

Supplementary Information: Machine Learned Potential for High-Throughput Phonon Calculations of Metal–Organic Frameworks

Alin Marin Elena^{1†}, Prathami Divakar Kamath^{2,3†}, Théo Jaffrelo Inizan^{2,3,4},
Andrew S. Rosen⁵, Federica Zanca¹, Kristin A. Persson^{2,3*}

¹Scientific Computing Department, Science and Technology Facilities Council, Keckwick Lane, Daresbury, WA4 1PT, Cheshire, UK.

²Department of Materials Science and Engineering, University of California, Berkley, CA 94720, California, USA.

³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, California, USA.

⁴Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, CA 94720, California, USA.

⁵Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, New Jersey, USA.

*Corresponding author(s). E-mail(s): kapersson@lbl.gov;

†These authors, ordered alphabetically, contributed equally. All others, except for the corresponding author*, are also ordered alphabetically.

Contents

Supplementary Note 1: Details of the 127 Curated Dataset and Sampling Strategy	2
Supplementary Note 2: DFT Data for training and benchmarking	4
Supplementary Note 3: Details of GFN1-xTB Calculations	5
Supplementary Note 3.1: Geometry Optimization	5
Supplementary Note 3.2: Phonon calculations with GFN1-xTB	5
Supplementary Note 4: Details of Benchmarking	6
Supplementary Note 4.1: Benchmarking of energy, forces and stress	6
Supplementary Note 4.2: Details of the out-of-sample 70 test structures	6
Supplementary Note 4.3: Benchmarking of vibrational properties	7
Supplementary Note 5: Details of capturing equilibrium positions and partial density of states	10
Supplementary Note 6: Computational efficiency of MACE-MP-MOF0	11

MOF IDs	N_{atoms}	Reduced Formula	Crystal System	Space Group	Synthesized
HKUST-1	156	Cu ₂ H ₂ (C ₃ O ₂) ₂	cubic	Fm-3m	TRUE
IRMOF-10	166	Zn ₄ H ₂₄ C ₄₂ O ₁₃	triclinic	P1	TRUE
IRMOF-14	190	Zn ₄ H ₂₄ C ₅₄ O ₁₃	cubic	Fm-3m	TRUE
JUK-8 (cp)	288	ZnH ₂₄ C ₃₄ N ₆ O ₇	monoclinic	C2/c	TRUE
MIL-125	240	Ti ₂ H ₇ (C ₄ O ₃) ₃	triclinic	P1	TRUE
MIL-53 (lp)	38	AlH ₅ C ₈ O ₅	orthorhombic	Imma	TRUE
MOF-5	106	Zn ₄ H ₁₂ C ₂₄ O ₁₃	cubic	Fm-3m	TRUE
MOF-650	142	Zn ₄ H ₁₈ C ₃₆ O ₁₃	trigonal	R3m	TRUE
MOF-74	54	ZnHC ₄ O ₃	trigonal	R-3	TRUE
NOTT-300	72	AlH ₄ C ₈ O ₅	tetragonal	I4 ₁ 122	TRUE
UiO-66	114	Zr ₃ H ₁₄ (C ₃ O ₂) ₈	cubic	F-43m	TRUE
UiO-67	174	Zr ₃ H ₂₆ (C ₂₁ O ₈) ₂	cubic	F23	TRUE
UiO66Zr-Br	114	Zr ₃ H ₁₁ C ₂₄ Br ₃ O ₁₆	triclinic	P1	TRUE
UiO66Zr-Cl	114	Zr ₃ H ₁₁ C ₂₄ Cl ₃ O ₁₆	triclinic	P1	TRUE
UiO66Zr-I	114	Zr ₃ H ₁₁ C ₂₄ I ₃ O ₁₆	triclinic	P1	TRUE
UiO66ZrCH ₃	132	Zr ₃ H ₂₀ C ₂₇ O ₁₆	triclinic	P1	TRUE
UiO66ZrNH ₂	126	Zr ₃ H ₁₇ C ₂₄ N ₃ O ₁₆	triclinic	P1	TRUE
UiO66ZrOH	120	Zr ₃ H ₁₄ C ₂₄ O ₁₉	triclinic	P1	TRUE
UiO66ZrSH	120	Zr ₃ H ₁₄ C ₂₄ S ₃ O ₁₆	triclinic	P1	TRUE

Supplementary Table 1: Some of the experimentally widely studied MOFs

Supplementary Note 1: Details of the 127 Curated Dataset and Sampling Strategy

The curated dataset includes 19 MOFs that are widely experimentally studied, tabulated in Supplementary Table 1. In addition to these, we sampled 108 more MOFs from the QMOF database [1, 2] using a farthest point sampling (FPS) approach. The MACE descriptors (invariant= False) are computed for all the atoms of the unit cell. For each structure, three descriptors were computed, per atom, per species and per system and each of them were used for selection with FPS. The atom-wise descriptors are grouped into average element-wise descriptors in the MOF. The total descriptor for MOF is thus obtained by averaging over the element-wise descriptors. The FPS strategy is used on this total descriptor to sample diverse MOFs. This strategy selects the first MOF by randomly selecting a total descriptor, and the next descriptor which is the farthest in magnitude to the initial descriptor is selected. This is done in an iterative way to sample the 108 diverse MOFs.

The computed total MACE descriptors along with the key properties of these MOFs are summarized in Supplementary Table 2. Supplementary Figure 1 shows the diverse functional groups covered in the linkers of the 127 MOFs in the curated dataset. The reported symmetries for the MOFs are obtained with pymatgen’s [3] space group analyzer using a symprec of 10^{-3} Å. The input structure files are uploaded in the github repository.

MOF IDs	N_{atoms}	Reduced Formula	Crystal System	Space Group	Synthesized	invariants false
qmof-008aa12	84	MgC ₇ H ₃ O ₃	trigonal	R3	FALSE	0.0003878775451
qmof-00f0897	32	ZnC ₆ H ₃ INO ₄	triclinic	P-1	FALSE	8.88E-05
qmof-0116630	50	AlC ₁₂ Cl ₃ O ₆ S ₃	hexagonal	P6	FALSE	0.001330784606
qmof-01be70a	84	ZnC ₁₂ F ₄ N ₄	tetragonal	I4/mmm	FALSE	0.0009686209828
qmof-02c07e3	44	AlC ₂₁ H ₁₂ NO ₉	trigonal	P3	FALSE	0.0009555107003
qmof-038aa2d	62	AlBr ₃ C ₁₅ H ₆ O ₆	trigonal	P321	FALSE	0.00077466085
qmof-03c6ba9	50	AlBr ₃ C ₁₂ O ₆ S ₃	hexagonal	P6	FALSE	0.001092108467
qmof-06caa98	150	InC ₁₂ H ₄ O ₈	cubic	I-43d	TRUE	0.0003928277946
qmof-09e46be	72	AlC ₈ H ₄ O ₅	tetragonal	I4 ₁ 122	TRUE	0.0006111207319
qmof-09f0270	56	Al ₂ C ₂₄ H ₃ N ₃ O ₁₈ S ₆	trigonal	P3	FALSE	0.001069140617
qmof-0c2559f	86	Zn ₃ C ₂₂ Cl ₄ H ₂ O ₁₂	orthorhombic	Imm2	FALSE	0.0005460350179
qmof-0de3389	76	GaC ₈ H ₅ O ₅	orthorhombic	Pnma	TRUE	0.0006280954695
qmof-1269b58	30	ZnC ₈ F ₆ H ₄ N ₄ O ₂ S ₄ Si	triclinic	P1	FALSE	0.001399723273
qmof-131c306	60	ZnC ₂₄ F ₆ H ₁₆ N ₄ O ₈ Si	tetragonal	P4/mmm	FALSE	0.0005893840842
qmof-1339fce	90	ZnC ₄ H ₄ N ₆	cubic	Fd-3m	TRUE	0.001576859357
qmof-15bd34b	86	AlBr ₃ C ₂₄ H ₉ O ₆	hexagonal	P6	FALSE	0.0005611541738
qmof-15f1bb8	70	Zn ₄ C ₁₂ H ₆ N ₆ O ₇	cubic	F-43m	TRUE	0.0008381884648
qmof-15fda22	96	ZnC ₃ H ₇ NO ₃ P	trigonal	P3 ₁ 21	TRUE	0.0006240746661

Continued on next page

MOF IDs	N_{atoms}	Reduced Formula	Crystal System	Space Group	Synthesized	invariants false
qmoF-1db67f1	106	Hf3C20H8N6O16	triclinic	P1	FALSE	0.0002671534533
qmoF-204707c	42	ZnC10H5NO4	monoclinic	P2	FALSE	0.0002486021949
qmoF-2063599	82	Zn4C18H6O13	orthorhombic	C222_1	TRUE	0.0003850523433
qmoF-2264c2b	72	MgC5HO5	trigonal	R-3	TRUE	0.000619171926
qmoF-2a5ad82	38	ZnC4H6N8	monoclinic	Cc	TRUE	0.001476508996
qmoF-2b7dbda	90	Zn4C20F8O13	tetragonal	I4/mmm	FALSE	0.0003990564989
qmoF-303620d	106	Zn4C24Cl12O13	cubic	Fm-3m	FALSE	0.0008028901748
qmoF-305bca1	106	Hf3C20H8N6O16	triclinic	P1	FALSE	0.000326737975
qmoF-311d91b	90	CdC4H4N6	cubic	Fd-3m	TRUE	0.001325784509
qmoF-32894ad	86	AlC21F12N3O6	hexagonal	P6	FALSE	0.001063528508
qmoF-3630fc5	154	Zn4C24N12O37	cubic	F432	FALSE	0.0008594648253
qmoF-38e57cf	41	AlC21Cl6H6NO6	trigonal	P312	FALSE	0.001253026183
qmoF-3b672a9	28	ZnC8F6H8N4Si	tetragonal	P4/mmm	TRUE	0.0008675236636
qmoF-3cd6c37	72	GaC8H4O5	tetragonal	I4_122	TRUE	0.0007119420077
qmoF-3ed2303	38	AlC8FH4O5	monoclinic	C2	FALSE	0.0006351898131
qmoF-3ed9eea	178	Zn4Br12C48H12O13	cubic	Fm-3m	FALSE	0.0003689764411
qmoF-4304595	76	SnC6H8N2S2	orthorhombic	Fdd2	TRUE	0.001054459923
qmoF-4333405	72	ZnC5H2N10	orthorhombic	P2_12_12_1	TRUE	0.001784799812
qmoF-456d03e	90	CeC22H14O8	monoclinic	C2	TRUE	0.0004221730703
qmoF-4a4ca58	94	AlC24H15N2O5	orthorhombic	Cmmm	TRUE	0.0008136870127
qmoF-4b5a943	90	MgC7H3O3S	trigonal	R3	FALSE	0.0005462764226
qmoF-4cd6b6d	96	Zn2C5H6N8O3	orthorhombic	Pnma	TRUE	0.001370777806
qmoF-50150b4	104	CeC24H16N07S3	monoclinic	Cc	TRUE	0.000686591613
qmoF-52ef7e9	86	AlC21H12N3O6	hexagonal	P6	FALSE	0.0008311548277
qmoF-536d29b	138	ZnC8H10N4	cubic	I-43m	TRUE	0.001000874524
qmoF-5414b76	92	InC24H14N2O5	orthorhombic	Cmmm	TRUE	0.0006981945426
qmoF-555ceef	162	Zn5C36Cl4H12N18O6	cubic	Fm-3m	TRUE	0.001307822367
qmoF-5763612	82	CdC18H12N6O4	orthorhombic	Pnc2	TRUE	0.0007274964476
qmoF-5a5fe74	102	Zn3Br8C24N4O12	tetragonal	I422	FALSE	0.0005463151798
qmoF-5b5489b	68	CeC9H16O8	triclinic	P-1	TRUE	0.0002145280049
qmoF-637e6e5	82	ZnBrC22H12NO4	orthorhombic	Pmmm	TRUE	0.0005622421611
qmoF-63c5667	178	Zn4C48F24O13	cubic	Fm-3m	FALSE	0.0005597239522
qmoF-697208e	96	ZnC48F6H24N8O8Si	tetragonal	P4/mmm	TRUE	0.0006395713449
qmoF-6f53f75	106	Zn4Br12C24O13	cubic	Fm-3m	FALSE	0.0003132854093
qmoF-700d2d4	99	InC16H8O8	hexagonal	P6_222	TRUE	0.0004171118097
qmoF-72b3537	49	AlC27F12H3O6	monoclinic	C2	FALSE	0.001035681318
qmoF-7601c86	184	CeC5H7O10	orthorhombic	Pbcn	TRUE	0.0008989056743
qmoF-7616b2d	41	Al2C18H6O12S3	trigonal	P321	FALSE	0.0008300985119
qmoF-7979b73	80	ZnC7H6N4O2	tetragonal	I-4	TRUE	0.0009141177655
qmoF-79d8511	96	ZnC5ClH3N4O2	trigonal	R-3	TRUE	0.001349068628
qmoF-7e6da38	49	Al2C22F2H10O11S2	triclinic	P1	FALSE	0.0007625644577
qmoF-7f34f70	68	ZnC24F6H24N12Si	tetragonal	P422	FALSE	0.001063045191v
qmoF-801a6e7	92	AlC24Cl12H3O6	hexagonal	P6	FALSE	0.00128350626
qmoF-810dc2c	108	CdC12H8N2O4	tetragonal	P4_12_12	TRUE	0.0004452133008
qmoF-82276a2	92	AlC8H5O9	orthorhombic	Pnma	FALSE	0.0006629947591
qmoF-8573227	46	AlC24Cl6H6N3O6	triclinic	P1	FALSE	0.00109575062
qmoF-86777db	76	AlC8H5O5	orthorhombic	Pnma	TRUE	0.0005206021693
qmoF-887c43e	56	AlC12H3O9S3	hexagonal	P6	FALSE	0.0008646717303
qmoF-8e45c8d	71	Al2C33H24O12	trigonal	P3	FALSE	0.0007121689446
qmoF-925a580	92	AlC21H9N3O12	hexagonal	P6	FALSE	0.0007245725753
qmoF-946ef24	100	Zn3C27Cl2H6O12	orthorhombic	Imm2	FALSE	0.0003497524204
qmoF-95aa0a6	90	MgC5H5O4	trigonal	R-3	TRUE	0.000308486483
qmoF-95ec286	40	Zn4C13H7N3O13	triclinic	P1	FALSE	0.0004281694265
qmoF-99abeac	86	AlC24Cl3H9O6	hexagonal	P6	FALSE	0.000692683487

Continued on next page

MOF IDs	<i>N_atoms</i>	Reduced Formula	Crystal System	Space Group	Synthesized	invariants false
qmof-9d582e8	124	CeC34H19O8	triclinic	P-1	TRUE	0.0006418395273
qmof-9e6c541	46	AlC10H7O5	orthorhombic	<i>I</i> 2_12_12_1	FALSE	0.0006480367019
qmof-a2d8489	86	AlC24Cl12O6	hexagonal	P622	FALSE	0.001242926984
qmof-a2d95c3	106	Zn4C24H12O13	cubic	Fm-3m	TRUE	0.0002865821869
qmof-a30571c	178	Zn4C48F12H12O13	cubic	Fm-3m	FALSE	0.0004293359544
qmof-aa922ec	54	MgC4HO3	trigonal	R-3	TRUE	0.0002297545827
qmof-ab8d1a0	77	Zn3C30F3H21N14O6	trigonal	R3m	TRUE	0.0008793604563
qmof-ab953dd	92	AlC24H12O9	hexagonal	P6	FALSE	0.0005889440273
qmof-b41b038	126	CdC10H6N4	cubic	I-43m	TRUE	0.0006402615622
qmof-b60a8d6	62	AlC12N3O12S3	hexagonal	P6	FALSE	0.001029449053
qmof-ba8f2fc	68	AlC15H9O6S3	hexagonal	P6	FALSE	0.0009007667261
qmof-bc64fc7	52	ZnC24F14H8N4Si	tetragonal	P4/ <i>mmm</i>	FALSE	0.0006891653851
qmof-c07a718	52	ZnC24F6H16N4Si	tetragonal	P4/ <i>mmm</i>	FALSE	0.0006675406518
qmof-c09dfd8	90	MgC4H4N6	cubic	Fd-3m	TRUE	0.00144323264
qmof-c2c0d1c	92	AlC21H15N3O6	hexagonal	P6	FALSE	0.0009696140125
qmof-c5776f5	124	CeC34H19O8	triclinic	P-1	TRUE	0.0006231984566
qmof-ce0e7d6	90	MgC7H3NO3	trigonal	R-3	FALSE	0.0005080620351
qmof-d08b9c2	80	InC8H6O5	tetragonal	<i>I</i> 4_122	TRUE	0.0003444877584
qmof-d1ccea	190	Zn4C54H24O13	cubic	Fm-3m	TRUE	0.0004732471874
qmof-db4f7fe	92	AlC24Cl3H12O6	hexagonal	P6	FALSE	0.000798310838
qmof-dbb18cc	178	Zn4C48Cl12H12O13	cubic	Fm-3m	FALSE	0.0005924072866
qmof-dbfac13	94	ZnC18H12N14O2	orthorhombic	Fddd	TRUE	0.001102626607
qmof-de28178	76	MgC8H5O5	tetragonal	<i>I</i> 4_122	TRUE	0.0004331245168
qmof-de5e312	154	Zn4C24H24N12O13	cubic	F432	FALSE	0.0009449258517
qmof-df82873	38	GaC8H5O5	orthorhombic	Imma	TRUE	0.0006108831426
qmof-e300fed	84	ZnC10H6O4	orthorhombic	Cmm2	FALSE	0.0004823973201
qmof-e77426e	78	Zn3C20H4O12	tetragonal	I422	FALSE	0.0001194943756
qmof-e8a0478	92	SrC2H8N8O4	orthorhombic	Pcca	TRUE	0.001200377815
qmof-ed20cf9	86	AlC21Cl6H6N3O6	trigonal	P3	FALSE	0.00111536554
qmof-f226470	102	Zn3C24H8O16	tetragonal	I422	FALSE	0.000296201794
qmof-f3f8bc5	86	AlC24F6H6O6	hexagonal	P622	FALSE	0.0006333304413
qmof-f4d2cb4	86	AlC24F12O6	hexagonal	P622	FALSE	0.0007630434156
qmof-f61d5ce	72	InC8H4O5	tetragonal	<i>I</i> 4_122	TRUE	0.0005064372208
qmof-f75eb5a	34	ZnC2H4N10	monoclinic	Cc	TRUE	0.001922275326
qmof-fbd20d4	41	AlC21Cl3H9NO6	trigonal	P3	FALSE	0.001033014473
qmof-fbd219f	47	AlC21H15N4O6	trigonal	P3	FALSE	0.00118114542

Supplementary Table 2: MOFs sampled from the QMOF database which are diverse in their nodes, linkers and topology

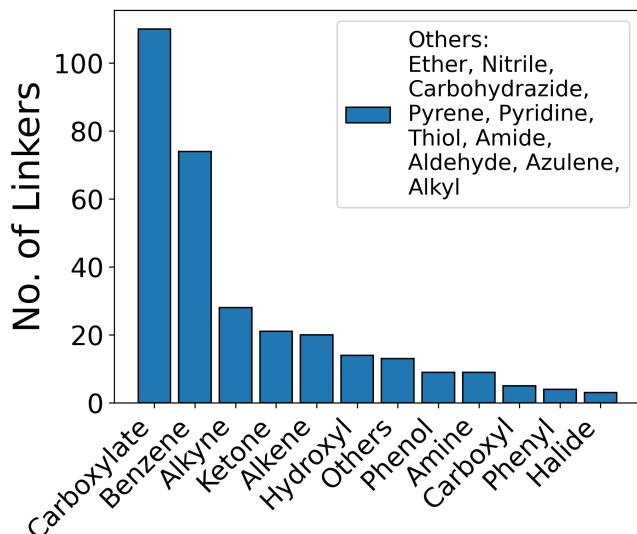
Supplementary Note 2: DFT Data for training and benchmarking

The DFT calculations for generating the training data were performed using the VASP package [4]. The MPGGA workflows in atomate2 [5] were used to perform static, relaxation, equation of state and atomic reference calculations to have the VASP input set compatible with the MPtrj dataset [6] with the following INCAR updates to the base input set.

```

user_incar_settings={
  "EDIFF": 1e-5,
  "LCHARG": False,
  "NELM": 200,
  "ISMEAR": 0,
  "LREAL": False,
  "IVDW": 12,
  "EDIFFG": -0.02 #for relaxation
}

```



Supplementary Figure 1: Diversity of the functional groups in the linkers of the 127 MOFs in the curated dataset

Symmetry Search Tolerance (Å)	Number of Symmetry Inequivalent Displacements
10^{-5}	318
10^{-4}	318
10^{-3}	67
10^{-2}	67
10^{-1}	67

Supplementary Table 3: The number of symmetry inequivalent displacements needed in the equilibrium structure depending on the symprec value used for identifying its symmetry.

The phonon DFT data used for MOF-5, MIL-53 (lp) and UiO-66 for benchmarking in the main text was obtained from the band.yaml files made public in [7]. The band.yaml for MOF-74 was generated from the force constants file made public in [8]. The functional used in the reference phonon data is consistent with the functional we used for training i.e. PBE +D3(BJ) [9, 10]. The band structures of MACE-MP-MOF0 models were generated along the same band path of the reference DFT data to show the overlap between the two methods.

Supplementary Note 3: Details of GFN1-xTB Calculations

Supplementary Note 3.1: Geometry Optimization

The lattice parameters reported in Table 2 of the main text with GFN1-xTB are obtained by performing a full cell relaxation on the input structures with the DFTB+ package [11] using the GFN1-xTB Hamiltonian and Rational optimizer until the maximum force component is 10^{-8} a.u. The space group of the resultant optimized geometry is determined with the default symprec of 10^{-5} Å with pymatgen for all methods in Table 1 of the main text. The change in the number of symmetry inequivalent displacements with loosening the symmetry search tolerance for phonon calculations with GFN1-xTB is reported for MOF-5 in Supplementary Table 3

Supplementary Note 3.2: Phonon calculations with GFN1-xTB

In order to preserve symmetry of the MOF, geometry optimization for phonon calculations with GFN1-xTB is done by only allowing the atoms in the unit cell to relax with the same convergence criteria. The resultant optimized geometry symmetry is identified with a symprec of 10^{-3} Å. The parameters for obtaining the phonons from the optimized structure with GFN1-xTB are the same as those used with MACE-MP-MOF0 workflow described in the main text.

Supplementary Note 4: Details of Benchmarking

Supplementary Note 4.1: Benchmarking of energy, forces and stress

The benchmarking of energies, forces and stress obtained with MACE-MP-MOF0 and MACE-MP-MOF0-v2 are in very good agreement with DFT for the training, test and validation sets indicated by the R2 score ≥ 0.99 in Supplementary Figure 2

The element-wise mean absolute errors (MAE) were obtained for all the 24 elements present in the curated dataset of 127 MOFs with the MACE-MP-0b and MACE-MP-MOF0-v2 models against DFT energies as shown in Supplementary Figure 3. We observe a systematic improvement in the MAE of the elements obtained with MACE-MP-MOF0-v2 by including a few reference configurations in the training dataset as compared to the MACE-MP-0b model. The errors with MACE-MP-0b are lower than the MACE-MP-MOF0-v2 predictions for Cu and I elements due to insufficient training data as there are 2 or less MOFs constituting these elements. Cu is also a magnetic element in HKUST-1, the only Cu-based MOF in the dataset. Therefore, the reference DFT data for Cu was generated with spin-polarization (ISPIN =2). This leads to reference configurations of the same MOF with different spin states for Cu, which leads to poor training of the model for spin-polarized systems. Hence, MLPs for spin-polarized systems need to be trained with care, and is not the focus of this work. The presented model, with a focus on closed-shell metal ion-based MOFs, shows a significant improvement in the MAE with MACE-MP-MOF0 model for the elements of interest in this work.

Supplementary Note 4.2: Details of the out-of-sample 70 test structures

The main characteristics of the additional 70 MOFs apart from the curated dataset of 127 structures sampled from the QMOF database are summarized in Supplementary Table 4 to test the out-of-sample errors for the models.

MOF IDs	N_{atoms}	Reduced Formula	Crystal System	Space Group	Synthesized
qmof-f3e009b	58	Zn4C12O13	triclinic	C2	FALSE
qmof-727c7c1	88	Zn4Br3C15H3N6O13	triclinic	P1	FALSE
qmof-574737f	38	Zn4H6N8	monoclinic	Cc	TRUE
qmof-293c249	67	Zn4C22H26O15	triclinic	P1	FALSE
qmof-5ad780f	50	Zn4C21H11NO13	triclinic	P1	FALSE
qmof-4513e39	49	Zn4C22H10O13	triclinic	P1	FALSE
qmof-f7ee91d	72	Zn4C33H19N3O13	triclinic	P1	FALSE
qmof-8ee1311	90	Zn4C43H29NO13	triclinic	P1	FALSE
qmof-831c36b	57	Zn4C27H12O14	triclinic	P1	FALSE
qmof-b839d0c	56	Zn4C27H12O13	triclinic	P1	FALSE
qmof-f23efa9	73	Zn4C34FH19N2O13	triclinic	P1	FALSE
qmof-d29cec2	70	Zn4C35H14O17	triclinic	P1	FALSE
qmof-78b934a	98	Zn4C24F4H4O13	monoclinic	C2/m	FALSE
qmof-8bfe542	58	ZnC14H5I4NO4	triclinic	P1	FALSE
qmof-0f8a04b	90	ZnC20H14N2O8	triclinic	P1	FALSE
qmof-9b1d1ca	62	ZnBr2C16H7NO4	triclinic	P1	FALSE
qmof-e6e911b	86	ZnC17H18NO6	triclinic	P1	FALSE
qmof-b98c123	54	ZnBr2C11H8NO4	triclinic	P1	FALSE
qmof-260d2c2	50	ZnC12H5I2NO4	triclinic	P1	FALSE
qmof-dca24c7	54	ZnC9H8N5O4	triclinic	P1	FALSE
qmof-0625810	40	ZnC10F2H2NO4	triclinic	P-1	FALSE
qmof-90a3f60	84	ZnC18H8N3O10S2	triclinic	P1	FALSE
qmof-4bdce90	80	ZnC17Cl2H14N2O4	triclinic	P1	FALSE
qmof-6c8f362	54	ZnC13H6NO6	triclinic	P1	FALSE
qmof-6822f6b	74	ZnC15H10I4NO6	triclinic	P1	FALSE
qmof-bc6d0a0	54	Zn2Br2C26F3H11N2O8	triclinic	P1	FALSE
qmof-c93885a	94	ZnC17H19N2O8	triclinic	P1	FALSE
qmof-3b418e4	92	ZnC23H14N3O5	triclinic	P1	FALSE
qmof-e252678	70	ZnC16Cl3H8N3O4	triclinic	P1	FALSE

Continued on next page

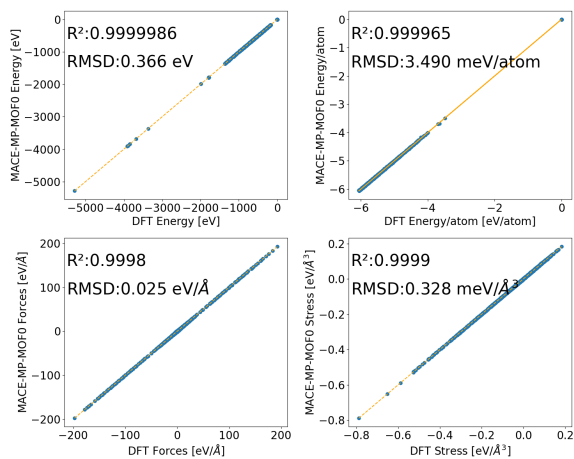
MOF IDs	N_{atoms}	Reduced Formula	Crystal System	Space Group	Synthesized
qmof-07e6831	67	Zn ₂ C ₃₅ H ₁₉ N ₃ O ₈	triclinic	P1	FALSE
qmof-fc68b48	50	Zn ₂ C ₂₄ FH ₁₃ N ₂ O ₈	triclinic	P1	FALSE
qmof-1b95eee	90	Zn ₃ C ₁₈ F ₂ H ₆ I ₄ O ₁₂	triclinic	C2/m	FALSE
qmof-2d1ab29	78	ZnC ₁₉ H ₉ I ₅ N ₄ O ₄	triclinic	P1	FALSE
qmof-8b33928	72	ZnC ₁₇ H ₉ N ₃ O ₆	triclinic	P-1	FALSE
qmof-abdc322	64	Al ₂ C ₁₂ H ₅ N ₁₀ O ₁₂	triclinic	P2 ₁	FALSE
qmof-2f5ad9e	99	Al ₄ C ₄₅ H ₂₈ N ₂ O ₂₀	triclinic	P1	FALSE
qmof-e53eab4	91	Al ₄ C ₃₇ H ₂₈ O ₂₂	triclinic	P1	FALSE
qmof-74ace65	98	Zn ₂ C ₂₂ Cl ₇ F ₇ H ₃ N ₈	triclinic	P1	FALSE
qmof-d38923a	54	ZnC ₂ FH ₂ N ₃	monoclinic	R-3c	TRUE
qmof-7a6a1bc	92	ZnC ₁₀ H ₆ N ₆	monoclinic	P2 ₁ /c	TRUE
qmof-cda9fd1	84	AlC ₈ H ₄ N ₅ O ₅ S ₂	triclinic	P1	FALSE
qmof-738d0ce	50	Al ₂ C ₂₄ ClF ₂ H ₁₁ O ₁₀	triclinic	P1	FALSE
qmof-87d1c3f	56	AlC ₁₂ H ₉ O ₆	triclinic	P-1	FALSE
qmof-38c31e8	48	AlC ₁₀ H ₅ N ₇ O ₇	triclinic	P-1	FALSE
qmof-ba7e598	60	AlC ₁₃ H ₉ N ₅ O ₅ S	triclinic	P1	FALSE
qmof-099c53d	60	AlC ₁₄ H ₉ N ₅ O ₅	triclinic	P1	FALSE
qmof-7559c55	56	AlBrC ₁₄ FH ₆ O ₅	triclinic	P1	FALSE
qmof-a5088d3	92	AlC ₁₀ F ₂ H ₅ O ₅	triclinic	P1	FALSE
qmof-7131987	52	AlC ₁₁ H ₇ O ₆ S	triclinic	P1	FALSE
qmof-a1e0b3b	44	AlC ₁₀ Cl ₃ H ₂ O ₅ S	triclinic	P1	FALSE
qmof-f944e8f	48	AlC ₁₀ FH ₅ N ₅ O ₅ S	triclinic	P1	FALSE
qmof-b2a2ee0	92	AlC ₂₄ Cl ₆ H ₉ O ₆	hexagonal	P6	FALSE
qmof-ef4a02d	92	AlC ₂₄ F ₁₂ H ₃ O ₆	hexagonal	P6	FALSE
qmof-6afa476	98	AlC ₂₁ H ₁₂ N ₃ O ₁₂	trigonal	P3	FALSE
qmof-916c1c5	86	AlC ₂₄ Cl ₆ H ₆ O ₆	triclinic	P622	FALSE
qmof-d9b204d	92	AlC ₂₁ H ₁₅ N ₃ O ₆	trigonal	P6	FALSE
qmof-b589eac	62	AlC ₁₂ H ₆ N ₃ O ₆ S ₃	hexagonal	P6	FALSE
qmof-0e578b6	74	AlC ₁₂ H ₃ O ₁₅ S ₆	triclinic	P6	FALSE
qmof-978268d	53	AlC ₂₁ H ₁₈ N ₇ O ₆	trigonal	P312	FALSE
qmof-4ede352	49	AlC ₂₇ Cl ₁₂ H ₃ O ₆	triclinic	C2	FALSE
qmof-9e4a54b	79	AlC ₄₅ F ₃ H ₂₄ O ₆	triclinic	P1	FALSE
qmof-233f752	65	AlC ₂₇ H ₆ N ₁₃ O ₁₈	triclinic	P1	FALSE
qmof-db2cdc7	77	AlC ₂₇ N ₁₉ O ₃₀	triclinic	P1	FALSE
qmof-38ae9d1	58	AlC ₂₄ H ₁₈ N ₉ O ₆	triclinic	P1	FALSE
qmof-bd7e56c	70	AlC ₂₄ H ₂₄ N ₁₅ O ₆	triclinic	P1	FALSE
qmof-c052686	72	Zn ₂ C ₃₆ Cl ₁₇ H ₇ N ₂ O ₈	triclinic	P1	FALSE
qmof-65294ad	86	ZnBr ₄ C ₂₀ H ₇ N ₃ O ₈	triclinic	P1	FALSE
qmof-54da05e	64	ZnC ₁₁ H ₁₂ N ₃ O ₅	triclinic	P1	FALSE
qmof-dd62091	28	ZnC ₈ F ₈ H ₂ N ₄ S ₄ Si	monoclinic	Cm	FALSE
qmof-d4f01fe	28	ZnC ₈ F ₁₀ N ₄ S ₄ Si	monoclinic	C2/m	FALSE

Supplementary Table 4: 70 diverse structures sampled from the QMOF database for benchmarking forces, stress and energies

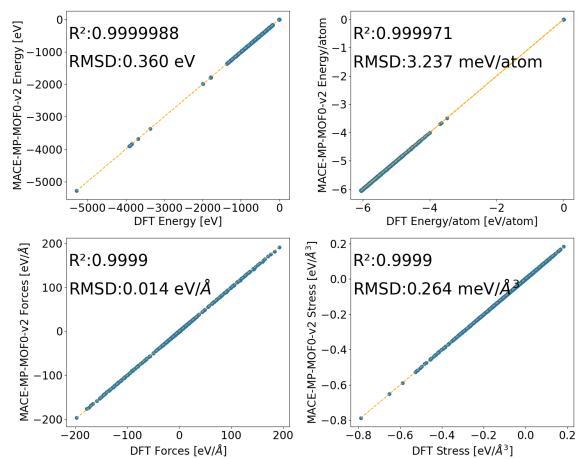
Supplementary Note 4.3: Benchmarking of vibrational properties

The Density of states in Supplementary Figure 4 are computed with the linear tetrahedron method in Phonopy [13] with a frequency pitch of 0.01 THz for each of the methods.

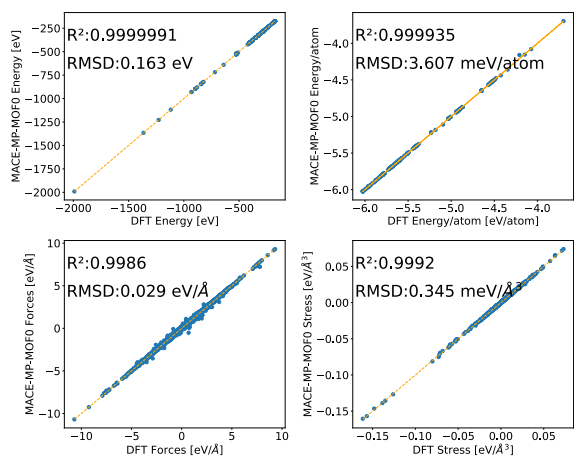
The Normalized Partial Density of States was computed for MOF-74 with a frequency pitch of 0.05 on a $11 \times 11 \times 11$ grid to study the contribution of each species of the MOF to the entire phonon spectrum. The normalization is done such that the summing the area under the curve for each species gives the total area under



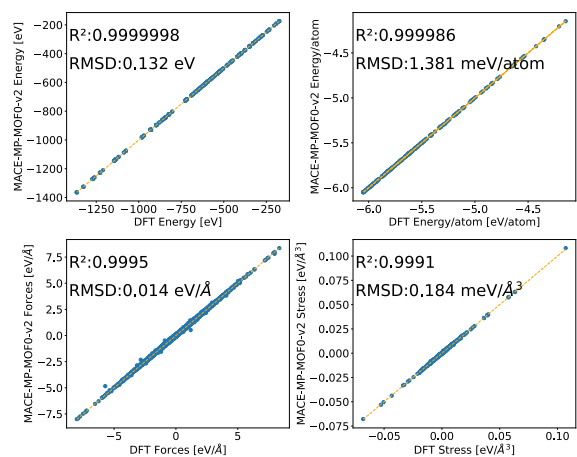
(a)



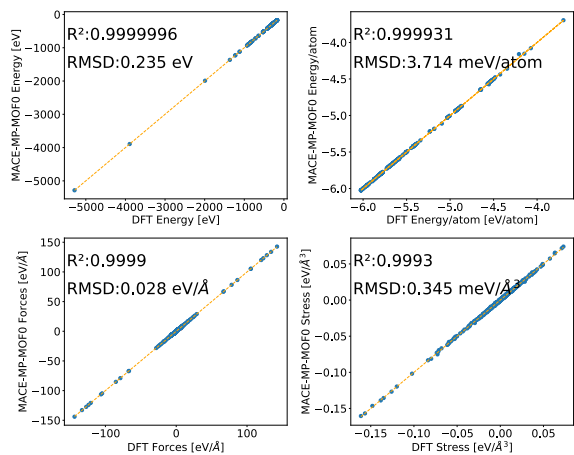
(b)



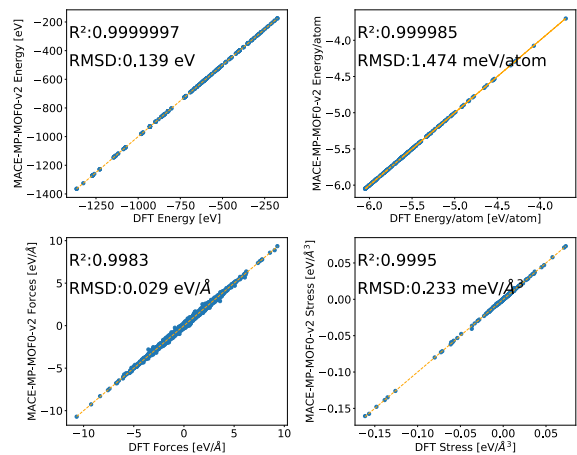
(c)



(d)

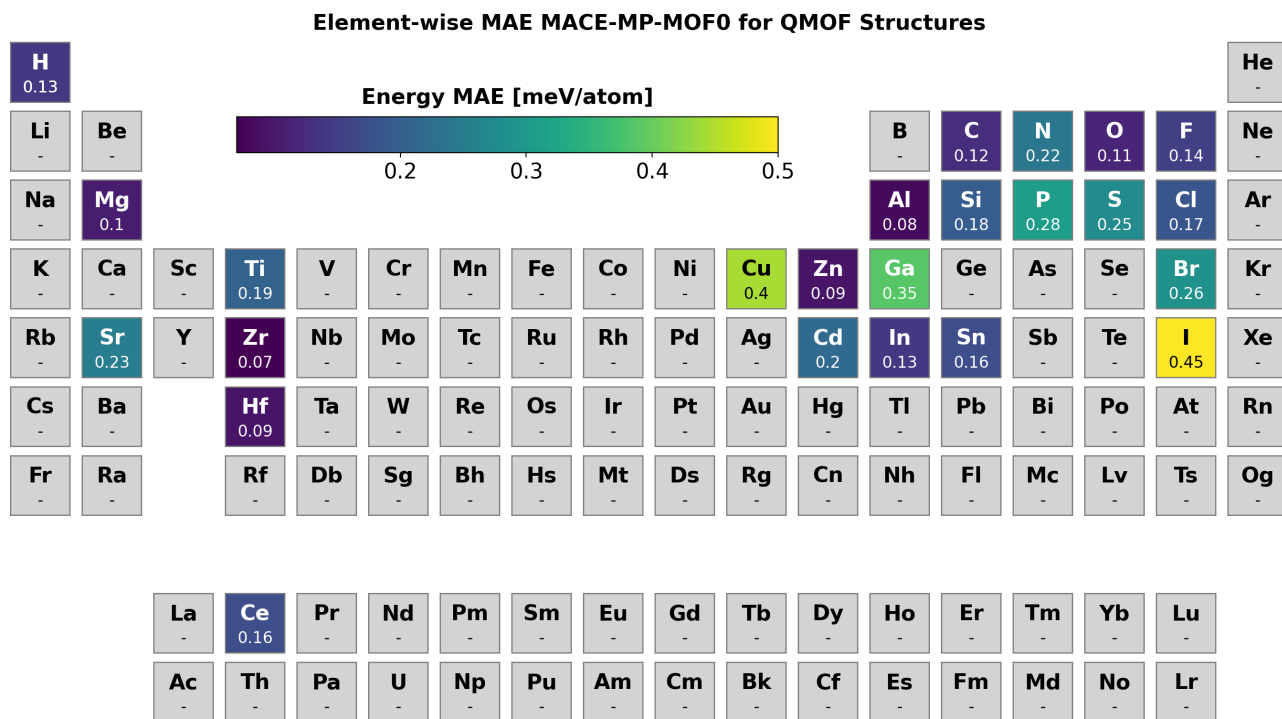


(e)

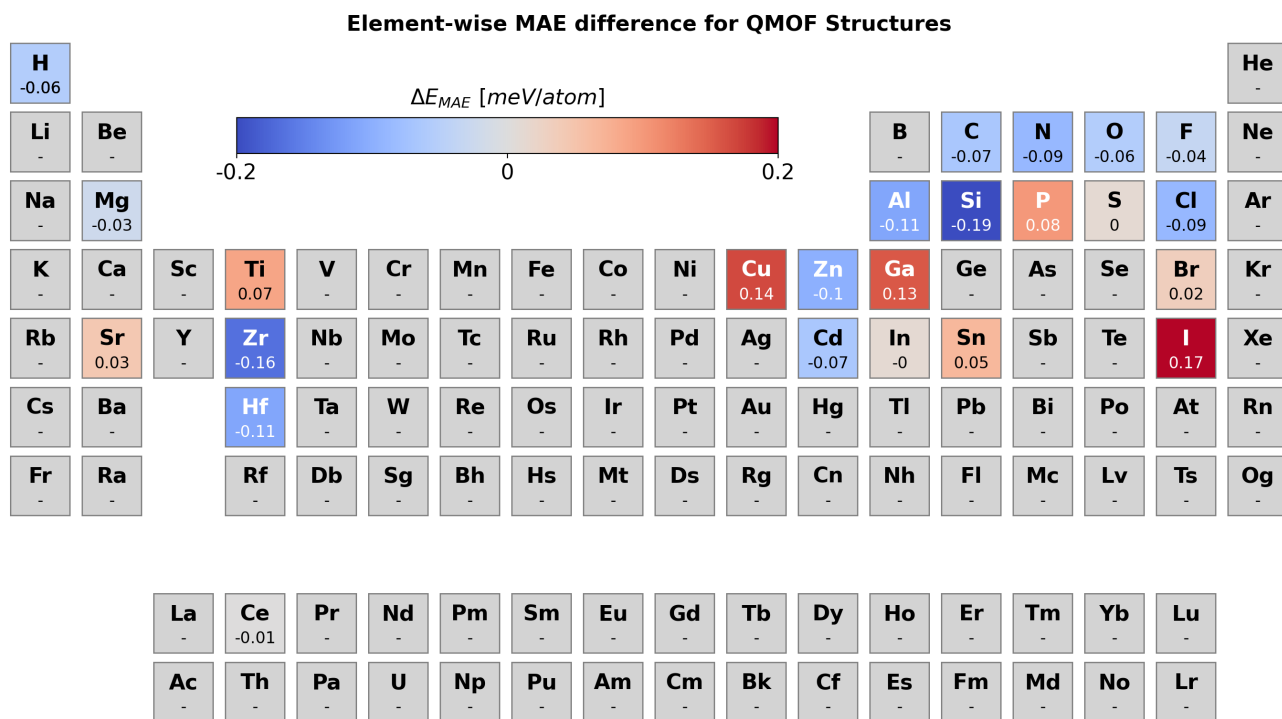


(f)

Supplementary Figure 2: Prediction of energies, forces and stress obtained with MACE-MP-MOF0 (a,c,e) and MACE-MP-MOF0-v2 (b,d,f) against DFT with a linear fit ≥ 0.99 R2 score

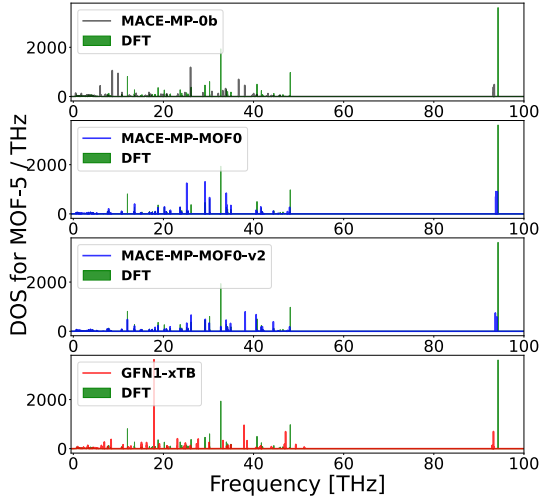


(a)

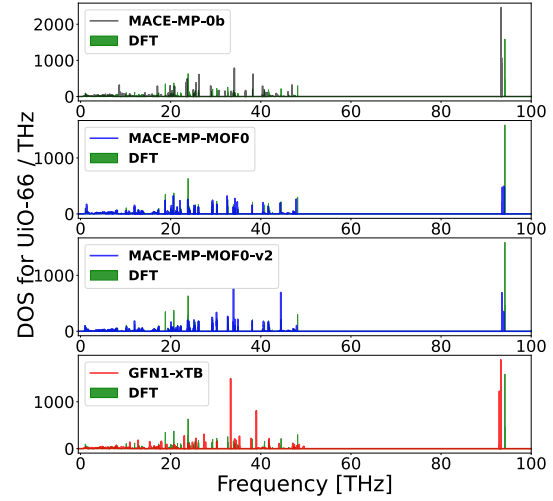


(b)

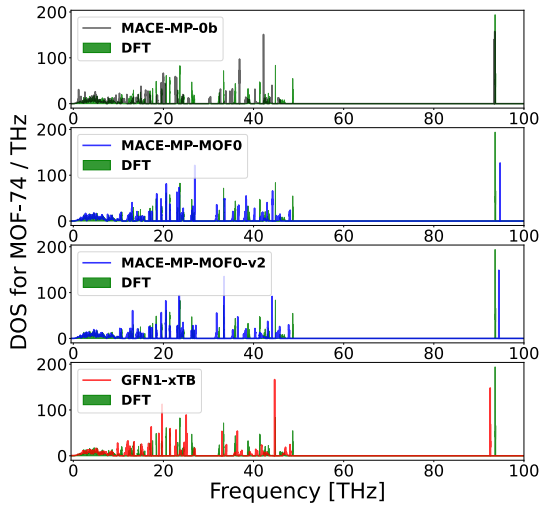
Supplementary Figure 3: The heatmap of element-wise MAEs in energies relative to DFT obtained with (a) the MACE-MP-MOF0-v2 model and (b) their difference between the the MACE-MP-MOF0-v2 and the MACE-MP-0b model predictions



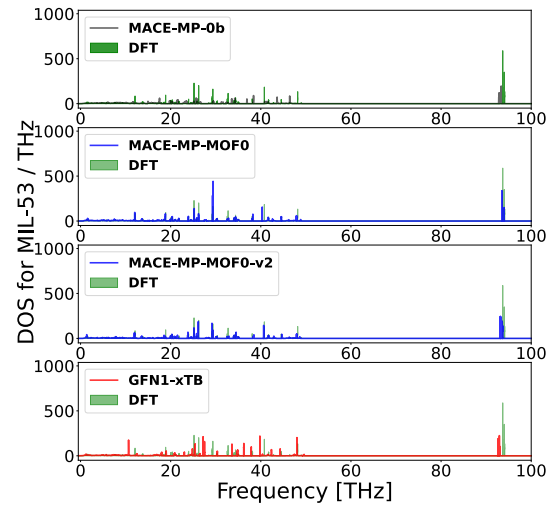
(a)



(b)



(b)



(b)

Supplementary Figure 4: Density of states (DOS) in the full frequency range for the investigated MOFs predicted by the respective methods overlaid on DFT data. DFT DOS were calculated from force constants available in [8, 12] on a $11 \times 11 \times 11$ mesh

the curve for the total density of states.

$$\text{Normalized PDOS}_{\text{element}}(E) = \text{PDOS}_{\text{element}}(E) \times \frac{\text{Total DOS}}{\text{Total PDOS}_{\text{element}}}$$

Supplementary Note 5: Details of capturing equilibrium positions and partial density of states

The Supplementary Table 5 provides the comparison of different methods in accurately obtaining the fractional atomic coordinates of the optimized geometry for MOF-5, UiO-66 and MIL-53. While the MACE-MP-MOF0 consistently achieves significant improvement in capturing the interactions of all the elements, the improvement

MOF	Method	$RMSD_{Node} \times 10^3$	$RMSD_H \times 10^3$	$RMSD_C \times 10^3$	$RMSD_O \times 10^3$
MOF-5 (Zn)	GFN1-xTB	3.7	0.9	1.3	2.1
	MACE-MP-0b	0.6	0.3	9.1×10^{-2}	0.2
	MACE-MP-MOF0	3.9×10^{-3}	8.8×10^{-2}	3.8×10^{-2}	2.5×10^{-2}
	MACE-MP-MOF0-v2	2.9×10^{-3}	5.6×10^{-3}	2.5×10^{-2}	1.40×10^{-2}
UiO-66 (Zr)	GFN1-xTB	5.2	5.1	4.7	3.1
	MACE-MP-0b	1.1	17.4	15.7	2.6
	MACE-MP-MOF0	1.2×10^{-3}	0.5	0.5	5.2×10^{-2}
	MACE-MP-MOF0-v2	4.0×10^{-2}	0.3	0.3	6.6×10^{-2}
MIL-53 (Al)	GFN1-xTB	1.3	7.4	2.4	5.3
	MACE-MP-0b	1.1	159.4	30.6	52.1
	MACE-MP-MOF0	0.3	5.1	2.9	4.3
	MACE-MP-MOF0-v2	0.3	6.2	3.3	4.6

Supplementary Table 5: Root Mean Square Deviations (RMSD)s of optimized fractional atomic positions with the respective methods relative to DFT

in performance with MACE-MP-MOF0 becomes even more prominent for covalent interactions between C, H and O in the linkers.

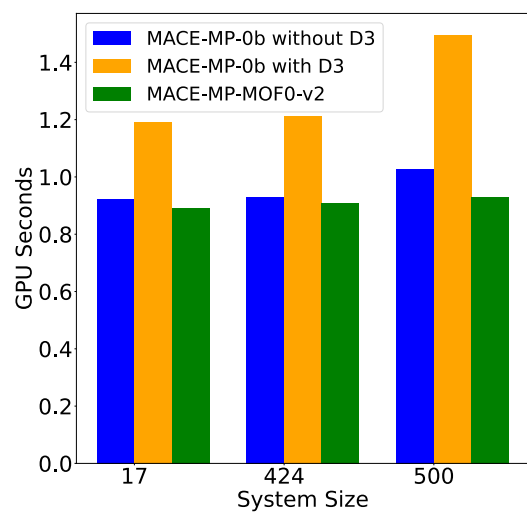
Supplementary Note 6: Computational efficiency of MACE-MP-MOF0

The MACE-MP-MOF0 and MACE-MP-MOF0-v2 models were trained on DFT data with vdW corrections instead of explicitly adding dispersion corrections on the foundation model (MACE-MP-0b without D3) to increase the speed of the model and eliminate some spurious phonon instabilities. Hence, the computational efficiency in performing a static calculation of the MACE-MP-MOF0-v2 model was tested against the GFN1-xTB, MACE-MP-0b (with D3 correction) and MACE-MP-0b(without D3) on three MOFs of varying size, the smallest being 17 atoms and the largest being 500 atoms in the unit cell. As shown in Supplementary Figure 5, the MACE-MP-MOF0-v2 model is 50% faster than the MACE-MP-0b model with D3 correction. All the results reported in the main text for the MACE-MP-0b model are with D3 correction, demonstrating the the MACE-MP-MOF0-v2 model not only shows an improvement in accuracy but also in speed. As compared to the Semi-Empirical Quantum Mechanical method GFN1-xTB, the MACE-MP-MOF0-v2 model is significantly cheaper by 90%. It must be noted that these times are for single static calculation, and for phonons, when hundreds of such static calculations need to be performed on supercells containing large number of atoms, the better computational efficiency of MACE-MP-MOF0-v2 becomes more evident.

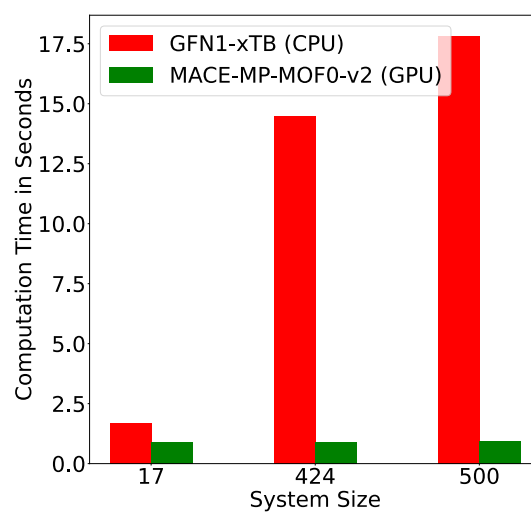
References

- [1] Rosen, A. S. *et al.* High-throughput predictions of metal–organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. *npj Comput Mater* 8, 112 (2022).
- [2] Rosen, A. S. *et al.* Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* 4, 1578–1597 (2021).
- [3] Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* 68, 314–319 (2013). URL <https://www.sciencedirect.com/science/article/pii/S0927025612006295>.
- [4] Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* 54, 11169–11186 (1996). URL <https://link.aps.org/doi/10.1103/PhysRevB.54.11169>.
- [5] Ganose, A. *et al.* atomate2. URL <https://github.com/materialsproject/atomate2>.

- [6] Deng, B. *et al.* Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **5**, 1031–1041 (2023). URL <https://doi.org/10.1038/s42256-023-00716-3>.
- [7] Wieser, S. & Zojer, E. Machine learned force-fields for an ab-initio quality description of metal-organic frameworks (2023). [2308.01278](https://doi.org/10.26434/chemrxiv-2023-2308).
- [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. *Advanced Theory and Simulations* **5**, 2200031 (2022). URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/adts.202200031>.
- [9] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996). URL <https://link.aps.org/doi/10.1103/PhysRevLett.77.3865>.
- [10] 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. *The Journal of Chemical Physics* **132**, 154104 (2010). URL <https://doi.org/10.1063/1.3382344>.
- [11] Hourahine, B. *et al.* Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics* **152**, 124101 (2020). URL <https://doi.org/10.1063/1.5143190>.
- [12] Wieser, S. & Zojer, E. Machine learned force-fields for an ab-initio quality description of metal-organic frameworks. *npj Computational Materials* **10**, 18 (2024). URL <https://doi.org/10.1038/s41524-024-01205-w>.
- [13] Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scripta Materialia* **108**, 1–5 (2015). URL <https://www.sciencedirect.com/science/article/pii/S1359646215003127>.



(a)



(b)

Supplementary Figure 5: Supplementary Table 5: The computational time for performing static calculations on MOFs with different number of atoms in the unit cell