



# D11.5 – Demonstrate transfer of select model(s) to novel battery materials/chemistry

## VERSION

VERSION	DATE
1	19/2-2024

## PROJECT INFORMATION

GRANT AGREEMENT NUMBER	957189
PROJECT FULL TITLE	Battery Interface Genome - Materials Acceleration Platform
PROJECT ACRONYM	BIG-MAP
START DATE OF THE PROJECT	1/9-2020
DURATION	3 years
CALL IDENTIFIER	H2020-LC-BAT-2020-3
PROJECT WEBSITE	big-map.eu

## DELIVERABLE INFORMATION

WP NO.	WP11
WP LEADER	Arghya Bhowmik
CONTRIBUTING PARTNERS	CID, CSIC, NIC, DTU
NATURE	Demonstrator
AUTHORS	Beñat Larrarte, Francisco Javier Méndez, Elixabete Ayerbe,
CONTRIBUTORS	Deyana Stoytcheva, Rosa Palacin, Jan Bitenc, Robert Dominko, Sabine Ploteau, Tammo Schwietert, Vishank Kumar, Xueping Qin, Arghya Bhowmik
CONTRACTUAL DEADLINE	30/12-2023
DELIVERY DATE TO EC	19/2-2024
DISSEMINATION LEVEL (PU/CO)	PU

## ACKNOWLEDGMENT



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189. The project is part of BATTERY 2030+, the large-scale European research initiative for inventing the sustainable batteries of the future.



## ABSTRACT

This deliverable unveils a groundbreaking shift in predictive models, seamlessly transferring expertise from lithium-ion to emerging systems like sodium and magnesium-based technologies. Addressing lithium-ion limitations, our approach accelerates sustainable energy storage solutions through machine learning models and tailored experimental techniques. Demonstrating remarkable adaptability, our methodologies contribute significantly to the overall progress in energy storage technologies.

## TABLE OF CONTENTS

<b>1. INTRODUCTION .....</b>	<b>2</b>
<b>2. LITHIUM AND SODIUM BATTERY LIFETIME GENERATOR USING REGRESSION APPROACH AND ELECTRODE EMBEDDING .....</b>	<b>3</b>
<b>3. DEEP NEURAL NETWORK FOR BATTERY CAPACITY FACE PREDICTION .....</b>	<b>12</b>
<b>4. FROM LI-ION TO MG METAL ANODE-ORGANIC BATTERY .....</b>	<b>15</b>
<b>5. EXPERIMENTAL APPROACHES TOWARDS TRANSFERABILITY TO NOVEL BATTERY CHEMISTRIES.....</b>	<b>17</b>
<b>6. ELECTROLYTE SYSTEMS (LI/NA/MG) BASED ON AIMD AND MACHINE LEARNING SIMULATIONS.....</b>	<b>19</b>

## 1. Introduction

The evolving landscape of battery technologies demands predictive models that transcend traditional lithium-ion systems. This formal deliverable showcases the successful transfer of models designed for lithium-ion to emerging frontiers like sodium- or magnesium-based systems, addressing the limitations of lithium-ion technology and advancing sustainable energy storage solutions.

This deliverable covers machine learning models for battery lifetime prediction and tailored experimental approaches to assess transferability. Demonstrating adaptability beyond the original scope underscores the versatility of our methodologies, contributing to the overall advancement of energy storage technologies.

The subsequent sections delve into specific aspects of transferability, including the lithium and sodium battery lifetime generator using regression approach and electrode embedding, deep neural network for battery capacity face prediction, the transition from Li-ion to magnesium metal anode-organic battery, experimental approaches towards transferability to novel battery chemistries and electrolyte systems (Li/Na/Mg) based on AIMD and machine learning simulations. Each section



provides a comprehensive understanding of the methods employed and outcomes achieved in extending predictive models to diverse energy storage technologies, shaping the trajectory of future advancements in the field.

## 2. Lithium and sodium battery lifetime generator using regression approach and electrode embedding

### 2.1 Introduction

The current landscape of predicting battery lifetime demands the evolution of machine learning (ML) models capable of transcending specific chemical compositions and technological variations. Building upon Rieger et al.'s (13) foundational work on remaining useful lifetime prediction models (WP11), this study explores innovative methodologies to address this crucial requirement. Traditionally, prevalent models in this field rely on time series forecasting, using cycle estimation to construct input sliding windows. However, departing from this norm, our research introduces a novel approach focused on a regression-based model transfer from lithium-ion (Li-ion) technology to diverse chemical compositions and technological frameworks. This pioneering paradigm incorporates essential cell and testing characteristics, enabling the forecasting of battery lifetime across various chemistries and technologies.

Utilizing an extensive dataset from WP6 and a synthesis of scholarly literature, our study enhances its neural network with multifaceted data, encompassing diverse chemistry and technology variants, battery designs, and various testing conditions. This comprehensive dataset forms the bedrock for developing an adaptable and versatile predictive model. Additionally, our research introduces an automated data postprocessing methodology, streamlining and refining the dataset through techniques such as time series outlier detection, correction, smoothing, and undersampling. This meticulous approach enriches the dataset and enhances the robustness and accuracy of subsequent predictive models.

In terms of empirical findings, our newly devised model competes vigorously with existing state-of-the-art models and demonstrates remarkable performance across a wide array of scenarios. Unlike conventional models restricted to specific circumstances, our model's adaptability and competence across diverse scenarios underscore its unprecedented versatility. The structure of this section unfolds across distinct sections, meticulously examining battery data detailing the methodology, including model approach, electrode embedding techniques, feedforward neural network architecture, and the proposed Feedforward Neural Network (FNN) model tailored for lifetime generation. Subsequent sections elaborate on the empirical results obtained, leading to a comprehensive conclusion summarizing the findings and their broader implications within the field of battery lifetime prediction.

### 2.2 Battery data

#### 2.2.1 Data acquisition

Within the framework of BIG-MAP, this investigation systematically compiles approximately 1000 cycling test outcomes by harnessing data derived from public repositories and scholarly publications. This expansive dataset encompasses parameters incorporating 11 distinct positive electrodes, four disparate negative electrodes, and various cell designs, such as coin, cylindrical, and



pouch configurations. The dataset comprehensively addresses static and dynamic charging and discharging profiles from C/10 to 8C, encompassing a wide range of working voltage windows and accommodating varying temperature conditions. The primary aim, facilitated by this comprehensive dataset, is developing a sophisticated machine learning model proficient in predicting battery lifespan across a spectrum of scenarios.

The significance of this diversity extends to substantial benefits and practical applications. Firstly, the broad spectrum of cell chemistry enhances the model's utility and contributes to discovering novel electrode compositions. This diminishes the necessity for extensive experimental testing and augments the exploratory potential for new electrode compositions through electrode embedding. Secondly, incorporating dynamic charging and discharging profiles based on state of charge (SOC) evolution aligns with recent advancements in the field, wherein these protocols have garnered popularity due to their efficacy. Thirdly, the inclusion of cell design proves to be a pivotal aspect in facilitating the scalability of cells, ranging from coin to pouch configurations. The resultant table concisely summarizes the data employed in this methodology, adhering to established best practices in the composition of scientific papers.

**Table 1.** Summary of all experimental data used in this model.

Author	Positive Electrode	Negative Electrode	Cell Design	Charging Profile	Discharging Profile	Voltage Min	Voltage Max	Temperature
Attia et al. [1]	LFP	Gr	Cylindric	Dynamic	4	2	3.6	30
Preger et al. [2]	LFP, NCA, NMC811	Gr	Cylindric	0.5	0.5, 1, 2, 3	2, 2.5	3.6, 4.2	15, 25, 35
Zhang et al. [3]	LNO	Gr, Gr/Si	Coin	1	1	2.5	4.2	23
Birkl et al. [4]	LCO	Gr	Pouch	2	1.84	2.5	4.2	40
He et al. [5]	LCO	Gr	Prismatic	0.5, 1	0.5, 1	2.7	4.2	23
CAMP [6]	LMNO, NMC532, NMC111, NMC622, NMC, NMC811	LTO, Gr, Gr/Si	Pouch	btw 0.1, 2	btw 0.1, 2	btw 2.5, 3.5	btw 3, 4.8	23
Severson et al. [7]	LFP	Gr	Cylindric	Dynamic	4	2	3.6	30
CID & BM WP6 [8]	LNO, NMC811	Gr	Coin	1	1, 2, 3	2.5	4.1, 4.2, 4.3	15, 25, 45
CID internal [9]	NCA	Gr/Si	Cylindric	0.5	0.5, 1	2.5	4.2, 4.35, 4.5	10, 25, 45
Wang et al. [10]	NVPF	HC	Coin	5	5	2	4.5	25



Wei et al. [11]	NVZP	HC	Cylindric	0.2, 0.5	0.2, 0.5	1	4.5	25
Uthpala et al. [12]	NVZP	HC	Coin, Cylindric	0.5	1	1, 1.5	4.05, 4.2	28,40

### 2.2.2 Data processing

In the realm of large databases for machine learning, data postprocessing is pivotal. As datasets grow, so does complexity, introducing noise that can impact model performance. Postprocessing systematically refines data, enhancing quality to optimize machine learning model training. Its importance lies in cleansing data and directly boosting model accuracy and adaptability to complex datasets. In this context, postprocessing becomes essential for unlocking the full potential of machine learning in analyzing vast and intricate datasets, ensuring accurate and reliable insights. This way, an automatic methodology is built based on three steps:

- *Step 1 - Outlier detection and correction:* The benefits of outlier detection lie in enhancing data quality. Identifying and correcting unusual data points improves the accuracy of analyses and models. This process ensures that anomalies that could adversely affect decision-making are addressed.
- *Step 2 - Time series smoothing:* Time series smoothing offers the advantage of revealing underlying patterns and trends more clearly by reducing noise and variations. This improved clarity aids in better understanding the inherent behavior of the data, facilitating more accurate trend identification and pattern recognition.
- *Step 3 - Time series undersampling:* Undersampling provides a balanced dataset by selectively reducing the number of data points, particularly from the majority class in imbalanced datasets. This balance mitigates the dominance of the majority class and prevents bias in machine learning models, ultimately improving predictive accuracy. In the specific context of this work, where each state of health (SoH) percentile has only one associated cycle, undersampling ensures a distributed learning approach across the entire life spectroscopy, contributing to a more comprehensive and unbiased model.

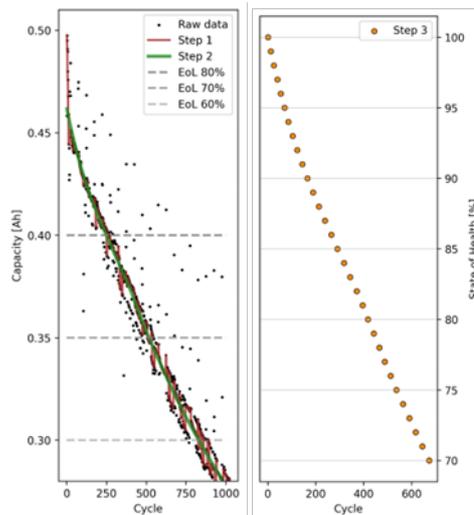


Figure 1. An example of the automatic postprocessing.

## 2.3 Methodology

### 2.3.1 Model approach

Most models within the field address lifetime prediction by employing time series forecasting techniques, wherein historical data from various cycling cells are utilized to forecast the behavior of another. These models learn from patterns, seasonality, trends, and relevant factors observed in the degradation evolution of other cells to predict a new one. However, they encounter challenges in making accurate long-term predictions due to several drawbacks:

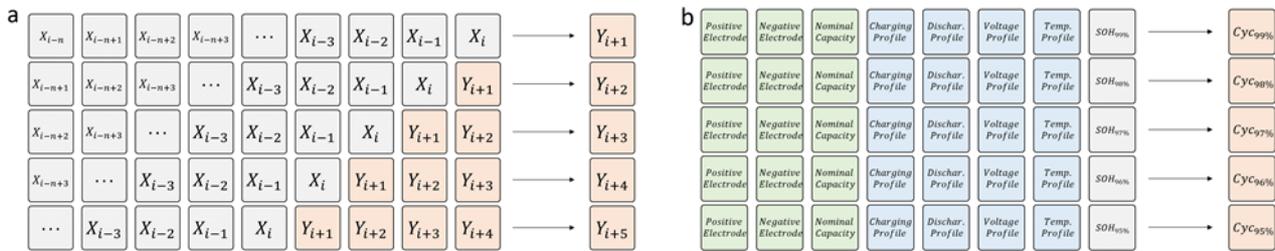
- *Accumulation of errors:* As predictions from previous cycles become part of the input sequence, error propagation and uncertainty in long-term simulations become challenging to control.
- *Limited historical context:* Models may struggle to forecast beyond the range of the historical data they were trained on, lacking insight into potential future unforeseen events or shifts in underlying patterns. This challenge is particularly highlighted when predicting the evolution of a new cell based on similarities to others unless the model can compare among cells and testing conditions.
- *Sensitivity to inputs:* While mid-term patterns can be forecasted from the initial set of cycles, predicting long-term patterns and behavior becomes challenging due to the extreme sensitivity of these input cycles.

Thus, although machine learning models can effectively predict short to medium-term patterns by leveraging historical data, their accuracy diminishes for longer-term forecasts due to input sensitivity and uncertainties related to future events and patterns. Consequently, we propose utilizing a regression approach for battery lifetime forecasting, integrating domain knowledge encompassing cell chemistry, design, and testing conditions. This approach addresses the limitations of time series forecasting by incorporating more comprehensive domain-specific information.

The regression approach employed here is intended to enrich the machine learning model by incorporating detailed information about the specific cell characteristics and the varied conditions under which testing is conducted. This strategic inclusion of diverse data aims to render the solution adaptable and transferable, ensuring its efficacy across a spectrum of scenarios and applications.



Furthermore, the essence of the battery lifetime forecast hinges on the prediction of cycle numbers associated with particular SoH percentiles. This approach essentially aims to discern and forecast the conclusion points for each percentile, providing insight into when a specific SoH percentile is anticipated to reach its culmination. The model offers precise estimations regarding the termination phase of individual SoH percentiles, enhancing the comprehensive understanding of battery lifetime prognosis.

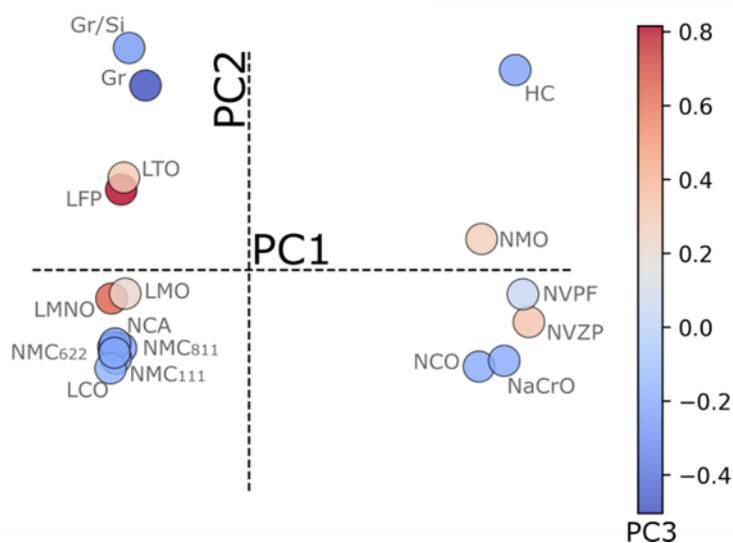


**Figure 2.** a) Illustration of battery lifetime prediction based on time series, b) battery lifetime prediction based on regression approach.

### 2.3.2 Electrode embedding

A significant challenge emerged in the pursuit of integrating comprehensive data related to cell properties and testing conditions, particularly concerning the intricate and pivotal role played by the positive and negative electrodes within the cell behavior framework. Recognizing the complexity and criticality of this facet in our study prompted the development of a specialized embedding technique, drawing inspiration from the concept of word embedding utilized in Natural Language Processing.

Unlike words, electrodes possess measurable and detailed characteristics. Leveraging this distinction, we established a customized electrode mapping methodology akin to word embedding. This approach incorporates principal component analysis (PCA) to effectively reduce dimensionality, facilitating a concise representation of electrode characteristics. Consequently, this electrode mapping technique is a foundational mechanism for comprehensively capturing and analyzing the nuanced attributes of electrodes within the cell behavior domain. By leveraging this custom embedding strategy, the machine learning process becomes adept at juxtaposing various electrodes, thus enabling the establishment of a comprehensive map illustrating similarities between them. This methodology facilitates a nuanced comparison among different electrodes, consequently engendering an intricate map delineating their respective similitudes.



**Figure 3.** The electrode map is based on principal component analysis (PCA) and visualized using the most important principal components (PC).

### 2.3.3 Proposed neural network model

The feedforward neural network, a foundational structure within artificial neural networks, orchestrates unidirectional information flow—originating from input nodes, traversing potential hidden layers, and culminating at output nodes without cyclic connections. Compared to more intricate counterparts like recurrent or convolutional neural networks, feedforward neural networks work accurately in solving regression problems thanks to their simplicity.

Architecturally, it consists of three distinct layers: input, hidden, and output. Each layer comprises individual neurons interconnected through weighted links. The input layer, aligned with input data dimensions, transmits inputs to subsequent layers. Hidden layers, functioning as computational cores, compute weighted sums of preceding layer outputs, apply activation functions, and transmit outcomes. Finally, the output layer generates network outputs based on provided inputs.

Neurons in each layer establish complete interconnections with those in the subsequent layer, with connection strength represented by weights. Learning involves iterative weight updates responding to output errors.

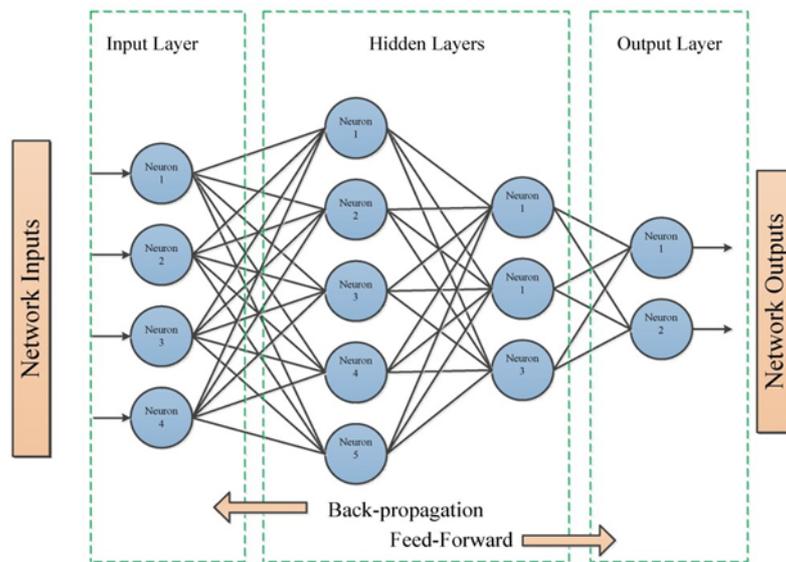


Figure 4. An illustration of the Feedforward Neural Network model.

## 2.4 Results

The outcomes of our comprehensive study, enabled by a meticulously curated database and an intricate post-processing methodology, underscore the model's remarkable forecasting precision across the entire spectrum of electrodes and other input parameters. This uniform accuracy holds for the entire range of input data on which the model has been trained. Consequently, there is no imperative to adapt or employ distinct methodologies when working with lithium-ion or sodium-ion technologies, given that all experimental data is centralized.

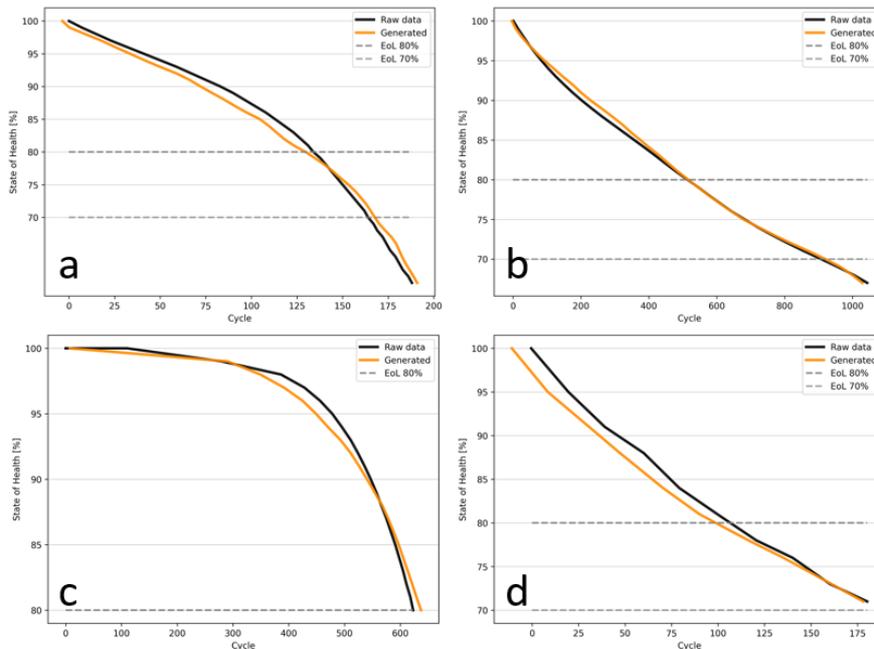


Figure 5. a) NMC811/Gr b) NMC532/Gr-Si c) LFP/Gr d) NVZP/HC.



Furthermore, our battery lifetime generation mechanism simulates the model multiple times with identical inputs, varying solely the SOH input value to forecast the cycle number. This approach facilitates the expeditious generation of SOH evolution through a limited number of simulations, eliminating the need to forecast every cycle. This not only reduces the potential for error or uncertainty but also enhances the efficiency of the process. Conversely, a singular simulation can accurately determine the cycle number at which the model reaches 80 % SOH, providing valuable insights with a minimal computational burden. The model interpretation reveals its proficiency:

- Accurate comprehension of capacity decay without relying on time series.
- The distinction between different capacity evolution shapes.
- Precise forecasting in diverse scenarios involving various cell chemistry, cell designs, or testing conditions.
- Correct predictions across a range of cycling lengths, from 50 to 5000.
- Applicability to both lithium-ion and sodium-ion cells

The model's accuracy is estimated by using mean absolute percentage error (MAPE) to understand the error magnitude easily. Nevertheless, the model comparison is not straightforward due to the MAPE limitations regarding small denominators or data differentiation among different models.

**Table 2.** Comparison of different battery lifetime prediction models.

Publication	Date	Model used	MAPE [%]
CIDETEC approach	2024	FNN	6.6
Rieger et al. (13)	2023	LSTM	10.6
Severson et al. (7)	2019	Regression models	8.6
Fuzhan et al. (14)	2023	LSTM Encoder-Decoder	9
Choi et al. (15)	2019	LSTM	1.08
Park et al. (16)	2020	LSTM	1.02

## 2.5 Conclusion

Adopting a regression-based approach for battery lifetime prediction significantly departs from the prevalent time series methodology. While delivering comparable accuracy to existing state-of-the-art models, our model distinguishes itself by its heightened flexibility and remarkable adaptability across diverse chemistries and technologies. Notably, one pivotal advantage of this approach lies in its independence from requiring initial experimental cycles to forecast cell lifetimes. This breakthrough allows the prediction of non-assembled or untested cells based solely on cell properties and historical data within the battery testing domain.

Introducing an advanced automatic postprocessing methodology is a cornerstone achievement, specifically designed to manage vast datasets and address the challenges inherent in harnessing big data for future endeavors. This robust methodology streamlines data processing, setting the stage for more efficient analysis and modeling and positioning our approach at the forefront of handling extensive data challenges.

While our model exhibits commendable accuracy and robustness, continued enhancement is achievable by including a more diverse array of scenarios encompassing varied cell chemistries, designs, and testing conditions. Broadening the scope of our dataset will undoubtedly fortify the model's predictive capabilities, ensuring its efficacy across a wider spectrum of real-world scenarios.



However, an identified limitation lies in electrode embedding based on PCA, where adding new electrodes alters the positions of existing ones, impeding the insertion of new electrodes. To circumvent this challenge, we propose leveraging the encoder-decoder neural network architecture, mitigating input penalization by creating an embedding after the encoder part. This innovative approach addresses the electrode insertion obstacle and exhibits promise in battery lifetime prediction and other deep learning fields.

Lastly, envisaged improvements involve incorporating information and properties related to electrode manufacturing or cell-forming steps into the model. This augmentation aims to empower the model to adapt and evolve alongside changes in battery lifetime, further enhancing its predictive accuracy and relevance in real-world applications.

## 2.6 References

- (1) ATTIA, Peter M., et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 2020, vol. 578, no 7795, p. 397-402.
- (2) PREGER, Yuliya, et al. Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *Journal of The Electrochemical Society*, 2020, vol. 167, no 12, p. 120532.
- (3) ZHANG, Bojing, et al. Robotic cell assembly to accelerate battery research. *Digital Discovery*, 2022, vol. 1, no 6, p. 755-762.
- (4) BIRKL, Christoph. Oxford battery degradation dataset 1. 2017.
- (5) HE, Wei, et al. Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method. *Journal of Power Sources*, 2011, vol. 196, no 23, p. 10314-10321.
- (6) Ward, L., Kubal, J., Babinec, S. J., Lu, W., Dunlop, A., Trask, S., Polzin, B., Jansen, A., & Paulson, N. H. Dataset of NMC battery Tests from CAMP, 2023 Release [Data set]. Materials Data Facility.
- (7) SEVERSON, Kristen A., et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 2019, vol. 4, no 5, p. 383-391.
- (8) CIDETEC work at BIGMAP WP6
- (9) CIDETEC private data (internal project)
- (10) WANG, Mingxue, et al. Synthesis and electrochemical performances of Na<sub>3</sub>V<sub>2</sub>(PO<sub>4</sub>)<sub>2</sub>F<sub>3</sub>/C composites as cathode materials for sodium ion batteries. *RSC advances*, 2019, vol. 9, no 53, p. 30628-30636.
- (11) WEI, Meng, et al. Remaining useful life prediction for 18650 sodium-ion batteries based on incremental capacity analysis. *Energy*, 2022, vol. 261, p. 125151.
- (12) SUBASINGHE, Lihil Uthpala, et al. A Study on the Capacity Degradation in Na<sub>3</sub> 2V1. 8Zn0. 2 (PO<sub>4</sub>)<sub>3</sub> Cathode and Hard Carbon Anode Based Sodium-Ion Cells. *Journal of The Electrochemical Society*, 2022, vol. 169, no 8, p. 080507.
- (13) RIEGER, Laura Hannemose, et al. Uncertainty-aware and explainable machine learning for early prediction of battery degradation trajectory. *Digital Discovery*, 2023, vol. 2, no 1, p. 112-122.
- (14) RAHMANIAN, Fuzhan, et al. Attention towards chemistry agnostic and explainable battery lifetime prediction. 2023.



- (15) Y. Choi, S. Ryu, K. Park and H. Kim, "Machine Learning-Based Lithium-Ion Battery Capacity Estimation Exploiting Multi-Channel Charging Profiles," in *IEEE Access*, vol. 7, pp. 75143-75152, 2019.
- (16) K. Park, Y. Choi, W. J. Choi, H. -Y. Ryu and H. Kim, "LSTM-Based Battery Remaining Useful Life Prediction With Multi-Channel Charging Profiles," in *IEEE Access*, vol. 8, pp. 20786-20798, 2020.

## 3. Deep Neural Network for Battery Capacity Fade Prediction

### 3.1 Introduction

The capacity of rechargeable Li-ion batteries decreases over time due to several degradation mechanisms. Cells must be cycled for enough cycles to study capacity loss over longer timescales, resulting in test times that can take up to several months. This requires a large time allocation on the battery cyclers. Being able to predict the curve based on early cycling data without having to perform all these cycles is, therefore, a major advantage in operational efficiency for battery material producers such as Umicore.

The degradation of batteries is a complicated process that involves multiple mechanisms interacting at different timescales. As a result, it is challenging to define a physics-based model that accurately predicts the capacity curve. However, recent advancements in artificial intelligence have encouraged data-driven approaches (e.g., Saxena et al. (1) or Rieger et al.(2)), which have shown promise in predicting these types of non-linear problems. In this context, CIDETEC has developed a deep neural network (DNN) that predicts the capacity fade of a battery based on the type of electrodes and electrolyte used, the cell design, the nominal capacity of the cell, the C-rate used for charging and discharging, the voltage window used for cycling, and the test temperature. Here, Umicore has internalized this architecture and developed this deep neural network with the help of CIDETEC, performed various tests on its internal data, and evaluated the model's performance. This work thus contributes to this deliverable on accessing the transferability of such models to different chemistry (from LFP to NMC) and to a different environment from CIDETEC to Umicore.

### 3.2 Methods

#### 3.2.1 Input of the neural network

As a cathode active material manufacturer, Umicore is particularly interested in predicting battery capacity curves for different cathode materials. The test protocols are fixed for the batteries used in this study, and the parameters used by CIDETEC, such as C-rate or voltage window, do not vary significantly. This is why Umicore retains only three inputs used by CIDETEC: the nominal capacity of the cell, the temperature, and the cycle number. Additionally, Umicore adds two parameters derived from the voltage curves: the difference between the capacity obtained at the 3<sup>rd</sup> cycle and the 100<sup>th</sup> cycle and the difference in the area under the voltage curve between the 3<sup>rd</sup> and 100<sup>th</sup> cycle.



### 3.2.2 Dataset

The deep neural network model (DNN) is tested on an internal dataset from Umicore. The dataset consists of batteries with NMC cathode active material where different coatings, doping, and synthesis conditions are used; the batteries are tested at varying temperatures.

## 3.3 Results

### 3.3.1 Results of the DNN

The DNN is trained using 60 % of the data as a training set, 15 % of the data as a validation set, and 25 % of the data as a testing set. The architecture and its parameters are optimized to obtain the results with the lowest prediction error (i.e., the mean squared error (MSE) between the actual and the predicted curve regarding the SOH values is minimized).

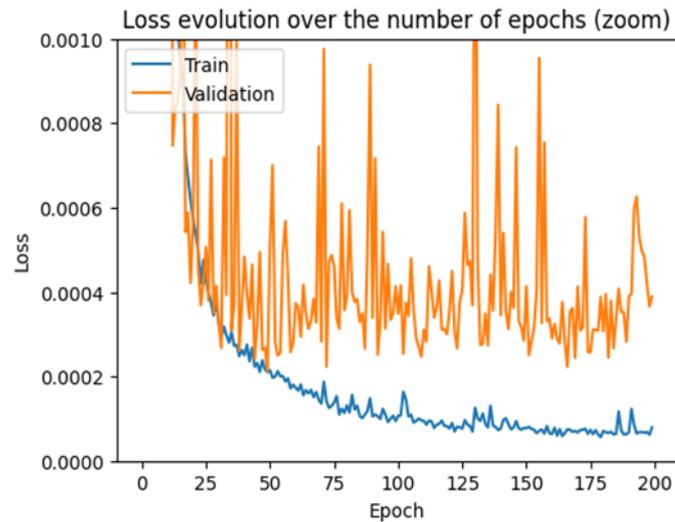
The parameters optimized for the model are the activation function (a function that transforms the input between two consecutive layers), the loss function (a function that calculates the error between the actual value and predicted value), and the features explained in section 3.2.1.

Different activation functions have been applied to find the optimal activation function. Using a “swish” or “relu” activation function to train the model gives the best results on the test set, i.e., the error between the actual curves and the ones predicted is minimized. Especially using the “relu” activation, the performance increases for curves without knee point (the point where the slope of the capacity curve changes and the cell decays faster). Moreover, the end-of-life prediction, which is defined as 80% of initial capacity, improves when using the “relu” activation function.

Similarly, models using different loss functions training the DNN have been carried out. It is observed that the best performance is obtained by using the ‘mean squared error between the actual and the predicted curves’ as a loss function.

Finally, experiments have been performed to assess the model’s performances using different features. It is shown that removing the “area” and the “last point” of the voltage curve as features (explained in section 3.2.1) decreases the model performances on the test set (RMSE multiplication of  $\sim 2.5$  and MAE multiplication of  $\sim 2$ ). Moreover, the addition of features used by Severson et al. (3) (skewness, kurtosis, slope, and intercept of the linear fit to the first point) allows for improvement in the overall model’s performance.

To train the model with the parameters described above, it has been observed, as shown in Figure 6, that about 200 epochs are needed:



**Figure 6.** The train and validation loss.

To quantify the results, two error metrics are calculated:

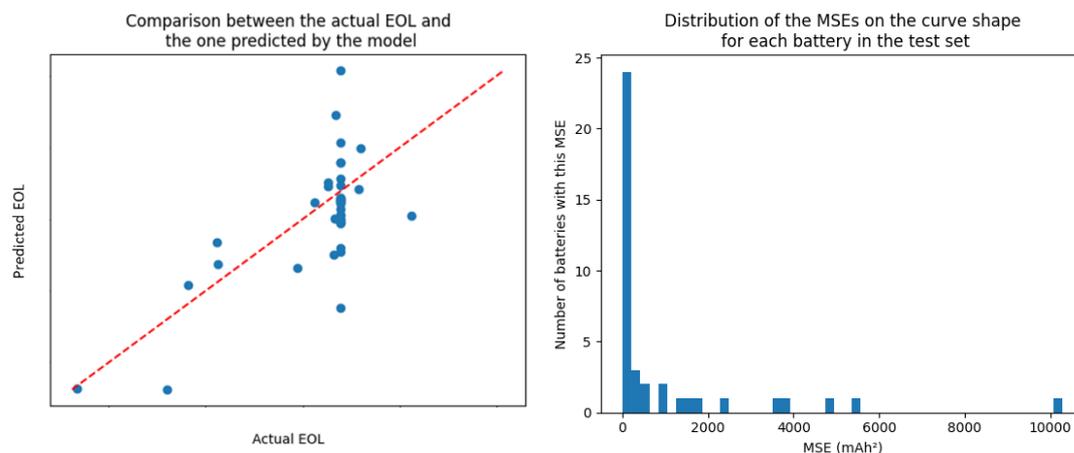
- The root mean square error (RMSE) of the capacity curve is calculated for each battery in the test set, and for each cycle, the squared difference between the predicted capacity curve and an experimentally measured capacity curve is calculated. The RMSE is obtained from this by taking the root of the averaged squared differences. An RMSE percentage is obtained by comparing the RMSE by the capacity window (initial capacity – EOL capacity criteria).
- The RMSE of the capacity curve is calculated for each battery in the test set, and for each cycle, the squared difference between the predicted capacity curve and experimentally measured capacity curve is calculated. The RMSE is obtained from this by taking the root of the averaged squared differences. An RMSE percentage is obtained by comparing the RMSE by the capacity window (initial capacity – EOL capacity criteria).

Graphically, most of the predicted curve shapes have similar profiles to the actual curves. However, some outliers do not capture the right curve shape.

The following metrics give the best results:

- RMSE on the capacity curve: 6.5 %.
- MAPE on the end-of-life: 10 %.

So far, cross-validation has not been carried out. The results of the model are shown in Figure 7.



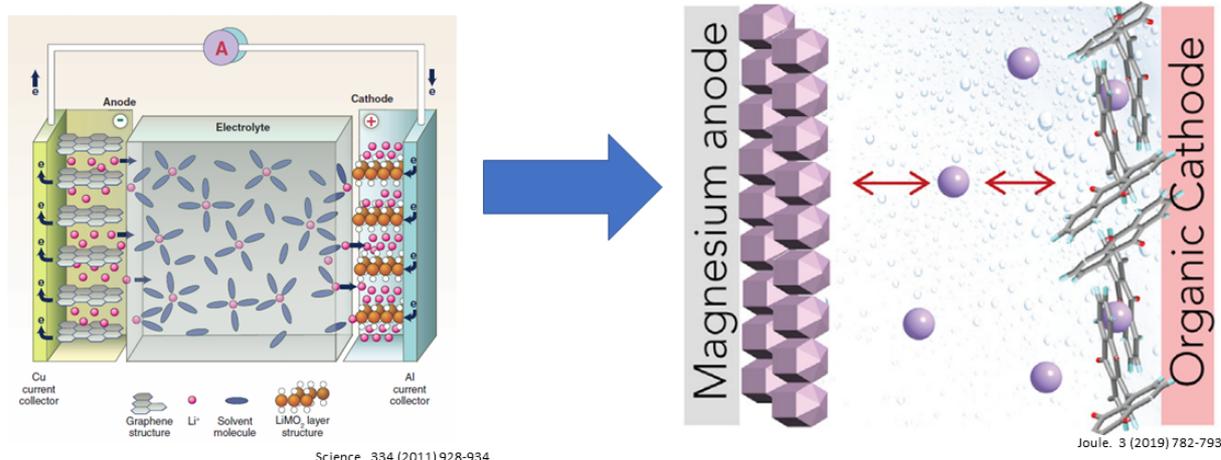
**Figure 7.** Results of the DNN on Umicore data. Left the actual model. Actual vs predicted EOL and MSE distribution. Right, histogram of a number of batteries as a function of MSE.

### 3.4 References

- (1) SAXENA, Saurabh, et al. A convolutional neural network model for battery capacity fade curve prediction using early life data. *Journal of Power Sources*, 2022, vol. 542, p.231736. <https://doi.org/10.1016/j.jpowsour.2022.231736>
- (2) RIEGER, Laura Hannemose, et al. Uncertainty-aware and explainable machine learning for early prediction of battery degradation trajectory. *Digital Discovery*, 2023, vol. 2, no 1, p. 112-122.
- (3) SEVERSON, Kristen A., et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 2019, vol. 4, no 5, p. 383-391.

## 4. From Li-ion to Mg metal anode-organic battery

Mg metal batteries are attracting significant research interest due to the high gravimetric and volumetric capacity of Mg metal anodes. However, major challenges can be identified as a lack of practically applicable Mg electrolytes and cathode materials (1). In recent decades, significant improvements have been achieved in the field of Mg electrolytes, while the standard inorganic cathode material is - still more than two decades old - the Chevrel phase,  $\text{Mo}_6\text{S}_8$  (2). Significant progress has been achieved through the utilization of organic cathode materials. Organic materials have significantly more adaptable structures that accommodate multivalent cations with larger charge densities (3). Hence, direct transfer of Li-ion type rocking chair battery designs is not possible, and we utilized a Mg metal anode setup with an n-type organic cathode based on conjugated carbonyl active materials (Figure 8).



### Reactions at specific electrode:

De-/Intercalation

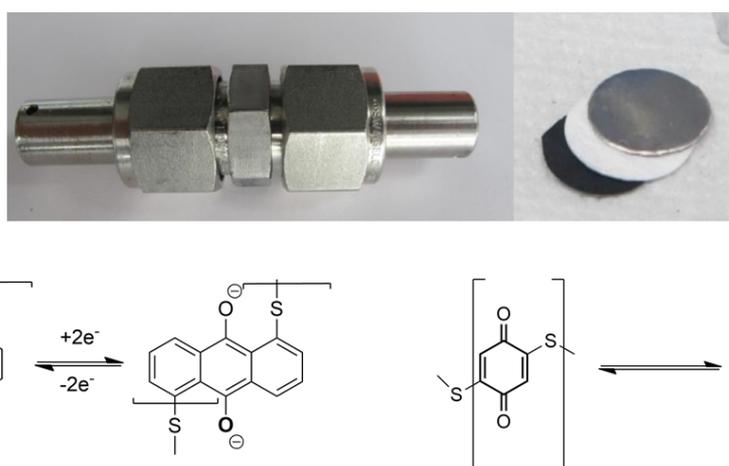
De-/Insertion

Metal plating/stripping

De-/Coordination

**Figure 8.** The conceptual shift from conventional Li-ion battery (left) compared to Mg metal-organic battery (right) utilized in this work, recreated from references 4,5.

In our work, we utilized a half-cell setup using the Mg metal foil anode counter electrode, organic working electrode, and glass fiber separator in a Swagelok-type housing (Figure 9). The organic electrode was composed of 60 % active material, 30 % Printex carbon black, and 10 % PTFE binder. The used electrolyte was based on the combination of  $\text{Mg}(\text{TFSI})_2$  and  $\text{MgCl}_2$  salts in DME solvent (6). The electrochemical characterization was performed using galvanostatic cycling with a defined potential limitation. Two different active materials were utilized: anthraquinone-based poly (anthraquinonyl sulfide)-PAQS and benzoquinone-based poly (hydroquinonyl-benzoquinonyl sulfide)-PHBQS. Both compounds operate through the reduction of carbonyl into an alkoxy group (Figure 9 (7)).



**Figure 9.** Used Swagelok cell setup (top) and proposed electrochemical mechanism of PAQS (bottom left) and PHBQS (bottom right)



## 4.1 References

- (1) Canepa, P.; Sai Gautam, G.; Hannah, D. C.; Malik, R.; Liu, M.; Gallagher, K. G.; Persson, K. A.; Ceder, G. Odyssey of Multivalent Cathode Materials: Open Questions and Future Challenges. *Chem. Rev.* **2017**, *117* (5), 4287–4341. <https://doi.org/10.1021/acs.chemrev.6b00614>.
- (2) Mohtadi, R.; Tutusaus, O.; Arthur, T. S.; Zhao-Karger, Z.; Fichtner, M. The Metamorphosis of Rechargeable Magnesium Batteries. *Joule* **2021**, *5* (3), 581–617. <https://doi.org/10.1016/j.joule.2020.12.021>.
- (3) Ding, C.; Li, C.; Tian, H.; Tong, Y.; Huang, W.; Zhang, Q. Recent Progress on Organic Electrode Materials for Multivalent (Zn, Al, Mg, Ca) Secondary Batteries. *Batter. Supercaps* **2022**, *5* (7). <https://doi.org/10.1002/batt.202200160>.
- (4) Dunn, B.; Kamath, H.; Tarascon, J.-M. Electrical Energy Storage for the Grid: A Battery of Choices. *Science (80-. )*. **2011**, *334* (6058), 928–935. <https://doi.org/10.1126/science.1212741>.
- (5) Dong, H.; Liang, Y.; Tutusaus, O.; Mohtadi, R.; Zhang, Y.; Hao, F.; Yao, Y. Directing Mg-Storage Chemistry in Organic Polymers toward High-Energy Mg Batteries. *Joule* **2019**, *3* (3), 782–793. <https://doi.org/10.1016/j.joule.2018.11.022>.
- (6) Shterenberg, I.; Salama, M.; Yoo, H. D.; Gofer, Y.; Park, J.-B.; Sun, Y.-K.; Aurbach, D. Evaluation of (CF<sub>3</sub>SO<sub>2</sub>)<sub>2</sub>N<sup>-</sup> (TFSI) Based Electrolyte Solutions for Mg Batteries. *J. Electrochem. Soc.* **2015**, *162* (13), A7118–A7128. <https://doi.org/10.1149/2.0161513jes>.
- (7) Pavčnik, T.; Bitenc, J.; Pirnat, K.; Dominko, R. Electrochemical Performance of Mg Metal-Quinone Battery in Chloride-Free Electrolyte. *Batter. Supercaps* **2021**, *4* (5), 815–822. <https://doi.org/doi.org/10.1002/batt.202000293>.

## 5. Experimental approaches towards transferability to novel battery chemistries

Machine learning (ML) potentials are revolutionizing computational chemistry by predicting interatomic energies and forces, accelerating atomic structure optimization, and molecular dynamics simulations. This opens doors for property and structure prediction, reaction analysis, and materials discovery. However, unreliable results can occur without a reliable assessment of prediction confidence. Uncertainty quantification (UQ) methods address this issue by enhancing the robustness and reliability of ML potentials.

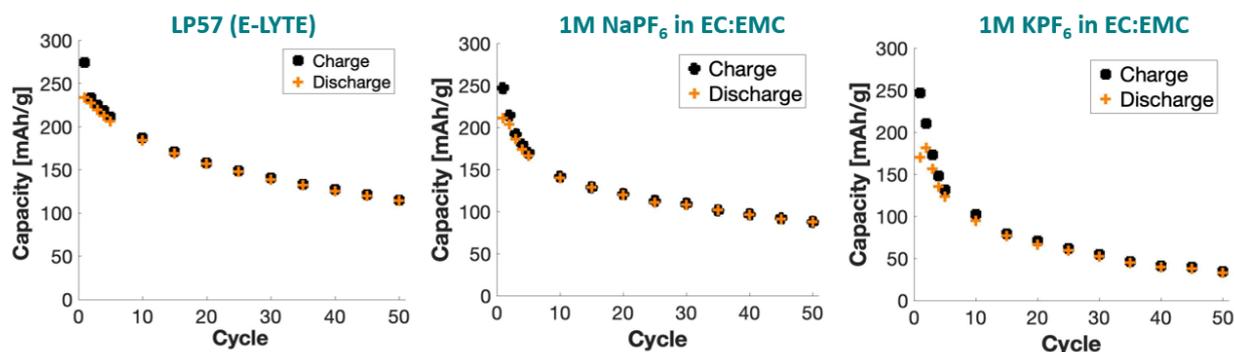
The pathway towards transferability on the experimental side is complex due to the fact that battery technology involves a myriad of parameters, and the electrode and electrolyte components do have compatibility issues as well as interdependencies. A plan was defined to generate experimental data starting from the BIG-MAP standard Li-ion technology and gradually moving to new chemistries, keeping as many parameters as possible constant. Yet, changing a single parameter at a time is not possible, and compromises have to be achieved.

The positive electrode, LiNiO<sub>2</sub>, was kept, as it has a layered structure prone to intercalation of diverse species, and isostructural compounds with Na, K, or even Ca instead of Li are known (1,2).



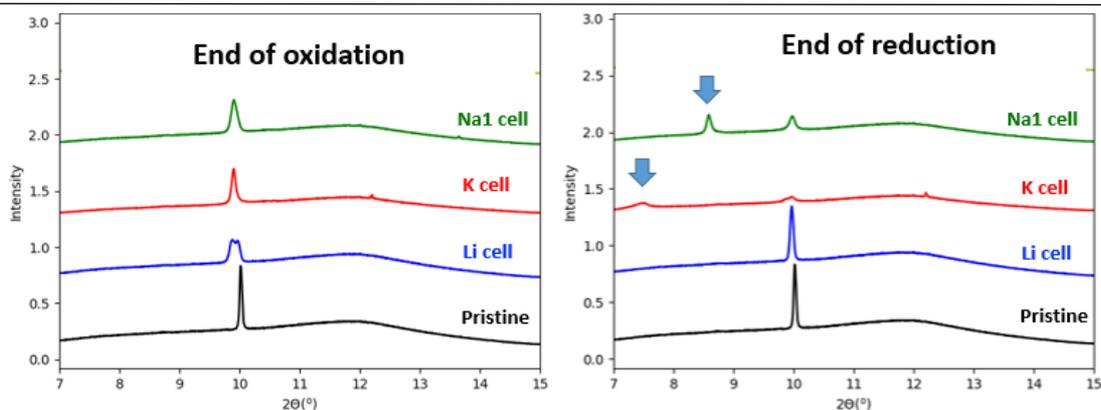
After the first charge, which involves full oxidation and extraction of the  $\text{Li}^+$  ions from the crystal structure, one could expect that other cations present in the electrolyte could intercalate instead. On the negative side, the situation is more complex, as graphite is known not to intercalate naked  $\text{Na}^+$  ions. Thus, a common activated carbon negative electrode was used in all tests. Note that the redox mechanism is not faradaic but capacitive, and hence, it has to be ensured that it is present in a large excess for suitable capacity balancing of the cell. In addition, since its redox potential changes significantly with the state of charge, using a three-electrode setup cell was compulsory to assess its behavior, and modification of voltage limits was needed. With respect to the electrolyte, the initial plan was to use ATFSI or  $\text{B}(\text{TFSI})_2$  for  $\text{A}=\text{Li}, \text{Na}, \text{K}$  or  $\text{B}=\text{Mg}, \text{Ca}$  as salts in order to have a common anion and use a single high dielectric constant solvent to minimize solubility issues. Yet, this came with the associated well-known problem of TFSI not passivating the Al current collector. Thus, corrosion tests were carried out, and a Ti current collector was selected instead. New laminated electrodes were prepared on Ti, but adhesion problems induced a significant performance degradation, even in the case of Li.

A contingency plan was then designed to focus only on monovalent charge carrier ions (Li, Na, K) and keep the electrolyte formulation as 1M  $\text{APF}_6$  in EC/EMC, testing  $\text{LiNiO}_2$  electrodes cast on Al vs. activated carbon counter electrodes in three-electrode cells. The results of cycling at C/10 can be seen in Figure 10. Capacity fading is significant, even in the case of Li, as a result of a non-optimized setup and seen to be larger in the case of  $\text{KPF}_6$ .



**Figure 10.** Capacity vs. cycle number for three identical Swagelok cells using  $\text{LiNiO}_2$  on Al as working electrode, activated carbon as counter electrode (capacity excess >100%), and a silver wire as quasi reference electrode. Electrolyte composition is depicted, the amount present in the cell being 300  $\mu\text{l}$ .

Operando X-ray diffraction experiments were carried out at ALBA synchrotron (MSPD beamline), which enabled following the first three cycles for all cells in Figure 11 are depicted the patterns at the end of oxidation and at the end of reduction where appearance of peaks at low angles can be seen for Na and K, which confirm its involvement in the redox mechanism, despite contribution of the  $\text{Li}^+$  present in the electrolyte after 1st oxidation (calculated to be 14 %) cannot be excluded.



**Figure 11.** Comparison of operando X-ray diffraction patterns taken at ALBA MSPD beamlines on  $\text{LiNiO}_2$ //activated carbon cells using Li, Na, and K electrolytes where the peaks at low angles appear derived from the intercalation of large size cations in the layered  $\text{LiNiO}_2$  structure.

## 5.1 References:

- (1) DELMAS, C. The  $\text{NiO}_2$  Slab: A Very Convenient Structural Unit for Chimie Douce Reactions. *En Materials Science Forum*. Trans Tech Publications Ltd, 1994. p. 131-136.
- (2) CUSHING, B.L.; FALSTER, A.U.; SIMMONS, W.B.; WILEY, J.B. A multivalent ion exchange route to lamellar calcium cobalt oxides,  $\text{Ca}_x\text{CoO}_2$  ( $x \leq 0.5$ ) 1996 p. 2635-2636.

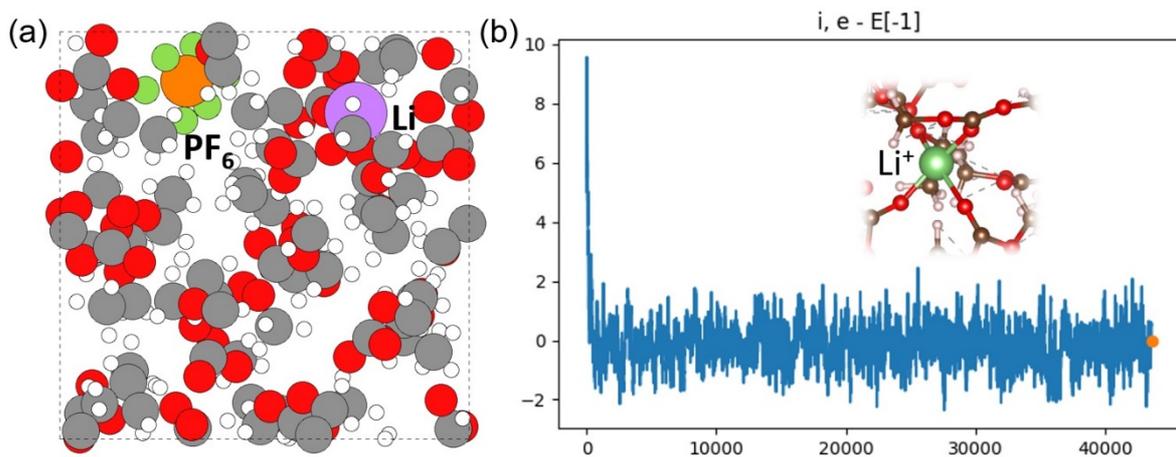
## 6. Electrolyte systems (Li/Na/Mg) based on AIMD and machine learning simulations

We have explored the electrolyte systems, including  $\text{NaPF}_6$  and  $\text{Mg}(\text{ClO}_4)_2$ , in the solvent mixture of EC/EMC = 3:7 (LP57), beyond Li-ion batteries with  $\text{LiPF}_6$ .  $\text{LiPF}_6$  is a widely employed electrolyte in Li-ion batteries, valued for its high ionic conductivity and electrochemical stability. Nevertheless, challenges such as susceptibility to moisture and thermal instability exist. Recent research indicates that augmenting  $\text{LiPF}_6$  electrolytes with specific solvents and co-solvents can enhance their performance. To advance the understanding of  $\text{LiPF}_6$  in solvents, including the solvation properties of metal cations, is essential for optimizing its performance and exploring upgraded or alternative electrolytes. This knowledge is not only applicable to Li-ion batteries but also holds significance for Na and Mg batteries, necessitating a comprehensive exploration of properties in reference to Li cations.

We conducted atomic-scale simulations using ab initio molecular dynamics (AIMD) for three distinct systems featuring Li, Na, and Mg cations. The aim of these simulations was to delve into the solvation properties of metal cations, offering valuable insights into the diffusion movement of cations within widely used electrolytes. Each system was subjected to AIMD simulations, generating trajectories of at least 20 picoseconds (40000 steps with a time step of 0.5 fs). Subsequently, these trajectories were employed as the input dataset for active learning, a methodology designed to accelerate molecular dynamics simulations.

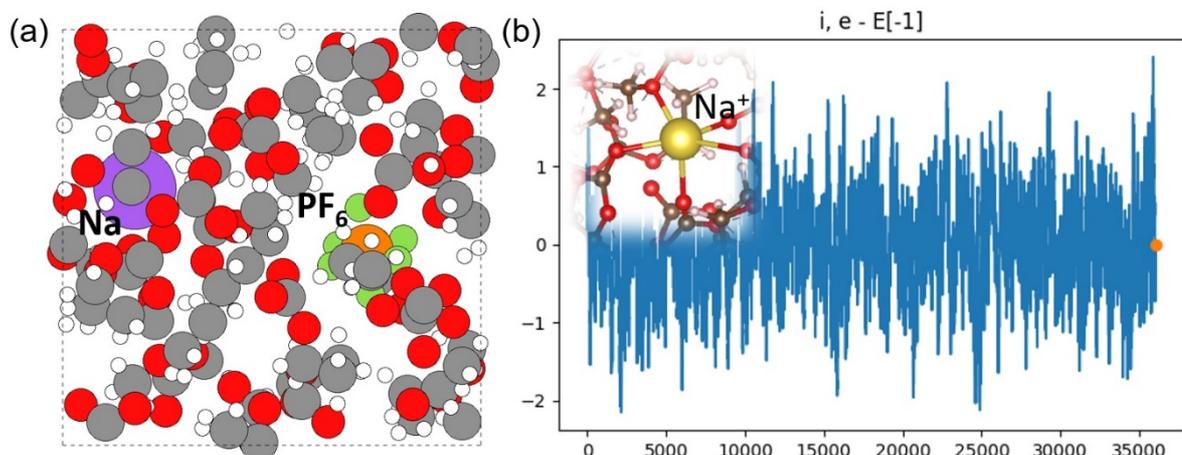


This approach enables us to leverage the information gathered during the initial simulations, optimizing the efficiency of subsequent simulations. By employing active learning, we seek to enhance our understanding of the intricate dynamics and interactions within electrolytes containing Li, Na, and Mg cations. The ultimate goal is to contribute to the development of improved electrolyte formulations and advance the performance of batteries, considering the unique challenges associated with each metal cation.



**Figure 12.** (a) Simulation model of LiPF6; (b) Energy fluctuations with Li solvation (inserted).

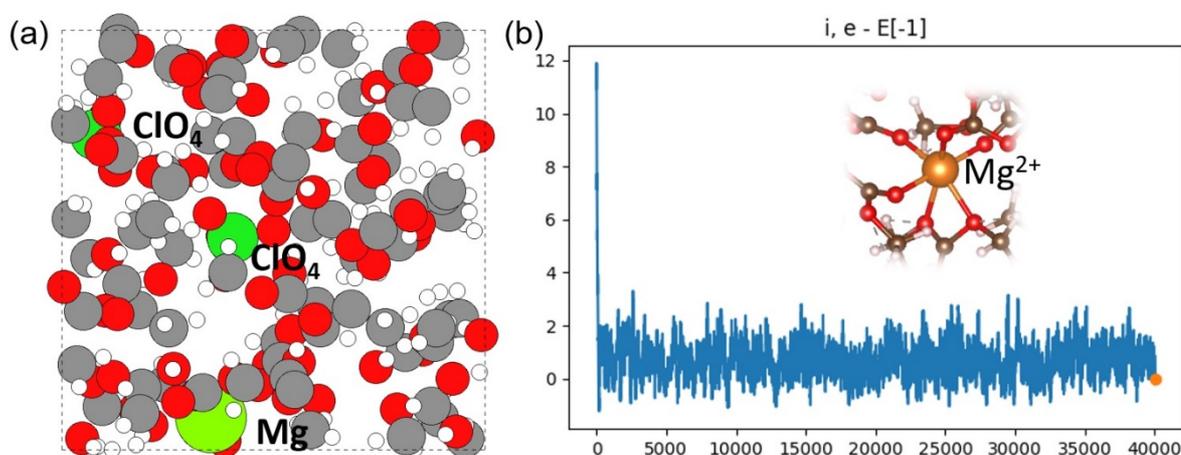
In the case of LiPF6, as illustrated in Figure 12, the dynamic trajectory reveals a separation between the Li cation and PF6 anion (Figure 12a). Throughout the 20 ps AIMD simulations, the system energies exhibit notable stability (Figure 1b). The Li solvation structure, intricately involved in this stability, is detailed in Figure 12b, showcasing a fourfold Li-O coordination between the Li cation and the surrounding solvents, which include 1 ethylene carbonate (1EC) and 3 ethyl methyl carbonate (3EMC). This observed coordination pattern provides crucial insights into the solvation dynamics of Li cations within the LiPF6 electrolyte. The stable trajectory and well-defined solvation structure contribute to understanding the complex interplay between Li cations and electrolyte components. Such detailed analyses are instrumental in refining our comprehension of electrolyte behavior guiding efforts to enhance battery performance and reliability.



**Figure 13.** (a) Simulation model of NaPF6; (b) Energy fluctuations with Na solvation (inserted).

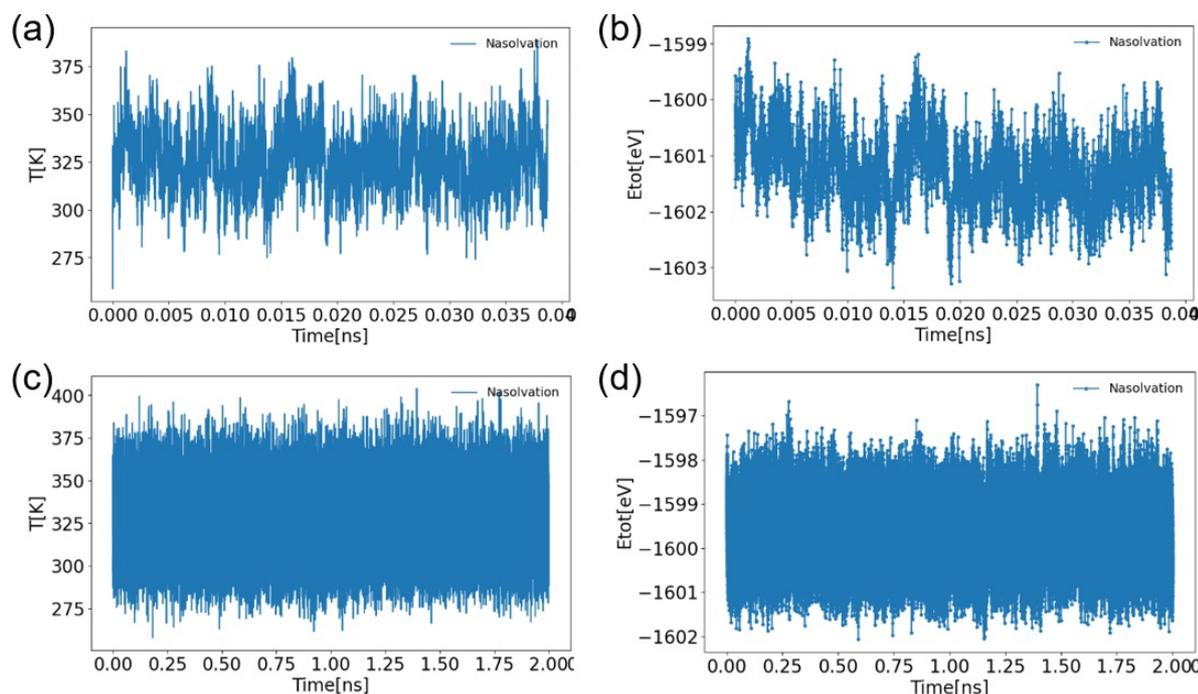


For NaPF<sub>6</sub>, Figure 13a displays the solvent box, illustrating the spatial arrangement of Na and PF<sub>6</sub> molecules amidst the surrounding environment. In Figure 13b, energy fluctuations associated with Na solvation structures are depicted over time. This figure provides a comprehensive view of the dynamic interactions within the NaPF<sub>6</sub> system, focusing on the nuanced behavior of the Na<sup>+</sup> ion. The inset image in Figure 13b offers a detailed perspective on the coordination of the Na<sup>+</sup> ion, revealing a network of six Na-O bonds (2EC and 3EMC). This coordination pattern is essential to understanding the microscopic solvation structure, a critical factor in electrolyte design. Recent studies have highlighted the significance of the weak coulombic interaction between PF<sub>6</sub> and Na, suggesting its role in promoting the transport number of Na<sup>+</sup> ions in bulk liquid and influencing Na storage kinetics. These findings contribute valuable insights to the ongoing efforts aimed at optimizing electrolyte formulations for enhanced performance in sodium-based energy storage systems.



**Figure 14.** (a) Simulation model of Mg(ClO<sub>4</sub>)<sub>2</sub>. (b) Energy fluctuations with Mg solvation (inserted).

For Mg(ClO<sub>4</sub>)<sub>2</sub>, Figure 14 presents both the structural arrangement and energy fluctuations associated with Mg solvation structures. The solvation box in Figure 14a illustrates the spatial configuration of ions and molecules featuring ClO<sub>4</sub> and Mg. In Figure 14b, the dynamic nature of solvation structures is revealed through energy fluctuations over time. Notably, these structures demonstrate stability over the 20 ps molecular dynamic trajectories. Delving into the microscopic solvation structure, as depicted in the inset of Figure 14b, Mg<sup>2+</sup> is found to be stabilized by six Mg-O coordination bonds involving 2EC and 3EMC, a coordination pattern akin to Na cations. Recent studies have emphasized the efficacy of a cooperative solvation/surface engineering approach in achieving reversible Mg plating/stripping within conventional carbonate electrolytes. These insights contribute to the ongoing exploration of strategies for optimizing Mg-based energy storage systems.



**Figure 15.** ML-MD temperature and energy fluctuations in NaPF<sub>6</sub> system for iteration 0 (a and b) and iteration 3 (c and d).

Based on the previously collected AIMD dataset, active learning workflows can accelerate molecular dynamics simulations and provide a better/more accurate understanding of electrolyte interactions. For the Na-PF<sub>6</sub> system, Figure 15 shows the temperature and total energy analysis in Figures 15a and 15b for machine-learning molecular dynamics in iteration 0 and Figures 15c and 15d for iteration 3. It appears that iteration 3 is more accurate and efficient since every iteration of graph neural network (GNN) training gives better results about molecular dynamics. Figures 4a and 4b show a more erratic pattern, indicating less stability in temperature and energy levels at iteration 0. However, as seen in Figures 15c and 15d, there is a noticeable reduction in fluctuations by iteration 3, showcasing enhanced accuracy and efficiency due to the iterative training of the GNN. It also shows that the molecular dynamics can be largely accelerated where 2 ns trajectories can be obtained within two days, which is impossible/impractical for AIMD simulations. This improvement underscores the capability of GNNs to refine molecular dynamics predictions with each training cycle. We will continue this work on Na and Mg ion batteries to help with electrolyte design.

Utilizing the previously collected AIMD dataset, the active learning workflow significantly accelerates molecular dynamics simulations, providing a more nuanced and accurate understanding of electrolyte interactions. Focusing on the Na-PF<sub>6</sub> system, Figure 4 presents a temperature and total energy analysis in Figures 15a and 15b for machine-learning molecular dynamics in iteration 0 and in Figures 15c and 15d for iteration 3. In the initial iteration (iteration 0), Figures 15a and 15b depict a more erratic pattern, suggesting less stability in temperature and energy levels. However, as the graph neural network (GNN) undergoes iterative training, iteration 3 demonstrates enhanced accuracy and efficiency, with a noticeable reduction in fluctuations, as illustrated in Figures 15c and 15d. This iterative refinement highlights the capability of GNNs to improve molecular dynamics predictions with each training cycle.

Furthermore, the results indicate a substantial acceleration of molecular dynamics, allowing the generation of 2 ns trajectories within two days. This represents a significant improvement over



BIG-MAP

## Battery Interface Genome - Materials Acceleration Platform



the time-consuming nature of AIMD simulations, emphasizing the practicality and efficiency of the GNN approach. These advancements reinforce the potential of GNNs in refining molecular dynamics predictions and underscore their role in advancing electrolyte design. Our ongoing work continues and extends to Na and Mg ion batteries, contributing to the broader understanding of electrolyte behavior for improved battery performance.