Supportive Information – Phonon-accurate machine-learning potentials from automated workflows

Christina Ertural[1], Aakash A. Naik[1,3], Yuanbin Liu[2], Jonas Grandel[1], Natascia Fragapane[2], Joe D. Morrow[2], Daniel F. Thomas du Toit[2], Yuxing Zhou[2], Philipp Benner[1], Volker L. Deringer[2,*], and Janine George[1,3,*]

[1]Federal Institute for Materials Research and Testing (BAM), Unter den Eichen 87, 12205 Berlin, Germany.
[2]Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, UK
[3]Friedrich Schiller University Jena, Institute of Condensed Matter Theory and Solid-State Optics, Max-Wien-Platz 1, 07743 Jena, Germany.
[*]Corresponding authors: Volker L. Deringer (volker.deringer@chem.ox.ac.uk), Janine George (janine.george@bam.de)

ABSTRACT

The supportive information for the main article.

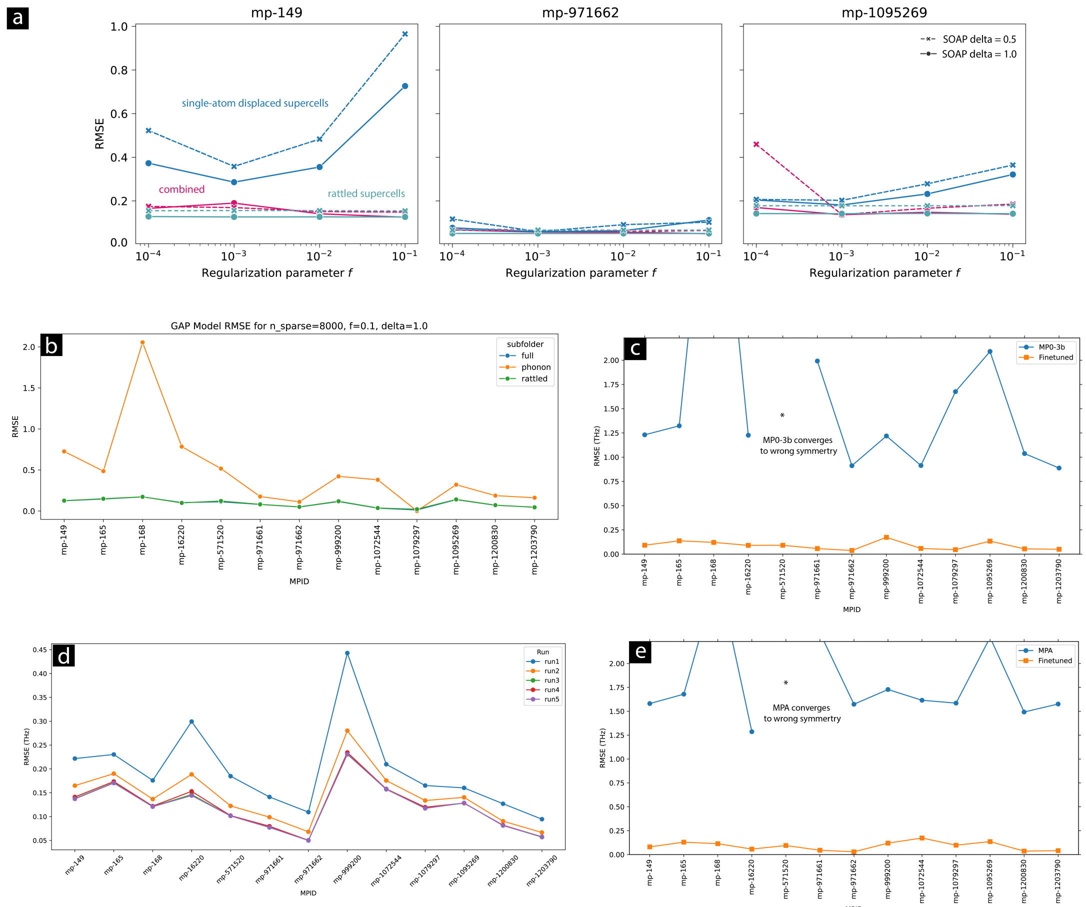


Figure S1. Si convergence tests. a) Regularization parameter f vs. RMSE for mp-149, mp-971662, mp-1095269. b) Si-MPIDs vs. RMSE for GAP with SOAP-delta = 1.0, n_sparse = 8000, $f = 0.1$. c) Si-MPIDs vs. MP0-3b and fine-tuned MP0-3b. d) Si-MPIDs vs. MACE from scratch for increasing run steps (run1 to run5) with different epoch times and training settings (see method section of main text). e)Si-MPIDs vs. MPA and fine-tuned MPA.

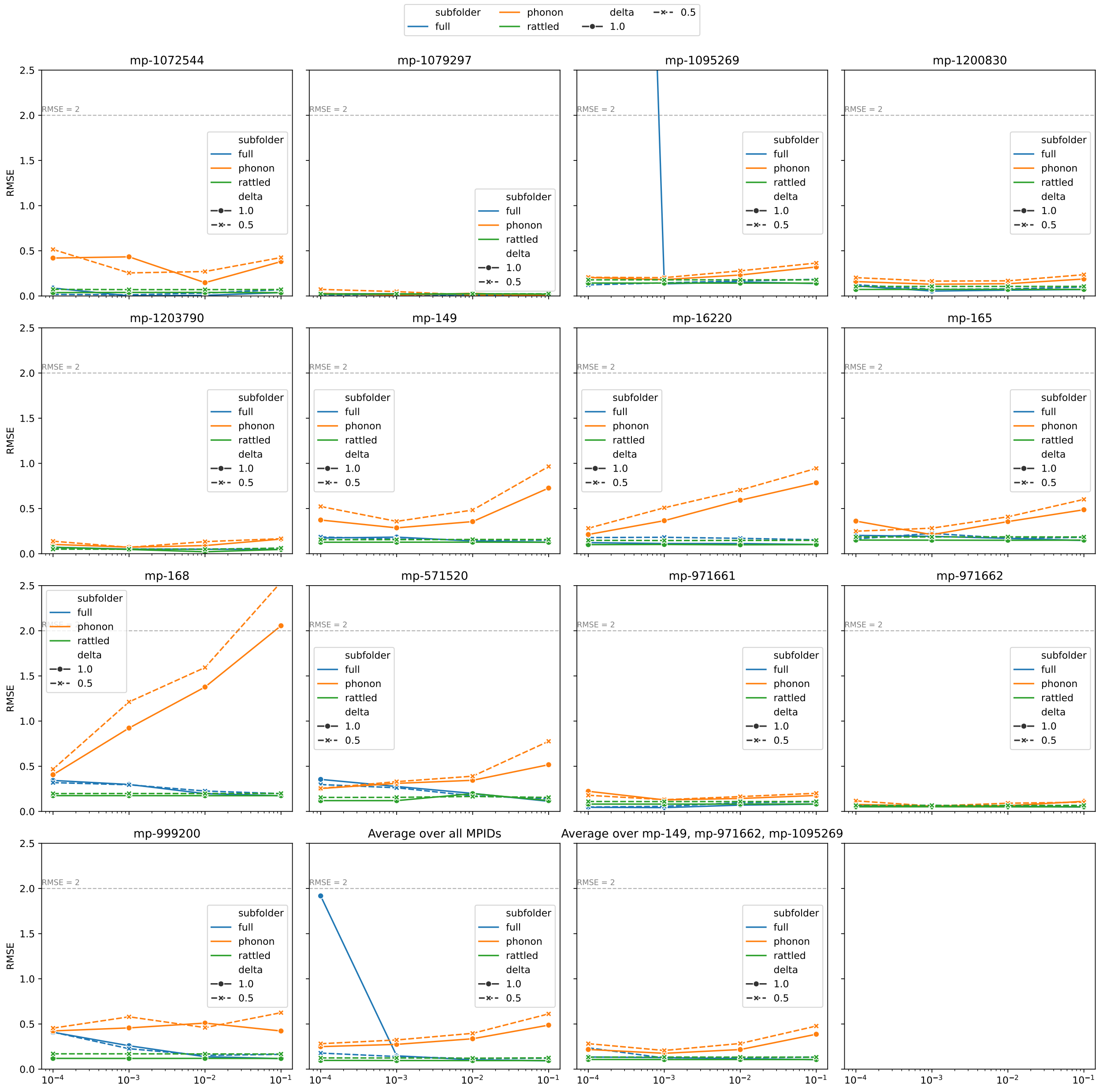


Figure S2. Si convergence tests. a) Regularization parameter (reg) vs. RMSE for n_sparse = 7000.

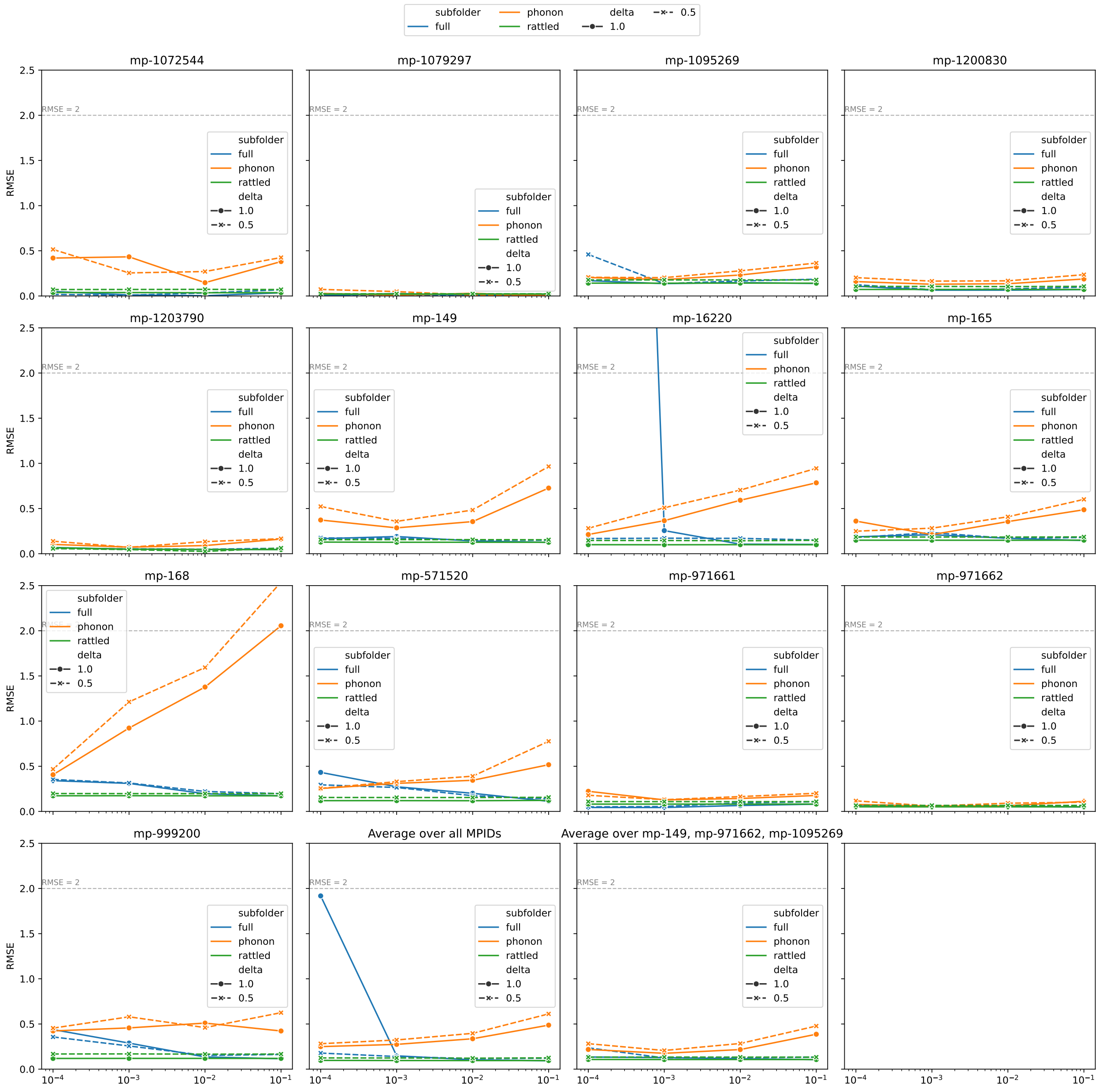


Figure S3. Si convergence tests. a) Regularization parameter (reg) vs. RMSE for n_sparse = 8000.

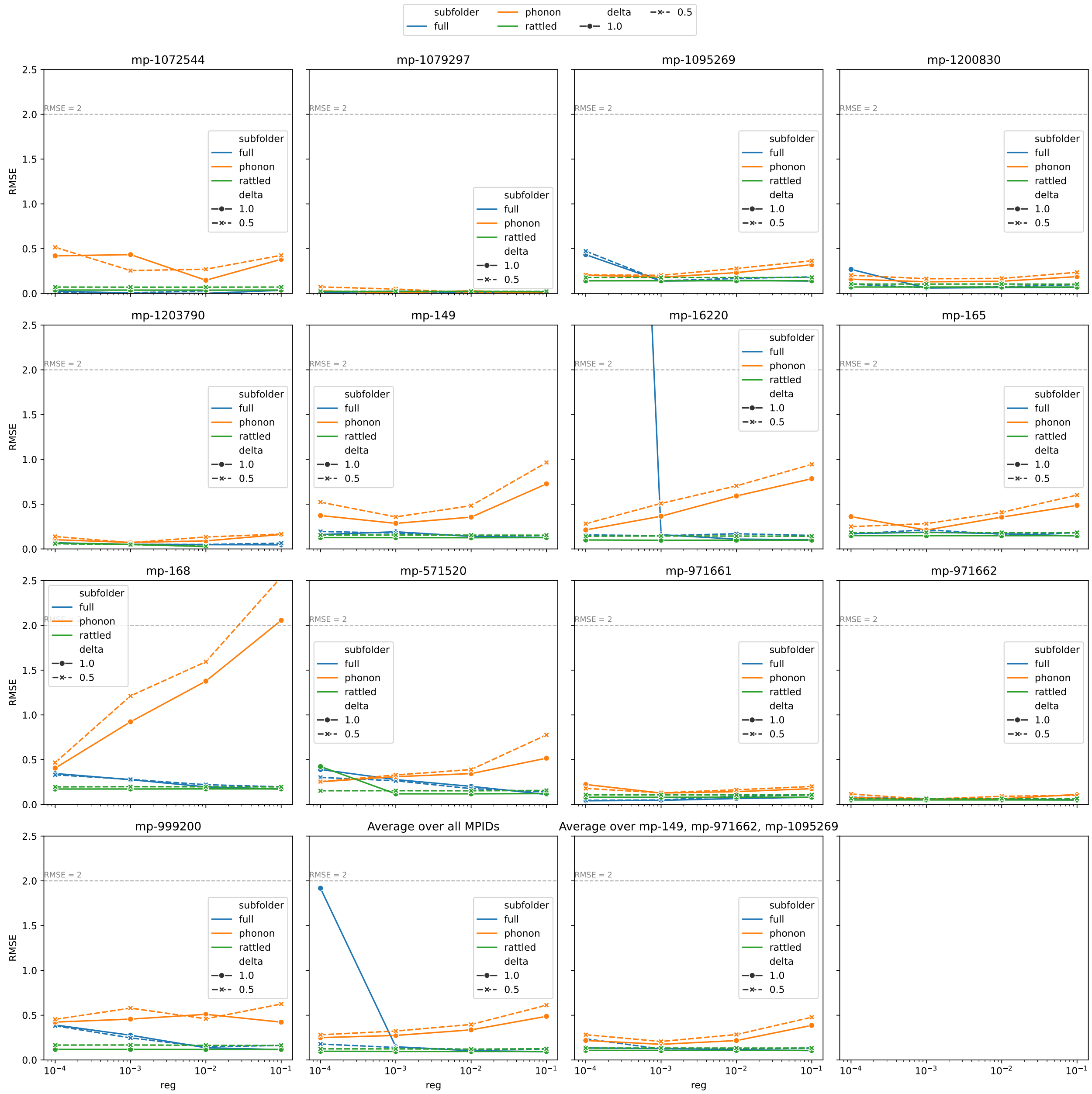


Figure S4. Si convergence tests. a) Regularization parameter (reg) vs. RMSE for n_sparse = 9000.

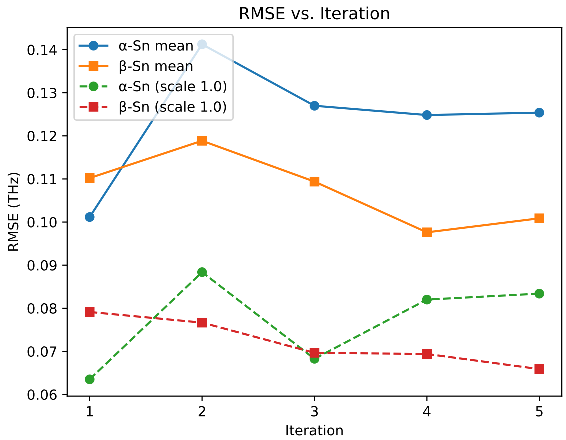
Advanced benchmark

Next, we test the robustness of the Si potentials that we acquired from `autoplex`. Si MPIDs have been filtered for suitable candidates (regarding supercell size) with a function? in `autoplex` from which five MPIDs have been tested, and three of them produced sensible results. Our test structures are mp-34, mp-644693, and mp-988210. We test the accuracy of the different potentials as a kind of "stress test".

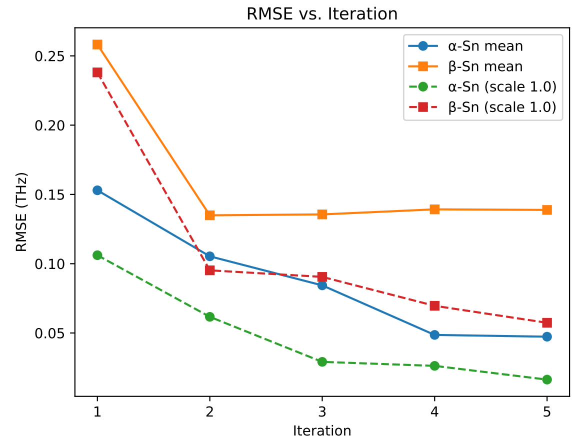
Potential	**MPID**	**RMSE (THz)**
GAP	mp-34	8.80
GAP	mp-644693	1.18
GAP	mp-988210	0.61
MACE finetuned	mp-34	1.98
MACE finetuned	mp-644693	1.10
MACE finetuned	mp-988210	—
MP0-3b	mp-34	2.81
MP0-3b	mp-644693	—
MP0-3b	mp-988210	2.76
MACE from scratch	mp-34	10.90
MACE from scratch	mp-644693	1.39
MACE from scratch	mp-644693	—

Table 1. RMSE values per MPID and potential type. There is no entry when the structure did not converge to the correct symmetry with the respective potential.

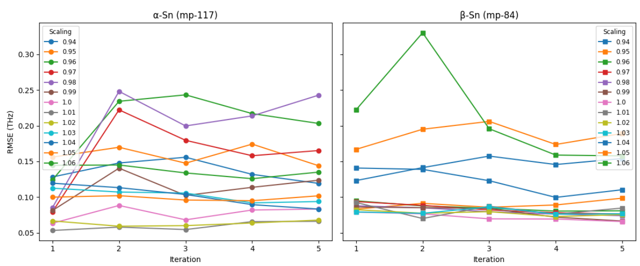
The overall outcome of this further benchmark is rather negative, as the (average) RMSE is relatively high, and the structures do not converge to the correct symmetry. Only mp-988210 yields acceptable results with the GAP potential. The reason for this outcome could be that the suitable MPIDs already went into the data generation, and mp-644693 and mp-988210 are polycrystalline/amorphous structures, which are not represented in the training data at all. Still, in the case of mp-34, it is evident that the foundation model MP0-3b performs better than the potentials trained from scratch and that the RMSE can be further reduced by fine-tuning.



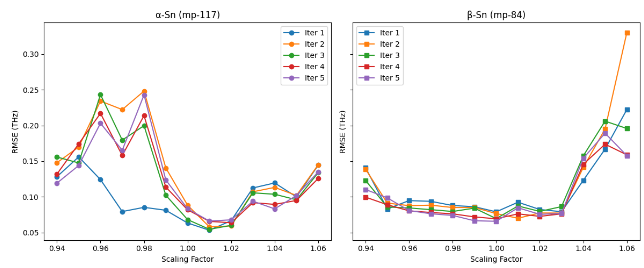
(a) Average RMSE vs. iteration GAP.



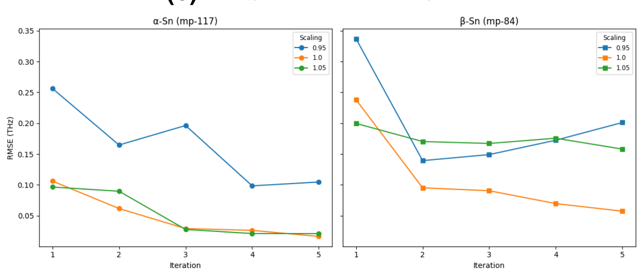
(b) Average RMSE vs. iteration MACE.



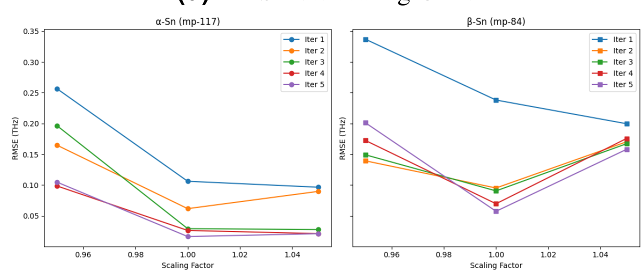
(c) RMSE vs. iteration GAP.



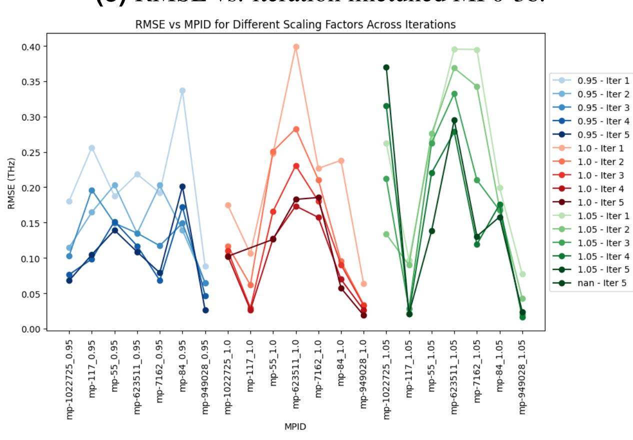
(d) RMSE vs. scaling GAP.



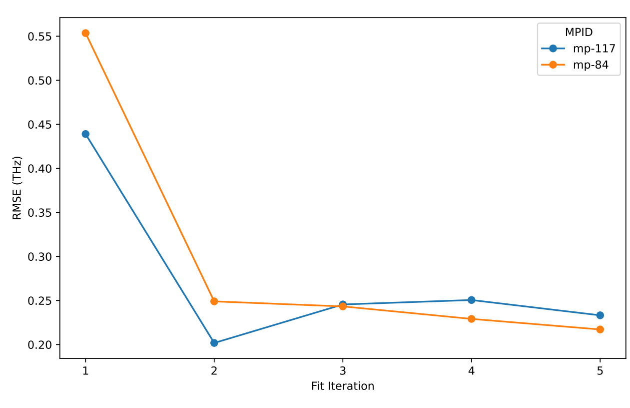
(e) RMSE vs. iteration finetuned MP0-3b.



(f) RMSE vs. scaling finetuned MP0-3b.

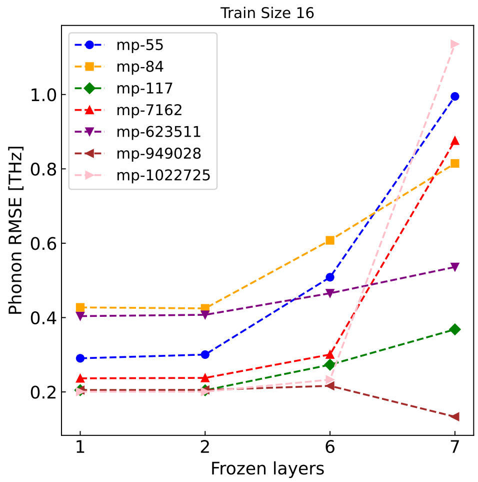


(g) RMSE vs. MPIDs finetuned MP0-3b.

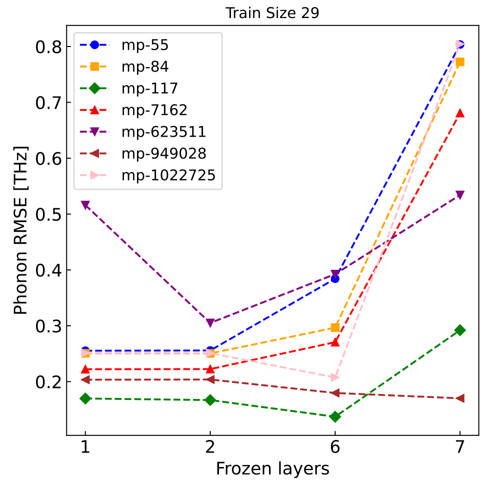


(h) RMSE vs. fit iteration for finetuned MPA.

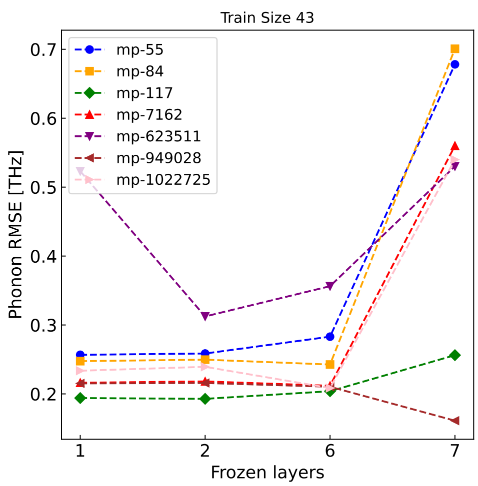
Figure S5. Sn plots.



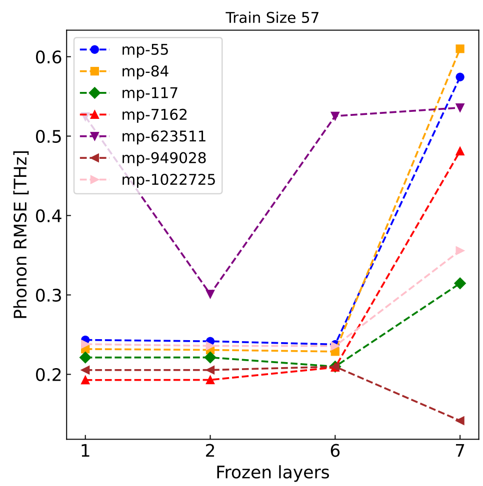
(a) RMSE vs. frozen layers for training size 16.



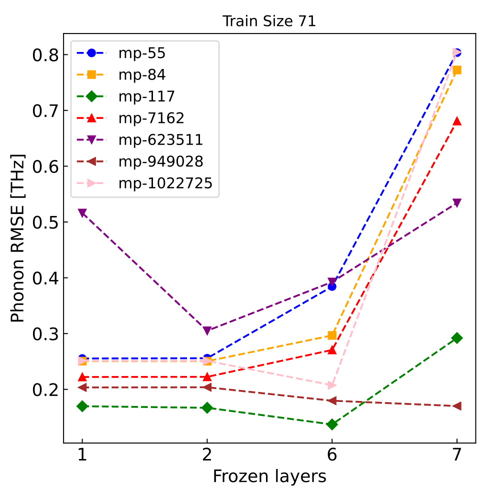
(b) RMSE vs. frozen layers for training size 29.



(c) RMSE vs. frozen layers for training size 43.

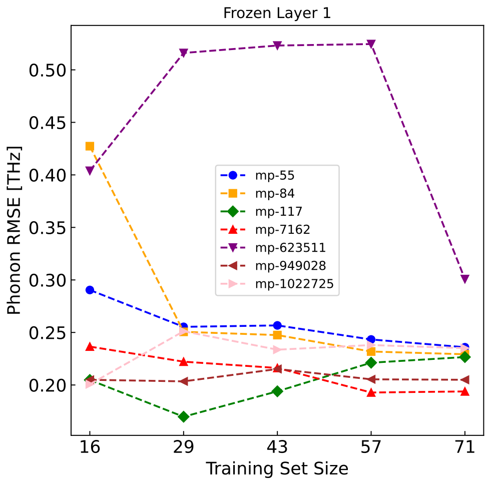


(d) RMSE vs. frozen layers for training size 57.

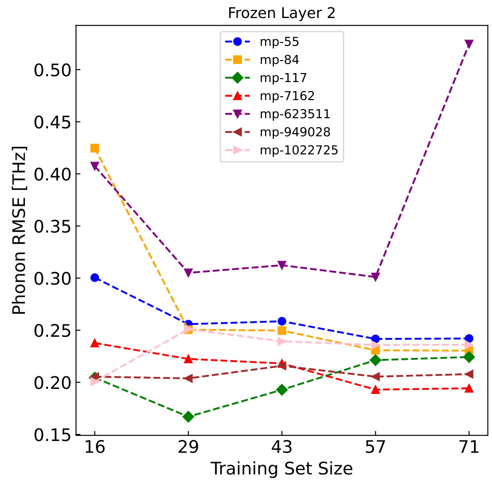


(e) RMSE vs. frozen layers for training size 71.

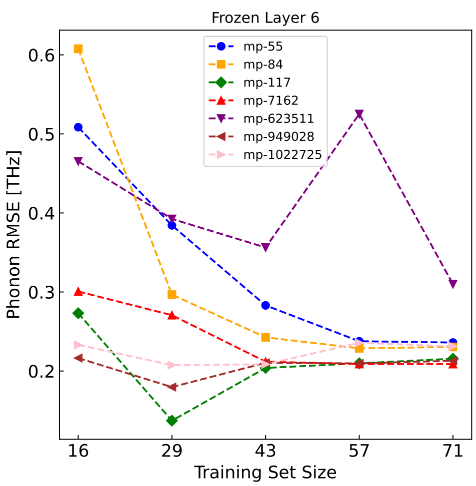
Figure S6. Sn plots MACE frozen layer parameter.



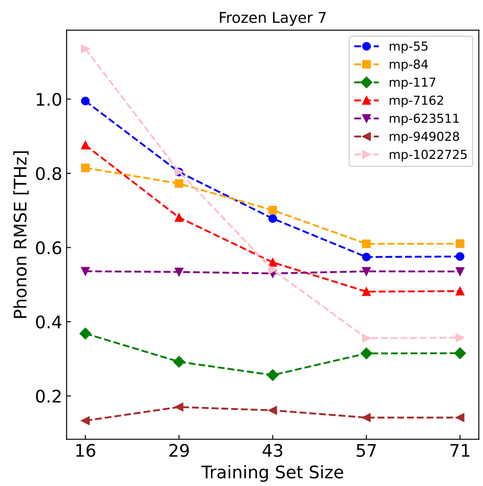
(a) RMSE vs. frozen layers for frozen layer 1.



(b) RMSE vs. frozen layers for frozen layer 2.



(c) RMSE vs. frozen layers for frozen layer 6.



(d) RMSE vs. frozen layers for frozen layer 7.

Figure S7. Sn plots MACE frozen layer parameter.

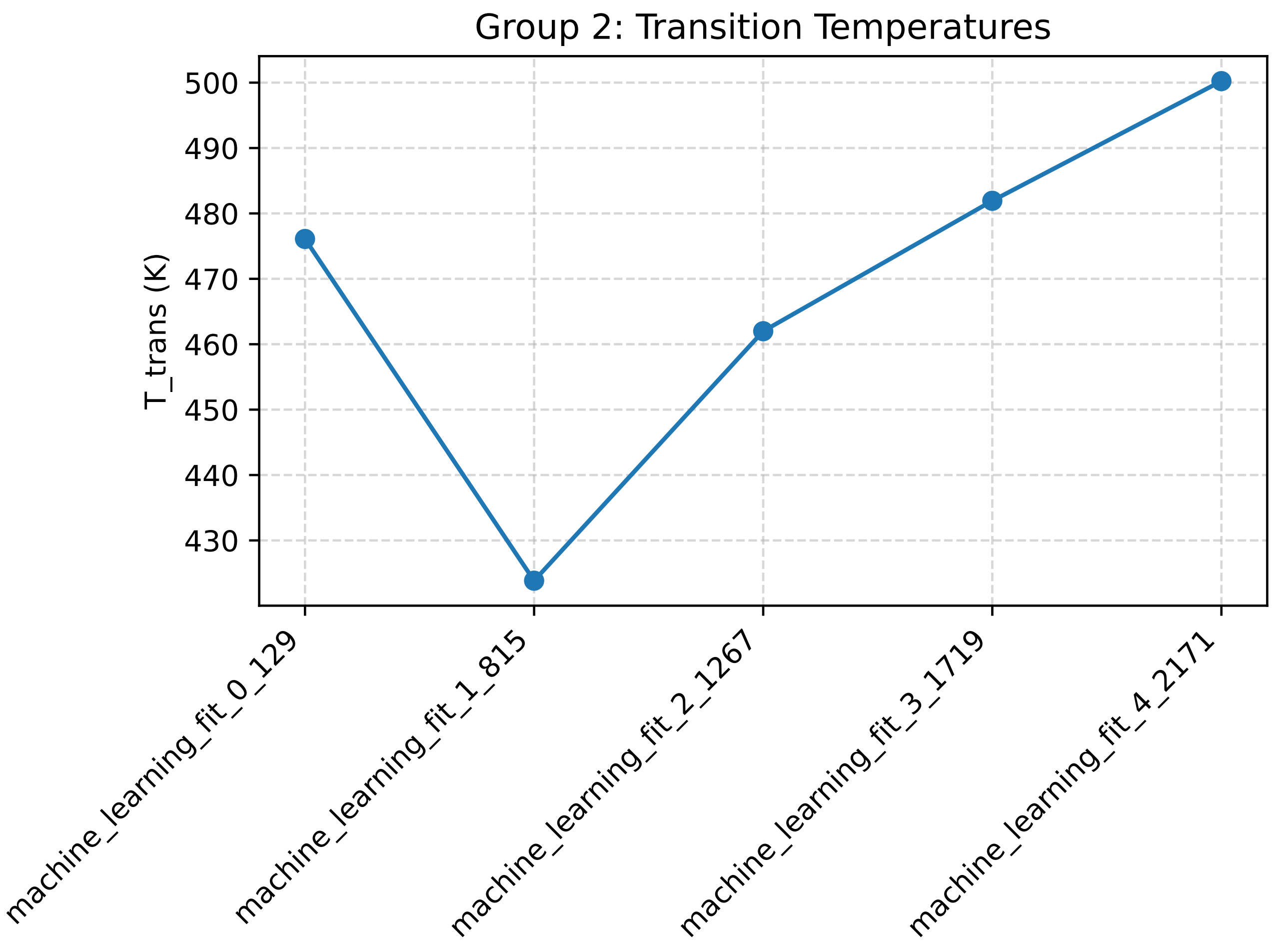
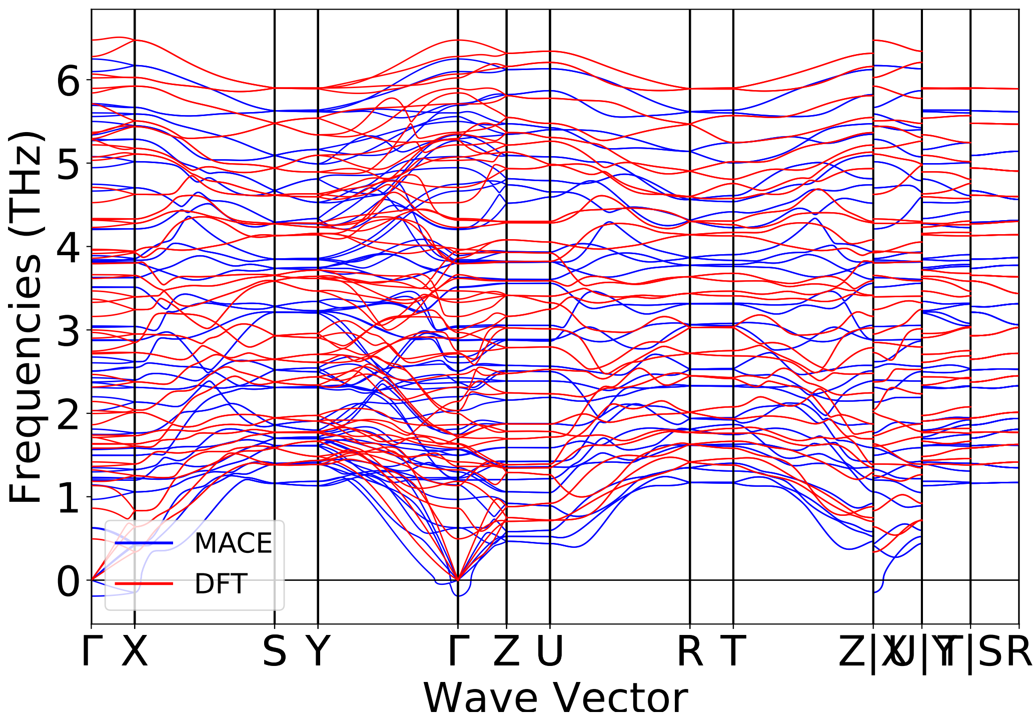
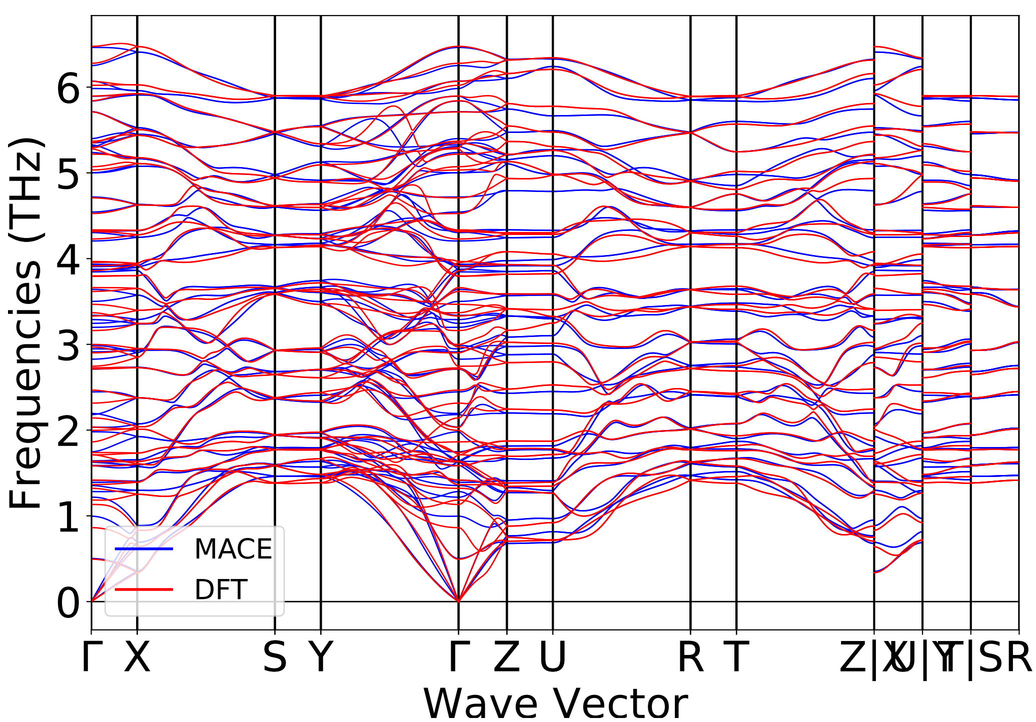


Figure S8. Sn transition temperature in dependence on iterations.

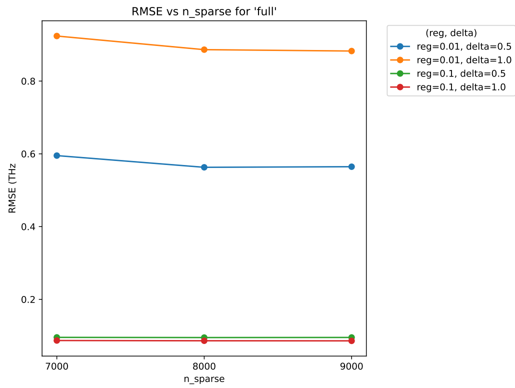


(a) MPA, RMSE = 0.35 THz.

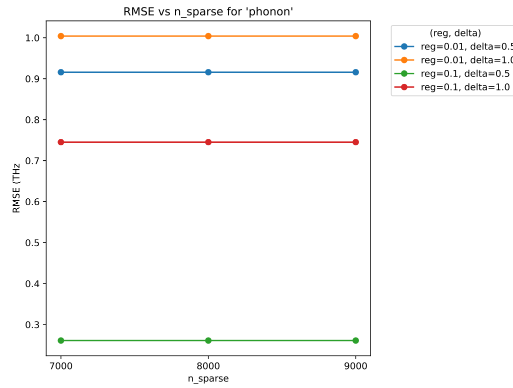


(b) MPA fine-tuned, RMSE = 0.04 THz.

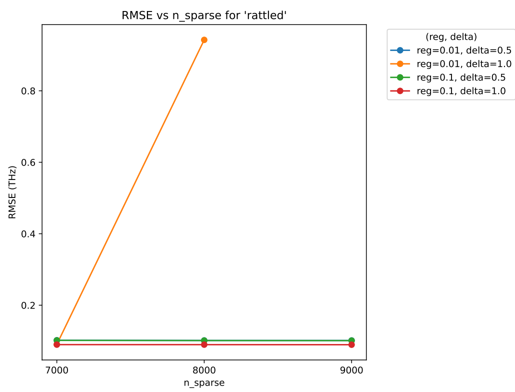
Figure S9. Sb_2Se_3 MPA and MPA fine-tuned based phonon structures. Note that the MPA is at DFT PBE level, while the DFT and MPA fine-tuned are at DFT PBE-sol level.



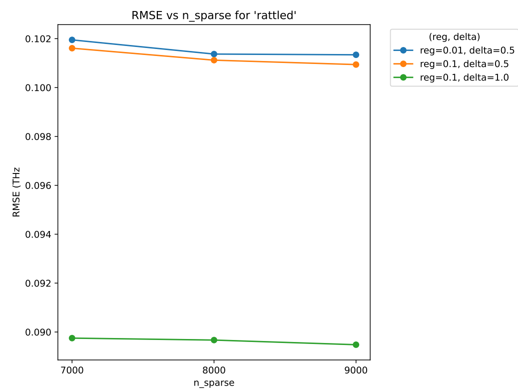
(a) RMSE vs. n_sparse for the full dataset.



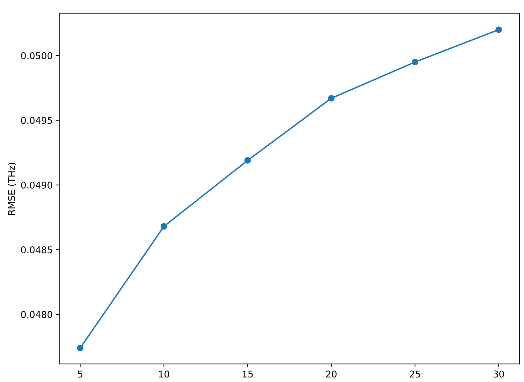
(b) RMSE vs. n_sparse for the single-atom displaced sub-dataset.



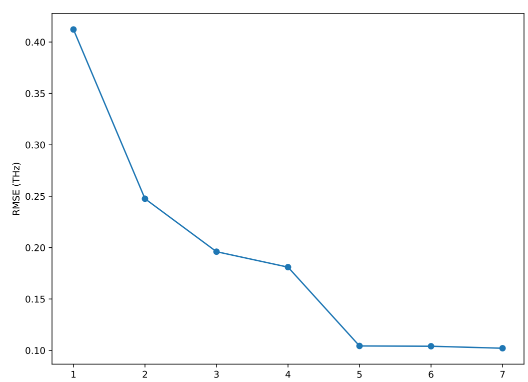
(c) RMSE vs. n_sparse for the rattled sub-dataset.



(d) RMSE vs. n_sparse for the rattled sub-dataset excluding extreme outliers.

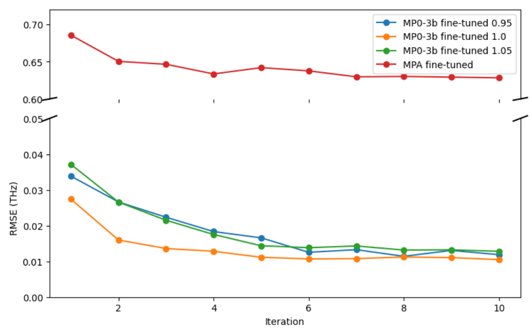


(e) RMSE vs. batch size of the MACE fine-tuning.

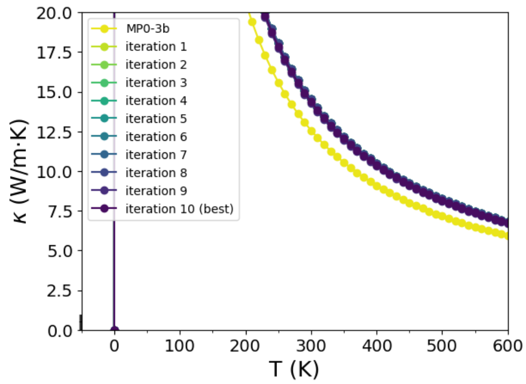


(f) RMSE vs. run iteration of the MACE from scratch generation.

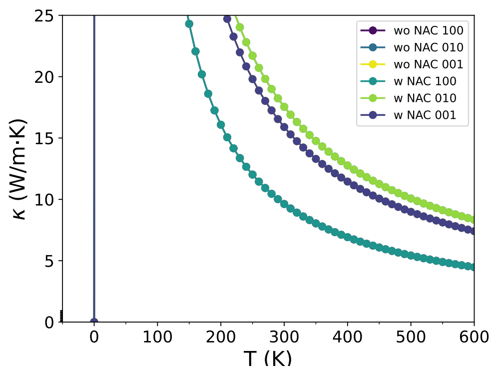
Figure S10. Sb_2Se_3 convergence tests.



(a) RMSE vs. fine-tune iteration of MP0-3b and MPA



(b) Total κ (RTA) in dependence of fit iteration.



(c) Anisotropic κ (RTA, with NAC = w NAC, without NAC = wo NAC).

Figure S11. Ga_2O_3 convergence tests.

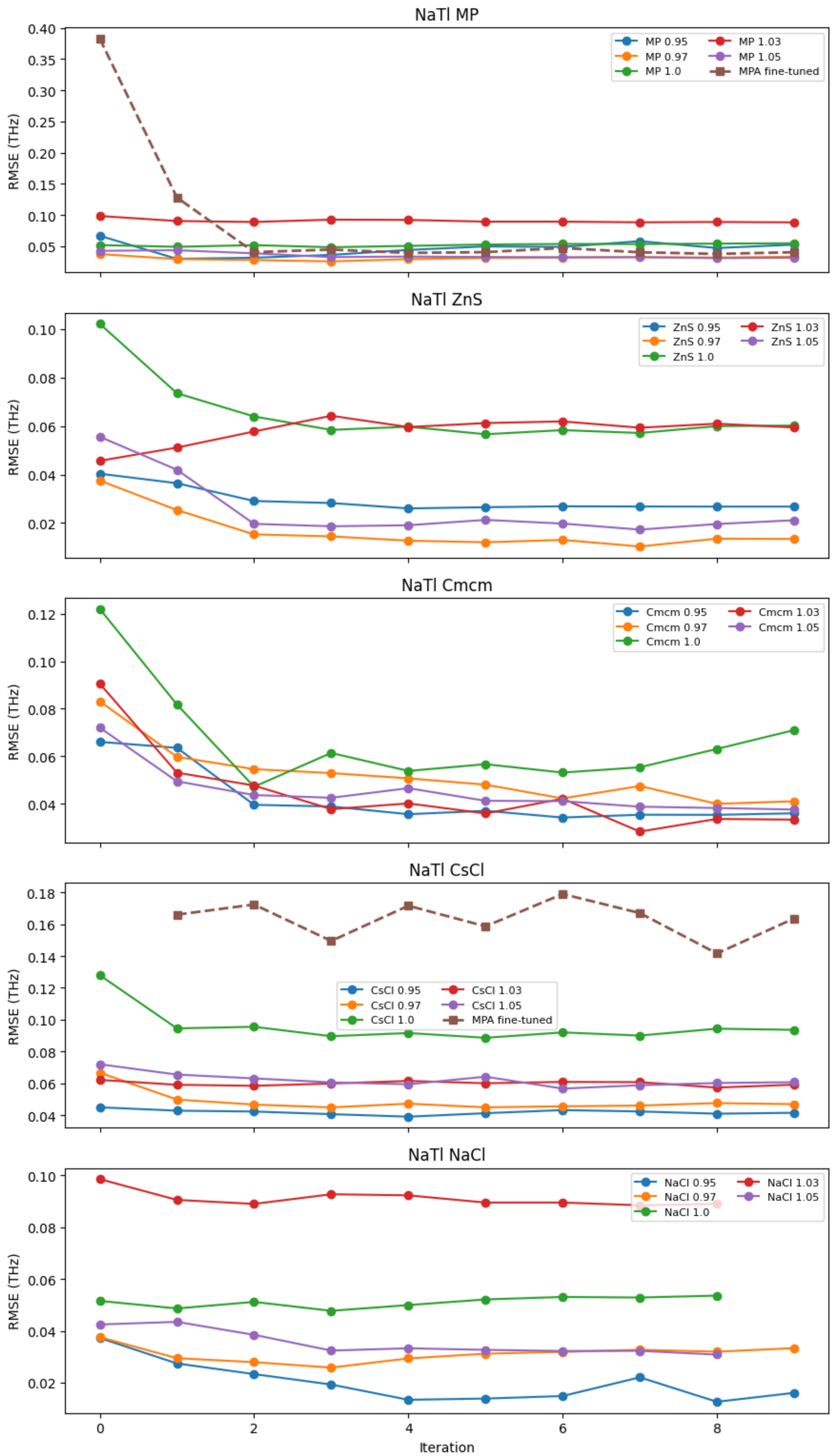


Figure S12. NaTl convergence tests.

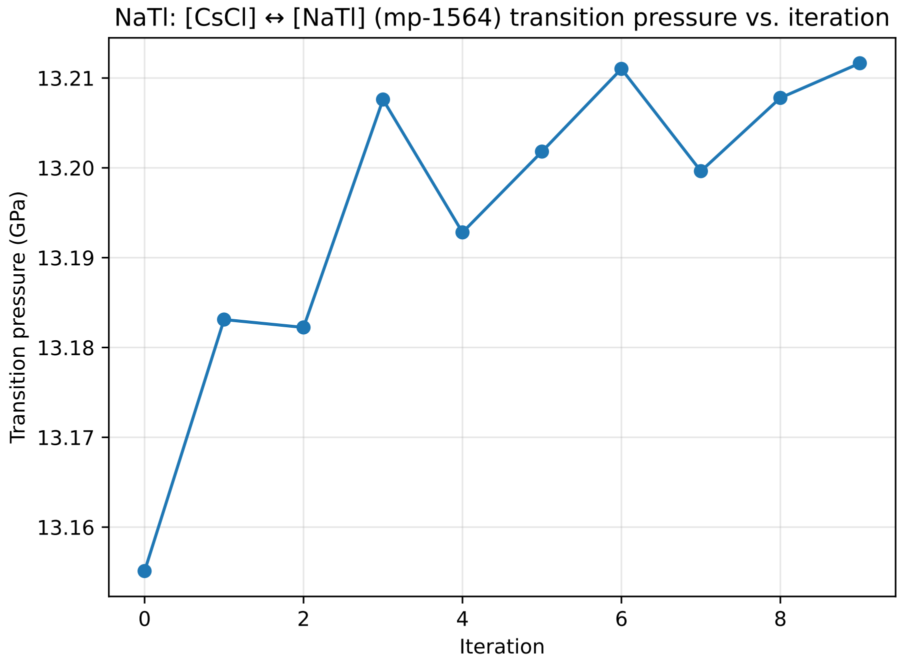


Figure S13. NaTl convergence test for MP0-3b fie-tuning all iterations.

The MACE hyperparameters varied during the MACE from scratch potential generation for the convergence tests for Si and Sb_2Se_3 are listed below in Tab.3.

Run	Max Epochs	Energy Weight	Forces Weight	Stress Weight
run1	5000	1000.0	1000.0	1.0
run2	1000	1000.0	1200.0	1.0
run3	1000	1000.0	1500.0	1.0
run4	1000	1500.0	1000.0	1.0
run5	1000	1000.0	2000.0	1.0

Table 2. Overview of MACE from scratch run parameters for Si.

Run	Max Epochs	Energy Weight	Forces Weight	Stress Weight
run1	5000	1000.0	1500.0	1.0
run2	5000	1000.0	1200.0	1.0
run3	5000	1000.0	1300.0	1.0
run4	2000	1500.0	1000.0	1.0
run5	2000	1000.0	2000.0	1.0
run6	1000	1000.0	1500.0	1.0
run7	2000	1000.0	1200.0	1.0

Table 3. Overview of MACE from scratch run parameters for Sb_2Se_3.

Fit	Train	Test	% Train Data
Sn (GAP)			
machine_learning_fit_0_81	44	22	38%
machine_learning_fit_1_869	61	31	54%
machine_learning_fit_2_1410	79	39	69%
machine_learning_fit_3_1951	96	48	84%
machine_learning_fit_4_2492	114	56	100%
Sn (fine-tuned MACE, α/β-Sn)			
machine_learning_fit_0_15	5	2	38%
machine_learning_fit_1_209	9	4	69%
machine_learning_fit_2_340	13	6	100%
Sn (fine-tuned MACE, all Sn)			
machine_learning_fit_0_129	15	7	21%
machine_learning_fit_1_815	29	14	41%
machine_learning_fit_2_1267	43	21	61%
machine_learning_fit_3_1719	57	28	80%
machine_learning_fit_4_2171	71	35	100%
NaTl (fine-tuned MACE)			
machine_learning_fit_0_153	18	9	8%
machine_learning_fit_1_1034	35	17	16%
machine_learning_fit_2_1604	99	49	46%
machine_learning_fit_3_2174	116	57	54%
machine_learning_fit_4_2744	133	65	61%
machine_learning_fit_5_3314	150	73	69%
machine_learning_fit_6_3884	166	82	77%
machine_learning_fit_7_4454	183	90	84%
machine_learning_fit_8_5024	200	98	92%
machine_learning_fit_9_5594	217	106	100%
Ga_2O_3 (fine-tuned MACE)			
machine_learning_fit_0_21	21	10	5%
machine_learning_fit_1_261	41	20	10%
machine_learning_fit_2_414	61	30	15%
machine_learning_fit_3_567	81	40	20%
machine_learning_fit_4_720	101	49	25%
machine_learning_fit_5_873	121	59	30%
machine_learning_fit_6_1026	282	138	71%
machine_learning_fit_7_1179	320	158	80%
machine_learning_fit_8_1332	360	178	90%
machine_learning_fit_9_1485	400	198	100%

Table 4. Overview of data per iteration for Sn, NaTl, and Ga_2O_3, including relative train set sizes.