# High-throughput ML Screening of Doped Cathode Active Materials

**Lucas Pinède**[1 2] **Cory Hargus**[1] **Richard Tran**[1] **Mathilde L. D. Franckel**[1 3] **Nikita Hall**[1]
**Hannah Bull**[1] **Sagar Pal**[1] **Alexandre Duval**[1]
[1]*Entalpic, Paris, France* [2]*Chimie ParisTech, PSL University, Paris, France* [3]*Imperial College, London, United-Kingdom*.
Correspondence to: Lucas Pinède lucas.pinede@entalpic.ai, Cory Hargus cory.hargus@entalpic.ai.

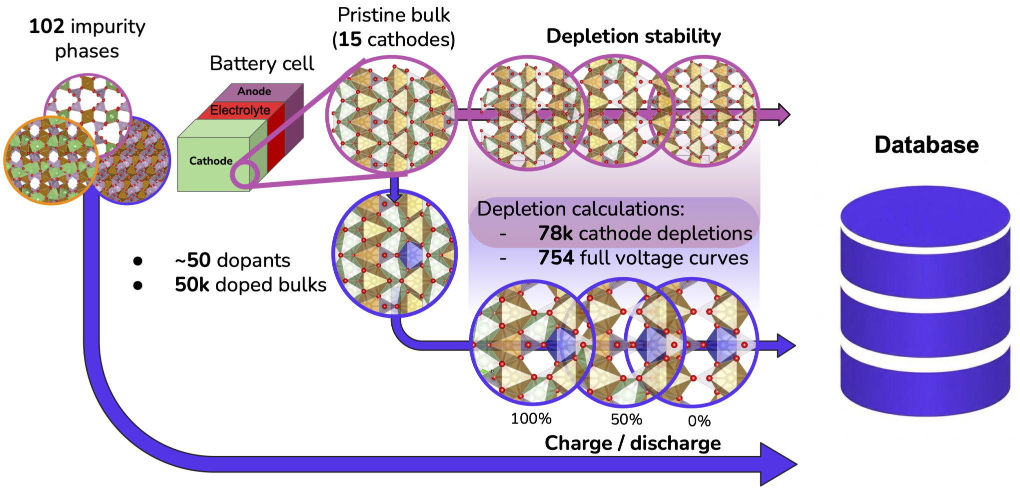

Fig. 1: High-throughput pipeline for discovering novel CAMs. (Center) After determining stable dopant configurations, the full voltage curve is computed by sequentially assessing the most stable cation vacancy until full delithiation. (Left) In parallel, formation energies of pristine and doped impurity phases are computed to construct the convex hull for phase stability assessment. All calculations are accelerated using an off-the-shelf universal Machine Learning Interatomic Potential (MLIP).

## 1. Introduction

Battery technology is essential to the clean energy transition, with applications ranging from consumer electronics to large-scale grid-scale energy storage. Lithium-ion batteries, in particular, have proven durable, safe, and low-cost. Outstanding challenges in their design include simultaneously optimizing energy density, stability, and ionic and electronic conductivity.

One approach to improving performance has been "doping" the Cathode Active Material (CAM) with extrinsic elements, similar to common practices in the semiconductor industry [1, 2, 3, 4]. Yet, the vast combinatorial space of CAMs, dopants and even co-dopants makes brute-force experimental of this space inaccessible. In the past two decades, computational chemistry modeling has emerged as a new materials discovery paradigm, accelerating screening of large chemical spaces, particularly for battery applications [5, 6, 7]. However, workhorse models such as Density Functional Theory (DFT) still suffer from high cost and poor (cubic) scaling with system size. The large amount of DFT data generated through these efforts has recently enabled training universal Machine Learning Interatomic Potentials (MLIPs) with high accuracy and generalizability across a broad range chemical systems, with dramatically reduced computational expense.

In this work, we present an automated high-throughput screening pipeline that combines first-principles DFT calculations and universal MLIPs to identify promising doped CAMs. While DFT provides state-of-the-art predictive capabilities, its computational cost prohibits large-scale screening of co-dopants, limiting its application to targeted studies. Therefore, we leverage MLIPs to emulate DFT at a fraction of the cost, accelerate structural relaxations, and energy computations. This approach unlocks the phase stability and full voltage characterization of 15 CAMs across 50 dopants and co-dopants ($\sim 1,000,000$ MLIP calls per CAM), achieving strong agreement with DFT. Furthermore, our ML workflow bridges the gap to experimental conditions by scaling to large supercells, allowing for the modeling of realistic dopant concentrations.

We validate our computational screening by presenting experimental synthesis and X-ray Diffraction (XRD) characterization of the most promising doped materials identified by our workflow. This cross-validation of these results highlights the precision of our approach and the potential of the candidates for next-generation battery technology. Additionally comparing with DFT methods, we demonstrate the ability of ML workflows to accelerate scientific discovery in highly combinatorial search spaces with good predictive accuracy.

## 2. Methodology

We enable the high-throughput screening of doped CAMs using an ML-accelerated pipeline. We employ UMA [8], a foundational MLIP pretrained on diverse atomic structures ranging from molecules to materials. To generalize across varying DFT levels of theory, UMA utilizes a Mixture of Linear Experts (MoLE) architecture conditioned on the specific task. For inference, we select the *omat* task expert, as the Open Materials (OMat24) [9] dataset domain aligns with our target inorganic structures and reference DFT settings.

Our framework is illustrated Figure 1. After screening for thermodynamic stability we subsequently subsequently evaluate electrochemical performance.

We start by identifying the input bulk CAM's most stable doped structures by enumerating interstitial and substitution doping sites and calculating the corresponding defect formation energy, in a parallel and high-throughput fashion.

$$E^f(X) = E_{\text{tot}}(X) - E_{\text{tot}}(\text{bulk}) - \sum_i n_i \mu_i \qquad (1)$$

where $E_{\text{tot}}[\text{X}]$ is the total energy of the doped cell in its relaxed geometry. $E_{\text{tot}}[\text{bulk}]$ is the total energy of the pristine cathode material and $\mu(\text{x})$ is the chemical potential of the species exchanged.

From the most stable doped configurations, we compute the full voltage profile by iteratively depleting the mobile cation and allowing relaxation of the atomic structure. The voltage profile is then directly derived from the energies of the intermediate structures:

$$V = -\frac{\mu_{Li}^{cathode} - \mu_{Li}^{anode}}{zF} \qquad (2)$$

Here, $\mu_{Li}^{cathode}$ and $\mu_{Li}^{anode}$ represent the chemical potential of lithium in the cathode and anode materials, respectively, while $z$ is the number of electrons transferred per lithium ion and $F$ is the Faraday constant.

After computing the full voltage profile, we ensure that the doped delithiated structure remain stable throughout the charge by computing the energy above hull considering possible impurity phases.

## 3. Results

For each CAM, our ML-accelerated workflow enables the enumeration of doping and co-doping sites in the bulk structure representing 300k MLIP calls per, as illustrated in Fig. 1. These initial computations determine the most stable dopant structures for electrochemical properties computations. Computing the full voltage profile requires about 1 million MLIP calls as we need to sequentially delithiate the structure by enumerating each possible next vacancy site and relaxing there structure. We generated this screening data for 15 CAMs across 50 dopants, demonstrating that universal MLIPs enable accurate, cost-effective large-scale screening of complex properties.

We highlight the first step of our workflow Figure 2, where we show the formation energy computed for substitutional or interstitial doping of bulk $LiFePO_4$. Our pipeline identifies stable substitutional doping sites across a large portion of the periodic table. We subsequently compute the full voltage profile for delithiation of the cathodes revealing that doping can improve electrochemical performances. We identify among the doped structures the one combining negative formation energy (stable doped configuration) and improved (higher) voltage that remain stable throughout delithiation. After validating our prediction running our pipeline with DFT energy predictions and relaxations, we select the top candidates for experimental synthesis and characterization. We measure XRD spectra to confirm the synthesized material's structure, and characterize the electrochemical properties of the doped cathodes.

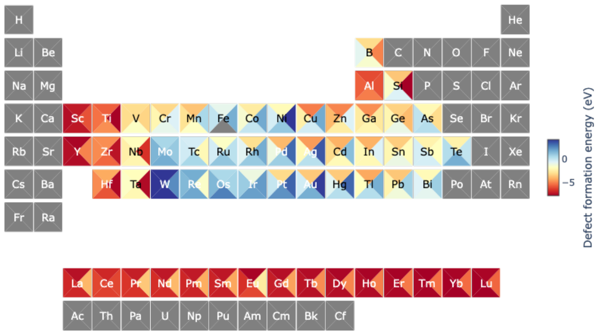

Fig. 2: Formation energies of dopants in $LiFePO_4$ at different lattice sites. Heatmap showing the formation energy (eV) for each dopant occupying four possible positions: Fe substitution (top), P substitution (right), Li substitution (bottom), or interstitial site (left). Colors indicate energy magnitude in electron-volt, with lower values corresponding to more favorable dopants incorporation: warm colors indicate favorable dopant incorporation.

## 4. Related work

Previous screening efforts leveraging DFT for CAMs have often been limited to a narrow set of doping sites [10], or a small number of dopant species [11, 12, 13] and cathodes.

Early works have validated the accuracy of MLIPs against DFT for fundamental properties such as formation energies and lattice parameters [14], as well as for more complex tasks like Li diffusion and X-ray spectrum prediction during delithiation [15]. Furthermore, MLIPs have facilitated the study of phase transformations via molecular dynamics [16], and the high-throughput screening of solid-state electrolytes [17, 18] is a good example of a succesfull aplication of MLIPs.

However, despite these advances, most existing studies rely on potentials trained on tailored datasets or limit "universal" MLIPs evaluation to specific chemical systems [19], while large scale transferability has yet to be demonstrated for doped CAMs. In this work, we propose a scaled-up, high-throughput framework leveraging foundational models of matter. We demonstrate their ability to provide accurate descriptions of chemical phenomena leading to the discovery of better Li-ion batteries.

## References

[1] Steven C Erwin, Lijun Zu, Michael I Haftel, Alexander L Efros, Thomas A Kennedy, and David J Norris. Doping semiconductor nanocrystals. *Nature*, 436(7047):91–94, 2005.

[2] Juan An, Liyi Shi, Guorong Chen, Musen Li, Hongjiang Liu, Shuai Yuan, Shimou Chen, and Dengsong Zhang. Insights into the stable layered structure of a li-rich cathode material for lithium-ion batteries. *Journal of Materials Chemistry A*, 5(37):19738–19744, 2017.

[3] Wenjun Jiang, Chunxiao Zhang, Yuzhang Feng, Bo Wei, Libao Chen, Ruifeng Zhang, Douglas G Ivey, Peng Wang, and Weifeng Wei. Achieving high structure and voltage stability in cobalt-free li-rich layered oxide cathodes via selective dual-cation doping. *Energy Storage Materials*, 32:37–45, 2020.

[4] Huaming Qian, Haoqi Ren, Ying Zhang, Xianfeng He, Wenbin Li, Jingjing Wang, Junhua Hu, Hong Yang, Hirbod Maleki Kheimeh Sari, Yu Chen, et al. Surface doping vs. bulk doping of cathode materials for lithium-ion batteries: a review. *Electrochemical Energy Reviews*, 5(4):2, 2022.

[5] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

[6] Alexander Urban, Dong-Hwa Seo, and Gerbrand Ceder. Computational understanding of li-ion batteries. *npj Computational Materials*, 2(1):1–13, 2016.

[7] Qiu He, Bin Yu, Zhaohuai Li, and Yan Zhao. Density functional theory for battery materials. *Energy & Environmental Materials*, 2(4):264–279, 2019.

[8] Brandon M Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R Kitchin, Daniel S Levine, et al. Uma: A family of universal models for atoms. *arXiv preprint arXiv:2506.23971*, 2025.

[9] Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, C Lawrence Zitnick, and Zachary W Ulissi. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771*, 2024.

[10] Hongliang Zhang, Yang Gong, Jie Li, Ke Du, Yanbing Cao, and Jiaqi Li. Selecting substituent elements for limnpo4 cathode materials combined with density functional theory (dft) calculations and experiments. *Journal of Alloys and Compounds*, 793:360–368, 2019.

[11] Lingling Fang, Ming Wang, Qinghua Zhou, Huihui Xu, Wei Hu, and Huili Li. Suppressing cation mixing and improving stability by f doping in cathode material linio2 for li-ion batteries: First-principles study. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 600:124940, 2020.

[12] Muhammad Hilmy Alfaruqi, Seokhun Kim, Sohyun Park, Seulgi Lee, Jun Lee, Jang-Yeon Hwang, Yang-Kook Sun, and Jaekook Kim. Density functional theory investigation of mixed transition metals in olivine and tavorite cathode materials for li-ion batteries. *ACS applied materials & interfaces*, 12(14):16376–16386, 2020.

[13] Zahra Moradi, Amir Heydarinasab, and Farshid Pajoum Shariati. First-principle study of doping effects (ti, cu, and zn) on electrochemical performance of li2mno3 cathode materials for lithium-ion batteries. *International Journal of Quantum Chemistry*, 121(4):e26458, 2021.

[14] Dohyeong Kwon and Duho Kim. Machine learning interatomic potentials in engineering perspective for developing cathode materials. *Journal of Materials Chemistry A*, 12(35):23837–23847, 2024.

[15] Miran Ha, Amir Hajibabaei, Dong Yeon Kim, Aditya Narayan Singh, Jeonghun Yun, Chang Woo Myung, and Kwang S Kim. Al-doping driven suppression of capacity and voltage fadings in 4d-element containing li-ion-battery cathode materials: machine learning and density functional theory. *Advanced Energy Materials*, 12(30):2201497, 2022.

[16] Peichen Zhong, Bowen Deng, Shashwat Anand, Tara Mishra, and Gerbrand Ceder. Modeling phase transformations in mn-rich disordered rocksalt cathodes with machine-learning interatomic potentials. *Physical Review Materials*, 9(10):105404, 2025.

[17] Ryohto Sawada, Kosuke Nakago, Chikashi Shinagawa, and So Takamoto. High-throughput investigation of stability and li diffusion of doped solid electrolytes via neural network potential without configurational knowledge. *Scientific Reports*, 14(1):11602, 2024.

[18] Xingyu Guo, Zhenbin Wang, Ji-Hui Yang, and Xin-Gao Gong. Machine-learning assisted high-throughput discovery of solid-state electrolytes for li-ion batteries. *Journal of Materials Chemistry A*, 12(17):10124–10136, 2024.

[19] Nian Ran, Chengbo Li, Qinwen Cui, Dezhen Xue, and Jianjun Liu. Dynamic oxygen-redox evolution of cathode reactions based on the multistate equilibrium potential model. *npj Computational Materials*, 11(1):208, 2025.