
ACCURATE PREDICTION OF VOLTAGE OF BATTERY ELECTRODE MATERIALS USING ATTENTION BASED GRAPH NEURAL NETWORKS

A PREPRINT

Steph-Yves Louis, Edirisuriya M. Dilanga Siriwardane

Department of Computer Science and Engineering
University of South Carolina
Columbia, SC, 29201, USA

Rajendra P. Joshi*

Pacific Northwest National Laboratory
Richland, WA, 99352, USA
rajendra.joshi@pnnl.gov

Sadman Sadeed Omeel

Department of Computer Science and Engineering
University of South Carolina
Columbia, SC, 29201, USA

Neeraj Kumar

Pacific Northwest National Laboratory
Richland, WA, 99352, USA

Jianjun Hu †

Department of Computer Science and Engineering
University of South Carolina
Columbia, SC, 29201, USA
jianjunh@cse.sc.edu

January 1, 2022

ABSTRACT

Performing first principle calculations to discover electrodes' properties in the large chemical space is a challenging task. While machine learning (ML) has been applied to effectively accelerate those discoveries, most of the applied methods ignore the materials' spatial information and only use pre-defined features: based only on chemical compositions. We propose two attention-based graph convolutional neural network techniques to learn the average voltage of electrodes. Our proposed method, which combines both atomic composition and atomic coordinates in 3D-space, improves the accuracy in voltage prediction by 17% when compared to composition based ML models. The first model directly learns the chemical reaction of electrodes and metal-ions to predict their average voltage, whereas the second model combines electrodes' ML predicted formation energy (E_{form}) to compute their average voltage. Our models demonstrates improved accuracy in transferability from our subset of learned metal-ions to other metal-ions.

Keywords Electrodes · Graph neural network · Voltage · Deep learning

1 Introduction

Batteries are the dominant source of energy for diverse applications and main work-horse for portable electronics [1, 2]. Common examples where batteries are increasingly adopted are electric vehicles and grid energy storage [3, 4]. Besides their wide use, there are still big interests in improving these batteries' performance for more improved reliability in devices demanding large energy density. But to develop next-generation batteries, accurate and efficient

*Corresponding author: R.PJ

†Corresponding author: J.H. (<http://www.cse.sc.edu/~jianjunh>)

exploration of large chemical space is necessary and predicting their performance represents the first step towards this goal. Traditionally, the exploration of these batteries or electrodes’ properties was done using time-consuming physics based simulations [5] and/or by using resource-intensive experiments [6]. Mainly, examining each material in the large chemical space while searching for robust electrodes imposes great difficulties with these traditional methods. Hence, machine learning has been used as an alternative for their impressive performance and are increasingly adopted in the battery community for predicting the performance metrics of battery components including intercalation potentials or voltage [2, 7–14].

Recently multiple ML-based approaches to predict the voltage of electrodes materials have been used [8, 15, 16]. Such approaches include models based on ML-potentials and also other simple deep neural network, which are the most accurate. ML-potentials based models are trained to learn the potential energy surface of solids using data from physics based simulations [8]. Though promising, the accuracy of these ML-potentials is limited to particular types of materials with specific atomic compositions and have weak transfer-ability. Another important limitation to these ML-potentials is that generating the dataset for their model from density functional theory (DFT), for each composition space, is extremely challenging. For instance, Viswanathan et al. adapted a ML-potentials to predict voltage, only for Li-graphite based electrodes, using a relatively small DFT based dataset [8]. While the accuracy of Viswanathan’s model compares to that of DFT simulations for Li-graphite based electrodes, its transfer-ability to Na-graphite or K-graphite is impractical. This impracticability results from the fact that the Li-graphite specific ML-potentials model performs poorly on other electrodes type. Since such data does not exist for each materials for other metal-ion battery, ML models that will work well with diverse metal-ion batteries and transfers equally well is necessary.

For other models including deep neural network, ML-models are trained on nearly 5000 electrodes materials from materials project database with density functional theory calculated voltage as target [15, 16]. Nonetheless, these models have poorer predictive ability than other ML-models used for properties of solids. [17] This poor performance can be attributed to both the small size of the dataset and chemical diversity within such small dataset for the different metal-ion batteries. Also, more advanced ML methods, such as graph based ML approaches, had not been applied yet in the domain of electrode materials. The majority of existing literature mostly employs pre-defined calculated features as input information for the materials. Solely relying on these pre-calculated atomic composition features implies that the model ignores any 3D-structural environment information within the materials and thus misses the key attributes used in reference physics based simulations [15, 16]. However, we anticipate that combining both chemical composition along with some 3D-structural information inside of a more complex deep learning method may lead to a significantly more accurate voltage prediction; as it has been shown for other crystals’ properties predictions [17–19].

In this work, we adapt a graph-convolutional neural network that learns the robust representation of electrodes materials from the atom types and corresponding 3D-coordinates only. We show that, our method improves the accuracy in voltage prediction by nearly more than 17 % when compared to other works in the literature. Our proposed method shows great transferability to new metal-ion battery chemistry as it outperforms all related published works on voltage prediction by more than 20 %. Our voltage-prediction solution includes two techniques. The first consists of a model which learns the chemical reaction of input electrodes and outputs their average voltage. The other technique involves using a trained model that predicts E_{form} of individual electrodes. These electrodes’ E_{form} are subsequently used within our derived formula to output the voltage-prediction. Comparing our performance to those found in the literature, we show that calculating the voltage from formation energy of electrodes is ideal way to predict voltage in the scenario where there is limited data to train ML models for intercalation reactions.

2 Methods

2.1 Data

DFT computed voltages and structures of electrodes materials for 4,402 battery systems were collected from the Material Project (MP) database using Pymatgen Materials Genomes (pymatgen) [20]. The distribution of the dataset, which consists of ten different metal-ions (Cs, Y, Zn, Ca, Li, Mg, Na, K, Al, and Rb) batteries is shown in Fig. 1 [21]. Due to its high popularity as charge carrier, Li (2291) has the highest number of battery systems. Ca-based systems (484) appear in the dataset more than Mg (393), Na (328), and Zn (385) based electrodes. The lowest numbers in Fig. 1 are available from Rb (50) and Cs (39).

To develop the two proposed methods below described, we use the previously mentioned dataset of 4,402 electrodes for the chemical reaction-based model and another dataset of about 60,000 materials downloaded from the MP database for E_{form} -based model [22]. The two datasets are mutually exclusive of their materials. For each dataset, we split the data into 3 sets: with 85% used for training, 7.5% for testing, and 7.5% for validation.

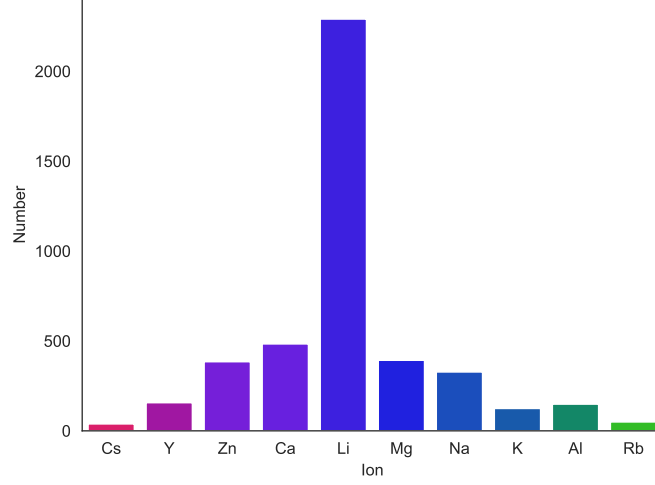
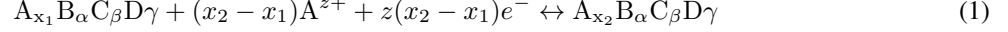


Figure 1: Distribution of the numbers of battery frameworks collected from the MP database for each metal-ion.

2.2 Voltage

Understanding the chemical reaction of an intercalation battery framework is essential to learn the voltage of corresponding system. As an example, intercalation reaction of the hypothetical layered material $A_{x_1}B_{\alpha}C_{\beta}D\gamma$ and A ions to form $A_{x_2}B_{\alpha}C_{\beta}D\gamma$ can be represented as follows.



Here, the electrode at the left-hand side, which reacts with the cation, exhibits higher potential than the electrode on the right-hand side. Thus, we label the left-hand and right-hand side electrodes as high-potential and low-potential electrodes, respectively. To calculate the voltage in DFT, we estimate the Gibbs free energy of individual electrodes defined as $G = \Delta E + P\Delta V' - T\Delta S$, where ΔE is the internal energy change, P is the pressure, $\Delta V'$ is the volume change, T is the temperature and ΔS is the entropy difference of the system. However, $P\Delta V' \approx 10^{-5}$ eV and $T\Delta S \approx 25$ meV at room temperature. Therefore by neglecting those two terms, we are able to calculate the voltage (V) by only considering the internal energy change as shown in Eq.2. Here, the terms $E[\eta_{x_i}](i = 1, 2)$ are the total energy of the chemical formula (η) with x_1 and x_2 contents of the ion (A), and z is the valency of the intercalating metal-ion. For instance, $\eta_{x_1} = A_{x_1}B_{\alpha}C_{\beta}D\gamma$ and $\eta_{x_2} = A_{x_2}B_{\alpha}C_{\beta}D\gamma$ for the chemical reaction of Eq.1. For the ions mentioned in Fig. 1, $z = 1$ for Li, Na, K, Rb and Cs, $z = 2$ for Ca, Mg and Zn, and $z = 3$ for Al and Y.

$$V \approx \frac{1}{z(x_2 - x_1)e} (E[\eta_{x_1}] - E[\eta_{x_2}] + (x_2 - x_1)E(A)) \quad (2)$$

$$E_{\text{form}_{x_i}} = E[\eta_{x_i}] - x_i E(A) - \alpha E(B) - \beta E(C) - \gamma E(D) \quad (3)$$

The E_{form} per unit formula of $A_{x_1}B_{\alpha}C_{\beta}D\gamma$ electrode is given by Eq.3, where $E(A)$, $E(B)$, $E(C)$ and $E(D)$ represents the energy of each atom in their bulk phase. By computing the difference between formation energies of high and low electrodes, we can show that voltage can be determined using Eq. 4.

$$V \approx \frac{1}{z(x_2 - x_1)e} (E_{\text{form}_{x_1}} - E_{\text{form}_{x_2}}) \quad (4)$$

However with ML, we can calculate the voltage in two different ways. (I) By training the ML models directly to learn the voltage and (II) by training ML models to learn the formation energy of involved high-potential and low potential electrodes. We compare the performance of model trained on each case. For the former case, input to the ML models is chemical reactions (hence labelled reaction-based model) whereas for the latter input is the structure of each

compounds to learn the corresponding formation energy (labelled E_{form} -based model). The voltage is then calculated from formation energies using Equation 4.

2.3 Graph Neural Network Architecture

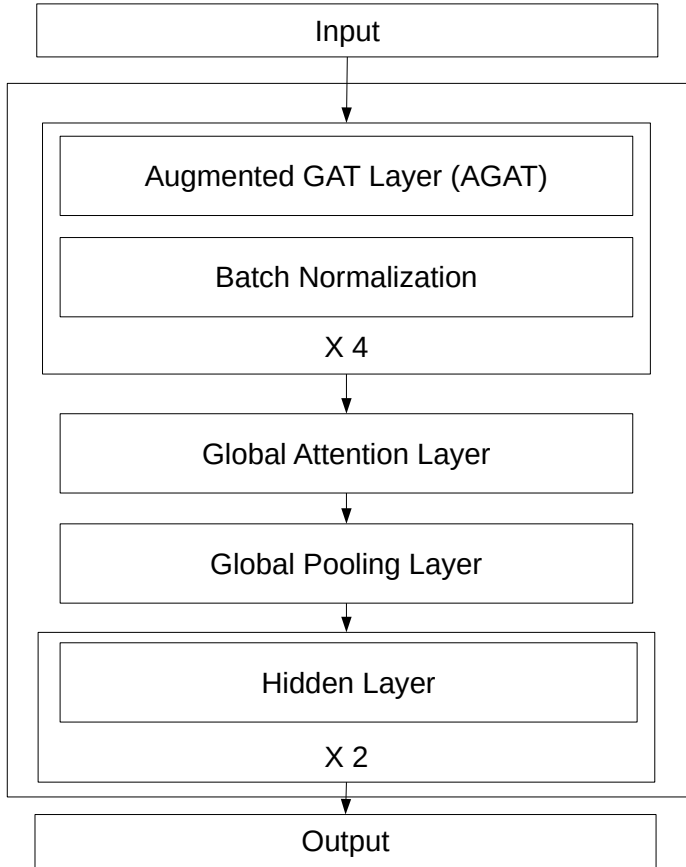


Figure 2: An overview of the GATGNN architecture.

Graph neural networks (GNN) are deep neural networks that have been applied to effectively learn latent features from network or graph data [23]. Our models are based on a subset of these GNNs which adapt the technique of the attention mechanism to GNN [24–26]. Particularly, we adapt the technique of GATGNN introduced by Louis et. al. [27] for this study. Compared to other GNN applied in the domain of materials, GATGNN learns each atom’s contribution both locally (within a local atomic space) and globally (with respect to all atoms in the material) [17, 18, 28, 29]. Particularly, our approach efficiently captures the atoms’ local importance through its augmented graph attention layers (AGAT) and then the atoms’ global importance using its global attention layer. For the local attention, our model consist of 4 AGAT layers of 4 attention-heads, each consisting of 64 neurons. The local soft-attention $\alpha_{i,j}$ between a node i and a neighbor j can be represented by the as:

$$\alpha_{i,j} = \frac{\exp(a_{i,j})}{\sum_{k \in N_i} \exp(a_{i,k})} \quad (5)$$

N_i denotes the neighborhood of node i and $a_{i,j}$ is the parameterized weight coefficient between nodes i and j , which denotes the importance of node j to node i in in Equation 5.

For the global attention, we use a single global composition layer. The atom’s global attention is learned from the encoded atom’s type [27]. In our model, each atom communicates with the sixteen nearest neighboring atoms, to create a 16×9 -dimensional feature vector for each bond, and each atom is embedded by a 16 dimensional feature vector [27].

The global attention g_i can be described as following:

$$g_i = \frac{(\mathbf{x}_i \parallel \mathbf{E}) \cdot \mathbf{W}}{\sum_{x_c \in \mathbf{X}} (\mathbf{x}_c \parallel \mathbf{E}) \cdot \mathbf{W}} \quad (6)$$

$\mathbf{x} \in \mathbb{R}^F$ denotes a learned embedding, \mathbf{E} denotes a compositional vector of the crystal, $\mathbf{W} \in \mathbb{R}^{1 \times (F+|E|)}$ denotes a parameterized matrix, and x_c denotes the learned embedding of any atom c within the crystal in Equation 6.

The overall architecture of our GATGNN is shown in Figure 2. We implement all the components in our proposed approach using deep learning libraries of Pytorch and the library of Pytorch-Geometric [30, 31]. The same loss function SmoothL1 loss function is used to train both models [32].

2.4 Chemical Reaction-based Voltage Predictor

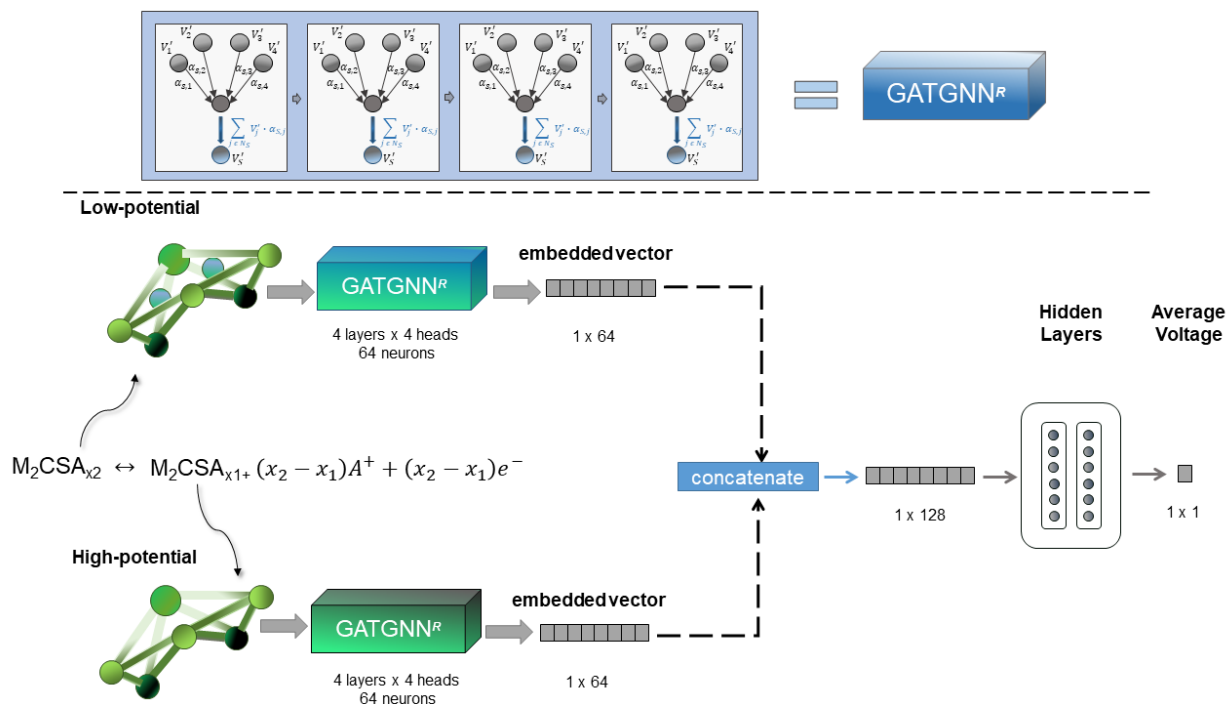


Figure 3: Architecture of our reaction-based average voltage model. Top panel shows the underlying GATGNN modules used in the work.

Due to recent advances in ML models, they have been increasingly and accurately adopted for learning the properties of chemical reactions of molecules[33, 34]. These models however have not been used for reactions involving crystals. Here, we develop a graph neural network model which considers the chemical reaction of electrodes and metal-ions as the input. Our proposed reaction-based model consists of 2 graph neural networks arranged in parallel, which then are both followed by series of hidden fully connected layers. Each graph neural network (each GATGNN) takes for input the structure of an electrode encoded as a graph and outputs a single embedded vector representation of an electrode. Figure 3 illustrates the framework of our proposed model.

For the reaction based model, we feed each involved electrode through GATGNN block and concurrently learn from corresponding low and high potential electrodes. Following the graph convolutions, the output of both blocks are subsequently concatenated into a 128 dimensional vector to be fed to 2 fully connected layers. The final predicted average voltage is calculated by learning the chemical interaction of the two electrodes. Hyperparameters were optimized for all the models used in this work. We train the model for 500 epochs with early-stopping, using a learning rate of 1×10^{-3} , weight-decay of 5×10^{-3} , and batch size of 128.

2.5 Formation Energy-based Voltage Predictor

Our proposed E_{form} -based model is based on an optimized pre-trained model of a GATGNN of 4 layers 128 neurons with 4 attention heads to model crystals to their E_{form} . As illustrated in Figure 4, the proposed method consists of 1

GATGNN which independently outputs the E_{form} for low and high potential electrode. This optimized formation-energy GATGNN model was trained for 300 epochs early-stopping, using a learning rate of 1×10^{-3} , weight-decay of 5×10^{-3} , batch size of 64. Following the prediction of the electrodes' E_{form} , their average voltage is subsequently obtained using Equation 4.

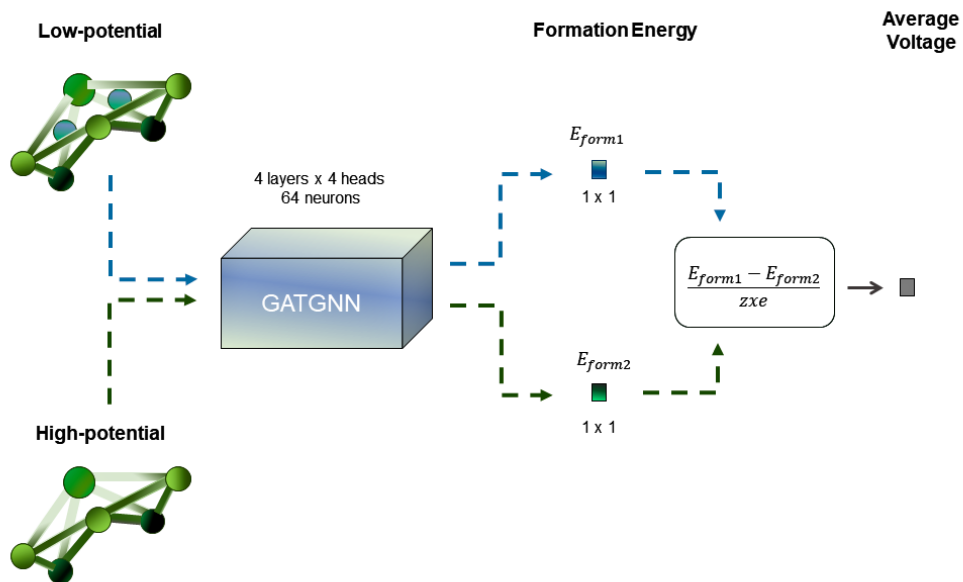


Figure 4: Architecture of E_{form} -based average voltage model

3 Results and Discussion

The accuracy of the trained models was investigated using MAE and R^2 as the statistical measures.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N (V_i^{\text{DFT}} - V_i^{\text{ML}}) \quad (7)$$

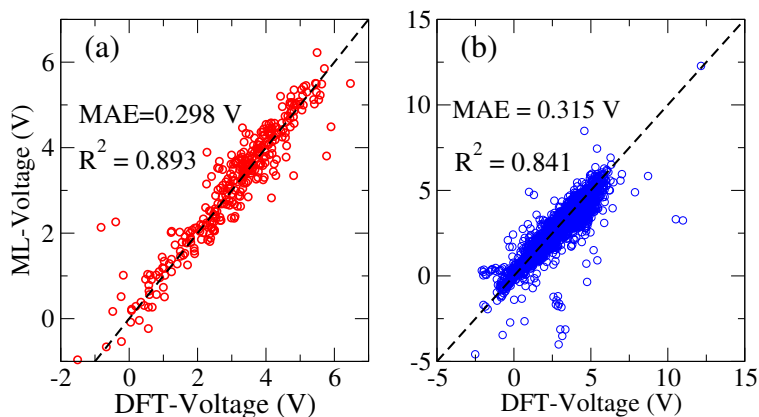


Figure 5: Voltage calculated from DFT Vs voltage predicted by machine learning graphs from (a) reaction-based and (b) E_{form} -based models (4000 data points). Here, part (a) has only a 329 data points since rest of the data were used to train the reaction-based model.

$$R^2 = 1 - \frac{\sum_{i=1}^N (V_i^{\text{DFT}} - V_i^{\text{ML}})^2}{\sum_{i=1}^N (V_i^{\text{DFT}} - \bar{V})^2} \quad (8)$$

Here, MAE is the mean absolute error and R^2 is the coefficient of determination. V_i^{DFT} represents the voltage computed from DFT for i^{th} battery framework, and V_i^{ML} indicates the corresponding machine learning predicted voltage. \bar{V} and N are the number of data and the average voltage in the test set, respectively. In Fig. 5, the voltage predicted using machine learning models (ML-Voltage) vs the voltage calculated using DFT (DFT-Voltage) is shown for both reaction- and E_{form} -based models.

For the reaction-based model, the test set consist of 329 battery systems comprising of 177 Li, 35 Ca, 3 Cs, 7 Rb, 11 K, 10 Y, 20 Na, 11 Al, 29 Zn, and 26 Mg battery systems. For the E_{form} -based model, we used all 4,402 electrodes of intercalation frameworks as the test set. It is clear from Fig. 5 that both models provide almost the same MAE even though the both training and test set sizes are significantly different. Among two approach, reaction-based model is most accurate with MAE of 0.298 V. Moreover, the R^2 value of reaction-based model is about 6% higher than that of the E_{form} -based model. One of the advantages of the reaction-based model over the other model is that it achieves almost the same accuracy as its counterpart although was trained using only 4073 materials. On the contrary, the E_{form} -based model trained with 60,000 materials to reached an MAE of 0.315 V. We note that, for training the E_{form} -based model, we excluded the battery materials from the training set. Our method provides nearly 17 % improvement when compared to previous approaches on the same data[15, 16].

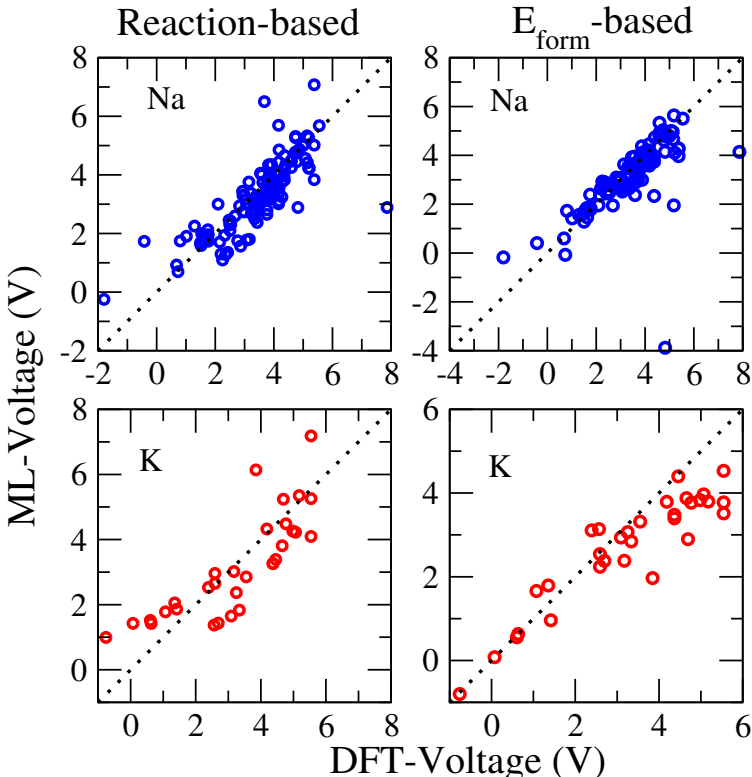


Figure 6: Performance of reaction-based and E_{form} -based models to predict the voltage of Na and K-ion based electrodes.

To examine the transferability of our models, we studied the accuracy of predicting the voltage in Na, and K ions-based electrodes. For this, we replaced the Li ions in known Li electrodes structures from materials Project with Na and K-ions. For such structures, geometry optimization was not performed to calculate voltage from our ML models. This approach allows us to examine the effect of geometry optimization on the performance of our models while predicting the voltage. Our approach is motivated by benchmark ML models used for molecules, which have achieved chemical accuracy with empirically obtained 3D-coordinates from SMILES strings of molecules [35]. For this, we extracted only DFT voltage of those Na and K-ions-based electrodes from MP database, which share the same structure (symmetry, space groups) as corresponding Li-based electrodes. As listed in Table 3, Na-ion shares relatively larger amount of Li

electrode structures (127), while K-ion electrodes has only 32. We replaced Li in those Li-based electrodes with Na and K-ions and calculated voltage using our ML models. We compared such voltage for Na and K-based electrodes with the corresponding voltage taken from MP. We found that, E_{form} -based model can predict voltage with MAE of 0.47 V which is nearly 20 % improvement over similar studies performed in the literature to examine the transferability of the model [16]. We obtained relatively large MAE of 0.56 V from reaction based model, which again is similar to previous work reported in literature. For K-ion battery electrodes, we obtained relatively large error from both reaction based and formation energy based models. We found that, formation energy based models are more accurate for predicting voltage for K-ion battery with MAE of 0.70 V.

Table 1: The number of Na, and K battery frameworks which share the same structures of Li frameworks (Test set size) and the MAE values in V for predicting voltage of the battery systems, where Li ions were substituted by those alkali ions.

Alkali ion	Test set size	MAE (GATGNN) Reaction-based	MAE (GATGNN) E_{form} -based
Na	127	0.56	0.47
K	32	0.87	0.70

In summary, we developed and compared the performance of two attention based graph neural networks to predict the voltage of the battery electrode materials. The first method predicts the voltage by considering the chemical reaction between a high-potential electrode and a metal ion to form a low-potential electrode. The second model predicts the E_{form} of individual electrodes before being used to compute the voltage. Our reaction-based model is 17% more accurate than the model explored in literature on the same MP dataset. Our reaction-based model consistently provides lower MAE and higher R^2 values compared to that from E_{form} -based model. Moreover, we show that relative to known models in literature, our models demonstrate high transfer-ability of performance when applied to Na and K-ion electrodes.

4 Author Contributions

S.L. and E.S. contributed equally to this work. Conceptualization, R.P.J., E.S., J.H.; methodology, S.L., E.S., J.H.; software, S.L., E.S.; validation, S.L., E.S., R.P.J.; investigation, S.L., E.S., J.H.; resources, J.H. and N.K.; data curation, R.P.J. and E.S.; writing—original draft preparation, E.S., S.L., R.P.J., and J.H.; writing—review and editing, N.K., J.H.; supervision, J.H. and R.P.J.; funding acquisition, J.H. and N.K.

5 Acknowledgement

Research reported in this work was supported in part by NSF under grants 1940099 and 1905775. The views, perspective, and content do not necessarily represent the official views of NSF. R.P.J. and N.K were supported by Laboratory Directed Research and Development Program and Mathematics for Artificial Reasoning for Scientific Discovery investment at the Pacific Northwest National Laboratory, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC06-76RLO.

References

- [1] Li M, Lu J, Chen Z, Amine K. 30 years of lithium-ion batteries. *Advanced Materials*. 2018;30(33):1800561.
- [2] Guo H, Wang Q, Stuke A, Urban A, Artrith N. Accelerated Atomistic Modeling of Solid-State Battery Materials With Machine Learning. *Frontiers in Energy Research*. 2021;9:265. Available from: <https://www.frontiersin.org/article/10.3389/fenrg.2021.695902>.
- [3] Deng J, Bae C, Denlinger A, Miller T. Electric Vehicles Batteries: Requirements and Challenges. *Joule*. 2020;4(3):511–515. Available from: <https://www.sciencedirect.com/science/article/pii/S254243512030043X>.
- [4] Goel S, Sharma R, Rathore AK. A review on barrier and challenges of electric vehicle in India and vehicle to grid optimisation. *Transportation Engineering*. 2021;4:100057. Available from: <https://www.sciencedirect.com/science/article/pii/S2666691X21000130>.
- [5] Jain A, Shin Y, Persson KA. Computational predictions of energy materials using density functional theory. *Nature Reviews Materials*. 2016;1(1):1–13.

- [6] Gandomi YA, Aaron DS, Houser JR, Daugherty MC, Clement JT, Pezeshki AM, et al. Critical Review—Experimental Diagnostics and Material Characterization Techniques Used on Redox Flow Batteries. *Journal of The Electrochemical Society*. 2018;165(5):A970–A1010. Available from: <https://doi.org/10.1149/2.0601805jes>.
- [7] Natarajan AR, Van der Ven A. Machine-learning the configurational energy of multicomponent crystalline solids. *npj Computational Materials*. 2018;4(1):1–7.
- [8] Houchins G, Viswanathan V. An accurate machine-learning calculator for optimization of Li-ion battery cathodes. *The Journal of Chemical Physics*. 2020;153(5):054124.
- [9] Allam O, Kuramshin R, Stoichev Z, Cho B, Lee S, Jang S. Molecular structure–redox potential relationship for organic electrode materials: density functional theory–Machine learning approach. *Materials Today Energy*. 2020;17:100482.
- [10] Artrith N, Urban A, Ceder G. Constructing first-principles phase diagrams of amorphous Li x Si using machine-learning-assisted sampling with an evolutionary algorithm. *The Journal of chemical physics*. 2018;148(24):241711.
- [11] Shandiz MA, Gauvin R. Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries. *Computational Materials Science*. 2016;117:270–278.
- [12] Cubuk ED, Sendek AD, Reed EJ. Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data. *The Journal of chemical physics*. 2019;150(21):214701.
- [13] Cunha RP, Lombardo T, Primo EN, Franco AA. Artificial intelligence investigation of NMC cathode manufacturing parameters interdependencies. *Batteries & Supercaps*. 2020;3(1):60–67.
- [14] Dixit MB, Verma A, Zaman W, Zhong X, Kenesei P, Park JS, et al. Synchrotron Imaging of Pore Formation in Li Metal Solid-State Batteries Aided by Machine Learning. *ACS Applied Energy Materials*. 2020;3(10):9534–9542.
- [15] Joshi RP, Eickholt J, Li L, Fornari M, Barone V, Peralta JE. Machine Learning the Voltage of Electrode Materials in Metal-Ion Batteries. *ACS Applied Materials & Interfaces*. 2019;11(20):18494–18503. Available from: <https://doi.org/10.1021/acsmi.9b04933>.
- [16] Moses IA, Joshi RP, Ozdemir B, Kumar N, Eickholt J, Barone V. Machine Learning Screening of Metal-Ion Battery Electrode Materials. *ACS Applied Materials & Interfaces*. 0;0(0):null. PMID: 34160211. Available from: <https://doi.org/10.1021/acsmi.1c04627>.
- [17] Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*. 2019;31(9):3564–3572.
- [18] Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*. 2018;120(14):145301.
- [19] Banjade HR, Hauri S, Zhang S, Ricci F, Gong W, Hautier G, et al. Structure motif–centric learning framework for inorganic crystalline systems. *Science Advances*. 2021;7(17). Available from: <https://advances.sciencemag.org/content/7/17/eabf1754>.
- [20] Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*. 2013;68:314–319. Available from: <https://www.sciencedirect.com/science/article/pii/S0927025612006295>.
- [21] Zhou F, Cococcioni M, Marianetti CA, Morgan D, Ceder G. First-principles prediction of redox potentials in transition-metal compounds with LDA+U. *Physical Review B*. 2004 Dec;70:235121. Available from: <http://link.aps.org/doi/10.1103/PhysRevB.70.235121>.
- [22] Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*. 2013;1(1):011002. Available from: <http://link.aip.org/link/AMPADS/v1/i1/p011002/s1&Agg=doi>.
- [23] Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*. 2020;32(1):4–24.
- [24] Louis SY, Nasiri A, Rolland FJ, Mitro C, Hu J. NODE-SELECT: A Graph Neural Network Based On A Selective Propagation Technique. *arXiv preprint arXiv:210208588*. 2021.
- [25] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *arXiv preprint arXiv:170603762*. 2017.
- [26] Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv preprint arXiv:171010903*. 2017.

- [27] Louis SY, Zhao Y, Nasiri A, Wang X, Song Y, Liu F, et al. Graph convolutional neural networks with global attention for improved materials property prediction. *Physical Chemistry Chemical Physics*. 2020;22(32):18141–18148.
- [28] Omprakash P, Manikandan B, Sandeep A, Shrivastava R, Viswesh P, Panemangalore DB. Graph representational learning for bandgap prediction in varied perovskite crystals. *Computational Materials Science*. 2021;196:110530.
- [29] Rosen AS, Iyer SM, Ray D, Yao Z, Aspuru-Guzik A, Gagliardi L, et al. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*. 2021;4(5):1578–1597.
- [30] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in pytorch. *Neural Information Processing Systems*. 2017.
- [31] Fey M, Lenssen JE. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:190302428*. 2019.
- [32] Girshick R. Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*; 2015. p. 1440–1448.
- [33] Gallarati S, Fabregat R, Laplaza R, Bhattacharjee S, Wodrich MD, Corminboeuf C. Reaction-based machine learning representations for predicting the enantioselectivity of organocatalysts. *Chem Sci*. 2021;12:6879–6889. Available from: <http://dx.doi.org/10.1039/D1SC00482D>.
- [34] Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Central Science*. 2017;3(5):434–443. PMID: 28573205. Available from: <https://doi.org/10.1021/acscentsci.7b00064>.
- [35] Hirohara M, Saito Y, Koda Y, Sato K, Sakakibara Y. Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. *BMC Bioinformatics*. 2018 12;19.