

ARTICLE OPEN



Machine learned force-fields for an Ab-initio quality description of metal-organic frameworks

Sandro Wieser¹ and Egbert Zojer¹✉

Metal-organic frameworks (MOFs) are an incredibly diverse group of highly porous hybrid materials, which are interesting for a wide range of possible applications. For a meaningful theoretical description of many of their properties accurate and computationally highly efficient methods are in high demand. These would avoid compromises regarding either the quality of modelling results or the level of complexity of the calculated properties. With the advent of machine learning approaches, it is now possible to generate such approaches with relatively little human effort. Here, we build on existing types of machine-learned force fields belonging to the moment-tensor and kernel-based potential families to develop a recipe for their efficient parametrization. This yields exceptionally accurate and computationally highly efficient force fields. The parametrization relies on reference configurations generated during molecular dynamics based, active learning runs. The performance of the potentials is benchmarked for a representative selection of commonly studied MOFs revealing a close to DFT accuracy in predicting forces and structural parameters for a set of validation structures. The same applies to elastic constants and phonon band structures. Additionally, for MOF-5 the thermal conductivity is obtained with full quantitative agreement to single-crystal experiments. All this is possible while maintaining a very high degree of computational efficiency. The exceptional accuracy of the parameterized force field potentials combined with their computational efficiency has the potential of lifting the computational modelling of MOFs to the next level.

npj Computational Materials (2024)10:18; <https://doi.org/10.1038/s41524-024-01205-w>

INTRODUCTION

Metal-organic frameworks (MOFs) are highly porous hybrid materials consisting of inorganic metal oxide nodes connected by organic linkers. Since their first realization, a wide range of potential applications has been discussed including catalysis^{1–3}, gas storage and separation^{4–6}, electronic devices^{7–9}, as well as drug encapsulation and delivery^{10,11}. In view of the seemingly limitless number of possible MOFs, finding the material with properties best suited for a specific application is a sizable challenge. The situation is further complicated by the frequently encountered difficulties when trying to measure the relevant materials parameters, e.g., due to often small crystallite sizes or the inclusion of foreign molecules in the MOF pores. Therefore, computational methods provide an excellent tool to not only predict properties but also to support the development of dependable structure-to-property-relationships. Due to the large number of atoms in MOF unit cells compared to conventional crystalline materials, ab-initio methods like density-functional theory (DFT) are frequently too computationally expensive for that task. This is particularly true when modelling properties at elevated temperatures or for properties that depend on the dynamics of the MOF constituents, like thermal expansion or thermal conductivity. Then it is necessary to resort to lower levels of theory, like classical force field potentials (FFPs), which can speed up calculations by many orders of magnitude^{12–16}, albeit at the price of a significantly reduced accuracy.

The most commonly applied and most easy to use force field potentials are transferable force-fields, which can be straightforwardly applied to modelling the majority of the elements in the periodic table. These include the Dreiding¹⁷ force field or the universal force field (UFF)¹⁸, where for the latter a variant exists

that is specifically adapted for the description of MOFs (UFF4MOF)^{19–21}. Especially UFF4MOF has been frequently used in literature for modelling MOFs^{22–25}, as it is readily available and particularly convenient for rapid structure prediction²⁰. Such highly transferable potentials are, however, not designed for an accurate description of dynamical properties²⁰ and usually result in sizable errors. In the case of MOFs, poor agreement between UFF4MOF-based simulations and DFT-based methods were explicitly demonstrated, for example, for the elastic properties of MIL-53²⁶ and for the thermal conductivity of MOF-5. In the latter case, UFF4MOF overestimated the single-crystal derived experimental data by a factor of 2.6²⁷. More advanced potentials focus on a more accurate description of a particular materials class, like GAFF²⁸ or COMPASS²⁹ for organic molecules or BTW-FF¹³ for MOFs. However, even these potentials with reduced transferability tend to have difficulties in accurately predicting vibrational properties³⁰, which are notoriously challenging to describe³¹, but which crucially influence many materials parameters.

A higher degree of accuracy (at the price of a further reduced transferability) is observed for potentials individually parameterized for specific molecular fragments. This, for example, applies to part of the MOF-FF family of potentials, which are parameterized for specific organic and inorganic MOF building blocks^{12,32}. Potentials like MOF-FF^{12,27,33} or QuickFF¹⁵ can also be parameterized fully system-specifically giving up on transferability to achieve an even higher degree of precision. The functional form of these force fields is designed relying on chemical intuition, for example, distinguishing between bonding and non-bonding interactions, which is not always ideal for MOFs, in which the dynamic adsorption and desorption of guest molecules can be important. This triggered the development of more flexible potentials like ReaxFF³⁴, but often at the price of an overall reduced accuracy

¹Institute of Solid State Physics, Graz University of Technology, NAWI Graz, Petersgasse 16, 8010 Graz, Austria. ✉email: egbert.zojer@tugraz.at

combined with problems regarding energy conservation and thermal stability³⁵. Another (major) disadvantage that we experienced when system-specifically parametrizing classical force fields like MOF-FF was that for achieving the required level of accuracy (compared to DFT and experimental data) a careful tailoring of the chosen set of potential terms, reference data, and fitting algorithms is often crucial^{27,33}. This makes the parametrization process extremely cumbersome.

With increasing computational power and advances in machine learning approaches, more convenient to use and often more accurate machine learned potentials (MLPs) emerged. In recent years, they started to see widespread use for successfully describing many materials^{36–39}. So far, for MOFs, MLP approaches have been largely limited to neural network potentials, which already show a promising performance in terms of precision^{26,40–46}. They were employed to reliably describe computationally expensive quantities like the thermal conductivity of several MOFs⁴⁴. However, many of the said investigations required a relatively large number of DFT reference data for training^{40,45} and neural network potentials are still computationally less efficient than traditional FFPs when tested on the same computer architecture^{40,47}. In this context it should be noted that some of the implementations of these potentials like GPUMD⁴⁸ and DeePMD-kit⁴⁹ feature parallelization on graphical processing units, making them also an attractive option for large simulations^{44,48}. Other types of machine learned potentials like Gaussian approximation potentials (GAPs)⁵⁰ or moment tensor potentials (MTPs)⁵¹ have seen widespread use for conventional materials^{52–54}, but their performance has, to the best of our knowledge, not yet been evaluated systematically for MOFs. Importantly, the sizable amount of DFT reference data required for machine learning based approaches can be cut down significantly by either only training molecular fragments of a system^{40,41,55} or by employing strategies to more efficiently sample phase space^{26,56}.

This raises the questions, how machine learned potentials can be most efficiently used for modelling MOFs, how they can be efficiently parametrized, how they perform for describing elastic and phonon-related properties, and to what extent their more complex form slows down the computations compared to traditional force field potentials. To answer these questions, we employed existing implementations of machine learned potentials. This makes the presented approach as widely applicable as possible, as these tools can be used essentially out-of-the-box with little human development effort required. In particular, we will focus on the kernel-based potentials (derivatives of GAPs) available within the wide-spread Vienna ab-initio Simulation Package (VASP)⁵⁷, which we will refer to as VASP MLPs, and on the MTPs implemented within the freely available MLIP⁵⁸ (machine learning interatomic potential) package. By efficiently combining both methods, we will present an efficient strategy for obtaining machine learned force field potentials, which allow the description of the structural and dynamic properties of MOFs at a level of quality that hitherto has been limited to quantum-mechanical simulations. In that spirit, the goal of the present work is not to develop a new type of machine-learned force field, but to efficiently apply the combination of two particularly promising approaches described in literature to the materials class of MOFs. In this way, an easy to implement roadmap is provided for how to lift the computational modelling of structural and dynamic MOF properties to the next level.

RESULTS

Overview of the computational approach

In this section, the fundamental aspects of the proposed approach for generating highly efficient machine-learned potentials for modelling MOFs are described. It also contains details necessary

for understanding the discussion of the obtained results. The more technical aspects of the theoretical approaches and details on the applied procedures that are not of immediate relevance for understanding the presented results are summarized in the Methods section, which will, thus, be referred to whenever appropriate. Further details are provided in the Supplementary Information.

Active learning approaches are a particularly promising strategy for sampling phase space to generate reference structures needed for learning force field potentials. However, there have only been few applications of such approaches to MOFs²⁶; moreover, especially in early implementations the training still demanded a rather high number of DFT calculations⁴⁵. A strategy that has the potential to severely cut down the required computational effort is available in conjunction with a kernel-based MLP in the VASP code. A detailed description of that approach is provided in ref. ^{56,59}. In the following, the key elements of the approach shall be summarized briefly to put the employed strategy into perspective: the active learning approach in VASP trains a kernel-based machine learned potential in the course of a molecular dynamics (MD) run. The potential is built from a basis set depending on local reference configurations (atoms with their surrounding neighbourhoods), which is dynamically expanded during the simulation^{56,57,60} and which defines the possible parameters for the potential. When the estimated Bayesian error of the forces in a specific MD step exceeds a given threshold, the concurrent force field is not considered accurate enough to continue the molecular dynamics simulation. In that case, the energies, forces, and stresses for the current time step are recalculated using DFT. These DFT calculations are also used to expand the set of reference data, from which the local reference configurations for the various atoms are selected, as described in the Methods section. The accuracy threshold is dynamically adjusted in the course of the molecular dynamics simulation based on the average of the 10 previous Bayesian errors estimates at time steps at which the force field was retrained⁵⁶. This leads to larger thresholds at higher temperatures, where the absolute values of the forces are higher. Such an adjustment is especially required when training with a temperature gradient, as a fixed threshold would add too few configurations at low temperatures and too many configurations at high temperatures. When enough new reference data have been accumulated, the basis set of the force field is updated with new local reference configurations, as explained in more detail in the Methods section. This approach has already been successfully employed to obtain accurate potentials for a range of materials^{56,57,60}. A possible drawback is that Gaussian approximation potentials have the reputation of being comparably slow⁴⁷, although, strictly speaking, this has been shown for other variants than the VASP MLPs used here. Moreover, advancements in the field of machine learned potentials occur at a rapid pace potentially also affecting their computational efficiency. This is exemplified by recent implementations of VASP (version 6.4.0 and upward), which promise substantial improvements of the computational efficiency of the used potentials. As described in the 'Evaluating the efficiency of the potentials' section, the latter is achieved by technical improvements of the implementation and a reduction of the computational overhead during production runs compared to the generation of training data.

Still, the aforementioned moment tensor potentials have been shown to lead to a particularly high accuracy-to-cost ratio in force-field based simulations (compared, for example, to GAPs)⁴⁷. Therefore, it appears useful to test, whether they could provide an even improved performance compared also to the latest VASP MLP variants. To ease the understanding of the later discussions, the key properties of MTPs are summarized in the following. Again, more information is provided in the methods section and a detailed account of the underlying theory of MTPs can be found in

the seminal paper by Shapeev et al.⁵¹. Unlike the VASP MLPs, MTPs describe the atomic energies based on a fixed basis set with a static number of parameters independent of the reference data used. The a priori choice of that basis set allows a more straightforward adjustment of their accuracy and speed^{51,58}. On more technical grounds, the basis sets of MTPs are built from moment tensor descriptors consisting of a radial and an angular part. The radial part is represented by a series of polynomials up to a given order, while the angular part is represented by a tensor up to a certain rank, which primarily determines the many-body character of the potential. Combinations of those moment tensors then define the basis functions, for which the parameters are trained. How many combinations of moment tensors are included in the potential, and, therefore, to what extent many-body interactions are considered, is defined by a 'level' parameter, which represents the primary way to adjust the speed and accuracy of an MTP. A second parameter to be chosen by the user is the radial basis set size, which defines the maximum order of the used polynomials. Increasing the MTP level and the basis set size leads to an increased accuracy of the MTP, but this comes at the price of a reduced computational efficiency. Unfortunately, for complex systems MTPs are not well compatible with active-learning strategies, like the one sketched above, as the required frequent reparameterization of the MTPs in MLIP becomes very time consuming. This is a consequence of the different mathematical strategies necessary for determining the parameters of the potentials (see Methods section). Further details can also be found in Supplementary Note 2.2.

Another relevant choice regarding the functional form of the MTPs and VASP MLPs is, how many chemically different species are considered in their parametrization. The minimum number of such atom types is given by the number of different chemical elements present in the material of interest. Larger numbers of atom types are obtained when also differentiating between identical elements in different chemical environments. This separation of atom types increases the number of parameters in the potential and allows to describe variations in chemical environments more accurately. However, the computational cost is adversely affected due to the larger required basis set. This increase in computational costs is significantly larger for the VASP MLP than for the MTP, which is related to their functional form, as discussed in the methods section.

As default approach for generating the data presented below, we generated separate parameters for atoms in different chemical environments, and, thus, employed a 'full' separation of atom types with further details provided in Supplementary Note 2. In selected cases, to better assess the trade-off between performance and efficiency, we also performed parametrization runs choosing only a single atom type per element.

Strategy: parametrization of the potentials

The parametrization of the MTPs and VASP MLPs requires DFT calculated total energies, forces on atoms, and stresses on unit cells for a sufficiently large number of reference configurations. The necessary reference structures were generated employing the on-the-fly machine-learning force field methodology implemented in VASP^{56,59}, whose conceptual approach has been outlined in the previous section. Details of the applied procedure like the chosen temperature intervals and simulation lengths were optimized in prior tests based on MOF-5⁶¹. In short, the 0 K optimized structures of the respective systems containing between 54 and 114 atoms in the unit cell were used as starting geometries. A molecular dynamics (MD) simulation was initiated in an NPT ensemble using a Parrinello-Rahman barostat⁶² at zero pressure in combination with a Langevin thermostat⁶³ starting at 50 K and heating the system up to 900 K over the course of 50,000 time steps of 0.5 fs. Despite the temperature of interest being

around room temperature, we chose a rather high maximum temperature of 900 K, because including larger atomic displacement amplitudes at higher temperatures in the reference data set is beneficial for machine learned potentials. This a consequence of such potentials performing well in an interpolation regime, while showing a poor ability to generalize to unknown situations⁵⁸.

To assess a potentially beneficial impact of increasing the size of the reference data sets, we also created 'extended reference data sets' by performing additional learning runs in an NPT ensemble at a constant temperature of 300 K for 100,000 time steps starting from the already obtained potential. Here, the Bayesian error threshold was fixed at a value of 0.02 eVÅ⁻¹ for all systems, which is slightly lower than the threshold value at 300 K during the training runs with continuous heating.

Besides generating reference structures, the above procedure (either including or excluding the 'extended reference data set') already produces first versions of the VASP MLPs. Nevertheless, the sampled reference data were used to retrain new VASP MLPs for the production runs. This was done to increase their level of accuracy with optimized settings for the selection of local reference configurations and to increase the computational efficiency by generating potentials excluding the on-the-fly error prediction. The exact details regarding the used settings during the retraining can be found in the Methods section.

In a next step, the DFT reference data generated in the VASP active learning runs were used to train also moment tensor potentials⁵¹, similar to the approach applied already in ref. ⁶⁴ to interfaces. The fitting of the MTP parameters was achieved by minimizing a cost function built from deviations in energies, forces, and stresses between DFT and MTP-predicted data and employing the MLIP package⁵⁸. Due to the stochastic nature of the initialization of the MTP fits, we performed several fits for each data set. The optimal potential for the respective system was then chosen based on its description of a validation set of reference structures. The associated technical details and numerical settings are again provided in the Methods section including a discussion of the computational cost of the fitting process. A further benefit of using the same reference data sets for the MTPs and VASP MLPs is that then one avoids any reference-bias, which allows a more rigorous comparison of the performance of the different machine learned potentials.

Strategy: benchmarking of the potentials

To demonstrate that the proposed strategy is truly useful, it is crucial to benchmark the obtained potentials for a sufficiently large range of materials and physical properties. Therefore, we tested the VASP MLP and the MTP approaches on several complementary MOF structures extensively described in literature and discussed in more detail in the next section. Here, it should be mentioned that the focus of the present paper will be on the properties of chemically stable and homogeneous materials. Therefore, at this stage, we do not aim at modelling, e.g., chemical reactions, or the diffusion of guest molecules or solvents. Thus, quantities of interest for the benchmarking comprise static and dynamic structural properties of MOFs and derived quantities. These include unit-cell parameters, energies, forces, stresses on strained cells, elastic constants, thermal expansion coefficients, and thermal conductivities. Of major relevance for several of those properties are phonons, where we will not limit ourselves to the Γ -point but will rather study phonons in the entire first Brillouin zone. This is important, as properties like the thermodynamic stability⁶⁵, as well as heat and charge-transport properties crucially depend on off- Γ phonons⁶⁶.

The benchmarking will be done primarily via a comparison to reference data generated by DFT and only secondarily via a comparison to experimental data (e.g., when a quantity of interest is not accessible to DFT due to its computational complexity). The

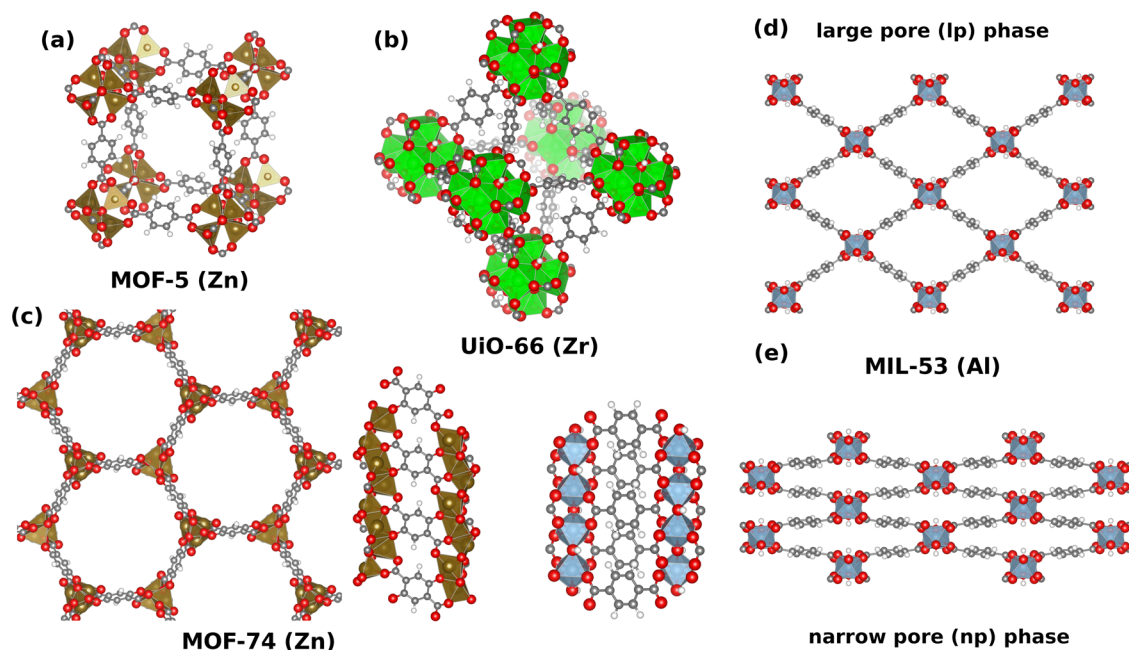


Fig. 1 Atomic structures of the investigated metal-organic frameworks. The following systems are considered: **a** MOF-5 (Zn), **b** UiO-66 (Zr), **c** MOF-74 (Zn), **d** large and **e** narrow pore phase of MIL-53 (Al). For the anisotropic systems MOF-74 and MIL-53 (np) side- and top-views are provided. The structures were visualized using the VESTA package¹¹⁹. Colour coding: Zn: brown, Zr: green, Al: blue, C: grey, H: white, O: red.

reason for that is that the goal of the present study is to assess the quality of the parametrization of the potentials and the suitability of their functional form. A comparison to experiments would additionally depend on the performance of the specific DFT methodology used in the simulations. A general DFT benchmarking for MOFs is, however, not in the focus of the present study. In fact, if alternative DFT-based approaches (relying, e.g., on different functionals or van der Waals corrections) were found to be more suitable for describing certain quantities, one could easily adopt them in the suggested parametrization process, as long as they were efficient enough for the active learning strategy. Other complications with experimental data are (i) that for many of the benchmarked quantities they do not exist and (ii) that at realistic experimental conditions MOF-based materials contain an often-unknown concentration of guest molecules and/or defects. Moreover, they typically consist of crystallites of limited size. This frequently makes the experimental determination of truly intrinsic materials properties of pristine MOF materials highly challenging.

To put the obtained agreement between the properties obtained with machine-learned potentials and with DFT into perspective, we also include a comparison to results obtained with UFF4MOF^{18–20}, as an example for an off-the-shelf transferrable force fields widely used for modelling MOFs. As UFF4MOF has not been specifically parametrized for the considered materials, larger errors are to be expected. Still this comparison reveals, what level of improvement in the results can be expected from machine-learned potentials obtained via the suggested approach. It will also answer the question, whether calculating the above-mentioned quantities with force fields like UFF4MOF is sensible at all. In order to not ignore system-specifically parametrized traditional force field potentials, we will also provide a comparison to properties calculated using a MOF-FF³² potential variant, which has previously been parametrized for MOF-5²⁷ against static DFT calculations similar to those performed here.

Materials of interest

The systems of interest for the present investigation share two features: first, they are widely studied due to their high thermal

stability and/or promising properties and, secondly, their unit cells are small enough that they are accessible to DFT simulations with properly converged numerical settings on contemporary supercomputers. Moreover, the chosen MOFs fundamentally differ in their topologies, shapes, dimensionalities, and flexibilities of the pores, and in the nature of the metal ions. Regarding the latter, we focus on closed-shell metal ions to avoid spin-order phenomena as an additional complication. The structures of the systems are shown in Fig. 1.

The first system is MOF-5⁶¹ consisting of Zn₄O nodes and 1,4-benzene-dicarboxylate (bdc) linkers. It forms a stable face-centred cubic structure with space group Fm-3m (225) containing 106 atoms in its primitive unit cell. Being one of the first published MOFs⁶¹, it has been intensively studied in the past, and is also included here as it is one of the very few MOFs for which reliable, single crystal thermal conductivity measurement are available⁶⁷. The second material is UiO-66⁶⁸. It consists of 12-coordinated ZrO based nodes connected by bdc linkers. Despite its different topology, the system still shows the same general symmetry as MOF-5 with space group Fm-3m (225). UiO-66 is one of the most commonly investigated MOFs⁶⁹ primarily due to its exceptionally high thermal stability⁷⁰. Another system of interest is MOF-74⁷¹, consisting of 1D extended metal-oxide pillar nodes connected by dobdc (2,5-dioxido-1,4-benzenedicarboxylate) linkers. They are aligned such that a honeycomb structure of hexagonally shaped 1D-extended pores is formed. This leads to a rhombohedral crystal structure with space group R-3 (148) with 54 atoms in the primitive unit cell. The system is known for its exceptional CO₂ uptake⁷². MOF-74 exists for various metals including Ca, Mg, Zn, Mn, Fe, Co, Cu and combinations thereof⁷³. Here, we picked the prominently investigated Zn variant. MOF-74 is particularly interesting due to its pronounced anisotropy. The fourth and final system is MIL-53⁷⁴, a MOF consisting of metal-oxide node pillars in an octahedral arrangement connected by bdc linkers in the perpendicular direction forming a ‘wine rack’ shaped structure. This MOF is known to occur in two primary phases: a low temperature ‘narrow’ pore phase and a high temperature ‘large’ pore phase⁷⁵. The large pore phase is orthorhombic with space group Imma (74), while the narrow pore phase is a monoclinic

Table 1. Total number of generated reference structures for the on-the-fly training of the VASP machine learned potentials.

system	$N_{\text{atoms}}^{\text{UC}}$	$N_{\text{references}}^{\text{initial}}$	$\text{error}_{\text{Bayesian}}^{\text{last}}$	$N_{\text{references}}^{\text{final}}$
MOF-5	106	974	0.055	1110
UiO-66	114	739	0.077	1373
MOF-74	54	998	0.089	2549
MIL-53 (lp)	76	783	0.080	2832
MIL-53 (np)	76	827	0.088	3073

The numbers are given for each system after the initial training run (heating of the system from 50 to 900 K over 25 ps), $N_{\text{references}}^{\text{initial}}$, and after the extension of the reference data set (training at constant temperature of 300 K over 50 ps with a fixed Bayesian error threshold of $0.02 \text{ eV}\text{\AA}^{-1}$), $N_{\text{references}}^{\text{final}}$; additionally, the numbers of atoms in the unit cell used for the training, $N_{\text{atoms}}^{\text{UC}}$, and the last Bayesian error threshold in the initial training, $\text{error}_{\text{Bayesian}}^{\text{last}}$ (given in $\text{eV}\text{\AA}^{-1}$), are shown.

system with space group Cc (9)⁷⁴. For the purpose of this work, we will consider both phases as separate systems, referring to them as MIL-53 (lp) and MIL-53 (np). MIL-53 has also been synthesized for a wide range of different metals. Here, we study the AI version of MIL-53, as one of the most commonly investigated variants⁷⁶.

Overview of the simulations

All quantities discussed in the following were calculated for all MOFs with a VASP MLP, with an MTP of level 22 (and a radial basis-set size of 10). The bulk of the presented data are based on force fields trained on the initial set of reference configurations. Only when including the extended reference data had a significant impact, this will be discussed explicitly. In the main manuscript, the focus will be on the results for MOF-5 and MOF-74 as prototypical examples for an isotropic and an anisotropic MOF with data on other MOFs usually mentioned only briefly. The entirety of the results for all MOFs and all reference data sets are provided in the Supplementary Notes 3.1, 3.3, and 3.5.

The reference data required for the training of the machine learned potentials were obtained using forces, stresses and energies calculated by density functional theory (DFT), as described above and in the Method section. This yielded between 739 and 998 reference configurations in the initial training, as summarized in Table 1 (together with the Bayesian error threshold at the end of the training). The latter is lower in MOF-5 compared to the other systems. This indicates a somewhat higher degree of accuracy when describing the forces for this system, which is confirmed when observing the evolution of the actual error during training (see Supplementary Note 2.1). The number of reference structures is in a similar range as for other machine-learned force field potentials for MOFs, including available accurate neural network potentials^{26,44} that either employ an incremental learning approach²⁶, or that are trained on structures derived from ab-initio MD runs⁴⁴.

Table 1 also contains the number of structures for the 'extended reference data set' discussed above. As can be seen in Table 1, for MOF-5, where the automatically set Bayesian threshold was already relatively low after the initial training, only 136 additional reference structures were added in the extended training run. In contrast, the number of structures in the datasets increased by a factor of two to three for the other systems, where the error threshold was higher initially. Further details on the training including the time evolution of the error threshold and the errors of the VASP MLPs compared to the training set are provided in the Supplementary Note 2.1.

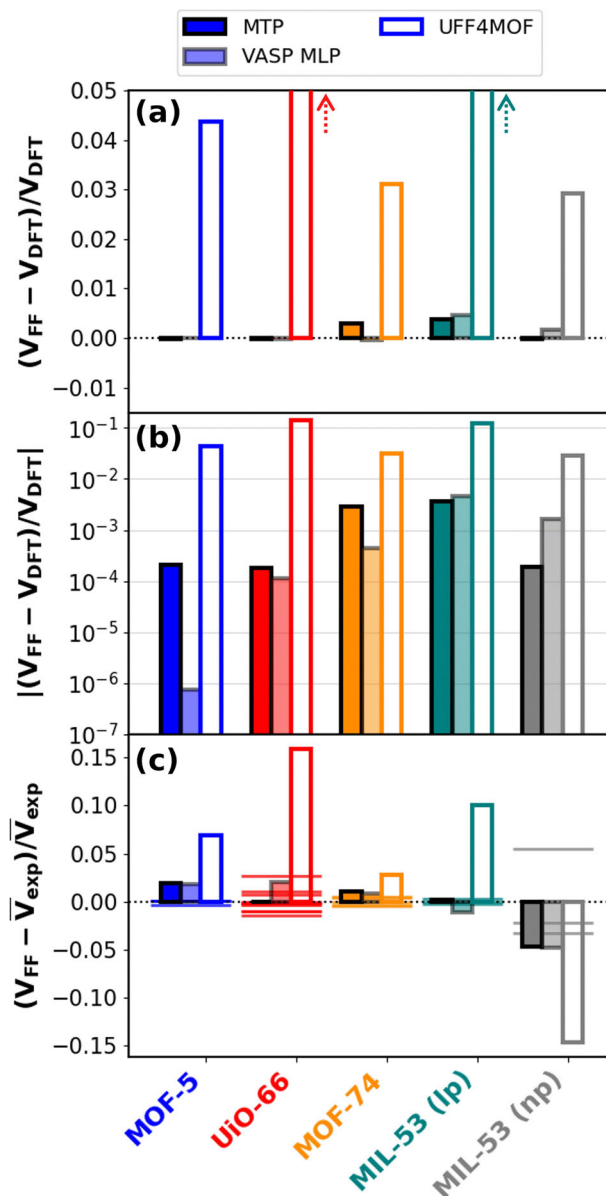


Fig. 2 Relative deviations of the volumes obtained with the force field potentials compared to DFT and experimental reference data. For the DFT comparison in (a, b) 0 K optimized unit cells are considered, while for the comparison with experimental data in (c) the values at room temperature are shown (see main text for details). For all systems (MOF-5: blue, UiO-66: red, MOF-74: yellow, MIL-53 (lp): teal, MIL-53 (np): grey) the comparison of the volumes is shown for machine learned potentials trained on the initial reference data set. MTP results are displayed by strongly coloured bars, VASP MLPs results by lightly coloured bars, and UFF4MOF results by empty bars. The simulated volumes in (c) were obtained via NPT simulations at 300 K. They are compared to the average of the experimentally obtained volumes. The individual experimental results for each system are indicated as vertical lines in the respective colour.

Benchmarking: crystallographic unit cells

As a first step in the benchmarking of the obtained machine-learned potentials, their ability to predict unit-cell parameters is assessed. We start with an analysis of the unit-cell volumes. Overall, the agreement of cell volumes between the VASP MLPs, the MTPs, and DFT is excellent. This is evidenced by the relative deviations between force-field and DFT data plotted in Fig. 2a, b

on a linear and on a logarithmic scale. For the isotropic systems MOF-5 and UiO-66 the relative deviation between both machine learned potentials and the DFT calculated volumes is below 0.2%. For MOF-5, the error of the VASP MLPs is reduced even further by approximately two orders of magnitude. This is in stark contrast to the situation for UFF4MOF, where deviations amount to up to 14%. The errors of the machine learned force fields are somewhat larger for the anisotropic MOFs, but they are still comparably low (around 0.3%). Here, the errors of the UFF4MOF volumes are at least one order of magnitude larger. A similar trend emerges, when considering the unit-cell vectors, with negligible deviations for the isotropic systems and somewhat larger errors for the c-parameters in MOF-74 and MIL-53 (see Supplementary Note 3.1.1 and Supplementary Table 12).

When comparing the simulations to experimental unit-cell volumes, one has to keep in mind that the majority of the experiments for MOF-5^{77–80}, UiO-66^{68,80–82}, MOF-74^{83–86}, MIL-53 (lp)^{75,87,88} and MIL-53 (np)^{74,75,88} were performed at or around room temperature (around 300 K). Thus, for this comparison we did not merely optimize the geometries using the respective force fields but performed NPT simulations at 300 K for the machine learned potentials and for the UFF4MOF force field (for details see method section). Thus, here the force fields not only need to be able to correctly predict '0 K' structures, but also need to capture the thermal expansion of the MOF crystals. As will be discussed in the section on Future Challenges, this is a particular challenge, as also for the underlying DFT approach modelling thermal expansion coefficients is a sizable challenge⁸³. Moreover, there usually is a rather significant spread in the experimental data. A comparison with the average experimental cell volumes is contained in Fig. 2c. The deviations of the cell volumes of individual measurements from the average values are indicated by the horizontal lines. A summary of the experimental data sets and the simulated 300 K lattice parameters can be found in Table S18. Again, for most systems the values obtained with the MTPs and VASP MLPs are close to the values averaged over comparable experiments (with deviations clearly below 2%). Only for MIL-53 (np) the situation appears to be somewhat worse with an underestimation of the average experimental volume by nearly 5%. In this context, one, however, must note that also experimental values vary by $\pm 4\%$ around the average. For UFF4MOF the agreement is much worse also for the comparison to experimental cell volumes, as can be inferred from Fig. 2c.

Benchmarking: prediction of energies, forces, and stresses

The probably most important benchmark for the performance of the machine learned potentials is how well they describe total energies, forces, and stresses. To ensure that the potentials properly predict situations beyond the reference structures they were trained on, it is important to analyse their accuracy based on an independent set of validation data. Here, the validation set consists of 100 DFT computed displaced structures obtained via an active learning molecular dynamics simulation in VASP at a temperature of 300 K. The deviations of the force-field predicted total energies, forces, and stresses from the DFT simulations are shown in Fig. 3 for MOF-5 and MOF-74. In neighbouring panels, we, on the one hand, compare MTP with VASP MLP results and, on the other hand, MTP with UFF4MOF data, which necessitates very different scales. Corresponding plots for UiO-66, MIL-53 (lp), and MIL-53 (np) portraying a similar situation are contained in Supplementary Fig. 15. The energies and forces for the validation data set were criteria for the choice of the optimal MTPs amongst all generated ones (see above). Thus, in addition to the root mean square deviations (RMSDs) for the chosen MTPs, also average RMSD values for all trained MTPs are reported in parentheses. The individual RMSD values for each generated potential can be found in Supplementary Table 7.

For the energies of the displaced structures relative to the equilibrium structures (first line of plots in Fig. 3), one observes an excellent performance of both machine learned potentials (panels a and c) where for the MTPs and VASP MLPs almost all deviations amount to $<1\%$ of the absolute values. Notably, for both machine learned potentials the largest differences in total energies for MOF-5 are around $0.3 \text{ meV}(\text{atom})^{-1}$, which is well below the chosen convergence criterion for the underlying DFT calculations, which amounts to $1 \text{ meV}(\text{atom})^{-1}$. In sharp contrast, for UFF4MOF the errors are more than two orders of magnitude larger with deviations of up to $50 \text{ meV}(\text{atom})^{-1}$ (see Fig. 3b).

For MOF-74, the MTP and VASP MLP errors are somewhat larger than for MOF-5, but also for this system they are two orders of magnitude below the UFF4MOF values. The trends for the maximum errors in energies are fully consistent with those of the corresponding RMSD values (see above).

For the forces in the second row of Fig. 3, the distributions of the errors are symmetric around zero. Consistent with the situation for the energies, the spread for the machine-learned potentials is similarly small for MTPs and VASP MLPs, while the transferrable force field at best displays a mediocre performance. For example, for MOF-5 the RMSD for the forces calculated with UFF4MOF amounts to $1.03 \text{ eV}\text{\AA}^{-1}$, while it is only $0.02 \text{ eV}\text{\AA}^{-1}$ for the MTP and the VASP MLP. In view of the particular relevance of force predictions, we considered also the Dreiding force field as an example for a traditional potential ignorant of MOFs in its parametrization. It performs even worse than UFF4MOF with an RMSD of $2.13 \text{ eV}\text{\AA}^{-1}$ for MOF-5. These errors are severe considering that the mean absolute value of all the forces in the validation set for MOF-5 amounts to only $0.59 \text{ eV}\text{\AA}^{-1}$. However, not all traditional force field potentials perform poorly. For example, employing a traditional, but system-specifically parameterized MOF-FF potential²⁷ for MOF-5 one obtains a substantially improved RMSD of $0.09 \text{ eV}\text{\AA}^{-1}$, which is an order of magnitude lower than the UFF4MOF value (albeit still larger than the values for the machine learned potentials). This (not entirely unexpectedly) suggests that the most important step in improving the performance of a force-field is its system-specific parametrization.

For the errors of the stresses in the last row of Fig. 3, we observe slightly larger values for the VASP MLPs than for the MTPs. This is possibly related to the details in the MTP training, where increased fitting weights for the stress contributions compared to the weights for the contributions of the forces were used. This was not done in the training of the VASP MLPs. For UFF4MOF, the error is again by two orders of magnitude larger. A peculiarity are the significantly larger stress errors for MOF-74 in Fig. 3k compared to MOF-5 in Fig. 3i for all force fields. This can at least in part be attributed to the absolute stress values being distinctly higher in MOF-74.

Notably, we observed a clear reduction of the force and stress errors of the VASP MLPs, when employing the extended reference data set in the parametrization, such that the initial RMSDs for MOF-74 of $0.27 \text{ eV}\text{\AA}^{-1}$ and 0.17 kbar decreased to $0.22 \text{ eV}\text{\AA}^{-1}$ and 0.12 kbar , respectively. Conversely, for the MTPs the use of the extended reference data sets has only a minor impact on the errors for forces, energies and stresses (see Table S11).

Concerning the other investigated systems (see Supplementary Fig. 15 and discussion in Supplementary Note 3.1), for the cubic UiO-66 MOF the situation for force, stress and energy errors is comparable to the MOF-5 case, while the anisotropic MIL-53 behaves more like MOF-74 with somewhat higher deviations for forces and stresses.

Benchmarking: vibrational properties of the studied MOFs

After benchmarking the quality of the properties directly provided by the machine learned potentials, it is useful to assess derived quantities. Here, we will focus on vibrational properties, which for

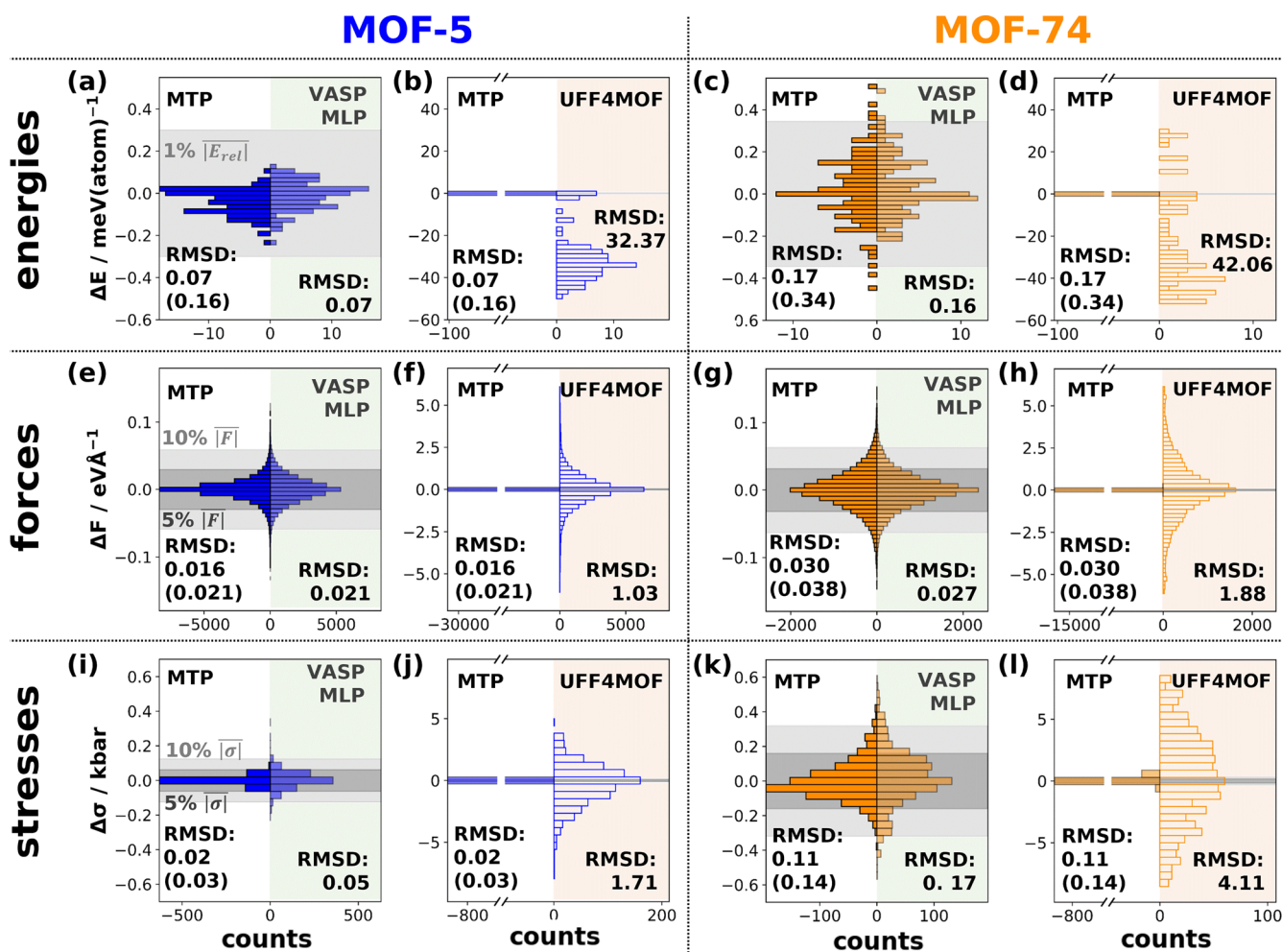


Fig. 3 Histograms detailing the deviations of the energies, forces, and stresses. Quantities obtained with the MTP (dark shading, all panels), VASP MLP (light shading; **a**, **c**, **e**, **g**, **i**, **k**) and UFF4MOF (no shading; **b**, **d**, **f**, **h**, **j**, **l**) have been subtracted from DFT reference data for the validation set. Deviations in energies relative to the situation at equilibrium geometry, ΔE , are shown in (**a–d**), differences in the components of the forces on individual atoms, ΔF , are shown in (**e–h**), and differences in the components of the stresses $\Delta\sigma$ are contained in (**i–l**). The areas shaded in grey indicate the ranges with errors within 5% and 10% of the absolute DFT values for forces and stresses and within 1% for the energy differences. The RMSD values of the respective quantities for various FFs compared to DFT are noted in the panels, where for the MTPs the RMSD values for the ideal force field and the average RMSD values for all parametrized force fields (in parentheses) are provided. The figure contains only the data for MOF-5 and MOF-74, with the data for all other studied MOFs contained in Supplementary Fig. 15.

crystalline materials are determined by phonons. They crucially impact phononic heat transport, charge transport (via the strong electron-phonon coupling), and elastic properties. As a starting point, Fig. 4a–d compares the low frequency phonon band structures obtained with DFT and the MTPs for MOF-5, UiO-66, MOF-74 and MIL-53 (lp). In Fig. 4e, the comparison between the VASP MLP and DFT for MOF-5 is shown. Figure 4 does not contain data for MIL-53 (np), as for this system it has been challenging to compute a phonon band structure without imaginary modes when using DFT (see Supplementary Note 1.4 for details).

The low frequency region in the focus of Fig. 4 is insofar particularly interesting, as it contains the phonon modes most relevant for the above-mentioned transport properties as well as for the relative thermal stability of different phases. It is immediately evident from the plots in Fig. 4 (and in Supplementary Fig. 24 for most VASP MLP cases) that both types of machine learned potentials reproduce the DFT-calculated phonon band structures almost perfectly. Especially for MOF-5 the largest deviations amount to $<5\text{ cm}^{-1}$, with this excellent agreement persisting for UiO-66 for the MTPs. For MOF-74 and MIL-53 (lp), we see some larger deviations for individual phonon bands, however,

especially for the MTPs the agreement is still extremely good and outperforms the results for any traditional force field potentials that we are aware of by a large margin. As shown in Supplementary Fig. 24, for the VASP MLPs, the agreement for MIL-53 (lp) and UiO-66 is slightly worse than for the MTPs, but can be significantly improved using the extended reference data set (see Supplementary Note 3.5).

The good agreement between force field and DFT-calculated phonon band structures is lost when employing UFF4MOF, which is illustrated for MOF-5 in Fig. 4f. The situation becomes even worse for MOF-74, as illustrated in Supplementary Fig. 26. UFF4MOF substantially overestimates the dispersion of the acoustic modes, which results in too high group velocities. This is one of the origins of the severely overestimated thermal conductivity for MOF-5 when calculating it with UFF4MOF²⁷. Another peculiarity of UFF4MOF is that it yields a severe overestimation of the frequencies of virtually all optical bands.

While the agreement of the phonon band structures is very reassuring, Fig. 4 displays only the situation for the low frequency phonons. As plotting the band structures over an extended frequency range would not be useful due to the sheer number of

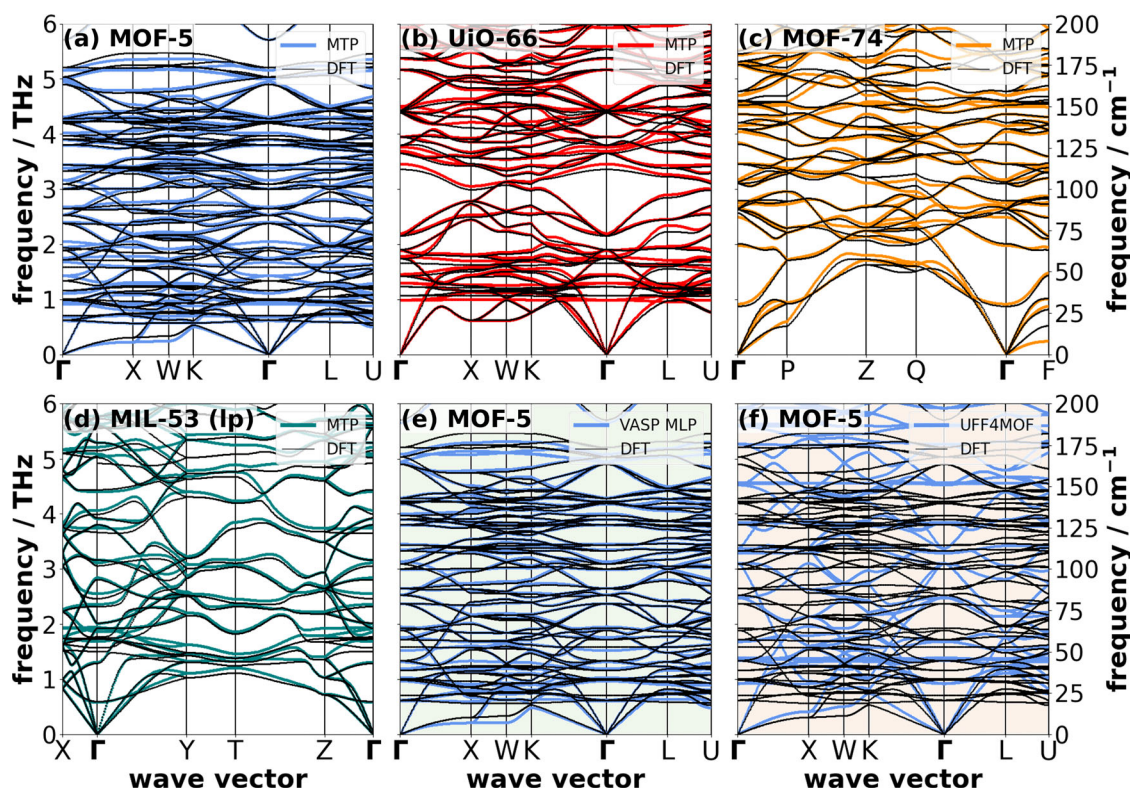


Fig. 4 Comparison of low frequency phonon band structures. They were computed using the ‘ideal’ MTPs (a–d) trained on the initial reference data set (coloured lines) and employing DFT (black lines) for the following systems: MOF-5 in **a**, UiO-66 in **b**, MOF-74 in **c** and MIL-53 (lp) in **d**. The comparison between the VASP MLP and DFT for MOF-5 is given in panel **e**. Panel **f** contains for MOF-5 the comparison between UFF4MOF and DFT. High symmetry point names correspond to the space group dependent conventions detailed in ref. ¹²⁰.

bands, in Fig. 5 the phonon densities of states (DOSs) calculated with the force fields are compared to those obtained with DFT for MOF-5 and MOF-74 (data for the other MOFs can be found in Supplementary Fig. 16). In the left panels, phonon DOSs in the same low frequency region covered also in Fig. 4 are displayed, while the right panels show the DOSs for the entire frequency range in which phonons exist in the studied materials. As expected from the preceding discussion, for both the MTPs and the VASP MLPs in the low frequency region one sees mostly minor deviations from the DFT data, which are somewhat more significant for MOF-74 than for MOF-5. Interestingly, this trend also prevails, when plotting DOSs for the full frequency range. The situation deteriorates for UFF4MOF, where it becomes virtually impossible to correlate specific features in the DOSs calculated with the force field with those calculated with DFT. Additionally, UFF4MOF massively underestimates the magnitude of the frequency gap found in the DFT simulations between 50 THz and 90 THz.

To quantify the quality of the description of phonons by the three types of force fields, Table 2 lists the RMSDs of the Γ -point frequencies for all materials in the low-frequency and in the full frequency ranges. Finally, for MOF-5, Table 2 also contains the results for the MOF-FF force field. Especially in the low frequency region, an excellent agreement with the DFT data is achieved for all machine learned potentials with only single digit wavenumber errors between 1.1 and 3.9 cm^{-1} . Even for the full frequency range and when using only the initial training set, MTP RMSD values for most systems are around an amazing 3 cm^{-1} with a maximum of 7.6 cm^{-1} for MIL-53 (np). For the VASP MLPs, deviations in the entire frequency range are somewhat larger ranging from 4.2 to 10.4 cm^{-1} . However, especially for the VASP MLP, the frequency error can be improved substantially when extending the reference data set, as shown in the Supplementary Note 3.1. The MOF-5

frequency error for the conventional and system-specific MOF-FF potential is substantially higher than for the machine-learned counterparts. However, MOF-FF still performs better by more than one order of magnitude compared to UFF4MOF, where the frequency RMSDs can reach several hundred wavenumbers over the entire frequency region.

Benchmarking: elastic properties of the MOFs

Another relevant quantity related to the properties of acoustic phonons (via the Christoffel equation⁸⁹) is the elastic stiffness tensor, C . The symmetry-inequivalent elements of C obtained from DFT and from the force field potentials are shown in Fig. 6 for MOF-5 and MOF-74. For the MTPs the agreement is close to ideal for both materials with maximum deviations of 1 GPa. For the VASP MLPs, the situation in MOF-5 is similar but in MOF-74 the agreement is slightly worse. Again, it can be improved by considering the extended reference data set. The lower degree of accuracy for MOF-74 is presumably caused by the more demanding description of the elastic constants in that highly anisotropic system⁹⁰. Additionally, as mentioned already in the context of the stresses on strained cells discussed above, the MTPs were trained with higher relative weights for the stress contributions compared to the forces, which could explain the improved level of accuracy in the MTP-calculated elastic constants.

Not unexpectedly, the performance of UFF4MOF is much worse. An extreme example is the C_{33} constant in MOF-74, which when using UFF4MOF is calculated to be >5 times as high as in the DFT reference (see Fig. 7f). Interestingly, the bulk modulus for MOF-5 calculated via the Voigt average from the components of the stiffness tensor is rather similar when calculated with UFF4MOF (16.8 GPa) and with DFT (16.1 GPa). These values are consistent with the literature¹⁴, but considering that the individual

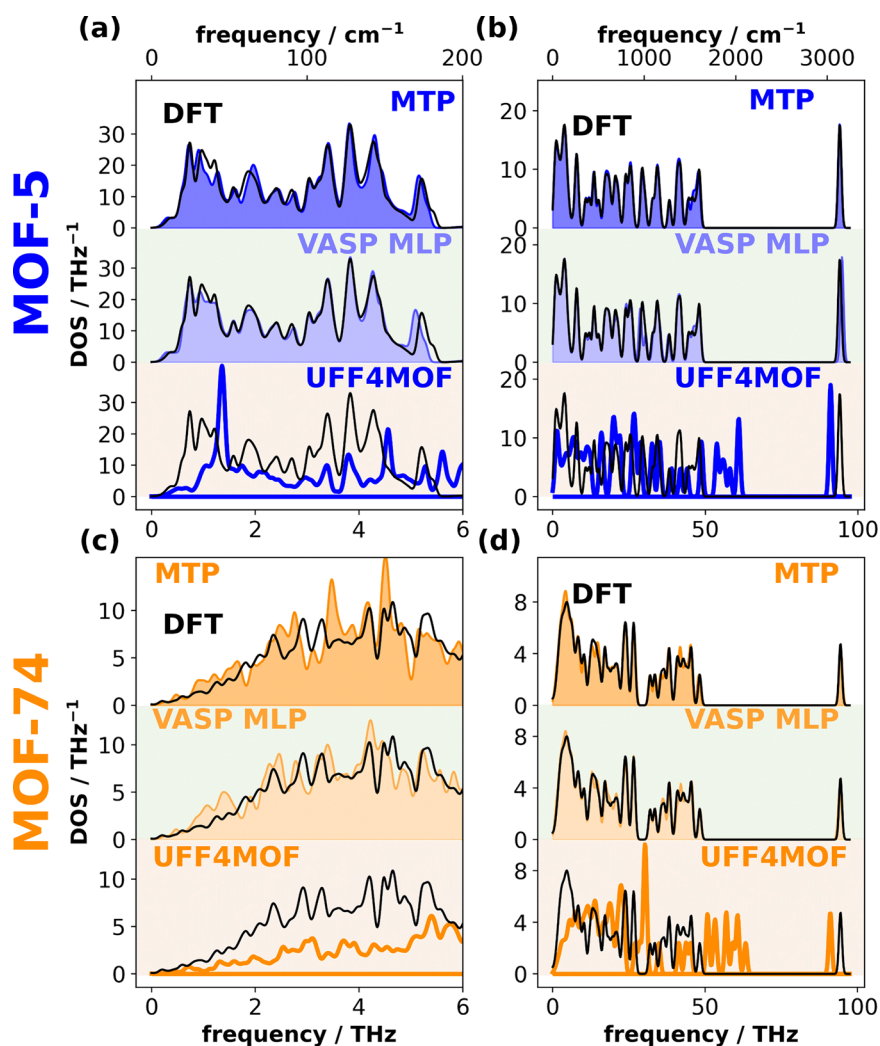


Fig. 5 Phonon densities of states (DOSs) calculated with DFT and various force fields. Comparisons of phonon DOSs are shown for DFT calculations (black lines) and the force field potentials (shaded areas/coloured lines) for the low frequency phonon modes in (a, c) and for the entire phonon spectrum in (b, d). The plots contain the comparisons for MOF-5 (a, b, blue) and for MOF-74 (c, d, orange) for the MTP, the VASP MLP, and the UFF4MOF. The DOSs were obtained for $11 \times 11 \times 11$ q-point meshes used to sample the first Brillouin zone and a Gaussian smearing is applied to broaden the modes. The width of that smearing is set to 0.05 THz for the low frequency region in (a, c) and to 0.5 THz for the entire frequency range in (b, d).

components of the stiffness tensor differ substantially between the two approaches, the seemingly good agreement for UFF4MOF for the bulk modulus has to be attributed to a fortuitous cancellation of errors.

The numerical values of the tensor components as well as plots for UiO-66 and MIL-53 are contained in Supplementary Note 3.1.4. The situation for the machine learned potentials is again comparable to that in MOF-5 and MOF-74 with deviations being largest for MIL-53 (np) with an RMSD for all tensor elements of 1.2 GPa. However, this is still quite very good, given the higher absolute values of the tensor elements in that material.

Benchmarking: thermal conductivity

As a final benchmark quantity, we consider the thermal conductivity. It is again intimately related to phonon properties, where now also anharmonic effects play a decisive role. As calculating the thermal conductivity of MOFs employing DFT is far beyond present computational possibilities, the comparison is restricted to experimental data and here to MOF-5, as to the best of our knowledge this is the only materials amongst the ones studied here for which high-quality single crystal data exist.

According to ref. ⁶⁷, the experimental room temperature value of the thermal conductivity of MOF-5 amounts to 0.32 W(mK)^{-1} and it has already been shown that the UFF4MOF and the Dreiding force fields fail in reproducing that value (yielding thermal conductivities of 0.847 W(mK)^{-1} and 1.102 W(mK)^{-1} , respectively), while the system-specifically parametrized MOF-FF variant provides a sensible value of 0.29 W(mK)^{-1} ²⁷. Here, non-equilibrium molecular dynamics simulations for MOF-5 were performed using a level 18 MTP. Notably, for these simulations we used a lower level of the MTP to reduce the computational cost of correcting for finite size effects, which requires simulations for various supercell lengths^{91,92}. This is, however, not expected to have a major impact, as the difference in the uncorrected thermal conductivities for selected cell lengths between level 22 and level 18 MTPs are minimal as shown in Supplementary Note 4.2. The resulting thermal conductivity obtained from NEMD simulations for cell lengths ranging from 208 to 416 Å perfectly reproduces the experiment, yielding a value of 0.32 W(mK)^{-1} . As a technical detail, it should be mentioned that this result has been obtained applying the traditional approach of determining the temperature gradient in the NEMD simulation in the region of a linear

Table 2. Root mean square deviations (RMSDs) for the phonon frequencies of the investigated systems.

System	Method	types	RMSD _{full} / cm ⁻¹	RMSD ₂₀₀ / cm ⁻¹
MOF-5	VASP MLP	7	9.0	1.5
	MTP	7	3.3	1.6
	UFF4MOF	5	203.2	51.0
	MOF-FF ³³	7	14.1	7.8
UiO-66	VASP MLP	9	10.4	2.7
	MTP	9	3.7	1.6
MOF-74	VASP MLP	8	4.2	1.1
	MTP	8	3.1	1.4
	UFF4MOF	4	242.8	116.9
MIL-53 (lp)	VASP MLP	8	8.8	3.2
	MTP	8	5.2	2.8
MIL-53 (np)	VASP MLP	8	6.9	2.5
	MTP	8	7.6	3.9

The values are given for Γ -point phonons obtained by the MTPs and VASP MLPs for the full frequency range, RMSD_{full}, and for the low frequency modes up to 200 cm⁻¹ (=6 THz), RMSD₂₀₀. Additionally, values for UFF4MOF^{19,20}, Dreiding¹⁷ and a system-specifically parameterized MOF-FF^{32,33} are provided for selected systems. The number of atom types used in each respective force field model is also given.

temperature decrease. When instead employing the method of Li et al., determining the temperature gradient from the temperature difference between the hot and cold thermostats⁹¹ (which we used previously^{27,33}), a somewhat lower value of 0.26 W(mK)⁻¹ is obtained. Therefore, we also employed the ‘approach to equilibrium molecular dynamics’, AEMD, method as a complementary molecular dynamics-based strategy for obtaining thermal conductivities. This again yielded a value of 0.32 W(mK)⁻¹, confirming the excellent agreement of the MTP simulations with the experimental data. A full discussion and further details regarding the heat transport simulations are provided in Supplementary Note 4.2. At the time of the calculations for this paper, NEMD or AEMD simulations were not accessible using VASP, hence no comparison between the VASP MLPs and MTPs can be provided for the thermal conductivity.

Future challenges

While the above discussion highlights the superior performance of the considered machine learned potentials for predicting various observables, we will next discuss thermal expansion as a process for which the performance of the machine learned potentials appears to be less satisfactory. In the case of thermal expansion, one again needs to compare the calculated thermal expansion coefficients to experimental data, as simulating the thermal expansion of MOFs with DFT is extremely challenging and prone to errors (see discussion in the Supporting Information of ref. ⁸³). As far as the available experimental data are concerned, some of the issues were already mentioned in the context of the crystallographic unit cells at 300 K for the narrow pore phase of MIL-53, where the experimental results also show a substantial spread. The situation is also problematic for MOF-74, where our own measurements yielded only very minor changes of the unit-cell volume with temperature^{83,93}. In view of these uncertainties, thermal expansion shall only be briefly discussed in the following, with significantly more details provided in Supplementary Note 4.1.

For example, for MOF-74, the MTPs/VASP MLPs predict a very small positive (1.6 · 10⁻⁶ K⁻¹ and 1.2 · 10⁻⁶ K⁻¹/2.0 · 10⁻⁶ K⁻¹ and

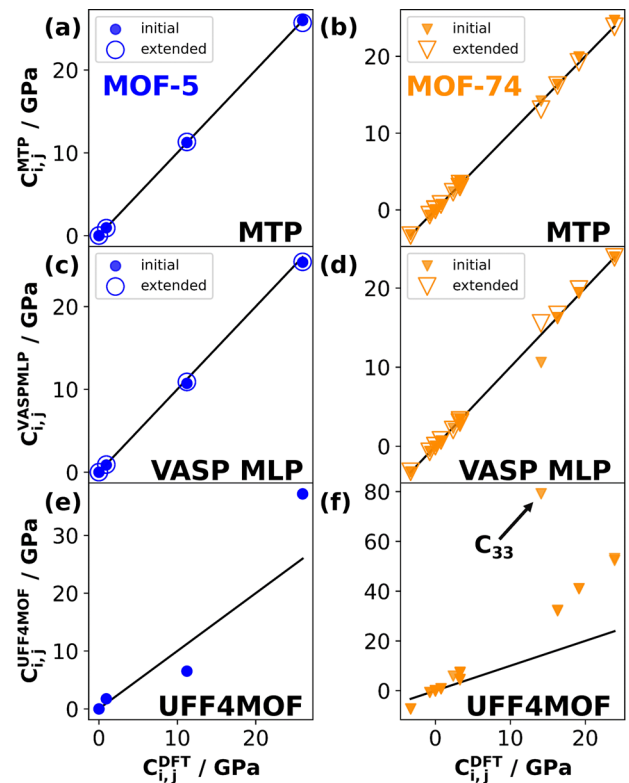


Fig. 6 Comparison of the elements of the elastic stiffness tensor. Shown are the tensor elements C_{ij} calculated with DFT and with the MTPs in (a, b), with DFT and with the VASP MLPs in (c, d), as well as with DFT and UFF4MOF in (e, f). The plots shown here contain data for MOF-5 in (a, c, d, blue) and MOF-74 in (b, d, f, orange) with equivalent plots for the other studied MOFs provided in Supplementary Fig. 17. In the upper and central panels, small filled symbols indicate values obtained with potentials trained on the initial reference data set and large empty symbols refer to the values for potentials trained on the extended reference data set. The solid black lines pass through the origin and have a slope of 1 such that they indicate an ideal agreement between force field and DFT simulations.

6.9 · 10⁻⁶ K⁻¹) thermal expansion coefficient in linker directions instead of a very small negative value (−2 · 10⁻⁶ K⁻¹) suggested by the experiments. Due to the generally limited accuracy of the experimental and computational methods this difference might still be acceptable. What is, however, more concerning is that along the pore direction, the MTPs/VASP MLPs predict a thermal expansion value of 31.0 · 10⁻⁶ K⁻¹/25.3 · 10⁻⁶ K⁻¹, which is nearly an order of magnitude larger than the experimental result of ~4 · 10⁻⁶ K⁻¹¹⁸³.

There are, however, also cases in which the agreement between simulations and experiments is better, like for MOF-5, where the simulated thermal expansion coefficients of −11.3 · 10⁻⁶ K⁻¹ for the MTP and −11.5 · 10⁻⁶ K⁻¹ for the VASP MLP are rather close to the experimental values, which range between −13.1 · 10⁻⁶ K⁻¹ and −15.3 · 10⁻⁶ K⁻¹^{178,79,94}. As already mentioned above, at this point it is unclear, whether the varying quality of the agreement between theory and experiments arises from the machine learned potentials, or whether the problems are at least in part inherent to the DFT approach used for the reference data generation. An indication that the chosen DFT methodology does have a significant impact is that for MOF-74 it has been observed that when applying the numerically more stable mode Grüneisen theory of thermal expansion, the sign of the thermal expansion coefficient in linker direction would actually depend on the chosen functional (PBE vs. PBEsol)⁸³.

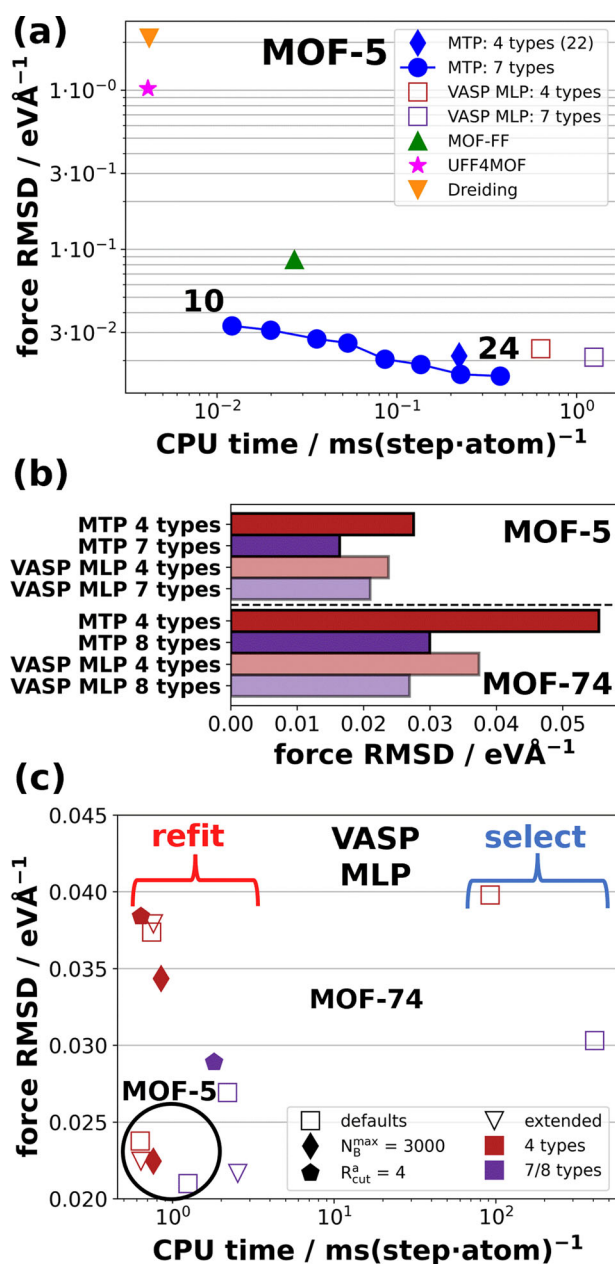


Fig. 7 Evaluation of the accuracy and speed of different variants of the tested force field potentials obtained for different numbers of atom types. In **a** the RMSDs of the forces for the validation set between DFT and various force-field approaches are shown as a function of the CPU time used per time step and atom for MOF-5. Data points are included for: MTPs at various levels trained with 7 atom types (blue squares), a level 22 MTP trained with only 4 atom types (blue diamond), two variants of the VASP MLP trained with 4 atom types (red square) and with 7 atom types (purple square), a system-specifically trained MOF-FF variant³³ (green triangle), and the transferable UFF4MOF¹⁹ (pink star) and Dreiding¹⁷ force fields (orange down triangle). The speed was obtained by performing an MD simulation in the NPT ensemble with further details provided in the main text. In **b**, the impact of including the separation of atom types on the validation set force RMSD is shown for the MTP and the VASP MLP for MOF-5 and MOF-74. **c** contains the values of force RMSDs and CPU times for different variants of the VASP MLPs for MOF-5 and MOF-74. The colour coding refers to the number of used atom types when training the potential (purple symbols: 7 atom types for MOF-5 and 8 for MOF-74; red symbols: 4 atom types). Inverted triangles indicate potentials trained on the extended reference data set, all other potentials were trained on the initial reference data set. Empty symbols indicate MLPs obtained retraining the potential using the default settings of VASP 6.4.1 and filled symbols represent MLPs obtained with adjusted settings. Symbols on the left side ('refit' bracket) indicate MLPs after a refit designed for production runs. Additional variation in the settings comprise: increasing the maximum number of local reference configurations from 1500 to 3000 for the 4 typed potential ($N_{\text{B}}^{\text{max}}$, diamond) and reducing the cutoff of the angular descriptors from 5 to 4 ($R_{\text{cut}}^{\text{a}}$, pentagons). In the lower left circle, results for MOF-5 are shown, while the rest of the panel only contains data for MOF-74. The CPU time used was tested for a 31104 atom sized supercell of MOF-74 with otherwise the same settings as for MOF-5.

steps even at room temperature. Notably, the situation turned out to be significantly more problematic for the MTPs than for the VASP MLPs, as discussed in detail for the MTPs in Supplementary Note 2.2. We attribute the less satisfactory performance of MTPs in the context of thermal stability to their poorer ability to generalize compared to GAP-type potentials, as already suggested previously⁵⁴. This has been attributed to the unregularized global basis functions of the MTP⁵⁴. The MTPs for MIL-53 were particularly problematic, as the material features a comparatively flat energy landscape. One possibility to mend this problem was to train several MTPs using the same reference data. In that case, due to variations originating from the stochastic nature of the initialization of the training procedure, some potentials turned out to be substantially more stable than others. However, this is not an ideal solution, as additional extensive MD test runs would be required for each MTP to judge the thermal stability. Moreover, training a large number of accurate MTPs is computational costly. Notably, also the extended reference data set obtained based on the molecular dynamics simulations at 300 K did not substantially improve the situation. However, adding 1009 DFT reference structures based on active learning runs at 400 K at a reduced error threshold did result in MTPs for MIL-53 (lp) that were all stable at least up to room temperature.

These results suggest that at least under certain circumstances adding more strongly displaced reference structures obtained by higher-temperature learning steps helps overcoming the problem of long-term stability of the force field. Unfortunately, this is associated with a significantly increased learning effort, where additional reference data beyond the first 500–1000 structures do not substantially improve the accuracy of the MTPs. Additionally, for MTPs it could be beneficial, to further improve the data sampling approach or to augment the force fields with potential

The other major challenge when applying the MTP approach to MOFs is to train the potential such that a force field is obtained that is fully stable for the targeted simulation conditions. Stable here means that the system in question remains structurally intact also over hundreds of thousands (or even millions) of molecular dynamics time steps that one would, for example, apply when performing NEMD simulations. A structural disintegration of the material is not possible for simple force field potentials characterized, e.g., by harmonic bonding potentials, but for approaches based on mathematically more complex models, like machine learned potentials, this can become a problem. This is especially true when in the stochastic description of the atomic motion at finite temperatures large displacements occur for which the potential has not been sufficiently well trained.

Improving the long-term stability of the force fields was one of the main motivations for increasing the final temperature in the on-the-fly training runs to 900 K. Nevertheless, not all trained force fields turned out to be stable over hundreds of thousands of time

terms that prevent a structural disintegration of the studied system.

In general, the stability issues are much less pronounced for the VASP MLPs, which use a basis set more suitable for generalization to unknown situations. Based on our stability tests in an NPT ensemble, all VASP MLPs trained on the initial reference data set including atom type separation were stable up to 700 K, which is the maximum temperature included in the tests (see also Supplementary Note 4.1.2). Still, at very high temperatures some most likely unphysical cell deformations took place for some of the systems, but at least the structure never disintegrated.

Evaluating the efficiency of the potentials

So far, the discussion focused on the precision of the machine learned potentials, demonstrating that they show an excellent performance when compared to DFT data. What has not been touched upon so far is an evaluation of the computational efficiency of the potentials. Here, the MTPs allow a straightforward adjustment of the number of parameters of the potentials and, thus, of their numerical efficiency. The primary hyperparameter to tune in this context is the so-called 'level' of the MTP, which controls the degree to which many-body interactions are included and, thus, substantially impacts the number of parameters. Consequently, it has a much more profound impact on both precision and performance than the number of radial basis functions, which, thus, has been set to 10 for all presented data. The other setting of the MTPs that substantially affects the computational efficiency is the cutoff for the interactions. However, if one deviates too much from the default value of 5 Å (chosen also here), the precision of the potentials deteriorates severely. Therefore, using the same reference data as above we systematically varied the level of the MTPs from 10 to 24 and compared the accuracy (in terms of force RMSDs) and the speed of the force fields in Fig. 7. In panel a, the computational speed of the machine learned potentials is correlated with their accuracy and compared to more traditionally force fields for MOF-5. For this comparison we use three different types of conventional potentials: a system-specifically parameterized MOF-FF potential for MOF-5 used in our previous work^{32,33}, the transferable universal force field¹⁸ extended for MOFs (UFF4MOF)^{19,20}, and the transferable Dreiding¹⁷ potential. For the VASP MLPs, panel (a) contains only the data for the force field refitted on the initial reference data set, as described in the computational approach section (using VASP 6.4.1). The computational speed was evaluated for NPT simulations on $4 \times 4 \times 4$ conventional supercells of MOF-5 carried out on 64 cores of a dual socket AMD EPYC 7713 (Milan) node of the supercomputer VSC-5 (see <https://vsc.ac.at/systems/vsc-5/>). It should be mentioned that differences in the speeds of the force fields are caused by differences in scaling with the number of nodes (see Supplementary Note 4.1.4) and can be affected by using different libraries when compiling the VASP and the LAMMPS codes. Additionally, we stress that the tests performed here are far from comprehensive and are restricted to the force fields considered in this study. There certainly are other options, like, for example, the GPUMD⁴⁸ force fields, for which a very high speed at good accuracy has been reported for MOFs⁴⁴.

The traditional UFF4MOF and Dreiding potentials are the most computationally efficient options, but they also suffer from a poor description of the forces with RMSD values above 1 eV\AA^{-1} . As discussed already above, the system-specifically parameterized MOF-FF potential hugely improves the situation with an RMSD value of 0.09 eV\AA^{-1} . However, the more complex functional form including cross terms and an optional, more accurate treatment of long-range electrostatic interactions via a particle-particle particle-mesh solver⁹⁵ slows down MOF-FF by a factor of ca. 5 compared to the more simple force fields.

Significantly more accurate are MTPs, where the highest levels show the best degree of accuracy among the tested potentials for MOF-5 with a force RMSD of 0.016 eV\AA^{-1} . Interestingly, reducing the level of the MTP from 24 to 10 yields a speedup by a factor of 28, but at the same time less than doubles the force RMSD (to 0.033 eV\AA^{-1}) such that the level 10 MTP still clearly outperforms MOF-FF. In fact, that MTP is computationally only about twice as expensive as UFF4MOF, but with a force RMSD reduced by a factor of 34. For the level 22 MTP, we also tested the performance without separating the atom types: as can be clearly seen in Fig. 7a, this has hardly any impact on the computational efficiency, but decreases the degree of accuracy (with the force RMSD increasing from 0.016 eV\AA^{-1} to 0.021 eV\AA^{-1} for 7 and 4 atom types, respectively).

The VASP MLPs are of a similar accuracy as the mid- to high-level MTPs, but they are at least a factor of 2–3 slower than the highest level MTPs. At least in part this can be attributed to the rather unfavourable scaling of the parallelization of the molecular dynamics part of VASP, as shown in Supplementary Note 4.1.4. For example, when using only up to 16 cores, the VASP MLP with 4 atom types is only about 50% slower than the level 22 MTP. We expect this scaling to be improved in future implementations.

Figure 7a also shows that the VASP MLP with 7 atom types is about a factor of 2 slower than the one with 4 types while only offering a moderately increased degree of accuracy. Conversely, the gain when considering more atom types in the VASP MLP is much more significant for MOF-74, as shown in Fig. 7b. Thus, in the case of the VASP-generated potential, whether or not it is useful to consider more atom types is system dependent. As indicated already above, the situation for the MTPs is quite different: there, the benefit of considering more atom types is generally larger, while the additional computational cost is negligible. Thus, for MTPs differentiating between atoms in chemically different environments is definitely advisable for the considered systems. We tentatively attribute this difference between the MTPs and VASP MLPs to the fact that a larger number of atom types significantly increases the number of adjustable parameters in MTPs, while in VASP MLPs the size of the basis set is adjusted dynamically, which at least in part anticipates the benefit of more atom types for the accuracy.

In the course of our work, we learned that the speed and accuracy of the VASP MLPs retrained employing the VASP 6.4.1 default settings clearly outperformed VASP MLPs trained with the default settings of earlier versions of the code both in terms of speed and especially accuracy. The latter resulted primarily from including a larger number of local reference configurations in a reselection process (datapoints below the 'select' bracket, which serve as the reference data for the following comparison). A more in-depth discussion of the dependence of the performance of the VASP MLPs on the version of the code can be found in Supplementary Note 3.7. Another factor hugely impacting the speed of the VASP MLP is whether or not the live error estimation in the VASP MLP is performed. As shown in Fig. 7c, when this is not the case (like for the potentials used to generate the symbols under the 'refit' bracket), the computational cost is reduced by about two orders of magnitude. For the 'refit' calculations also the accuracy of the VASP MLP is slightly improved due to the use of a singular value decomposition approach⁹⁶, when refitting the parameters (in contrast to the less accurate but more efficient LU factorisation during training). When additionally increasing the maximum number of reference configurations retained in the VASP MLP (red diamond), a higher level of accuracy is achieved, albeit at the cost of a reduced computational speed. To again improve the speed of the potential, reducing the radial cutoff of the angular descriptors appears to be a viable option. However, in the performed tests, (see pentagons) the difference compared to choosing the default value is only minor. The most accurate VASP MLP for MOF-74 is achieved combining a large number of atom

types with an extended reference data set (see purple down-triangle). Overall, for the tests under the 'refit' bracket one observes that, whenever one increases the accuracy of the force prediction by accordingly setting one of the rather diverse parameters explored for the VASP MLPs, this also decreases the overall computational speed.

DISCUSSION

In summary, in this work we describe a strategy for obtaining machine learned potentials that allow the description of the properties of MOFs at essentially DFT quality with hugely reduced computational costs. This strategy uses the active learning approach of VASP^{56,97} for sampling the configuration space and for generating reference structures. On these reference data, machine learned potentials can already be trained in VASP, which are highly accurate and efficient (especially when using the latest releases of the code). The resulting force fields are nearly two orders of magnitude more accurate, but also clearly slower than traditional potentials like UFF4MOF. If speed is the priority, the effective reference data generation procedure of VASP can be combined with the numerical efficiency and also high accuracy of moment tensor potentials (MTPs)^{51,58}. They allow flexible adjustments to their speed and accuracy by manually adjusting the size of the basis set leading to potentials, where the faster variants are only less than an order of magnitude slower than traditional transferable potentials while still maintaining an exceptionally high degree of accuracy. As all necessary codes (the VASP and MLIP software packages) are readily available, a general use of the proposed strategy should be comparably straightforward.

A systematic comparison of predicted forces, energies, stresses, phonon band-structures, and thermal transport coefficients shows that the such-trained MTPs and VASP MLPs yield a truly amazing agreement with DFT reference data (as the primary benchmark quantities) and with experimental data in cases where DFT data are not available. The said tests were performed for a representative selection of isotropic as well as anisotropic MOFs including MOF-5, MOF-74, UiO-66, and the narrow and large pore phases of MIL-53. Compared to transferrable force fields like UFF4MOF the improvements in prediction quality is enormous with errors for certain quantities dropping by several orders of magnitude and this at only very moderately increased computational costs. Only when attempting to describe thermal expansion processes, distinct deviations between the machine learned potentials and experiments are observed, where there are some indications that in that case the reference DFT methodology might be at least partly at fault. Notably, in the case of MTPs for some systems also the thermal stability of the force fields needs to be further improved. This can become an issue when performing extended molecular dynamics runs comprising hundreds of thousands of time steps, but for the systems considered here the problem can be mediated by including reference data from dedicated high temperature runs. The VASP generated potentials suffer substantially less from issues of thermal stability. This makes them a highly promising option especially for flexible materials with flat potential energy surfaces.

Overall, the data presented here and other recent advancements in MOF force field development^{26,44} clearly point towards machine-learned potentials as becoming a game-changer for the modelling of this complex class of materials. This particularly applies to the simulation of dynamical MOF properties, which are often computationally not accessible to ab-initio methods like DFT but for whose description traditional force fields are either too inaccurate or too tedious to parametrize. Here, the use of VASP MLPs or MTPs can speed up the simulations by many orders of magnitude keeping essentially the accuracy level of the reference method used in their parametrization.

METHODS

In this section, we will discuss the most important settings and technical details relevant for this work. A more detailed comparison of the nature of MTPs and VASP MLPs is also provided. Additional technical information and convergence tests are contained in the Supplementary Information.

The DFT approach

For the required reference ab-initio calculations, density functional theory (DFT) simulations were performed with the Vienna ab-initio Software Package (VASP)^{98–103} employing the Perdew-Burke-Ernzerhof (PBE) functional^{104,105} and applying Grimme's D3 correction with Becke-Johnson damping^{106,107} to account for dispersive forces. For all training runs a plane-wave energy cutoff of 900 eV was chosen after careful convergence tests based on total energies, vibrational frequencies and elastic moduli. The energy convergence for the self-consistent field approach was set to 10^{-8} eV. The structures for each system were optimized until the maximum absolute force component in the system reached at least $<10^{-3}$ eVÅ⁻¹ using a quasi-Newton optimization algorithm. Further details on the VASP settings including the system-specific chosen k-space sampling mesh are provided in Supplementary Note 1.

Functional form of the machine learned potentials

This section introduces aspects regarding the functional form of the potentials that are relevant for understanding the fundamental differences between the kernel-based VASP MLPs and the moment tensor potentials. It also summarizes the chosen settings used throughout this work. A comprehensive description of the VASP MLPs and of the MTPs beyond specific aspects relevant to assess the results presented in this work can be found in refs. ^{51,56}.

The energies for each atom E_i in the VASP MLPs⁵⁶ are computed from kernels K which represent a measure for the similarity between local reference configurations, i_B , and the local configuration of interest, i .

$$E_i = \sum_{i_B=1}^{N_B} w_{i_B} K(\mathbf{X}_i, \mathbf{X}_{i_B}) \quad (1)$$

The kernel is built from the expansion coefficients of the radial and angular descriptors containing information regarding the geometry of the local configurations. The descriptors are defined up to a certain cut-off, which was set to 5 Å during active learning. For the reselection of the reference configurations and subsequent refits, the cut-off for the radial descriptors was set to 8 Å. The weights, w_{i_B} , represent the parameter that needs to be trained for each local reference configuration i_B . This is done not only for the energies but also for the forces and stresses using the derivatives of the kernel with respect to the atom positions and cell parameters. Notably, the number of local reference configurations, i_B , increases with the number of reference structures contained in the training set (up to a maximum value).

Conversely, the basis set of the MTPs⁵¹ is built from moment tensor descriptors, M , describing the neighbourhood, n , around an atom i :

$$M_{\mu,\nu}(n_i) = \sum_j f_{\mu}(|\mathbf{r}_{ij}|, z_i, z_j) \underbrace{\mathbf{r}_{ij} \otimes \dots \otimes \mathbf{r}_{ij}}_{\nu \text{ times}} \quad (2)$$

These descriptors are constructed from a radial part f_{μ} and an angular part represented by a series of ν outer products of the distance vector \mathbf{r}_{ij} . The radial part consists of a sum of radial basis functions defined up to a distance cut-off for each pair of atom types z_i and z_j , where the primary component is a polynomial, Q^{β} , up to a maximum order N_Q . This number is usually referred to as the number of radial basis functions. This means that N_Q

parameters c_{μ,z_i,z_j}^β are associated with the radial part of the descriptor:

$$f_\mu(|\mathbf{r}_{ij}|, z_i, z_j) = \sum_{\beta=1}^{N_Q} c_{\mu,z_i,z_j}^\beta Q^\beta(|\mathbf{r}_{ij}|). \quad (3)$$

The indices μ and ν define the ‘level’⁵⁸ of the MTP with level($M_{\mu,\nu}$) = 2 + 4 μ + ν . The level specifies, which combinations of moment tensor descriptors are included to form the basis functions $B_{\alpha\nu}$, which then lead to the energy of an atom around a neighbourhood using

$$E_i = \sum_{\alpha} \xi_{\alpha} B_{\alpha}(n_i). \quad (4)$$

Thus, the parameters that need to be trained are the radial parameters c_{μ,z_i,z_j}^β for each atom pair and value of μ and the parameters ξ_{α} for each basis function. The primary difference compared to the VASP MLPs is that here the number of parameters does not depend on the amount of training data and can be adjusted freely based on the needs of the user by setting the values for the level and N_Q . Since especially the angular component of the moment tensor descriptors can be reused for many different basis functions, a high level of computational efficiency is maintained.

On-the-fly training of the VASP MLPs

As indicated in the previous section, the energy, force and stress weights for each of the basis functions in a VASP MLP need to be trained on the reference data. For this, the reference data are generated on-the-fly in an active learning approach, which is based on molecular dynamics simulations. These were carried out using a time step of 0.5 fs. The friction coefficient for the Langevin thermostats and barostats was chosen to be 10 ps⁻¹ and a fictitious mass of 1000 amu was assigned to it for all systems, as required for the Langevin barostat method of Parrinello and Rahman¹⁰⁸. In the initial training runs, the temperature was set to gradually increase from 50 K to 900 K over 50,000 time steps. To generate the independent test sets to validate energies, atomic forces and stress tensors of the individual force field potentials, another set of completely independent on-the-fly machine learning force field runs was performed from scratch in an NPT ensemble at 300 K with a time step of 0.5 fs. These runs continued until 100 DFT calculations had been performed, which then served as system-specific test sets. In the context of generating the test set, the machine learning approach was used primarily to improve the sampling of phase space compared to, e.g., picking every n^{th} configuration from a conventional ab-initio MD run.

The complexity of the selection and training procedure evidently leads to a number of parameters, whose choice leaves a large room for optimizations. This is also reflected by advancement of the VASP code, where version 6.4.1 led to significant improvements over version 6.3.0. Since the training runs for this work were carried out with version 6.3.0, the local reference configurations were reselected with version 6.4.1 with the ML_MODE=select setting. Therefore, when discussing actual numerical values for the parameters in the following, we give them for the training run and for the situation after the reselection of local configurations in all cases in which they differ (see also the Supplementary Note 3.7 and ref. ⁵⁶ for more details).

All training runs were performed using only 4 different atom types (corresponding to the chemical elements) for each system. Subsequently, for the reselection of the reference configurations the number of atom types was increased to differentiate between chemically different environments. Another important aspect of the active learning approach is to decide when a first principles calculation needs to be performed. This is done when the highest

Bayesian errors on the forces exceeds a certain threshold. This threshold starts at 0.002 eVÅ⁻¹ and is allowed to adjust dynamically based on the average Bayesian error from previous 10 steps in which the potential was retrained. The Bayesian error on the forces of individual atoms additionally decides, whether the neighbourhood around that atom is added to the local reference configurations. The total number of local reference configurations can be controlled using a multiplicative factor (set by ML_SCLC_CTIFOR) acting on the Bayesian error threshold, which only affects the selection process but not the decision when first principle calculations are performed. During active learning, this factor was set to 1.0, while for the reselection of local reference configurations for the final VASP MTP it was reduced to 0.6, so that more local reference configurations were picked (see Supplementary Note 3.7 for specific details). As a further sparsification step, a CUR algorithm¹⁰⁹ selects configurations to be discarded when the eigenvalues of the kernel matrix fall below a certain threshold, as then the similarity between individual reference configurations would be too large. During the training runs, the CUR sparsification threshold parameter was set to 10⁻¹⁰ and for the reselection of the local configurations and the subsequent refitting procedure, it was set to 10⁻¹¹ to obtain fast potentials. The selection and sparsification procedures also allow setting the maximum number of local reference configurations to limit the computational cost. In VASP, this can be done with the ML_MB tag for each atom type and was set to 3000 for each atomic species during active learning. For the reselected VASP MLPs it was set to 1500 per atomic species to speed up the potential. For Fig. 7, tests were also performed using the refined potentials with a maximum of 3000 configurations per atom types. To speed up the training during active learning, the potential is retrained only after 5 first principles steps had been added to the reference data or when the Bayesian error exceeded the threshold by a factor of 2 (4 in the case of the reselection).

The weights of the VASP MLPs are trained by solving a linear system of equations containing the weights, the kernels, and dimensionless energies, forces and stresses for the structures in the reference data set. For this, Bayesian regression is used, where the likelihood, prior and posterior distributions are described via multivariate Gaussian distributions⁵⁶, with the optimum weights located at the centre of the distributions. These optimum weights can then be obtained by solving a new system of equations. During on-the-fly training this is performed via an efficient LU factorization procedure¹¹⁰. In postprocessing, a more expensive and more accurate approach leading to more accurate results of the weights using a singular value decomposition (SVD) method with Tikhonov regularisation⁹⁶ is used (set with ML_MODE=refit), where the regularisation precision parameter ML_SIGW0 is set to 10⁻⁷. Additionally, a speedup of the potential is achieved by disabling the error prediction during the production runs via ML_MODE=refit. In the Supplementary Note 3.7 more details regarding the different regression methods and additional comparisons regarding the accuracy are provided.

MTP training

To train the MTPs, the reference data generated during the VASP on-the-fly learning runs were used. For determining the MTP parameters, it is necessary to minimize a cost function built from the energies, forces and stresses in the training set⁵⁸. The minimization was performed using a BFGS (Broyden-Fletcher-Goldfarb-Shanno)¹¹¹ algorithm. The initial parameters were chosen randomly. Then an initial training was performed for 75 iterations. Subsequently the main minimization procedure was started and performed until the differences of the cost function in the previous 50 iterations reached <2 · 10⁻³. Said cost function is built from a weighted sum over the square deviations of forces, energies and stresses. The weights in the cost function for stresses

and energies were chosen to be 1.0 and for forces to be 0.01. These are the default values for forces and energies. For stresses, the weight was increased compared to the default value of 0.001, as we found that it improves the accuracy of the stresses without substantially impacting the precision of the forces or energies.

For most of the training runs, a few thousand iterations were required to reach the convergence criterion (which are 2–3 orders of magnitude more iterations than regression steps done for the training of the VASP MLPs). However, it should be noted that complications occurred for some of the resulting MTPs, where the minimization algorithm managed to escape the lowest minimum found during the process to ultimately reach a local minimum with a larger cost function. This can be remedied by either choosing looser convergence criteria or by choosing the intermediate solution with the lowest cost function. The latter approach was employed in this work for the cases where this issue was observed and no further minimum was found within 10,000 optimization steps. For additional details and convergence curves for the MTP training, see the Supplementary Note 2.2.

Since we wish to show how well the MTPs reproduce the reference data, we use a relatively large basis set as a baseline with a level of 22 and a radial basis set size of 10. However, tests were also performed varying the level of the MTP between 10 and 24 for each system (see Supplementary Note 3.6 for detailed results) to show the impact of modifying the number of parameters contained in the MTPs. The MTPs were generally fitted including the separation of atom types since there is barely any extra computational cost involved when applying the MTPs. However, to test the importance of introducing this additional chemical knowledge and of increasing the number of parameters, for MOF-5 and MOF-74 MTPs at level 22 were also trained when only using 4 atom types (one for each chemical element).

Due to the stochastic nature of the initialization of the training procedure, multiple potentials were trained for each system and reference data set. A possible strategy for choosing the final MTP would have been to pick the one with the smallest cost function. However, it is common practice to judge the quality of a fitted potential based on quantities independent of the parametrization procedure, like the properties of a validation set¹¹². In this way, one can judge the performance of the interpolation and not just the description of the training set, which could also be described well by the MTP due to overfitting. Therefore, for the further analysis and for the calculation of physical properties, the potentials were chosen that best managed to reproduce the forces, energies and stresses in the validation set and that were stable up to at least 300 K. In this context it should be mentioned that for high level MTPs (like the level 22 MTPs used for most parts of this manuscript), the variations in the properties by the multiply fitted potentials are typically rather small and picking the MTPs with the smallest cost functions would have made an only marginal difference for the force, energy and stress errors and would not have impacted any of the trends discussed here. The only exception is the thermal stability of the potentials, where some substantial deviations were found between the differently initialized MTPs. This is why training several MTPs can represent a workaround for issues related to the thermal stability. Corresponding details are shown in Supplementary Table 8. The data in that table also illustrate that it can happen that certain force fields provide a clearly worse description of the validation data without having a significantly increased cost function in the original parametrization, which is the primary reason why a multitude of MTPs were trained for each data set.

Evaluation of benchmark properties

When creating a new force field potential, proper validation is crucial and thus we aim to provide a large array of properties to benchmark based on DFT and experimental reference data. In

general, the single point and MD simulations for the DFT references and the VASP MLP were performed directly using VASP and for the MTPs using the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) package¹¹³ interfacing MLIP⁵⁸.

Structure optimizations at 0 K for the machine learned potentials were carried out starting from the DFT relaxed structures and for the VASP MLPs the same optimization settings were used as for the actual DFT calculations. For the MTPs an iterative process was used, where first the atomic positions and lattice parameters were optimized using a conjugate gradient algorithm. This was followed by an optimization of only the atomic positions and both steps were repeated 10 times. The end of either of the steps was defined as the situation, when either 1000 force evaluations had been performed (or technically also when the excessive convergence criterion of force changes falling to $<10^{-8} \text{ eV}\text{\AA}^{-1}$ had been reached).

The elastic stiffness tensors were computed using a finite differences approach based on the stress-strain relationship⁹⁰, where the lattice was strained anisotropically by a fraction of 0.01. For each strain direction, an optimization of atomic positions was carried out with the MTP and the resulting stresses were used for the evaluation. For DFT and the machine learned potential in VASP, the strained cells were not individually optimized, but rather the contributions from relaxing the ions were computed from the second-order force constants in the harmonic approximations to save computation time (as implemented in VASP using the IBRION = 6 setting).

Phonon band structures were computed using a supercell-based finite-differences approach as implemented in phonopy¹¹⁴. For the required atomic displacements, a displacement distance of 0.01 Å was used. The used supercells were carefully converged. Whenever possible, this was done based on DFT results. However, due to the large unit cells in MOFs, for some of the systems, supercell convergence tests were possible only based on the force-field results. For specific details regarding these tests and the supercells used for each system, view Supplementary Notes 1.4 and 3.4.

To evaluate the thermal expansion coefficients, molecular dynamics simulations in the NPT ensemble were performed for the MTPs at temperatures ranging between 100 K to 700 K. This was done over 100,000 time steps for supercells of the respective system (details can be found in Supplementary Note 4.1). In addition to providing temperature dependent lattice parameters, this also allows a rough assessment of the thermal stability of the obtained force field potentials. To obtain the actual thermal expansion coefficients linear fits (sometimes in restricted temperature ranges) were performed over the averaged lattice parameters from the thermally stable MD simulations.

To obtain the thermal conductivity of MOF-5 for comparing the experimental data to the MTP simulations, non-equilibrium molecular dynamics (NEMD)⁹² and ‘approach to equilibrium molecular dynamics’ (AEMD)^{115,116} simulations were performed. For NEMD, a supercell based on the lattice parameters at the desired temperature of 300 K obtained from the preceding NPT simulations was constructed, as detailed below. First, an equilibration at 300 K was performed in an NVT ensemble for 25 ps. Afterwards, the system was treated in an NVE ensemble while employing the Müller-Plathe algorithm¹¹⁷, which periodically swaps the kinetic energies of the 16 highest energy atoms in the first slab of the supercell with the kinetic energies of the 16 lowest energy atoms in the centre slab of the supercell every 4800 time steps. Here, each slab was chosen to be 12.98 Å long, encompassing one node and one linker of MOF-5. This leads to the creation of a hot and a cold region while maintaining the total energy in the system. Consequently, a heat flux from the hot to the cold region emerges allowing the straightforward computation of the thermal conductivity using Fourier’s law in combination

with the temperature gradient. Finite size effects due to scattering at the thermostat boundaries were accounted for by computing the thermal conductivity for several different cell lengths and performing an extrapolation to the infinite size limit^{91,92}. Perpendicular to the heat-flow direction, the supercell comprised 2 conventional unit cells. For the AEMD simulations, we used a similarly shaped cell, but this time one half of the cell was set to a higher temperature of 350 K and the other half of the cell was set to a lower temperature of 250 K. This is realized by using a thermostat on one half of the supercell, while not solving the equations of motion for atoms in the other half. This step was repeated, but now equilibrating the other half of the supercell to its starting temperature, while the equations of motions were not solved for the already thermalized half of the supercell. Then the thermostats were turned off and the entire system was treated in an NVE ensemble. The temperature difference between both halves of the system was recorded as a function of time during the molecular dynamics simulation. The temperature decay can be fitted to a function containing the thermal diffusivity, which in turn leads to the thermal conductivity¹¹⁶. Similar to NEMD, AEMD is also prone to finite size effects, which can be corrected using a slightly different extrapolation approach than for NEMD derived from the Boltzmann transport equation¹¹⁸. For more information regarding thermal conductivity calculations (including a justification for the used supercell sizes) see Supplementary Note 4.2.

To evaluate the computational efficiency, separate simulations in an NPT ensemble were performed at 300 K for at least 2000 time steps using large supercells comprising 27136 atoms for MOF-5 and 31104 atoms for MOF-74. Here, care was taken not to include the initialization overhead and not to lose a significant amount of time to outputting the trajectories.

For obtaining observables based on the traditional force field potentials for MOF-5 and MOF-74, LAMMPS was employed using the same methods as specified for the MTPs. To assign the parameters for the Dreiding and UFF4MOF potentials to the structure files, the lammps-interface code written by Boyd et al. was used¹⁴. This code has been developed to benchmark various force fields for several MOFs like MOF-5, UiO-66 and IRMOF-10. For a more specific discussion and analysis regarding UFF4MOF, see the Supplementary Notes 3.1 and 3.5. To compare the machine-learned potentials to a system-specifically fitted traditional force fields, a MOF-FF type FFP was used. The parametrization of that force field is described in detail in our previous work³³.

DATA AVAILABILITY

The datasets generated and/or analysed during the current study are available in the TU Graz Repository; <https://doi.org/10.3217/wyc7s-8en40>.

CODE AVAILABILITY

VASP can be acquired from the VASP Software GmbH (see www.vasp.at/); LAMMPS is available at www.lammps.org; MLIP is available at mlip.skoltech.ru/download; the lammps-mlip interface is available at github.com/ashapeev/interface-lammps-mlip-2; Phonopy is available at [phonopy.github.io/phonopy](https://github.com/phonopy/phonopy); further scripts used to compute some of the properties of the MTPs, like the elastic stiffness tensors, are publicly available at <https://doi.org/10.3217/wyc7s-8en40>.

Received: 3 August 2023; Accepted: 9 January 2024;

Published online: 20 January 2024

REFERENCES

- Ranocchiari, M. & Bokhoven, J. A. Van Catalysis by metal-organic frameworks: fundamentals and opportunities. *Phys. Chem. Chem. Phys.* **13**, 6388–6396 (2011).
- Dhakshinamoorthy, A., Li, Z. & Garcia, H. Catalysis and photocatalysis by metal organic frameworks. *Chem. Soc. Rev.* **47**, 8134–8172 (2018).
- Li, D., Xu, H. Q., Jiao, L. & Jiang, H. L. Metal-organic frameworks for catalysis: state of the art, challenges, and opportunities. *EnergyChem* **1**, 100005 (2019).
- Li, H. et al. Recent advances in gas storage and separation using metal-organic frameworks. *Mater. Today* **21**, 108–121 (2018).
- Fan, W., Zhang, X., Kang, Z., Liu, X. & Sun, D. Isorecticular chemistry within metal-organic frameworks for gas storage and separation. *Coord. Chem. Rev.* **443**, 213968 (2021).
- Jia, T., Gu, Y. & Li, F. Progress and potential of metal-organic frameworks (MOFs) for gas storage and separation: a review. *J. Environ. Chem. Eng.* **10**, 108300 (2022).
- Stassen, I. et al. An updated roadmap for the integration of metal-organic frameworks with electronic devices and chemical sensors. *Chem. Soc. Rev.* **46**, 3185–3241 (2017).
- Allendorf, M. D. et al. Electronic devices using open framework materials. *Chem. Rev.* **120**, 8581–8640 (2020).
- Zhang, L., Zhou, Y. & Han, S. The role of metal-organic frameworks in electronic sensors. *Angew. Chem.* **133**, 15320–15340 (2021).
- Velásquez-Hernández, M. et al. Towards applications of bioentities@MOFs in biomedicine. *Coord. Chem. Rev.* **429**, 213651 (2021).
- Lawson, H. D., Walton, S. P. & Chan, C. Metal-organic frameworks for drug delivery: a design perspective. *ACS Appl. Mater. Interfaces* **13**, 7004–7020 (2021).
- Dürholt, J. P., Fraux, G., Coudert, F. X. & Schmid, R. Ab initio derived force fields for zeolitic imidazolate frameworks: MOF-FF for ZIFs. *J. Chem. Theory Comput.* **15**, 2420–2432 (2019).
- Bristow, J. K., Tiana, D. & Walsh, A. Transferable force field for metal-organic frameworks from first-principles: BTW-FF. *J. Chem. Theory Comput.* **10**, 4644–4652 (2014).
- Boyd, P. G., Moosavi, S. M., Witman, M. & Smit, B. Force-field prediction of materials properties in metal-organic frameworks. *J. Phys. Chem. Lett.* **8**, 357–363 (2017).
- Vanduyfhuys, L. et al. QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *J. Comput. Chem.* **36**, 1015–1027 (2015).
- Bristow, J. K., Skelton, J. M., Svane, K. L., Walsh, A. & Gale, J. D. A general forcefield for accurate phonon properties of metal-organic frameworks. *Phys. Chem. Chem. Phys.* **18**, 29316–29329 (2016).
- Mayo, S. L., Olafson, B. D. & Goddard, W. A. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909 (1990).
- Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
- Addicoat, M. A., Vankova, N., Akter, I. F. & Heine, T. Extension of the universal force field to metal-organic frameworks. *J. Chem. Theory Comput.* **10**, 880–891 (2014).
- Coupry, D. E., Addicoat, M. A. & Heine, T. Extension of the universal force field for metal-organic frameworks. *J. Chem. Theory Comput.* **12**, 5215–5225 (2016).
- Yang, Y., Ibikunle, I. A., Gallis, D. F. S. & Sholl, D. S. Adapting UFF4MOF for heterometallic rare-earth metal-organic frameworks. *Appl. Mater. Interfaces* **14**, 54101–54110 (2022).
- Islamov, M. et al. High-throughput screening of hypothetical metal-organic frameworks for thermal conductivity. *NPJ Comput. Mater.* **9**, 11 (2023).
- Baxter, S. J. et al. Tuning thermal expansion in metal-organic frameworks using a mixed linker solid solution approach. *J. Am. Chem. Soc.* **141**, 12849–12854 (2019).
- Anderson, R. & Gómez-Gualdrón, D. A. Increasing topological diversity during computational “synthesis” of porous crystals: how and why. *CrystEngComm* **21**, 1653–1665 (2019).
- Yang, Y. & Sholl, D. S. A systematic examination of the impacts of MOF flexibility on intracrystalline molecular diffusivities. *J. Mater. Chem. A* **10**, 4242–4253 (2022).
- Vandenhoute, S., Cools-Ceuppens, M., DeKeyser, S., Verstraelen, T. & Van Speybroeck, V. Machine learning potentials for metal-organic frameworks using an incremental learning approach. *NPJ Comput. Mater.* **9**, 053804 (2023).
- Wieser, S., Kamencek, T., Schmid, R., Bedoya-Martínez, N. & Zojer, E. Exploring the impact of the linker length on heat transport in metal-organic frameworks. *Nanomaterials* **12**, 2142 (2022).
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
- Sun, H. COMPASS: An ab initio force-field optimized for condensed-phase applications—overview with details on alkane and benzene compounds. *J. Phys. Chem. B* **102**, 7338–7364 (2002).
- Kamencek, T. et al. Evaluating computational shortcuts in supercell-based phonon calculations of molecular crystals: the instructive case of naphthalene. *J. Chem. Theory Comput.* **16**, 2716–2735 (2020).

31. Nejat Pishkenari, H. & Ghaf Ghanbari, P. Vibrational properties of C60: a comparison among different inter-atomic potentials. *Comput. Mater. Sci.* **122**, 38–45 (2016).
32. Bureekaew, S. et al. MOF-FF - A flexible first-principles derived force field for metal-organic frameworks. *Phys. Status Solidi Basic Res.* **250**, 1128–1141 (2013).
33. Wieser, S. et al. Identifying the bottleneck for heat transport in metal-organic frameworks. *Adv. Theory Simul.* **4**, 2000211 (2021).
34. Van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
35. Furman, D. & Wales, D. J. Transforming the accuracy and numerical stability of reaxFF reactive force fields. *J. Phys. Chem. Lett.* **10**, 7215–7223 (2019).
36. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
37. Behler, J. & Csányi, G. Machine learning potentials for extended systems: a perspective. *Eur. Phys. J. B* **94**, 142 (2021).
38. Dragoni, D., Daff, T. D., Csányi, G., Marzari, N. & Achieving, D. F. T. Accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron. *Phys. Rev. Mater.* **2**, 013808 (2018).
39. Mueller, T., Hernandez, A. & Wang, C. Machine learning for interatomic potential models. *J. Chem. Phys.* **152**, 050902 (2020).
40. Eckhoff, M. & Behler, J. From Molecular fragments to the bulk: development of a neural network potential for MOF-5. *J. Chem. Theory Comput.* **15**, 3793–3809 (2019).
41. Tayfuroglu, O., Kocak, A. & Zorlu, Y. A neural network potential for the IRMOF series and its application for thermal and mechanical behaviors. *Phys. Chem. Chem. Phys.* **24**, 11882–11897 (2022).
42. Yang, C. T. et al. Deep learning neural network potential for simulating gaseous adsorption in metal-organic frameworks. *Mater. Adv.* **3**, 5299–5303 (2022).
43. Yu, Y., Zhang, W. & Mei, D. Artificial neural network potential for encapsulated platinum clusters in MOF-808. *J. Phys. Chem. C* **126**, 1204–1214 (2022).
44. Ying, P. et al. Sub-micrometer phonon mean free paths in metal-organic frameworks revealed by machine-learning molecular dynamics simulations. *Appl. Mater. Interfaces* **15**, 36412–36422 (2023).
45. Achar, S. K., Wardzala, J. J., Bernasconi, L., Zhang, L. & Johnson, J. K. Combined deep learning and classical potential approach for modeling diffusion in UiO-66. *J. Chem. Theory Comput.* **18**, 3593–3606 (2022).
46. Goeminne, R., Vanduyffhuys, L., Van Speybroeck, V. & Verstraelen, T. DFT-Quality adsorption simulations in metal-organic frameworks enabled by machine learning Potentials. *J. Chem. Theory Comput.* **19**, 6313–6325 (2023).
47. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
48. Fan, Z. et al. Neuroevolution machine learning potentials: combining high accuracy and low cost in atomistic simulations and application to heat transport. *Phys. Rev. B* **104**, 104309 (2021).
49. Wang, H., Zhang, L., Han, J. & E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178–184 (2018).
50. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
51. Shapeev, A. V. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
52. Byggmästar, J., Nordlund, K. & Djurabekova, F. Gaussian approximation potentials for body-centered-cubic transition metals. *Phys. Rev. Mater.* **4**, 093802 (2020).
53. Deringer, V. L. et al. Gaussian process regression for materials and molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
54. Rosenbrock, C. W. et al. Machine-learned interatomic potentials for alloys and alloy phase diagrams. *NPJ Comput. Mater.* **7**, 24 (2021).
55. Herbold, M. & Behler, J. Machine learning transferable atomic forces for large systems from underconverged molecular fragments. *Phys. Chem. Chem. Phys.* **25**, 12979–12989 (2023).
56. Jinnouchi, R., Karsai, F. & Kresse, G. On-the-fly machine learning force field generation: application to melting points. *Phys. Rev. B* **100**, 014105 (2019).
57. Verdi, C., Karsai, F., Liu, P., Jinnouchi, R. & Kresse, G. Thermal transport and phase transitions of zirconia by on-the-fly machine-learned interatomic potentials. *npj Comput. Mater.* **7**, 156 (2021).
58. Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn. Sci. Technol.* **2**, 025002 (2021).
59. Jinnouchi, R., Lahnsteiner, J., Karsai, F., Kresse, G. & Bokdam, M. Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with Bayesian inference. *Phys. Rev. Lett.* **122**, 225701 (2019).
60. Liu, P., Verdi, C., Karsai, F. & Kresse, G. A-B phase transition of zirconium predicted by on-the-fly machine-learned force field. *Phys. Rev. Mater.* **5**, 1–11 (2021).
61. Yaghi, O. M., O’Keeffe, M., Eddaoudi, M. & Li, H. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
62. Evans, D. J. Computer ‘experiment’ for nonlinear thermodynamics of Couette flow. *J. Chem. Phys.* **78**, 3297–3302 (1983).
63. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17**, 1302–1322 (1978).
64. Liu, P. et al. Combining machine learning and many-body calculations: coverage-dependent adsorption of CO on Rh(111). *Phys. Rev. Lett.* **130**, 78001 (2023).
65. Pallikara, I., Kayastha, P., Skelton, J. M. & Whalley, L. D. The physical significance of imaginary phonon modes in crystals. *Electron. Struct.* **4**, 033002 (2022).
66. Ziman, J. M. *Electrons and Phonons: The Theory of Transport Phenomena in Solids*. Vol. 540 (Oxford University Press, 2001).
67. Huang, B. L. et al. Thermal conductivity of a metal-organic framework (MOF-5): Part II measurement. *Int. J. Heat. Mass Transf.* **50**, 405–411 (2007).
68. Cavka, J. H. et al. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *J. Am. Chem. Soc.* **130**, 13850–13851 (2008).
69. Winarta, J. et al. A decade of UiO-66 research: a historic review of dynamic structure, synthesis mechanisms and characterization techniques of an archetypal metal-organic framework. *Cryst. Growth Des.* **20**, 1347–1362 (2020).
70. Valenzano, L. et al. Disclosing the complex structure of UiO-66 metal organic framework: a synergic combination of experiment and theory. *Chem. Mater.* **23**, 1700–1718 (2011).
71. Tranchemontagne, D. J., Hunt, J. R. & Yaghi, O. M. Room temperature synthesis of metal-organic frameworks: MOF-5, MOF-74. *Tetrahedron* **64**, 8553–8557 (2008).
72. Choe, J. H., Kim, H. & Hong, C. S. MOF-74 type variants for CO₂ capture. *Mater. Chem. Front.* **5**, 5172–5185 (2021).
73. Kim, H. & Hong, C. S. MOF-74-type frameworks: tunable pore environment and functionality through metal and ligand modification. *CrystEngComm* **23**, 1377–1387 (2021).
74. Loiseau, T. et al. A rationale for the large breathing of the porous aluminum terephthalate (MIL-53) upon hydration. *Chem. - A Eur. J.* **10**, 1373–1382 (2004).
75. Liu, Y. et al. Reversible structural transition in MIL-53 with large temperature hysteresis. *J. Am. Chem. Soc.* **130**, 11813–11818 (2008).
76. Tomar, S. & Singh, V. K. Review on synthesis and application of mil-53. *Mater. Today Proc.* **43**, 3291–3296 (2021).
77. Lock, N. et al. Elucidating negative thermal expansion in MOF-5. *J. Phys. Chem. C* **114**, 16181–16186 (2010).
78. Lock, N. et al. Scrutinizing negative thermal expansion in MOF-5 by scattering techniques and AB initio calculations. *J. Chem. Soc. Dalton Trans.* **42**, 1996–2007 (2013).
79. Zhou, W., Wu, H., Yildirim, T., Simpson, J. R. & Walker, A. R. H. Origin of the exceptional negative thermal expansion in metal-organic framework-5 Zn₄O (1,4-benzenedicarboxylate) **3**. *Phys. Rev. B - Condens. Matter Mater. Phys.* **78**, 054114 (2008).
80. Burtch, N. C. et al. Negative thermal expansion design strategies in a diverse series of metal-organic frameworks. *Adv. Funct. Mater.* **29**, 1904669 (2019).
81. Yot, P. G. et al. Exploration of the mechanical behavior of metal organic frameworks UiO-66(Zr) and MIL-125(Ti) and their NH₂ functionalized versions. *Dalt. Trans.* **45**, 4283–4288 (2016).
82. Germann, L. S. et al. Real-Time in situ monitoring of particle and structure evolution in the mechanochemical synthesis of UiO-66 metal-organic frameworks. *Cryst. Growth Des.* **20**, 49–54 (2020).
83. Kamenec, T., Schrode, B., Resel, R., Ricco, R. & Zojler, E. Understanding the origin of the particularly small and anisotropic thermal expansion of MOF-74. *Adv. Theory Simul.* **5**, 2200031 (2022).
84. Rosi, N. L. et al. Rod packings and metal-organic frameworks constructed from rod-shaped secondary building units. *J. Am. Chem. Soc.* **127**, 1504–1518 (2005).
85. Molina, M. A., Manjón-Sanz, A. & Sánchez-Sánchez, M. On the contribution of pair distribution function (PDF) to the characterization of nanocrystalline MOFs: the case of M-MOF-74. *Microporous Mesoporous Mater.* **319**, 110973 (2021).
86. Mayoral, A., Sanchez-Sanchez, M., Alfayate, A., Perez-Pariente, J. & Diaz, I. Atomic observations of microporous materials highly unstable under the electron beam: the cases of Ti-doped AlPO₄-5 and Zn-MOF-74. *ChemCatChem* **7**, 3719–3724 (2015).
87. Nanthamathee, C., Ling, S., Slater, B. & Atfield, M. P. Contradistinct thermo-responsive behavior of isostructural MIL-53 type metal-organic frameworks by modifying the framework inorganic anion. *Chem. Mater.* **27**, 85–95 (2015).
88. Yot, P. G. et al. Metal-organic frameworks as potential shock absorbers: the case of the highly flexible MIL-53(Al). *Chem. Commun.* **50**, 9462–9464 (2014).
89. Jaeken, J. W. & Cottenier, S. Solving the Christoffel equation: phase and group velocities. *Comput. Phys. Commun.* **207**, 445–451 (2016).

90. Kamencek, T. & Zojer, E. Understanding the anisotropic elastic properties of metal–organic frameworks at the nanoscale: the instructive example of MOF-74. *J. Phys. Chem.* **125**, 24728–24745 (2021).
91. Li, Z. et al. Influence of thermostatting on nonequilibrium molecular dynamics simulations of heat conduction in solids. *J. Chem. Phys.* **151**, 234105 (2019).
92. Schelling, P. K., Phillpot, S. R. & Keblinski, P. Comparison of atomic-level simulation methods for computing thermal conductivity. *Phys. Rev. B - Condens. Matter Mater. Phys.* **65**, 144306 (2002).
93. Queen, W. L. et al. Site-specific CO₂ adsorption and zero thermal expansion in an anisotropic pore network. *J. Phys. Chem. C* **115**, 24915–24919 (2011).
94. Lock, N., Wu, Y., Christensen, M., Iversen, B. B. & Kepert, C. J. Elucidating negative thermal expansion in metal-organic frameworks. *Acta Crystallogr. Sect. A Found. Crystallogr.* **114**, 16181–1686 (2008).
95. Hockney, R. *Computer Simulation Using Particles* Vol. 540 (CRC Press, 1989).
96. Bharadwaj, Prakash, K. B. & Kanagachidambaresan, G. R. *Pattern Recognition and Machine Learning* Vol. 738 (Springer Science & Business Media, 2021).
97. Jinnouchi, R., Miwa, K., Karsai, F., Kresse, G. & Asahi, R. On-the-fly active learning of interatomic potentials for large-scale atomistic simulations. *J. Phys. Chem. Lett.* **11**, 6946–6955 (2020).
98. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
99. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metalamorphous– semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
100. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
101. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B - Condens. Matter Mater. Phys.* **54**, 11169–11186 (1996).
102. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B - Condens. Matter Mater. Phys.* **59**, 1758–1775 (1999).
103. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
104. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
105. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **78**, 1396–1396 (1997).
106. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
107. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
108. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
109. Mahoney, M. W. & Drineas, P. CUR matrix decompositions for improved data analysis. *Proc. Natl Acad. Sci. USA* **106**, 697–702 (2009).
110. Golub, G. H. & Van Loan, C. F. *Matrix Computations* 3rd Edn., Vol. 784 (Johns Hopkins Univ. Press, 1996).
111. Fletcher, R. *Practical Methods of Optimization* Vol. 436 (John Wiley & Sons, Ltd, 2000).
112. Morrow, J. D., Gardner, J. L. A. & Deringer, V. L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **158**, 121501 (2023).
113. Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
114. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
115. Lampin, E., Palla, P. L., Francioso, P. A. & Cleri, F. Thermal conductivity from approach-to-equilibrium molecular dynamics. *J. Appl. Phys.* **114**, 033525 (2013).
116. Melis, C., Dettori, R., Vandermeulen, S. & Colombo, L. Calculating thermal conductivity in a transient conduction regime: theory and implementation. *Eur. Phys. J. B* **87**, 96 (2014).
117. Müller-Plathe, F. A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. *J. Chem. Phys.* **106**, 6082–6085 (1997).
118. Zaoui, H., Palla, P. L., Cleri, F. & Lampin, E. Length dependence of thermal conductivity by approach-to-equilibrium molecular dynamics. *Phys. Rev. B* **94**, 054304 (2016).
119. Momma, K. & Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **44**, 1272–1276 (2011).
120. Setyawan, W. & Curtarolo, S. High-throughput electronic band structure calculations: challenges and tools. *Comput. Mater. Sci.* **49**, 299–312 (2010).

ACKNOWLEDGEMENTS

The authors thank Lukas Legenstein, Lukas Reicht, Nina Strasser and Florian Lindner for insightful discussions and Daniel Fashing for a preliminary benchmarking of MTPs. The authors also thank the group of Georg Kresse, in particular the machine learned force field team Ferenc Karsai, Andreas Singraber and Jonathan Lahnsteiner for their help in the training of the VASP MLPs. S.W. is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute of Solid State Physics (26163). This research was funded in part, by the Austrian Science Fund (FWF) [Grant-DOI: 10.55776/P33903-N]. For the purpose of open access, the authors have applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission. Partial funding by the Graz University of Technology through the Lead Project Porous Materials @ Work (for Sustainability) (LP-03) is also acknowledged. The computational results presented have been generated using the Vienna Scientific Cluster (VSC).

AUTHOR CONTRIBUTIONS

SW and EZ conceptualized the presented research. SW performed all simulations and continuously discussed the outcomes with EZ and his group (see Acknowledgements). The first draft of the manuscript was written by SW and was revised by him and EZ in several iterations including input from the Kresse group. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-024-01205-w>.

Correspondence and requests for materials should be addressed to Egbert Zojer.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024