Importance of Electronic Entropy for Machine Learning Interatomic Potentials

Martin Hoffmann Petersen [1 2] **Steen Lysgaard** [1] **Kedar Hippalgaonkar** [2 3] **Arghya Bhowmik** [1] **Juan Maria Garcia Lastra** [1]

[1]*Technical University of Denmark, Department of Energy Conversion and Storage, Lyngby, Denmark 2800* [2]*School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798* [3]*Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544*. Correspondence to: Martin Hoffmann Petersen mahpe@dtu.dk.

1. Introduction

Machine Learning Interatomic Potentials (MLIPs) are increasingly used to accelerate computational materials discovery, enabling simulations at scales far beyond those accessible to first-principles methods [1, 2]. However, most MLIPs rely exclusively on atomic positions and chemical species, neglecting explicit electronic information [3, 4]. To mitigate this limitation, newer MLIPs incorporate the Latent Ewald Summation (LES) framework [5] to account for long-range electrostatic interactions and environment-dependent charge effects[6, 7]. While this approach significantly improves the treatment of global electrostatics, the absence of explicitly resolved atomic charges can still limit the accurate description of local chemical environments and charge-state-dependent phenomena. This limitation is particularly pronounced during structural optimization, where inaccuracies in force predictions increase the likelihood of converging to local minima associated with incorrect charge distributions.

As a consequence, the omission of an explicit electronic description can lead to systematic errors in systems where identical atomic species exist in multiple charge states. In such cases, the local chemical environment alone may be insufficient to uniquely determine the underlying electronic structure. This limitation is particularly critical in battery cathode materials, where electrochemical performance is governed by redox processes and the coexistence of multiple charge states [8, 9]. One representative example is the olivine-type cathode material $NaFePO_4$, which spans a wide range of sodium concentrations and involves a Fe^{2+}/Fe^{3+} redox couple during (de)sodiation.

In this work, we investigate the impact of electronic entropy (defined here as the spatial distribution of Fe^{2+}/Fe^{3+} redox pairs) on MLIP performance using $NaFePO_4$ as a model system. In addition to redox activity, $NaFePO_4$ exhibits significant Na/vacancy disorder and undergoes an experimentally verified composition-driven phase transition at $Na_{2/3}FePO_4$, making it a stringent benchmark for charge-sensitive modeling [10, 11, 12].

We compare four representative MLIPs: MACE[13], PaiNN[14], CHGNet[15], and the atomic charge-aware cPaiNN[16], with particular focus on their ability to describe charge-state-dependent energetics, atomic disorder, and phase stability. This gives a powerful comparison between different architectures without implicit charge correction (PaiNN and MACE) as well as architectures which account for charge effects indirectly using magnetic moments (CHGNet) and explicitly using atomic charges using bader charge analysis[17] (cPaiNN).

2. Methods and Results

CHGNet (v0.4.1)[15], MACE (v0.3.13) [13], PaiNN (v1.0.0), and cPaiNN (v1.0.0)[16] were initially trained on the polyanion sodium cathode materials dataset[18]. The trained models and correpsonding data splits used in this work are available at [19].

To explicitly train the models to account for electronic entropy, all four MLIPs were additionally trained on a chemically restricted subset of the same dataset containing only structures composed of Na, Fe, P, Si, S, and O. This subset excludes compounds containing other transition-metal species while retaining a diverse set of relevant cathode materials across varying sodium concentrations, including olivine-type $NaFePO_4$, maricite-type $NaFePO_4$, Na_2FeSiO_4, and $Na_{2.56}Fe_{1.72}(SO_4)_3$. Restricting the chemical space in this manner enables a more controlled and systematic evaluation of the ability of MLIPs to distinguish between the Fe^{2+} and Fe^{3+} charge states.

Even without explicitly considering different charge states, cathode materials exhibit significant atomic site disorder as a function of sodium concentration, arising from partial occupancy of Na ions and vacancies at the Na Wyckoff sites [8, 20, 21]. Due to the combinatorial complexity of Na/vacancy arrangements, low-energy configurations at each sodium concentration are efficiently sampled using a Genetic Algorithm (GA) [22, 23] driven by MLIP-predicted energies. This approach enables the construction of composition–energy convex hulls and corresponding Open-Circuit Voltage (OCV) profiles, providing a crucial link between computational predictions and experimental measurements [24].

Fig. 1 shows the results of the GA sampling for $NaFePO_4$ using the trained MACE model on the polyanion sodium cathode materials dataset, compared against Density Functional Theory (DFT)–optimized structures for the three lowest- and highest-energy compositions identified by the GA. The comparison clearly demonstrates that the MACE model fails to reproduce the DFT energetic ordering and structural energetics across the sampled compositions, highlighting limitations in its ability to capture charge- and disorder-driven

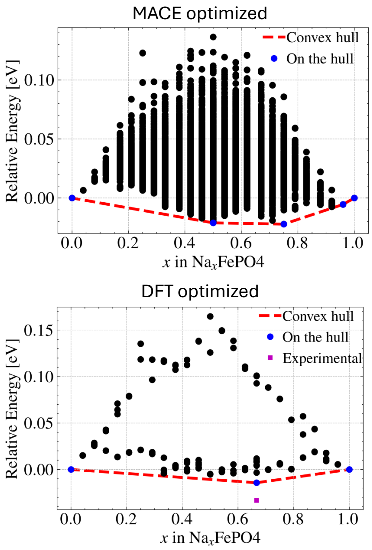

Fig. 1: Structures obtained from the GA sampling driven by the MACE model, compared with DFT–optimized structures for the three lowest- and highest-energy compositions identified by the GA, together with DFT optimization of the experimentally verified structure [11]. In addition, seven extra lower-energy compositions identified by the GA are DFT-optimized at the experimental sodium concentration of 66%.

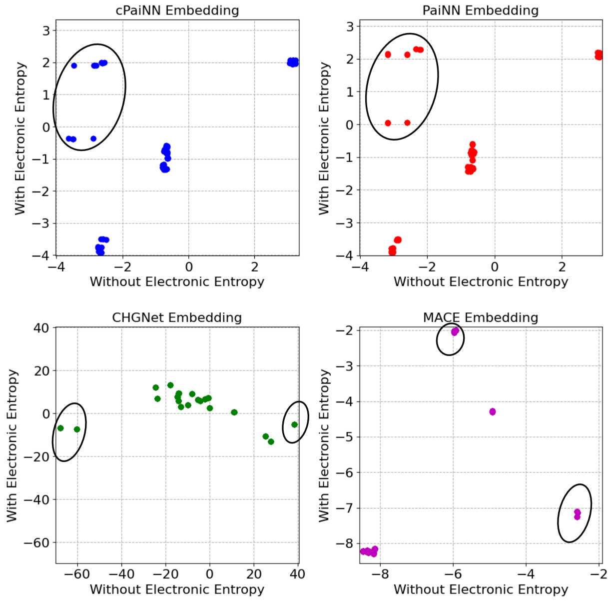

Fig. 2: Node embeddings used to predict the total energy for the four MLIPs, compared with and without explicit inclusion of electronic entropy. The Fe^{2+}/Fe^{3+} redox pair embeddings are highlighted to illustrate changes in the learned potential energy landscape.

effects in this system. Moreover, the GA sampling did not identify the experimentally verified phase, suggested by DFT to be clearly the lowest in energy. Our results show that models lacking indirect charge information fail to reproduce experimental observations. While indirectly charge-aware models yield improved agreement, they still do not fully capture the experimental behavior.

Figure 1 shows the results of genetic algorithm (GA) sampling for $NaFePO_4$ using the trained MACE model on the polyanionic sodium cathode materials dataset, compared with density functional theory (DFT)–optimized structures for the three lowest and highest-energy compositions identified by the GA. In addition, seven lower-energy compositions identified by the GA are further DFT-optimized at the experimental sodium concentration of 66% to assess whether the GA sampling captures the experimentally observed phase. The comparison demonstrates that the MACE model fails to reproduce the DFT energetic ordering and structural energetics across the sampled compositions, highlighting its limited ability to capture charge- and disorder-driven effects in this system. Moreover, the GA sampling, and thus the MACE model, does not identify the experimentally verified phase, which is predicted by DFT to clearly be the lowest-energy structure.

More generally, our results show that models lacking explicit charge information fail to reproduce experimental observations. While indirectly charge-aware models exhibit improved agreement, they still do not fully capture the experimental behavior.

Incorporating electronic entropy directly into all four MLIPs, using the chemically cleaned dataset together with our electronic-entropy embedding, leads to a substantial improvement in performance across all models. This improvement is further reflected in clear differences between the learned potential energy landscapes of MLIPs trained with and without electronic entropy, as revealed by the architecture-specific embeddings shown in Fig. 2. The embeddings are extracted from the MLIPs during optimization of the experimentally verified structure. The resulting clustering highlights subtle yet crucial distinctions arising from the Fe^{2+}/Fe^{3+} redox pairs.

These findings underscore the importance of explicitly incorporating multiple charge states into MLIP architectures. Notably, charge-aware MLIPs achieve superior accuracy compared to charge-agnostic models once electronic entropy is considered. Beyond establishing a stringent benchmark for MLIPs, this work provides practical guidance for the development of next-generation models capable of faithfully capturing charge-state-dependent chemistry and atomic disorder in complex materials.

References

[1] Maksim Kulichenko, Benjamin Nebgen, Nicholas Lubbers, Justin S. Smith, Kipton Barros, Alice E. A. Allen, Adela Habib, Emily Shinkle, Nikita Fedik, Ying Wai Li, Richard A. Messerly, and Sergei Tretiak. Data generation for machine learning interatomic potentials and beyond. *Chemical Reviews*, 124(24):13681–13714,

November 2024.

[2] Arun Mannodi-Kanakkithodi and Maria K.Y. Chan. Computational data-driven materials discovery. *Trends in Chemistry*, 3(2):79–82, February 2021.

[3] Oliver T. Unke, Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12(1), December 2021.

[4] Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, March 2021.

[5] Bingqing Cheng. Latent ewald summation for machine learning of long-range interactions. *npj Computational Materials*, 11(1), March 2025.

[6] Dongjin Kim, Xiaoyu Wang, Peichen Zhong, Daniel S. King, Theo Jaffrelot Inizan, and Bingqing Cheng. A universal augmentation framework for long-range electrostatics in machine learning interatomic potentials, 2025.

[7] Dongjin Kim and Bingqing Cheng. Long-range electrostatics for machine learning interatomic potentials is easier than we thought, 2025.

[8] Simon Loftager, Steen Brian Schougaard, Tejs Vegge, and Juan María García-Lastra. Density functional theory study of redox potential shifts in lixmnyfe1–ypo4 battery electrodes. *The Journal of Physical Chemistry C*, 123(1):102–109, December 2018.

[9] Ting Jin, Huangxu Li, Kunjie Zhu, Peng-Fei Wang, Pei Liu, and Lifang Jiao. Polyanion-type cathode materials for sodium-ion batteries. *Chemical Society Reviews*, 49(8):2342–2377, 2020.

[10] Fumiyasu Nozaki, Shaoning Zhang, Martin Hoffmann Petersen, Jinkwang Hwang, Jin Hyun Chang, Juan María García-Lastra, and Kazuhiko Matsumoto. Oxygen-powered sustainable fepo4 preparation for sodium metal batteries with li acetate recovery. *Energy amp; Environmental Science*, 18(3):1408–1417, 2025.

[11] Florent Boucher, Joël Gaubicher, Marine Cuisinier, Dominique Guyomard, and Philippe Moreau. Elucidation of the na2/3fepo4 and li2/3fepo4 intermediate superstructure revealing a pseudouniform ordering in 2d. *Journal of the American Chemical Society*, 136(25):9144–9157, June 2014.

[12] Martin Hoffmann Petersen, Arghya Bhowmik, Jinhyun Chang, and Juan Maria García Lastra. Polyanion sodium cathode materials dataset, 2025.

[13] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems*, 35:11423–11436, 2022.

[14] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

[15] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, September 2023.

[16] Martin Hoffmann Petersen, Chiku Paradi, Jonas Busk, , Jinhyun Chang, Arghya Bhowmik, and Juan Maria García Lastra. Limitations of foundation models in energy materials simulations: A case study in polyanion sodium cathode materials, 2026.

[17] W Tang, E Sanville, and G Henkelman. A grid-based bader analysis algorithm without lattice bias. *Journal of Physics: Condensed Matter*, 21(8):084204, January 2009.

[18] Martin Hoffmann Petersen, Jin Hyun Chang, Arghya Bhowmik, and Juan Maria García Lastra. Dataset exploring the atomic scale structure and ionic dynamics of polyanion sodium cathode materials. *Scientific Data*, 12(1), August 2025.

[19] Martin Hoffmann Petersen, Chiku Parida, Jonas Busk, Jinhyun Chang, Arghya Bhowmik, and Juan Maria García Lastra. Additional data for the polyanion sodium cathode materials dataset, 2025.

[20] Mila AI4Science, Alex Hernandez-Garcia, Alexandre Duval, Alexandra Volokhova, Yoshua Bengio, Divya Sharma, Pierre Luc Carrier, Yasmine Benabed, Michał Koziarski, and Victor Schmidt. Crystal-gfn: sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*, 2023.

[21] Maha Ismail, Oier Lakuntza, Javier Carrasco, Damien Saurel, Montse Casas-Cabanas, Marine Reynaud, and Amaia Saracibar. Probing phase formation and structural transformations in sodium extraction and insertion of nafe1–ymnypo4 through first-principles calculations. *Inorganic Chemistry*, 63(43):20541–20550, October 2024.

[22] Steen Lysgaard, David D. Landis, Thomas Bligaard, and Tejs Vegge. Genetic algorithm procreation operators for alloy nanoparticle catalysts. *Topics in Catalysis*, 57(1–4):33–39, October 2013.

[23] Paul C. Jennings, Steen Lysgaard, Jens Strabo Hummelshøj, Tejs Vegge, and Thomas Bligaard. Genetic algorithms for computational materials discovery accelerated by machine learning. *npj Computational Materials*, 5(1), April 2019.

[24] Simon Loftager, Juan María García-Lastra, and Tejs Vegge. A density functional theory study of the ionic and electronic transport mechanisms in lifebo3 battery electrodes. *The Journal of Physical Chemistry C*, 120(33):18355–18364, July 2016.