Supplementary Information (SI) for Journal of Materials Chemistry C.

Machine Learning Assisted Screening of Intrinsic Rattling Compounds with Large Atomic Displacement

Kunpeng Yuan[1*], Zhaoxuan Feng[2], Xiaoliang Zhang[3], Dawei Tang[3]

[1]*College of New Energy, China University of Petroleum (East China), Qingdao 266580, China*.
[2]*State Key Laboratory of Heavy Oil Processing, College of Chemistry and Chemical Engineering, China University of Petroleum (East China), Qingdao 266580, China*.
[3]*Key Laboratory of Ocean Energy Utilization and Energy Conservation of Ministry of Education, School of Energy and Power Engineering, Dalian University of Technology, Dalian 116024, China.*

[*]*Corresponding author. yuankp@upc.edu.cn*

1. Input descriptors for traditional machine learning models

The descriptors for materials are the key factors to influence the prediction capability of machine learning models. Wang et al.[1] tested the performance of different types of descriptors through screening crystalline materials with ultralow thermal conductivity. They found that the descriptors based on crystal properties and composition-weighted properties have the best performance. The first part of descriptors includes fundamental properties of crystals, such as space group, number of species, number of atoms per cell, volume per atom, volume per cell, and mass density. The second part of descriptors are composition-weighted elemental properties based on the chemical composition of a given compound. It is defined as

$$\langle \mathrm{Q} \rangle = \sum_i Q_i x_i \tag{1}$$

Where Q_i represents the elemental property and x_i is the atomic fraction of the element i. The employed elemental properties are listed in Table S1.

Table S1. The list of features.

Category	Features
Crystal properties	Space group, number of species, number of atoms per cell, volume per atom, volume per cell, mass density.
Composition-weighted properties	Atomic number, valence electrons, atomic mass, group, period, electronegativity, Mendeleev number, global hardness, the orbital exponent of Slater-type orbitals, polarizability, electrophilicity indices, van der Waals radii, covalent radii, absolut radii, electron affinity, molar volume, first ionization energy, boiling point, boiling point, thermal conductivity, atomization enthalpy, fusion enthalpy, vaporization enthalpy, binding energy, atomic density.

2. Spearman's rank correlation

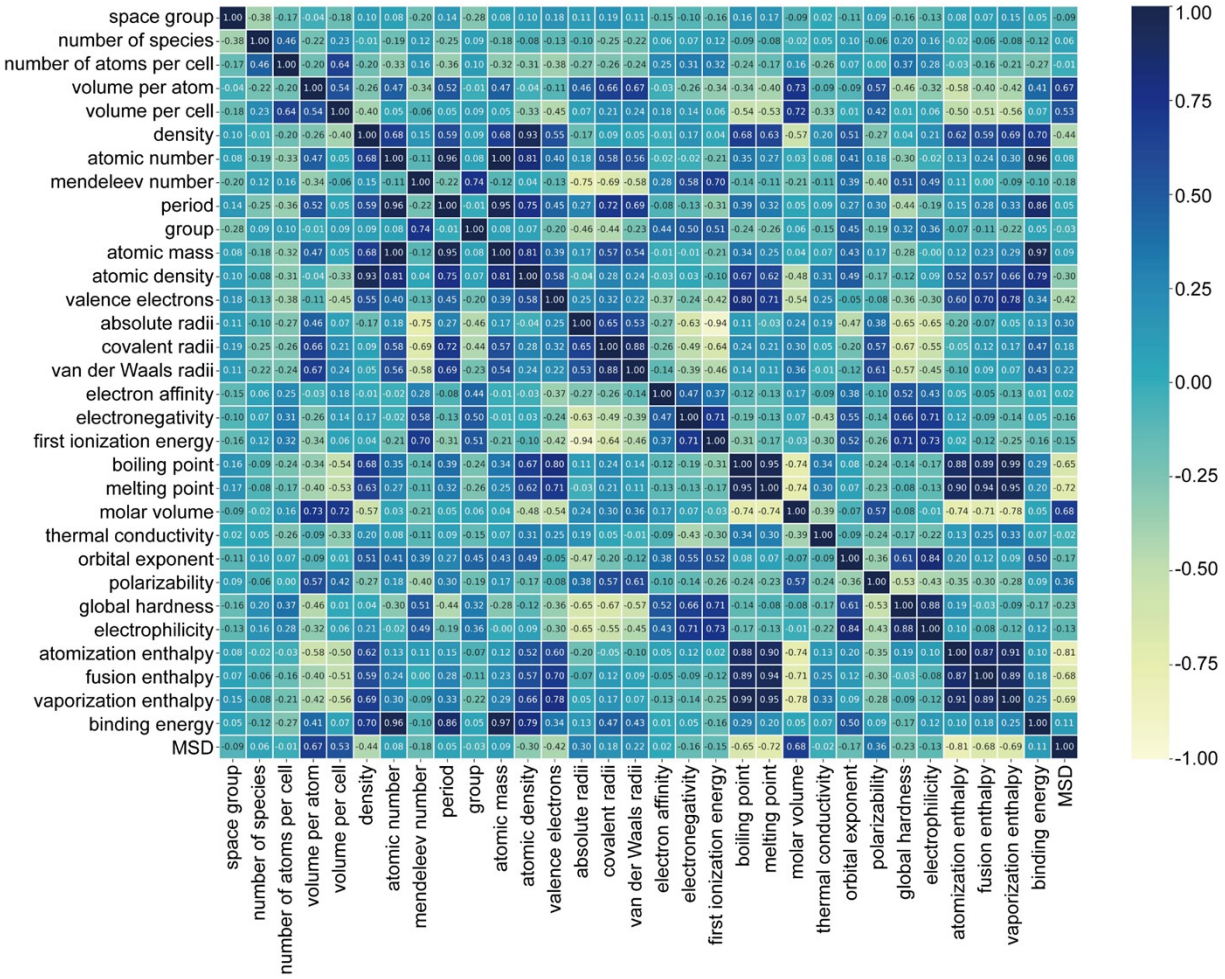

Figure S1. The Spearman's rank correlation coefficient between input descriptors and MSD.

3. Plot of MSD for space group 216, 221, 225, and 227

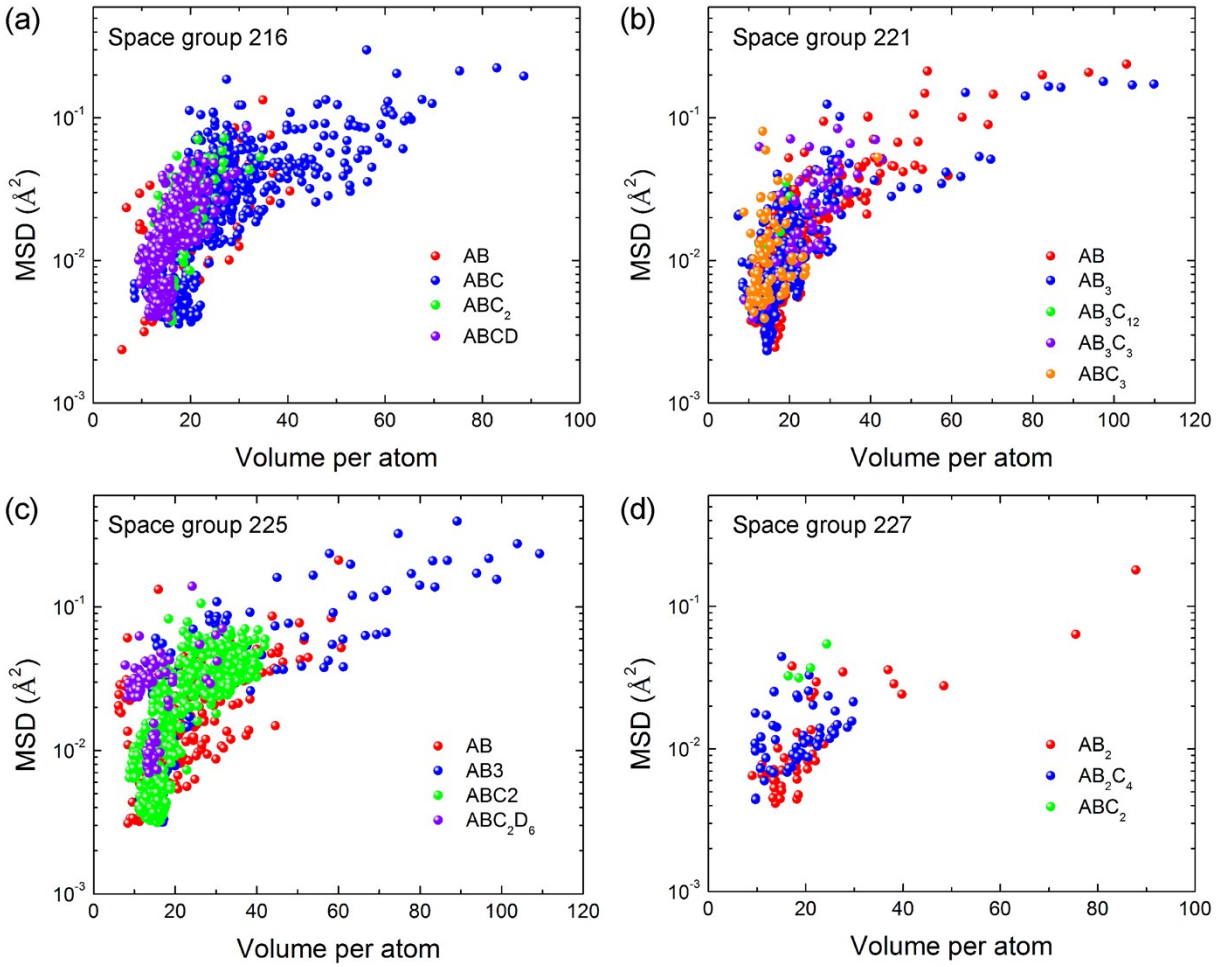

Figure S2. Variation of MSD with volume per atom for space group (a) 216, (b) 221, (c) 225, and (d) 227. The structure prototypes are marked by different colors.

4. Plot of MSD for triclinic crystal systems

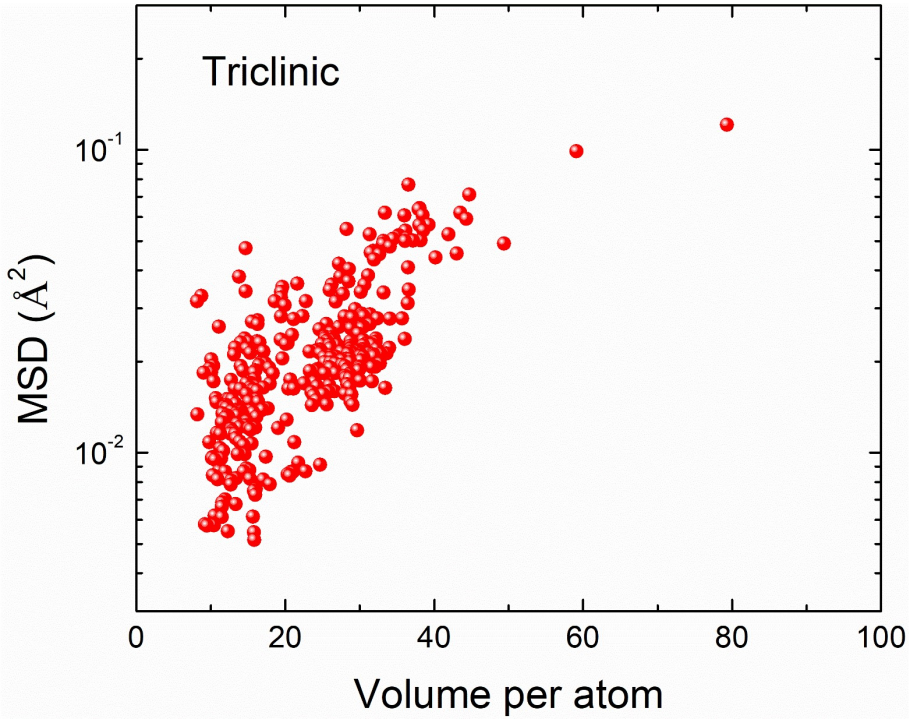

Figure S3. Variation of MSD with volume per atom for triclinic crystal systems in the prediction dataset.

5. List of top 20 materials with high MSD

Table S2. Top 20 materials with high MSD predicted by our machine learning model.

OQMD_ID	Formula	MSD_Prediction ($Å^2$)	Dynamical stability	MSD_DFT ($Å^2$)
309018	Cs3Li	0.3278	Negative	-
344386	Rb3Na	0.2634	Negative	-
348759	Rb3Li	0.2620	Negative	-
308891	Rb3Li	0.2515	Negative	-
348014	Cs3Na	0.2482	Negative	-
305111	RbNa	0.2421	Negative	-
880244	CsRbLi	0.2173	Negative	-
312986	CsNa3	0.2152	Negative	-
882056	CsKLi	0.2124	Negative	-
304929	CsLi	0.2037	Negative	-
945688	CsNaHg	0.2036	Negative	-
881789	CsNaLi	0.2019	Negative	-
346261	K3Li	0.2015	Negative	-
348856	CsNa3	0.1946	Negative	-
307975	Cs3H	0.1897	Negative	-
881355	KRbLi	0.1883	Positive	0.2541
880160	RbNaLi	0.1842	Negative	-
310561	Cs3Ca	0.1834	Positive	0.1927
307303	RbLi	0.1833	Negative	-
984823	CsLiCd	0.1824	Negative	-

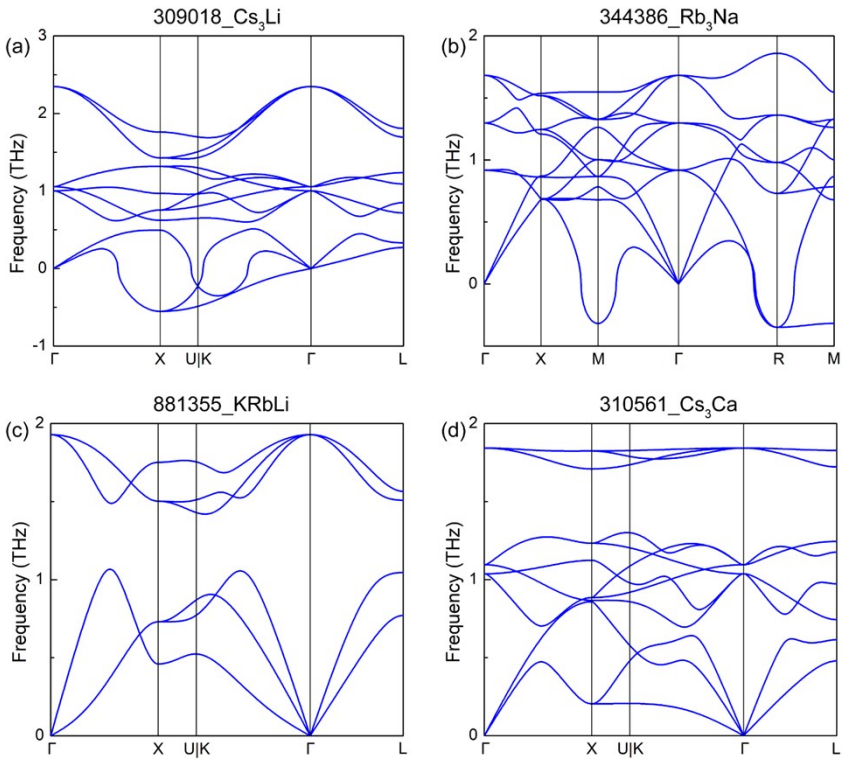

Figure S4. Representative phonon dispersions of materials in prediction dataset. (a) Cs_3Li, (b) Rb_3Na, (c) KRbLi, and (d) Cs_3Ca.

6. Validation list of MSD by first principles calculations

Table S3. MSD predicted by machine learning models and calculated by first principles for the selected 20 double perovskite materials.

OQMD_ID	Formula	MSD_Prediction (Å^2)	MSD_DFT (Å^2)
1546965	$LiIrRb_2Cl_6$	0.1099	0.1166
1549329	$HgPtCs_2Cl_6$	0.0958	0.0567
1720995	$LiAlRb_2Cl_6$	0.0943	0.0851
1433001	$LiTlCs_2I_6$	0.0938	0.0356
1433028	$LiHgCs_2I_6$	0.0890	0.0512
1546761	$TlRhCs_2Br_6$	0.0863	0.0907
1575131	$InRhCs_2Br_6$	0.0846	0.0451
1344196	$LiBiCs_2Cl_6$	0.0841	0.0723
1432958	$NaLiCs_2I_6$	0.0791	0.0629
1575778	$KIrCs_2Cl_6$	0.0769	0.0698
1721624	$InIrCs_2Cl_6$	0.0756	0.0431
1543168	$KRhCs_2Cl_6$	0.0744	0.0737
1575133	$AlAgCs_2Br_6$	0.0715	0.0406
1549349	$KRuCs_2Cl_6$	0.0707	0.0642
1575138	$AgOsCs_2Br6$	0.0697	0.0673
1543122	$NaCoRb_2Cl_6$	0.0638	0.0556
1575765	$InRhCs_2Cl_6$	0.0579	0.0373
1368340	$NaCdCs_2F_6$	0.0293	0.0275
1431038	$NaLiCs_2F_6$	0.0241	0.0253
1706276	$NaClBa_2O_6$	0.0147	0.0103

7. Phonon dispersion and avoided crossing in $LiIrRb_2Cl_6$

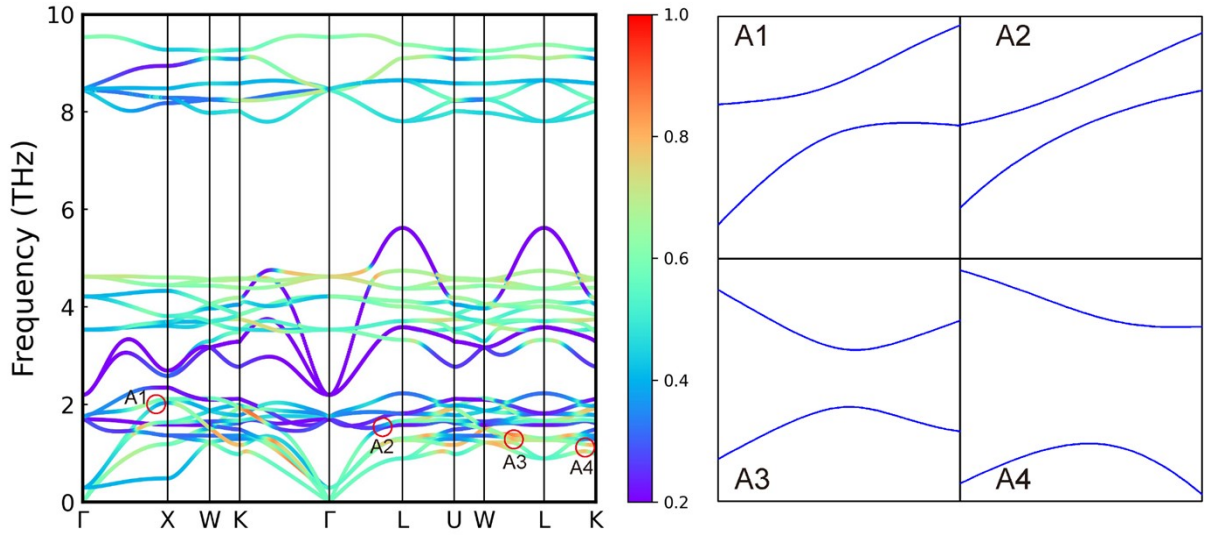

Figure S5. Phonon dispersion of $LiIrRb_2Cl_6$ colored by participation ratio and zoomed-in regions around avoided crossing points.

References

1. Wang, X.; Zeng, S.; Wang, Z.; Ni, J., Identification of Crystalline Materials with Ultra-Low Thermal Conductivity Based on Machine Learning Study. *The Journal of Physical Chemistry C* 2020, 124, 8488-8495.