



Cite this: *Phys. Chem. Chem. Phys.*,  
2026, **28**, 4459

# MACE foundation models for lattice dynamics: a benchmark study on double halide perovskites

Jack Yang, \*<sup>a</sup> Ziqi Yin,<sup>a</sup> Lei Ao <sup>abc</sup> and Sean Li <sup>a</sup>

Recent developments in materials informatics and artificial intelligence have led to the emergence of foundational energy models for material chemistry, as represented by the suite of MACE-based foundation models, bringing a significant breakthrough in universal potentials for inorganic solids. As to all method developments in computational materials science, performance benchmarking against existing high-level data with a focus on specific applications, is critically needed to understand the limitations in the models, thus facilitating the ongoing improvements in the model development process, and occasionally, leading to significant conceptual leaps in materials theory. Here, using our own published DFT (density functional theory) database of room-temperature dynamic stability and vibrational anharmonicity for ~2000 cubic halide double perovskites, we benchmarked the performances of four different variants of the MACE foundation models for screening the dynamic stabilities of inorganic solids. Our analysis shows that, as anticipated, the model accuracy improves with more training data. The dynamic stabilities of weakly anharmonic materials (as predicted by DFT) are more accurately reproduced by the foundation model, than those of highly anharmonic and dynamically unstable ones. The predominant source of error in predicting dynamic stability arises from the amplification of errors in atomic forces when predicting the harmonic phonon properties through the computation of the Hessian matrix, less so is the contribution from possible differences in the range of the configurational spaces that are sampled by DFT and the foundation model in molecular dynamics. We hope that our present findings will stimulate future studies towards more physics-inspired approaches in assessing the accuracy of foundation models for atomistic modelling.

Received 2nd December 2025,  
Accepted 20th January 2026

DOI: 10.1039/d5cp04693a

[rsc.li/pccp](http://rsc.li/pccp)

## 1 Introduction

Atomistic modelling plays a pivotal role in modern materials physics and chemistry, and is complementary to the experimental endeavours in discovering new materials for structural, electronic, energy harvesting and many other applications. Primarily, the key information to be extracted from atomistic modelling, particularly DFT<sup>1</sup> (density functional theory), which is the workhorse for modern computational materials science, is the total energy of a material with a specific atomistic structure. This information is of profound importance in materials discoveries, because it is one of the key indicators for materials' stabilities (and to some extent, synthesizabilities<sup>2,3</sup>). From here, other physical properties, such as electronic structures, magnetic ground states and optical responses, can be acquired as

auxiliaries to a DFT calculation, since it solves a good approximation to the fundamental physical equation that governs the quantum mechanical behaviours of electrons in materials.

One of the key drawbacks of DFT is its  $\mathcal{O}(N^3)$  scaling behaviour to the system size measured as the number of atoms  $N$ , which makes it computationally very expensive to apply in large-scale modellings, such as for chemically disordered high-entropy materials,<sup>4</sup> and systems in which many-body interactions significantly dictate their physical properties.<sup>5</sup> Traditionally, this bottleneck was overcome by employing atom-atom force fields<sup>6</sup> that are designed with fast-to-calculate analytical functions based on known physics (*e.g.* harmonic potential for bond stretching) with parameters fitted to a single or a specific set of material(s). However, this also imposes significant limitations in applying these tailor-made force fields to correctly model exotic materials behaviours such as anharmonic phonon vibrations and Coulomb interactions between polarised charge densities, thus restricting the model transferability to systems that have not been parameterised. This makes the development of universal force field a long-term challenge in materials modelling. This, nevertheless, is not a problem for DFT.

<sup>a</sup> School of Material Science and Engineering, University of New South Wales, Sydney, New South Wales 2052, Australia. E-mail: [jianliang.yang1@unsw.edu.au](mailto:jianliang.yang1@unsw.edu.au)

<sup>b</sup> Jiangxi Provincial Key Laboratory of Advanced Electronic Materials and Devices, Jiangxi Science and Technology Normal University, Nanchang 330018, China

<sup>c</sup> School of Physics, University of Electronic and Technology of China, Chengdu 610054, China



Over the past two decades, machine-learning interatomic potentials (MLIPs) have emerged and rapidly developed to bridge the gap between DFT- and force-field-based energy models. The early successes in this endeavour are largely rooted in using kernel regressions based on hand-crafted and physically inspired descriptors for local atomic environments.<sup>7,8</sup> This knowledge has been fueled into the recent development of deep-learning potential models such as Schnet,<sup>9</sup> NequIP,<sup>10</sup> MACE<sup>11</sup> and So3krates.<sup>12</sup> Some of them<sup>12</sup> have further incorporated deep-learning architectures, such as the attention mechanism<sup>13</sup> from the large language models to capture long-ranged atom–atom interactions in materials, demonstrating the cross-paradigm nature in this field of research. In the meantime, the continuous expansions of large computational materials databases, such as the Materials Project<sup>14</sup> and OMAT,<sup>15</sup> have provided the community with rich resources of structural, energetic and property data of materials that are generated in a consistent level of theory. The scale and diversity of hundreds of millions of first-principles calculations provided by these databases unlock our capabilities to develop a transferable universal foundation energy model for (solid-state) materials across a significant portion of the existing chemical space.<sup>16</sup>

This significant milestone can be exemplified by the recent achievement behind the releases of a suite of foundation models<sup>17</sup> based on the MACE<sup>11</sup> (message-passing atomic cluster expansion) architecture, which is the focus of this study. More specifically, to learn the total atomic interaction energies in chemical systems, MACE combines the graph neural network<sup>18</sup> that models chemical structures as graphs and utilises the message-passing mechanism<sup>19</sup> to exchange chemical bonding information across multiple message-passing layers in the network, together with the atomic cluster expansion<sup>20</sup> formalism to ensure that the equivariance of the local atomic environment is preserved as the messages are passed through the network.

In the first release,<sup>17</sup> dubbed as mp-0, the foundation model was trained on the MPtrj<sup>21</sup> dataset, which contains a large number of static calculations and structural optimisation trajectories for inorganic solids at the PBE+*U* level of theory. This includes approximately 1.5 M structures with 90% of them having less than 70 atoms per unit cell. With this level of coverage, the mp-0 model had been applied to demonstrate its applicability to 30 different categories of atomistic simulations, ranging from ice structures, metal organic frameworks, heterogeneous catalysts, amorphous structures, to complex liquid–solid interfaces.

However, even at this training scale, the intrinsic problem associated with any MLIPs cannot be overlooked in the

foundation model, that is, at its best, the model accuracy can only be as good as the underlying theory that was applied to generate the training data. This issue has already been addressed,<sup>17</sup> for example, the DFT setting for generating the MPtrj dataset is less tight compared to that required for accurate phonon calculations, as such, the error in reproducing the DFT phonon bandwidth with mp-0 is  $\sim 1\text{--}2$  THz, which is an order of magnitude larger than the results from a highly specialised model.<sup>22</sup> Overcoming this kind of shortage is undoubtedly a key driving force for the ongoing improvement of the MACE-foundation models (Table 1). This is because many key physical properties of materials, such as dynamic stabilities,<sup>23,24</sup> electron dynamics and superconductivities,<sup>5</sup> all share strong links to the phononic behaviours of the materials. A notable improvement in predicting phonon properties is expected with the latest iteration of the model that was trained on the OMAT<sup>15</sup> database.

This is an interesting development, as the major improvement from MPtrj to OMAT was not necessarily on DFT settings that improved the phonon accuracy, but an expansion in the dataset size which contains, for example, rattled crystal structures sampled from Boltzmann distributions as well as molecular dynamic trajectories. This highlights the importance of including more training data that can closely trace the topologies of the underlying DFT potential energy surfaces (PES) for different materials as a key strategy for developing foundational models for materials chemistry.

A particularly relevant case is anharmonic<sup>27</sup> solids, for which vibrating atoms tend to traverse a PES with complex topology that notably deviates from the idealised parabolic shape. A representative material system is the cubic perovskites, for which the high-symmetry cubic structure is a saddle point on a double-well-shaped PES, that can be expressed as a fourth-order polynomial.<sup>28</sup> Solving the eigenvalue equation for the dynamical matrix of harmonic phonons for these systems typically leads to imaginary phonon frequencies<sup>29</sup> at the high symmetry points in the reciprocal space, which correspond to the structurally-related antiferrodistortive<sup>30</sup> or electronically-related ferroelectric<sup>31</sup> instabilities. Distorting the high-symmetry cubic perovskite structure along the eigenvectors of these imaginary phonon eigenvectors corresponds to symmetry-breaking events that will drive the structure into an energetically more stable state. Physically, the depth of the double-well potential has a strong contribution towards the degree of vibrational anharmonicities. The latter is strongly related to the chemical constituents and bonding characteristics of the materials.

The above idea inspires our present investigation, in which we use our unique harmonic phonon and room-temperature

Table 1 Overview of the MACE foundation models that are benchmarked in this study

Model name	Elements covered	Training dataset	Level of theory	Notes
matpbs-pbe-omat-ft	89	MATPES-PBS <sup>25</sup>	DFT (PBE)	No + <i>U</i> correction
mpa-0-medium	89	MPtrj <sup>21</sup> + sAlex <sup>26</sup>	DFT (PBE+ <i>U</i> )	Improved accuracy particularly high pressure stabilities
mp-0b3-medium	89	MPtrj	DFT (PBE+ <i>U</i> )	Improved phonon properties
omat-0-medium	89	OMAT <sup>15</sup>	DFT (PBE+ <i>U</i> )	Excellent phonon properties



*ab initio* molecular dynamics (AIMD) database of  $\sim 2000$  halide double perovskites (HDPs),<sup>32</sup> covering a diverse range of material dynamic stabilities and vibrational anharmonicities,<sup>27</sup> while maintaining structural homogeneity (all having  $Fm\bar{3}m$  space group symmetry), to benchmark the performances of the MACE foundation models (Table 1) in tracing the topologies of PES across a range of different degrees of anharmonicity.

Our detailed analysis reveals the following. The previously established anharmonicity score<sup>27</sup> is fundamentally equivalent to a measurement of force-fitting residue,<sup>33</sup> which can be used to reveal (a) part of the chemical space where the foundational models performed well (poorly) in reproducing the DFT–PES, as well as (b) regions of the DFT–PES for an individual material that are well (poorly) reproduced by the foundation model. Overall, it shows that the highly anharmonic part of the chemical space and the DFT–PES for individual material are generally less well reproduced by the foundation model. Nevertheless, if both the harmonic and anharmonic contributions to the atomic forces are computed consistently with the same energy model, it should provide a reasonably good indication of the dynamic stabilities of a material that is quantitatively aligned with the DFT result, suggesting these foundation models<sup>17</sup> are indeed sufficient for accelerating large-scale screening of finite-temperature materials stabilities, which is a critical component in the theory-driven materials discoveries.

## 2 Methodologies

### 2.1 HDP database

For details of the DFT calculations that are used to generate the database, as well as the chemical space covered, we refer the readers to our original publication.<sup>32</sup> Details for accessing the database are provided in the SI. All DFT calculations were performed at the PBE (Perdew–Burke–Ernzerhof)<sup>34</sup> level of theory without Hubbard  $U$  correction, which is broadly consistent with the parameterisation level of the MACE foundation models (Table 1). With respect to the current study, the most important DFT data for each HDP that is contained in this database includes:

(1) Harmonic force constant matrix  $\Phi_{ij}^{\alpha\beta}$  computed from the finite-displacement approach in real space.<sup>35</sup> Physically, each matrix element of the force constant matrix corresponds to the force appearing on atom  $i$  along the  $\alpha$  Cartesian direction when atom  $j$  is displaced along the  $\beta$  direction. The availability of the force constant matrix enables us to surrogate a  $(3N + 1)$ -dimensional (with  $N$  being the number of atoms in the simulation supercell) harmonic approximation to the PES in the vicinity of the local minimum that corresponds to the high-symmetry  $Fm\bar{3}m$  structure of HDP. Diagonalising the Fourier transformation of  $\Phi$  will provide us with the phonon eigenfrequencies  $\{\omega(\mathbf{q},n)\}$ , where  $\mathbf{q}$  is the phonon wavevector in the first Brillouin zone, and  $n$  is the band index for a given  $\mathbf{q}$ . The phonon dispersion relationship can be obtained by connecting  $\{\omega(\mathbf{q},n)\}$  with the same  $n$  across all symmetrically unique  $\mathbf{q}$ -points in the first Brillouin zone, from which one can

compute the corresponding phonon group velocities as  $\mathbf{v}_g(\mathbf{q},n) = \partial\omega(\mathbf{q},n)/\partial\mathbf{q}$ .

(2) AIMD trajectory which contains a set of time-dependent atomic coordinates and forces  $\{\mathbf{R}(t), \mathbf{F}(t)\}$  that are sampled at 300 K for up to 1.6 ps at 1 fs time step under the NVT ensemble. AIMD simulations enable us to sample a wider (higher-energy) portion of the energy basin that is centred around the  $Fm\bar{3}m$  local minimum. Since DFT does not take any assumption on the topology of the underlying PES (as opposed to traditional force fields), but solely determines the local PES gradient (encapsulated in  $\mathbf{F}(t)$ ) by solving the electronic structure at the given atomic configuration  $\mathbf{R}(t)$ , it is able to capture the nonparabolic aspect in the topology of the PES, especially at distant to the local minimum.

### 2.2 Anharmonicity score

By combining the information of harmonic force constants and AIMD trajectories, Knoop *et al.*<sup>27</sup> proposed the following score to measure the degree of vibrational anharmonicity of a material at a given temperature  $T$ :

$$\sigma^{(2)} = \sqrt{\frac{\sum_{i,\alpha} \langle (F_i^{\alpha,A})^2 \rangle_T}{\sum_{i,\alpha} \langle (F_i^\alpha)^2 \rangle_T}}. \quad (1)$$

Essentially, the anharmonicity is measured by comparing the standard deviation of the total ( $F$ ) and anharmonic ( $F^A$ ) atomic forces sampled across the AIMD trajectory. Here,  $F_i^\alpha$  ( $F_i^{\alpha,A}$ ) is the total and anharmonic force on the  $i$ -th atom in the simulation cell along the  $\alpha$ -Cartesian direction, and they are related to each other *via*  $F_i^{\alpha,A} = F_i^\alpha - F_i^{\alpha,(2)}$ , in which the harmonic component of the atomic force can be computed from the force constant  $\Phi_{ij}^{\alpha\beta}$  as  $F_i^{\alpha,(2)} = -\sum_{j,\beta} \Phi_{ij}^{\alpha\beta} \mathbf{u}_j^\beta$ , with  $\mathbf{u}_i^\alpha$  being the atomic

displacement from its equilibrium position. Summation over all atoms and three Cartesian directions for a given AIMD frame gives the time-dependent  $\sigma^{(2)}(t)$ , which provides a measure of anharmonicity for the particular frame, whereas taking the average  $\langle \sigma^{(2)}(t) \rangle_t$  over the entire AIMD trajectory provides a single numerical measure of the anharmonicity of a given material at  $T$ . The later also determines the dynamical stability of the materials, as those with  $\langle \sigma^{(2)}(t) \rangle_t > 1$  are considered as unstable at the simulated temperature  $T$ .

There are two important aspects of the anharmonicity score, which is rooted from its definition eqn (1). Firstly, as the harmonic force is directly proportional to the atomic displacements,  $\sigma^{(2)}(t)$  can be treated as a single-valued proxy to gauge the range of the phase space being sampled in an MD simulation.<sup>36</sup> Secondly, eqn (1) shows that the anharmonicity score is fundamentally a measure of standard deviation, which is also a measure of force fitting accuracy in all MLIP developments,<sup>33</sup> hence, as shall be shown below, it can provide us with more physical and diagnostic insight into the chemical and structural phase spaces in which the foundation models performed well or extrapolate poorly in practical simulations.



### 2.3 Configurational space analysis

Unsupervised machine learning provides a powerful way to compare the configurational spaces sampled with two different energy models, in this case, the DFT and MACE foundation model (more specifically, the omat-0-medium model), which will enable us to understand more deeply the discrepancies in the dynamic stabilities of HDPs that are acquired from these two different energy models. For this purpose, we first mathematically encode each MD frame with the SOAP<sup>37</sup> (smoothed overlap of atomic positions) structure descriptor. Technically, all atoms in the simulation supercells were included as the ‘centres’ on which their surrounding atomic environments are considered in constructing the structure descriptor for the crystal, with the periodic boundary conditions taken into account. The radial cut-off distance for finding the neighbouring atoms to each centre is  $r_c = 5 \text{ \AA}$ . Each atom is modelled as a normal distribution centred at its Cartesian coordinates, with a standard deviation of  $\sigma_c = 0.1 \text{ \AA}$ . The number of basis functions used to expand the radial and angular distribution of the atomic environment around each centre are set to  $n_{\text{max}} = 7$  and  $\ell_{\text{max}} = 6$ , respectively. The REMatch<sup>38</sup> (regularized entropy match) similarity metric is employed to compute the similarity between two multiatomic MD frames, from here, the similarity kernel, which encodes the pairwise structural similarities among all MD frames in the trajectory can be constructed. The SOAP-REMatch kernel is then subsequently used to construct a two-dimensional map with the Kernel principal component analysis (KPCA), enabling us to visually compare the configurational spaces being sampled by AIMD and MACE-MD. The SOAP-REMatch kernel is computed using the dscribe<sup>39</sup> package, and the KPCA analysis is performed with scikit-learn.<sup>40</sup>

## 3 Results and discussion

### 3.1 Harmonic phonons

We first examine the performances of the MACE foundation models in reproducing the key phononic characteristics of

solids computed from periodic DFT. For this purpose, we first recompute the harmonic force constant matrix using the same finite-displacement approach with the same size of  $(2 \times 2 \times 2)$  supercell for each HDP as in our previous study,<sup>32</sup> except now the atomic forces on each finite-displaced supercell structure are computed with the MACE foundation models, from which the harmonic constant matrix  $\Phi_{\text{MACE}}$  can be determined. As detailed in the Methodologies section, diagonalising  $\Phi_{\text{MACE}}$  will give us a set of phonon eigenfrequencies  $\{\omega_{\text{MACE}}(\mathbf{q}, n)\}$  and group velocities  $\{v_{\text{MACE}}(\mathbf{q}, n)\}$ , from which the following two root-mean-squared-errors (RMSE) metrics were applied to gauge the deviation of the MACE predicted phononic properties from those computed with DFT: (1) RMSE in  $\omega^2$ , defined as

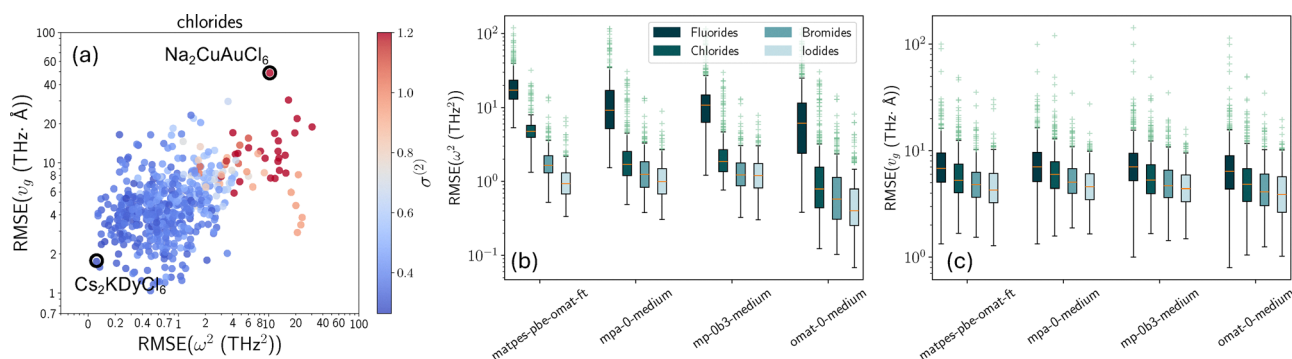
$$\text{RMSE}(\omega^2) = \sqrt{\frac{1}{N_q N_n} \sum_{\mathbf{q}, n} \|\omega_{\text{MACE}}^2(\mathbf{q}, n) - \omega_{\text{DFT}}^2(\mathbf{q}, n)\|}, \quad (2)$$

which eliminates the possible complication of comparing a real and an imaginary phonon eigenfrequency with the same combination of  $\{\mathbf{q}, n\}$ , here  $N_q$  is the total number of wavevectors sampled in the first Brillouin zone and  $N_n$  is the total number of eigenstates for a given eigenvector  $\mathbf{q}$ . Physically, the magnitudes of the phonon eigenfrequencies provide good indications on the mechanical strengths of a solid. (2) RMSE in  $v_g$ , defined as

$$\text{RMSE}(v_g) = \sqrt{\frac{1}{N_q N_n} \sum_{\mathbf{q}, n} (v_{\text{MACE}}(\mathbf{q}, n) - v_{\text{DFT}}(\mathbf{q}, n))^2}, \quad (3)$$

which provides a good indication on reproducing the shape of the DFT-phonon dispersion relationship.

As an example, Fig. 1(a) shows the relationship between  $\text{RMSE}(\omega^2)$  and  $\text{RMSE}(v_g)$  for chloride HDPs computed with the omat-0-medium model. To showcase what these two error metrics reflect, we also show in Fig. S2 of the SI the corresponding comparisons of the phonon dispersion curves computed with the omat-0-medium model and DFT for the



**Fig. 1** Accuracies of the MACE foundation models in predicting the harmonic phonon properties for HDPs. (a) Correlations between the RMSEs in predicting the phonon eigenfrequencies and group velocities with respect to the DFT results for chloride HDPs using the omat-0-medium foundation model. Each data point is colour-coded according to its anharmonicity score  $\sigma^{(2)}$  computed from DFT.<sup>32</sup> Two extrema with the best (lower left) and worst (upper right) prediction accuracies are highlighted, the corresponding phonon dispersion relationships for which are shown in Fig. S2 of the SI. (b) and (c) The box plots that present the ranges of RMSEs in predicting the phonon eigenfrequencies and group velocities using all four foundation models for each group of HDPs as categorised by the halide anion. The orange line indicates the medium RMSE. The box limits represent the 1st and 3rd quartiles. The whiskers show the range of RMSEs within  $1.5 \times$  the interquartile range of the box limits, and the outliers are denoted by light cyan plus symbols.



two extrema ( $\text{Cs}_2\text{KDyCl}_6$  and  $\text{Na}_2\text{CuAuCl}_6$ ) identified in Fig. 1(a). For  $\text{Cs}_2\text{KDyCl}_6$  which has the lowest  $\text{RMSE}(\omega^2)$  value, it can be seen from Fig. S2 (left) that the phonon dispersion curves computed from the MACE model are well overlapped with the DFT ones, except some deviations near 0 THz around the  $X$ -high symmetry point. For the worst case of  $\text{Na}_2\text{CuAuCl}_6$ , Fig. S2 (right) shows the MACE underestimates the imaginary phonon frequencies which consequently led to a flatter dispersion curve that increases  $\text{RMSE}(\mathbf{v}_g)$ . In this particular case, the material that is deemed unstable on the DFT-PES became less dynamically unstable on the MACE-PES.

By further highlighting each point on Fig. 1(a) with the anharmonic score  $\sigma^{(2)}$  computed with DFT for the corresponding HDP structure, a more intriguing trend is revealed, which shows that the accuracy of the phononic properties predicted by the MACE foundation model are strongly correlated with  $\sigma^{(2)}$ , such that structures with high dynamical stabilities (lower  $\sigma^{(2)}$ ) exhibit lower RMSEs, and *vice versa*. This is a systematic trend that occurs in all four halide systems investigated, regardless on which datasets the foundation models had been trained [Fig. S3–S6]. It is not a coincidence, as discussed in the Methodology section, that  $\sigma^{(2)}$  is also a RMSE-type measurement, but with a fundamental geometric insight that captures the deviation of the shape of the PES from a hyperparabola.

In Fig. 1(b) and (c), we show the box plots that provide a more summative view over our benchmark results on the harmonic phonon properties for HDPs. It can be seen that, across the chemical space from fluorides to iodides, the accuracy in predicting the harmonic phonon properties increases as the atomic masses of the halide anions increase. This effect is particularly pronounced in reproducing the phonon eigenfrequencies [Fig. 1(c)]. As shown in our previous work,<sup>32</sup> the vibrational anharmonicities of HDPs do exhibit a systematic decrease from light to heavier halides [Fig. S1(a)], hence the chemical trends behind the RMSE values shown in Fig. 1(b) and (c) are largely consistent with the trend shown in Fig. 1(a) for chlorides with respect to the variations in  $\sigma^{(2)}$ . The observed chemical trend is largely unchanged across all four parameterisations of the MACE foundation models, with the *omat-0-medium* being the best performing model, showing an order of magnitude improvement in  $\text{RMSE}(\omega^2)$  compared to the worst performing *matpes-pbe-omat-ft* model.

On a more fundamental level, the chemical trend observed in RMSEs can be further correlated with the phonon bandwidths (how wide  $\omega$  spans, which can be equivalently reflected from the averaged phonon eigenfrequencies  $\langle\omega\rangle$ ) for HDPs with different halide anions. As shown in Fig. S1(b) extracted from our previous study,<sup>32</sup>  $\langle\omega\rangle$  increases systematically from iodides to fluorides. This indicates that, for lighter halides, the vibrating ions experience larger restoring forces that originate from a steeper topology of the PES. From the perspective of training atom–atom force fields,<sup>41</sup> it is generally understood that steep or sharp rising parts of the PES are more challenging to be accurately trained, which would require more training data and/or more tailored functional forms to reduce the training complexity. In the domain of fully data-driven MLIPs, the

quality of the potential energy model becomes more critically dependent on the breadth of the configuration space covered in the training set. Whilst the *OMAT*<sup>15</sup> dataset already contains rattled atomic structures sampled according to the Boltzmann distribution up to 1000 K, the number of the rattled structures per compound was fixed. Our present findings suggest that, moving forward, a more adaptive scheme in constructing the training sets, particularly applying a weighting scheme to include more rattled structures following the high-frequency phonon modes for systems containing light elements, would be an interesting path to explore for increasing the accuracy of foundational energy models.

### 3.2 Anharmonicity of AIMD-sampled configurations from the MACE foundation models

Whilst harmonic phonon properties are often applied first as a key determinant for materials' dynamical stability, in many cases, they are insufficient for fully characterising the finite-temperature phase stabilities of materials. For example, as shown in Fig. S2 of the SI, the presence of imaginary phonon frequencies (from calculations performed at 0 K) is often considered as an indicator of mechanical instability. This is a typical feature in many perovskites, however, upon the rise of temperature, the collective vibrations of ions in the material change the crystal potential that is experienced by the vibrating ions, a physical effect that can be captured by MD simulations. Consequently, the imaginary phonon frequencies become thermally 'renormalised' into real ones,<sup>42</sup> *i.e.* the material is thermally stabilised at the finite temperature. This shows that MD simulations are essential for fully characterising the finite-temperature phase stabilities of materials, and the capability for MLIPs to generate an ensemble of configurations at a given temperature stably is an important criterion to benchmark the quality of MLIPs.<sup>43</sup>

Nevertheless, directly comparing an AIMD trajectory with another one that is independently sampled with a different energy model, in this case, MLIP, it is often difficult to come up with good interpretations that can lead to direct and meaningful physical insights into the qualities of MLIPs. Fundamentally, this is because two different energy models correspond to two different PES, and even a small difference in the PES topologies can shift the equilibrium positions, barrier heights and transition states, such that the two MD trajectories may cover completely different configuration spaces.

To overcome this kind of complication, in this section, we shall first take the existing AIMD trajectory for each HDP<sup>36</sup> (a total of 1682 valid ones), to recompute the atomic forces for each frame in every trajectory with the MACE foundation model, from which a new  $\langle\sigma^{(2)}\rangle_t^{\text{MACE}}$  metric ( $\sigma_{\text{MACE}}$  for short-handed notation) can be attained, that is to be directly compared with  $\langle\sigma^{(2)}\rangle_t^{\text{DFT}}$  ( $\sigma_{\text{DFT}}$  for short-handed notation). More specifically, for each AIMD trajectory and MACE foundation model that we benchmarked, we compute  $\sigma^{\text{MACE}}$  with two different approaches to obtain the harmonic component of the atomic forces ( $F_i^{\alpha,(2)} = -\Phi_{ij}^{\alpha\beta} u_j^\alpha$ ) *via* the force constant  $\Phi_{ij}^{\alpha\beta}$ : (a)  $\Phi_{\text{DFT}}$  approach, where the atomic forces for each displaced configuration generated from the finite-displacement method<sup>35</sup> are



computed with DFT<sup>32</sup> to construct the force constant  $\Phi$ , and (b)  $\Phi_{\text{MACE}}$ -approach, with the atomic forces computed with the MACE foundation models.

Geometrically, the  $\Phi_{\text{DFT}}$ -approach can be considered as a way of providing a direct measure of the ability of the MACE energy models to exactly reproduce the topologies of the DFT-PES around the local minimum. When  $\sigma_{\text{MACE}} > \sigma_{\text{DFT}}$ , the MACE model overestimates the total forces, leading to a more anharmonic PES compared to the DFT baseline, which is the other way around when  $\sigma_{\text{MACE}} < \sigma_{\text{DFT}}$ . In other words, the discrepancy between  $\sigma_{\text{MACE}}$  and  $\sigma_{\text{DFT}}$  provided an absolute measure on the errors in predicting the total atomic forces. When  $\sigma_{\text{MACE}} = \sigma_{\text{DFT}}$ , the DFT energy landscape can be fully reconstructed by the MACE energy models over all  $\{\mathbf{u}\}$ . For the  $\Phi_{\text{MACE}}$ -approach,  $\sigma_{\text{MACE}}$  takes no reference to the DFT-energy landscape, thus it reflects the degree of anharmonicity of the MACE-energy landscape itself. When  $\sigma_{\text{MACE}} = \sigma_{\text{DFT}}$ , it means that the relative contributions from the (an)harmonic force components to the total atomic forces are the same between the MACE and DFT energy models, and the MACE and DFT-PES differ from each other globally by some constant multiplicative factor.

Physically, comparing  $\sigma_{\text{MACE}}$  with  $\sigma_{\text{DFT}}$  provides a key indication to whether the degrees of finite-temperature dynamic stabilities of materials predicted by the MACE foundation models agree with those predicted by the DFT. In particular, the  $\Phi_{\text{MACE}}$ -approach provides a looser criterion in making this judgement as it only requires the relative contributions of the anharmonic forces to the total one being the same as predicted from MACE models and DFT, which may benefit from error cancellations in subtracting  $\Phi_{ij}^{\alpha\beta} u_i^\alpha$  from total atomic force  $F_i^\alpha$ , when both terms are predicted from the MACE models. In contrast, the  $\Phi_{\text{DFT}}$ -approach is more strict, which would require the total atomic force predicted from the MACE-models to closely match those computed from DFT. This information is presented with the confusion matrices shown in Fig. 3. In each confusion matrix, the dynamic stabilities are characterised into three categories<sup>33</sup> [Fig. 2]: (a)  $\sigma \in (0, 0.5)$ , corresponding to highly stable structures with weak vibrational anharmonicity that is predominantly contributed by three-phonon scatterings (encapsulated by the third-order force constant  $\Phi_{ijk}^{\alpha\beta\gamma}$ ). (b)  $\sigma \in [0.5, 1]$ , meaning the phase is still stable at the simulated temperature  $T$  but with stronger vibrational anharmonicity that is dominated by the force constants from fourth-order and above. And finally, (c)  $\sigma > 1$ , meaning the phase is unstable at  $T$ .

Results in Fig. 3 show that the confusion matrices are dominated by the diagonal elements, meaning that the consensus in predicting the phase stabilities at 300 K between the MACE foundation models and DFT are generally acceptable across a wide range of stability regimes, supporting the argument that MACE foundation models is useful for fast pre-screening filter for unstable materials.<sup>17</sup>

In the  $\Phi_{\text{DFT}}$ -approach, one sees a clear trend of increasing the number of correctly predicting materials' dynamic stabilities from the matpes-pbs-omat-ft to the omat-0-medium model, which is in line with the observations above. The likelihood of the MACE models to overly stabilise (destabilise) the DFT-predicted unstable

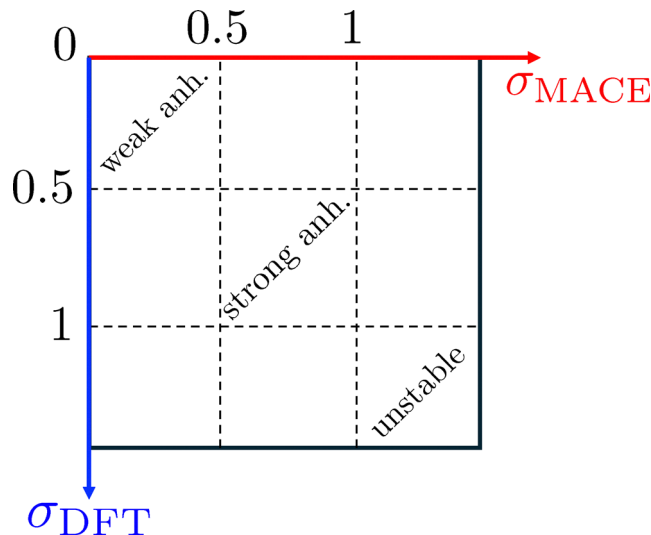


Fig. 2 Ranges of anharmonic scores computed from DFT and MACE models that are covered by different blocks of the confusion matrix. In particular, the diagonal blocks represent cases where the MACE model predicts the same dynamical stabilities as from DFT calculations, and they are categorised into cases corresponding to systems that are (a) weakly anharmonic ( $\sigma < 0.5$ ), (b) strongly anharmonic ( $\sigma \in [0.5, 1]$ ) and (c) dynamically unstable ( $\sigma > 1$ ), following the results from our previous high-throughput screening studies on single perovskites.<sup>33</sup> This convention is followed when presenting the results in Fig. 3 and 4.

(stable) materials, *i.e.*  $\sigma_{\text{MACE}} > 1$  for  $\sigma_{\text{DFT}} < 1$  or *vice versa* is generally low. Except for the matpes-pbs-omat-ft and mp-0b3-medium models, for which we see a significant number of weakly anharmonic HDPs being classified as strongly anharmonic by the MACE models.

The  $\Phi_{\text{MACE}}$ -approach reveals a different outcome. In this case, the number of HDPs that have their dynamical stabilities being correctly identified remain almost unchanged across different MACE models. As discussed above, this means that, across a large part of the chemical space, the relative anharmonic contributions to the overall topologies of the PES remain invariant from DFT to the different MACE models. However, we also observed from Fig. 3 that when the  $\Phi_{\text{MACE}}$  is used to extract the harmonic components of the atomic forces, the number of HDPs being placed in the lower off-diagonal parts of the confusion matrices increased significantly, meaning that when the MACE model is used solely to compute  $\sigma$ , it tends to over-stabilise the highly anharmonic and unstable HDPs (see Fig. S7(d) in the SI for an example). This observation can be better reflected when we plot the distributions of  $\langle \sigma_t^{(2)} \rangle^{\text{MACE}} - \langle \sigma_t^{(2)} \rangle^{\text{DFT}}$  in Fig. S9 of the SI, which show strong tailing in the negative part. As discussed in the harmonic phonon section, this reflects the poorer generalisability of the MACE foundational models in capturing the anharmonic features of the PES, particularly for materials with low dynamical stabilities.

### 3.3 Anharmonicity of HDPs computed solely from the omat-0-medium model

In the practical applications where MLIP is used to determine the dynamic stabilities of materials, both the harmonic force



	matpes-pbs-omat-ft	mpa-0-medium	mp-0b3-medium	omat-0-medium																																				
$\Phi_{\text{DFT}}$	<table border="1"> <tr><td>717</td><td>248</td><td>1</td></tr> <tr><td>19</td><td>390</td><td>65</td></tr> <tr><td>0</td><td>8</td><td>234</td></tr> </table>	717	248	1	19	390	65	0	8	234	<table border="1"> <tr><td>915</td><td>55</td><td>0</td></tr> <tr><td>52</td><td>399</td><td>26</td></tr> <tr><td>0</td><td>24</td><td>219</td></tr> </table>	915	55	0	52	399	26	0	24	219	<table border="1"> <tr><td>856</td><td>114</td><td>0</td></tr> <tr><td>45</td><td>399</td><td>33</td></tr> <tr><td>0</td><td>20</td><td>223</td></tr> </table>	856	114	0	45	399	33	0	20	223	<table border="1"> <tr><td>940</td><td>29</td><td>0</td></tr> <tr><td>49</td><td>414</td><td>14</td></tr> <tr><td>1</td><td>27</td><td>215</td></tr> </table>	940	29	0	49	414	14	1	27	215
717	248	1																																						
19	390	65																																						
0	8	234																																						
915	55	0																																						
52	399	26																																						
0	24	219																																						
856	114	0																																						
45	399	33																																						
0	20	223																																						
940	29	0																																						
49	414	14																																						
1	27	215																																						
$\Phi_{\text{MACE}}$	<table border="1"> <tr><td>925</td><td>41</td><td>0</td></tr> <tr><td>155</td><td>314</td><td>5</td></tr> <tr><td>26</td><td>114</td><td>102</td></tr> </table>	925	41	0	155	314	5	26	114	102	<table border="1"> <tr><td>960</td><td>10</td><td>0</td></tr> <tr><td>164</td><td>304</td><td>9</td></tr> <tr><td>28</td><td>102</td><td>113</td></tr> </table>	960	10	0	164	304	9	28	102	113	<table border="1"> <tr><td>926</td><td>44</td><td>0</td></tr> <tr><td>145</td><td>317</td><td>15</td></tr> <tr><td>19</td><td>89</td><td>135</td></tr> </table>	926	44	0	145	317	15	19	89	135	<table border="1"> <tr><td>962</td><td>7</td><td>0</td></tr> <tr><td>150</td><td>316</td><td>11</td></tr> <tr><td>23</td><td>102</td><td>118</td></tr> </table>	962	7	0	150	316	11	23	102	118
925	41	0																																						
155	314	5																																						
26	114	102																																						
960	10	0																																						
164	304	9																																						
28	102	113																																						
926	44	0																																						
145	317	15																																						
19	89	135																																						
962	7	0																																						
150	316	11																																						
23	102	118																																						

Fig. 3 Table of confusion matrices showing how well the MACE models (across the columns) reproduce the vibrational anharmonicity of DHPs computed from DFT (across the rows), which is represented as the number of DHPs that fall into each category. The top (bottom) row presents the anharmonicity scores evaluated using force constant matrix computed from DFT (corresponding MACE models shown on the top row).

constants and the finite-temperature MD sampling would have been conducted with the same MLIP, with little or no reference to prior DFT results. Hence, in this section, we shall present some further results and analysis on determining the vibrational anharmonicity of HDPs solely based on the omat-0-medium, which was shown to be the best performing model from the previous sections.

Computationally, the MD samplings using the MACE foundation model, dubbed as MACE-MD, are carried out as follows. For each HDP, we randomly selected 2 frames from the previously sampled AIMD trajectory as the starting points to perform 2 independent MACE-MD samplings. This choice of starting points for MACE-MD imposes a weak constraint that the configurational space that is sampled by the MACE-MD should have some overlap with the configurational space sampled from AIMD, at least in the initial stage of the MACE-MD sampling. Each MACE-MD simulation was run for 2 fs at a 1 ps time step using the MD engine from the atomic simulation environment.<sup>44</sup> The same as in our previous work,<sup>32</sup> the Andersen thermostat<sup>45</sup> with a collision probability of 0.5 was

employed to maintain the simulation temperature at the target value of 300 K. The corresponding  $\langle \sigma_t^{(2)} \rangle_{\text{MACE}}$  was averaged over all 4000 MACE-MD frames using the force constant  $\Phi_{\text{MACE}}$  computed with the same omat-0-medium model to extract the harmonic components of the atomic forces.

Fig. 4 presents the confusion matrices that compare the numbers of HDPs that share the same/different classifications of their dynamic stabilities as solely determined from either the DFT or the omat-0-medium foundation model. Similar to the results shown in Fig. 3, the confusion matrix [Fig. 4(a)] is still dominated by the diagonal elements, meaning that the performance in determining the room-temperature dynamic stabilities using the omat-0-medium foundation model alone is generally acceptable. Counting the numbers in the lower diagonal part of the confusion matrix, we found 38% chance of categorising HDPs with low stabilities to be more stable ones. In contrast, the upper diagonal part of the confusion matrix leads to only 6% chance of misplacing stable materials to be less stable, which predominantly comes from the fluorides [Fig. 4(b)]. This shows that the omat-0-medium model leads

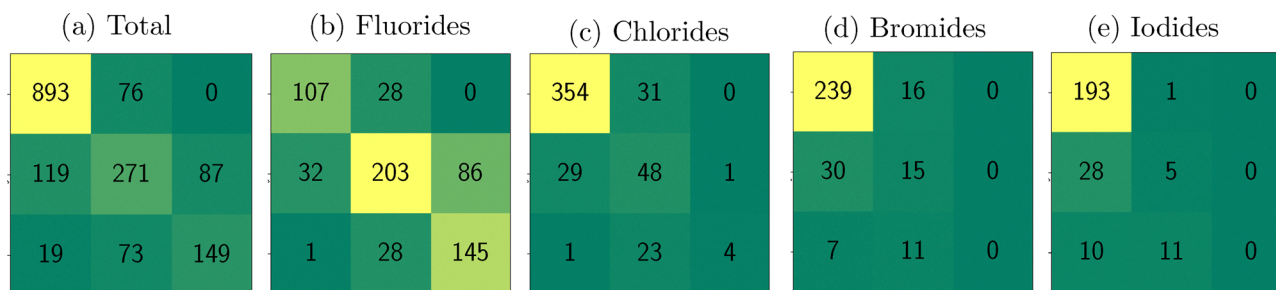


Fig. 4 Confusion matrices showing the reproducibility of the DFT-computed vibrational anharmonicity for DHPs using the MACE model. Here, the omat-0-medium model is used for both computing the harmonic force constants, as well as the molecular dynamics samplings.



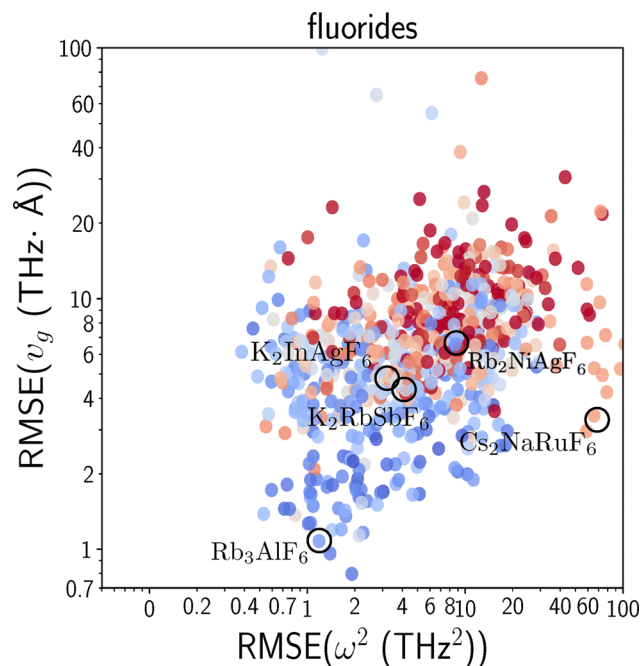
to more false positive cases than false negative one. This means that the chance of missing stable materials is lower, compared to including more unstable materials, when it comes to (pre-)screening dynamically stable materials using the omat-0-medium model, whereby more accurate models (such as DFT) can be used subsequently to further filter out the false positive results.

As mentioned above, comparing two energy models using results from MD simulations may introduce bias because the subtle differences in the PES topologies underpinned by the two energy models may cause MD simulations to sample two distinctly different configurational spaces that intrinsically exhibit different properties. To assess the extent of this bias that could have contributed to the results that are presented in Fig. 4, we have selected five extreme cases among the fluoride compounds (Table 2) and performed unsupervised machine learning to compare the similarities in the configurational spaces that are sampled by the AIMD and MACE-MD [see the Methodologies section for details]. Results from such analysis (Fig. 6) show that, first of all, the way we selected the initial structures for running the MACE-MD simulations did mitigate some of the bias by enforcing the configurational spaces sampled by the two different energy models to (at least partially) overlap with each other. The system that exhibits the largest overlap in the configuration spaces sampled by the two energy models is  $\text{Rb}_3\text{AlF}_6$ , of which the computed  $\sigma_{\text{MACE}}$  is literally identical to  $\sigma_{\text{DFT}}$  (Table 2). In this case, we can consider the DFT-PES around the local minimum for the cubic  $\text{Rb}_3\text{AlF}_6$  has been well reproduced by the omat-0-medium model. In contrast,  $\text{K}_2\text{RbSbF}_6$  represents the other extreme case where the configuration space sampled by the MACE-MD diverges quite significantly from the one sampled with AIMD. The other three compounds listed in Table 2 are somewhere between these two extreme cases, as revealed in Fig. 6.

To check that the above observations are not necessarily biased by the longer trajectories that are sampled by the numerically more efficient MLIP, we performed extra simulations which extended the original AIMD trajectories to 4 ps in length, and re-performed the same KPCA analysis. With the longer AIMD trajectory, Fig. S11 of the SI shows the divergence between AIMD and MACE-MD trajectories for  $\text{K}_2\text{RbSbF}_6$  has reduced, but for  $\text{Rb}_2\text{NiAgF}_6$ , the divergence between the two trajectories increased. Taking into account the stochasticity of MD simulations when interpreting the KPCA maps that are shown in Fig. 6 and Fig. S12 of the SI, we can say that the topologies of the PES underpinned by DFT and the omat-0-medium model for most

**Table 2** Selected fluoride HDPs, the molecular dynamics trajectories of which sampled from AIMD and MACE-MD, will be compared with the SOAP-REMatch kernel

Compound	$\sigma_{\text{DFT}}$	$\sigma_{\text{MACE}}$	$ \sigma_{\text{DFT}} - \sigma_{\text{MACE}} $
$\text{Rb}_3\text{AlF}_6$	<0.5	<0.5	0.000375
$\text{Rb}_2\text{NiAgF}_6$	<0.5	[0.5, 1]	0.400
$\text{Cs}_2\text{NaRuF}_6$	[0.5, 1]	<0.5	0.663
$\text{K}_2\text{InAgF}_6$	[0.5, 1]	[0.5, 1]	0.00102
$\text{K}_2\text{RbSbF}_6$	[0.5, 1]	>0.5	0.440



**Fig. 5** Selected fluoride HDPs, the molecular dynamics trajectories of which sampled from AIMD and MACE-MD, will be compared with the SOAP-REMatch kernel.

materials under the current investigation should be very similar, rendering sufficient similarities in the configuration spaces that are sampled from these two models. As such, dissimilarities in the sampled configuration spaces are not believed to be strongly contributing to the discrepancies in  $\sigma_{\text{DFT}}$  and  $\sigma_{\text{MACE}}$ .

By further colouring each point on the KPCA maps with the anharmonicity score for the corresponding MD snapshot (Fig. S10 and S12 of the SI), it becomes clear that when  $|\sigma_{\text{DFT}} - \sigma_{\text{MACE}}|$  is large, it can be generally attributed to a systematic error in which  $\sigma_{\text{MACE}}$  computed for the entire MACE-MD trajectory is collectively and significantly different from  $\sigma_{\text{DFT}}$  even in the regions of the configurational space where the overlap between those sampled from AIMD and MACE-MD is significant. This suggests that the error in computing  $\sigma$  must be inherited from the error in computing the Hessian matrix  $\Phi$ , which is indeed supported by Fig. 5 showing that low (high) errors in predicting the phonon group velocities and frequencies generally translate to low (high) discrepancies between predicted values of  $\sigma_{\text{MACE}}$  and  $\sigma_{\text{DFT}}$ . The reason for this, is that, understandably, just as the higher accuracy that is required in calculating the atomic forces for determining the phononic properties with DFT, even small errors in predicting the atomic forces with the foundational models could translate into large notable differences in the phonon dispersion relationship, as the errors are amplified in the calculations of the derivatives of forces with respect to the atomic positions.

## 4 Conclusions and outlooks

Using our own unique database that has characterised the degree of vibrational anharmonicity and room-temperature



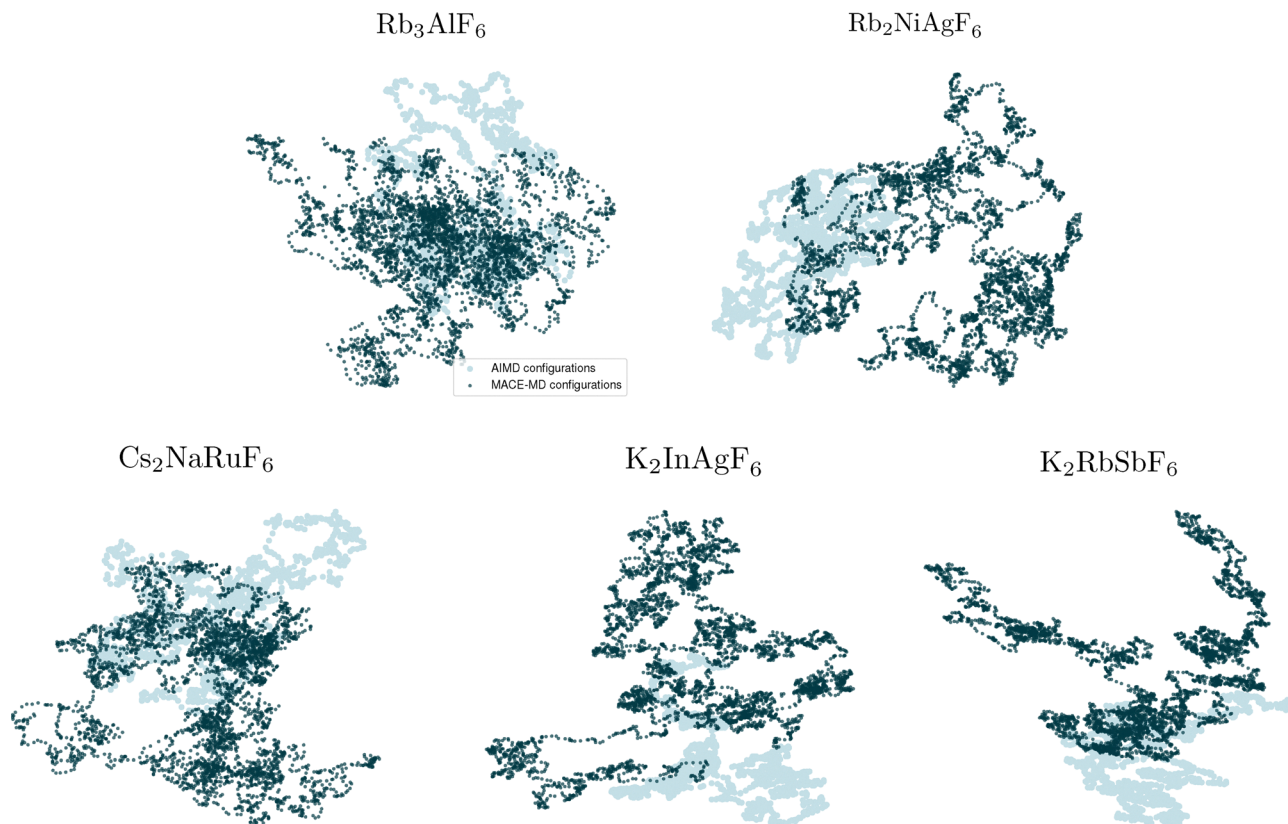


Fig. 6 KPCA maps for the five fluoride HDPs listed in Table 2 that compare the configuration spaces sampled by AIMD<sup>32</sup> and MACE-MD with the omat-0-medium model. Each point on the map corresponds to a configuration in the MD trajectory.

dynamic stabilities of  $\sim 2000$  halide double perovskites, which includes both the harmonic phonon and 300 K-MD simulation data computed at the DFT level of theory, in this study, we have systematically benchmarked four recent variants of the MACE foundation models for inorganic solids, on their capabilities and accuracies in determining materials' dynamic stabilities. This is an important application scenario for the foundation models in computational material discoveries, where phase stabilities predicate all subsequent endeavours of discovering new exotic physical and chemical applications of new materials. Out of the four variants of the MACE foundation models, it is found that the omat-0-medium model performs the best in reproducing both the 0 K-harmonic phonon properties, as well as the room-temperature dynamic stabilities of HDPs that were determined from DFT simulations. Mathematically, the anharmonicity score shares a highly similar form as the standard deviations that are employed to measure the accuracy of the MLIPs, thus it is reasonable to observe that the errors in predicting the harmonic phonon properties using the MACE foundation models showed strong correlation with the structures' anharmonicity scores, whereby weakly (strongly) anharmonic materials exhibit higher (lower) accuracies in such predictions. This suggests that including more data, such as high-temperature MD trajectories, metastable materials, or even hypothetical materials that may be unstable, is important in further developing and/or fine-tuning foundation models to achieve broad applicability and better generalisability.

Based on the above primary findings, we further extended our benchmark by computing the anharmonicity scores for HDPs with both the harmonic force constants and MD samplings solely based on the omat-0-medium model. It is found that the dynamic stabilities determined using this kind of approach correlate well with the DFT results, suggesting that the omat-0-medium model is suitable for accelerating the screening of dynamic materials stabilities for materials discoveries. In more careful examinations of the HDP systems, on which the omat-0-medium model performed well (poorly) in reproducing the anharmonicity scores as determined by DFT, we think it is reasonable to believe that the topologies of the DFT PES are generally well reproduced by the MACE foundation model, whereby considerable overlaps in the configuration spaces sampled by the two approaches can be observed. The (large) discrepancies between the foundation-model- and DFT-predicted anharmonicity scores can be predominantly attributed to the amplification of the errors in predicting the atomic forces with the foundation model when calculating the Hessian matrices. This also shows the possible need of including Hessians in training materials' foundation models, to promise their applications in scenarios where high numerical precision is a must in atomistic materials' modelling.

We hope that the findings presented in this study have provided interesting and useful insights to facilitate the ongoing developments and fine-tunings of materials foundation



models. For instance, proposing metrics such as the anharmonic score is particularly interesting, as it can not only be used for quantifying the model quality, but also be interpreted based on materials' properties to provide more physical insights into understanding the model performances. We envisage that the continuous evolution of the foundation models will further advance the important statistical physics tools in configurational space sampling, particularly in tackling the challenge of meeting the ergodic condition, which bears implications in computing and understanding a wide range of physical and chemical properties of functional materials, such as thermal expansions, lattice thermal conductivities, catalytic activities under realistic (*e.g.* solvated) environments, and many others.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The original DFT data used for benchmarking the MACE foundation models are available on the Harvard Dataverse, for which the URLs to the dataset have been provided explicitly in the supplementary information (SI). Supplementary information is available. See DOI: <https://doi.org/10.1039/d5cp04693a>.

## Acknowledgements

JY acknowledges the support of the Australian Research Council Discovery Projects of DP250100778. SL acknowledges the support of the Australian Research Council Discovery Projects of DP250104633 and Laureate Fellowship of FL250100143. AL acknowledges the support of the National Natural Science Foundation of China under grant no. 12264017.

## References

- W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.*, 1965, **140**, A1133.
- C. J. Bartel, S. L. Millican, A. M. Deml, J. R. Rumptz, W. Tumas, A. W. Weimer, S. Lany, V. Stevanović, C. B. Musgrave and A. M. Holder, Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry, *Nat. Commun.*, 2018, **9**, 4168.
- M. J. McDermott, S. S. Dwaraknath and K. A. Persson, A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis, *Nat. Commun.*, 2021, **12**, 3097.
- C. Oses, C. Toher and S. Curtarolo, High-entropy ceramics, *Nat. Rev. Mater.*, 2020, **5**, 295.
- F. Giustino, Electron-phonon interactions from first principles, *Rev. Mod. Phys.*, 2017, **89**, 015003.
- J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg and B. H. Morrow, Review of force fields

- and intermolecular potentials used in atomistic computational materials research, *Appl. Phys. Rev.*, 2018, **5**, 031104.
- A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, *Phys. Rev. Lett.*, 2010, **104**, 136403.
- R. Jinnouchi, F. Karsai and G. Kresse, On-the-fly machine learning force field generation: Application to melting points, *Phys. Rev. B*, 2019, **100**, 014105.
- K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, Schnet-a deep learning architecture for molecules and materials, *J. Chem. Phys.*, 2018, **148**, 241722.
- S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, *Nat. Commun.*, 2022, **13**, 2453.
- I. Batatia, D. P. Kovacs, G. Simm, C. Ortner and G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, *Adv. Neural Inf. Process. Syst.*, 2022, **35**, 11423.
- J. T. Frank, O. T. Unke and K.-R. Müller, So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems, *arXiv*, 2022, preprint, arXiv:2205.14276, DOI: [10.48550/arXiv.2205.14276](https://doi.org/10.48550/arXiv.2205.14276).
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Attention is all you need, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.
- A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, *et al.*, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.*, 2013, **1**, 011002.
- L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick and Z. W. Ulissi, Open materials 2024 (omat24) inorganic materials dataset and models, *arXiv*, 2024, preprint, arXiv:2410.12771, DOI: [10.48550/arXiv.2410.12771](https://doi.org/10.48550/arXiv.2410.12771).
- A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, Scaling deep learning for materials discovery, *Nature*, 2023, **624**, 80.
- I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Crare, M. Ceriotti, S. Chong, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Eljoius, Z. El-Machachi, E. Fako, F. Falcioni, A. C. Ferrari, J. L. A. Gardner, M. J. Gawkowski, A. Genreith-Schriever, J. George, R. E. A. Goodall, J. Grandel, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. H. Ho, S. Hofmann, C. Holm, J. Jaafar, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, P. Kourtis, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, C. Lin, J. T. Margraf, I.-B. Magdu, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner,



- K. A. Persson, K. Reuter, A. S. Rosen, L. A. M. Rosset, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, C. Sutton, V. Svahn, T. D. Swinburne, J. Tilly, C. van der Oord, S. Vargas, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, T. Wolf, F. Zills and G. Csányi, A foundation model for atomistic materials chemistry, *J. Chem. Phys.*, 2025, **163**, 184110.
- 18 A. Duval, S. V. Mathis, C. K. Joshi, V. Schmidt, S. Miret, F. D. Malliaros, T. Cohen, P. Lio, Y. Bengio and M. Bronstein, A hitchhiker's guide to geometric GNNs for 3D atomic systems, *arXiv*, 2023, preprint, arXiv:2312.07511, DOI: [10.48550/arXiv.2312.07511](https://doi.org/10.48550/arXiv.2312.07511).
- 19 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, *Proceedings of the 34th International Conference on Machine Learning*, 2017, vol. 70, pp. 1263–1272.
- 20 R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, *Phys. Rev. B*, 2019, **99**, 014104.
- 21 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, *Nat. Mach. Intell.*, 2023, **5**, 1031.
- 22 J. George, G. Hautier, A. P. Bartók, G. Csányi and V. L. Deringer, Combining phonon accuracy with high transferability in Gaussian approximation potential models, *J. Chem. Phys.*, 2020, **153**, 044104.
- 23 R. P. Stoffel, C. Wessel, M.-W. Lumey and R. Dronskowski, Ab initio thermochemistry of solid-state materials, *Angew. Chem., Int. Ed.*, 2010, **49**, 5242.
- 24 L. Monacelli, R. Bianco, M. Cherubini, M. Calandra, I. Errea and F. Mauri, The stochastic self-consistent harmonic approximation: calculating vibrational properties of materials with full quantum and anharmonic effects, *J. Phys.:Condens. Matter*, 2021, **33**, 363001.
- 25 A. D. Kaplan, R. Liu, J. Qi, T. W. Ko, B. Deng, J. Riebesell, G. Ceder, K. A. Persson and S. P. Ong, A foundational potential energy surface dataset for materials, *arXiv*, 2025, preprint, arXiv:2503.04070, DOI: [10.48550/arXiv.2503.04070](https://doi.org/10.48550/arXiv.2503.04070).
- 26 M. J. Mehl, D. Hicks, C. Toher, O. Levy, R. M. Hanson, G. Hart and S. Curtarolo, The AFLOW library of crystallographic prototypes: part 1, *Comput. Mater. Sci.*, 2017, **136**, S1–S828.
- 27 F. Knoop, T. A. Purcell, M. Scheffler and C. Carbogno, Anharmonicity measure for materials, *Phys. Rev. Mater.*, 2020, **4**, 083809.
- 28 J. Yang, Composition-dependent chemical and structural stabilities of mixed tin-lead inorganic halide perovskites, *Phys. Chem. Chem. Phys.*, 2020, **22**, 19787.
- 29 I. Pallikara, P. Kayastha, J. M. Skelton and L. D. Whalley, The physical significance of imaginary phonon modes in crystals, *Electron. Struct.*, 2022, **4**, 033002.
- 30 J. Klarbring and S. I. Simak, Nature of the octahedral tilting phase transitions in perovskites: A case study of CaMnO<sub>3</sub>, *Phys. Rev. B*, 2018, **97**, 024108.
- 31 I. B. Bersuker, Pseudo-Jahn-Teller Effect: A Two-State Paradigm in Formation, Deformation, and Transformation of Molecular Systems and Solids, *Chem. Rev.*, 2013, **113**, 1351.
- 32 J. Yang, J. Fan and S. Li, Mapping the room-temperature dynamic stabilities of inorganic halide double perovskites, *Chem. Mater.*, 2022, **34**, 9072.
- 33 J. Yang and S. Li, An atlas of room-temperature stability and vibrational anharmonicity of cubic perovskites, *Mater. Horiz.*, 2022, **9**, 1896.
- 34 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- 35 A. Togo and I. Tanaka, First principles phonon calculations in materials science, *Scr. Mater.*, 2015, **108**, 1.
- 36 J. Yang, Mapping temperature-dependent energy-structure-property relationships for solid solutions of inorganic halide perovskites, *J. Mater. Chem. C*, 2020, **8**, 16815.
- 37 A. P. Bartók, R. Kondor and G. Csányi, On representing chemical environments, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
- 38 S. De, A. P. Bartók, G. Csányi and M. Ceriotti, Comparing molecules and solids across structural and alchemical space, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13754.
- 39 L. Himanen, M. O. Jäger, E. V. Morooka, F. F. Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, DSCRIBE: Library of descriptors for machine learning in materials science, *Comput. Phys. Commun.*, 2020, **247**, 106949.
- 40 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825.
- 41 M. J. Van Vleet, A. J. Misquitta, A. J. Stone and J. R. Schmidt, Beyond Born-Mayer: Improved models for short-range repulsion in *ab initio* force fields, *J. Chem. Theory Comput.*, 2016, **12**, 3851.
- 42 T. Tadano and S. Tsuneyuki, First-principles lattice dynamics method for strongly anharmonic crystals, *J. Phys. Soc. Jpn.*, 2018, **87**, 041015.
- 43 X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli and T. Jaakkola, Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations, *arXiv*, 2022, preprint, arXiv:2210.07237, DOI: [10.48550/arXiv.2210.07237](https://doi.org/10.48550/arXiv.2210.07237).
- 44 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms, *J. Phys.:Condens. Matter*, 2017, **29**, 273002.
- 45 H. C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.*, 1980, **72**, 2384.

