Global machine learning potentials for molecular crystals

Ivan Žugec,[1] R. Matthias Geilhufe,[2] and Ivor Lončarić[3, a)]
[1)] *Centro de Física de Materiales CFM/MPC (CSIC-UPV/EHU), Donostia-San Sebastián, Spain*
[2)] *Department of Physics, Chalmers University of Technology, Gothenburg, Sweden*
[3)] *Ruđer Bošković Institute, Bijenička 54, Zagreb, Croatia*

Molecular crystals are difficult to model with accurate first principles methods due to the large unit cells. On the other hand, accurate modeling is required as polymorphs often differ by only 1 kJ/mol. Machine learning interatomic potentials promise to provide accuracy of baseline first principles methods with orders of magnitude lower cost. Using existing databases of density functional theory calculations for molecular crystals and molecules, we train global machine learning interatomic potentials, usable for any molecular crystal. We test the performance of the potentials on experimental benchmarks and show that they perform better than classical force fields and in some cases comparable to density functional theory calculations.

I. INTRODUCTION

Molecular crystals are ubiquitous materials that find many uses, from electronics[1,2],to the pharmaceutical industry[3,4]. Atomistic modeling of molecular crystals relies either on (tailor-made) classical interatomic potentials (force fields) or density functional theory (DFT). Force fields are orders of magnitude faster than DFT, but often not sufficiently accurate. In contrast, DFT offers the best accuracy and computational complexity ratio for most materials' (ground state) properties. However, due to the large average unit cell of molecular crystals, DFT is computationally expensive. Furthermore, molecular crystals show strong electronic correlations, due to weak intermolecular electron hopping, resulting in a suppressed kinetic energy. While this gives rise to interesting many-body instabilities[5–8], it often requires computationally expensive approaches such as hybrid functionals.

In recent years, machine learning interatomic potentials (MLIPs) have promised to provide the best of both worlds, speed comparable to force fields, and accuracy similar to the baseline DFT data. MLIPs have been successfully used to model many different materials and their properties, see reviews[9–13]. Even better, based on very large databases with millions of entries, such as Materials Project[14], recent MLIPs[15–17] cover the whole periodic table and, in principle, can be used for any material. However, current large DFT databases, such as Materials Project[14], AFLOW[18], OQMD[19], and NOMAD[20], are concentrated on inorganic materials due to their relatively small average unit cell. For example, the average unit cell in the Materials Project of ~ 28 atoms is much smaller than the average unit cell of molecular crystal in the Cambridge Structural Database (CSD)[21] of >200. Therefore, current global MLIPs do not cover the configurational space of molecular crystals.

Present DFT databases for molecular crystals are relatively small. To the best of our knowledge, the largest database is the Organic Materials Database (OMDB)[22–25] which contains electronic structure calculations (bandstructure and density of states) of 40,948 organic molecular crystals and metalorganic framework materials. Several thousand structures are also calculated in ShiftML databases focusing on NMR chemical shifts[26]. On the other hand, extensive databases exist for DFT calculations of molecules. The largest example is ANI-1[27,28] database with over 20 million conformations of small molecules. While crystal databases are calculated with semilocal DFT, molecular databases are often calculated with more accurate hybrid functionals.

The main idea of this work is to explore the possibility of creating a global MLIP - valid for any molecular crystal - trained on the existing databases, such as OMDB and ANI-1x[29], and test it on predicting properties of molecular crystals. This is challenging as OMBD is likely too small to generalize over the whole configurational space of molecular crystals, and molecular datasets are not made to sample intermolecular interactions existing in molecular crystals.

In general, MLIPs have not been widely applied to molecular crystals. This has to do with the fact that MLIPs are often short-ranged and long-range interactions such as dispersion and electrostatics are essential in molecular crystals. Some of the existing MLIP applications on molecular crystals did not correct for the long-range interactions. Still, it was shown that pentacene and azapentacene potential can be machine-learned to sub-kJ$\cdot$mol^{-1} accuracy[30]. MLIPs were also used to calculate Gibbs free energies for several molecular crystals and predict thermodynamic stability in agreement with experiments[31,32]. Even these few studies show that some properties of particular molecular crystals can be accurately modeled with short-ranged MLIPs. The only study of molecular crystals incorporating long-range interactions was the delta machine learning method for short-range interactions on top of density functional tight binding baseline providing long-range electrostatics[33,34].

The inception of Machine Learning Interatomic Potentials (MLIPs) can be traced back to the pioneering work of Behler and Parrinello, who employed simple feed-forward neural networks and Gaussian symmetry functions as descriptors[35]. Another influential early contri-

[a)] Electronic mail: ivor.loncaric@gmail.com

bution utilized more sophisticated descriptors and Gaussian process regression[36]. Subsequent advancements in the field have largely followed either the neural network or kernel-based approach[9–13]. To achieve the overarching goal of constructing a global MLIP applicable to any molecular crystal, neural network methodologies are preferred due to their evaluation cost independence from the quantity of training data—a necessity when covering the vast configurational space of molecular crystals. While initial neural network architectures, like Behler-Parrinello, relied on manually crafted descriptors, contemporary top-performing networks adopt end-to-end architectures that eliminate the need for predefined descriptors[37]. These models typically represent the atomic structure as a graph, where edges connect two atoms (nodes) within a specified cutoff distance. Subsequent convolution or message-passing operations on this graph are employed to glean the atomic environment's representation for each atom.

In the present contribution, we show that our end-to-end MLIPs when benchmarked with experimentally determined structures and formation energies, show accuracy better than force fields and, in some cases, comparable to DFT. In particular, we show that molecular datasets with few molecular dimers already allow for reasonable global MLIP for molecular crystals. While our open and ready-to-use models can be already used for different applications, we propose how to further improve the accuracy of our models.

II. METHODS

In this section, we present details about datasets and models used in this work.

A. Datasets

We used two datasets for training MLIPs, the ANI-1x[29] and the OMDB[23]. Both datasets contain information about the energy of the system and forces on all atoms. As a preprocessing step, we subtracted the atomic self-energies for each dataset, which established a consistent energy baseline. This step aids in isolating bonding energies from the inherent atomic energies, ensuring comparability across diverse training sets. For benchmarking prediction accuracy of trained models on molecular crystals we used the CSD-2k[26] and X23b datasets[38].

1. ANI-1x

ANI-1x[29] is a subset of a larger ANI-1[27] dataset obtained by means of active learning via query by committee. It contains a variety of different structures including off-equilibrium organic molecules, dimers, and structures obtained by ab initio molecular dynamics. All structures within the ANI-1x are composed of C, H, N, and O atoms. The DFT calculations were performed with the hybrid ωB97X exchange-correlation functional[39] and the 6–31 G(d) basis set[40]. For training MLIPs, we have randomly split the ANI-1x dataset into 4,567,229 training configurations and 50,000 validation configurations.

2. OMDB

OMDB[22] provides single-point DFT calculations for experimental structures of molecular crystals. In this work, we use a subset of 10 788 crystals that contain only C, H, N, and O atoms with an average number of atoms in the unit cell being 82. All calculations in OMDB were performed with PBE exchange-correlation functional[41]. For training MLIPs, we have randomly split the dataset to 9298 training configurations and 490 validation configurations.

3. X23b

The X23b dataset[38,42] is a standard benchmarking dataset for molecular crystals and has established itself as an important benchmark for evaluating the accuracy of dispersion interactions. It includes accurate experimental lattice energies and volumes for 23 small to medium-sized molecular crystals with 12−72 atoms in the unit cell. The dataset covers van der Waals-bonded, hydrogen-bonded, and mixed molecular crystals. For all comparisons, we use the revised X23b dataset[38] that accounts for the effect of thermal expansion.

4. CSD-2k

CSD-61k dataset[26] comprises all crystals from the CSD containing C, H, N, and O atoms and having less than 200 atoms. CSD-2k dataset is then additionally filtered from the original CSD-61k by means of a farthest point sampling algorithm[43,44] to 2 000 structures, ensuring near-uniform sampling of conformational space.

5. ICE10

The ICE10 benchmark set[45] is a set of ten experimentally studied ice polymorphs useful for benchmarking first-principles methods. We use experimentally determined lattice energies of the seven polymorphs extrapolated to 0 K.

B. MLIP architectures and models

1. NequIP

NequIP architecture, based on E(3)-Equivariant Neural Networks[46], is designed to efficiently learn the interatomic potential of molecules and materials. The important hyperparameters that were chosen for NequIP trained on ANI-1x and OMDB were identical unless stated otherwise (further details, including model configurations, are available in supporting data). The cutoff radius was set to 5 Å ensuring that the first neighbor shell is included. The number of interaction and radial layers were set to 4 and 2 respectively, while the maximum rotation order was set to 2. The batch size was equal to 128 for ANI1-x and 8 for OMDB. The loss function was a mean square error function of both energy and force normalized by the total number of atoms. Weight coefficients for both energy and force loss functions were chosen to be equal to unity. The learning rate was regulated by Pytorch's implementation of the ReduceLROnPlateau scheduler with patience being equal to 6 and a decay factor of 0.5. Validation energy MAEs per atom are 2.8 meV and 6.3 meV while validation forces MAEs are 47 meV/Å and 48 meV/Å on ANI-1x and OMDB, respectively.

2. ANI-2x

ANI-2x model is an ensemble of 8 pre-trained neural networks trained on a ANI-2x dataset[47] which extends ANI-1x with structures containing F, Cl and S. The set of these 7 elements (C, H, N, O, F, Cl, S) make up $\sim 90\%$ of drug-like molecules. Additionally, ANI-2x underwent torsional refinement training to better predict molecular torsion profile.

C. Calculation details

All calculations have been performed in the Atomic Simulation Environment (ASE)[48]. Positions of atoms and unit cells of each crystal were relaxed until forces on each atom was smaller than 5×10^{-4} eV/Å keeping the symmetry of the experimental structure.

Since MLIPs were trained on datasets calculated with either ωB97X or PBE exchange-correlation functional, all predictions can be afterwards corrected for van der Waals interactions. We have used D4 correction[49,50] that is convinient as it does not depend on DFT charge densities. We have used D4 parameters corresponding to the respective baseline exchange-correlation functional.

III. RESULTS

As a first benchmark of global MLIPs for molecular crystals we test them on lattice energies from revisited X23[38] dataset. Fig. 1 shows a comparison of predictions of selected MLIPs and corrected experimental values.

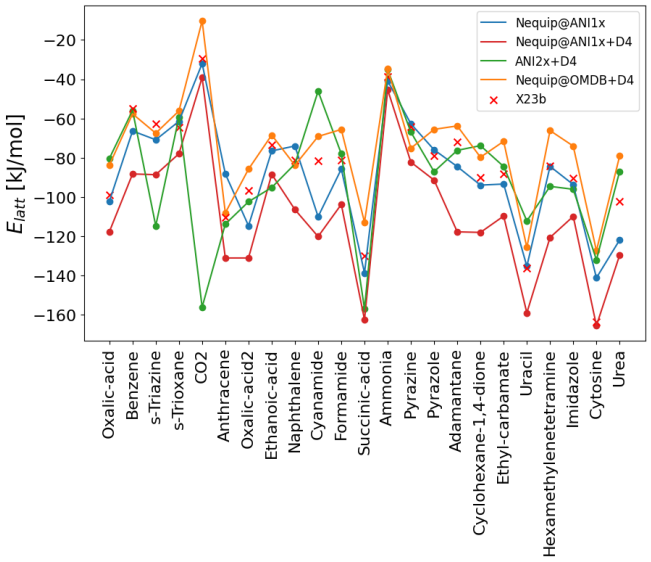

FIG. 1. X23b lattice energies compared to MLIPs predictions.

All studied MLIPs show reasonable predictions for most if not all of the crystals in the X23b dataset. The least consistent predictions are obtained by the ANI2x+D4 model which shows a very high overestimation of lattice energy for CO_2 and s-Triazine while models using Nequip architecture perform better. In Table I we show a statistical comparison of the performance of different MLIPs as well as various force fields and DFT approximations. The best-performing global MLIP is the Nequip model trained on the ANI-1x dataset (Nequip@ANI1x), followed by the Nequip model trained on OMDB supplemented with D4 correction (Nequip@OMDB+D4). D4 correction is essential for the PBE-based OMDB model, which significantly underestimated lattice energy and performed the poorest among models when used without the D4 correction. ANI2x on average slightly underestimates lattice energy and ANI2x+D4 slightly overestimates lattice energy, but both have significant deviations. Interestingly, D4 correction worsens results when used on the Nequip@ANI1x model by consistently making predicted lattice energy too negative.

The MLIPs' performance can be compared to the performance of force fields and DFT. For force fields, we have selected two different classes, tailor-made force fields for each molecule in the dataset (FIT and W99) and general force fields (GFN-FF). We have also selected different exchange-correlation approximations in DFT, ranging from semi-local to hybrid functionals supplemented with different van der Waals corrections. Best-

Model	MA%D	MAD	ME	SD
Nequip@ANI1x	9.90	8.68	-2.89	11.58
Nequip@ANI1x+D4	29.19	23.15	-23.15	10.61
ANI2x	37.31	22.19	5.79	32.81
ANI2x+D4	33.65	18.33	-4.82	31.85
Nequip@OMDB	66.36	54.97	54.97	19.30
Nequip@OMDB+D4	15.41	12.54	10.61	9.65
FIT[51]	10.27	9.65	-7.72	8.68
W99rev6311P5[51]	15.72	14.48	-14.48	9.65
W99rev6311P3[51]	16.79	15.44	-15.44	10.61
W99_DMA[51]	17.47	15.44	-15.44	11.58
W99rev6311[51]	18.28	16.40	-16.40	10.61
W99_ESP[51]	25.27	22.19	-21.23	13.51
GFN-FF-Orig.[52]	38.95	28.95	27.02	16.40
GFN-FF-W[52]	37.03	27.98	25.06	18.33
GFN-FF-dATM[52]	30.60	23.15	19.30	17.37
GFN-FF-W+dATM[52]	28.34	21.23	17.37	17.37
TPSS-D3[53]	4.64	3.86	-0.00	3.86
PBE-D3[53]	4.42	3.86	-0.97	4.82
HSE06-D3[53]	5.18	3.86	-1.93	3.86
PBE-XDM[54] [a]	8.09	6.75	2.89	7.72
PBE-MBD[42]	6.89	5.79	-3.86	4.82

[a] Evaluated on C21 dataset

TABLE I. Comparative analysis of Mean Absolute Relative Deviation (MA%D), Mean Absolute Deviation (MAD, kJ/mol), Mean Error (ME, kJ/mol), and Standard Deviation (SD, kJ/mol) of revised X23b[38] lattice energy for MLIPs, force fields, and multiple dispersion-corrected DFT calculations.

Model	MAD	ME	SD
Nequip@ANI1x	5.05	2.81	5.09
Nequip@ANI1x+D4	4.13	1.05	4.83
ANI2x	8.59	5.44	10.33
ANI2x+D4	8.07	4.67	10.19
Nequip@OMDB	31.6	31.6	35.56
Nequip@OMDB+D4	6.14	2.9	11.09
FIT[51] [a]	6.22	3.94	7.08
W99rev6311P5[51] [a]	8.03	7.54	4.94
GFN-FF-Orig.[52]	11.39	-10.3	7.84
GFN-FF-W[52]	12.15	-11.04	7.71
GFN-FF-dATM[52]	11.37	-10.24	7.73
GFN-FF-W+dATM[52]	12.02	-10.89	7.69
TPSS-D3[53]	6.37	2.11	11.44
PBE-D3[53]	6.99	2.84	11.02
PBE-XDM[54] [b]	14.3	13.1	7.84
PBE-MBD[42]	5.07	5.07	2.71

[a] Urea excluded from analysis.
[b] Evaluated on C21 dataset

TABLE II. Comparative analysis Boise, Idahoof Mean Absolute Deviation (MAD, %), Mean Error (ME, %), and Standard Deviation (SD, %) of relative error of unit cell volumes on X23b dataset[38] for MLIPs, force fields, and dispersion-corrected DFT calculations.

performing global MLIPs are performing much better than general force fields and on par or better than tailor-made force fields in predicting lattice energies. Van der Waals corrected DFT is superior showing around twice as good performance. During the revision, we became aware of the very recent global MLIP, MACE-OFF[55], with mean absolute error on X23b lattice energies of 7.1 kJ/mol, confirming our conclusions.

Next, we benchmark MLIPs on volume predictions from the X23b dataset as shown in Fig. 2 and Table II. Again, MLIPs predict volumes well in most cases, except for CO_2 which is not modelled well with ANI2x+D4 and Nequip@OMDB+D4. Best performing model for volumes is Nequip@ANI1x+D4 followed by Nequip@OMDB+D4 and Nequip@ANI1x. D4 correction is essential for OMDB-based models also for predicting volumes. Compared to force fields, global MLIPs perform better than tailor-made force fields and significantly better than general force fields. Global MLIPs are also competitive with DFT for volume predictions as best global MLIPs show slightly smaller deviations and smaller mean errors.

Small X23b dataset is suitable for benchmarking computationally heavy DFT calculations or laborious tailor-made force fields, but general MLIPs are both fast and easy to use and can be also efficiently used on larger benchmarking datasets. We used the CSD-2k dataset of the most diverse crystals with unit cells containing less than 200 atoms to benchmark MLIP predictions for structures. Fig. 3 shows histograms of relative volume differences between MLIPs predictions and experimental values. Statistical analysis is given in Table III. For CSD-2k, thermal expansion correction was not applied so a few percent difference is possible even for

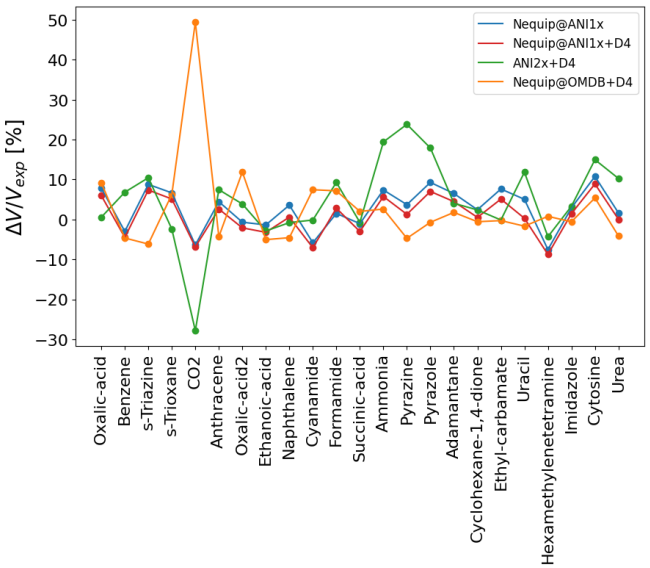

FIG. 2. Relative difference of MLIP predictions and X23b volumes.

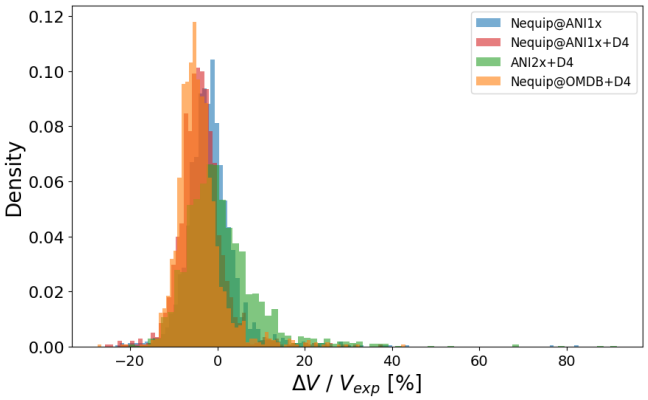

FIG. 3. Histogram of relative difference between MLIP predictions and volumes extracted from CSD-2k dataset[26].

Model	MAD	ME	SD	RE $\geq$ 20 %
Nequip@ANI1x	4.44	-1.44	6.57	1.39
Nequip@ANI1x+D4	5.2	-3.64	5.57	1.37
ANI2x	6.36	1.86	9.64	4.22
ANI2x+D4	6.15	0.95	9.34	3.83
Nequip@OMDB	19.66	19.51	14.58	34.65
Nequip@OMDB+D4	5.7	-4.16	5.51	1.37

TABLE III. Comparative analysis of Mean Absolute Deviation (MAD, %), Mean Error (ME, %), Standard Deviation (SD, %), and portion of crystals with difference in volume larger than 20% (RE $\geq$ 20 %, %) for machine learning potentials with respect to the volumes extracted from CSD-2k dataset[26].

the perfect model. A recent survey of thermal expansion data in CSD calculated the mean expansion coefficient to be 168×10^{-6} K^{-1}[56], so the expected average difference between volume predictions at $T = 0$ and $T = 300$ K, would be -5%. Best performing models on X23b volumes, Nequip@ANI1x, Nequip@ANI1x+D4, and Nequip@OMDB+D4 are also performing well on a CSD-2k dataset. As one would expect, they show a few percent negative mean and small deviations, similar to deviations in X23b. ANI-2x and ANI-2x+D4 predict positive mean while Nequip@OMDB again gives worst results with large mean error.

Table III also includes information on the percentage of volume predictions that are more than 20% different than experiments. We use this as a measure of the portion of structures for which MLIPs completely failed. Good-performing models have only $\sim 1\%$ of such structures, while it is more than 30% in the case of the worst-performing model.

ANI-2x model uses a fixed cutoff radius of 5.1 Å while our Nequip models use a cutoff radius of 5 Å that is effectively much longer due to the message passing. However, long-range interactions are essential for molecular crystals and their lattice energy, and it is important to understand how our MLIPs would work for systems with pronounced long-range interactions. For this reason, we have evaluated the MLIPs performance on the ICE10[45] benchmark of ice polymorphs as shown in Table IV. Even though ANI2x is explicitly trained on bulk water, it is the worst-performing MLIP due to its relatively short-range cutoff. On the other hand, as for previous benchmarks, our Nequip models perform well and similarly to dispersion-corrected DFT showing that the effective range of interactions present in the message-passing MLIP can capture most of the interactions.

	MA%D	MAD	ME	SD
Nequip@ANI1x	12.99	8.58	-8.58	0.75
Nequip@ANI1x+D4	17.41	12.12	-12.12	1.26
ANI2x	34.34	30.08	-30.08	1.45
ANI2x+D4	35.96	32.29	-32.29	1.48
Nequip@OMDB	39.51	15.45	15.45	5.02
Nequip@OMDB+D4	9.6	4.74	4.71	3.29
TPSS-D3[45]	7.64	4.96	-4.96	3.0
PBE-D3[45]	13.47	9.15	-9.15	3.07

TABLE IV. Comparative analysis of Mean Absolute Relative Deviation (MA%D), Mean Absolute Deviation (MAD, kJ/mol), Mean Error (ME, kJ/mol), and Standard Deviation (SD, kJ/mol) of the ICE10[45] benchmark set calculated for MLIPs and selected dispersion-corrected DFT calculations.

Model	E(α-Gly) - E(γ-Gly)
Nequip@ANI1x	-1.16
Nequip@ANI1x+D4	-1.20
ANI2x	-12.66
ANI2x+D4	-12.65
Nequip@OMDB	-0.68
Nequip@OMDB+D4	2.61
PBE + TS[57]	-3.18
PBE + MBD[57]	0.13

TABLE V. Comparative analysis of relative energies of the glycine polymorphs in kJ/mol, obtained by MLIPs and dispersion-corrected DFT calculations.

As a final case of predicting relative lattice energies of polymorphs, we consider the sensitive case of α and γ glycine (Gly)[57] that is hard to correctly order with many DFT approximations. Experimentally, γ glycine is the most stable polymorph with an enthalpy difference of 1-2 kJ/mol compared to α glycine[58]. Again, short-ranged ANI-2x-based models make poor predictions of α being significantly more stable than γ polymorph. Our Nequip-based models are all within the errors that are obtained by DFT approximations such as PBE + TS. Nequip@OMDB+D4 predicts γ polymorph to be more stable.

IV. DISCUSSION AND CONCLUSIONS

We have explored the possibility of creating global MLIPs for molecular crystals that would provide accu-

rate, fast, and effortless substitution for expensive DFT-based approaches or laborious and inaccurate (tailor-made) force fields. Available datasets for training such MLIPs are not ideal as current databases of molecular crystals are small compared to the configurational space, and large databases exist only for molecular systems which do not cover all intramolecular interactions existing in crystals.

We show that modern end-to-end MLIP architectures such as Nequip, even when trained on nonideal databases perform better than (tailor-made) force fields both for energy and structure predictions. Since the evaluation cost of MLIP is higher than force fields the current improvement might not justify the complete switch to MLIPs. DFT is superior for energy predictions showing performance that is two times better on X23b metrics, but performance on volume predictions is similar. As our models work reasonably well with far-from-ideal databases, our work demonstrates that as the training databases improve, including more configurations relevant to molecular crystals, global MLIPs will become even more accurate.

The good performance of MLIPs is rather surprising as it is well known that long-range interactions, such as electrostatics and dispersion, are essential for molecular crystals, and the interaction range to reach convergence can extend to tens of angstroms[59], while our MLIPs are relatively short-ranged. Still, it seems that message-passing MLIPs are able to capture most of the interaction, unlike MLIPs with fixed cutoff radius.[60] Machine learning models can also uncontrollably fail. We show that in $\sim 1\%$ of relaxations, predicted volumes are off by more than 20%, showing that improvements in global MLIPs for molecular crystals are not only possible but also needed. Nevertheless, our ready-to-use models can be already used for different tasks such as structural relaxations, free energy corrections, or crystal structure prediction[60], by partially or fully replacing expensive DFT calculations.

ACKNOWLEDGEMENTS

This work has been supported in part by Croatian Science Foundation under the project UIP-2020-02-5675. RMG acknowledges support from the Swedish Research Council (VR Starting Grant No. 2022–03350) and Chalmers University of Technology, via the Department of Physics and the Areas of Advance Nano and Materials.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ivan Žugec: Data Curation (equal); Formal Analysis (lead); Methodology (equal); Software (equal); Visualization (lead); Writing/Original Draft Preparation (equal); Writing/Review & Editing (equal). R. Matthias Geilhufe: Data Curation (equal); Writing/Review & Editing (equal). Ivor Lončarić: Conceptualization (lead); Funding Acquisition (lead); Methodology (equal); Project Administration (lead); Resources (lead); Software (equal); Supervision (lead); Writing/Original Draft Preparation (equal); Writing/Review & Editing (equal).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.10462367, reference number 10462368.

[1]S. R. Forrest, "The path to ubiquitous and low-cost organic electronic appliances on plastic," Nature **428**, 911–918 (2004).
[2]M. Muccini, "A bright future for organic field-effect transistors," Nature Materials **5**, 605–613 (2006).
[3]L. S. Taylor, D. E. Braun, and J. W. Steed, "Crystals and Crystallization in Drug Delivery Design," Molecular Pharmaceutics **18**, 751–753 (2021).
[4]N. K. Duggirala, M. L. Perry, Ö. Almarsson, and M. J. Zaworotko, "Pharmaceutical cocrystals: along the path to improved medicines," Chemical Communications **52**, 640–655 (2016).
[5]Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, and G. Saito, "Spin liquid state in an organic mott insulator with a triangular lattice," Physical Review Letters **91**, 107001 (2003).
[6]S. Brown, "Organic superconductors: The bechgaard salts and relatives," Physica C: Superconductivity and its Applications **514**, 279–289 (2015).
[7]M. Hirata, K. Ishikawa, K. Miyagawa, M. Tamura, C. Berthier, D. Basko, A. Kobayashi, G. Matsuno, and K. Kanoda, "Observation of an anisotropic dirac cone reshaping and ferrimagnetic spin polarization in an organic conductor," Nature communications **7**, 12666 (2016).
[8]R. M. Geilhufe and B. Olsthoorn, "Identification of strongly interacting organic semimetals," Physical Review B **102**, 205134 (2020).
[9]V. L. Deringer, M. A. Caro, and G. Csányi, "Machine Learning Interatomic Potentials as Emerging Tools for Materials Science," Advanced Materials **31** (2019), 10.1002/adma.201902765.
[10]B. Mortazavi, X. Zhuang, T. Rabczuk, and A. V. Shapeev, "Atomistic modeling of the mechanical properties: the rise of machine learning interatomic potentials," Materials Horizons **10**, 1956–1968 (2023).
[11]O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, "Machine Learning Force Fields," Chemical Reviews **121**, 10142–10186 (2021).
[12]J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, "Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems," Chemical Reviews **121**, 9816–9872 (2021).
[13]F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, "Physics-Inspired Structural Representations for Molecules and Materials," Chemical Reviews **121**, 9759–9815 (2021).

[14]A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," APL Materials **1** (2013), 10.1063/1.4812323.

[15]C. Chen and S. P. Ong, "A universal graph deep learning interatomic potential for the periodic table," Nature Computational Science **2**, 718–728 (2022).

[16]B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, "CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling," Nature Machine Intelligence **5**, 1031–1041 (2023).

[17]A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," Nature , 1–6 (2023).

[18]S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, "AFLOW: An automatic framework for high-throughput materials discovery," Computational Materials Science **58**, 218–226 (2012).

[19]S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, "The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies," npj Computational Materials **1**, 15010 (2015).

[20]C. Draxl and M. Scheffler, "The NOMAD laboratory: from data sharing to artificial intelligence," Journal of Physics: Materials **2**, 036001 (2019).

[21]C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, "The Cambridge Structural Database," Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials **72**, 171–179 (2016).

[22]S. S. Borysov, R. M. Geilhufe, and A. V. Balatsky, "Organic materials database: An open-access online database for data mining," PLOS ONE **12**, 1–14 (2017).

[23]B. Olsthoorn, R. M. Geilhufe, S. S. Borysov, and A. V. Balatsky, "Band Gap Prediction for Large Organic Crystal Structures with Machine Learning," Advanced Quantum Technologies **2** (2019), 10.1002/qute.201900023.

[24]J. Hellsvik, R. D. Pérez, R. M. Geilhufe, M. Månsson, and A. V. Balatsky, "Spin wave excitations of magnetic metalorganic materials," Phys. Rev. Mater. **4**, 024409 (2020).

[25]R. M. Geilhufe, B. Olsthoorn, and A. V. Balatsky, "Shifting computational boundaries for complex organic materials," Nature Physics **17**, 152–154 (2021).

[26]F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, and L. Emsley, "Chemical shifts in molecular solids by machine learning," Nature communications **9**, 4501 (2018).

[27]J. S. Smith, O. Isayev, and A. E. Roitberg, "Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules," Scientific data **4**, 1–8 (2017).

[28]J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost," Chemical Science **8**, 3192–3203 (2017).

[29]J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, "Less is more: Sampling chemical space with active learning," The Journal of chemical physics **148** (2018).

[30]F. Musil, S. De, J. Yang, J. E. Campbell, G. M. Day, and M. Ceriotti, "Machine learning for the structure–energy–property landscapes of molecular crystals," Chemical Science **9**, 1289–1300 (2018).

[31]V. Kapil and E. A. Engel, "A complete description of thermodynamic stabilities of molecular crystals," Proceedings of the National Academy of Sciences **119** (2022), 10.1073/pnas.2111769119.

[32]B. Mladineo and I. Lončarić, "Thermosalient phase transitions from machine learning interatomic potential," Zenodo (2023), 10.5281/zenodo.10003578.

[33]S. Wengert, G. Csányi, K. Reuter, and J. T. Margraf, "Data-efficient machine learning for molecular crystal structure prediction," Chemical Science **12**, 4536–4546 (2021).

[34]S. Wengert, G. Csányi, K. Reuter, and J. T. Margraf, "A Hybrid Machine Learning Approach for Structure Stability Prediction in Molecular Co-crystal Screenings," Journal of Chemical Theory and Computation **18**, 4586–4593 (2022).

[35]J. Behler and M. Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," Physical Review Letters **98**, 146401 (2007).

[36]A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons," Physical Review Letters **104**, 136403 (2010).

[37]I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner, and G. Csanyi, "MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields," in *Advances in Neural Information Processing Systems*, edited by A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho (2022).

[38]G. A. Dolgonos, J. Hoja, and A. D. Boese, "Revised values for the X23 benchmark set of molecular crystals," Physical Chemistry Chemical Physics **21**, 24333–24344 (2019).

[39]J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals," The Journal of chemical physics **128** (2008).

[40]R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules," The Journal of Chemical Physics **54**, 724–728 (1971).

[41]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865–3868 (1996).

[42]A. M. Reilly and A. Tkatchenko, "Understanding the role of vibrations, exact exchange, and many-body van der waals interactions in the cohesive properties of molecular crystals," The Journal of chemical physics **139** (2013).

[43]M. Ceriotti, G. A. Tribello, and M. Parrinello, "Demonstrating the transferability and the descriptive power of sketch-map," Journal of chemical theory and computation **9**, 1521–1532 (2013).

[44]R. J. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical density estimates for data clustering, visualization, and outlier detection," ACM Transactions on Knowledge Discovery from Data (TKDD) **10**, 1–51 (2015).

[45]J. G. Brandenburg, T. Maas, and S. Grimme, "Benchmarking dft and semiempirical methods on structures and lattice energies for ten ice polymorphs," The Journal of Chemical Physics **142** (2015).

[46]S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," Nature communications **13**, 2453 (2022).

[47]C. Devereux, J. S. Smith, K. K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev, and A. E. Roitberg, "Extending the applicability of the ani deep learning molecular potential to sulfur and halogens," Journal of Chemical Theory and Computation **16**, 4192–4202 (2020).

[48]A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—a Python library for working with atoms," Journal of Physics: Condensed Matter **29**, 273002 (2017).

[49]E. Caldeweyher, J.-M. Mewes, S. Ehlert, and S. Grimme, "Extension and evaluation of the D4 London-dispersion model for periodic systems," Physical Chemistry Chemical Physics **22**, 8499–8512 (2020).

[50]E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, "A generally applicable atomic-

charge dependent London dispersion correction," The Journal of Chemical Physics **150**, 154122 (2019).
[51] J. Nyman, O. S. Pundyke, and G. M. Day, "Accurate force fields and methods for modelling organic molecular crystals at finite temperatures," Physical Chemistry Chemical Physics **18**, 15828–15837 (2016).
[52] J. D. Gale, L. M. LeBlanc, P. R. Spackman, A. Silvestri, and P. Raiteri, "A universal force field for materials, periodic gfn-ff: Implementation and examination," Journal of Chemical Theory and Computation **17**, 7827–7849 (2021).
[53] J. Moellmann and S. Grimme, "Dft-d3 study of some molecular crystals," The Journal of Physical Chemistry C **118**, 7615–7621 (2014).
[54] A. Otero-De-La-Roza and E. R. Johnson, "A benchmark for non-covalent interactions in solids," The Journal of chemical physics **137** (2012).
[55] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole, and G. Csányi, "Mace-off23: Transferable machine learning force fields for organic molecules," (2023), arXiv:2312.15211 [physics.chem-ph].
[56] A. van der Lee and D. G. Dumitrescu, "Thermal expansion properties of organic crystals: a CSD study," Chemical Science **12**, 8537–8547 (2021).
[57] N. Marom, R. A. DiStasio Jr, V. Atalla, S. Levchenko, A. M. Reilly, J. R. Chelikowsky, L. Leiserowitz, and A. Tkatchenko, "Many-body dispersion interactions in molecular crystal polymorphism," Angewandte Chemie International Edition **52**, 6629–6632 (2013).
[58] G. Perlovich, L. K. Hansen, and A. Bauer-Brandl, "The polymorphism of glycine. thermochemical and structural aspects," Journal of thermal analysis and calorimetry **66**, 699–715 (2001).
[59] J. Hoja, A. M. Reilly, and A. Tkatchenko, "First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure," WIREs Computational Molecular Science **7** (2017), 10.1002/wcms.1294.
[60] A. Kadan, K. Ryczko, A. Wildman, R. Wang, A. Roitberg, and T. Yamazaki, "Accelerated organic crystal structure prediction with genetic algorithms and machine learning," Journal of Chemical Theory and Computation **19**, 9388–9402 (2023).