sensor.

In addition to collecting data from UPSs, the gateway performs another important task: it also collects data from temperature sensors. Monitoring temperature is crucial for preventing overheating and maintaining optimum operating conditions for the entire system, especially in those systems, such as UPSs, that use batteries. Through a combination of data from UPSs and temperature sensors, the gateway becomes a key element for advanced monitoring and control of the electrical infrastructure, helping to ensure business continuity and security of connected devices and equipment.

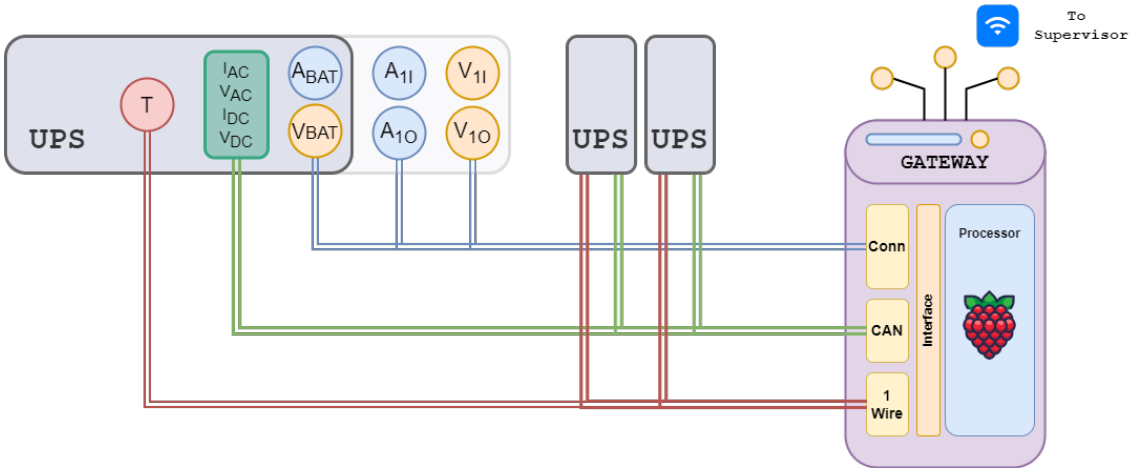


Figure 5.6: Sensing Board and Gateway connection Structure

The microcontroller section is implemented using an ESP32, which handles the measurement of various analog channels for voltage and current on the UPS board. It also implements the CAN Bus for data collection from various measurement module boards. Additionally, the ESP32 communicates via the SPI bus with a Raspberry Pi 4 for data collection, potential processing, and transfer over the internet.

Figure 5.8 illustrates the detailed circuitry for measuring the mains voltage using an isolated operational amplifier. In addition to the input divider comprising R1, R2, R4, R7, and R9, there is the isolated operational amplifier TLP7920, which is powered from the mains side using an isolated DC-DC converter. This amplifier has two outputs referenced to the ground, which connect to two analog inputs on the microcontroller. These outputs have an offset to enable measurement of the negative portion of the signal. Software processing can be applied

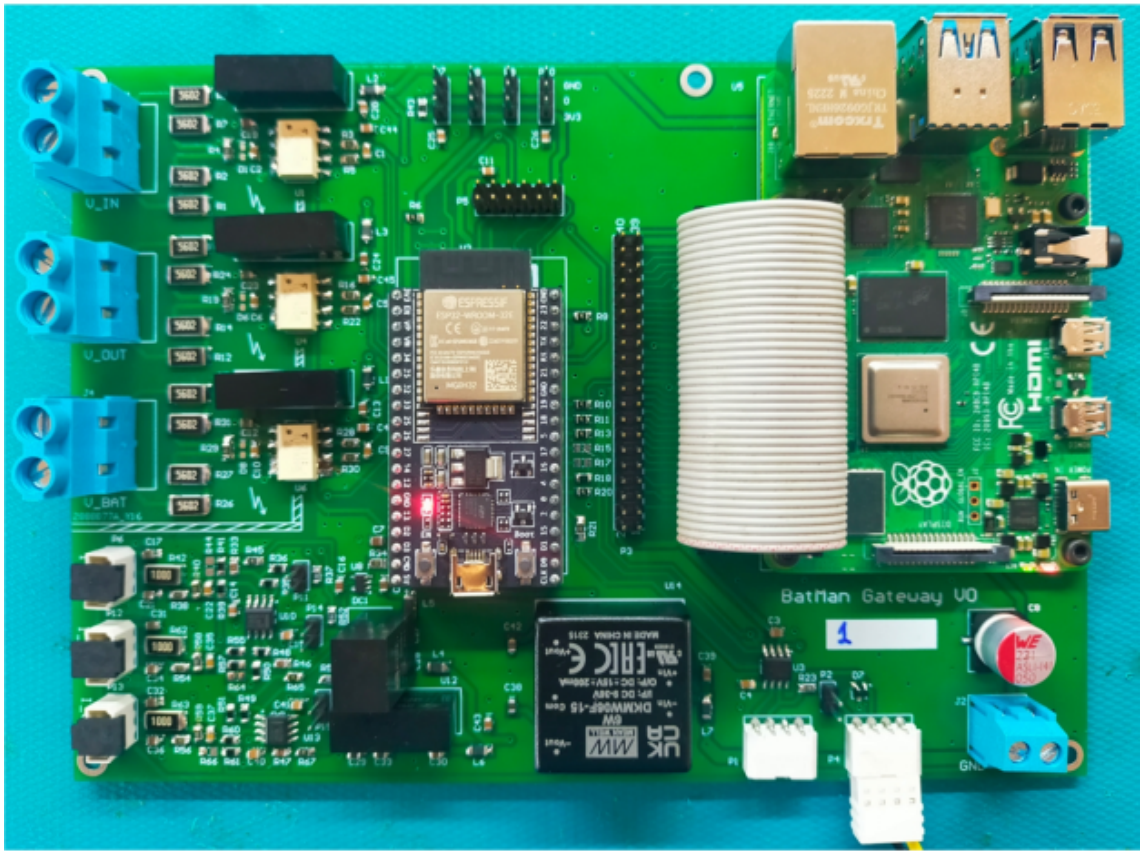


Figure 5.7: Gateway Board assembled

to calculate the difference, thereby reducing common-mode error. Three identical channels are present on the board.

Current measurements, see Figure 5.9, on the mains and UPS input/output are carried out using the LEM sensor, model LEM LA 55-P, which has a 50 A RMS rating. As shown in Figure 3, the LEM sensor is powered by a dual 15 V voltage supply and loaded by the 100 Ω measurement resistor, R42. Subsequently, there is a voltage divider consisting of 1 $k\Omega$ and 300 Ω resistors. The filtered output of this divider, through C22, is applied to a differential stage that introduces a voltage offset of 1.65V with a unity gain.

Figure 5.10 presents the schematic section related to interfacing with the "one-wire" DS18B20 sensors powered at 3.3 V.

5.3.1 Project Status

The system to be created is configured as an additional device based on a distributed architecture which has the purpose of acquiring "online" information useful for verifying in real time the functioning of the accumulators and estimating, on the basis of regression and classification techniques based on ML and on historical series also relating to previously occurred events, the state of health and the residual life of the accumulator and to activate any maintenance policies based on predictive analysis. The project is divided into the following points:

1. Definition of system requirements
2. Experimental characterization in a laboratory environment

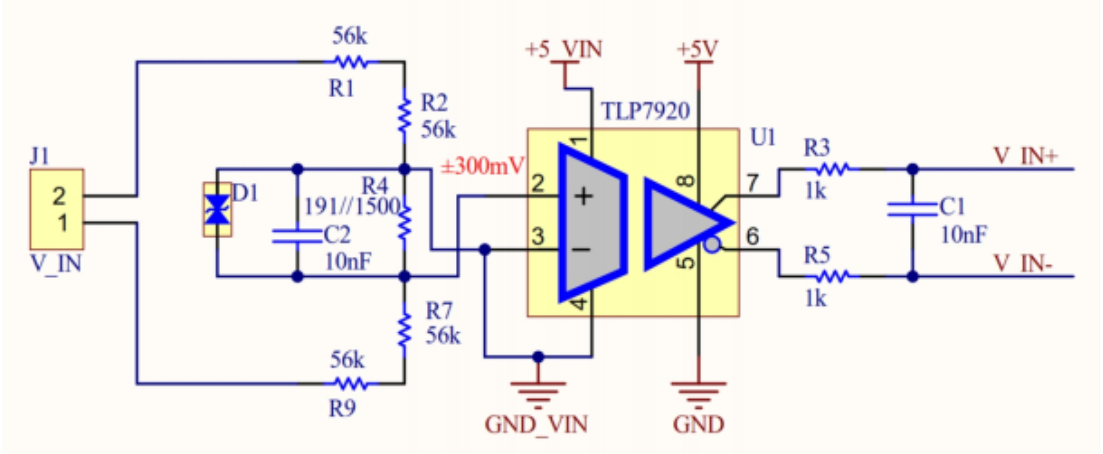


Figure 5.8: Grid Voltage measurement section

3. Software development
4. Construction and validation of the demonstrator
5. Implementation and acquisition of measurement data
6. Testing of artificial intelligence algorithms to identify possible trends toward failure situations.

At present, the *TRL*7 prototypes have been built and validated (point 4) and their implementation is underway in order to collect an adequate number of data.

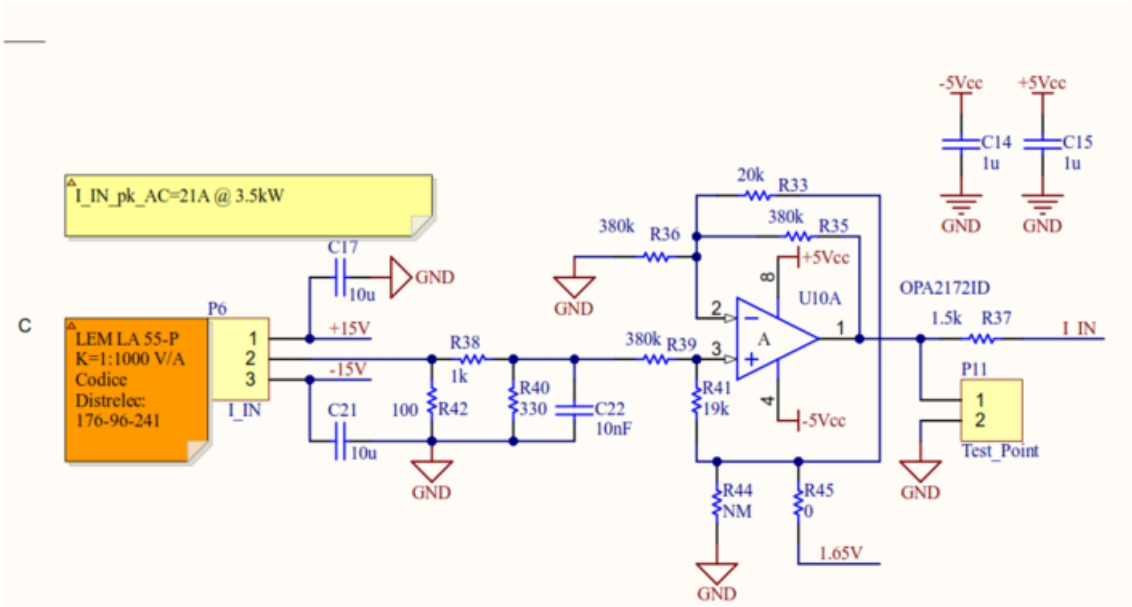


Figure 5.9: UPS input and output current measurement section

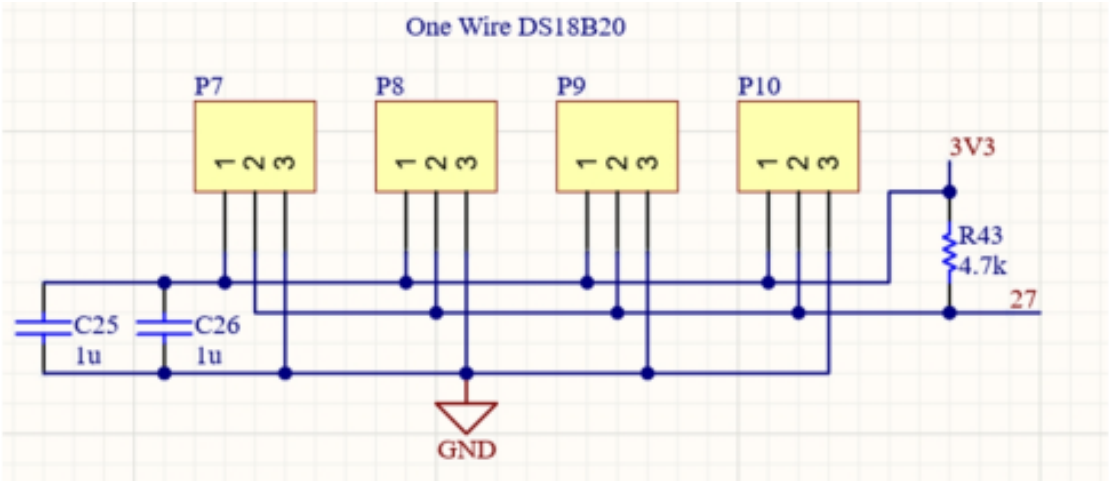


Figure 5.10: Temperature measurement section

Conclusions

This thesis work is conducted in the realm of predictive maintenance, where over time, we have progressed from a failure-driven maintenance approach, where maintenance is performed based on either incipient failures, or a time-based maintenance approach, where the maintenance is predetermined at time intervals, to a condition-based maintenance approach. Condition-based maintenance involves conducting maintenance based on the actual wear and tear of the monitored system and has assumed such significance that it rightfully holds a place among the themes of Industry 4.0. Predictive maintenance necessitates not only knowledge of the current state of a device but also its evolution over time. In the case of batteries or battery-powered devices, determining both the current state and the future state can be challenging. In both cases, knowledge of the state can only be derived from measuring the system's characteristic parameters and appropriate modeling them.

Characteristic parameters in batteries include internal parameters such as battery voltage, delivered or absorbed current, ambient and cell temperature, and impedance measured at various stimulus frequencies, as well as external or derived parameters like state of charge and state of health.

Regarding the data collection challenge, this thesis primarily focuses on impedance measurement of batteries, specifically impedance spectroscopy, which involves measuring impedance variations across different stimulus frequencies. Various battery stimulus signals, such as stepped sine or multisine, were analyzed in relation to acquisition system settings and data processing algorithms. It was observed that for the low frequencies used in battery stimulation, algorithms based on the Fast Fourier Transform significantly impact measurement times, as they require at least one full period to avoid the "scallop loss" issue. To address this, sine fitting-based algorithms, adaptable to sub-period acquisitions, were considered, allowing time savings but potentially sacrificing accuracy.

The experience gained in battery impedance measurement systems was applied to the BATMAN platform, a distributed monitoring system for the state of batteries within Uninterruptible Power Supplies (UPSs). This multisensory measurement device actively stimulates batteries and communicates all necessary information to a remote server to create a digital model of the system. Throughout the doctoral experience, several setups were established to acquire impedance measurement datasets under varying battery state of charge, which were subsequently shared with the scientific community. This data was also used to train machine learning algorithms to estimate battery state of charge based on impedance spectroscopy measurements. A methodology for Battery Management System (BMS) designers was created to determine the ideal stimulus frequency set for classification performance optimization while minimizing measurement time. An evolutionary algorithm was developed, incorporating a classifier within the fitness function. The results achieved reduced measurement time by over 90% while maintaining classification performance equal to or exceeding that obtained with all stimulus frequencies. The system was also observed to simplify the classifier's task by eliminating misleading features.

Additionally, within the realm of battery diagnostics measurement systems, a preliminary

study, in collaboration with the University of Newcastle, was conducted to estimate state of charge through non-invasive impedance measurements using a coil attached to one side of the battery. The innovation lies in the ability to estimate state of charge from a measurement that is galvanically separated from the battery, i.e., without direct contact with its terminals.

Modeling a system is fundamental to prognosis in predictive maintenance. In this thesis work, the focus was on predicting battery voltage using two methods. The first method employs established techniques in the literature based on the current and voltage characteristics during a pulsed battery discharge procedure, enabling the identification of equivalent circuit parameters for subsequent use in a battery digital twin capable of predicting terminal voltage based on absorbed current and battery state of charge. The second method utilizes an innovative system based on genetic programming, utilizing battery voltages and currents during a constant current discharge procedure. The genetic algorithm yielded an analytical function that links the State of Charge, discharge current, and battery temperature. The analytical function is particularly advantageous as it can be easily implemented in cost-effective systems and, in this specific case, can be forced to be invertible, allowing State of Charge to be obtained from voltage, current, and temperature measurements.

In conclusion, the accomplished studies have contributed to various aspects of predictive maintenance for batteries, providing methodologies and metrological tools for the accurate estimation of fundamental battery parameters.

Bibliography

[1] VRLA Batteries Introduction, January 2010.

[2] Mohammad Abareshi, Erfan Sadeghi, Mohsen Hamzeh, Mehrdad Saif, and Seyed Mohammad Mahdi Alavi. Multi-purpose controllable electrochemical impedance spectroscopy using bidirectional dc–dc converter. *Journal of Energy Storage*, 55:105750, 2022.

[3] A. Affanni, A. Bellini, G. Franceschini, P. Guglielmi, and C. Tassoni. Battery choice and management for new-generation electric vehicles. *IEEE Transactions on Industrial Electronics*, 52(5):1343–1349, 2005.

[4] Kofi Afrifa Agyeman, Sekyung Han, and Soohee Han. Real-time recognition non-intrusive electrical appliance monitoring algorithm for a residential building energy management system. *Energies*, 8(9):9029–9048, 2015.

[5] Parisa Amiribavandpour, Weixiang Shen, and Ajay Kapoor. Development of thermal-electrochemical model for lithium ion 18650 battery packs in electric vehicles. In *2013 IEEE Vehicle Power and Propulsion Conference (VPPC)*, pages 1–5. IEEE, 2013.

[6] D. Andre, M. Meiler, K. Steiner, Ch. Wimmer, T. Soczka-Guth, and D.U. Sauer. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation. *Journal of Power Sources*, 196(12):5334–5341, June 2011.

[7] Kélig Aujogue, Alban Pothérat, Ian Bates, François Debray, and Binod Sreenivasan. Little earth experiment: An instrument to model planetary cores. *Review of Scientific Instruments*, 87(8):084502, 2016.

[8] Iman Babaeiyazdi, Afshin Rezaei-Zare, and Shahab Shokrzadeh. State of charge prediction of ev li-ion batteries using eis: A machine learning approach. *Energy*, 223:120116, 2021.

[9] Stefano Barsali and Massimo Ceraolo. Dynamical models of lead-acid batteries: Implementation issues. *IEEE Transactions on energy conversion*, 17(1):16–23, 2002.

[10] Evgenij Barsoukov and J. Ross Macdonald, editors. *Impedance Spectroscopy: Theory, Experiment, and Applications*. Wiley, 1 edition, apr 2005.

[11] Nipun Batra, Manoj Gulati, Amarjeet Singh, and Mani B Srivastava. It's different: Insights into home energy consumption in india. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, pages 1–8, 2013.

[12] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM conference on embedded systems for energy-efficient buildings*, pages 80–89, 2014.

[13] G Berrettoni, C Bourelly, D Capriglione, L Ferrigno, and G Miele. Preliminary sensitivity analysis of combinatorial optimization (co) for nilm applications: Effect of the meter accuracy. In *2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI)*, pages 486–490. IEEE, 2021.

[14] Bernard A Boukamp. A linear kronig-kramers transform test for immittance data validation. *Journal of the electrochemical society*, 142(6):1885, 1995.

[15] C. Bourelly, M. Vitelli, F. Milano, M. Molinara, F. Fontanella, and L. Ferrigno. GA-Based Features Selection for Electro-chemical Impedance Spectroscopy on Lithium Iron Phosphate Batteries. In *2023 IEEE International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC)*, pages 1–6, Venice, Italy, March 2023. IEEE.

[16] C. Bourelly, M. Vitelli, F. Milano, M. Molinara, F. Fontanella, and L. Ferrigno. Ga-based features selection for electro-chemical impedance spectroscopy on lithium iron phosphate batteries. In *2023 IEEE International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC)*, pages 1–6, 2023.

[17] Carmine Bourelly, Michele Vitelli, Filippo Milano, Mario Molinara, Francesco Fontanella, and Luigi Ferrigno. Eis-based soc estimation: A novel measurement method for optimizing accuracy and measurement time. *IEEE Access*, 2023.

[18] Carmine Bourelly, Michele Vitelli, Filippo Milano, Mario Molinara, Francesco Fontanella, and Luigi Ferrigno. Eis-based soc estimation: A novel measurement method for optimizing accuracy and measurement time. *IEEE Access*, 11:91472–91484, 2023.

[19] Emanuele Buchicchio, Alessio De Angelis, Francesco Santoni, Paolo Carbone, Francesco Bianconi, and Fabrizio Smeraldi. Dataset on broadband electrochemical impedance spectroscopy of lithium-ion batteries for different values of the state-of-charge. *Data in Brief*, 45:108589, 2022.

[20] Nadezda Buneeva and Andreas Reinhardt. Ambal: Realistic load signature generation for load disaggregation performance evaluation. In *2017 ieee international conference on smart grid communications (smartgridcomm)*, pages 443–448. IEEE, 2017.

[21] CE. Directive 2006/32/EC of the European Parliament And of the Council. `https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32006L0032&from=LV`, 2022. Online; Accessed:14.03.2022.

[22] CE. Directive 2009/72/EC of the European Parliament And of the Council. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0072&from=EN`, 2022. Online; Accessed:14.03.2022.

[23] CE. Directive 2012/27/CE of the European Parliament And of the Council. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=IT`, 2022. Online; Accessed:14.03.2022.

[24] Hsueh-Hsien Chang. Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. *Energies*, 5(11):4569–4589, 2012.

[25] Dong Chen, David Irwin, and Prashant Shenoy. Smartsim: A device-accurate smart home simulator for energy analytics. In *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 686–692. IEEE, 2016.

[26] Min Chen and G.A. Rincon-Mora. Accurate electrical battery model capable of predicting runtime and I-V performance. *IEEE Transactions on Energy Conversion*, 21(2):504–511, June 2006.

[27] VG Choudhari, AS Dhoble, and TM Sathe. A review on effect of heat generation and various thermal management systems for lithium ion battery used for electric vehicle. *Journal of Energy Storage*, 32:101729, 2020.

[28] Jon P Christophersen. Battery test manual for electric vehicles, revision 3. Technical report, Idaho National Lab.(INL), Idaho Falls, ID (United States), 2015.

[29] COP26. COP26: The Glasgow Climate Pact. `https://ukcop26.org/wp-content/uploads/2021/11/COP26-Presidency-Outcomes-The-Climate-Pact.pdf`, 2022. Online; Accessed:10.03.2022.

[30] Carlos M Costa, João C Barbosa, Renato Gonçalves, Helder Castro, FJ Del Campo, and Senentxu Lanceros-Méndez. Recycling and environmental issues of lithium-ion batteries: Advances, challenges and opportunities. *Energy Storage Materials*, 37:433–465, 2021.

[31] Utpal Kumar Das, Prashant Shrivastava, Kok Soon Tey, Mohd Yamani Idna Bin Idris, Saad Mekhilef, Elmira Jamei, Mehdi Seyedmahmoudian, and Alex Stojcevski. Advancement of lithium-ion battery cells voltage equalization techniques: A review. *Renewable and Sustainable Energy Reviews*, 134:110227, 2020.

[32] G Di Capua, N Oliva, F Milano, C Bourelly, F Porpora, A Maffucci, and N Femia. A behavioral model for lithium batteries based on genetic programming. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2023.

[33] Yuanli Ding, Zachary P Cano, Aiping Yu, Jun Lu, and Zhongwei Chen. Automotive li-ion batteries: current status and future perspectives. *Electrochemical Energy Reviews*, 2:1–28, 2019.

[34] eLAMI. An Innovative Simulated Dataset of Electrical Loads for Advanced Smart Energy Applications. `https://dx.doi.org/10.21227/sj97-0828`, 2022. Online; Accessed:11/07/2022.

[35] Adrian Filip et al. Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research. In *2nd workshop on data mining applications in sustainability (SustKDD)*, volume 2012, 2011.

[36] Abbas Fotouhi, Daniel J. Auger, Karsten Propp, Stefano Longo, and Mark Wild. A review on electric vehicle battery modelling: From Lithium-ion toward Lithium–Sulphur. *Renewable and Sustainable Energy Reviews*, 56:1008–1021, April 2016.

[37] Linda Gaines. The future of automotive lithium-ion battery recycling: Charting a sustainable course. *Sustainable Materials and Technologies*, 1:2–7, 2014.

[38] Jurgen Garche, Chris Dyer, Patrick T Moseley, Zempachi Ogumi, David AJ Rand, and Bruno Scrosati. *Encyclopedia of electrochemical power sources*. Newnes, 2013.

[39] Suman Giri, Mario Bergés, and Anthony Rowe. Towards automated appliance recognition using an emf sensor in nilm platforms. *Advanced Engineering Informatics*, 27(4):477–485, 2013.

[40] S. Gold. A PSPICE macromodel for lithium-ion batteries. In *The Twelfth Annual Battery Conference on Applications and Advances*, pages 215–222, Long Beach, CA, USA, 1997. IEEE.

[41] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*, 2020.

[42] MA Hannan, Md Murshadul Hoque, Azah Mohamed, and Afida Ayob. Review of energy storage systems for electric vehicle applications: Issues and challenges. *Renewable and Sustainable Energy Reviews*, 69:771–789, 2017.

[43] George William Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

[44] Hongwen He, Rui Xiong, and Jinxin Fan. Evaluation of Lithium-Ion Battery Equivalent Circuit Models for State of Charge Estimation by an Experimental Approach. *Energies*, 4(4):582–598, March 2011.

[45] Simon Henriet, Umut Şimşekli, Benoit Fuentes, and Gaël Richard. A generative model for non-intrusive load monitoring in commercial buildings. *Energy and Buildings*, 177:268–278, 2018.

[46] Ari Hentunen, Teemu Lehmuspelto, and Jussi Suomela. Time-Domain Parameter Extraction Method for Thévenin-Equivalent Circuit Battery Models. *IEEE Transactions on Energy Conversion*, 29(3):558–566, September 2014.

[47] Ian R Hill and Ed E Andrukaitis. State-of-charge determination of lead-acid batteries using wire-wound coils. *Journal of power sources*, 162(2):870–877, 2006.

[48] Dickson N. T. How, M. A. Hannan, M. S. Hossain Lipu, and Pin Jern Ker. State of charge estimation for lithium-ion batteries using model-based and data-driven methods: A review. *IEEE Access*, 7:136116–136136, 2019.

[49] IEC. IEC 61000-4-7. `https://webstore.iec.ch/publication/4228`, 2022. Online; Accessed:07.02.2022.

[50] SM Rakiul Islam, Sung-Yeul Park, and Balakumar Balasingam. Circuit parameters extraction algorithm for a lithium-ion battery charging system incorporated with electrochemical impedance spectroscopy. In *2018 IEEE Applied Power Electronics Conference and Exposition (APEC)*, pages 3353–3358. IEEE, 2018.

[51] Farzad Jaliliantabar, Rizalman Mamat, and Sudhakar Kumarasamy. Prediction of lithium-ion battery temperature in different operating conditions equipped with passive battery thermal management system by artificial neural networks. *Materials Today: Proceedings*, 48:1796–1804, 2022.

[52] Shida Jiang and Zhengxiang Song. A review on the state of health estimation methods of lead-acid batteries. *Journal of Power Sources*, 517:230710, 2022.

[53] Du Jiani, Liu Zhitao, Wang Youyi, and Wen Changyun. A fuzzy logic-based model for li-ion battery with soc and temperature effect. In *11th IEEE International Conference on Control Automation (ICCA)*, pages 1333–1338, 2014.

[54] V.H. Johnson. Battery performance models in ADVISOR. *Journal of Power Sources*, 110(2):321–329, August 2002.

[55] Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2(1):1–14, 2015.

[56] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, 1995.

[57] Sahar Khaleghi, Md Sazzad Hosen, Danial Karimi, Hamidreza Behi, S Hamidreza Beheshti, Joeri Van Mierlo, and Maitane Berecibar. Developing an online data-driven approach for prognostics and health management of lithium-ion batteries. *Applied Energy*, 308:118348, 2022.

[58] Mojtaba Ahmadieh Khanesar, Mohammad Teshnehlab, and Mahdi Aliyari Shoorehdeli. A novel binary particle swarm optimization. In *2007 Mediterranean Conference on Control & Automation*, pages 1–6, 2007.

[59] Neeta Khare, Pritpal Singh, and John K Vassiliou. A novel magnetic field probing technique for determining state of health of sealed lead-acid batteries. *Journal of Power Sources*, 218:462–473, 2012.

[60] Jihyun Kim, Thi-Thu-Huong Le, and Howon Kim. Nonintrusive load monitoring based on advanced deep learning and novel signature. *Computational intelligence and neuroscience*, 2017, 2017.

[61] Christoph Klemenjak, Christoph Kovatsch, Manuel Herold, and Wilfried Elmenreich. A synthetic energy dataset for non-intrusive load monitoring in households. *Scientific Data*, 7(1):1–17, 2020.

[62] J Zico Kolter and Matthew J Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62, 2011.

[63] Weicong Kong, Zhao Yang Dong, Jin Ma, David J Hill, Junhua Zhao, and Fengji Luo. An extensible approach for non-intrusive load disaggregation with smart meter data. *IEEE Transactions on Smart Grid*, 9(4):3362–3372, 2016.

[64] Reiner Korthauer. *Lithium-ion batteries: basics and applications*. Springer, 2018.

[65] Markos Koseoglou, Evangelos Tsioumas, Dimitrios Papagiannis, Nikolaos Jabbour, and Christos Mademlis. A novel on-board electrochemical impedance spectroscopy system for real-time battery impedance estimation. *IEEE Transactions on Power Electronics*, 36(9):10776–10787, 2021.

[66] E. Kuhn, C. Forgez, P. Lagonotte, and G. Friedrich. Modelling Ni-mH battery using Cauer and Foster structures. *Journal of Power Sources*, 158(2):1490–1497, August 2006.

[67] Jiamei Lan, Ruichao Wei, Shenshi Huang, Dongping Li, Chen Zhao, Liang Yin, and Jian Wang. In-depth bibliometric analysis on research trends in fault diagnosis of lithium-ion batteries. *Journal of Energy Storage*, 54:105275, 2022.

[68] Thomas F Landinger, Guenter Schwarzberger, and Andreas Jossen. A novel method for high frequency battery impedance measurements. In *2019 IEEE International Symposium on Electromagnetic Compatibility, Signal & Power Integrity (EMC+ SIPI)*, pages 106–110. IEEE, 2019.

[69] Fredrik Larsson, Petra Andersson, Per Blomqvist, and Bengt-Erik Mellander. Toxic fluoride gas emissions from lithium-ion battery fires. *Scientific reports*, 7(1):10018, 2017.

[70] Alexandros Ch Lazanas and Mamas I Prodromidis. Electrochemical impedance spectroscopy - a tutorial. *ACS Measurement Science Au*, 2023.

[71] Dezhi Li, Licheng Wang, Chongxiong Duan, Qiang Li, and Kai Wang. Temperature prediction of lithium-ion batteries based on electrochemical impedance spectrum: a review. *International Journal of Energy Research*, 46(8):10372–10388, 2022.

[72] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *ACM Comput. Surv.*, 50(6), dec 2017.

[73] Matthew Li, Jun Lu, Zhongwei Chen, and Khalil Amine. 30 years of lithium-ion batteries. *Advanced Materials*, 30(33):1800561, 2018.

[74] Matthew Li, Jun Lu, Zhongwei Chen, and Khalil Amine. 30 years of lithium-ion batteries. *Advanced Materials*, 30(33):1800561, 2018.

[75] Xiaoyu Li, Jianhua Xu, Jianxun Hong, Jindong Tian, and Yong Tian. State of energy estimation for a series-connected lithium-ion battery pack based on an adaptive weighted strategy. *Energy*, 214:118858, 2021.

[76] Yuanyuan Li, Daniel-Ioan Stroe, Yuhua Cheng, Hanmin Sheng, Xin Sui, and Remus Teodorescu. On the feature selection for battery state of health estimation based on charging–discharging profiles. *Journal of Energy Storage*, 33:102122, 2021.

[77] Zhenyu Li, Zhongjie Wang, and Liang Wang. Discussion of the relationship between failure and fire of valve regulated lead acid battery. In *E3S Web of Conferences*, volume 185, page 01058. EDP Sciences, 2020.

[78] M.S. Hossain Lipu, M. A. Hannan, Aini Hussain, M.H.M. Saad, A. Ayob, and K. M. Muttaqi. Lithium-ion battery state of charge estimation method using optimized deep recurrent neural network algorithm. In *2019 IEEE Industry Applications Society Annual Meeting*, pages 1–9, 2019.

[79] Qianqian Liu, Chunyu Du, Bin Shen, Pengjian Zuo, Xinqun Cheng, Yulin Ma, Geping Yin, and Yunzhi Gao. Understanding undesirable anode lithium plating issues in lithium-ion batteries. *RSC advances*, 6(91):88683–88700, 2016.

[80] Tari Luca, Giuseppe Berrettoni, Carmine Bourelly, Gianni Cerro, Domenico Capriglione, and Luigi Ferrigno. elami—an innovative simulated dataset of electrical loads for advanced smart energy applications. *IEEE Access*, 10:91177–91191, 2022.

[81] Stephen Makonin, Bradley Ellert, Ivan V Bajić, and Fred Popowich. Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014. *Scientific data*, 3(1):1–12, 2016.

[82] Stephen Makonin, Fred Popowich, Lyn Bartram, Bob Gill, and Ivan V Bajić. Ampds: A public dataset for load disaggregation and eco-feedback research. In *2013 IEEE electrical power & energy conference*, pages 1–6. IEEE, 2013.

[83] James F. Manwell and Jon G. McGowan. Lead acid battery storage model for hybrid energy systems. *Solar Energy*, 50(5):399–405, May 1993.

[84] Egoitz Martinez-Laserna, Iñigo Gandiaga, Elixabet Sarasketa-Zabala, Julia Badeda, D-I Stroe, Maciej Swierczynski, and Ander Goikoetxea. Battery second life: Hype, hope or reality? a critical review of the state of the art. *Renewable and Sustainable Energy Reviews*, 93:701–718, 2018.

[85] Lukas Mauch and Bin Yang. A new approach for supervised power disaggregation by using a deep recurrent lstm network. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 63–67. IEEE, 2015.

[86] Alain Mauger and CM Julien. Critical review on lithium-ion batteries: are they safe? sustainable? *Ionics*, 23:1933–1947, 2017.

[87] Nina Meddings, Marco Heinrich, Frédéric Overney, Jong-Sook Lee, Vanesa Ruiz, Emilio Napolitano, Steffen Seitz, Gareth Hinds, Rinaldo Raccichini, Miran Gaberšček, and Juyeon Park. Application of electrochemical impedance spectroscopy to commercial li-ion cells: A review. *Journal of Power Sources*, 480:228742, 2020.

[88] Huixing Meng and Yan-Fu Li. A review on prognostics and health management (phm) methods of lithium-ion batteries. *Renewable and Sustainable Energy Reviews*, 116:109405, 2019.

[89] Andreas Mertens, Izaak C Vinke, Hermann Tempel, Hans Kungl, LGJ De Haart, Rüdiger-A Eichel, and Josef Granwehr. Quantitative analysis of time-domain supported electrochemical impedance spectroscopy data of li-ion batteries: reliable activation energy determination at low frequencies. *Journal of the Electrochemical Society*, 163(7):H521, 2016.

[90] Scott J. Moura, Federico Bribiesca Argomedo, Reinhardt Klein, Anahita Mirtabatabaei, and Miroslav Krstic. Battery state estimation for a single particle model with electrolyte dynamics. *IEEE Transactions on Control Systems Technology*, 25(2):453–468, 2017.

[91] Kiarash Movassagh, Arif Raihan, Balakumar Balasingam, and Krishna Pattipati. A critical look at coulomb counting approach for state of charge estimation in batteries. *Energies*, 14(14):4074, 2021.

[92] David Murray, Lina Stankovic, and Vladimir Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4(1):1–12, 2017.

[93] B Nagesha Rao. Diamagnetic susceptibility of sulphuric acid-water mixture. In *Proceedings of the Indian Academy of Sciences-Section A*, volume 3, pages 188–192. Springer India New Delhi, 1936.

[94] Pacific Smart Source. Pacific 360 AMX. `https://pacificpower.com/Resources/Documents/360AMX_Datasheet.pdf`, 2022. Online; Accessed:02.02.2022.

[95] Oliver Parson, Grant Fisher, April Hersey, Nipun Batra, Jack Kelly, Amarjeet Singh, William Knottenbelt, and Alex Rogers. Dataport and nilmtk: A building data set designed for non-intrusive load monitoring. In *2015 ieee global conference on signal and information processing (globalsip)*, pages 210–214. IEEE, 2015.

[96] Detchko Pavlov. Lead-acid batteries, science and technology , a handbook of lead-acid battery technology and its influence on the product, chapter 4-lead alloys and grids, grid des. *Grid Des. Princ.*, pages 169–243, 2017.

[97] Kollmeyer Phillip. Panasonic 18650PF Li-ion battery data. *Mendeley Data, V1*, 2018.

[98] David AJ Rand and Patrick T Moseley. Energy storage with lead–acid batteries. In *Electrochemical energy storage for renewable sources and grid balancing*, pages 201–222. Elsevier, 2015.

[99] V. Rao, G. Singhal, A. Kumar, and N. Navet. Battery model for embedded systems. In *18th International Conference on VLSI Design held jointly with 4th International Conference on Embedded Systems Design*, pages 105–110, Kolkata, India, 2005. IEEE Computer Soc.

[100] Juan Pablo Rivera-Barrera, Nicolás Muñoz-Galeano, and Henry Omar Sarmiento-Maldonado. Soc estimation for lithium-ion batteries: Review and future challenges. *Electronics*, 6(4):102, 2017.

[101] Michael A Roscher and Dirk Uwe Sauer. Dynamic electric behavior and open-circuit-voltage modeling of lifepo4-based lithium ion secondary batteries. *Journal of Power Sources*, 196(1):331–336, 2011.

[102] Antonio Ruano, Alvaro Hernandez, Jesus Ureña, Maria Ruano, and Juan Garcia. Nilm techniques for intelligent home energy management and ambient assisted living: A review. *Energies*, 12(11):2203, 2019.

[103] Hamed Sadegh Kouhestani, Xiaoping Yi, Guoqing Qi, Xunliang Liu, Ruimin Wang, Yang Gao, Xiao Yu, and Lin Liu. Prognosis and health management (phm) of solid-state batteries: Perspectives, challenges, and opportunities. *Energies*, 15(18):6599, 2022.

[104] Bhaskar Saha and Kai Goebel. Battery data set. *NASA AMES prognostics data repository*, 2007.

[105] Z.M. Salameh, M.A. Casacca, and W.A. Lynch. A mathematical model for lead-acid batteries. *IEEE Transactions on Energy Conversion*, 7(1):93–98, March 1992.

[106] Valerio Mario Salerno and Graziella Rabbeni. An extreme learning machine approach to effective energy disaggregation. *Electronics*, 7(10):235, 2018.

[107] Fritz Scholz et al. *Electroanalytical methods*, volume 1. Springer, 2010.

[108] Michael Schönleber, Dino Klotz, and E Ivers-Tiffée. A method for improving the robustness of linear kramers-kronig validity tests. *Electrochimica Acta*, 131:20–27, 2014.

[109] Philipp Schröer, Hedi Van Faassen, Thomas Nemeth, Matthias Kuipers, and Dirk Uwe Sauer. Challenges in modeling high power lithium titanate oxide cells in battery management systems. *Journal of Energy Storage*, 28:101189, April 2020.

[110] Kristen A Severson, Peter M Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H Chen, Muratahan Aykol, Patrick K Herring, Dimitrios Fraggedakis, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 2019.

[111] Tareq M. Shami, Ayman A. El-Saleh, Mohammed Alswaitti, Qasem Al-Tashi, Mhd Amen Summakieh, and Seyedali Mirjalili. Particle swarm optimization: A comprehensive survey. *IEEE Access*, 10:10031–10061, 2022.

[112] Pei Kang Shen, Chao-Yang Wang, Xueliang Sun, Jiujun Zhang, et al. *Electrochemical energy: advanced materials and technologies*. CRC press, 2018.

[113] Changho Shin, Eunjung Lee, Jeongyun Han, Jaeryun Yim, Wonjong Rhee, and Hyoseop Lee. The enertalk dataset, 15 hz electricity consumption data from 22 houses in korea. *Scientific data*, 6(1):1–13, 2019.

[114] Prashant Shrivastava, Tey Kok Soon, Mohd Yamani Idna Bin Idris, Saad Mekhilef, and Syed Bahari Ramadzan Syed Adnan. Comprehensive co-estimation of lithium-ion battery state of charge, state of energy, state of power, maximum available capacity, and maximum available energy. *Journal of Energy Storage*, 56:106049, 2022.

[115] Shikha Singh and Angshul Majumdar. Deep sparse coding for non–intrusive load monitoring. *IEEE Transactions on Smart Grid*, 9(5):4669–4678, 2017.

[116] Mandavilli Srinivas and Lalit M Patnaik. Genetic algorithms: A survey. *computer*, 27(6):17–26, 1994.

[117] Vladimir Stankovic, Jing Liao, and Lina Stankovic. A graph-based signal processing approach for low-rate energy disaggregation. In *2014 IEEE symposium on computational intelligence for engineering solutions (CIES)*, pages 81–87. IEEE, 2014.

[118] SmithBucklin Statistics. National recycling rate study. *Battery Council International, Chicago, Illinois*, 2014.

[119] Mattia Stighezza, Roberto Ferrero, Valentina Bianchi, and Ilaria De Munari. Machine learning and impedance spectroscopy for battery state of charge evaluation. In *2023 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, pages 24–29. IEEE, 2023.

[120] Carlo Taborelli and Simona Onori. State of charge estimation using extended kalman filters for battery management system. In *2014 IEEE International Electric Vehicle Conference (IEVC)*, pages 1–8. IEEE, 2014.

[121] S. Tamilselvi, S. Gunasundari, N. Karuppiah, Abdul Razak Rk, S. Madhusudan, Vikas Madhav Nagarajan, T. Sathish, Mohammed Zubair M. Shamim, C. Ahamed Saleel, and Asif Afzal. A Review on Battery Modelling Techniques. *Sustainability*, 13(18):10042, September 2021.

[122] Jiliang Tang, Salem Alelyani, and Huan Liu. *Feature selection for classification: A review*, pages 37–64. CRC Press, January 2014. Publisher Copyright: © 2015 by Taylor & Francis Group, LLC.

[123] Tektronix. High Voltage Differential Probes. `https://download.tek.com/manual/P5200A-High-Voltage-Differential-Probe-Instruction-Manual-077053602.pdf`, 2022. Online; Accessed:02.02.2022.

[124] Tektronix. Probe Power Supply 1103. https://download.tek.com/datasheet/1103_Probe_Power_Supply_Datasheet_60W-14843-1_2.pdf, 2022. Online; Accessed:02.02.2022.

[125] Tektronix. TCP202 DC Coupled Current Probe. https://download.tek.com/datasheet/TCP202_Current_Probe_Datasheet_51W-10736-7.pdf, 2022. Online; Accessed:02.02.2022.

[126] Marc Thele, Oliver Bohlen, Dirk Uwe Sauer, and Eckhard Karden. Development of a voltage-behavior model for NiMH batteries using an impedance-based modeling concept. *Journal of Power Sources*, 175(1):635–643, January 2008.

[127] TiePie. Handyscope HS5. https://www.tiepie.com/en/usb-oscilloscope/handyscope-hs5, 2022. Online; Accessed:02.02.2022.

[128] Jorn Tinnemeyer. 31-4: Diamagnetic measurements in lead acid batteries to estimate state of charge. *Journal of Power Sources04312010*, 4, 2010.

[129] Akshay SN Uttama Nambi, Antonio Reyes Lua, and Venkatesha R Prasad. Loced: Location-aware energy disaggregation framework. In *Proceedings of the 2nd acm international conference on embedded systems for energy-efficient built environments*, pages 45–54, 2015.

[130] Yi Wang, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, and Yun Zhao. Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, 20(2):117–129, 2015.

[131] Yung Fei Wong, Y Ahmet Şekercioğlu, Tom Drummond, and Voon Siong Wong. Recent approaches to non-intrusive load monitoring techniques in residential settings. In *2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG)*, pages 73–79. IEEE, 2013.

[132] Qian Wu and Fei Wang. Concatenate convolutional neural networks for non-intrusive load monitoring across complex background. *Energies*, 12(8):1572, 2019.

[133] Yinjiao Xing, Eden WM Ma, Kwok L Tsui, and Michael Pecht. Battery management systems in electric and hybrid vehicles. *Energies*, 4(11):1840–1857, 2011.

[134] RCM Yam, PW Tse, L Li, and P Tu. Intelligent predictive decision support system for condition-based maintenance. *The International Journal of Advanced Manufacturing Technology*, 17:383–391, 2001.

[135] Atsuo Yamada, Sai-Cheong Chung, and Koichiro Hinokuma. Optimized lifepo4 for lithium battery cathodes. *Journal of the electrochemical society*, 148(3):A224, 2001.

[136] Jihong Yan, Yue Meng, Lei Lu, and Lin Li. Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. *IEEE Access*, 5:23484–23491, 2017.

[137] Yokogawwa. WT3000 Yokogawa. https://tmi.yokogawa.com/solutions/discontinued/wt3000-precision-power-analyzer/, 2022. Online; Accessed:04.02.2022.

[138] K Zaghib, J Dubé, A Dallaire, K Galoustov, A Guerfi, M Ramanathan, A Benmayza, J Prakash, A Mauger, and CM Julien. Enhanced thermal safety and high power performance of carbon-coated lifepo4 olivine cathode for li-ion batteries. *Journal of Power Sources*, 219:36–44, 2012.

[139] Ying Zhang and Yan-Fu Li. Prognostics and health management of lithium-ion battery using deep learning methods: A review. *Renewable and Sustainable Energy Reviews*, 161:112282, 2022.

[140] Yunwei Zhang, Qiaochu Tang, Yao Zhang, Jiabin Wang, Ulrich Stimming, and Alpha A Lee. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature communications*, 11(1):1706, 2020.

[141] Bochao Zhao, Lina Stankovic, and Vladimir Stankovic. On a training-less solution for non-intrusive appliance load monitoring using graph signal processing. *IEEE Access*, 4:1784–1799, 2016.

[142] Jingyuan Zhao and Andrew F. Burke. Electric vehicle batteries: Status and perspectives of data-driven diagnosis and prognosis. *Batteries*, 8(10), 2022.

[143] Yanyan Zhao, Oliver Pohl, Anand I Bhatt, Gavin E Collis, Peter J Mahon, Thomas Rüther, and Anthony F Hollenkamp. A review on battery market trends, second-life reuse, and recycling. *Sustainable Chemistry*, 2(1):167–205, 2021.

[144] Ghassan Zubi, Rodolfo Dufo-López, Monica Carvalho, and Guzay Pasaoglu. The lithium-ion battery: State of the art and future perspectives. *Renewable and Sustainable Energy Reviews*, 89:292–308, 2018.