

# AI-Based Nano-Scale Material Property Prediction for Li-Ion Batteries

Subjects: Nanoscience & Nanotechnology

Contributor: Mohit Anil Lal, Akashdeep Singh, Ryan Mzik, Amirmasoud Lanjan, Seshasai Srinivasan

The status quo for techniques used in the discovery of new and novel materials to enhance battery technologies has progressed from expensive and time-consuming empirical trial and error methods to the more recent first principles approach of using quantum mechanics (QM), Monte Carlo simulations and molecular dynamics (MD). QM calculations evaluate electron–electron interactions by solving the complex Schrödinger equation, thereby enabling accurate results for a wide variety of properties. The emergence of ML, deep learning (DL) and artificial intelligence (AI) has helped alleviate the bottlenecks posed by QM and MD simulations and has made it possible to expand the scope of the search for novel materials in the chemical compound space (CCS).

Keywords: quantum mechanics ; molecular dynamics ; artificial intelligence ; machine learning ; nanotechnology

---

## 1. Introduction

The chemical compound space (CCS) is the theoretical space consisting of every possible compound known (and unknown) to us <sup>[1][2]</sup>. Even some of our largest databases consisting of approximately  $10^8$  known substances are a mere drop in the ocean compared with an estimated  $10^{180}$  substances that possibly make up the CCS <sup>[3][4]</sup>. Needless to say, the next big discovery of a compound that can revolutionize energy storage devices of the future is far from trivial.

The status quo for techniques used in the discovery of new and novel materials to enhance battery technologies has progressed from expensive and time-consuming empirical trial and error methods to the more recent first principles approach of using quantum mechanics (QM) <sup>[5][6][7][8][9]</sup>, Monte Carlo simulations and molecular dynamics (MD) <sup>[10][11][12][13][14]</sup>. QM calculations evaluate electron–electron interactions by solving the complex Schrödinger equation, thereby enabling accurate results for a wide variety of properties. However, the computational cost is a bottleneck for molecules larger than a couple hundred atoms. Hence, for multi-component or multi-layer structures such as the solid electrolyte interface layer, QM is not a feasible approach. Additionally, many battery components including ionic and polymer electrolytes, crystal structures and electrode–electrolyte interactions <sup>[11][15][16][17][18]</sup> are better analyzed on larger length and time scales that are inaccessible with QM. MD simulations simplify particle–particle interactions to five main types of interactions, namely nonbonded, bonded, angle, dihedral and improper interactions. These interactions, which can be obtained using a simple algebraic equation, reduce the computational cost significantly and are applicable to systems almost  $10^6$

times larger. To analyze ion migration in perovskite nickelate with 200 atoms, QM techniques, even using density functional theory (DFT) approximation to reduce computational costs, require about  $10^5$  core-hours of computational time in a picosecond range simulation. On the other hand, MD simulations with 105 atoms required only  $10^4$

core-hours of computational time <sup>[19]</sup>. Thus, MD simulations enable the analysis of a wide variety of properties and behavior of materials at the atomic scale, such as the crystal structure, thermal properties and mechanical properties, which are often too complex to model using QM calculations. In a recent review, Sun et al. <sup>[20]</sup> presented the use of MD simulations to optimize lithium metal batteries, investigating the transport structure of Li ions, the electrochemical process at the electronic, atomic or molecular level, the Li<sup>+</sup> transport mechanism and the Li deposition behavior in detail.

Though MD simulations are widely used to investigate the properties of materials at the atomic level, these simulations rely on experimentally derived interatomic potential parameters that determine the forces between particles <sup>[21]</sup>. This dependence on prior experimental data poses a challenge in using MD to design new and novel materials. To address this issue, Lanjan et al. <sup>[22]</sup> recently proposed a novel computational framework that couples QM calculations with MD simulations. This generates a wide range of crystal structures by varying a single system parameter (e.g., bond length) while keeping other parameters relaxed at their minimum energy level. The QM calculations are then used to evaluate the

system's energy as a function of these changes, and the resulting data points are used to fit the interaction equations to estimate the potential parameters for each type of particle–particle interaction. Employing this framework enables the study of crystal structures with the accuracy of QM calculations but at the speed and system sizes permissible by MD techniques. While this framework enhances nano-based computational methods, the QM calculations still need massive amounts of computational power, which can be significantly reduced with the AI-based technique proposed in this work.

The emergence of ML, deep learning (DL) and artificial intelligence (AI) has helped alleviate the bottlenecks posed by QM and MD simulations and has made it possible to expand the scope of the search for novel materials in the CCS. ML and DL algorithms are orders of magnitude faster than ab initio techniques. Unlike the QM-based simulations, which can take days to complete, ML algorithms can produce results within seconds. The use of AI has brought a paradigm shift in research related to improving battery technology as well as molecular property prediction and material discovery in general. For example, Sandhu et al. [23] used DL to examine the optimal crystal structures of doped cathode materials in lithium manganese oxide (LMO) batteries. Failed or unsuccessful synthesis data were used to predict the reaction success rate for the crystallization of templated vanadium selenites [24]. Using QM and ML techniques, Lu et al. [25] developed a method to predict undiscovered hybrid organic-inorganic perovskites (HOIPs) for photovoltaics. Their screening technique was able to shortlist six HOIPs with ideal band gaps and thermal stabilities from 5158 unexplored candidates. To identify material compositions with suitable properties, Meredig et al. [26] built an ML model trained on thousands of ground state crystal structures and used this model to scan roughly 1.6 million candidate compositions of novel ternary compounds to produce a ranked list of 4500 stable ternary compositions that would possibly represent undiscovered materials.

The broad approach employed when using AI-based property prediction models consists of three overarching components: a reference database consisting of relevant quantum mechanical data which is used to fit the AI model; a mathematical representation that not only uniquely describes the attributes of the reference materials but also enables effective model training; and finally a suitable AI model that can accomplish the learning task itself.

## 2. Database

The fundamental premise of AI is the ability to draw inferences from patterns in data and enable an accurate prediction in unknown domains. Hence, the data, which make up the training examples for our learning task, becomes a critical aspect for successful prediction. With the introduction of the Materials Genome Initiative in 2011 [27], the United States signaled the importance of unifying the infrastructure for material innovation and harnessing the power of material data. In lieu of the same goal, there has been an advent of various materials databases, such as the Inorganic Crystal Structure Database (ICSD) [28], the Open Quantum Materials Database (OQMD) [29], the Cambridge Structural Databases [30], the Harvard Clean Energy Project [31], the Materials Project [32] and the AFLOWLIB [33]. Specifically, the size of the training examples, the diversity of the dataset and the degrees of freedom all contribute to how effective the learning task for a specific objective can be [34]. In predicting properties such as the band gap energy and glass-forming ability for crystalline and amorphous materials, Ward et al. [35] methodically selected a chemically diverse set of attributes taken from the OQMD. Similarly, for electronic-structure problems, Schütt et al. [36] noted that the density of states at the Fermi energy is the critical property of concern. In predicting this property, around 7000 crystal structures from the ICSD were used, observing higher predicted variance for certain configurations and the need to extend the training set in these specific areas. The process of material discovery is complex and diverse, and it is not surprising that there is no one-size-fits-all database that can accurately predict the properties of all materials. The physical and chemical characteristics of materials vary widely, requiring different methods and techniques for precise analysis and prediction. Moreover, the current methodologies rely on the availability of well-curated data or the ability to manually generate such data, which is a daunting and often infeasible task, especially for new and unexplored materials. Thus, there is a need to develop generalizable and adaptable approaches that can efficiently handle a diverse range of materials, properties and configurations without the need for extensive data generation or curation.

## 3. Molecular Representation

ML algorithms draw inferences from data to establish a relationship between the atomic structure and the properties of a system. To enable the best possible structure-property approximation, a good representation of the material (also referred to as the 'fingerprint' or 'descriptor') is crucial. The first Hohenberg–Kohn theorem of DFT proves that the electron density of a system contains all the information needed to describe its ground state properties, and it is a 'universal descriptor' that can be used to predict these properties without knowledge of the details of the interactions between the electrons [37]. Crucially, for ML, a good molecular representation is invariant to rotation and translation of the system as well as permutation of atomic indices [38]. Therefore, unfortunately, the electronic density is not a universally suitable

representation of a system. Additionally, a good descriptor must be unique, continuous, compact and computationally cheap [38]. Often, there are multiple molecular geometries that possess similar values for a property. Hence, there is no single universal representation for all properties leading to hundreds of molecular descriptors that are suitable only for a small subset of the CCS and a small subset of properties [39]. A commonly used molecular representation that satisfies the above-mentioned criteria of a good representation is the 'Coulomb matrix'. It uses the same parameters that constitute the Hamiltonian for any given system, namely the set of Cartesian coordinates  $R_i$  and nuclear charges  $Z_i$  [40]. While the Coulomb matrix representation has shown tremendous success for property prediction in finite systems, it is unable to do the same for infinite periodic crystal structures [36]. Hansen et al. [41] proposed a new descriptor called 'bag-of-bonds' that performed better due to incorporating the many-body interactions of a system. In fact, the use of different descriptors in an ML endeavor for material property prediction is so common that there are open-source software packages that provide implementations for a myriad of different descriptors [38]. Unfortunately, a lack of clarity on the right descriptor makes the use of AI inaccessible to researchers that possess domain expertise but lack the needed knowledge of AI. Additionally, the lack of generalizability of a chosen descriptor makes the current AI-based techniques inaccurate and narrow in scope.

## 4. AI Model

In addition to an appropriate database and the precise molecular representation, a critical aspect in the material property prediction process is the choice of the AI algorithm. AI algorithms can be categorized into supervised learning, unsupervised learning and reinforcement learning. Supervised learning uses a standard fitting procedure that attempts to determine a mapping function between the known input features and the corresponding output labels. The goal is to make accurate predictions for new, unseen data. In contrast, unsupervised learning does not have prior knowledge of the desired output, and the goal is to find patterns and structures in this unlabeled data. Reinforcement learning uses an iterative trial-and-error process where the actions are determined based on reinforcement in the form of a reward-penalty system. The goal here is to maximize the cumulative reward over time. Supervised learning is the most widespread category of learning used in materials research. Different models may be better suited for certain types of materials or properties, and the choice of model often depends on the available data and the specific goals of the prediction task. Akbarpour et al. [42] found that artificial neural networks (ANNs) performed better in predicting the synthesis conditions of nano-porous anodic aluminum oxide at the interpore distance in comparison with both multiple linear regression and experimental studies. On the other hand, for the modeling and synthesis of zeolite synthesis, Manuel Serra et al. [43] found that support vector regression (SVR) outperformed ANNs and decision trees. Fang et al. [44] proposed a novel hybrid methodology for forecasting the atmospheric corrosion of metallic materials where the optimal hyperparameters for an SVR model were automatically determined using a generic algorithm. These examples highlight the need for AI expertise when choosing the right algorithm for a given application, which can be a barrier to making AI methods accessible for materials-based research.

---

## References

1. von Lilienfeld, O.A. First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties. *Int. J. Quantum Chem.* 2013, 113, 1676–1689.
2. Huang, B.; von Lilienfeld, O.A. Ab Initio Machine Learning in Chemical Compound Space. *Chem. Rev.* 2021, 121, 10001–10036.
3. Lemonick, S. Exploring chemical space: Can AI take us where no human has gone before? *Chem. Eng. News* 2020, 98, 30–35.
4. Ruddigkeit, L.; van Deursen, R.; Blum, L.C.; Raymond, J.L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* 2012, 52, 2864–2875.
5. Moradi, Z.; Heydarinasab, A.; Shariati, F.P. First-principle study of doping effects (Ti, Cu, and Zn) on electrochemical performance of Li<sub>2</sub>MnO<sub>3</sub> cathode materials for lithium-ion batteries. *Int. J. Quantum Chem.* 2021, 121, e26458.
6. Moradi, Z.; Lanjan, A.; Srinivasan, S. Multiscale Investigation into the Co-Doping Strategy on the Electrochemical Properties of Li<sub>2</sub>RuO<sub>3</sub> Cathodes for Li-Ion Batteries. *ChemElectroChem* 2021, 8, 112–124.
7. Tyagi, R.; Lanjan, A.; Srinivasan, S. Co-Doping Strategies to Improve the Electrochemical Properties of Li<sub>x</sub>Mn<sub>2</sub>O<sub>4</sub> Cathodes for Li-Ion Batteries. *ChemElectroChem* 2022, 9, e202101626.
8. Moradi, Z.; Lanjan, A.; Srinivasan, S. Enhancement of Electrochemical Properties of Lithium Rich Li<sub>2</sub>RuO<sub>3</sub> Cathode Material. *J. Electrochem. Soc.* 2020, 167, 110537.

9. Xia, W.; Sakurai, M.; Balasubramanian, B.; Liao, T.; Wang, R.; Zhang, C.; Sun, H.; Ho, K.M.; Chelikowsky, J.R.; Sellmyer, D.J.; et al. Accelerating the discovery of novel magnetic materials using machine learning-guided adaptive feedback. *Proc. Natl. Acad. Sci. USA* 2022, 119, e2204485119.
10. Lanjan, A.; Srinivasan, S. An Enhanced Battery Aging Model Based on a Detailed Diffusing Mechanism in the SEI Layer. *ECS Adv.* 2022, 1, 030504.
11. Lanjan, A.; Moradi, Z.; Srinivasan, S. Multiscale Investigation of the Diffusion Mechanism within the Solid–Electrolyte Interface Layer: Coupling Quantum Mechanics, Molecular Dynamics, and Macroscale Mathematical Modeling. *ACS Appl. Mater. Interfaces* 2021, 13, 42220–42229.
12. Lanjan, A.; Choobar, B.G.; Amjad-Iranagh, S. First principle study on the application of crystalline cathodes  $\text{Li}_2\text{Mn}_0.5\text{TM}_0.5\text{O}_3$  for promoting the performance of lithium-ion batteries. *Comput. Mater. Sci.* 2020, 173, 109417.
13. Lanjan, A.; Choobar, B.G.; Amjad-Iranagh, S. Promoting lithium-ion battery performance by application of crystalline cathodes  $\text{LiMn}_{1-x}\text{Fe}_x\text{PO}_4$ . *J. Solid State Electrochem.* 2020, 24, 157–171.
14. Stanev, V.; Choudhary, K.; Kusne, A.G.; Paglione, J.; Takeuchi, I. Artificial intelligence for search and discovery of quantum materials. *Commun. Mater.* 2021, 2, 105.
15. Aykol, M.; Herring, P.; Anapolsky, A. Machine learning for continuous innovation in battery technologies. *Nat. Rev. Mater.* 2020, 5, 725–727.
16. Ye, C.; Tan, R.; Wang, A.; Chen, J.; Comesaña Gándara, B.; Breakwell, C.; Alvarez-Fernandez, A.; Fan, Z.; Weng, J.; Bezzu, C.G.; et al. Long-Life Aqueous Organic Redox Flow Batteries Enabled by Amidoxime-Functionalized Ion-Selective Polymer Membranes. *Angew. Chem. Int. Ed.* 2022, 61, e202207580.
17. Spotte-Smith, E.W.C.; Blau, S.M.; Xie, X.; Patel, H.D.; Wen, M.; Wood, B.; Dwaraknath, S.; Persson, K.A. Quantum chemical calculations of lithium-ion battery electrolyte and interphase species. *Sci. Data* 2021, 8, 203.
18. Chattopadhyay, J.; Pathak, T.S.; Santos, D.M.F. Applications of Polymer Electrolytes in Lithium-Ion Batteries: A Review. *Polymers* 2023, 15, 3907.
19. Chan, H.; Narayanan, B.; Cherukara, M.J.; Sen, F.G.; Sasikumar, K.; Gray, S.K.; Chan, M.K.Y.; Sankaranarayanan, S.K.R.S. Machine Learning Classical Interatomic Potentials for Molecular Dynamics from First-Principles Training Data. *J. Phys. Chem. C* 2019, 123, 6941–6957.
20. Sun, Y.; Yang, T.; Ji, H.; Zhou, J.; Wang, Z.; Qian, T.; Yan, C. Boosting the Optimization of Lithium Metal Batteries by Molecular Dynamics Simulations: A Perspective. *Adv. Energy Mater.* 2020, 10, 2002373.
21. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* 2016, 145, 170901.
22. Lanjan, A.; Moradi, Z.; Srinivasan, S. A computational framework for evaluating molecular dynamics potential parameters employing quantum mechanics. *Mol. Syst. Des. Eng.* 2023, 8, 632–646.
23. Sandhu, S.; Tyagi, R.; Talaie, E.; Srinivasan, S. Using neurocomputing techniques to determine microstructural properties in a Li-ion battery. *Neural Comput. Appl.* 2022, 34, 9983–9999.
24. Raccuglia, P.; Elbert, K.C.; Adler, P.D.F.; Falk, C.; Wenny, M.B.; Mollo, A.; Zeller, M.; Friedler, S.A.; Schrier, J.; Norquist, A.J. Machine-learning-assisted materials discovery using failed experiments. *Nature* 2016, 533, 73–76.
25. Lu, S.; Zhou, Q.; Ouyang, Y.; Guo, Y.; Li, Q.; Wang, J. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* 2018, 9, 3405.
26. Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J.E.; Doak, J.W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* 2014, 89, 094104.
27. National Academies of Sciences, Engineering, and Medicine. NSF Efforts to Achieve the Nation's Vision for the Materials Genome Initiative: Designing Materials to Revolutionize and Engineer Our Future (DMREF); The National Academies Press: Washington, DC, USA, 2023.
28. Belsky, A.; Hellenbrandt, M.; Karen, V.L.; Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in support of materials research and design. *Acta Crystallogr. Sect. B Struct. Sci.* 2002, 58, 364–369.
29. Kirklin, S.; Saal, J.E.; Meredig, B.; Thompson, A.; Doak, J.W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *Npj Comput. Mater.* 2015, 1, 15010.
30. Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* 2016, 72, 171–179.
31. Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R.S.; Gold-Parker, A.; Vogt, L.; Brockway, A.M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening

- and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* 2011, 2, 2241–2251.
32. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* 2013, 1, 011002.
33. Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R.H.; Nelson, L.J.; Hart, G.L.; Sanvito, S.; Buongiorno-Nardelli, M.; et al. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* 2012, 58, 227–235.
34. Zhang, Y.; Ling, C. A strategy to apply machine learning to small datasets in materials science. *Npj Comput. Mater.* 2018, 4, 25.
35. Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *Npj Comput. Mater.* 2016, 2, 16028.
36. Schütt, K.T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K.R.; Gross, E.K.U. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Phys. Rev. B* 2014, 89, 205118.
37. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* 1964, 136, B864–B871.
38. Himanen, L.; Jäger, M.O.; Morooka, E.V.; Canova, F.F.; Ranawat, Y.S.; Gao, D.Z.; Rinke, P.; Foster, A.S. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* 2020, 247, 106949.
39. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* 2018, 361, 360–365.
40. Rupp, M.; Tkatchenko, A.; Müller, K.R.; von Lilienfeld, O.A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* 2012, 108, 058301.
41. Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O.A.; Müller, K.R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* 2015, 6, 2326–2331.
42. Akbarpour, H.; Mohajeri, M.; Moradi, M. Investigation on the synthesis conditions at the interpore distance of nanoporous anodic aluminum oxide: A comparison of experimental study, artificial neural network, and multiple linear regression. *Comput. Mater. Sci.* 2013, 79, 75–81.
43. Serra, J.M.; Baumes, L.A.; Moliner, M.; Serna, P.; Corma, A. Zeolite Synthesis Modelling with Support Vector Machines: A Combinatorial Approach. *Comb. Chem. High Throughput Screen.* 2007, 10, 13–24.
44. Fang, S.; Wang, M.; Qi, W.; Zheng, F. Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials. *Comput. Mater. Sci.* 2008, 44, 647–655.