Biannual Report

June 30, 2021

Ganesh Raghavendran
PI : Dr.Shirley Meng

1 High resolution microscopy image processing using AI

1.1 Introduction

In this information age, the current focus has shifted from collecting data to processing it. Presently, there is no dearth of data available in any research field and the main challenge is in processing it to retrieve meaningful information. This is no exception in case of materials field, where research scientists try to extract information from their research samples, using high resolution microscopy imaging[3]. The post-processing of these images has always been a challenge, owing to the complexity of the high resolution image and the need for prior knowledge of components involved. Also, multiple images of same sample are required for complete study, which adds to the load of post-processing.

Transmission electron microscopy(TEM) is an effective and indispensable tool for material scientists, in understanding the physics behind the transport phenomena and kinetics involved in any system. However, the post-processing of TEM images has always been a challenge, owing to the complexity of the high resolution image and the need for prior knowledge of components involved. Also, multiple TEM images of same sample are required for complete study, which adds to the load of post-processing. With a focus on tackling the challenges of post-processing, I propose a python based tool equipped with deep learning to process multiple TEM images as a batch and with high accuracy. I have worked with the HRTEM(High resolution TEM) images of Li metal battery anode with magnification order of 300kX or more.

Image processing has always been used in deciphering the microscopy images. The microscopy images are processed in a micrograph software and are manually analyzed by material experts to identify the components involved in the sample. With more complexity involved in high resolution images, there is always a possibility of human error. AI may be a solution to this problem and more recently, there has been a renaissance in post-processing, owing to the success of deep learning.[4][5][2] However, there are lot of limitations in these studies. These studies use datasets of a specific sample[4]. There is no single model to process and interpret all high resolution microscopy images, means that a new model is required for each and every type of sample being imaged.

The ubiquitousness of the usage of TEM for characterization has resulted in more data than that can be easily processed[3]. Owing to this, many studies involving segmentation methods to process TEM images, have been published[4]. I also used UNet semantic segmentation in this project, to segment the FFT (Fast Fourier Transform) images and obtain features with minimum signal to noise ratio. Thus obtained datainformation processed to construct TEM diffraction graphs. In this project, we consider processing of the FFT(Fast fourier transform) images of the high resolution microscopy images instead of the original microscopy images themselves[4]. This makes a unique difference in getting the diffraction graph of the sample and thus the components involved. Using FFT image instead of the microscopy image itself, can help to proces microscopy images of varied magnification and resolution. Several datasets of the FFT images are obtained from the

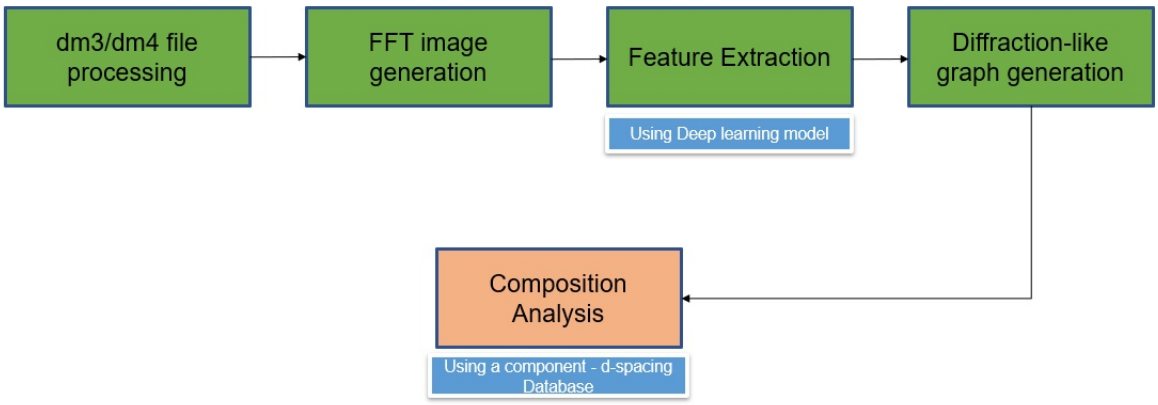


Figure 1: **Overview of the flow of the program :** The program flow from TEM image (.dm3/.dm4) file processing to composition analysis (Component distribution mapping)

python batch processing program developed by us, which uses the DM3Library to get microscopy image and its corresponding scale factor. Ground truth for these images are generated using Adobe photoshop and used to train the model to detect features of FFT image.

1.2 Project Goals

The proposed project aims to deliver a user-friendly python-based tool to process high resolution microscopy images and detect the components in the sample, with the processing rate more than 80 images/min. The critical success factors in achieving that goal include:

- Creating program module to process microscopy files and reconstructing TEM image
- Generating FFT image from the TEM image for feature extraction
- Creating a deep-learning model to identify features in FFT image
- Generating diffraction like graph with the detected features, using circle reintegration
- Creating/Obtaining the database of materials and its information
- Creating a heatmap of the component distribution on the TEM image

1.3 Program Development

For the initial study, we have considered HRTEM (High Resolution Transmission Electron Microscopy) images of Li-metal battery electrodes, with magnification factor of over

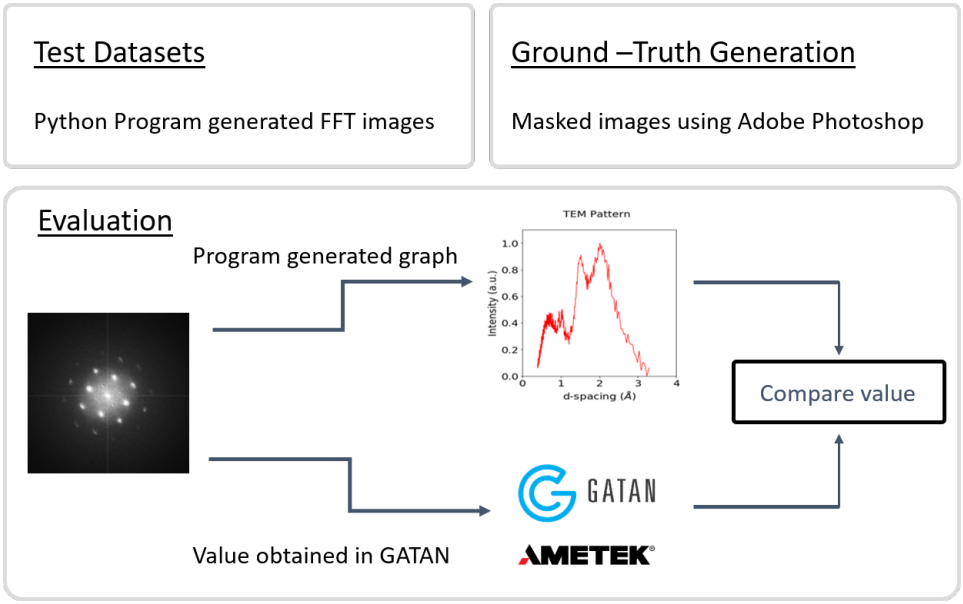


Figure 2: **Development of AI system for high resolution microscopy image processing :** Datasets are obtained from the python program and are used to design ground-truth in Adobe photoshop. Evaluation of the accuracy of the model is done by comparing the results obtained from our model and GATAN micrograph software

300kX. The .dm3/.dm4 files generated from GATAN micrography software are processed using the DM3Reader[1] library to extract TEM image of 4096x4096px dimension and it's corresponding scaling factor. The corresponding Reduced FFT (2048x2048px) image is generated using CV2 library and is denoised to match the quality of the image produced by the GATAN software.
The program flow is shown in Figure 1 and the feature detection in FFT images cannot be done using only image processing as noise in the image has a high chance of creating misinterpreted results(Fig.3). Owing to that fact, segmentation of the images are required to get an accurate detection of features in FFT image. Machine learning algorithms like Random forest classifier (Fig.4) and Deep learning models like UNet architecture (Fig.5) are also tested out and results are displayed.
The detected features are used to mask the FFT generated and corresponding diffraction-like graph is plotted. The d-spacing information is obtained from the peaks of the diffraction-like graph. The peak data are used for developing components distribution in the TEM sample, using a d-spacing database

1.3.1 Dice Coefficient:

Dice coefficient is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. Dice coefficient is a measure of precision and is used as the metric for our model optimization. Dice coefficient is given by the formula:

$$\text{Dice coefficient} = \frac{2 \text{ x Area of overlap between the predicted and ground truth}}{\text{Total area of predicted + ground truth}} \quad (1)$$

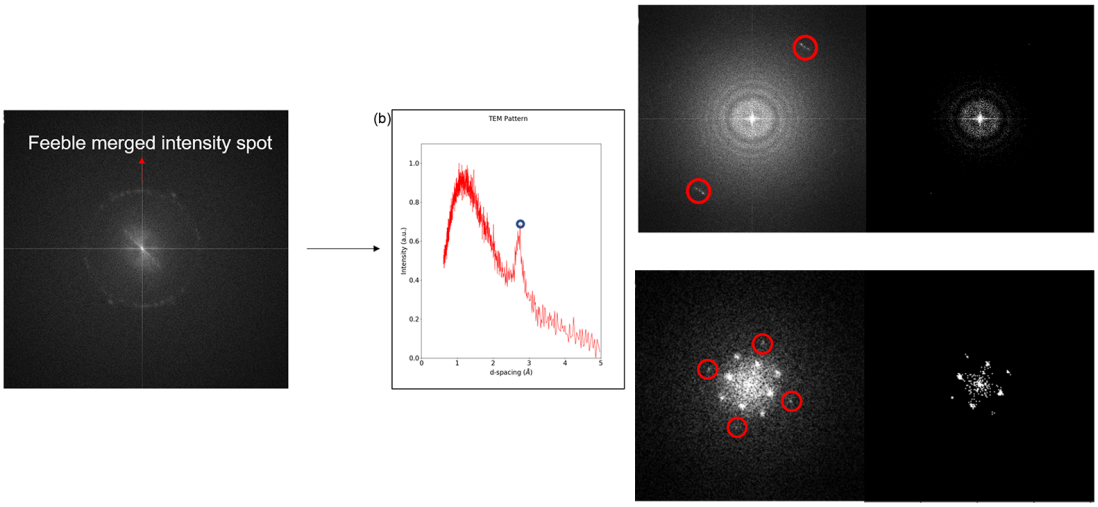


Figure 3: **Challenges in image processing :** From the left image, we can see that the noise in the FFT image has misinterpreted the highest peak (actual highest peak marked in blue circle). In both the images, some of the required features get removed by using normal noise removal techniques(marked in red circle)

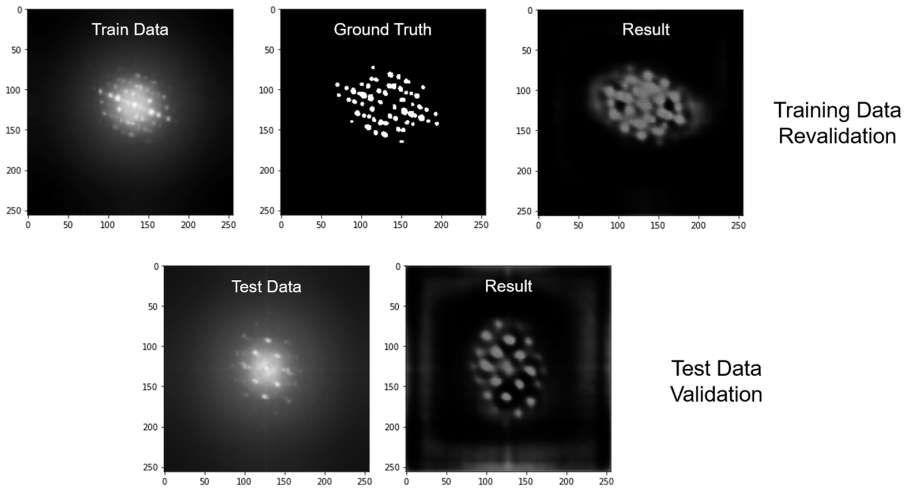
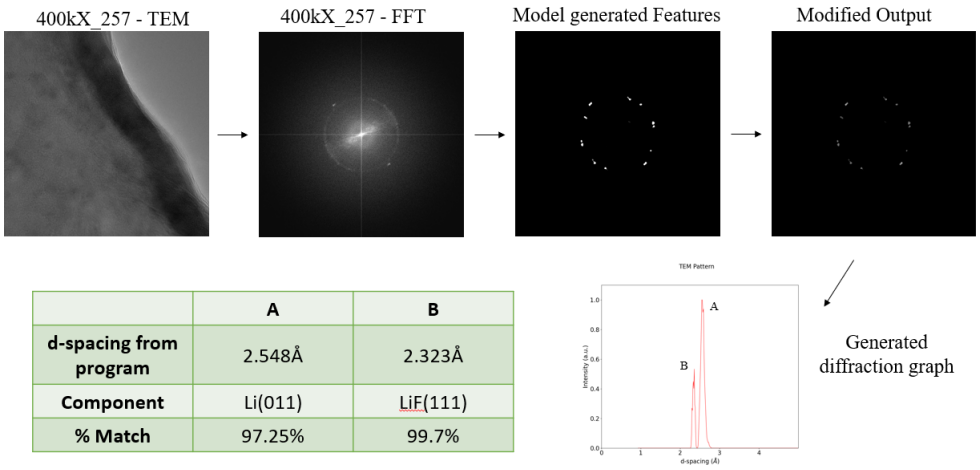


Figure 4: **Segmentation using UNet Architecture :** 60 images augmented to 413 images are trained in UNet model. The results obtained in the UNet segmentation model is more accurate than the random forest classifier but has more dispersed/fuzzy features

1.3.2 Initial testing:

A cryo-TEM .dm4 image of cycled Li metal sample of 400kx magnification was chosen for the initial testing (Fig.5). A UNet model trained with 1024px images and 512 filters was used for developing the model for detecting the FFT features. The complete program was developed using python 3.x and tensorflow 2.3.x. IFFT was developed from the detected features and watershed algorithm was used for getting the component distribution. A rudimentary d-spacing database was used to acquire the component information for the

corresponding peak position from the diffraction graph. The database used is developed from room temperature TEM image acquisition and analysis.



	A	B
d-spacing from program	2.548Å	2.323Å
Component	Li(011)	LiF(111)
% Match	97.25%	99.7%

Figure 5: **Program flow with an example data :** Program flow demonstrated with a cycled Li metal sample of 400kx magnification. Diffraction-like graph and d-spacing data are generated using the program, for a given TEM image

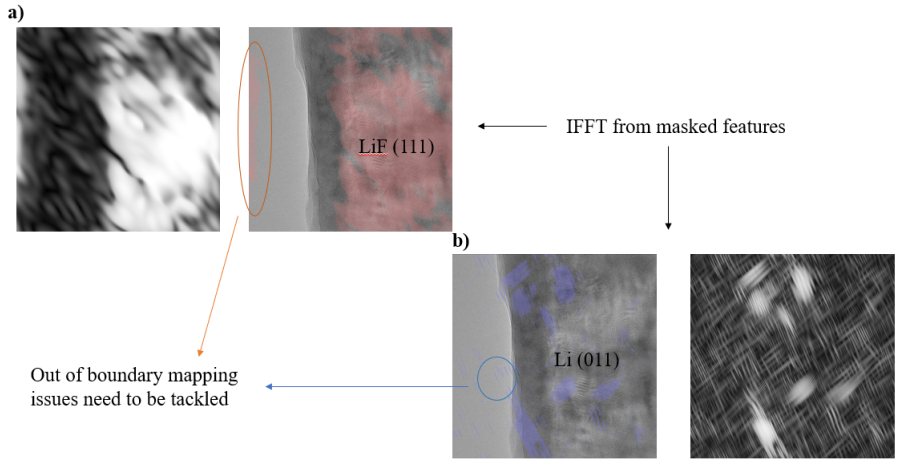


Figure 6: **Component distribution mapping :** Mapping of the crystalline components in the TEM image used in Fig.5 - a) LiF (111) and b) Li(011). The IFFT generated of the respective component is shown for comparison and the marked areas are the regions where out-of-boundary mapping is observed

Some of the main challenges involved are:

- The inability to process multiple files (not more than 4 at a time) due to bad memory allocation, because of deep learning model usage.
- Component mapping is not always perfect (Fig.6) and some incorrect mapping outside the sample boundary was observed

1.4 Future Plan:

As previously discussed in the section 1.3.2, the challenges will be tackled to obtain an optimized program with high throughput. Once fixed the program flow, it will be tested with multiple TEM images of magnification greater than or equal to 300kx magnification. A more detailed database will be generated to enable the usage of program not only for Li-based TEM samples but for more eclectic samples. To facilitate better user experience, a GUI as shown in Figure 7 will be developed.

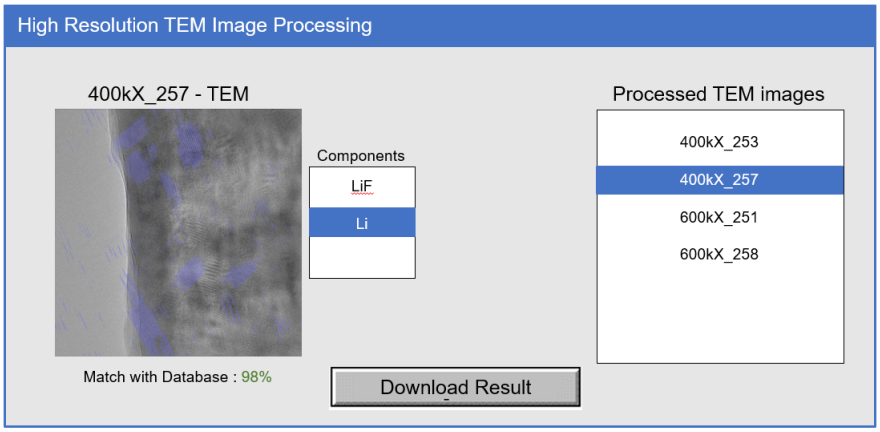


Figure 7: **GUI for TEM image Analysis :** Design concept for a TEM image analysis GUI

2 Liquified Gas Electrolytes - Na metal anode

2.1 Introduction

As global societies increase their dependence on energy, the demand for energy, consequently, increases. This increasing requirement of energy has led to geopolitical disagreements and even social instability. Present limitations of energy resources and climate concerns have spiked interest in more effective, efficient, and green means of power generation. However, this problem cannot be just solved by research on potential alternatives, but improvements in energy storage must be considered as a top priority.

LIBs (Lithium ion batteries) are gaining popularity and being used in EVs and grid storage applications. Despite being a popular battery chemistry, LIBs can't satisfy the energy demands for all sectors, owing to it's limited availability and uneven distribution. Sodium metal, being an eco-friendly, sixth most abundant alternative with high specific capacity of 1166 mAh/g, is a promising alternative for the grid-scale energy storage.

To enable Na metal, high rate capability and capacity are necessary. Main challenge in enabling Na metal anode is finding a suitable electrolyte. $NaPF_6$ (sodium hexafluorophosphate) in glymes (mono-, di-, and tetraglyme) and $NaBF_4$ in glymes show promising results with over 99.2% and 90% CE at 1-2mA cm^{-2} respectively. However, with a delicate balance between electrochemical stability, ionic conductivity, temperature, and safety, and also a need for grid-scale storage applications at lower temperatures, liquified gas electrolytes can be a promising alternative with higher conductivity, safety and lower viscosity. In this report, we will be focusing on NaFSI in dimethyl ether.

2.2 Experimental

2.2.1 Na metal soak test

Na metal soak test was done using a window cell with 1/4" punched sodium metal soaked in 1M NaFSI in Dimethyl ether. Two batches of soak test was done - one for 3 days and other for 6 days. The soaked sodium pieces were observed for color changes and was cut to check the SEI growth. It was found that minimal color change was observed on the surface of the soaked Na metal pieces and no visible SEI layer was observed in the cross-section (Fig), confirming the compatibility of the sodium with the electrolyte (1M NaFSI in Dimethyl ether) chosen.

2.2.2 Initial Electrochemical testing:

All Na-metal plating and stripping experiments were performed in the custom high-pressure stainless-steel coin cell. For Li-metal half-cell plating and stripping tests, a stainless-steel 316L working electrode was used, which was mechanically polished. In mechanical polishing, several grits were used to achieve a flat substrate, starting at 80 grit and ending at 1600 grit. After polishing the working electrode, the cleaned stainless-steel coin cell parts were transferred to the glovebox for assembly.

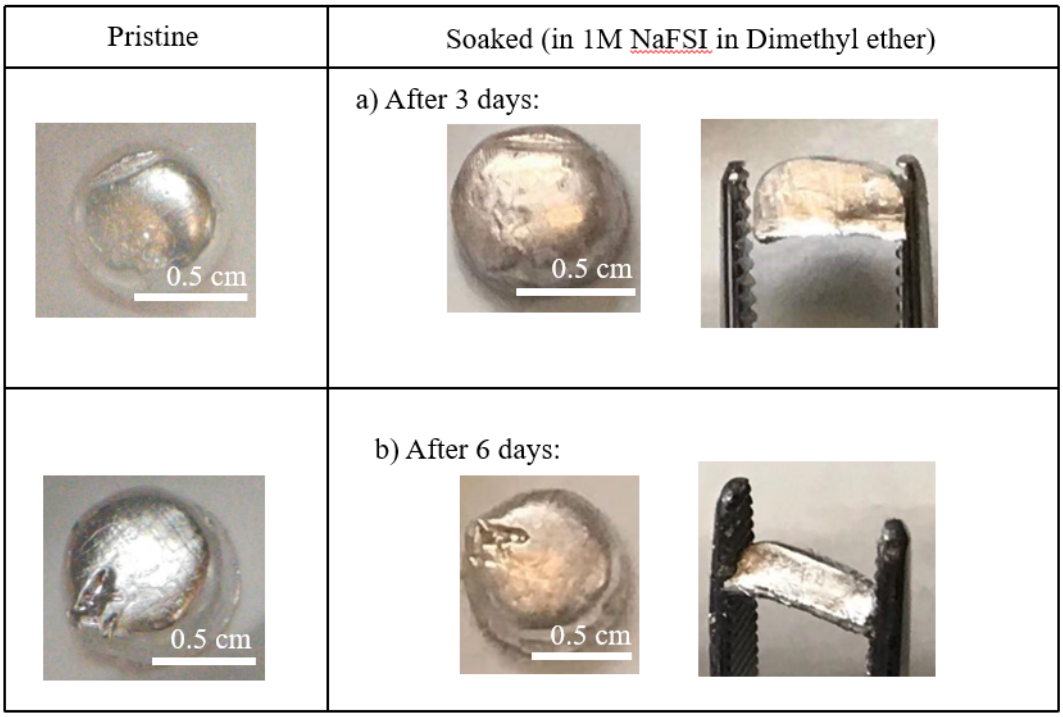


Figure 8: **Na metal soak test :** Na metal soaked in 1M NaFSI in Dimethyl ether for a) 3 days and b) 6 days, and checked for color change and SEI growth

A Na-metal half-cell was prepared using 1 mm thick Na-metal, scrapped clean with a glass slide and Na-metal rolled on copper foil. A 20 μm thick Celgard separator was placed between the Na-metal/Na-Cu and the aluminum current collector. To begin, half-cell cycling was performed with the 3cc of 1M NaFSI in Dimethyl ether electrolyte. The cell was given 10 hours of rest time after gas filling procedures to allow the salts to completely solubilize. Long term cycling was performed. The cells were cycled at a current density of 0.5 mA/cm^2 and at a capacity of 1 mAh/cm^2 . The Coulombic efficiency of the battery was calculated through the Na stripping capacity divided by the Na plating capacity during a single cycle and was kept consistent through this report.

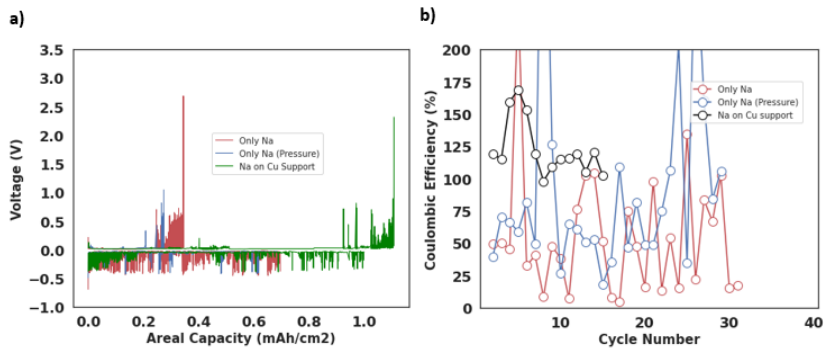


Figure 9: **Na-metal plating stripping in 1M NaFSI/Me_2O electrolyte :** Electrochemical data for plating stripping at 0.5 $mAcm^{-2}$ a) Voltage curve b) Coulombic Efficiency

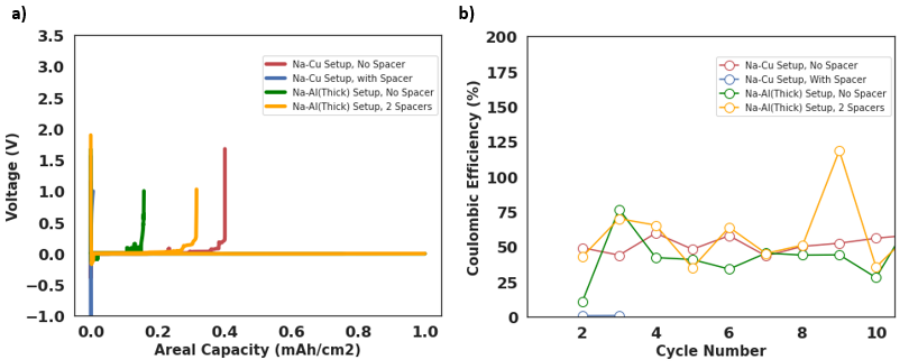


Figure 10: **Na-metal plating stripping in 1M $NaBF_4$/TEGDME electrolyte :** Electrochemical data for plating stripping at 0.5 $mAcm^{-2}$ a) Voltage curve b) Coulombic Efficiency

Additional Na-metal half-cell cycling tests were performed with 1M NaBF4 in TEGDME. Different cell setup with components like spacers, thick aluminum current collector (76 μm) were tried to tackle the contact issue observed in the cycling tests.

#	Anode	Cathode	Electrolyte System	Cycling Condition	Spacer	Notes
1	Na (5/16")	Thin Al foil (3/8")	1M NaFSI in Me_2O	**Plating** : 0.5 mA/cm^2 for 2h **Stripping** : 0.5mA/cm^2 till 1V	No	High Contact Resistance. Stripping voltage >1 V
2	Na/Cu (3/8")	Thin Al foil (3/8")	1M NaFSI in Me_2O	Same as 1	No	Better contact but still noisy and stripping voltage >1V
3	Na/Cu (3/8")	Thin Al foil (3/8")	1M NaBF4 in TEGDME	Same as 1	No	Similar to #2
4	Na/Cu (3/8",1/4")	Thick Al foil (3/8" , 1/4")	1M NaBF4 in TEGDME	Same as 1	No	Similar to 2,3 but stripping voltage ~1V
5	Na/Cu (3/8", 1/4")	Thick Al foil (3/8" , 1/4")	1M NaBF4 in TEGDME	Same as 1	Yes	Worse than #2,#3,#4

Figure 11: **Na metal electrochemical testing summary :** Summary of different parameters at which Na metal electrochemical testing was carried out

2.2.3 Future Plan - LGE cell modification

The contact issue observed in the cycling results shown are due to the improper contact between the current collectors and electrodes. To tackle the contact issue, inspired by the pressure cell setup, the LGE cell will be modified to provide adequate and uniform contact between the electrodes.
The cell component are made from 316L corrosion resistant stainless steel and are 0.1" thick to make them compatible with the existing cell design. 1/2" 2325 celgard will be used for this setup. The top current collector PTFE support length is reduced by approximately 0.3" to make the setup fit inside the cell. The cell components (Fig.12) are attached to the screws extending from the current collectors in both anode and

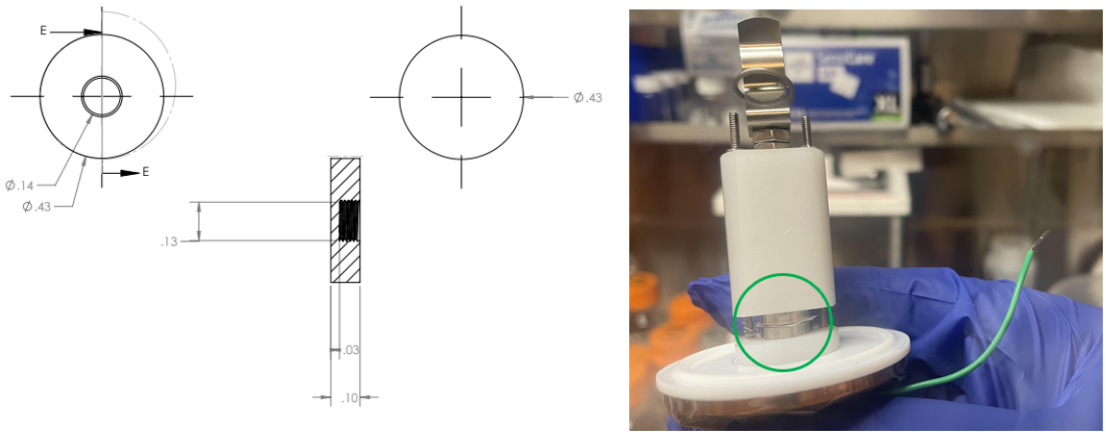


Figure 12: **LGE cell current collector modification :** LGE cell current collectors are modified by adding a cylindrical component (left image) of 0.43" radius, to provide higher and uniform area of contact for electrodes (marked in green circle in right image), and thus tackling the contact issue

cathode side to create a system as shown in figure. The performance of this modified cell setup will be checked and further study will be done.

3 Molecular Dynamics Simulations

3.1 Introduction

Prior to the computing age, the properties and behaviour of a system was studied solely based on the experiments and observations. With the advent of high performance computers, we can study the macroscopic systems at molecular scales. For the study of the fundamental building blocks of a system, the problem is approached as an N-body problem for which a general analytical solution doesn't exist and the behaviour of individual molecules should be considered. Molecular Dynamics (MD) simulations is a tool which enables scientists to tackle such problems.[art of md]

At the molecular level, MD simulations are usually carried out to get the most energy efficient system at given conditions. With no proper ordered structure and definite inter-particle interactions, liquids remain to be the most studied state of matter through MD simulations. The interactions or components that contribute to the total potential energy E_{total} of the system is given by,

$$E_{total} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{non-bonded} \tag{2}$$

where $E_{bonds}, E_{angles}, E_{dihedrals}$ are energy contributed to the system due to molecular bonds, bond angles and dihedral plane angles respectively. $E_{non-bonded}$, which corresponds to energy contributed due to interactions between non-bonding particles, is the most computationally intensive parameter to calculate. Non-bonded particle interactions is estimated using a Vanderwaals force parameter, lennard-jones potential V_{LJ} which is given by,

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3}$$

where r is the distance between two interacting particles, ϵ is the depth of the potential well and σ is the distance at which the particle-particle potential energy V is zero.

3.1.1 Force fields

GAFF - General Amber Force Fields:
GAFF was developed for organic molecules, compatible with existing AMBER force fields for nucliec acids and DNA. It has paramaters for most of the organic molecules composing H,C,N,O,P,S and halogens. It uses 'Class I' simple model for describing the the structures and nonbonded energies for organic and bioorganic systems -

$$E_{pair} = \sum_{bonds} k_r(r - r_{eq})^2 + \sum_{angles} k_\theta(\theta - \theta_{eq})^2 +$$
$$\sum_{dihedrals} \frac{v_n}{2}[1 + cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{_{ij}} \right]$$

where, r_{eq} and θ_{eq} are equilibration structural parameters; K_r, K_θ, V_n are force constants; n is multiplicity and γ is the phase angle for the torsional angle parameters. The A, B, and q parameters characterize the nonbonded potentials

OPLS - Optimized Potentials for Liquid Simulations:
The OPLS was developed to get good description of the organic molecules, with the force field focusing on both intramolecular terms for bond stretches, angle bends, and torsions, as well as the intermolecular and intramolecular nonbonded interactions. Bond stretch, angle bend, and torsional terms are adopted from the AMBER united-atom force field and the intramolecular interactions are focused in this force field type. The intermolacular nonbonded interactions between a and b are given by:

$$\Delta E_{ab} = \sum_{i}^{ona} \sum_{j}^{onb} \left(\frac{q_i q_j e^2}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right) \tag{4}$$

The intramolecular interaction of the nonbonded particles are same as the intermolecular interactions for all pairs of sites separated by more than three bonds

Polarizable Force Field:
For a more accurate and reliable force fields, many body interactions has to be taken into considerations, such as the polarizable dipoles. The polarizable force fields can be modeled using:

Inducible point dipole model : In this model a point dipole p_{ind} is considered to be induced by the electric field E^0 due to atomic charges and electric field E^p due to developed dipoles.

$$p_{ind} = \alpha(E^0 + E^p) \tag{5}$$

The contribution of the polarization energy E_{pol} to the total non bonded energy is given by the summation of all the induced dipole energies,

$$E_{pol} = -1/2 \Sigma p_i . E_i^0 \tag{6}$$

Fluctuating charge model : In this model, rather than having fixed induced dipoles, the charges are considered to be fluctuating due to its environment until instantaneous electronegativity equilibrium is achieved. For this model, the fluctuating charges are given fictitious masses and are treated as additional degrees of freedom in the equations of motion. It is more computationally expensive than the induced dipole model.

3.2 Current Plan

As initial training, MD simulations will be carried out for pure Dimethoxy ethane (DME) solvent. The resultant density of the system will be verified for the validity of the procedure followed. The simulation box (Fig.13) was created using packmol package. The simulation box contains 200 DME molecules inside a cube of side 154Å and space tol-

erance of 2Å. General Amber Force Field (GAFF) data will be used for calculating the force field parameters. For the above system, an initial energy minimization at 0K (en-

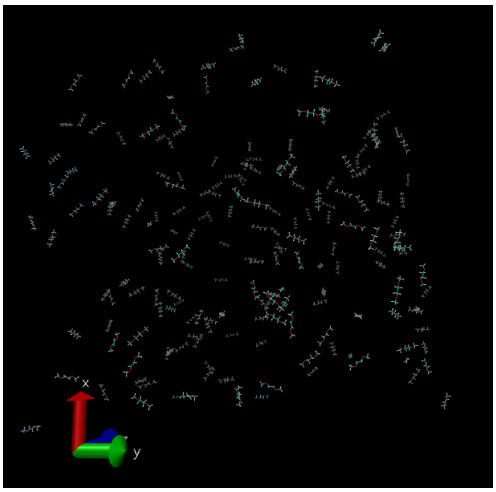


Figure 13: **Simulation box for pure DME :** Simulation box developed by using packmol package, with 200 DME molecules inside a cube of side 154Å and tolerance of 2Å. Molecule obtained using OpenBabel and visualisation is done using VMD software

ergy and force tolerances of 104) will be performed to obtain the ground-state structure. After this, the system will be slowly heated from 0K to room temperature at constant volume over 0.2ns using a Langevin thermostat, with a damping parameter of 100ps. The system will then be subjected to five cycles of quench-annealing dynamics, where the temperature will be slowly cycled between 298K and 894K over 0.8ns in order to eliminate the persistence of any meta-stable states. After annealing, the system will be equilibrated in the constant temperature (298K), constant pressure (1bar) (NpT ensemble) for 0.5ns before finally subjected to 5ns of constant volume, constant temperature dynamics.

4 Acknowledgement

I would also like to thank Professor Shirley Meng for the opportunity to work in LESC and for her training that allows me to grow as a researcher. Additionally, I would like to thank Prof.Tod Pascal for providing resources for training the deep learning model. Further, I would like to thank my ACE and LGE team members: Minghao, Weikang, Dylan, Ryo, Bingyu, Jackie, Baharak, and Alex. Jackie provided me with thorough initial trainings and is a reliable teammate. Outside of the LGE subgroup I would also like to thank John Holobeuk for his assistance with Molecular Dynamics.

References

[1] Dm3 python reader - https://bitbucket.org/piraynal/pydm3reader.

[2] G. Güven and A. B. Oktay. Nanoparticle detection from tem images with deep learning. pages 1–4, 2018.

[3] Mulholland G. Persson K. Seshadri R. Wolverton C. Meredig B Hill, J. Materials science with large-scale data and informatics: Unlocking new opportunities. *MRS Bulletin*, 41(05):399–409, 2016.

[4] Zakharov D. N. Mégret R. Stach E. A. Horwath, J. P. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *Npj Computational Materials*, 6(1), 2020.

[5] Philipp Fischer Olaf Ronneberger and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv.org*, 2015.