LeMat-Rho: High-Fidelity Charge Density Dataset for Machine Learning and Atomistic Materials Modeling

Richard Tran[1], Martin Siron[1], Georgia Channing[2], Mathilde L. D. Franckel[1,3], Guilherme Penedo[2], Alexandre Duval[1], Jonathan Schmidt[4]
[1]Entalpic, Paris, FR
[2]Hugging Face
[3]Department of Materials, Imperial College, London, UK
[4]Department of Materials, EPFL, Lausanne CH

m.franckel25@imperial.ac.uk

Over the last decade, several density-functional theory (DFT) crystal-structure databases have been created, largely using the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional [1]. Their recent consolidation into the LeMaterial repository has produced a comprehensive, unified collection of crystal structures. With LeMat-Rho, we aim to address two key challenges: (i) improving the fidelity of the LeMatBulk dataset with respect to exchange–correlation functional approximations, and (ii) overcoming the scarcity of charge-density data in large-scale materials dataset.

While the existing PBE LeMat-Bulk dataset provides a useful foundation for high-throughput searches the more expensive r²SCAN functional enables in significantly improved geometries and thermodynamic stability information. An update of the dataset to r²SCAN will lay a foundation for a next generation of materials discovery efforts and machine learning benchmarks.
In condensed matter, charge densities refer to the spatial distribution of electrons, providing insight into overarching electronic properties and chemical bonding. Practically, they form the starting and end-point of DFT calculations, the majority of the DFT cost resulting from the self-consistent cycle iterating from the starting to the final density. While most recent attention has been on, Machine Learning Interatomic Potentials (MLIPs) using only the energy and force data resulting from DFT the electronic structure data has been largely neglected, primarily due to the substantial storage and computational costs associated with charge-density data. However, since DFT as the state-of-the-art *ab-initio* method fundamentally operates on charge density as its central variable, direct access to large-scale, high-quality charge-density datasets and resulting ML models offer substantial advantages: they can enable more reliable initialization of DFT calculations and reduce the number of self-consistent iterations required for convergence. Despite its significance, the materials science community lacks a large, standardized, and high-accuracy database of electronic charge densities for crystal structures beyond the PBE functional.
In this work, we present a crystalline material dataset computed with the revised regularized strongly constrained and appropriately normed (r²SCAN) meta-GGA functional [2] for inorganic materials, including charge densities, aimed at accelerating high-throughput ab initio workflows and enabling improvements in charge-aware machine learning models.

Methods

Leveraging the LeMaterial repository [3], which contains the a comprehensive collection of crystal structures optimized with the widely used Perdew–Burke–Ernzerhof (PBE) functional [1], we perform two r²SCAN relaxations followed by a static calculation on a diverse subset of structures sampled from materials close to the LeMaterial convex hull of thermodynamic stability. This process is still ongoing with the end goal of providing a full r²SCAN convex hull. DFT settings were kept compatible with the MATPES dataset to allow for the unified usage of the datasets. The process was automated using *Atomate2*[4].
We also publish the electronic densities of the relaxed geometries.

Results and Analysis

The resulting dataset comprises of over 60k inorganic materials. The charge densities are compressed and stored in cloud-optimized Parquet format. Where possible, entries are integrated with the LeMaterial fetcher as open data under a CC-BY license interface to streamline access. LeMat-Rho supports scalable, serverless analysis and model training through cloud platforms.

Additionally, we evaluate the dataset as a resource for machine learning benchmarking. Existing charge-density prediction models, including ChargE3Net [5], are tested on the dataset to quantify predictive accuracy, generalization to unseen structures, and potential reductions in self-consistent DFT iterations. This establishes the dataset as a baseline for future ML studies of electronic densities. In parallel, we find that starting from our converged charge density, reduces DFT self-consistent iterations by approximately 26%.

To explore the chemical and physical significance of the charge densities, we perform post-processing analyses, such as Bader charge partitioning [6] and electron localization function (ELF) evaluation [7]. These analyses allow us to identify trends in charge distribution, bonding characteristics, and electronic structure across a diverse set of inorganic materials.

Lastly, we benchmarked the accuracy of formation energies and lattice volume calculated with r2SCAN against experimental quantities compiled in the OQMD database[8]. The formation energies calculated in this work illustrated a markedly smaller mean absolute error when compared to the formation energies computed using the PBE functionals from LeMat-bulk.

Outcome and Future Works

LeMat-Rho constitutes the largest r²SCAN-based materials database, the first open electronic density database at this level of theory. Ultimately, it aims to accelerate materials discovery, improves predictive modeling, and provides a foundational resource for next-generation quantum-accurate machine learning methods. Indeed, we hope to extend our work by fine-tuning universal interatomic potentials based on r²SCAN, improving the accuracy and efficiency of geometry optimization and energy predictions.

Figures

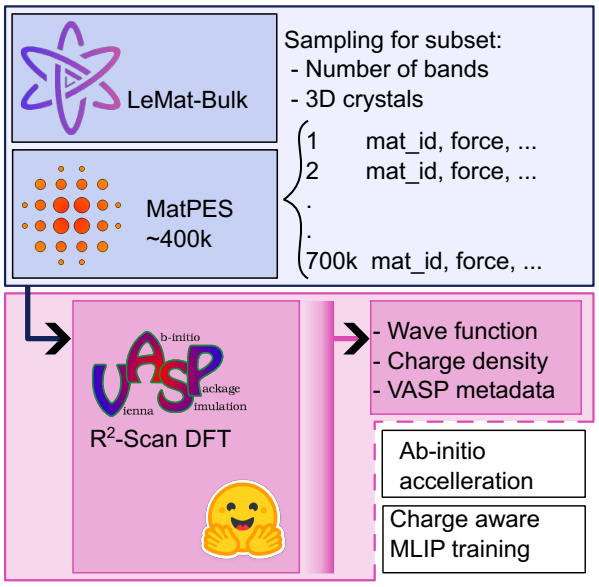

Figure 1. Illustration of the workflow for generating a crystalline material dataset with charge density. Starting from the data sources: LeMat-Bulk and MatPES**,** a select subset are computing utilising the VASP package to calculate material properties including wave function, and charge density.

References

[1] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.*, 77 (1996) 3865–3868

[2] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, *J. Phys. Chem. Lett.*, 11 (2020) 8208–8215

[3] LeMaterial Project: *Open Materials Database*, https://le-material.org

[4] A. M. Ganose, et al., *Digital Discovery*, 7 (2025) 1944-1973

[5] Koker, T., Quigley, K., Taw, E. *et al.*, *npj Comput Mater*, 161 (2024) 10

[6] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press (1990)

[7] D. Becke, K. E. Edgecombe, *J. Chem. Phys.*, 92 (1990) 5397–5403

[8] S. Kirklin, et al., *npj computational materials*, 11 (2015) 15