Plan Overview

A Data Management Plan created using DMPonline.be

Title: Learning the Physics of Dendrite Growth in Lithium-Ion Batteries: An Attention Mechanism Approach for Prevention and Mitigation

Creator: Nele Moelans

Principal Investigator: Nele Moelans

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Nele Moelans

Project abstract:

Lithium ion batteries (LIBs) are considered as the materials of the future when it comes to the efficient energy storage during utilization of renewable energy technologies. Lithium dendrites are responsible for problems like short circuits, catastrophic failures and fires, electrolyte decomposition, and loss of active lithium in these batteries. The formation of dendrites is an interfacial process spanning numerous length- and time scales; and regardless of decades of research, their composition, structure and formation still present a significant conundrum. The achievement of completely dendrite-free battery interfaces can be possible only through the correct understanding of the fundamental mechanisms governing the dendritic evolution. This research proposal presents a combination of approaches linking the microstructure-property and process-kinetics relationship in the different material phases of an electrochemical battery through integrated experiments, computations and artificial intelligence, to subsequently demystify the dendrite evolution mechanisms. Multi-physics finite element simulations and atomistic calculations will be combined. The computational datasets on structure, properties and behavior of battery materials, combined with the experimental datasets of temperature, voltage and other physical quantities, will be employed to construct attentive generative AI models.

ID: 212982

Start date: 01-01-2025

End date: 31-12-2028

Last modified: 24-02-2025

Learning the Physics of Dendrite Growth in Lithium-Ion Batteries: An Attention Mechanism Approach for Prevention and Mitigation

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Learning the Physics of Dendrite Growth in Lithium-Ion Batteries: An Attention Mechanism Approach for Prevention and Mitigation

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Learning the Physics of Dendrite Growth in Lithium-Ion Batteries: An Attention Mechanism Approach for Prevention and Mitigation
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
1) Codes	Phase field simulation codes and codes/scripts for pre- and posprocessing	generate new data	digital	software	.m, c++,.py	<100 MB	
2) Raw data sets	Outcomes from phase-field simulations	generate new data	digital	simulation data	.mat, .csv	< 5 TB	
3) Analysis data	Microstructure analysis data obtained after postprocessing of the simulation results	generate new data	digital	compiled data	.mat, .txt	< 10GB	
4) Publications/presentations	Publications/presentations	generate new data	digital	compiled data	.doc, .ppt, .tex, .pdf	< 1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

1)Codes
The c++ microstructure simulation models will be made available through Github as a new module to be used within MOOSE (https://mooseframework.inl.gov/ and https://github.com/idaholab/moose) with the name DENDRITEPHASE, using the common standards applied by the MOOSE developers; different versions will be named using the year of release + a serie number, eg 2025.v1; the codes will be documented extensively; a pdf and tex document describing the implemented models and techniques will be added to the folder.
The matlab codes will be made available in a Mendeley repository accompanied with a README.txt file. Moreover, all scripts will have extensive documentation explaining the meaning and format of input and output data and giving a general overview of the structure of the program. It will also give references to relevant papers explaining the model or method implemented.
2) Raw simulation data will be collected in separate folders per simulation test including a txt file with a clear description of what the data represent and how they were generated. The input-files used for the simulation will be kept insight the same folder. The name of the folder will contain the composition, temperature and a reference to the loading conditions of the considered material. A .txt file explaining the naming will be maintained.
3) Outcome from data analysis:
Outcomes from data analysis will be saved together with the raw data from which they originate. The file names will start with a part that is similar to the name of the raw data files, and a suffix referring to the property that is derived will be added to the name used for the raw data set. The meaning of these names will be collected in a README.txt document and published together with the data.
4) Articles and presentations : the name of the file will contain the title of the article/presentation and the journal/conference name, a date and a version number.
In all cases, the METADATA will also contain links to relevant publications, as well as a relevant contacting address (nele.moelans@kuleuven.be).

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

In the metadata, the following information will appear.
1) inputs/outputs outline
2) composition, temperature, voltage, time

3) composition, temperature, voltage, time, analysed quantity
4) date, topic, version

3. Data storage & back-up during the research project

Where will the data be stored?

1) Codes: Local institutional drives + staging/archive on HPC cluster + when a relevant publication is accepted in a Mendeley repository linked to the paper + github repository DENDRITEPHASE linked to the public github repository of MOOSE
2) Raw data: HPC archive + key data sets on which publications are based on Mendeley
3) Outcomes from data analysis : Local institutional drives + staging/archive on HPC cluster + when a relevant publication is accepted in a Mendeley repository linked to the paper
4) Papers/presentations: Local institutional drives, published open access through green or gold route, a postprint will be archived in LIRIAS

How will the data be backed up?

The local institutional drives and HPC archive/staging foresee automatic back up of the data.
Postprints of published papers and pdf's of presentations will be archived in LIRIAS.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

We will buy additional archive storage space (1- 5 TB) for the raw simulation data of this project.
The size of the other data is sufficiently low to be stored at the institutional drives at no cost for the individual researchers.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The access to the institutional drives and HPC archive/storage folders related to this project is restricted to the members working on the project.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The cost for HPC archive/staging is 20 euro/TB per year (https://icts.kuleuven.be/sc/onderzoeksgegevens/HPC-storage).
The cost will thus be at most 100 euo/year. This will be covered from the project's working budget.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data will be preserved for at least 10 years after the end of the project.

Where will these data be archived (stored and curated for the long-term)?

On the same locations as during the project.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

For storage of the raw data on HPC archive, we expect a cost of 100 euro/year. It should be possible to recover this from follow-up projects.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

1) codes : the matlab codes will be made available through Mendeley, the c++ codes through a Github repository connected to the MOOSE software under CC-BY license. A unique identifier will be appointed to sets of code.
2) The simulation data themselves will be saved on archive space at the VSC-KU Leuven (https://icts.kuleuven.be/sc/onderzoeksgegevens/HPC-storage). Because of their huge size, it is difficult to find an appropriate place to make them publicly available. We will give people access upon request. When needed, we can put the data temporarily on a place with fast access. We will make them available under a CC-BY license. Reduced key datasets underlying a publication (up to 50 GB) will be made available through Mendely under CC-BY license (or in a repository advised by the journal if possible).
3) The outcomes from data analysis of the raw data will be stored on the archive folders at the VSC-KU Leuven in the same folders as the raw data from which they are derived. Those underlying publications will also be submitted to Mendeley together with the relevant key data sets. They will be publicly available under CC-BY license and have a unique indentifier (doi)
4) Papers will be publisehd following the gold or green route (by submitting a postscript to the KU Leuven repository LIRIAS). Pdfs of presentations will also be archived in the LIRIAS repository.

If access is restricted, please specify who will be able to access the data and under what conditions.

/

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

1) Matlab and python codes on Mendeley, the C++ codes building on MOOSE in a github repository connected to the github repository from which the source files for MOOSE software can be downloaded
2)Archive space at the KU Leuven HPC facilities
3) Archive space at the KU Leuven HPC facilities + Mendeley repository when used in a publications
4) LIRIAS (KU Leuven), open access publications

When will the data be made available?

1),2),3) : when a paper based on the data is accepted or if it is decided not to use the data for a publication.

4) gold route and presentations: immediately; green route : possibly after an embargo time specified by the journal (we will aim at embargo times not longer than 6 months)

Which data usage licenses are you going to provide? If none, please explain why.

CC-BY

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

When data are published on Mendeley, a DOI is automatically appointed.
Also for the data in GitHub we will request a DOI.

What are the expected costs for data sharing? How will these costs be covered?

The use of Mendeley and LIRIAS is at no cost for the researchers.
The raw simulation data (as well as a copy of all the other data generated within the project) will be archived on the VSC-KU Leuven (https://icts.kuleuven.be/sc/onderzoeksgegevens/HPC-storage) at a price of 20 euro per Terrabyte per year. The price also includes 'staging' space, to where the data can be copied (relatively fast) temporarily to use it in computations. It is possible to give external researchers a temporary vsc account so that they can access and use these data on the VSC, avoiding the time and difficulties associated with copying a large data set to external infrastructure. We expect that by the end of the project, we will need approx.5 Terrabyte, i.e. a cost of 100 euro per year. During the project, this cost will be booked on the project. Since this project will most probably feed into future related projects, we will foresee budget allocation in these projects for long-term storage of the data.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

The researchers under supervision of the PI Nele Moelans

Who will manage data storage and backup during the research project?

The researchers will be requested to uploaded regularly their data to the archive folder at the HPC. The researchers will also be request to publish the codes and data underlying a paper on Mendeley. The files will be verified by Nele Moelans

Who will manage data preservation and sharing?

Nele Moelans

Who will update and implement this DMP?

Nele Moelans