Machine learned Force-Fields for an ab-initio Quality Description of Metal-Organic Frameworks

SUPPLEMENTARY INFORMATION

Sandro Wieser[1], Egbert Zojer[1]

[1]*Institute of Solid State Physics, Graz University of Technology, NAWI Graz, Petersgasse 16, 8010 Graz, Austria*

Table of Contents

Supplementary Discussion

Supplementary Note 1. DFT settings, convergence and computation of the benchmarking data

Supplementary Note 1.1 Overview of the chosen settings

To add more details regarding the used settings for the DFT simulations carried out by VASP, Supplementary Table 1 shows the chosen k-point sampling and energy cutoff for each system. Detailed convergence tests for all the settings can be found in Supplementary Note 1.2. Since the accuracy of the forces is vital for many of the investigated properties, the precision-mode in VASP[1–3] was set to PREC=Accurate. The electronic minimization algorithm was set to ALGO=FAST for all systems except for UiO-66, where ALGO=NORMAL was required to consistently achieve SCF convergence. A Gaussian smearing was used for the partial occupancies of the orbitals with a width of SIGMA = 0.05 eV. To account for dispersive forces, the D3 scheme by Grimme et al. with Becke-Johnson damping was used[4,5]. The used pseudopotentials required for the projector-augmented wave (PAW)[6] method are listed in

Supplementary Table 2 for each element. The evaluation of the projection operators was set to be performed in reciprocal space (LREAL=.FALSE.) for improved accuracy.

Supplementary Table 1: Chosen Γ-centered k-point mesh and energy cutoff for the investigated materials. The values were carefully chosen as described in Supplementary Note 1.2. For MIL-53 (lp) the same k-point sampling mesh as for MIL-53 (np) was used for the molecular-dynamics steps since a phase transition between lp and np would be technically possible. For MIL-53 (np) a denser k-point sampling mesh of 2×4×4 was used for the reduced unit cell containing only 38 atoms.

System	k-point sampling	Energy cutoff / eV
MOF-5	1×1×1	900
UiO-66	1×1×1	900
MOF-74	2×2×2	900
MIL-53 (lp)	1×2×1*	900
MIL-53 (np)	1×2×2 (2×4×4)	900

Supplementary Table 2: Used PAW pseudopotentials for each element. The VASP default sets were used. For the title, the names given after the 'TITEL' flag in the respective POTCAR files is listed.

Element name	PAW PBE title
Zn	PAW_PBE Zn 06Sep2000
Al	PAW_PBE Al 04Jan2001
Zr	PAW_PBE Zr_sv 04Jan2005
O	PAW_PBE O 08Apr2002
C	PAW_PBE C 08Apr2002
H	PAW_PBE H 15Jun2001

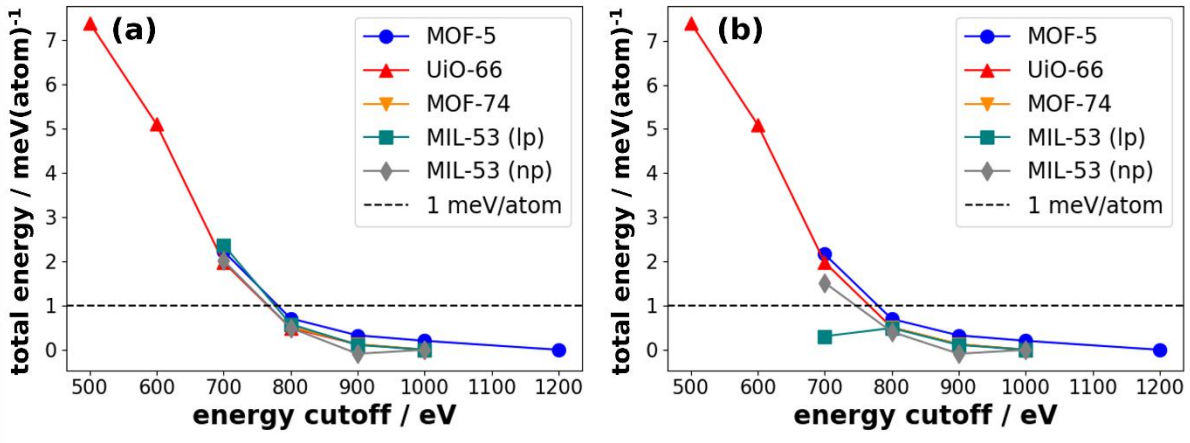
Supplementary Note 1.2 K-point and energy cutoff convergence and minimization of the structures

The k-point mesh and energy cutoff for all the systems were converged until the deviation in the total energy was below 1 meV(atom)$^{-1}$. However, this was not the only requirement, because it is far more important that the actually investigated properties are converged. Especially elastic tensor elements have been proven to require highly accurate DFT settings[7,8]. This is why we also provide convergence tests regarding the phonon properties and elastic constants.

For the relaxation procedure, symmetries were turned off in VASP and all constraints were removed. The energy was then minimized until the maximum absolute force component in the system was below 10^{-3} eVÅ^{-1}. Due to the very flat energy landscape of MIL-53, the initial relaxation for each phase at 900 eV energy cutoff and the k-point mesh specified in Supplementary Table 2 was performed as a 2-step procedure. In this case, a second relaxation run starting from the final structure from the first run was appended to minimize the effect of Pulay stresses. In the end, the result did not change much using this approach due to the relatively large plane-wave energy cutoff which already decreases the impact of Pulay stresses. This approach and the full relaxation from an experimentally obtained structure was only performed for one of the settings, for the other settings an already pre-relaxed cell was used as the input to save computation time.

Supplementary Figure 1 shows the convergence of the DFT obtained total energy for each system for two cases: the single point energy for the same atomic positions without further optimizations using the tested settings and also the total energy after a full relaxation. The k-point mesh used for these tests corresponds to Supplementary Table 1. It can be seen in (a), that beginning with an energy cutoff of 800 eV, the differences in energy are below 1 meV(atom)$^{-1}$ which is one of our chosen convergence criteria. In (b) the situation looks relatively similar overall, but for the large pore phase of MIL-53 the energy at a cutoff of 700 eV also seems very similar to the higher settings. However, this is just an artifact of the relaxation run finding a different minimum structure. In fact, one of the lattice parameters reaches a value of 18.93 Å at 700 eV instead of the 17.26 Å obtained for the higher

settings. This emphasizes the extremely flat energy landscape of MIL-53 and illustrates, why a careful convergence of the chosen settings is crucial.



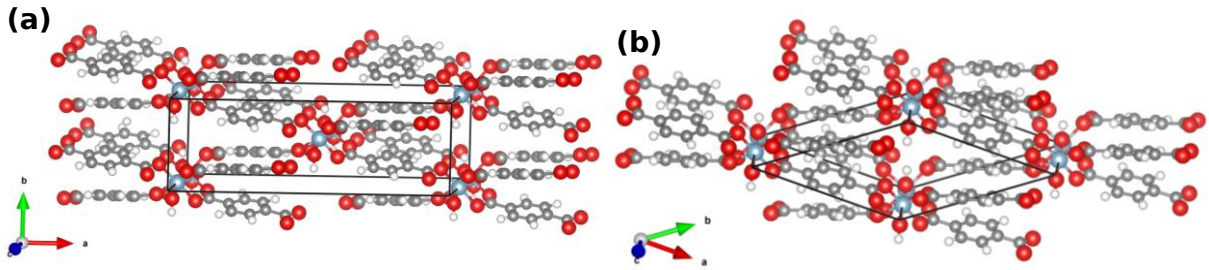
Supplementary Figure 1: Convergence of the total energy obtained from DFT as a function of the energy cutoff for the same unit cell (a) and after a full relaxation (b) for each system. The origin of the energy axis was defined as the total energy calculated for the highest energy cutoff. The black dashed line indicates the chosen convergence criterion of the energy differences being smaller than 1 meV(atom)$^{-1}$.

In Supplementary Table 3 the total energies for various k-point sampling meshes are shown. It can be seen that the finally chosen k-point meshes are less than 1 meV(atom)$^{-1}$ different compared to their denser counterparts for all of the systems. Another interesting aspect in this table is the small difference of the energies for the two phases of MIL-53 where the narrow pore phase shows a less than 3.8 meV(atom)$^{-1}$ lower energy.

Regarding the convergence of MIL-53 (np), it should be noted that a smaller unit cell was found based on the system symmetries after relaxation. This smaller cell contains only 38 atoms. The considerations in this Note apply to the larger cell with 76 atoms. However, the smaller cell was used in the phonon calculations and the evaluation of the elastic stiffness tensor. This will lead to a lower number of phonon modes and a differently oriented elastic tensor. This smaller cell shows a different orientation with a=10.42 Å, b=10.42 Å, c=6.66 Å, α=103.4°, β=103.4°, γ=36.6° and has space group C2/c (15). The difference between the two symmetry-equivalent cells is visualized in Supplementary Figure 2. Due to the smaller extent and the different orientation of the new cell, a different k-point sampling had to be used. Here, we chose a denser 2×2×4 grid to ensure convergence.

Supplementary Table 3: Total energies of the investigated systems for various k-point meshes at an energy cutoff of 900 eV. Total energies are provided for both a single point calculation of the same atomic positions and lattice parameters for each setting and also for the final setting-specific relaxed geometry. The k-point mesh reaching the desired maximum energy difference of 1 meV(atom)$^{-1}$ compared to the highest tested setting is indicated in bold.

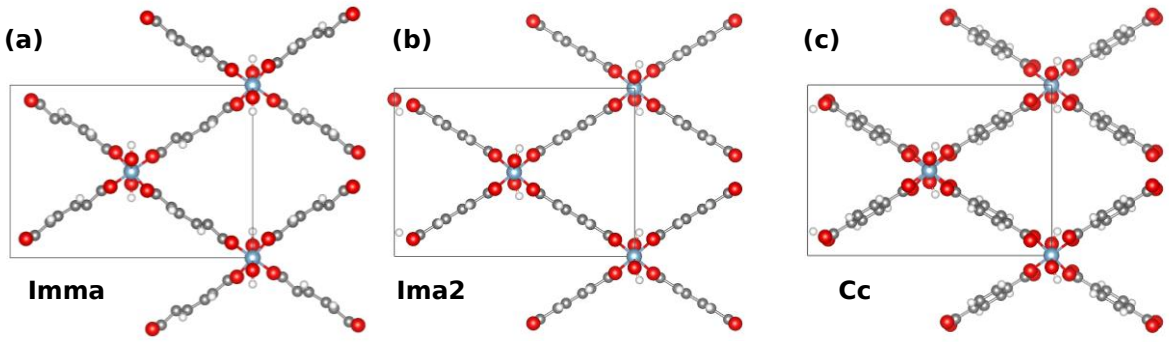
System name	k-point mesh	Energy cutoff / eV	Total energy / meV(atom)$^{-1}$	Total energy after relaxation / meV(atom)$^{-1}$
MOF-5	**1×1×1**	900	-6850.6	-6850.6
	2×2×2	900	-6850.6	-6850.6
UiO-66	**1×1×1**	900	-7525.0	-7525.0
	2×2×2	900	-7525.3	-7525.3
MOF-74	1×1×1	900	-6980.7	-6981.2
	2×2×2	900	-6987.0	-6987.0
	3×3×3	900	-6987.1	-6987.1
MIL-53 (lp)	1×1×1	900	-7121.8	-7130.9
	1×2×1	900	-7133.1	-7133.1
	2×3×2	900	-7133.1	-7133.1
MIL-53 (np)	**1×2×2**	900	-7136.9	-7136.9
	2×4×4	900	-7136.7	-7136.7
	2×6×6	900	-7136.7	-7136.7



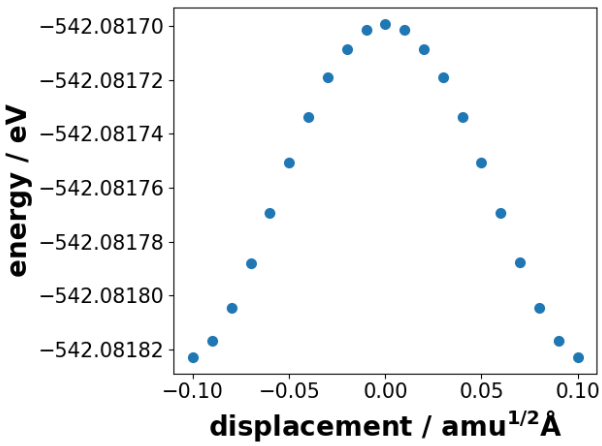
Supplementary Figure 2: Unit cell of MIL-53 (np) with 76 atoms (a) and the smaller but equivalent 38 atom unit cell (b).

Another oddity that occurred is the breaking of the orthorhombic symmetry during the relaxation of MIL-53 (lp) leading to a more energetically favorable structure with space group Cc (9). In this case, slight deviations from the 90° angles were observed for the lattice parameters amounting to ± 1° for two equivalent minima. To visualize this, a comparison of the orthorhombic structure and non-orthorhombic structure is shown in Supplementary Figure 3. Additionally depicted is the effect of the hydrogens in the nodes bending in a way that they do not align with the orientation of the node octahedra any more. The origin of that bending is that the straight geometry (central panel) represents

an energetic saddle point. The structure becomes energetically favorable when the hydrogens bend either way. This is shown in Supplementary Figure 4 by calculating the total energy of the system for displacing the structure along the relevant phonon eigenvector obtained for the structure with the central hydrogens (see Supplementary Figure 3 (a)). This leads to imaginary phonon frequencies for that structure. However, it should be stressed that the large pore phase of MIL-53 usually occurs only at higher temperatures, while the data in Supplementary Figure 4 represent tiny energy differences at 0 K. This means that in reality the experimentally measurable phase would be the average between the two minimum configurations. We confirmed this using our trained force field potentials where the slight tilt angle disappeared on average for molecular dynamics simulations for temperatures above 150 K, as shown in detail in Supplementary Note 4.1. This implies that what is observed here is merely a low temperature effect, but it is still important, as the lower degree of symmetry, for example, increases the number of non-zero elastic tensor elements compared to what has been reported in other publications[9,10].



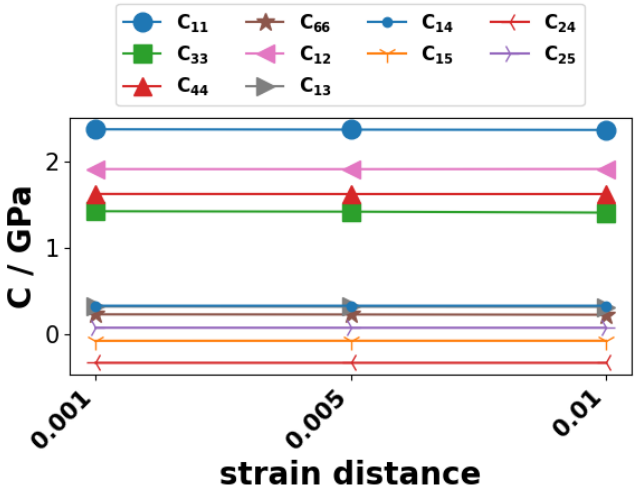
Supplementary Figure 3: MIL-53 (lp) at various stages of the energy minimization procedure. In (a) one sees the structure with space group Imma (that has also been reported in experiments[11,12]) with hydrogens aligned centrally between the linkers. In (b) a lower energy structure after further minimization is shown, where the hydrogens are bent in a consistent direction reducing the space group to Ima2 and in (c) a structure with even lower energy is shown, with a monoclinic angle of 91° and rotated linkers reducing the space group to Cc.



Supplementary Figure 4: Total energy as a function of displacement distance for the Imma structure of MIL-53 displaced along the imaginary phonon mode describing the bending motion of the central hydrogens.

Supplementary Note 1.3 Computation of the elastic tensor elements

In the following, convergence tests regarding the computation of the elastic tensor elements will be provided. The relevant quantity to converge is the applied fractional strain distance which was chosen as 0.01 in this work. The distance should, on one hand, be as small as possible so that the finite difference approach can still be applied but, on the other hand, should not be too small not to seriously suffer from numerical noise. Supplementary Figure 5 shows the elastic tensor elements in MOF-74 as a function of the strain distance. As can be seen, a reduction of this distance below 0.01 in the strain leads to no visible differences. The tests for slgihtly larger strain distances up to 0.02 were already performed in [7] and were also found to be inconsequential. Technical note: the test provided here only investigates the impact of the strain distance and not the displacement distance to obtain the interatomic force constants. When computing elastic constants as implemented in VASP (IBRION=6), the same parameter (POTIM) controls both the strain and interatomic displacement distances. If one changes this parameter, severe numerical issues occur at a displacement distance of 0.001 Å, while at 0.005 Å no significant differences were observed compared to 0.01 Å. Conclusively, for all displacement-based calculations of the force constants (which are also required to compute the phonons) in this work, we used a displacement distance of 0.01 Å. For all computations of elastic constants a relative strain of 0.01 was used.

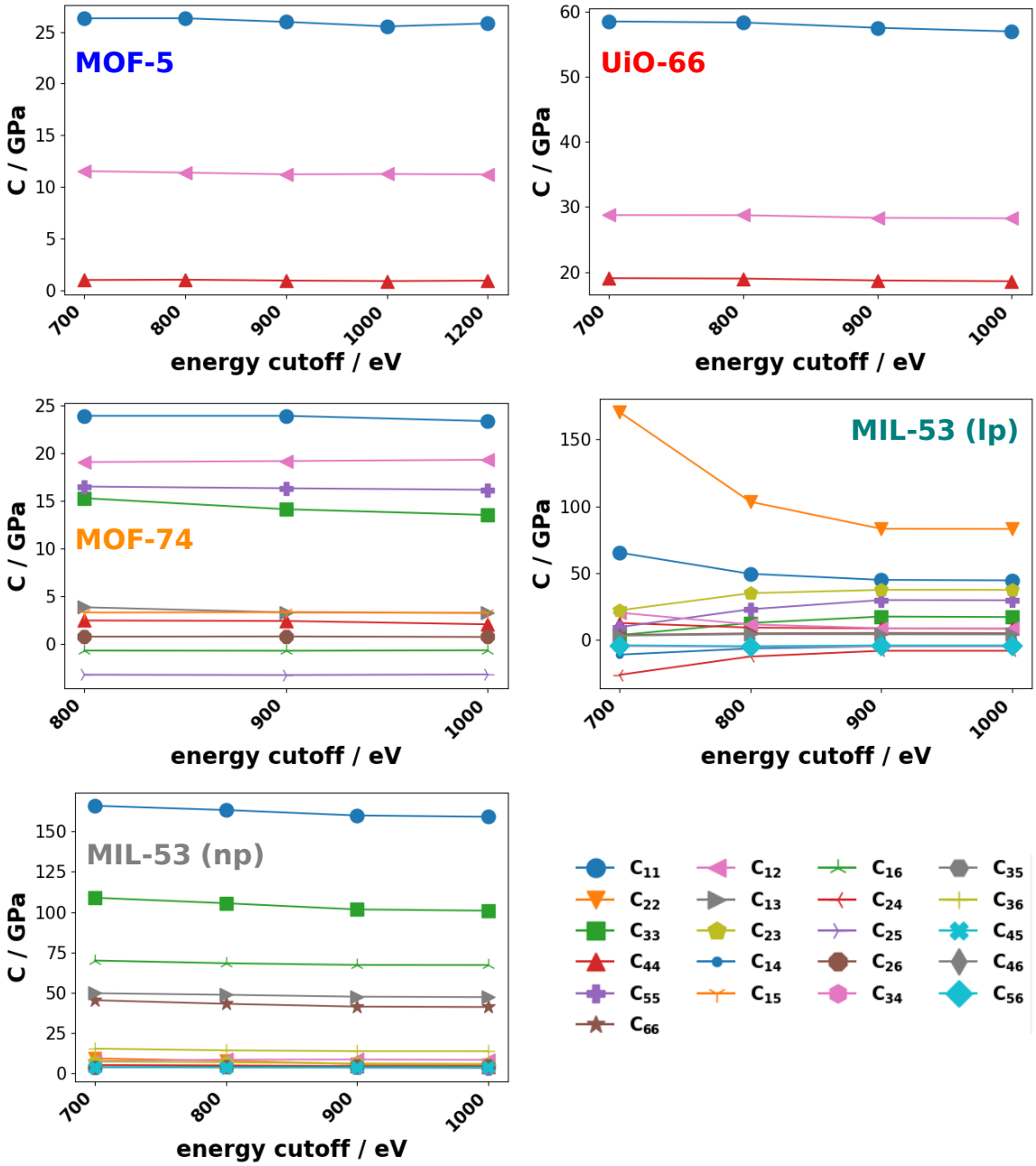


Supplementary Figure 5: Comparison of all unique elastic tensor elements in MOF-74 for different strain distances.

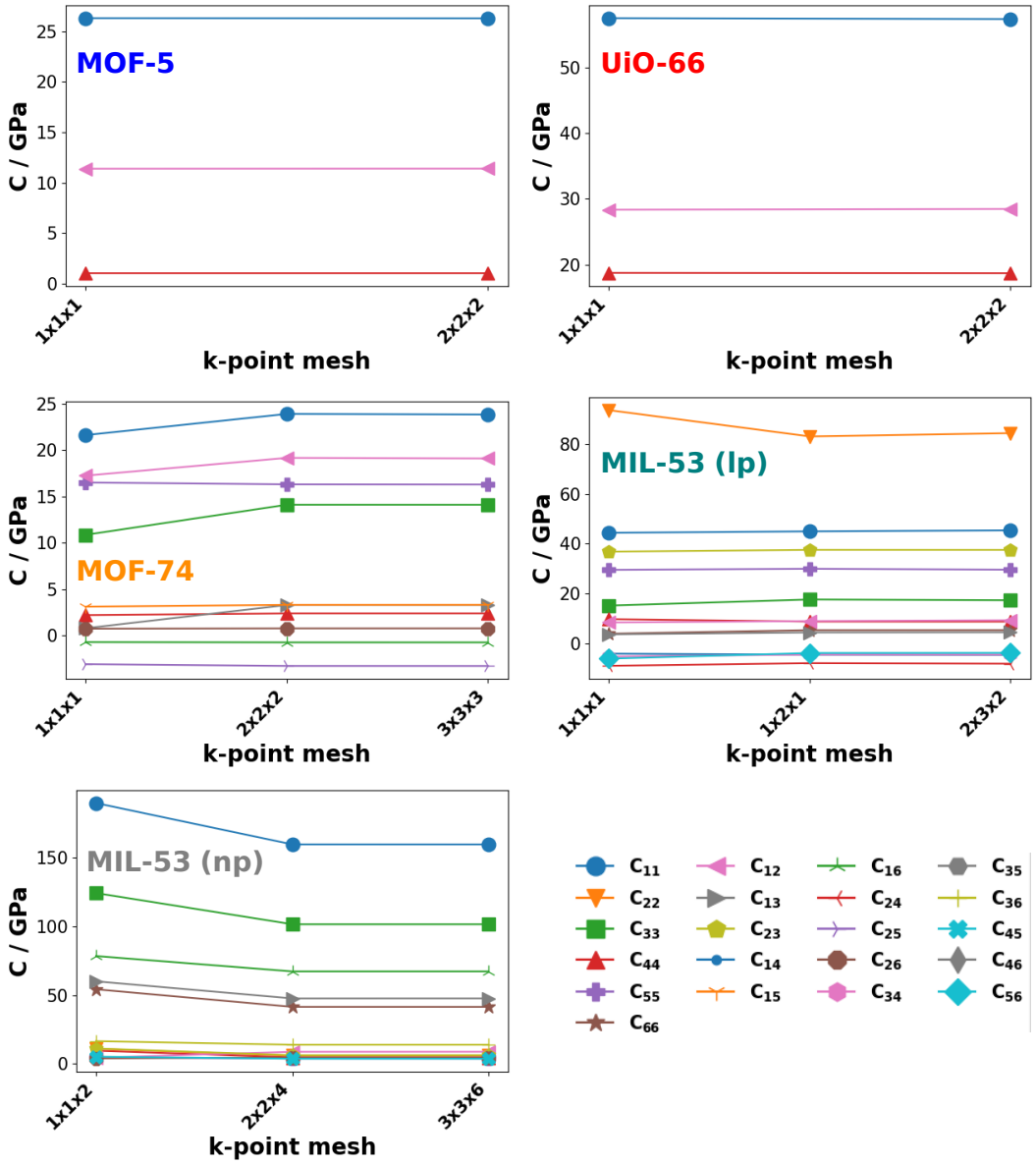
Additionally, we provide the energy cutoff convergence tests for the elastic tensor elements for each system in Supplementary Figure 6. This is done, despite the total energy changing only insignificantly after reaching 800 eV as shown in the previous Note, because the objective is to converge the actual properties of interest. The guideline of converging energies down to less than 1 meV(atom)$^{-1}$ is somewhat arbitrary and it depends on the objective of the calculation how severe the impact of the inaccuracy is. Here, the evaluation was performed using the optimized cell for the respective setting. In Supplementary Figure 6, we see that especially for MIL-53 (lp) and MOF-74 very high energy cutoffs of at least 900 eV are required to obtain properly converged elastic constants. For MOF-5, UiO-66 and MIL-53 (np) the difference for the different energy cutoffs is not that significant. However, small differences are still apparent even beyond a cutoff of 900 eV. So, if one would require extremely accurate elastic properties, going to even higher energy cutoffs might be necessary. For our purposes, however, we limit ourselves to 900 eV for each system as our main objective is not to provide the most accurate values possible, but rather to reproduce the DFT reference data with the machine learned potentials. In passing we note that in [7] it is shown for MOF-74 that 900 eV provide rather well converged values for the elastic tensor but that even higher cutoffs are required for converging the compliance tensor.

Supplementary Figure 7 shows the differences between the tested k-point sampling meshes. As can be seen, for all systems the k-meshes chosen based on the energy convergence listed in Supplementary Table 3 are also sufficient for converging the elastic constants. Here, it should be noted, that for MIL-53 (np) the smaller 38 atom unit cell was used for the computation of the elastic

tensor elements, which is the reason why the k-point convergence is very different than for MIL-53 (lp).



Supplementary Figure 6: Values of the unique elements of the elastic tensor for different energy cutoffs in the DFT calculations for each of the studied systems.



Supplementary Figure 7: Values of the unique elements of the elastic tensor for different k-point sampling meshes in the DFT calculations for each of the studied systems. For MOF-5, the elastic constants were computed for an energy cutoff of 800 eV and for all the other systems at 900 eV.

For the precise numerical values of the elastic tensor elements refer to the comparisons with the machine learned potentials in Supplementary Note 3.1. When computing the elastic tensor elements with DFT, the harmonic force constants are also provided. These can also be used to compute the phonon frequencies at the Γ point. So, we can also immediately provide the convergence tests for the vibrational frequencies. Due to the large number of modes in the MOFs, we will only report the RMSD values compared to the highest used energy cutoff in Supplementary Table 4. Additionally, tests for denser k-point sampling meshes are provided in Supplementary Table 4. It can be seen that for all

systems after reaching an energy cutoff of 900 eV and also for the denser k-point meshes than those that were ultimately used for the results, the RMSD values fall below a very small value of less than 1 cm^{-1}.

Supplementary Table 4: Deviations in the phonon frequencies at Γ for various k-point meshes and energy cutoffs. The RMSD values for results obtained with the same k-point sampling mesh (k-point mesh that was actually used for the presented results) are given compared to the entry in bold, which is for the highest energy cutoff used for the respective system. RMSD values for different k-point meshes are given compared to the results indicated with an asterisk, which is another calculation with the same energy cutoff for the k-point mesh that was actually used in the presented results.

System name	k-point mesh	Energy cutoff / eV	Γ-frequency RMSD / cm^{-1}
MOF-5	1×1×1	700	1.91
	1×1×1	*800	0.65
	1×1×1	900	0.36
	1×1×1	1000	0.38
	1×1×1	**1200**	-
	2×2×2	800	0.05
UiO-66	1×1×1	700	1.40
	1×1×1	800	0.64
	1×1×1	*900	0,16
	1×1×1	**1000**	**-**
	2×2×2	900	0.46
MOF-74	1×1×1	900	2.82
	2×2×2	800	0.80
	2×2×2	*900	0.16
	2×2×2	**1000**	**-**
	3×3×3	900	
MIL-53 (lp)	1×1×1	900	8.12
	1×2×1	700	4.02
	1×2×1	800	1.55
	1×2×1	*900	0.12
	1×2×1	**1000**	-
	2×3×2	900	0.35
MIL-53 (np)	1×1×2	900	20.54
	2×2×4	700	3.86
	2×2×4	800	1.68
	2×2×4	*900	0.09
	2×2×4	**1000**	-
	3×3×6	900	0.03

Supplementary Note 1.4 Computation of the phonon band structures

The phonon band structure calculations were performed employing phonopy[13] using a finite difference approach in a supercell of the respective material. The atomic displacement distance was

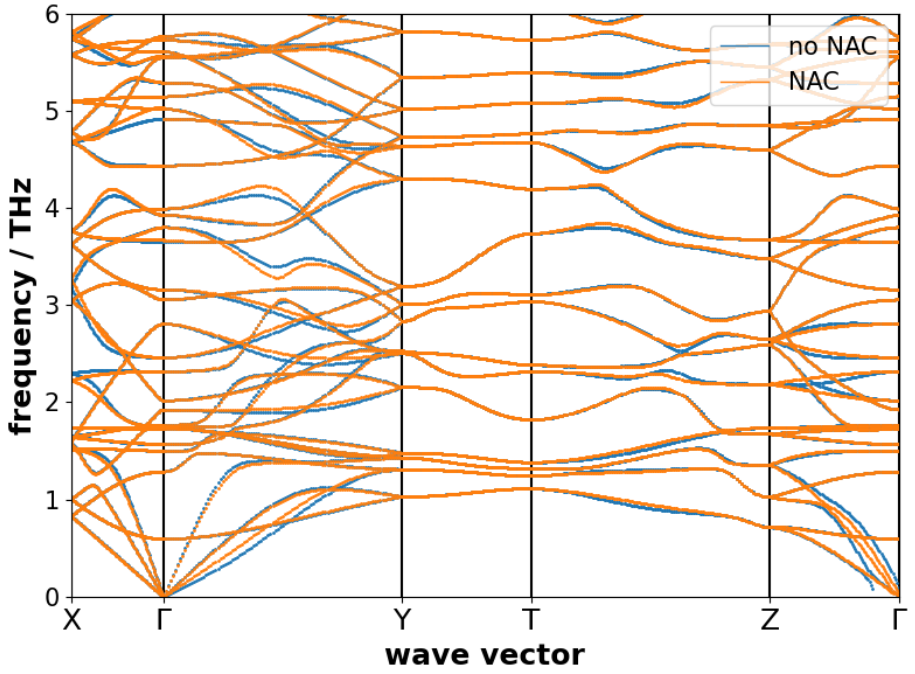
chosen as 0.01 Å and the supercell sizes are shown in Supplementary Table 5. Due to the large size of the supercells, the k-point mesh was set to Γ only for all systems. Due to the large computational expense of computing even larger supercells, convergence tests beyond these specified cells, were only carried out using the later trained MTPs which will be discussed in Supplementary Note 3.4. However, DFT includes longer range contributions to the forces, so the convergence is not entirely transferable. Therefore, some deviations might still occur. The remaining error is, however, expected to be small in view of the quite large supercells considered here. Additionally, we note that the supercell matrix chosen for MOF-5 and UiO-66 is used to create the cubic conventional unit cell from the primitive unit cell (for a visualization of the different unit cells for MOF-5, see the supporting information of [14]). The supercell for MOF-74 is used to create the hexagonal conventional unit cell from the primitive unit cell (for a more in-depth explanation and visualization of the different MOF-74 unit cells, see the supporting information of [7]). These differently shaped unit cells are then also used as a baseline to span larger supercells for the molecular dynamics simulations which were computed using the force field potentials.

Supplementary Table 5: Supercells used for the phonon band structure calculations for each of the investigated systems.

System name	Supercell
MOF-5	$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$
UiO-66	$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$
MOF-74	$\begin{pmatrix} 1 & 0 & 3 \\ -1 & 1 & 3 \\ 0 & -1 & 3 \end{pmatrix}$
MIL-53 (lp)	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$
MIL-53 (np)	$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

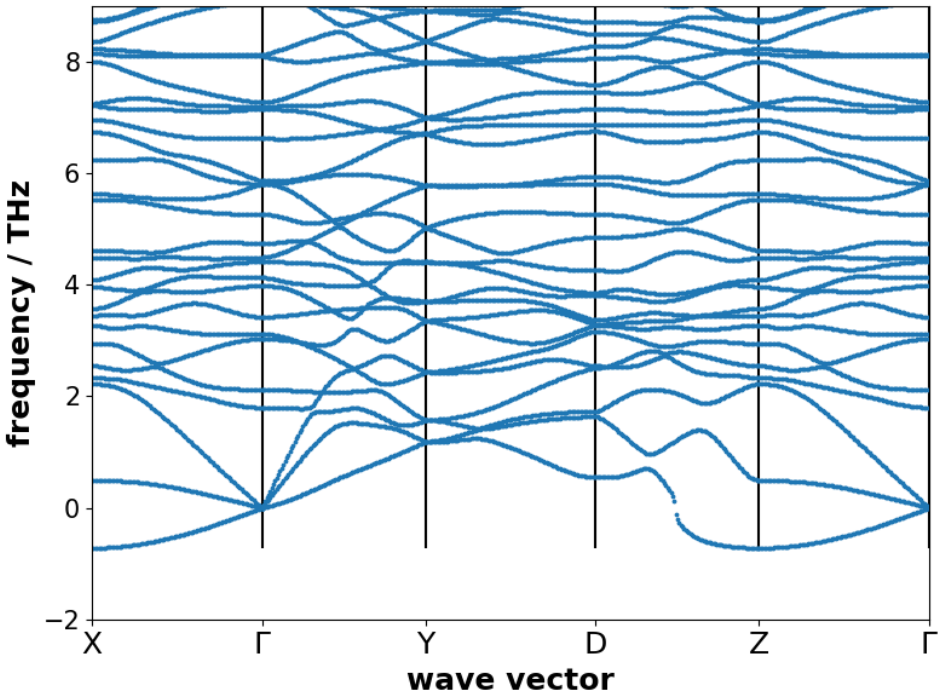
For the large pore phase of MIL-53, some small imaginary frequencies close to the Γ-point could be observed. However, this was corrected by applying a non-analytical term correction (NAC) [15–17] using the born-effective charges computed from the relaxed unit cell. The difference of the phonon bands with and without NAC correction is shown in Supplementary Figure 8. Beyond these imaginary

frequencies, the low frequency phonon band structures with and without NAC are very similar deviate only slightly.



Supplementary Figure 8: Comparison of MIL-53 (lp) low frequency phonon band structures without (blue) and with (orange) non-analytical term correction applied.

For the narrow pore phase of MIL-53 (np), calculating the phonon band structures led to imaginary modes beyond the Γ point as can be seen in Supplementary Figure 9 for a 2×4×2 supercell. This means, imaginary modes occur at the X point which for the chosen supercell should be exact, as it is commensurate with the used supercell. This could mean that actual crystal has a larger unit cell than the originally chosen one. So, we also tested a larger unit cell for building the supercells. This was the usually reported cell containing 76 atoms. After another relaxation and a new phonon computation, a new set of phonon bands was obtained, which once again showed imaginary modes at the X point. The procedure was repeated for a supercell of this larger cell – which was then used as the base unit cell – which did not resolve the issue either. Unfortunately, due to mounting computational cost it was deemed too expensive to consider even larger unit cells. A possible origin of the complications could be that the narrow-pore phase of MIL-53 is not stable at zero K and becomes a minimum of the free-energy landscape only at elevated temperatures. However, studying such aspects would go far beyond the scope of the present paper.



Supplementary Figure 9: Phonon band structure of MIL-53 (np) obtained from PBE forces showing imaginary phonon frequencies (here presented as negative frequencies) in particular at the X and Z points.

Supplementary Note 2. Training of the machine learned potentials

For the training of the machine learned potentials (except during the VASP active learning runs) the elements in the materials are further split into atom types depending on their neighborhood. For this, by utilizing a set of common covalent radii of the contained atoms, a list of immediate ('bonded') neighbors was created and atoms with an equivalent list were assigned an atom type each. For the specific types used, refer to Supplementary Table 6.

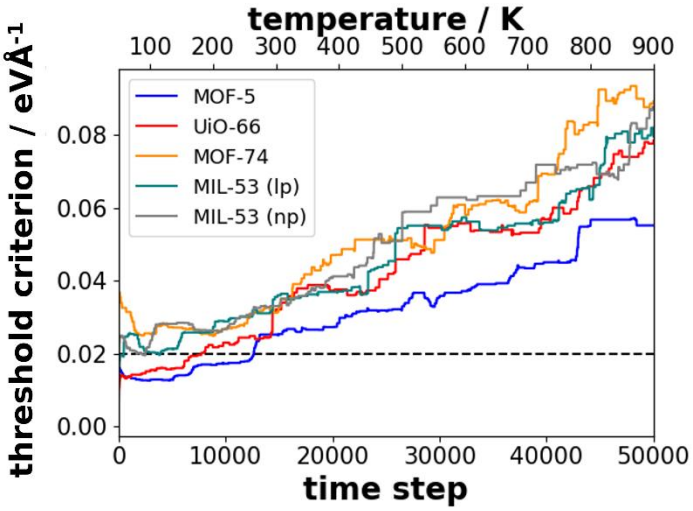
Supplementary Table 6: Atom types used for the machine learned potentials for each system. The types are given in a form, where the first symbol indicates the atom type element and the subsequent number specifies the number of atoms in its immediate neighborhood. After the underscore a list of element symbols is given fo*llowed by how many atoms of this element are in the base atom's immediate neighborhood.*

System name	Metal types	O types	C types	H types
MOF-5	Zn4_O4	O4_Zn4	C3_C1O2	H1_C1
		O2_C1Zn1	C3_C2H1	
			C3_C3	
UiO-66	Zr8_O8	O4_H1Zr3	C3_C1O2	H1_C1
		O3_Zr3	C3_C2H1	H1_O1
		O2_C1Zr1	C3_C3	
MOF-74	Zn4_O4	O3_C1Zn2	C3_C1O2	H1_C1
		O2_C1Zn1	C3_C2O1	
			C3_C2H1	
			C3_C3	
MIL-53	Al6_O6	O2_Al1C1	C3_C1O2	H1_C1
		O3_Al2H1	C3_C2H1	H1_O1
			C3_C3	

Supplementary Note 2.1 Performing the learning with VASP

As mentioned in the main manuscript, for the initial training runs the force threshold criterion during active learning was allowed to adjust dynamically over the course of the simulation based on the average Bayesian error over the previous 10 training steps. In Supplementary Figure 10 the value of this threshold criterion is shown as a function of time step. As can be seen, the value starts between 0.01-0.03 $eVÅ^{-1}$ and shows a steady increase over time (and, thus, temperature) for all systems. This is due to the larger displacements present at higher temperatures, which also show larger absolute force values making it more difficult to maintain the same absolute errors. It is also notable, that MOF-5 shows a substantially lower threshold criterion than all the other systems. A possible explanation for this is that MOF-5 is the most simple system considered here, without displaying a pronounced anisotropy like MOF-74 or MIL-53 and also with a smaller inorganic node than UiO-66. Here it should be noted that compared to the situation in MOF5, the larger node of UiO-66 contains more types of

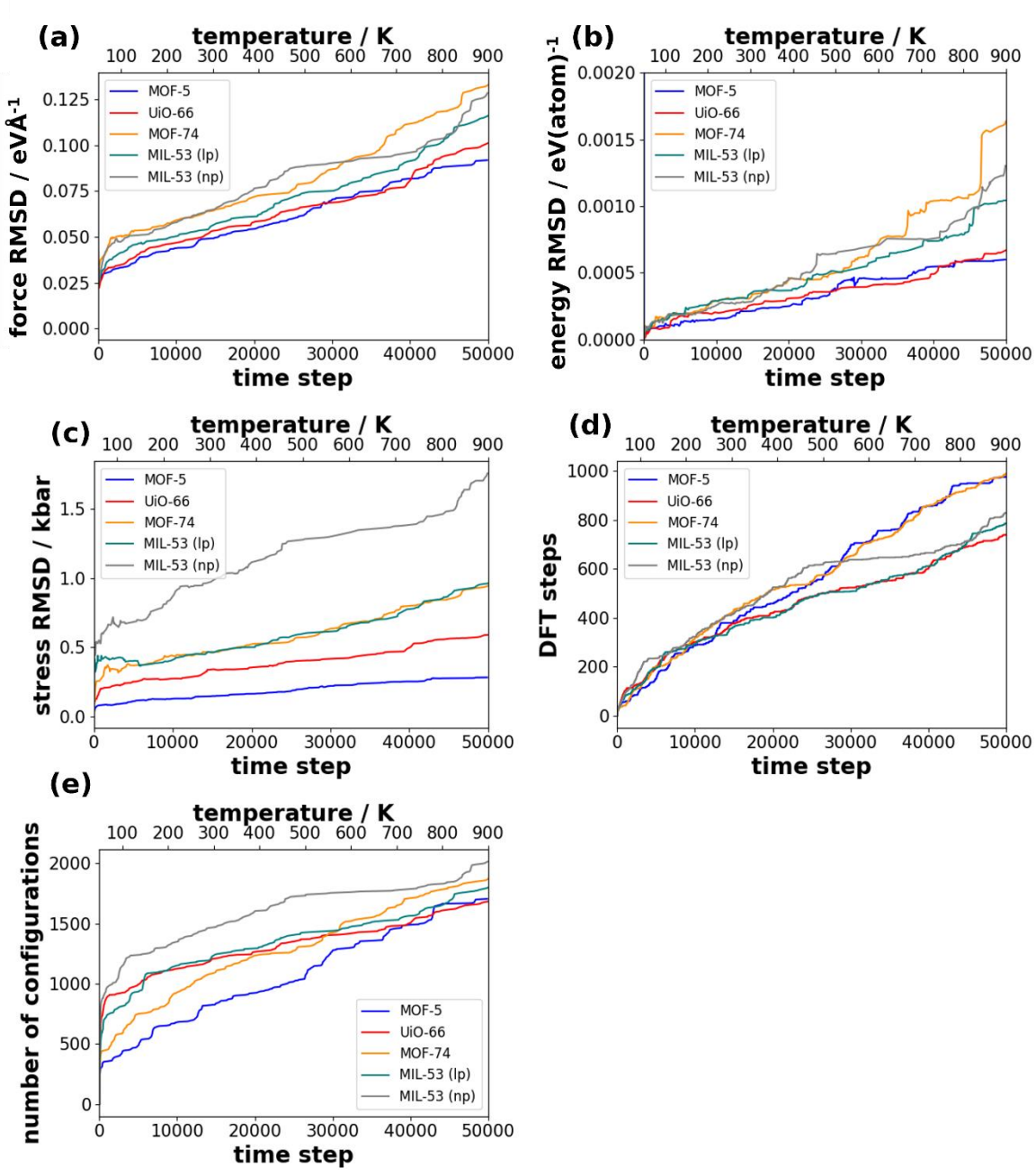
interactions that have to be modeled accurately. At the target temperature of 300 K, we see that the threshold criterion reaches values between 0.025 eVÅ^{-1} and 0.035 eVÅ^{-1} which is the main motivation why the fixed criterion was set to 0.02 eVÅ^{-1} for the extended training runs. A lower criterion has to be chosen to add additional reference configuration to the machine learned potential, but it should also not be chosen too low, such that the MD run would be performed nearly exclusively with DFT.



Supplementary Figure 10: Evolution of the dynamically adjusted threshold criterion during the initial VASP active learning runs for each investigated systems as a function of time step. The current system temperature is indicated which results from the heating during training. The black dashed line indicates the fixed threshold criterion for the extended training runs. In the course of the run, the temperature of the system has been increased steadily from 50 to 900 K.

Further details regarding the initial training procedure for the investigated systems are shown in Supplementary Figure 11. In panels a,b, and c the evolutions of the force/energy/stress RMSDs are shown as a function of time step. It can be seen that similar to the increase in threshold criterion, the RMSDs rise as more reference configurations are added at higher temperatures. While the trends are similar, for energy and forces, the anisotropic systems show a higher error than the isotropic MOF-5 and UiO-66. For the stresses more pronounced differences between the systems can be observed, which is largely a result of differences in the absolute values of the stresses experienced by the various systems. In Supplementary Figure 11 d the cumulative number of time steps computed with DFT is shown. All systems show a strong increase in added reference data in the early stage of the training compared to the later stages. Notably, for MOF-5 and MOF-74 the rate at which new reference structures are added at higher temperatures is larger than in the other systems. In Supplementary Figure 11 e the number of local reference configurations (basis function) used by the VASP MLP is shown. A large portion of the configurations is added immediately after the start of the training to

achieve a reasonable initial force field potential. Afterwards, we can observe a steady increase of added configurations for all systems. The different systems show a different number of configurations that were added during the initial training phase, where MOF-5 and MOF-74 show the lowest initial number. However, at the end of the run all systems show a relatively similar number of total local reference configurations.



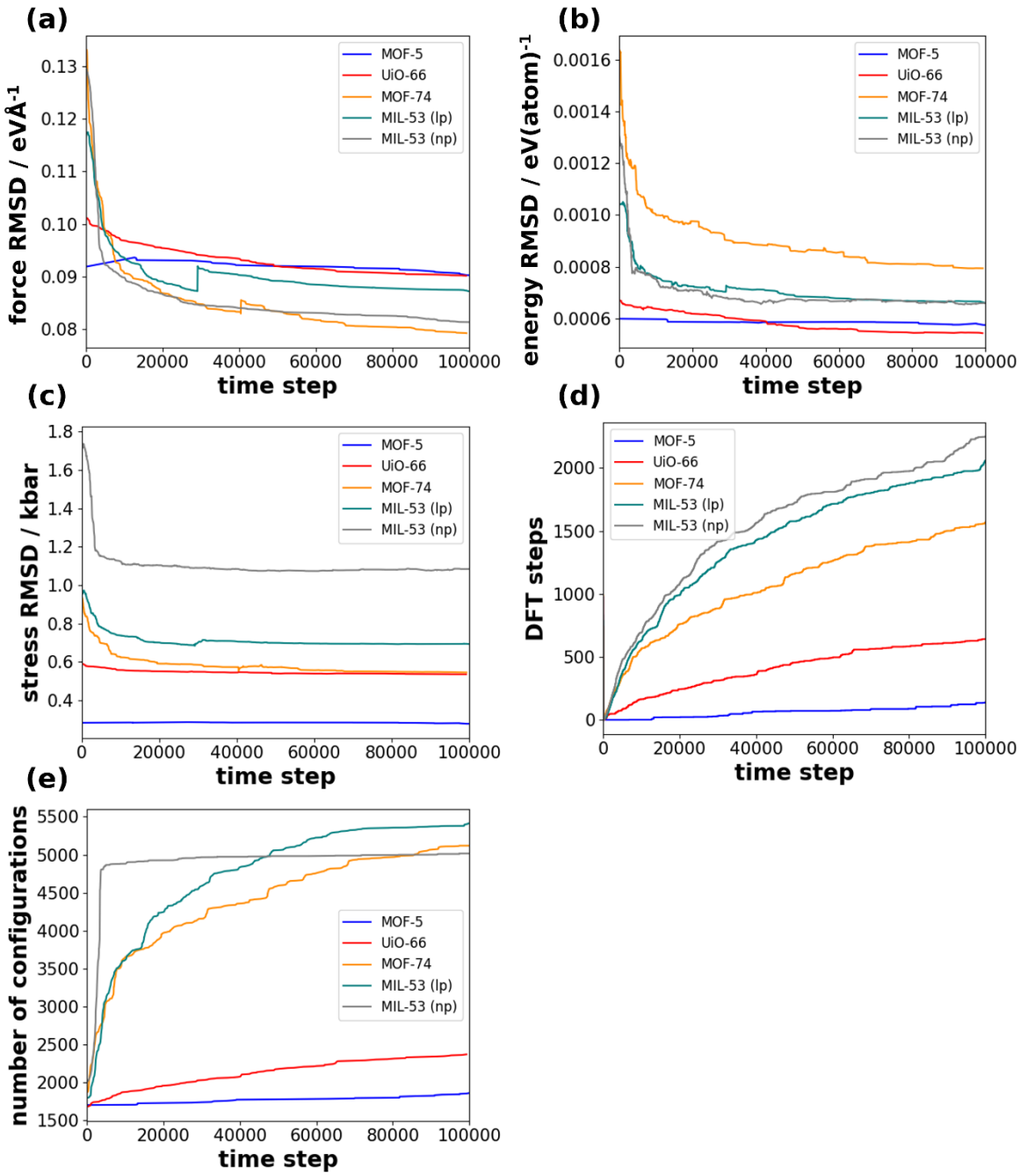
Supplementary Figure 11: Overview of the progress during the initial training runs for each of the investigated systems. The VASP MLP RMSDs for forces a, energies b and stresses c are shown as a function of simulation time step based on the reference data generated up to that point. In panel d the cumulative number of DFT steps computed over the course of the training is shown and panel e shows

the total number of local reference configurations (basis function) for the concurrent VASP MLP in the training run.

Supplementary Figure 12 is the analogue to Supplementary Figure 11 for the extended training run, where the threshold criterion was fixed at 0.02 eVÅ^{-1} and the temperature was kept constant at 300 K. For the force/energy/stress RMSDs in panels a, b, and c we see an immediate decrease of the errors followed by a slower drop for all systems except for MOF-5, where barely any changes can be seen. For UiO-66 the effect is also less pronounced. For MOF-74 and MIL-53 (np) a discontinuity can be observed in the RMSD values (especially in the forces), which happened when a restart of the training was initiated to increase the total number of reference configuration. For UiO-66, where a restart was also performed after 62,152 time steps, such a step cannot be discerned. It is also interesting that for MIL-53 and MOF-74 the force RMSD for the training set is actually lower than for UiO-66 and MOF-5 after the extended training set.

For the number of DFT steps in Supplementary Figure 12 d we now see large differences between the systems, with almost no new DFT steps performed for MOF-5, while especially MIL-53 sees a substantial increase in added reference data. For the local reference configurations in Supplementary Figure 12 (e) we see a strong immediate increase for MIL-53 and MOF-74, while UiO-66 shows a much weaker effect. Especially the narrow pore phase of MIL-53 appends a large number of reference configurations immediately. This correlates with the most substantial reduction in the RMSD values across all systems. The reason why proportionally fewer additional reference configurations were added in the later stage of the simulation is because the maximum number of reference configurations per atomic species was set to 3000. This limit was reached for the hydrogen atoms meaning that existing configurations were replaced with newer ones instead of merely appending new ones to the potential. Such a limit was initially imposed because it is known that lighter atom types, like hydrogen, tend to accumulate more configurations than required.

It should be noted, that for materials that undergo large deformations of the lattice while heating, it could be required to perform the heating over several individual simulations with restarts in between. This serves to prevent basis set errors as VASP only builds the basis at the beginning of a molecular dynamics run. In the investigated materials, the deformations of the cells during the MD runs were minor and basis set errors are less significant. This is not only a consequence of the limited extend of cell deformation over time but is also due to choosing a relatively high plane-wave energy cutoff.



Supplementary Figure 12: Overview of the progress during the extended training runs for each of the investigated systems. The RMSDs for forces a, energies b, and stresses c are shown as a function of the simulation time step for the reference data generated up to that point. In panel d the cumulative number of DFT steps computed over the course of the training is shown. Panel e shows the total number of local reference configurations (basis function) for the concurrent VASP MLP in the training run.

After the reference data had been obtained during the active learning run, new potentials were trained using a newer VASP version 6.4.1 (see methods section in main manuscript for details). This

was done for the initial reference data set with and without including separation of atom types as described in the beginning of this Note. For the extended reference data set, the training was for the most part only performed without further separation of atom types. An exception to this is MOF-74, where we also investigated the atom separated extended VASP MLPs alongside with many other settings, which are detailed in Supplementary Note 3.7. To discuss the impact of the retraining on the basis sets, Supplementary Table 7 lists the number of local reference configuration for each of the potentials presented in this work. There, it can clearly be seen that the number of local reference configurations increased substantially after retraining with VASP version 6.4.1. Especially the number of configurations for the metal atoms was very low for most of the systems after active learning. This means, that potentials retrained using a new VASP version, should improve in accuracy. Note, that the maximum number of local reference configurations for each atom type was set to 1500 during the retraining. In Fig. 7 in the main manuscript, we already discussed based on the example of MOF-74 that increasing this limit does only have a minor impact on the accuracy. And since the computational cost increases when including a larger number of reference configurations, the limit was left at 1500 for the VASP MLPs we typically show throughout this work.

Supplementary Table 7: Number of local reference configuration for each atom type of VASP MLPs from the active learning runs (with 6.3.0) and those retrained with VASP version 6.4.1 trained on the initial (init.) and extended (ext.) reference data sets. For VASP MLPS including separation of atom types, multiple lines are given.

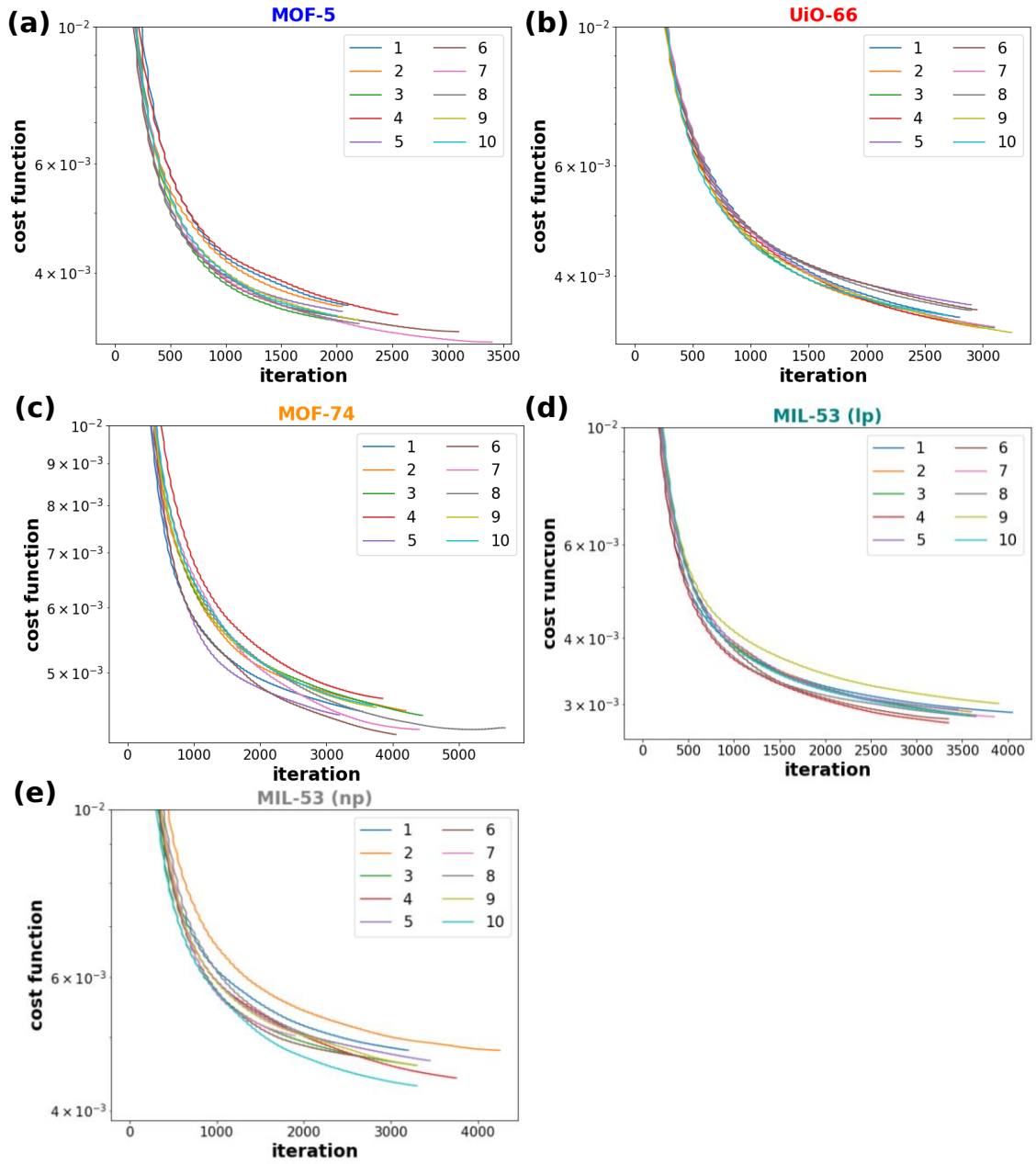
System	Training type	Metal conf.	O conf.	C conf.	H conf.	Total conf.
MOF-5	6.3.0	146	1141	378	43	1714.3
	6.4.1	764	1500	1352	319	3941.5
	6.4.1 7 types	1287	289	1317	289	5060
			1500	269		
				109		
	6.3.0 ext.	146	1287	387	43	1863
	6.4.1 ext.	783	1500	1406	333	4022
UiO-66	6.3.0	80	633	291	684	1694.3
	6.4.1	181	1500	850	1033	3570.5
	6.4.1 7 types	840	770	176	544	6133
			544	236	562	
			1500	961		
	6.3.0 ext.	81	732	293	1271	2377
	6.4.1 ext.	240	1500	1253	1500	4493
MOF-74	6.3.0	54	543	342	938	1883.3
	6.4.1	525	1500	1500	1305	4836.5
	6.4.1 7 types	830	1500	1156	1425	8942
			883	954		
				954		
				1240		
	6.3.0 ext.	54	1103	962	3000	5119
	6.4.1 ext.	737	1500	1500	1500	5237
	6.4.1 ext. 7 types	1047	1500	1500	1500	11350
			1303	1500		
				1500		
				1500		
MIL-53 (lp)	6.3.0	79	592	235	894	1806.3
	6.4.1	313	1500	837	1500	4156.5
	6.4.1 7 types	346	597	659	1235	6038
			1500	264	788	
				649		
	6.3.0 ext.	306	1855	253	3000	5414
	6.4.1 ext.	446	1500	1478	1500	4924
MIL-53 (np)	6.3.0	44	517	360	1097	2024.3
	6.4.1	257	1500	1500	1500	4763.5
	6.4.1 7 types	276	1500	1454	1500	8638
			400	1500	609	
				1399		
	6.3.0 ext.	121	1263	631	3000	5015
	6.4.1 ext.	545	1500	1500	1500	5045

Supplementary Note 2.2 Training the MTPs with MLIP

Since the starting parameters for the MTP training are initialized randomly, the trained potentials are also going to be slightly different. Supplementary Figure 13 shows the convergence of the training procedures for each system based on the initial training set. It is evident that all systems show a certain degree of variance in the value of the target function. Additionally, convergence is achieved only after several thousand fit iterations. In the shown cases, each fit took several hours to compute on one node of the VSC5 (see the main manuscript or https://vsc.ac.at//systems/vsc-5/ for architectural details of the supercomputer). This already indicates why using MTPs directly in an on-the-fly training approach is quite inconvenient. In fact, even when reducing the amount of time required for the fits by using a more sophisticated guess for the parameters (like the force field obtained in previous retraining iterations), they are still much more time consuming than the regression performed for the VASP MLPs, where in the later stages of an active learning run usually only 1-3 steps are required (and even there, the extra steps are only performed for the evidence approximation [18] to prevent overfitting). In a preliminary test of the active learning procedure implemented in MLIP for MOF-74, the performed fits still took 100-300 BFGS iterations at a late stage of a learning run. Note, that for systems with fewer atom types, this behavior should be significantly better. The objective function in Supplementary Figure 13 is somewhat arbitrarily constructed out of differences in forces/stresses and energies, so that it should be more telling to investigate the differences based on the chosen benchmark criteria. A numerical comparison of the final cost function values, and comparisons based on the Γ phonon frequencies, the elastic tensor, the forces and energies of the validation set as well as a maximum temperature for which the respective potential was stable, can be seen in Supplementary Table 8. The thermal stability was evaluated by performing molecular dynamics simulations up to a temperature of 700 K in 100 K steps for 100 ps. If the run was concluded, the MTP was considered to be ‘stable’ at that temperature. Please note, that this is not a definitive criterion as unphysical behavior can occur at random when the simulation reaches regions not as well sampled by the reference data set. That means, that the testing duration might not have been sufficient to ensure thermal stability at that temperature with an absolute certainty. It is evident, that even though some reference data up to 900 K were included in the training runs, none of the trained potential is stable up to even 700 K. This is the reason why such a high initial target temperature was aimed for, even though the potentials are intended to be used at room temperature. However, despite this, some of the systems, especially MIL-53, features a fair number of trained potentials that are not even stable up to room temperature. This can be remedied by adding additional training data to the reference data set at slightly higher than the target temperature as will be shown later in this Note. It is also apparent, that the thermal stability varies wildly with the differently trained MTPs and the maximum

temperature and lowest errors for the other criteria does not always correlate with the lowest cost function during training which makes choosing the 'best' potential somewhat challenging. Ideally more reference data can be added to make the potential more reliable, but this can be expensive depending on the system and the overall accuracy of these MTPs is already excellent for many applications. Alternatively, simulations in general can be performed at a reduced time step which also enhances the stability of the machine learned potentials. Such issues are the main drawback of machine learned approaches compared to conventional bonded force field potentials, which usually have well defined potential wells. This constitutes an area within the MTPs that can still be improved upon.

Aside of the thermal stability, Supplementary Table 8 shows that in some relatively rare cases poor potentials can emerge that show very high errors in the phonons or other properties. However, most of the other cases show relatively similar error values in all criteria. Indicated in bold in Supplementary Table 8 are those potentials, which are used for the comparisons including the initial reference data set MTPs in this work. This was done based on a low cost function during training and a thermal stability up to at least 300 K.



Supplementary Figure 13: Cost functions below 10^{-2} for multiple MLIP training runs based on the initial reference data set as a function of the numbers of iterations until convergence Has been achieved: a MOF-5, b UiO-66, c MOF-74, d MIL-53 (lp), and e MIL-53 (np). The individual training runs are each assigned a number to ease identification.

Supplementary Table 8: Variation of errors for multiple MLIP training runs performed for the initial training. The final values of the cost function, the RMSDs of the Γ-frequencies, the RMSDs of the elastic constants, and the RMSDs of the forces and energies based on the validation are listed. Also, the temperatures up to which the potentials were stable in the NPT simulations are provided. The potentials used for the remainder of the evaluation and for in-depth discussions in the main manuscript and in other chapters in the Supplementary Information are indicated in bold for each system.

System name	cost fct.	Γ freq. RMSD / cm^{-1}	$C_{i,j}$ RMSD / GPa	Force RMSD / $eVÅ^{-1}$	Energy RMSD / $meV(atom)^{-1}$	Stable until / K
MOF-5	0.00354	3.5	0.03	0.017	0.08	500
	0.00353	3.8	0.10	0.017	0.10	600
	0.00333	17.1	2.24	0.045	0.62	400
	0.00342	6.9	0.04	0.024	0.21	400
	0.00347	3.7	0.07	0.017	0.08	400
	0.00321	5.7	0.05	0.021	0.15	400
	0.00309	**3.3**	**0.02**	**0.016**	**0.07**	**500**
	0.00331	4.1	0.12	0.018	0.11	500
	0.00335	3.6	0.08	0.017	0.08	400
	0.00341	3.5	0.12	0.017	0.07	400
UiO-66	0.00344	4.4	0.20	0.017	0.07	300
	0.00330	3.9	0.22	0.016	0.05	200
	0.00331	3.7	0.13	0.016	0.07	300
	0.00333	**3.7**	**0.18**	**0.016**	**0.07**	**400**
	0.00360	4.7	0.30	0.017	0.06	200
	0.00353	4.3	0.29	0.017	0.08	300
	0.00332	3.6	0.13	0.016	0.08	300
	0.00352	3.4	0.29	0.017	0.09	400
	0.00325	3.8	0.11	0.016	0.06	300
	0.00346	3.6	0.25	0.017	0.07	300
MOF-74	0.00449	3.0	0.61	0.031	0.20	400
	0.00450	3.3	0.74	0.030	0.24	200
	0.00444	3.7	1.33	0.031	0.27	500
	0.00465	5.0	0.68	0.037	0.33	300
	0.00444	3.1	0.51	0.030	0.19	200
	0.00421	3.4	0.53	0.029	0.18	300
	0.00427	4.2	0.55	0.032	0.25	200
	0.00428	**3.1**	**0.30**	**0.030**	**0.17**	**600**
	0.00454	3.2	0.77	0.031	0.28	300
	0.00457	16.7	3.50	0.096	1.30	500
MIL-53 (lp)	0.00290	24.1	3.14	0.052	0.40	0
	0.00291	5.8	1.11	0.022	0.14	400
	0.00287	4.9	0.30	0.021	0.16	300
	0.00277	5.9	0.52	0.024	0.18	200
	0.00293	5.8	0.96	0.022	0.17	200
	0.00282	6.4	0.67	0.023	0.16	400
	0.00285	5.5	0.48	0.021	0.21	200
	0.00285	5.3	0.50	0.022	0.16	300
	0.00301	5.5	1.04	0.026	0.30	200
	0.00287	**5.2**	**0.52**	**0.022**	**0.17**	**300**
MIL-53 (np)	0.00481	9.5	0.47	0.029	0.12	100
	0.00480	9.4	1.10	0.033	0.24	100
	0.00459	7.3	0.70	0.028	0.13	300
	0.00442	7.3	0.94	0.027	0.14	100
	0.00466	9.8	0.58	0.028	0.15	400
	0.00471	6.6	0.75	0.029	0.15	200
	0.00503	7.6	0.77	0.030	0.15	600
	0.00492	7.0	0.64	0.029	0.17	400
	0.00459	**7.7**	**0.48**	**0.028**	**0.14**	**300**
	0.00432	8.4	0.73	0.028	0.12	200

A similar selection was performed for the extended reference data set MTP, as can be seen in Supplementary Table 9. Here, now no really 'bad' potentials with exceptionally poor properties are present any more and overall the larger reference data set leads to more consistent and usually

slightly lower error values among the different benchmark properties. Only the thermal stability is still quite inconsistent, especially for MIL-53 (lp), where the initial 5 performed training runs showed a quite poor stability. To investigate, whether this is simply a coincidence, 3 more MTP training runs with the same settings and different initializations were performed. All of them show an improved thermal stability. The randomness of the stability is somewhat concerning and even the extended reference data set did only slightly improve the situation for most systems. However, the extended reference data set was obtained for a constant temperature of 300 K and mostly served to analyze the improvements in accuracy at that particular temperature. To ensure the stability at a given temperature, one might expect that training the potential based on data generated at somewhat higher temperatures would be preferrable to sample also more strongly distorted structures during the training. To test this, we, in particular, generated 1009 additional reference configurations at a constant temperature of 400 K for MIL-53 (lp). The training settings other than the temperature and the initial potential (which was now including the initial and extended reference data sets) were the same as the regular extended reference data set. The results can be seen in Supplementary Table 10. All 5 of the trained MTPs on that reference data set were stable up to a temperature of 300 K. So, there definitely is an improvement. However, due to the variance involved in the stability issues and the relatively small sample size, we cannot be certain that such a training at 400K solves the issue entirely.

In general, the stability issues seem to be the primary weakness of the machine learned potentials considered here (they are also present for the VASP MLPs, as we discuss in Supplementary Note 4.1.2). In the interest of saving computation time, it would be beneficial to enhance the MTPs with some additional physical knowledge to make it less likely that the potential diverges in the unknown regime. Alternatively, different data sampling approaches could be used to reduce the total number of reference structures or one might consider adding auxiliary ‘solid-wall’ type potentials to prevent a decomposition of the studied systems. However, this would compromise the possibility to model adsorption or desorption processes or chemical reactions when using the potentials.

Supplementary Table 9: Variation of errors for multiple MLIP training runs performed for the extended training. The final values of the cost function, the RMSDs of the Γ-frequencies, the RMSDs of the elastic constants, and the RMSDs of the forces and energies based on the validation are listed. Also, the temperatures up to which the potentials were stable in the NPT simulations are provided. The potentials used for the remainder of the evaluation and for in-depth discussions in the main manuscript and in other chapters in the Supplementary Information, whenever data for the extended reference data set are shown, are indicated in bold for each system.

System name	cost fct.	Γ freq. RMSD / cm^{-1}	$C_{i,j}$ RMSD / GPa	Force RMSD / $eVÅ^{-1}$	Energy RMSD / meV(atom)$^{-1}$	Stable until / K
MOF-5	0.00359	4.8	0.05	0.020	0.16	500
	0.00347	4.0	0.08	0.018	0.08	600
	0.00497	4.3	0.16	0.020	0.11	400
	0.00349	6.4	0.02	0.023	0.16	500
	0.00318	**3.7**	**0.09**	**0.017**	**0.08**	**500**
UiO-66	**0.00289**	**2.8**	**0.32**	**0.016**	**0.05**	**400**
	0.00298	3.2	0.13	0.016	0.06	400
	0.00293	3.8	0.21	0.016	0.08	400
	0.00305	3.7	0.25	0.016	0.06	200
	0.00281	3.6	0.19	0.016	0.08	300
MOF-74	0.00356	2.8	0.54	0.028	0.21	500
	0.00328	2.8	0.43	0.028	0.17	300
	0.00337	2.9	0.62	0.028	0.20	400
	0.00355	**2.9**	**0.21**	**0.028**	**0.20**	**700**
	0.00356	2.9	0.61	0.028	0.15	700
MIL-53 (lp)	0.00242	5.1	1.36	0.024	0.17	200
	0.00239	5.3	1.69	0.022	0.19	200
	0.00226	5.0	0.62	0.020	0.12	100
	0.00255	4.6	0.50	0.020	0.14	100
	0.00254	**4.1**	**0.56**	**0.021**	**0.16**	**500**
	0.00247	4.6	1.14	0.021	0.15	400
	0.00216	5.2	0.46	0.019	0.13	300
	0.00243	5.6	1.40	0.022	0.22	400
MIL-53 (np)	**0.00422**	**6.4**	**0.41**	**0.027**	**0.14**	**500**
	0.00407	6.8	0.35	0.027	0.15	100
	0.00419	7.8	0.40	0.028	0.16	400
	0.00413	7.6	0.53	0.027	0.15	500
	0.00396	7.2	0.59	0.027	0.15	400

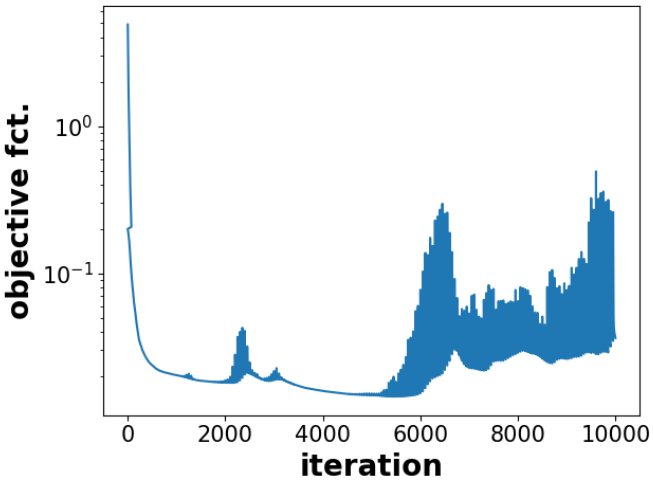
Supplementary Table 10: Variation of errors for multiple MLIP training runs performed for the additionally extended reference data set based on 400 K MD reference data for MIL-53 (lp). The final values of the cost function, the RMSDs of the Γ-frequencies, the RMSDs of the elastic constants, and the RMSDs of the forces and energies based on the validation are listed. Also, the temperatures up to which the potentials were stable in the NPT simulations are provided.

System name	cost fct.	Γ freq. RMSD / cm^{-1}	$C_{i,j}$ RMSD / GPa	Force RMSD / $eVÅ^{-1}$	Energy RMSD / $meV(atom)^{-1}$	Stable until / K
MIL-53 (lp)	0.00271	5.11	1.62	0.021	0.19	400
	0.00257	4.57	1.42	0.020	0.17	300
	0.00281	5.51	1.05	0.021	0.17	400
	0.00275	5.13	1.52	0.021	0.17	400
	0.00271	4.98	1.40	0.021	0.17	600

For the training for some of the systems, especially at lower MTP levels, we also encountered the problem that the training algorithm escaped a previously found minimum without finding a better minimum in a reasonable time frame. This issue is visualized in Supplementary Figure 14 for the example of MOF-5 at level 10 and is the reason why the convergence of the cost function needs to be observed carefully during the training. To solve the issue, the MTP for the minimum cost function can be chosen from the intermediate results produced during training However, as it stands, MLIP does not support outputting all intermediate training results, so it is up to the user to track and store away the output files. This can easily be realized by running a script as a background process during training. As an example, we provide the used bash script in the following:

```
#!/bin/bash
touch $1
date=$(stat -c %y "$1")
while sleep 0.01;
do
  date2=$(stat -c %y "$1")
  if [[ $date2 != $date ]];
  then
    cat $1 >> "all_$1"
    echo -e "\n" >> "all_$1"
  fi
  date=$date2
done
```

It takes one argument, which is the file name of the concurrent potential output by MLIP when using the '`--curr_pot_name=<file_name>`' option. Note, that for very fast fits for small systems it is necessary to adjust the time for the sleep command. Then, all intermediate results are written in the file all_<file_name> from which the intermediate result with the lowest cost function can be extracted.



Supplementary Figure 14: Example for the unsatisfactory convergence behavior of the cost function for training a level 10 MTP on the initial reference data set of MOF-5. The training was aborted after 10000 time steps.

Supplementary Note 3. Analysis of the potentials

Supplementary Note 3.1 Full evaluation of the errors

To support the results given in Fig. 3, 5 and 7 in the main manuscript, we show here a quantitative analysis of the errors. Supplementary Table 11 summarizes the root mean square errors in the energy-, force-, stress-, and Γ-point frequency calculations together with the root mean square errors of the elastic constants for all studied systems and for the initial and the extended training sets. Values are provided for the VASP MLPs and for the MTPs.

Supplementary Table 11: RMSD values of the VASP MLPs and MTPs trained on the initial and extended reference data set with the DFT reference for various quantities: the phonon frequencies at the Γ-point; the components of the elastic stiffness tensor, the components of the forces in the validation set; the energy differences between individual total energies of the structures in the validation set minus the total energy of the minimum structure; and the components of the stress tensor for structures in the validation set. The VASP MLPs trained on the initial reference data set are given for the case of including separation of atom types based on their neighborhood and for only 4 atom types as indicated by n_{types} (which denotes the number of atom types).

System	Method	n_{types}	f_Γ RMSD / cm^{-1}	$C_{i,j}$ RMSD / GPa	Force RMSD / $eVÅ^{-1}$	Energy RMSD / meV(atom) $^{-1}$	Stress RMSD / kbar
MOF-5	VASP MLP	4	8.00	0.20	0.024	0.15	0.05
	VASP MLP	7	8.96	0.27	0.021	0.07	0.05
	MTP	7	3.26	0.02	0.016	0.07	0.02
	VASP MLP extended	4	7.36	0.21	0.022	0.10	0.05
	MTP extended	7	3.70	0.09	0.017	0.01	0.02
UiO-66	VASP MLP	4	11.85	1.16	0.033	0.13	0.20
	VASP MLP	9	10.35	0.60	0.024	0.11	0.13
	MTP	9	3.71	0.18	0.016	0.07	0.06
	VASP MLP extended	4	7.82	0.29	0.025	0.11	0.15
	MTP extended	9	2.78	0.32	0.016	0.06	0.05
MOF-74	VASP MLP	4	5.67	1.26	0.037	0.25	0.24
	VASP MLP	8	4.72	0.62	0.027	0.20	0.17
	MTP	8	3.07	0.30	0.030	0.17	0.11
	VASP MLP extended	4	4.81	0.31	0.034	0.22	0.19
	VASP MLP extended	8	2.89	0.39	0.022	0.14	0.12
	MTP extended	8	2.91	0.21	0.028	0.20	0.09
MIL-53 (lp)	VASP MLP	4	12.51	0.91	0.033	0.28	0.27
	VASP MLP	8	8.83	0.78	0.027	0.19	0.19
	MTP	8	5.23	0.52	0.022	0.17	0.09
	VASP MLP extended	4	7.35	0.95	0.027	0.21	0.19
	MTP extended	8	4.98	1.40	0.021	0.17	0.06
MIL-53 (np)	VASP MLP	4	11.40	1.87	0.040	0.19	0.47
	VASP MLP	8	8.70	0.67	0.030	0.17	0.32
	MTP	8	7.64	0.48	0.028	0.14	0.24
	VASP MLP extended	4	10.78	1.53	0.038	0.21	0.44
	MTP extended	8	6.41	0.41	0.027	0.14	0.15

Supplementary Note 3.1.1 Evaluation of lattice parameters

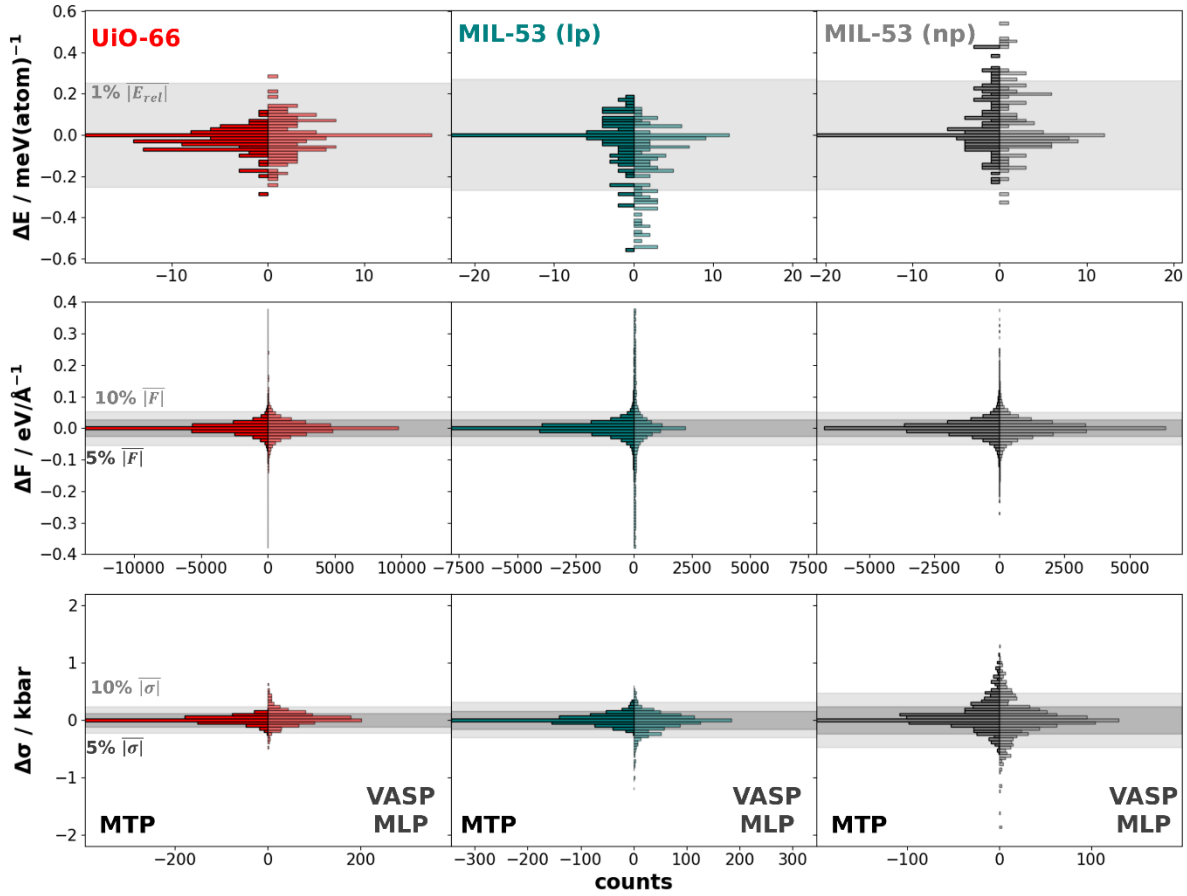
As we only provide volume deviations for the crystallographic unit cells in the main paper, all parameters of the unit cells calculated with the different methos for all systems are shown in Supplementary Table 12.

Supplementary Table 12: Optimized Lattice parameters for the investigated systems obtained for the various force field potentials and for the DFT reference. The angles not specified in the table are 90°. The number of atom types used to perform the calculations is indicated with n_{types}.

System	Method	n_{types}	a / Å	b / Å	c / Å	γ / °	V / Å³
MOF-5	MTP	7	26.066	26.066	26.066	90.00	17710.188
	MTP extended	7	26.065	26.065	26.065	90.00	17708.150
	VASP MLP	4	26.069	26.069	26.069	90.00	17715.051
	VASP MLP	7	26.068	26.068	26.068	90.00	17714.397
	VASP MLP extended	4	26.069	26.069	26.069	90.00	17716.105
	DFT	4	26.068	26.068	26.068	90.00	17714.265
	UFF4MOF	5	26.442	26.442	26.442	90.00	18487.701
UiO-66	MTP	9	20.898	20.898	20.898	90.00	9126.708
	MTP extended	9	20.898	20.898	20.898	90.00	9126.708
	VASP MLP	4	20.894	20.894	20.894	90.00	9121.881
	VASP MLP	9	20.898	20.898	20.898	90.00	9127.235
	VASP MLP extended	4	20.899	20.899	20.899	90.00	9128.174
	DFT	4	20.899	20.899	20.899	90.00	9128.019
MOF-74	MTP	8	26.138	26.138	6.704	120.00	3966.517
	MTP extended	8	26.139	26.139	6.704	120.00	3966.821
	VASP MLP	4	26.152	26.152	6.694	120.00	3965.438
	VASP MLP	8	26.136	26.136	6.690	120.00	3953.436
	VASP MLP extended	4	26.139	26.139	6.691	120.00	3959.878
	VASP MLP extended	8	26.133	26.133	6.689	120.00	3955.691
	DFT	4	26.078	26.078	6.718	120.00	3956.573
	UFF4MOF	4	26.293	26.351	6.798	120.01	4078.545
MIL-53 (lp)	MTP	8	17.206	6.666	12.360	91.04	1417.399
	MTP extended	8	17.308	6.670	12.217	90.92	1410.202
	VASP MLP	4	17.266	6.668	12.335	90.95	1417.025
	VASP MLP	8	17.310	6.689	12.250	90.16	1418.465
	VASP MLP extended	4	17.229	6.669	12.357	90.73	1419.833
	DFT	4	17.266	6.665	12.272	90.99	1412.025
MIL-53 (np)	MTP	8	19.818	6.664	6.547	104.37	837.592
	MTP extended	8	19.830	6.663	6.513	104.39	833.547
	VASP MLP	4	19.790	6.661	6.548	104.19	836.839
	VASP MLP	8	19.788	6.661	6.568	104.20	839.207
	VASP MLP extended	4	19.796	6.658	6.612	104.23	844.677
	DFT	4	19.783	6.659	6.561	104.16	838.052

Supplementary Note 3.1.2 Deviations of forces, energies and stresses based on the validation sets

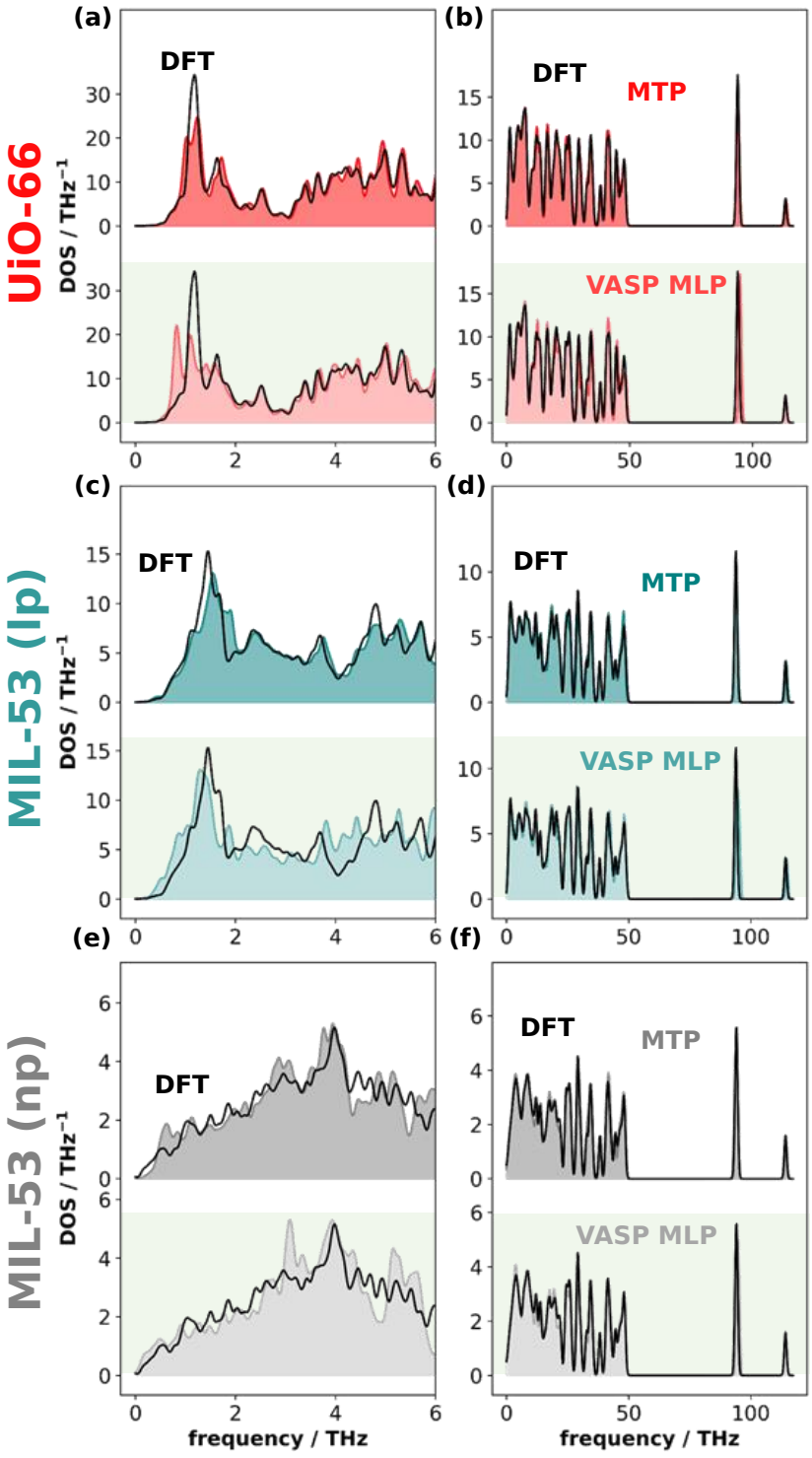
In the main manuscript, we focused on the force, energy and stress error evaluations for MOF-5 and MOF-74. In Supplementary Figure 15 the error distributions for the remaining systems can be seen for the VASP MLPs and the MTPs. It is evident that the deviations for all systems are in a similar range. Only for the stresses, we see larger deviations for the anisotropic systems, which is related to higher absolute values of the stresses in these systems.



Supplementary Figure 15: Histograms detailing the deviations (FF data subtracted from reference) of the energies (E), the components of the forces (F), and the components of the stresses (σ) obtained from DFT and from the VASP MLPs and the MTPs for UiO-66 (red), MIL-53 (lp) (teal) and MIL-53 (np) (grey). For the energies only the energy difference from the structure with the minimum energy is shown. The validation set structures were obtained from an active learning VASP MLP learning run at a temperature of 300 K. The areas shaded in gray 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.

Supplementary Note 3.1.3 Phonon densities of states

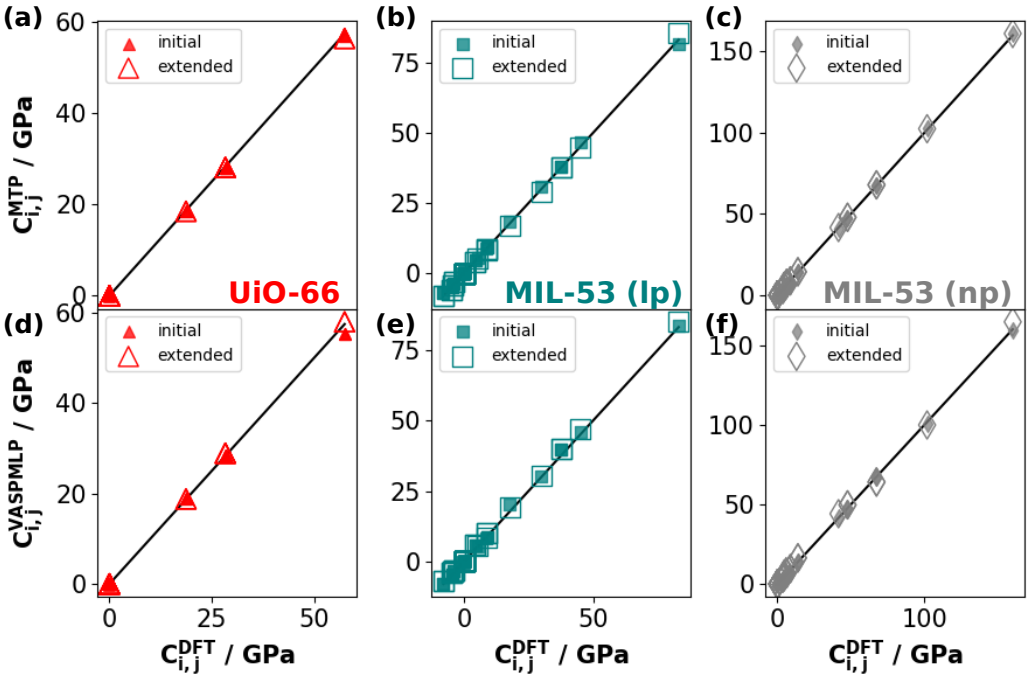
Similarly, for the phonon densities of states, we show the machine learned potential versus DFT comparison for UiO-66 and both phases of MIL-53 in Supplementary Figure 16. Like for MOF-5 and MOF-74, the agreement of the MTPs with the DFT data is excellent with only minor deviations, which are mostly noticeable in the zoom into the low frequency region. The VASP MLPs show a slightly worse agreement in the lower frequency region, but overall, the agreement is still excellent.



Supplementary Figure 16: Comparisons of the densities of states (DOS) obtained with DFT (indicated by black lines) and the machine learned force field potentials (shaded areas) for the low frequency phonon modes (a, c, e) and the entire frequency range (b, d, f). The comparisons are given for UiO-66 (a, b), MIL-53 (lp) (c, d) and MIL-53 (np) (e, f) for the MTP and the VASP MLP. The DOS was obtained for a 11x11x11 mesh sampled in the first Brillouin zone and a Gaussian smearing width of 0.05 THz was applied in (a, c, e) and of 0.5 THz in (b, d, f).

Supplementary Note 3.1.4 Elastic stiffness tensors

Supplementary Figure 17 shows a comparison of the elastic tensor elements obtained with DFT and the machine learned potentials for all systems. As can be seen, the agreement is generally good for both types of machine learned potentials for all system.



Supplementary Figure 17: Comparison of the elements of the elastic stiffness tensor. Shown are the results calculated with DFT and with the MTPs in a, b, c, with DFT and with the VASP MLPs in d, e, f as well as with DFT). The plots shown here contain data for UiO-66 in a, d, MIL-53 (lp) in b, e and MIL-53 (np) in c, f. 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 line pass through the origin and have a slope of 1 such that they indicate an ideal agreement between force field and DFT simulations.

The 21 elastic tensor elements for each system are given in Supplementary Table 13, Supplementary Table 14, and Supplementary Table 15, to allow a fully quantitative comparisons of the elastic tensor elements calculated with DFT and with the force fields . The data are split up in multiple tables due to space limitations.

In particular, for MIL-53 (np) several values are slightly non-zero for the MTPs while they are zero for PBE and the VASP MLPs. The reason for this is that for the MTPs the elastic tensor was computed

without applying any symmetry constraints, while for the VASP based calculations monoclinic symmetry was assumed (space groups Cc and C2/c).

Supplementary Table 13: 9 of the elements of the elastic tensor for each system evaluated by DFT, the VASP MLPs and the MTPs for the initial and the extended (ext.) reference data sets. The values of the tensor elements are given in GPa. Indices for C are given in Voigt notation which corresponds to the Cartesian directions as follows: 1=xx, 2=yy, 3=zz, 4=zy, 5=xz, 6=xy. The number of atom types in the used method is indicated with n_{type}.

System	Method	n_{type}	C_{11}	C_{22}	C_{33}	C_{44}	C_{55}	C_{66}	C_{12}	C_{13}	C_{23}
MOF-5	VASP MLP	4	25.39	25.39	25.39	0.94	0.94	0.94	10.95	10.95	10.95
	VASP MLP	7	25.35	25.34	25.35	0.87	0.87	0.87	10.73	10.73	10.73
	VASP MLP ext.	4	25.39	25.39	25.39	0.92	0.92	0.92	10.90	10.90	10.90
	MTP	7	25.99	25.99	25.99	0.95	0.95	0.95	11.25	11.25	11.25
	MTP ext.	7	25.70	25.70	25.70	0.89	0.89	0.89	11.31	11.31	11.31
	PBE	4	25.97	25.97	25.97	0.95	0.95	0.95	11.22	11.22	11.22
UiO-66	VASP MLP	4	53.89	53.89	53.89	18.35	18.35	18.35	27.11	27.11	27.11
	VASP MLP	9	55.42	55.42	55.42	18.84	18.84	18.84	28.34	28.34	28.34
	VASP MLP ext.	4	58.00	58.00	58.00	18.82	18.82	18.82	28.95	28.95	28.95
	MTP	9	56.99	56.99	56.99	18.66	18.66	18.66	28.10	28.10	28.10
	MTP ext.	9	56.43	56.43	56.43	18.52	18.52	18.52	28.14	28.14	28.14
	PBE	4	57.50	57.50	57.50	18.72	18.72	18.72	28.35	28.35	28.35
MOF-74	VASP MLP	4	27.48	27.48	13.93	16.53	16.53	2.19	23.10	3.65	3.65
	VASP MLP	8	23.92	23.92	10.58	16.26	16.26	2.25	19.42	2.68	2.68
	VASP MLP ext.	4	23.88	23.88	15.52	16.60	16.60	2.04	19.81	2.98	2.98
	MTP	8	24.67	24.68	14.18	16.35	16.34	2.35	19.92	3.72	3.72
	MTP ext.	8	23.82	23.85	13.05	16.19	16.19	2.29	19.22	2.97	2.98
	PBE	4	23.90	23.90	14.10	16.30	16.30	2.37	19.15	3.28	3.28
MIL-53 (lp)	VASP MLP	4	41.73	83.38	16.76	29.33	4.69	7.82	9.85	3.22	37.09
	VASP MLP	8	46.10	83.76	20.24	30.25	5.38	8.34	8.69	5.68	39.59
	VASP MLP ext.	4	47.06	85.07	19.23	30.24	5.57	8.41	10.02	5.84	39.77
	MTP	8	46.72	81.72	18.37	30.72	5.64	9.00	9.31	4.68	37.73
	MTP ext.	8	44.81	85.40	16.80	29.14	5.24	8.12	8.73	4.12	37.73
	PBE	4	45.02	83.20	17.62	29.95	5.27	8.71	8.92	4.34	37.57
MIL53 (np)	VASP MLP	4	167.04	7.73	100.76	3.44	42.92	5.58	9.93	51.63	7.70
	VASP MLP	8	159.50	6.71	100.21	2.88	41.63	3.41	7.15	47.51	5.47
	VASP MLP ext.	4	164.55	7.69	99.81	4.49	44.10	4.50	9.63	49.65	6.50
	MTP	8	161.01	5.41	102.47	3.96	40.83	4.49	7.36	47.63	5.71
	MTP ext.	8	161.02	6.04	102.37	3.48	41.49	4.89	8.68	47.79	7.14
	PBE	4	159.89	5.86	101.69	3.69	41.47	4.60	8.73	47.62	6.26

Supplementary Table 14: 6 of the elements of the elastic tensor for each system evaluated by DFT, the VASP MLPs and the MTPs for the initial and the extended (ext.) reference data sets. The values of the tensor elements are given in GPa. Indices for C are given in Voigt notation with 1=xx, 2=yy, 3=zz, 4=zy, 5=xz, 6=xy. Empty table entries indicate a value of zero. The number of atom types in the used method is indicated with n_{type}.

System	Method	n_{type}	C_{14}	C_{15}	C_{16}	C_{24}	C_{25}	C_{26}
MOF-74	VASP MLP	4	3.28	-0.58		-3.28	0.58	
	VASP MLP	8	3.26	-0.66		-3.26	0.66	
	VASP MLP ext.	4	3.24	-0.72		-3.24	0.72	
	MTP	8	3.27	-0.71		-3.25	0.71	
	MTP ext.	8	3.36	-0.69		-3.34	0.71	
	PBE	4	3.30	-0.75		-3.30	0.75	
MIL-53 (lp)	VASP MLP	4			-3.50			-6.18
	VASP MLP	8			-5.24			-7.91
	VASP MLP ext.	4			-4.14			-6.84
	MTP	8			-4.97			-6.96
	MTP ext.	8			-5.89			-7.99
	PBE	4			-4.59			-8.00
MIL-53 (np)	VASP MLP	4		63.69			5.62	
	VASP MLP	8		66.68			4.05	
	VASP MLP ext.	4		64.00			6.01	
	MTP	8	-0.10	67.32	-0.24	-0.02	4.79	-0.02
	MTP ext.	8	-0.20	67.81	-0.47	-0.04	5.22	-0.04
	PBE	4		67.29			4.54	

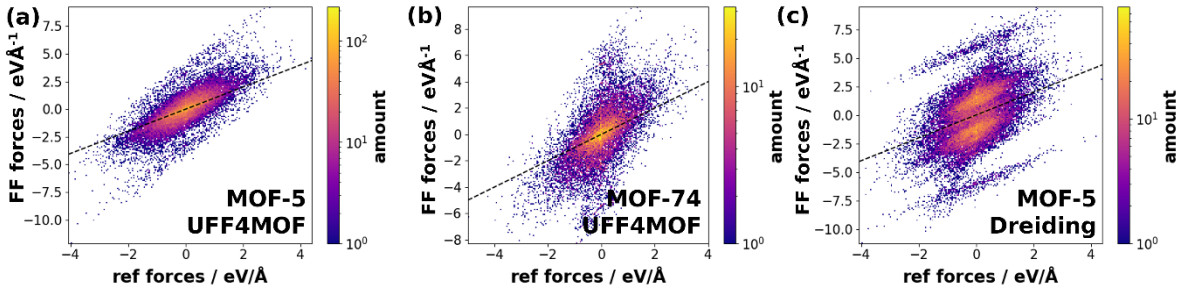
Supplementary Table 15: 6 of the elements of the elastic tensor for each system evaluated by DFT, the VASP MLPs and the MTPs for the initial and the extended (ext.) reference data sets. The values of the tensor elements are given in GPa. Indices for C are given in Voigt notation with 1=xx, 2=yy, 3=zz, 4=zy, 5=xz, 6=xy. Empty table entries indicate a value of zero. The number of atom types in the used method is indicated with n_{type}.

System	Method	n_{type}	C_{14}	C_{15}	C_{16}	C_{24}	C_{25}	C_{26}
MOF-74	VASP MLP	4					0.58	3.28
	VASP MLP	8					0.66	3.26
	VASP MLP ext.	4					0.72	3.24
	MTP	8					0.71	3.29
	MTP ext.	8					0.70	3.35
	PBE	4					0.75	3.30
MIL-53 (lp)	VASP MLP	4			-3.36	-3.80		
	VASP MLP	8			-4.62	-3.36		
	VASP MLP ext.	4			-3.71	-3.32		
	MTP	8			-4.24	-3.70		
	MTP ext.	8			-4.23	-3.22		
	PBE	4			-4.54	-3.93		
MIL-53 (np)	VASP MLP	4		14.13			3.72	
	VASP MLP	8		13.87			2.35	
	VASP MLP ext.	4		16.37			3.65	
	MTP	8	-0.03	13.96	-0.07	-0.06	2.86	-0.10
	MTP ext.	8	-0.05	14.22	-0.13	-0.13	3.46	-0.20
	PBE	4		13.88			3.66	

Supplementary Note 3.1.5 Further details regarding accuracy of the forces in the calculations relying on the UFF4MOF and Dreiding potentials

To put the performance of the machine learned potentials into perspective, we also evaluated the performance of UFF4MOF and Dreiding for the validation set. For the forces, we show the comparisons to DFT in Supplementary Figure 18 in the form of 2D Histograms. It can be seen that the deviations can be substantial and there is a trend to predict much higher forces by the UFF4MOF and Dreiding force fields than by DFT. For UFF4MOF in the case of MOF-5, there is still a somewhat reasonable agreement for a lot of the reference structure (but still much worse than any potential specifically fitted for MOF-5). For MOF-74 we see already a much worse agreement as the structure is substantially more complex. The worse performance primarily stems from the parameters used in UFF4MOF. One atom type has only one associated angle, which for the 4 coordinated Zn atoms (atom type Zn4+2 in UFF4MOF) amounts to 90°. This is close to the O-Zn-O angles in MOF-74 but there are

still deviations of several degrees. This leads to a distorted energy minimized structure for UFF4MOF with angles too close to 90°, but still resembling MOF-74 in its general shape. In that sense it is not surprising that UFF4MOF performs substantially worse for MOF-74. However, because of this one should be careful when applying these methods for high throughput screening. For the Dreiding potential the agreement is already extremely poor for MOF-5. Again, this is not at all surprising as that potential was never intended to be applied to MOFs. In fact, its creation even predates the discovery of the first MOFs.

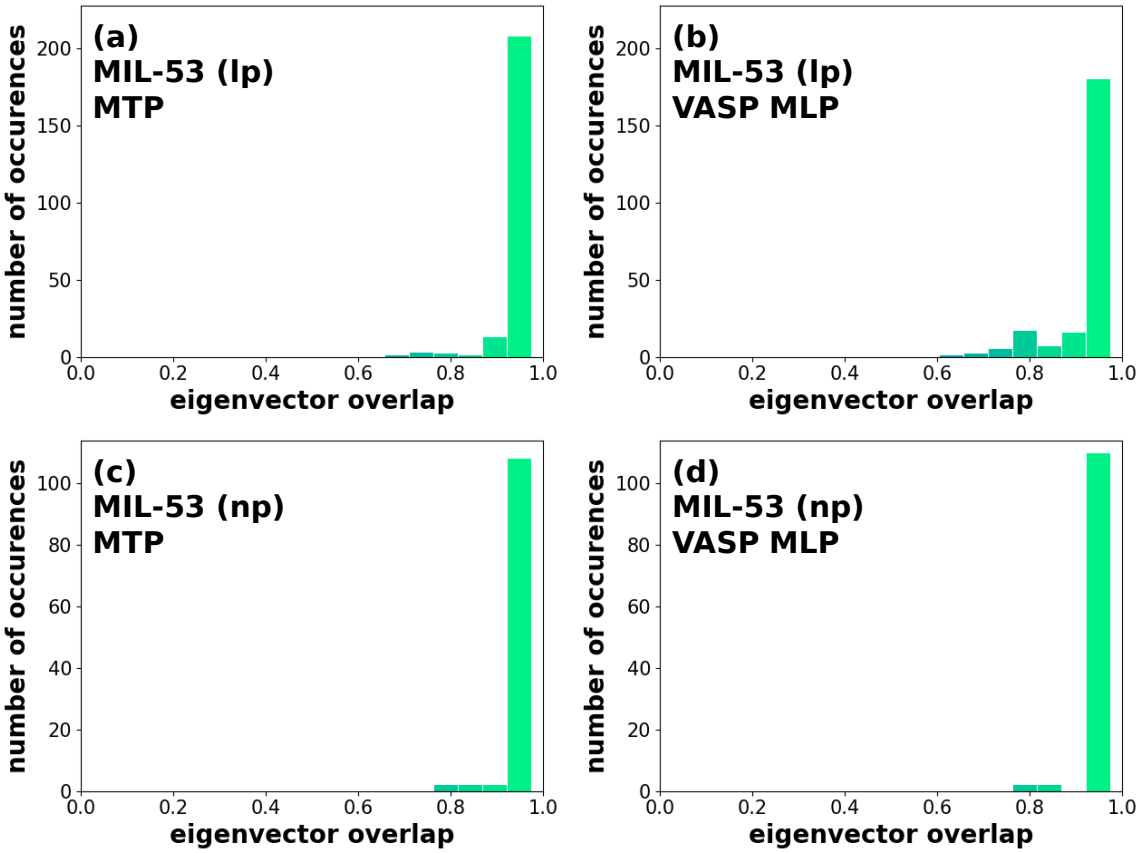


Supplementary Figure 18: 2D Histograms showing the correlation of the forces between UFF4MOF (in a, b) and Dreiding (in c) force field calculations and DFT data (ref) for MOF-5 in a, c and for MOF-74 in b. The black dashed line with a slope of one serves as a guide to the eye indicating ideal agreement.

Supplementary Note 3.2 Phonon eigenvector comparisons

Another aspect we checked for the vibrations and did not discuss in the main manuscript for it not to become too extensive is a quantitative evaluation of the actual equivalence of modes between the force field and DFT simulations. To test that, we calculated the dot products of the phonon mode eigenvectors, where values of 1 refer to a perfect agreement between the force-field and DFT calculated vibrations. This was done for all possible combinations of eigenvectors from both sets of modes. Then the agreement was optimized by solving an assignment problem to obtain the combination of eigenvectors with the highest overlap. In this way the actually matching modes – which might be ordered differently due to slight frequency shifts – are compared. A complication in that context is that many of the investigated MOFs belong to highly symmetric space groups. Consequently, they have large numbers of degenerate phonon modes, where any linear combination of the eigenvectors is a valid description of the motion, which prevents a meaningful evaluation of the dot product. MIL-53 is the only system, which does not show degenerate modes and where the comparison of all eigenvectors over the entire frequency range is sensible. Therefore, in the present comparison we focus on both phases of MIL-53, for which histograms of dot products are shown in

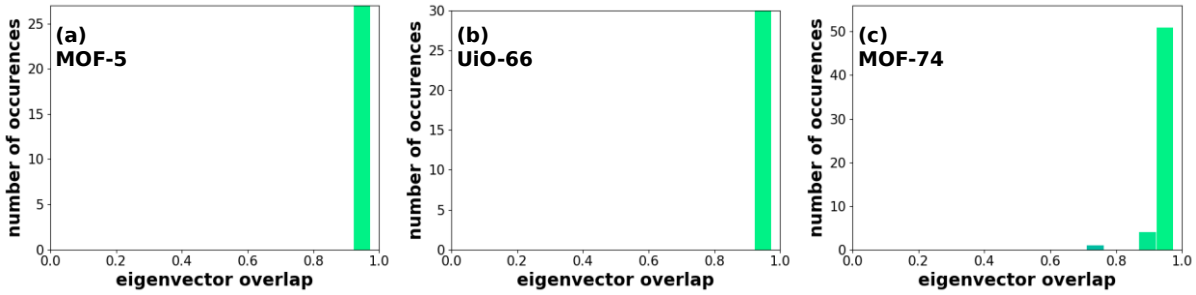
Supplementary Figure 19. For the MTPs and VASP MLPs, the agreement with DFT is almost perfect with the vast majority of the phonon modes being in the highest (0.95-1.00) bin.



Supplementary Figure 19: Histograms showing the optimized agreement of the overlap (among all possible combinations of eigenvectors from different modes) from the dot product of the phonon eigenvectors between DFT and MTP obtained phonon modes in a, c and between DFT and VASP MLP obtained phonon modes (b, d) for MIL-53 (lp) in a, b and MIL-53 (np) in c, d. The bins were distributed in intervals of 0.05. The range of the y axis was deliberately chosen to be equal to the total number of phonon modes in the system.

The other materials have a higher degree of symmetry and feature many two or three-fold degenerate modes. As any linear combination of such degenerate eigenvectors is a proper solution to the eigenvalue problem, the displacements for these modes are ambiguous and a comparison between MLP and DFT becomes rather meaningless. However, it is possible to use phonopy[13] to identify such degenerate modes based on symmetry considerations by evaluating the irreducible representation. Using this (often in combination with a visual inspection of the eigenmodes required by the rather 'messy' band structures), the degenerate modes can be eliminated and an evaluation of the agreement of the non-degenerate modes can be performed also for the other systems. The

corresponding histograms for the comparison of the MTPs based on the initial reference data set with the DFT reference are shown in Supplementary Figure 20. As can be seen, the non-degenerate modes fit very well for the MTPs, especially for MOF-5 and UiO-66, where all modes are within the 0.95 to 1.0 bin. However, the total number of clearly non-degenerate modes is relatively low making the comparison far less conclusive.



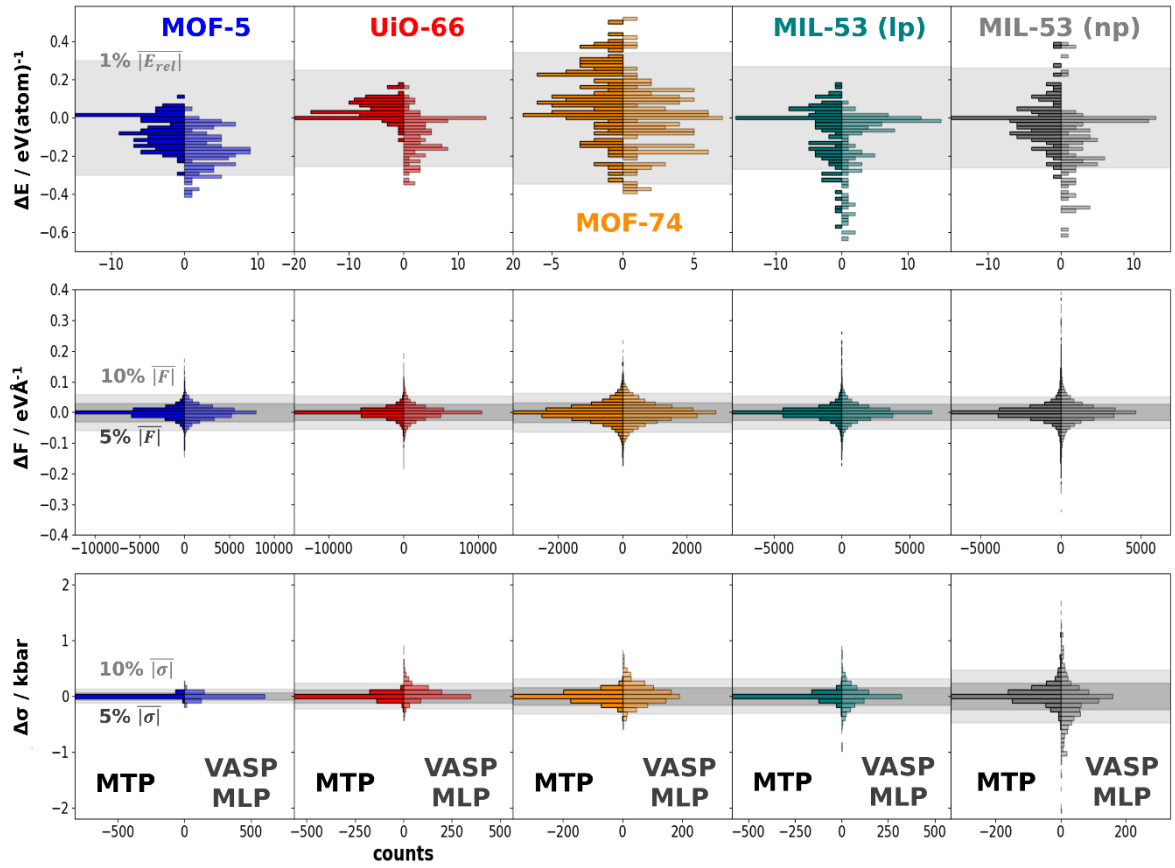
Supplementary Figure 20: Histograms showing the optimized overlap (among all possible combinations of eigenvectors from different modes) of the eigenvectors obtained from the MTPs (initial) and DFT of only the non-degenerate phonon modes.

Supplementary Note 3.3 Further analysis of the extended reference data set

To give a clearer overview of the impact of the extended reference data set, we provide a series of figures displaying relevant quantities in a style similar to the main manuscript for each system for both the MTP and the VASP MLPs. Generally, it should be noted that the VASP MLPs were only trained using 4 atom types for the extended reference data set. The exception for this represents MOF-74, for which results are shown in the main manuscript.

Supplementary Note 3.3.1 Deviation of forces, energies and stresses

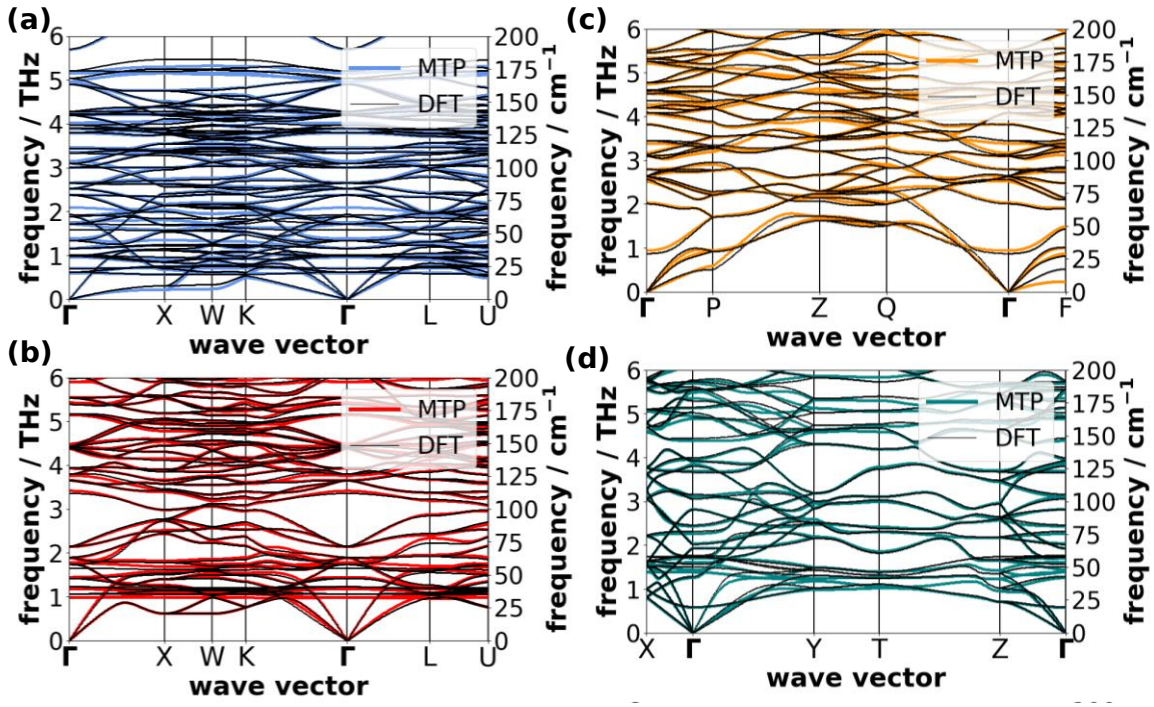
Supplementary Figure 21 contains the comparisons of forces, stresses and energies with the validation set. Compared to the initial reference data set, it can be seen that the differences are rather small. Note, that for the VASP MLPs the agreement now looks slightly worse, since they were only trained for four atom types on the extended reference data set. Numerically we only observed small improvements as already indicated in Supplementary Table 11.



Supplementary Figure 21: Histograms detailing the deviations (FF data subtracted from reference) of the energies (E), the components of the forces (F), and the components of the stresses (σ) obtained from DFT and from the VASP MLPs and the MTPs for MOF-5 (blue), UiO-66 (red), MOF-74 (orange), MIL-53 (lp) (teal) and MIL-53 (np) (grey). All force fields were obtained from the respective extended reference data set. For the VASP MLPs 4 atom types were used in this case. For the energies only the energy difference from the structure with the minimum energy is shown. The validation set structures were obtained from an active learning VASP MLP learning run at a temperature of 300 K. The areas shaded in gray 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.

Supplementary Note 3.3.2 MTP calculated phonon band structures for the extended reference dataset

In *Supplementary Figure 22* one can see the phonon band structures as obtained from MTPs trained on the extended reference data set. Also here, it is apparent that the agreement is excellent and especially for MIL-53 (lp) some improvements can be discerned over the initial reference data set. However, the changes are mostly minor



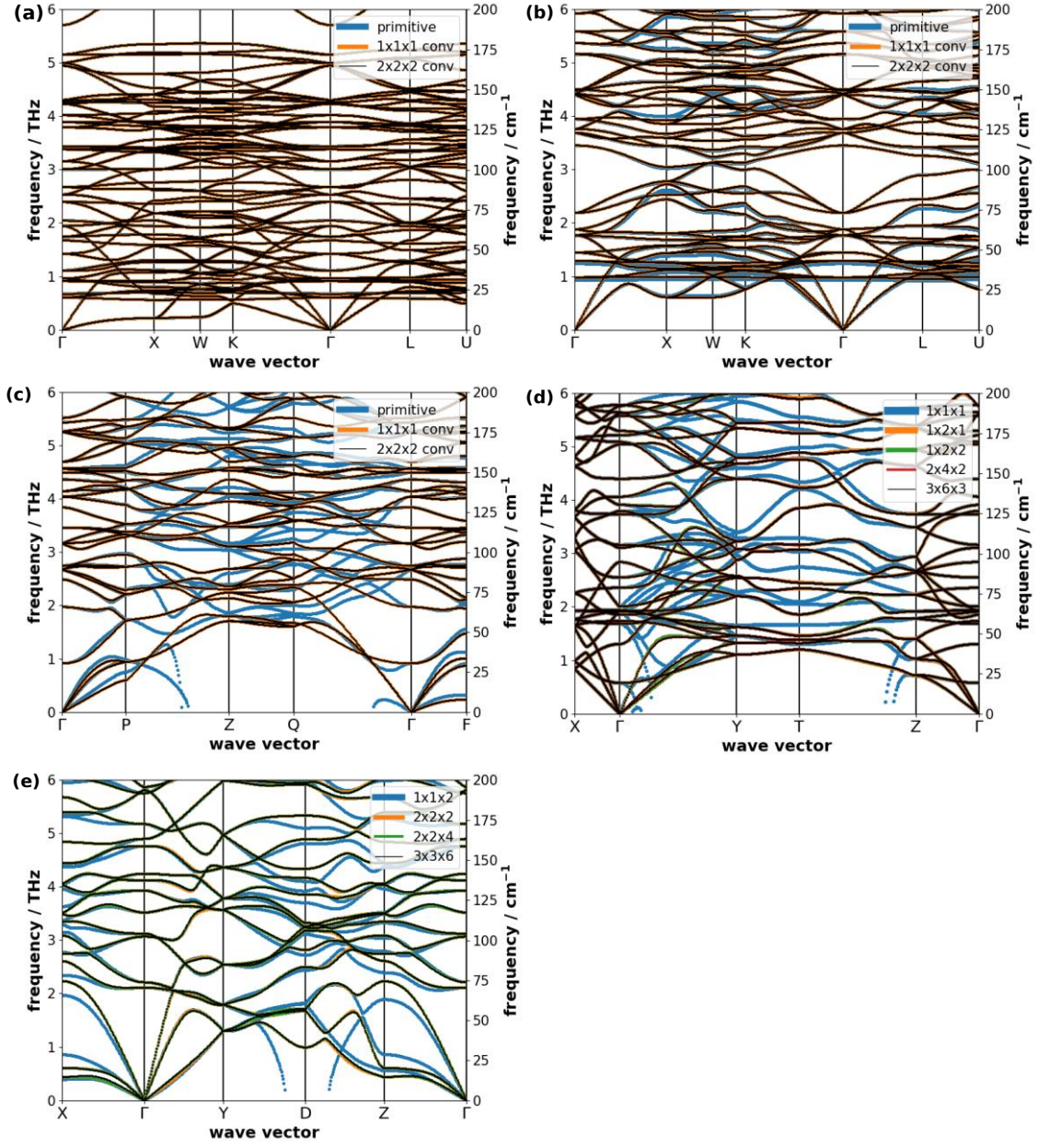
Supplementary Figure 22: Low frequency phonon band structures computed using the MTPs parametrized on the extended reference data sets) and computed using DFT for the following systems: MOF-5 in a, UiO-66 in b, MOF-74 in c, and the large pore phase of MIL-53 in d.

Supplementary Note 3.4 Convergence of the supercell size for the phonon band structures with the MTPs

When computing phonon band structures it is important to ensure a properly converged supercell. Supplementary Figure 23 shows the phonon band structures evaluated using MTPs for different supercells. Here, 'primitive' always refers to the smallest possible unit cell and 'conv' (conventional) means the $\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$ supercell for MOF-5 and UiO-66 and the $\begin{pmatrix} 1 & 0 & 3 \\ -1 & 1 & 3 \\ 0 & -1 & 3 \end{pmatrix}$ supercell for MOF-74. The bands are plotted on top of each other starting with the first one included in the caption. If no bands with a specific color can be seen that means that they essentially coincide with the bands for larger supercells.

For MOF-5 even the primitive unit cell is sufficient to obtain fully converged phonon band structures for the MTPs. For UiO-66 there are still some minor differences between the primitive and cubic conventional unit cell. For MOF-74, the primitive unit cell is clearly insufficient, but the regular conventional cell is already converged. For MIL-53 (lp), the 1×2×1 supercell already shows no visible

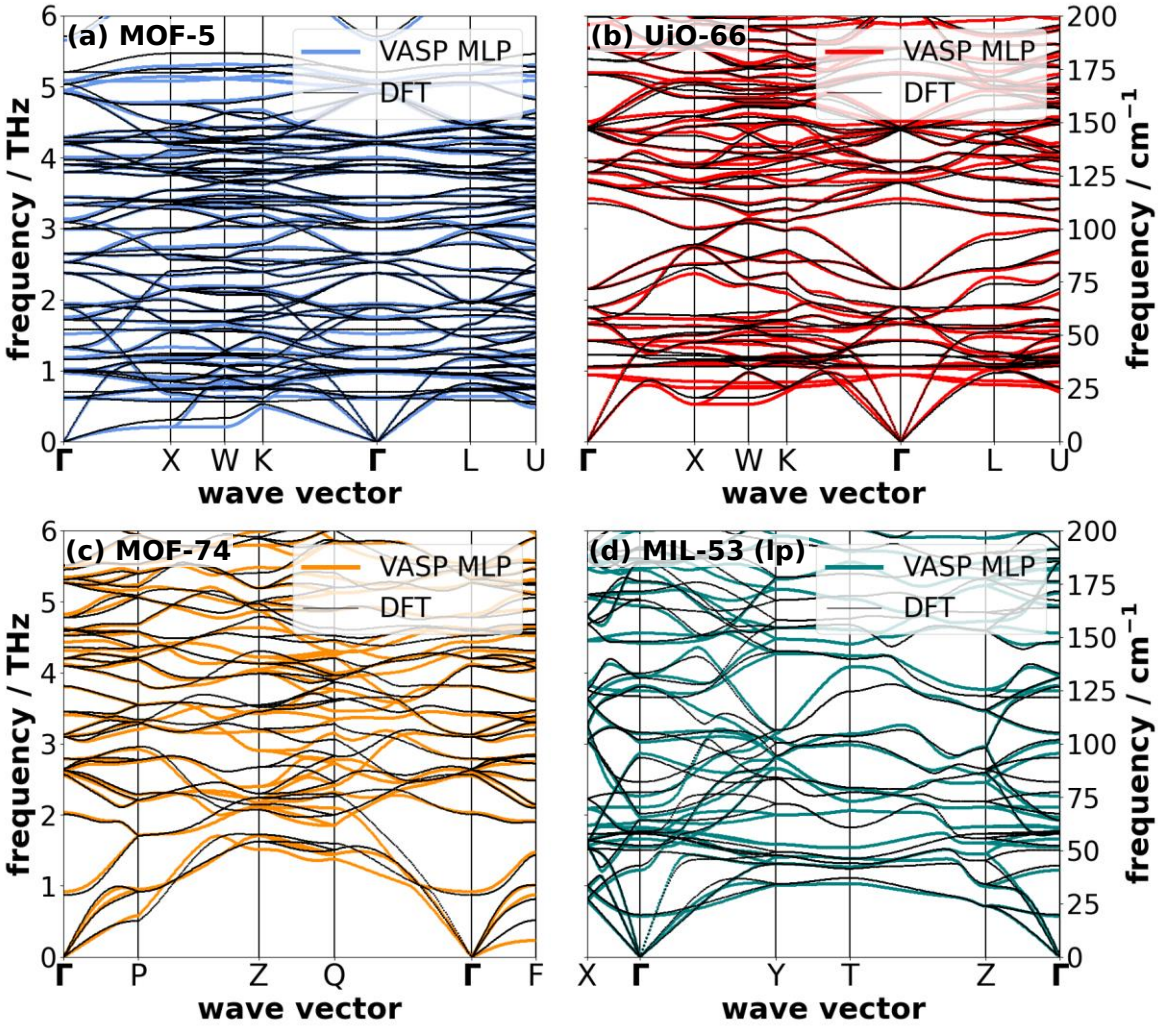
differences compared to the larger supercells. It should be noted that this is inconsistent with what was observed for DFT, where there were still substantial differences between 1×2×1 and 1×2×2 supercells for MIL-53 (lp). The reason for this is that the MTPs are cut off after already 5 Å, which should typically lead to a faster convergence behavior than DFT, where also longer-range interactions are included. For MIL-53 (np), we see convergence after using a 2×2×2 supercell with no imaginary frequencies (unlike DFT).



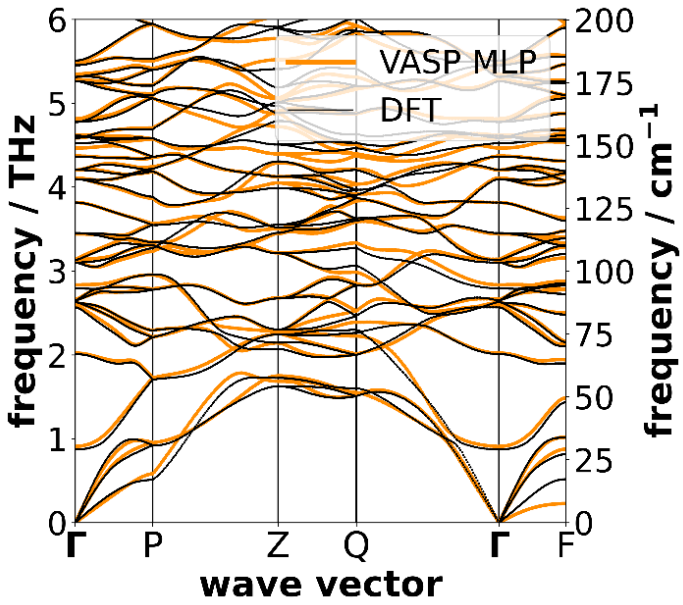
Supplementary Figure 23: Phonon band structures evaluated for different supercells for the following systems: MOF-5 in a, UiO-66 in b, MOF-74 in c, MIL-53 (lp) in d, and MIL-53 (np) in e. The MTPs based on the initial reference data set were used. The last entry in each legend represents the respective top layer of the drawn band structures.

Supplementary Note 3.5 Phonon band structures obtained with the VASP MLP and with UFF4MOF

In the main manuscript mostly the phonon band structures obtained with the MTPs are shown. For the VASP MLPs in particular only the phonon band structure for MOF-5 was shown. For the VASP MLPs trained on the initial reference data set, we provide the comparisons to the DFT-calculated band structures in Supplementary Figure 24 for 4 of the systems. Generally, the agreement is very good, maybe slightly worse than for the MTPs when compared with Fig. 4 in the main manuscript. There especially the optical mode around 1 THz in UiO-66 and some of the optical modes in the range of 2-5 THz for MOF-74 and MIL-53 show a somewhat better agreement for the MTPs. But as mentioned in the main manuscript: the VASP MLPs can still benefit substantially from an extension of the reference data set. This can be seen for MOF-74 in Supplementary Figure 25, where the phonon band structure is shown for a VASP MLP trained on the extended reference data set. There, the agreement is substantially improved compared to the initial reference data set and the somewhat poorly fitting modes around 2-4 THz present previously now show a good agreement.

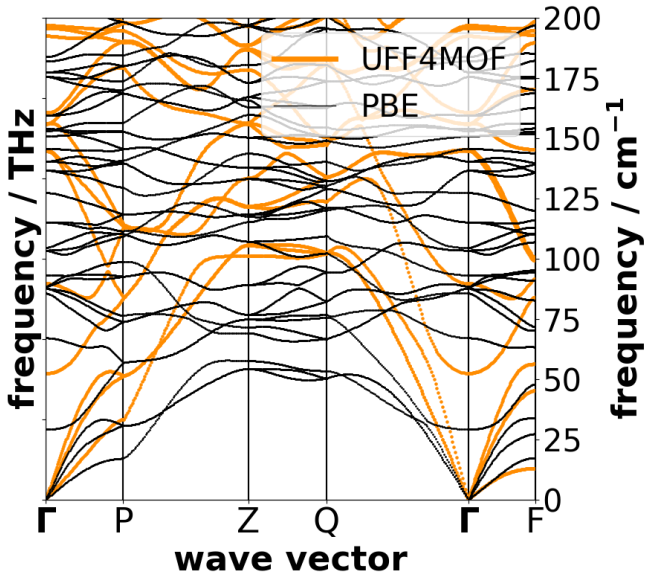


Supplementary Figure 24: Low frequency phonon band structures computed using the VASP MLPs (from the initial reference data set) and from DFT for the following systems: MOF-5 in a, UiO-66 in b, MOF-74 in c and the large pore phase of MIL-53 in d.



Supplementary Figure 25: Low frequency phonon band structures computed using the VASP MLP trained on the extended reference data set (including separation of atom types) and DFT for MOF-74.

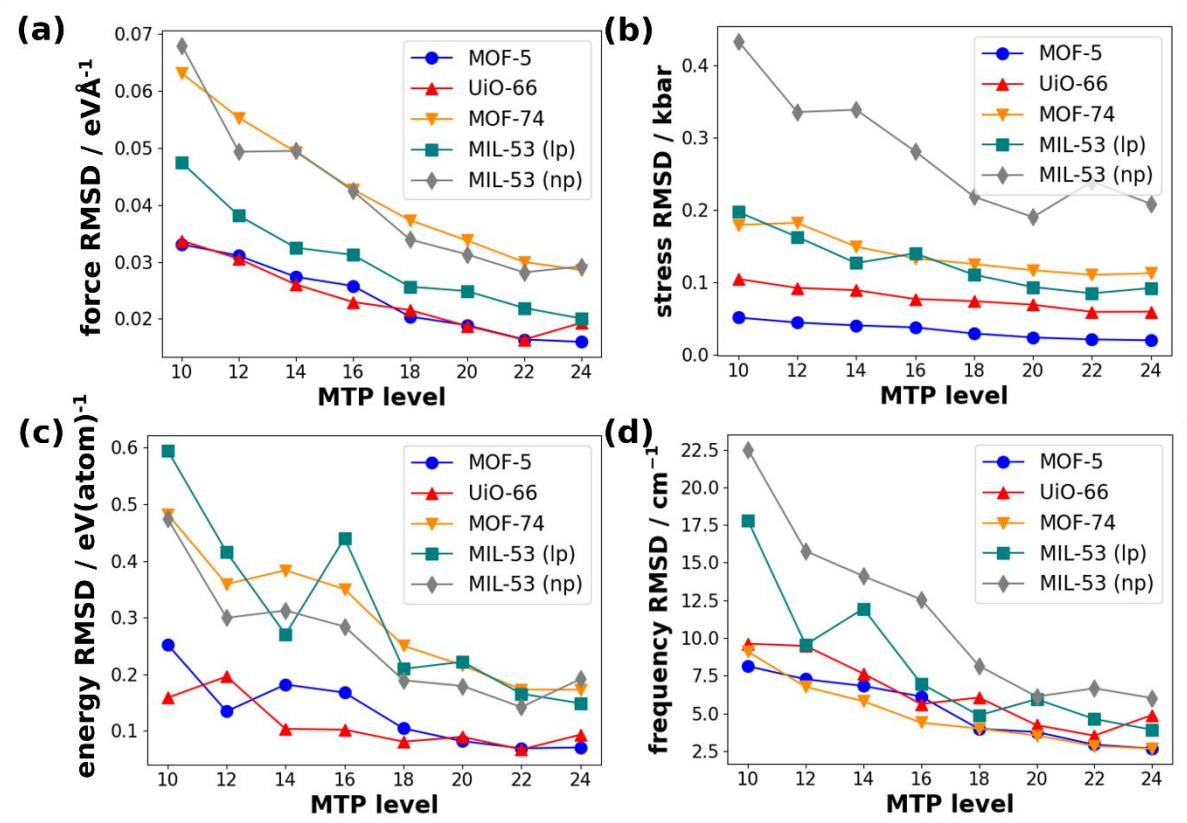
For UFF4MOF, the low frequency phonon band structure of MOF-74 and its comparison with DFT is shown in Supplementary Figure 26. It can be seen that especially in this frequency region the agreement is extremely poor – even worse than for MOF-5.



Supplementary Figure 26: Low frequency phonon band structures computed using UFF4MOF and DFT for MOF-74.

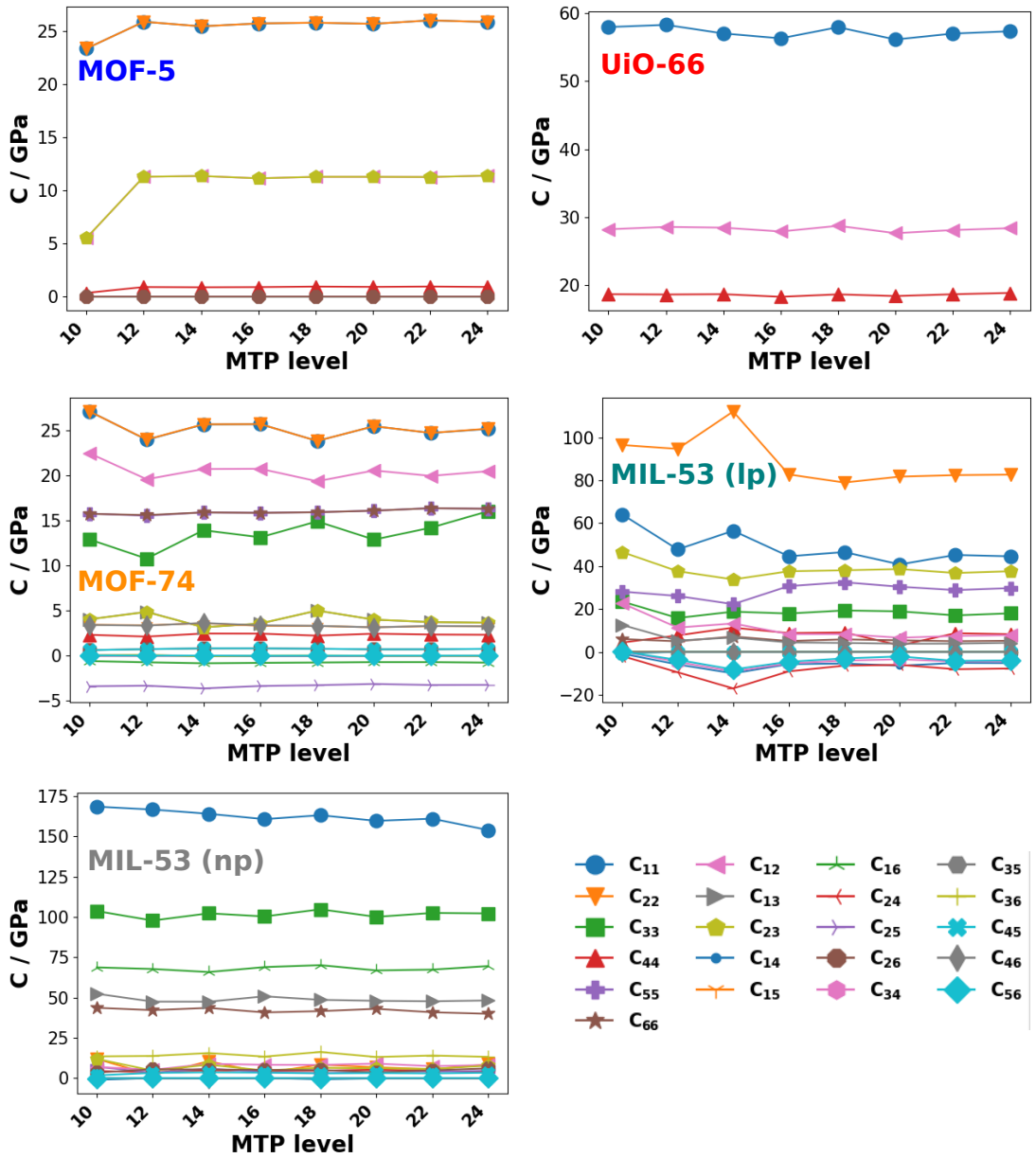
Supplementary Note 3.6 Details on the impact of the level of the MTPs

In this Note, we provide additional information on how the level of the MTP impacts the calculation of various of the observables. The impact on the RMSD errors for forces, stresses and energies for validation set and on for the frequencies at the Γ -point are shown in Supplementary Figure 27. Generally, all of the errors show a steady decline with the MTP level. Especially for the phonon frequencies the RMSD value can be reduced by a factor of three for some of the systems. For the stresses, the differences between the systems merely results from different absolute values. Generally, there is a certain degree of noise throughout the data points. This is a result of the fact that only the primarily used level 22 potentials the 'best' MTP of a set of fitted potentials was picked. For all other levels, only one MTP was fitted. This leads to a certain degree of variation in accuracy due to the stochastic initialization of the MTP fitting procedure. The energies show a larger degree of noise, which result from the errors already being low, even for the low level MTPs.



Supplementary Figure 27: The effect of the MTP level on the RMSDs of the forces in a, stresses in b and the energies in c in the validation set and also on the phonon frequencies at the Γ-point (in d) for a comparison between calculations with MTPs at various levels and with DFT. The MTPs have been parametrized based on the initial reference data set. The studied systems comprise: MOF-5 (blue), UiO-66 (red), MOF-74 (orange) and MIL-53 (lp) (teal) and MIL-53 (np) (gray).

Additionally, the comparisons for the elastic tensor elements are given in Supplementary Figure 28. In particular for MIL-53 (lp) and MOF-74 we see quite substantial deviations in the level 10-14 MTPs. These deviations do not occur for all the systems at the same levels, as the elastic tensor elements for MOF-5, UiO-66 and MIL-53 (np) are not that dissimilar from those for level 12 upwards. Overall, the results are quite system dependent and also depend on the particular tensor element. For example, for MOF-74, substantial deviations in C_{33} can be seen even at high levels of 20-24. For MIL-53 this applies to most tensor elements.



Supplementary Figure 28: Comparison of the elements of the elastic stiffness tensor for moment tensor potentials obtained for different levels (10-24) on the initial reference data set for each of the investigated systems.

Supplementary Note 3.7 Impact of VASP MLP training settings

In Fig. 7c in the main manuscript, the results for a number of different settings chosen for training MOF-74 VASP MLPs were shown. Here, we provide some additional information. Additionally, the role of some other settings, which have not resulted in large deviations, will be discussed. For this, Supplementary Table 16 provides an overview of the RMSD values for the forces in the validation set for various settings in combination with the CPU time per time step and atom for MOF-74. One of the primary differences that can be observed is between the default settings of VASP version 6.3.0 and 6.4.1, where the newer version promises much faster and more accurate potentials. In the field of machine learned potentials it is unsurprising that improvements are made at a rapid pace. In the following, we will discuss the reasons for said improvements.

To clarify the significance of the tested settings and the differences between the VASP versions, the method for fitting will be briefly discussed here. Note, that this explanation is not exhaustive and has been discussed in more detail elsewhere [19] (see also the VASP manual at https://www.vasp.at/wiki/index.php/The_VASP_Manual). The objective of the training is to obtain the weights $\mathbf{w}$, which yield the energies, forces and stresses combined in the vector $\mathbf{Y}$ via the design matrix $\mathbf{\Phi}$ containing the Kernels and their derivatives

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{w}.$$

To achieve this, the Bayesian theorem is used in combinations with multivariate Gaussian distributions describing the distribution of weights around a center $\bar{\boldsymbol{w}}$ in combination with a covariance matrix $\mathbf{\Sigma}$. Now the objective becomes to obtain these centers, which can be expressed as

$$\bar{\mathbf{w}} = s_v \mathbf{\Sigma} \mathbf{\Phi}^T \mathbf{Y}$$

$$\mathbf{\Sigma}^{-1} = s_w \mathbf{I} + s_v \mathbf{\Phi}^T \mathbf{\Phi}.$$

Here, $\mathbf{I}$ is the unit matrix, s_v is the noise parameter (ML_SIGV0) and s_w is the precision parameter (ML_SIGW0) of the regularization during the fit. During on-the-fly training, the weights are obtained using an LU factorization approach (ML_IALGO_LINREG=1) and during the refit to create the force fields for the production runs with a more expensive singular value decomposition (SVD, ML_IALGO_LINREG=3,4) approach. During LU factorization, the regularization parameters σ_w^2 and σ_v^2 are dynamically optimized (using an evidence approximation approach), where the eigenvalues λ of $s_v \mathbf{\Phi}^T \mathbf{\Phi}$ above a certain threshold are included in the optimization. This threshold is controlled by ML_EPS_REG. During the training runs with VASP 6.3.0 this value was set to 10^{-14} and in the refined training in the newer version to 10^{-15}. This evidence approximation is not used for the SVD fits for ML_IALGO_LINREG=4 due to the increased computational cost. In this case, s_w was manually set to 10^{-7}. In one of the tests in Supplementary Table 16, we see that increasing this value to 10^{-5} (by

adjusting ML_SIGW0) only makes a negligible difference in terms of accuracy and speed in the production runs.

Based on the high CPU time required given in Supplementary Table 16, it is clear that the potentials using ML_IALGO_LINREG 1 (Bayesian linear regression) and 2 (QR factorization) are not really suited for production runs. Therefore, one always should retrain the potential for a fast force field using either ML_IALGO_LINREG 3 or 4 (Singular value decomposition, SVD, where ML_IALGO_LINREG=4 is the default for VASP 6.4.1 using ML_MODE=refit). The accuracy in this case even gets better due to the use of the SVD method.

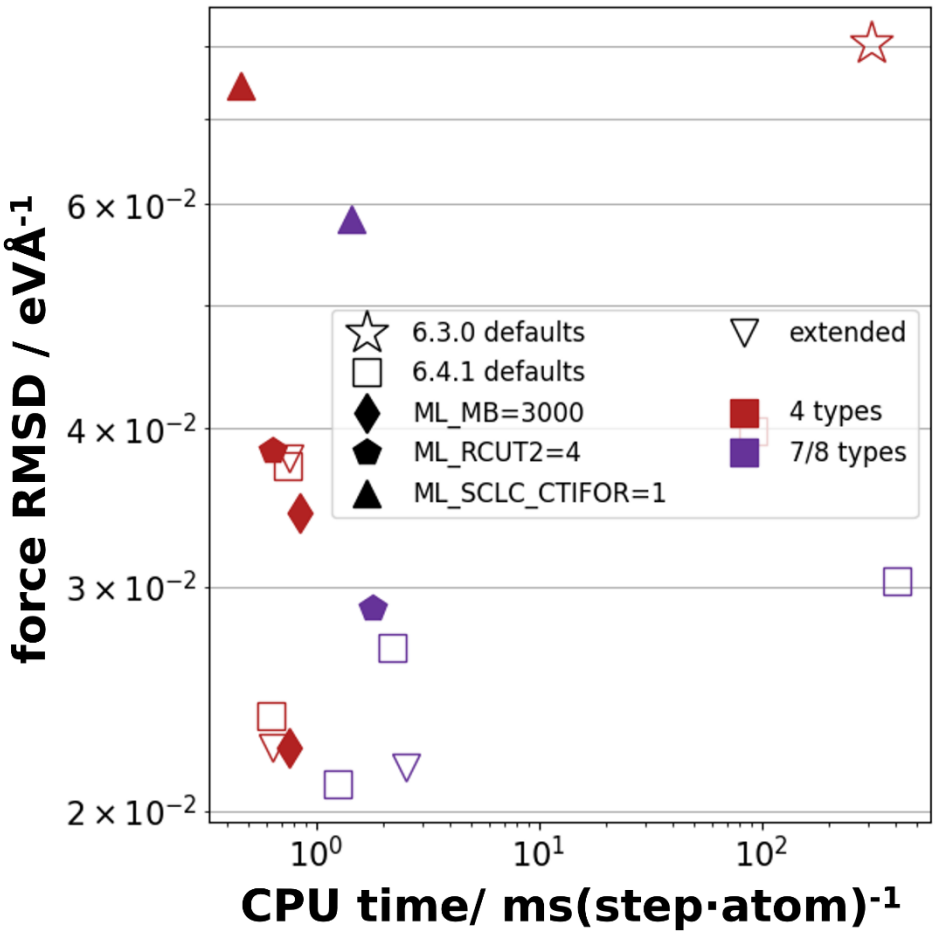
Aside of the settings affecting the fitting procedure, we also tested some settings related to the descriptors. This includes the cutoff radius of the angular descriptors already discussed in the main manuscript (ML_RCUT2), where only minor differences were observed. Also related to the angular descriptors, we also tested the effect of the ML_LMAX2 setting, which defines the maximum angular momentum quantum number of the spherical harmonics in the descriptor. However, the resulting potentials when setting the parameter to 2 performed essentially identical to the ones with the default value in VASP 6.4.1 of 3. When changing the cutoff of the radial descriptor (ML_RCUT1) from 8 (default in VASP 6.4.1) to 5 (default in VASP 6.3.0), only a small difference in accuracy and cost is observed.

The main reason why the new version of VASP yields much more accurate potentials due to including more local reference configurations, is the ML_SCLC_CTIFOR parameter, which was changed from 1 to 0.6 in the new version. In Supplementary Table 16 it becomes clear that when using 1.0 in the new version the number of reference configurations, and consequently the accuracy, is drastically reduced. However, this also leads to the fastest VASP MLPs as also shown visually in Supplementary Figure 29. The ML_SCLC_CTIFOR directly influences the number of reference configurations that are added during training, as the Bayesian error threshold is multiplied with this value to determine which local environments are included as local reference configurations. This makes it a convenient tool for controlling the accuracy and cost of the VASP MLPs.

Supplementary Table 16: Evaluation of the force RMSDs (with the validation sets) and required CPU time for different VASP MLPs for MOF-74. The number of atom types is given as n_{types} *and the number of local reference configuration as* n_{conf}*. Unless specified otherwise, the baseline of the force fields is always given as the VASP 6.4.1 defaults with the specified setting modified.*

Data set	setting	n_{types}	n_{conf}	Force RMSD / eVÅ^{-1}	CPU time / ms(step·atom)$^{-1}$
initial	VASP 6.3.0 defaults	4	1877	0.080	310.62

initial	6.4.1 defaults	4	4830	0.037	0.74
initial	ML_IALGO_LINREG=1	4	4830	0.040	92.14
initial	ML_IALGO_LINREG=2	4	4830	0.040	92.18
initial	ML_IALGO_LINREG=3	4	4830	0.037	0.75
initial	ML_SIGW0 = 1E-5	4	4830	0.038	0.74
initial	ML_RCUT2=4	4	3947	0.038	0.64
initial	ML_RCUT1=5	4	4967	0.038	0.70
initial	ML_LMAX2=2	4	4795	0.037	0.74
initial	ML_SCLC_CTIFOR=1	4	1770	0.075	0.46
initial	ML_MB=3000	4	5859	0.034	0.85
initial	6.4.1 defaults	8	8942	0.027	2.19
initial	ML_IALGO_LINREG=1	8	8942	0.030	408.73
initial	ML_IALGO_LINREG=2	8	8942	0.030	410.11
initial	ML_IALGO_LINREG=3	8	8942	0.027	2.21
initial	ML_SIGW0=1E-5	8	8942	0.027	2.30
initial	ML_RCUT2=4	8	6333	0.029	1.81
initial	ML_RCUT1=5	8	9564	0.028	2.24
initial	ML_LMAX2=2	8	8942	0.027	2.24
initial	ML_SCLC_CTIFOR=1	8	4242	0.058	1.45
extended	6.4.1 defaults	4	5237	0.038	0.76
extended	6.4.1 defaults	8	11350	0.022	2.53



Supplementary Figure 29: Errors of the forces on a validation set over CPU time required for MOF-74 obtained with a variety of different VASP MLPs, most of which were already shown analogously in Figure 7 in the main manuscript. Here, we also show the data point for the default settings of the old VASP version 6.3.0 as a red star. Additionally some potentials trained with VASP version 6.4.1 are shown, which are indicated with triangles and where ML_SCLC_CTIFOR has been set to 1.0, which was the default value for VASP 6.3.0, while for all the other data points with a higher degree of accuracy a value of 0.6 was used.

Supplementary Note 4. Molecular dynamics

Supplementary Note 4.1 Constant temperature and pressure simulations to evaluate the average lattice parameters and thermal expansion coefficients

To obtain the dependent lattice parameters and thermal expansion coefficients, the same style of NPT simulations was used as already mentioned in Supplementary Note 2.2. The supercells employed for each system are listed in Supplementary Table 17. The simulations were carried out in the temperature range between 100 and 700 K and all simulations for at least 100,000 time steps that

concluded successfully (i.e., for which the system did not disintegrate) were used in the evaluation of the thermal expansion coefficient. However, some data points were still excluded occasionally (see discussion below for details and individual explanation of the reasoning) due to a non-linear evolution of the lattice parameters/unit-cell volumes with temperature, which would make the evaluation of linear thermal expansion coefficients difficult. These coefficients were obtained by a linear fit of the dependence of individual lattice parameters on temperature.

Supplementary Table 17: Used supercells to obtain the thermal expansion coefficients using NPT molecular dynamics simulations for both the VASP MLP and MTP approaches.

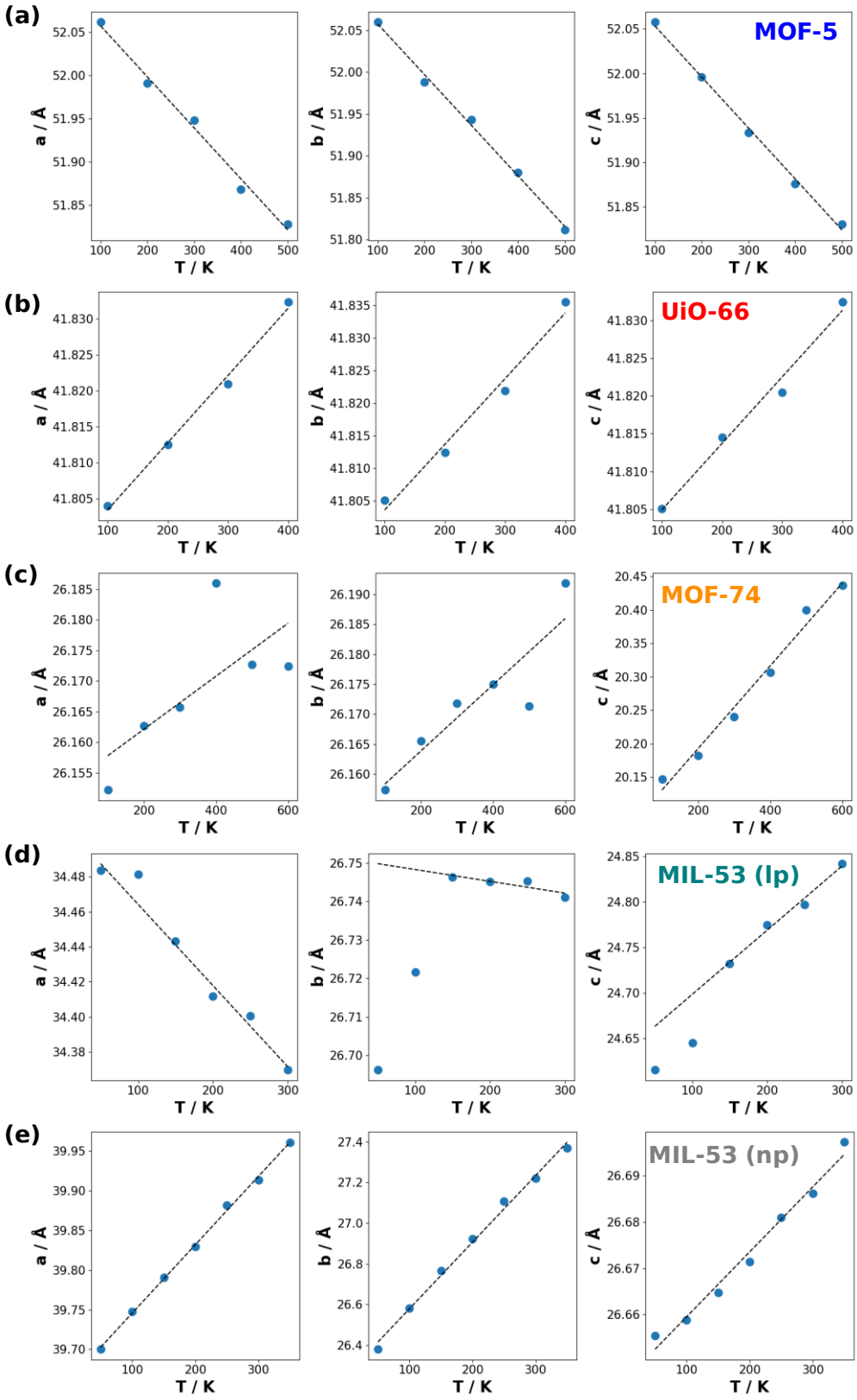
System	Supercell	Number of atoms
MOF-5	2x2x2 cubic conventional unit cells	3392
UiO-66	2x2x2 cubic conventional unit cells	3648
MOF-74	1x1x3 hexagonal conventional unit cells	486
MIL-53 (lp)	2x2x4 unit cells	1216
MIL-53 (np)	2x4x4 unit cells	2432

To add to the cells used for the thermal expansion calculations, the speed tests were performed for substantially larger supercells to properly test the parallelization on a high performance computation system. For MOF-5 4×4×4 conventional supercells (27,136 atoms) and for MOF-74 4×4×12 hexagonal conventional unit cells (31,104 atoms) were used. As an additional note, it should be said that in general for the tests of the computational efficiency time values were taken from the actual execution of the MD run steps. In LAMMPs this means that the time elapsed was taken from the performance breakdown after the 'run' command. In VASP the value specified as 'pes_ff' in the table of the Accumulative profile is taken, which is only output when the code was compiled using the '-DPROFILING' compiler option. This option also excludes the time lost due to outputting various quantities during the simulation, which can be a time bottleneck otherwise. Both strategies serve to exclude the initialization overhead from the actual evaluation of the required computation time per time step. This is the reason why we only performed the efficiency tests for 2000 time steps to save computation time for simulations on such large supercells.

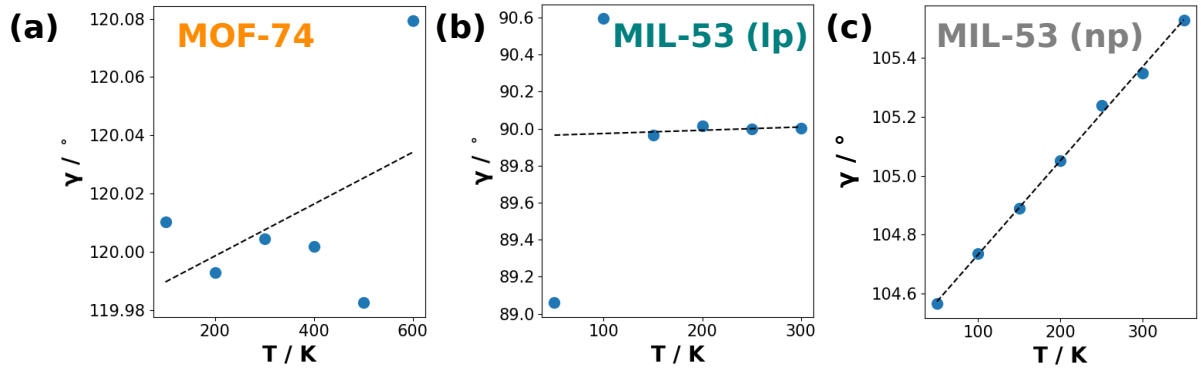
Supplementary Note 4.1.1 NPT simulations with the MTPs

Supplementary Figure 30 and Supplementary Figure 31 show the linear thermal expansion fits for the lattice parameters and the tilt angles for all the systems obtained from NPT simulations carried out by

the MTPs based on the initial reference data set. For the cubic systems MOF-5, MIL-53 (np) and UiO-66 the trends are clearly linear and the thermal expansion can be obtained straightforwardly. For MOF-74, we see some noisy datapoints in a and b direction which is mostly due to the thermal expansion being very small. For MIL-53 (lp), we see an odd step in the b parameter and a non-orthorhombic angle at low temperatures. This is related to the minimum structure at 0 K showing a slight deviation from the orthorhombic symmetry. This deviation amounts to two minima at +/- 1 ° with a weak energy barrier in between. At high temperatures, kinetic energies are high enough to overcome that barrier and one obtains an on average orthorhombic structure, which is consistent with experiment. Due to this, the thermal expansion for MIL-53 (lp) was only evaluated for temperatures of 150 K upwards. For MIL-53 (np) we also see a significant increase of the tilt angle γ with temperature amounting to 30.6 K^{-1}.



Supplementary Figure 30: Averaged lattice parameters of the investigated systems – MOF-5 (in a), UiO-66 (in b), MOF-74 (in c), MIL-53 (lp) (in d) and MIL-53 (np) (in e) – as a function of temperature (blue dots) obtained from NPT simulations using the MTPs based on the initial reference data sets. The black dashed line represents the linear fit used to obtain the thermal expansion coefficients.



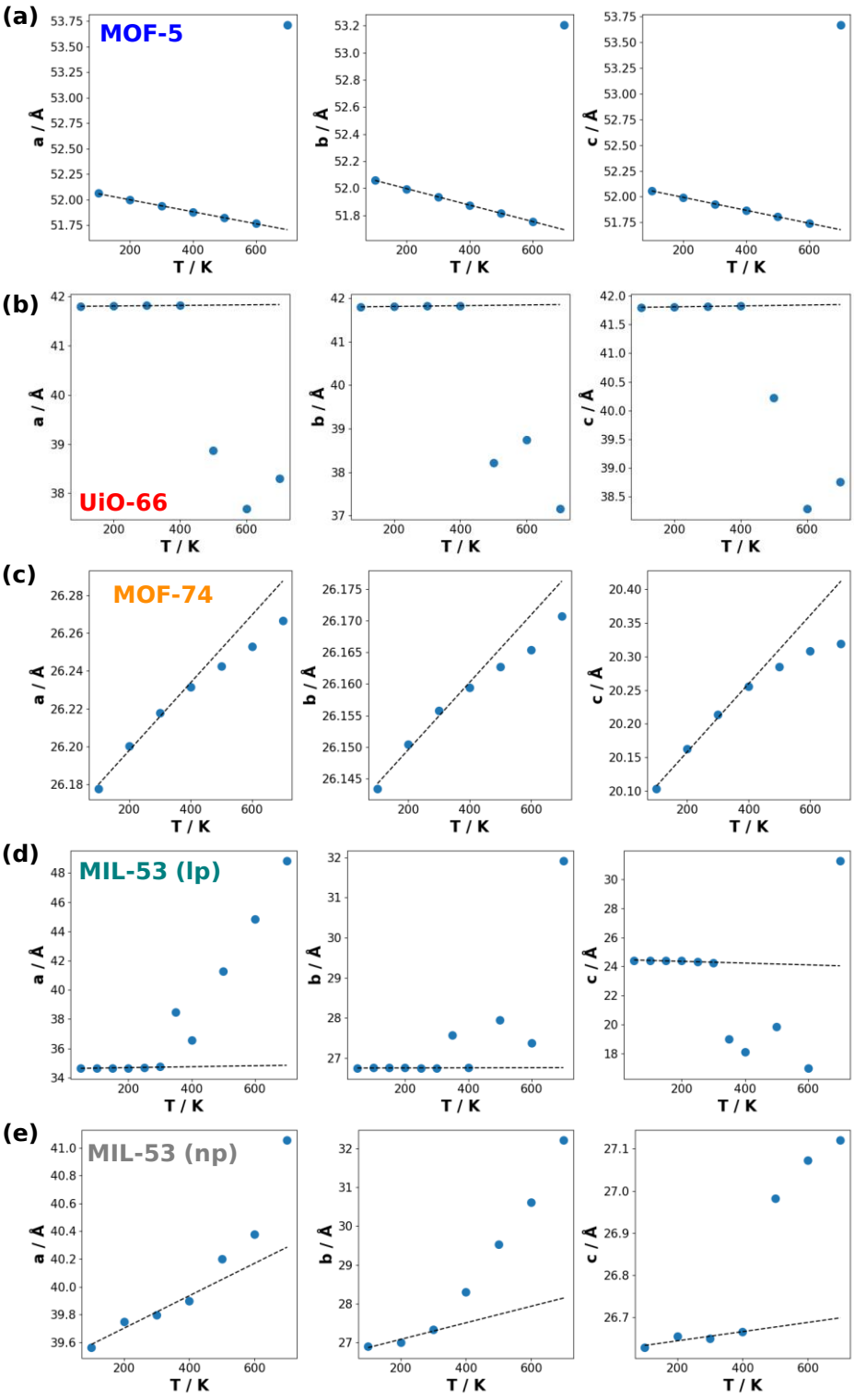
Supplementary Figure 31: Averaged tilt angles of the investigated systems – MOF-74 (in a), MIL-53 (lp) (in b) and MIL-53 (np) (in c) – as a function of temperature (blue dots) obtained from NPT simulations using the MTP based on the initial reference data sets. The black dashed line represents the linear fit used to obtain the change of the tilt angles with temperature.

It should be noted, that for MOF-74, it has been shown for DFT calculations and the Grüneisen theory of thermal expansion that a thermal expansion coefficient similar to the experiment is provided only by the PBEsol and not by the PBE functional [8]. Therefore, we also used this functional to train separate MTPs with an analogous approach as done throughout this work employing the PBE functional. A reference data set of 1034 DFT calculations was produced using a VASP active learning approach with temperatures reaching up to 700 K. However, the obtained MTPs in NPT simulations yielded thermal expansion coefficients as the PBE-trained MTP (e.g., thermal expansion coefficients for a level 22 MTP with a radial basis set size of 10: 5.5 10^{-6} K^{-1} in linker direction and 28.8 10^{-6} K^{-1} in stacking direction). Therefore, we conclude that the functional is not the primary reason for the discrepancy.

Supplementary Note 4.1.2 NPT simulations with the VASP MLPs

Also for the VASP MLPs, we performed NPT simulations up to a temperature of 700 K using the potential variant trained on the initial reference data set including separation of atom types. The lattice parameters as a function of temperature including the linear fits to obtain the thermal expansion coefficients, can be seen in *Supplementary Figure 32*. For the VASP MLPs all systems were technically stable up to a temperature of 700 K, but as can be clearly seen in the figure, at high temperatures large distortions of the unit cells occur for a lot of the systems. This is the reason why the linear fits were only performed up to a temperature, where the trend was still essentially linear. These temperatures were 600 K for MOF-5, 400 K for UiO-66, 400 K for MOF-74, 300 K for MIL-53 (lp)

and 300 K for MIL-53 (np). For MOF-74 we only imposed the limit due to the nonlinearity of the lattice parameters as a function of temperature and not due to discontinuities. For the other systems the behavior after the given temperature is most likely not physical.



Supplementary Figure 32: Averaged lattice parameters of the investigated systems – MOF-5 (in a), UiO-66 (in b), MOF-74 (in c), MIL-53 (lp) (in d) and MIL-53 (np) (in e) – as a function of temperature

(blue dots) obtained from NPT simulations using the VASP MLPs based on the initial reference data sets and using separation of atom types for the potentials. The black dashed line represents the linear fit used to obtain the thermal expansion coefficients.

Supplementary Note 4.1.3 Lattice parameters at room temperature and thermal expansion coefficients

Supplementary Table 18 provides a comparison of the calculated MTP and VASP MLP lattice parameters at room temperature (300 K for the simulations) with experiments. Generally, the values for the machine learned approaches are rather similar. For MOF-5, UiO-66 and MOF-74 the differences compared to experiment are relatively small. For MIL-53 (lp) we see a difference of up to 0.5 Å compared to the relatively consistent experimental data. This must be related to inaccuracies from the underlying DFT approach, as this large deviation from experiment was already present for the relaxed DFT unit cells. For MIL-53 (np) we also see large deviations between the MTP calculations and the experiment. However, for that system the available experimental data are also not particularly consistent.

Supplementary Table 19 summarizes the linear thermal expansion coefficients obtained with the MTPs and VASP MLPs for all systems as obtained by the procedure described in the preceding Note. In the main manuscript, only the thermal expansion coefficients for MOF-5 and MOF-74 were compared to experiment directly. Therefore, the remaining results are discussed here to expand upon the analysis of the lattice parameters at 300 K in Fig. 2. For UiO-66 in the linker direction the thermal expansion situation is similar to that in MOF-74 with an MTP/VASP MLP-calculated small positive thermal expansion coefficient ($2.2 \cdot 10^{-6}$ K^{-1}/ $1.8 \cdot 10^{-6}$ K^{-1}) compared to a slightly negative coefficient in the experiments ($-5.6 \cdot 10^{-6}$ K^{-1})[80]. For MIL-53 (lp) at least the qualitative trends are consistent for the MTP, with simulated anisotropic expansion values of -15.5, 4.4 and $40.2 \cdot 10^{-6}$ K^{-1} compared to -14.5, -1.0 and $24.2 \cdot 10^{-6}$ K^{-1} from experiment[87]. Conversely, the VASP MLPs predict qualitatively different linear thermal expansion coefficients of 9.3, 0.5 and $-24.9 \cdot 10^{-6}$ K^{-1} for MIL-53 (lp). Here, the differences between the machine learned potentials could be attributed to the difference in stress and force weights. However, further analysis would have to be carried out for a conclusive statement.

Supplementary Table 18: Room temperature lattice parameters obtained with MTPs and VASP MLPs (employing NPT simulations) and obtained from experiments.

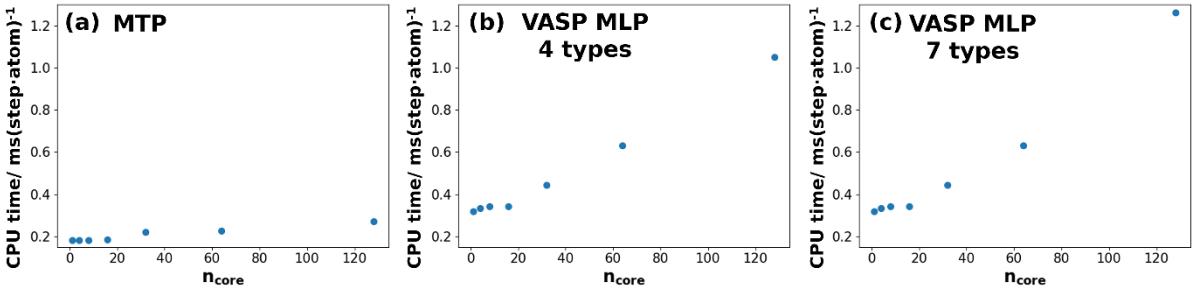
System	Method	a / Å	b / Å	c / Å	gamma / °	ref.
MOF-5	MTP	25.97	25.97	25.97	90.00	
	VASP MLP	25.97	25.97	25.96	90.00	
	exp.	25.82	25.82	25.82	90.00	[20]
		25.78	25.78	25.78	90.00	[21]
		25.82	25.82	25.82	90.00	[22]
UiO-66	MTP	20.91	20.91	20.91	90.00	
	VASP MLP	20.91	20.91	20.90	90.00	
	exp.	20.77	20.77	20.77	90.00	[22]
		20.70	20.70	20.70	90.00	[23]
		20.75	20.75	20.75	90.00	[24]
		20.67	20.67	20.67	90.00	[25]
		20.70	20.70	20.70	90.00	[25]
		20.96	20.96	20.96	90.00	[25]
		20.74	20.74	20.74	90.00	[25]
		20.76	20.76	20.76	90.00	[25]
		20.76	20.76	20.76	90.00	[25]
		20.82	20.82	20.82	90.00	[25]
		20.84	20.84	20.84	90.00	[25]
MOF-74	MTP	26.17	26.17	6.75	120.00	
	VASP MLP	26.22	26.16	6.74	120.00	
	exp.	25.93	25.93	6.80	120.00	[8]
		25.93	25.93	6.84	120.00	[26]
		25.93	25.93	6.83	120.00	[27]
		26.16	26.16	6.83	120.00	[28]
MIL-53 (lp)	MTP	17.18	6.69	12.42	90.00	
	VASP MLP	17.37	6.69	12.12	90.00	
	exp.	16.76	6.64	12.84	90.00	[11]
		16.76	6.63	12.79	90.00	[29]
		16.73	6.63	12.84	90.00	[30]
MIL-53 (np)	MTP	19.96	6.67	6.81	105.35	
	VASP MLP	19.90	6.66	6.83	105.24	
	exp.	20.76	6.61	7.06	113.58	[11]
		19.51	6.58	7.61	104.24	[12]
		19.63	6.56	7.16	104.7	[30]

Supplementary Table 19: Thermal expansion coefficients for the investigated systems obtained from NPT simulations performed with the MTPs and VASP MLPs compared to experimental values at room temperature.

System	Method	α_a / 10^{-6} K^{-1}	α_b / 10^{-6} K^{-1}	α_c / 10^{-6} K^{-1}	α_V / 10^{-6} K^{-1}
MOF-5	MTP	-11.0	-11.3	-11.6	-33.9
	VASP MLP	-11.7	-11.3	-11.6	-35.0
UiO-66	MTP	2.2	2.4	2.1	6.7
	VASP MLP	1.4	2.1	1.9	5.4
MOF-74	MTP	1.6	1.2	31.0	33.6
	VASP MLP	6.9	2.0	25.3	31.2
MIL-53 (lp)	MTP	-15.5	4.4	40.2	28.6
	VASP MLP	9.3	0.5	-24.9	-15.7
MIL-53 (np)	MTP	21.7	5.3	124.7	137.2
	VASP MLP	29.6	4.1	80.0	101.3

Supplementary Note 4.1.4 Multi-process scaling of the VASP MLPs and of the MTPs

In the main manuscript, we discussed in the context of Fig. 7 that a major reason why the VASP MLPs are slower than the MTPs is the unfavorable parallelization over multiple cores of a processing node. The values in the main manuscript are given for 64 cores contained in a dual socket AMD EPYC 7713 (Milan) node (i.e., using half of the node). Supplementary Figure 33 shows the required CPU time of a level 22 MTP, a VASP MLP trained with 4 atom types and a VASP MLP trained with 7 atom types for a 4×4×4 conventional supercell of MOF-5 (27136 atoms) starting from just using a single core up to the full 128 cores available on each node. It can be seen that in this comparison the difference in speed between the VASP MLPs and the MTPs is rather small (VASP MLPs with 4 species only about 50% slower) up to using 16 cores. However, the performance deteriorates for more cores. The improved parallelization of LAMMPS when performing molecular dynamics simulations is not particularly surprising, as the code was developed solely for performing such calculations. Conversely, VASP is primarily a DFT code. In passing we note that especially for the VASP MLP the scaling gets even much worse when using smaller unit cells.



Supplementary Figure 33: CPU time per step and atom as a function of the number of cores used tested. The tests have been performed for a level 22 MTP in a, a VASP MLP using 4 different atom types in b and a VASP MLP using 7 different atom types in c. The test was performed on a dual socket AMD EPYC 7713 (Milan) node for a 4×4×4 supercell of the conventional MOF-5 unit cell comprising 27136 atoms in an NPT ensemble.

Supplementary Note 4.2 Thermal conductivity

Supplementary Note 4.2.1 Non-equilibrium molecular dynamics

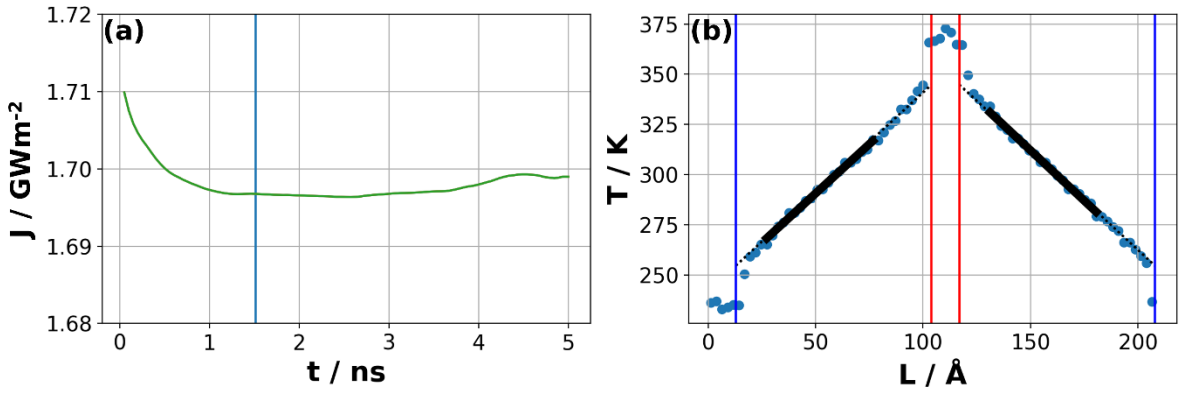
To compute the thermal conductivity, non-equilibrium molecular dynamics (NEMD) simulations were performed for MOF-5. Before the NEMD simulations were conducted, the simulation box, which is based on the unit cell parameters at 300 K obtained from the previously described NPT simulations, was equilibrated at the target temperature of 300 K in an NVT ensemble for 25.5 ps (using a Langevin thermostat with a damping parameter of 100 fs). Afterwards, additional time steps were performed until the concurrent time step showed a total energy corresponding to the average energy during the preceding 25 ps. This is done to prevent any adverse effects due to temperature oscillations caused by the thermostat and to precisely reach a temperature of 300 K for the rest of the simulation. Then, the thermostat was turned off and the simulation was continued in an NVE ensemble. To perform the actual NEMD simulation, a temperature gradient was imposed on a super cell of the material substantially elongated in one direction which leads to a heat flux from the hot to the cold region. The thermal conductivity can then be obtained after reaching steady state by employing Fourier's law. Due to the limited size of simulation boxes, there are still significant finite size effects present in this type of simulation, massively affecting the result [31]. These can be corrected to some extent by performing NEMD simulations for several different cell lengths to obtain a length-dependent thermal conductivity. The inverse of that length-dependent thermal conductivity can then be related to the inverse of the length. The thermal conductivity in the infinite size limit is then obtained by performing a linear fit through the available data points and obtaining the intercept value at an inverse length of 0.

As described already in the main manuscript, due to the sizable computational cost of the slowly converging NEMD simulations, lower level MTP (level 18) was used to carry out the entire finite size correction procedure. To show that the results for the slightly less accurate potential are consistent, three cell sizes were also computed using the same level 22 MTPs, which were used in the other parts of in this work. The required temperature gradients were created employing the Müller-Plathe heat exchange algorithm [32]. Here, the simulation box is divided into N slabs that correspond to twice the number of conventional unit cells of MOF-5 so that each slaps encompasses exactly 1 node and 1 linker. The first slab defines the cold region and the (N/2+1)st slab defines the hot region. Periodic boundary conditions were applied in all directions. During the simulation every 4800 time steps the kinetic energies of the 16 slowest (coldest) atoms in the hot slab were swapped with the kinetic energies of the 16 fastest (hottest) atoms in the cold slab. The reason for the infrequent energy swapping is to minimize the slight energy drift emerging throughout the simulation. A center-of-mass motion emerging in the system due to the arbitrary directions of the modified velocities of the atoms due to the Müller-Plathe algorithm is prevented by uniformly rescaling the velocities of all the atoms every 300 time steps while maintaining the relative velocities of the atoms relative to each other. The simulations were then performed for 5 ns in which the temperature profile over the simulation box and also the exchanged energy was recorded. The temperature profile and average heat flux for such a run using the level 18 MTP for an 8×2×2 supercell of the conventional unit cell (207×52×52 Å) of MOF-5 is depicted in Supplementary Figure 34. The heat flux, obtained from the exchanged energy from the Müller-Plathe algorithm, was averaged starting at the end of the total simulation time (so that, for example, the value at 10 % of the time is averaged over the remaining 90 %) to visualize and allow the easy exclusion of the initial time period, where the simulation has not reached steady state yet. This can clearly be seen by the different slope of the heat flux within the first ns of simulation time. Then, the actual value of the heat flux was determined by averaging over the last 3.5 ns of the simulation time. Supplementary Figure 34 (b) also reveals a step in temperatures at the boundaries between the thermalized regions and the bulk of the material. This arises due to phonon scattering at these boundaries. To obtain the thermal conductivity there are two approaches to determine the temperature gradient: to use the temperature gradient in the linear region in the bulk of the material or to use the absolute temperature difference between the thermalized regions. The former method has been traditionally employed in the majority of the literature [31] while the latter has recently been shown to be more consistent with results from more sophisticated methodologies that do not suffer from scattering at the thermostat boundaries, like the HNEMD (homogeneous non-equilibrium molecular dynamics) method [33]. This approach was tested for a slightly different style of NEMD

simulation using two thermostats (Langevin or Nosé-Hoover) and was shown to be particularly accurate with the Langevin thermostat. However, there is currently a technical issue of the implementation of the energy tallying using the MLIP-lammps interface in conjunction with the Langevin thermostats leading to nonsensical energy tallying in LAMMPS. Therefore, as described above, the Müller-Plathe algorithm was instead chosen to generate the temperature difference. Since the scattering effects at the thermostat boundaries were similar as what we already observed for the Langevin approach using traditional force field potentials [34], we still started by using the method of Li et al. [33] to obtain the temperature gradient, even though that method was specifically developed for Langevin or Nosé-Hoover thermostats.

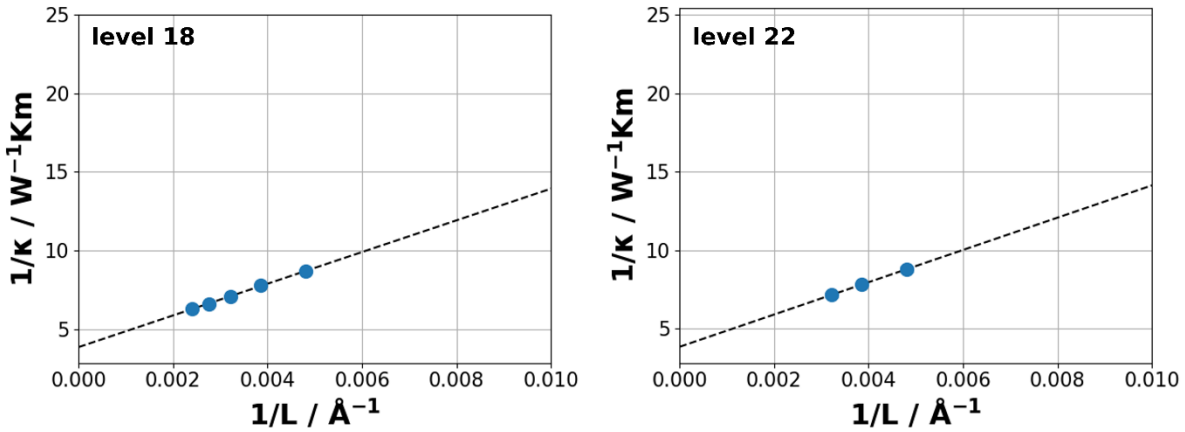
Using this approach, we obtained the thermal conductivity values for the level 18 and level 22 MTPs, whose inverse are shown in Supplementary Figure 35 in tandem with the extrapolation to the infinite size limit. This led to a thermal conductivity of 0.260 ± 0.011 $W(mK)^{-1}$ for the level 18 MTP and 0.261 ± 0.002 $W(mK)^{-1}$ for the level 22 MTP. Here, the provided error values reflect only the statistical standard error of the linear fit and should by no means be treated as the total error of the simulation. These values are somewhat below the experimental value of 0.32 $W(mK)^{-1}$ [35] and also below the value of 0.29 $W(mK)^{-1}$ we obtained with our traditional MOF-FF potentials [34].

A possible reason for the underestimation of the thermal conductivity in the present calculations could be that the approach by Li et al. [33] is not well justified for the Müller-Plathe algorithm. Therefore, as a second approach, we also determined the temperature gradient in the region of constant slope in the plots of the position dependent averaged temperatures away from the thermalized regions, which has been the more traditional approach for NEMD simulations. For the level 18 MTP this resulted in a finite size corrected thermal conductivity of 0.32 ± 0.02 $W(mK)^{-1}$ and for the level 22 MTP of 0.33 ± 0.02 $W(mK)^{-1}$. Again, both MTPs agree well with each other, and the values are now extremely close to the experimentally obtained value at room temperature.



Supplementary Figure 34: Evaluation of a NEMD simulation for a 8×2×2 supercell of MOF-5 using the level 18 MTP trained on the initial reference data set. The evolution of the averaged heat flux as a

function of time is depicted as the green line in a, where here the average was formed over all values starting at the end of the simulation (so that, for example, the value at 10 % of the time is averaged over the remaining 90 %). This serves to easily determine the amount of time it takes to reach the steady state. The blue vertical line indicates the time at which the averaged heat flux value was taken that was actually used for the evaluation of the thermal conductivity in Fourier's law (i.e., the heat flux was averaged between 1.5 ns and 5ns). In b the time averaged temperature profile is shown. The red vertical lines enclose the hot region and the blue vertical lines the cold region where the Müller-Plathe algorithm was invoked.



Supplementary Figure 35: Extrapolation of the inverse thermal conductivity to the infinite size limit for MOF-5 using level 18 and level 22 MTPs to carry out the individual NEMD simulations. The supercells used for the level 18 MTP are 8, 10, 12, 14, 16×2×2 and for the level 22 MTP 8, 10, 12×2×2 times the conventional cubic unit cells of MOF-5.

Supplementary Note 4.2.2 Approach to equilibrium molecular dynamics

Additionally, approach to equilibrium molecular dynamics (AEMD) simulations were performed for a level 18 MTP of MOF-5. Conceptually, the approach is rather similar: one uses long supercells (we used conventional supercells of MOF-5 ranging from 4×2×2 to 20×2×2) of which one half is heated, while the other half is cooled down compared to the target temperature. Convergence of the AEMD simulations was reached (meaning no difference in the evaluated thermal diffusivity as a function of time) at a simulation time of 1 ns for all system lengths. For cells smaller than 35 nm, we considered the temperature evolution over 0.5 ns, as the equilibrium was reached substantially faster. The instantaneous temperature difference ΔT between the halves of the system is tracked as a function of time t. Generally, the temperatures were computed classically from the average kinetic energies of

the atoms contained in the respective halves. This evolution is plotted in Supplementary Figure 36. Subsequently the following function is fitted to obtain the thermal diffusivity $\bar{\kappa}$[36]:

$$\Delta T(t) = \sum_{n=1}^{\infty} C_n e^{-\alpha_n^2 \bar{\kappa} t}$$

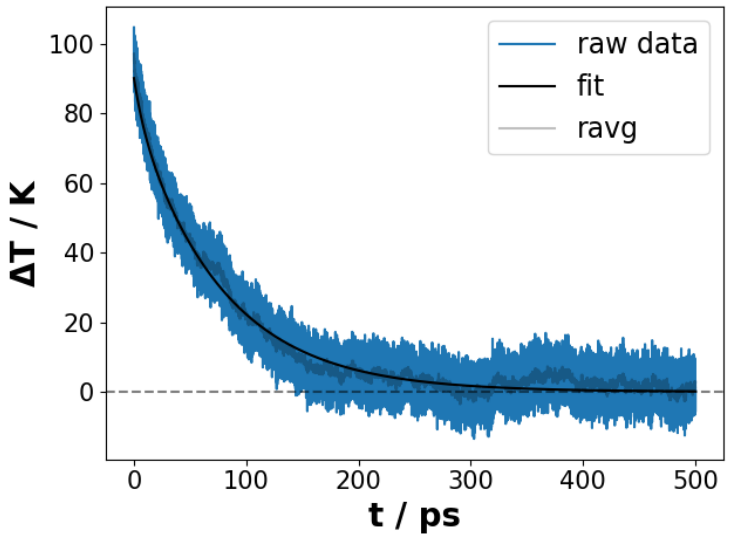
with

$$C_n = 8(T_1 - T_2)\frac{\left[\cos\left(\frac{\alpha_n L_z}{2}\right) - 1\right]^2}{\alpha_n^2 L_z^2}$$

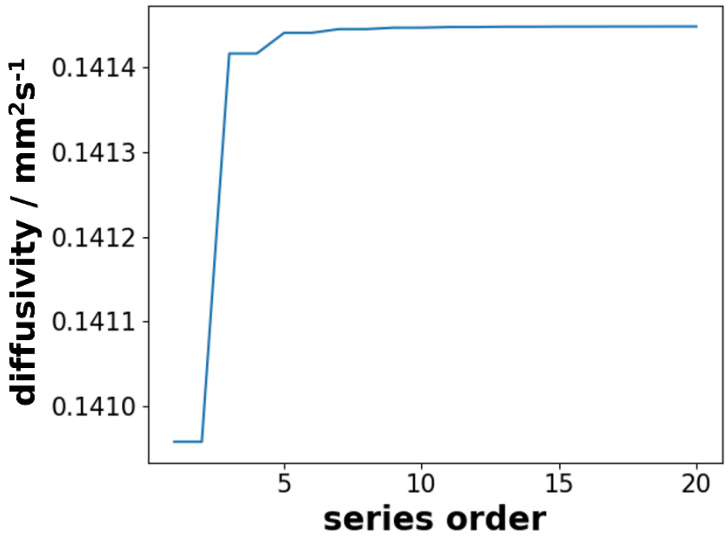
and

$$\alpha_n = \frac{2\pi n}{L_z}.$$

Here, L_z is the length of the unit cell, T_1 and T_2 are the starting temperatures of each half of the unit cell and n is the order of the exponential. Since it is impossible to fit this curve up to an infinite order, it is important to check the convergence of the series expansion as is done in Supplementary Figure 37. There, one can see that the dependence on exponential order is extremely small. Nonetheless, we use an order of 10 (i.e., including n between 1 and 10) for all our calculations to be on the safe side.



Supplementary Figure 36: Evolution of the temperature difference of the hot half and the cold half of a system during an AEMD simulation for an 8×2×2 conventional supercell of MOF-5. The blue line represents the raw data tracked at every time step, the grey line is the running average over the raw data using the closest 500 data points and the black line is the fit to obtain the thermal diffusivity.



Supplementary Figure 37: Dependence of the diffusivity for an 8×2×2 conventional supercell of MOF-5 on the exponential series order used to perform the fit over the temperature difference in an AEMD simulation.

From the fit, we obtain the thermal diffusivity $\bar{\kappa}$ for each supercell. From the thermal diffusivity, the thermal conductivity κ can be obtained via

$$\kappa = \frac{\bar{\kappa} C_p}{V},$$

with C_p being the heat capacity and V the volume of the supercell. Since we are doing classical MD simulations, we will always be in the Dulong-Petit limit, so the heat capacity at constant pressure when neglecting thermal expansion is

$$C_p \approx 3 N_A k_B$$

With k_B the Boltzmann constant and N_A the number of atoms in the system. Technically quantum corrections can now be applied due to the limited phonon occupation at the target temperature. The Dulong-Petit limit is definitely far from being reached in any MOF at 300 K. However, only the low frequency modes significantly contribute to the thermal conductivity due to the generally flat bands in the high frequency region (meaning a low group velocity) and the massive decay of phonon lifetimes at higher frequencies in many materials. Such a decay has previously observed in MOFs [37]. As it is mostly the high-frequency phonons that would be affected by quantum effects at higher temperatures, employing purely classical considerations appears reasonable.

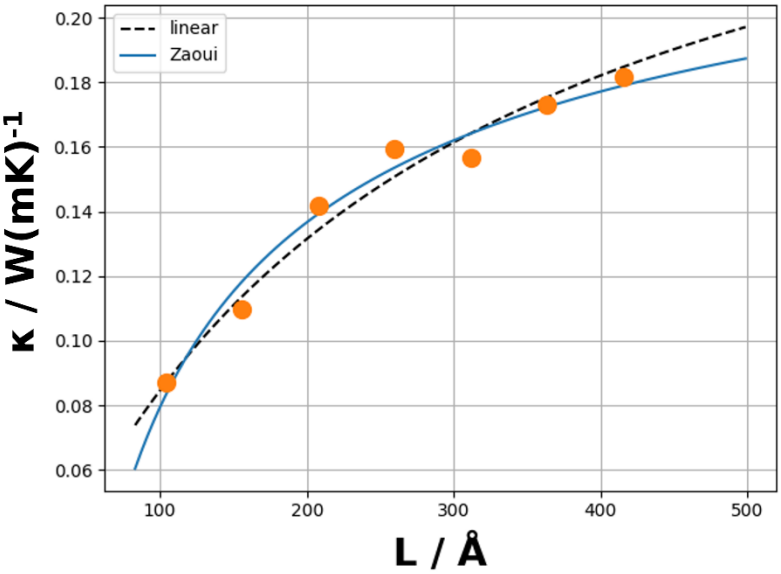
As a final step, also in the AEMD simulations finite size corrections have to be employed. The linear finite size correction we used in NEMD [31],

$$\frac{1}{\kappa(L)} = \frac{1}{\kappa_{\infty,linear}} \left(1 + \frac{\lambda}{L}\right)$$

where L is the length of the simulation box in heat transport direction, λ is a fitting parameter and κ_∞ is the thermal conductivity in the infinite size limit, was originally justified due to scattering at the thermostat boundaries. However, there are no thermostat boundaries in an AEMD simulation. Therefore, Zaoui et al. developed a new finite size correction scheme[38], which is derived from the Boltzmann transport equation:

$$\kappa(L) = \kappa_{\infty,Zaoui}\left(1 - \sqrt{\frac{\Lambda}{L}}\right).$$

Here Λ is also a fitting parameter. We employed this approach to obtain the thermal conductivity in the infinite size limit and the fitting curve with all the data points for different supercell sizes can be seen in Supplementary Figure 38. Here, we also added the linear fitting approach in the visualization. The results are actually extremely similar with $\kappa_{\infty,linear} = 0.302$ W(mK)$^{-1}$ and $\kappa_{\infty,linear} = 0.317$ W(mK)$^{-1}$.



Supplementary Figure 38: Extrapolation of the thermal conductivity to the infinite size limit for the AEMD simulations of MOF-5 using the (inversely) linear extrapolation approach (black dashed line) and the approach by Zaoui et al. which was specifically developed for AEMD (blue line). The orange points represent the results from the individual AEMD runs.

Supplementary Note 4.2.3 Widths of the supercells in the heat-transport simulations

In the above discussion, we were primarily concerned with the effect of the cell length. However, the cell thickness perpendicular to the heat transport direction is also important to converge NEMD and AEMD simulations [34]. We already discussed the convergence for NEMD and MOF-5 in [34]. There we found that using ×2×2 conventional supercells (×52×52 Å) is sufficient. However, we are now using

a different force field and also the alternative AEMD approach. It would be reasonable to assume that AEMD shows the same convergence behavior as NEMD due to the similarity of the approaches. Additionally, since the MTPs used in this work actually have a shorter cutoff for the atom-atom interactions than the MOF-FF potentials used in ref. [34], it is unlikely that the difference in convergence behavior is very different. Especially given that the MOF-FF potentials already modeled the phonons in MOF-5 very well.

Nonetheless, to be on the safe side, we perform some convergence tests for using thicker cells perpendicular to the transport directions in AEMD simulations. Supplementary Table 20 shows the thermal conductivities from AEMD simulations carried out on 16×2×2, 16×4×4, 32×2×2 and 32×4×4 supercells. For both of the cell lengths the difference in thermal conductivities between the thinner and thicker cells is extremely small. This is why we conclude that a 52 Å thick 2×2 cell is also sufficient for the AEMD simulations for MOF-5.

However, it should be mentioned that recently Green-Kubo simulations obtained a much higher value for the thermal conductivity of MOF-5 with 0.61 $W(mK)^{-1}$ [39] also using highly accurate machine learned potentials. There, they needed to use 5×5×5 supercells to achieve convergence, but this was the case for Green-Kubo (equilibrium molecular dynamics) [31] and homogeneous non-equilibrium molecular dynamics [40] approaches, which are fundamentally different from NEMD and AEMD simulations. As can be seen in Supplementary Table 20, the difference in the results between their and our work cannot be explained by different cell thicknesses. The exact origin of this difference needs to be investigated further and it could be connected to difference between non-equilibrium and equilibrium based methods.

Supplementary Table 20: Thermal conductivities κ obtained from AEMD simulations for MOF-5 for different supercell lengths $n_{fluxdir}$ (in Å: $L_{fluxdir}$) in heat transport direction and for different supercell thicknesses $n_{perpend.}$ (in Å: $L_{perpend}$) in both perpendicular directions.

$n_{fluxdir}$	$L_{fluxdir}$ / Å	$n_{perpend.}$	$L_{perpend.}$ / Å	κ / $W(mK)^{-1}$
16	415.5	2	51.9	0.182
16	415.5	4	103.9	0.185
32	831.0	2	51.9	0.240
32	831.0	4	103.9	0.239

Supplementary References

1. Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558–561, doi:10.1103/PhysRevB.47.558.
2. Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50, doi:10.1016/0927-0256(96)00008-0.
3. 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.* **1996**, *54*, 11169–11186, doi:10.1103/PhysRevB.54.11169.
4. 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.* **2010**, *132*, 154104, doi:10.1063/1.3382344.
5. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465, doi:10.1002/jcc.21759.
6. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775, doi:10.1103/PhysRevB.59.1758.
7. 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.* **2021**, *125*, 24728–24745, doi:10.1021/acs.jpcc.1c07882.
8. Kamencek, T.; Schrode, B.; Resel, R.; Ricco, R.; Zojer, E. Understanding the Origin of the Particularly Small and Anisotropic Thermal Expansion of MOF-74. *Adv. Theory Simulations* **2022**, *5*, 2200031, doi:10.1002/adts.202200031.
9. Vandenhaute, 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.* **2023**, *9*, 053804, doi:10.1038/s41524-023-00969-x.
10. Ortiz, A.U.; Boutin, A.; Fuchs, A.H.; Coudert, F.X. Anisotropic elastic properties of flexible metal-organic frameworks: How soft are soft porous crystals? *Phys. Rev. Lett.* **2012**, *109*, 1–5, doi:10.1103/PhysRevLett.109.195502.
11. Liu, Y.; Her, J.H.; Dailly, A.; Ramirez-Cuesta, A.J.; Neumann, D.A.; Brown, C.M. Reversible structural transition in MIL-53 with large temperature hysteresis. *J. Am. Chem. Soc.* **2008**, *130*, 11813–11818, doi:10.1021/ja803669w.
12. Loiseau, T.; Serre, C.; Huguenard, C.; Fink, G.; Taulelle, F.; Henry, M.; Bataille, T.; Férey, G. A Rationale for the Large Breathing of the Porous Aluminum Terephthalate (MIL-53) Upon

Hydration. *Chem. - A Eur. J.* **2004**, *10*, 1373–1382, doi:10.1002/chem.200305413.

13. Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **2015**, *108*, 1–5, doi:10.1016/j.scriptamat.2015.07.021.
14. Wieser, S.; Kamencek, T.; Dürholt, J.P.; Schmid, R.; Bedoya-Martínez, N.; Zojer, E. Identifying the Bottleneck for Heat Transport in Metal–Organic Frameworks. *Adv. Theory Simulations* **2021**, *4*, 2000211, doi:10.1002/adts.202000211.
15. Pick, R.M.; Cohen, M.H.; Martin, R.M. Microscopic theory of force constants in the adiabatic approximation. *Phys. Rev. B* **1970**, *1*, 910–920, doi:10.1103/PhysRevB.1.910.
16. Gonze, X.; Charlier, J.C.; Allan, D.C.; Teter, M.P. Interatomic force constants from first principles: The case of α-quartz. *Phys. Rev. B* **1994**, *50*, 13035–13038, doi:10.1103/PhysRevB.50.13035.
17. Gonze, X.; Lee, C. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1997**, *55*, 10355–10368, doi:10.1103/PhysRevB.55.10355.
18. Jinnouchi, R.; Asahi, R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. *J. Phys. Chem. Lett.* **2017**, *8*, 4279–4283, doi:10.1021/acs.jpclett.7b02010.
19. Jinnouchi, R.; Karsai, F.; Kresse, G. On-the-fly machine learning force field generation: Application to melting points. *Phys. Rev. B* **2019**, *100*, 014105, doi:10.1103/PhysRevB.100.014105.
20. 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.* **2008**, *114*, 16181–1686, doi:10.1107/s0108767308095950.
21. 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 Zn4 O (1,4-benzenedicarboxylate)3. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2008**, *78*, 054114, doi:10.1103/PhysRevB.78.054114.
22. Burtch, N.C.; Baxter, S.J.; Heinen, J.; Bird, A.; Schneemann, A.; Dubbeldam, D.; Wilkinson, A.P. Negative Thermal Expansion Design Strategies in a Diverse Series of Metal–Organic Frameworks. *Adv. Funct. Mater.* **2019**, *29*, 1904669, doi:10.1002/adfm.201904669.
23. Cavka, J.H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K.P. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *J. Am. Chem. Soc.* **2008**, *130*, 13850–13851, doi:10.1021/ja8057953.
24. Yot, P.G.; Yang, K.; Ragon, F.; Dmitriev, V.; Devic, T.; Horcajada, P.; Serre, C.; Maurin, G. Exploration of the mechanical behavior of metal organic frameworks UiO-66(Zr) and MIL-125(Ti) and their NH2 functionalized versions. *Dalt. Trans.* **2016**, *45*, 4283–4288,

doi:10.1039/c5dt03621f.

25. Germann, L.S.; Katsenis, A.D.; Huskić, I.; Julien, P.A.; Užarević, K.; Etter, M.; Farha, O.K.; Friščić, T.; Dinnebier, R.E. Real-Time in Situ Monitoring of Particle and Structure Evolution in the Mechanochemical Synthesis of UiO-66 Metal-Organic Frameworks. *Cryst. Growth Des.* **2020**, *20*, 49–54, doi:10.1021/acs.cgd.9b01477.
26. Rosi, N.L.; Kim, J.; Eddaoudi, M.; Chen, B.; O'Keeffe, M.; Yaghi, O.M. Rod packings and metal-organic frameworks constructed from rod-shaped secondary building units. *J. Am. Chem. Soc.* **2005**, *127*, 1504–1518, doi:10.1021/ja045123o.
27. 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* **2015**, *7*, 3719–3724, doi:10.1002/cctc.201500617.
28. 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.* **2021**, *319*, 110973, doi:10.1016/j.micromeso.2021.110973.
29. Nanthamathee, C.; Ling, S.; Slater, B.; Attfield, M.P. Contradistinct thermoresponsive behavior of isostructural MIL-53 type metal-organic frameworks by modifying the framework inorganic anion. *Chem. Mater.* **2015**, *27*, 85–95, doi:10.1021/cm503311x.
30. Yot, P.G.; Boudene, Z.; Macia, J.; Granier, D.; Vanduyfhuys, L.; Verstraelen, T.; Van Speybroeck, V.; Devic, T.; Serre, C.; Férey, G.; et al. Metal–organic frameworks as potential shock absorbers: The case of the highly flexible MIL-53(Al). *Chem. Commun.* **2014**, *50*, 9462–9464, doi:10.1039/c4cc03853c.
31. 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.* **2002**, *65*, 144306, doi:10.1103/PhysRevB.65.144306.
32. Müller-Plathe, F. A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. *J. Chem. Phys.* **1997**, *106*, 6082–6085, doi:10.1063/1.473271.
33. Li, Z.; Xiong, S.; Sievers, C.; Hu, Y.; Fan, Z.; Wei, N.; Bao, H.; Chen, S.; Donadio, D.; Ala-Nissila, T. Influence of thermostatting on nonequilibrium molecular dynamics simulations of heat conduction in solids. *J. Chem. Phys.* **2019**, *151*, 234105, doi:10.1063/1.5132543.
34. 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* **2022**, *12*, 2142, doi:10.3390/nano12132142.
35. Huang, B.L.; Ni, Z.; Millward, A.; McGaughey, A.J.H.; Uher, C.; Kaviany, M.; Yaghi, O. Thermal conductivity of a metal-organic framework (MOF-5): Part II. Measurement. *Int. J. Heat Mass*

Transf. **2007**, *50*, 405–411, doi:10.1016/j.ijheatmasstransfer.2006.10.001.

36. Melis, C.; Dettori, R.; Vandermeulen, S.; Colombo, L. Calculating thermal conductivity in a transient conduction regime: theory and implementation. *Eur. Phys. J. B* **2014**, *87*, 96, doi:10.1140/epjb/e2014-50119-0.
37. Wang, X.; Guo, R.; Xu, D.; Chung, J.; Kaviany, M.; Huang, B. Anisotropic Lattice Thermal Conductivity and Suppressed Acoustic Phonons in MOF-74 from First Principles. *J. Phys. Chem. C* **2015**, *119*, 26000–26008, doi:10.1021/acs.jpcc.5b08675.
38. Zaoui, H.; Palla, P.L.; Cleri, F.; Lampin, E. Length dependence of thermal conductivity by approach-to-equilibrium molecular dynamics. *Phys. Rev. B* **2016**, *94*, 054304, doi:10.1103/PhysRevB.94.054304.
39. Ying, P.; Liang, T.; Xu, K.; Zhang, J.; Xu, J.; Zhong, Z.; Fan, Z. Sub-micrometer phonon mean free paths in metal-organic frameworks revealed by machine-learning molecular dynamics simulations. *Appl. Mater. Interfaces* **2023**, *15*, 36412–36422, doi:10.1021/acsami.3c07770.
40. Fan, Z.; Dong, H.; Harju, A.; Ala-Nissila, T. Homogeneous nonequilibrium molecular dynamics method for heat transport and spectral decomposition with many-body potentials. *Phys. Rev. B* **2019**, *99*, 1–9, doi:10.1103/PhysRevB.99.064308.