Electronic Supplementary Material (ESI) for Chemical Communications.
This journal is © The Royal Society of Chemistry 2024

Supporting Information

Unveiling the Physical Mechanisms Driving Delafossite Crystals (ABX_2) Formation Through Interpretable Machine Learning

Ning Xu,[a] Zheng Li,[a] Xiaolan Fu,[a] Xiaojuan Hu,*[bc] Wenwu Xu*[a] and Zhong-Kang Han*[c]

[a] Department of Physics, School of Physical Science and Technology, Ningbo University, Ningbo, 315211, China

[b] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany

[c] School of Materials Science and Engineering, Zhejiang University, Hangzhou, 310027, China

Methodology

(1) Density functional theory calculations

Spin-polarized DFT calculations have been conducted using the Vienna Ab initio Simulation Package (VASP), a plane wave-based code. The core electrons were described with projector augmented wave potential, [1] while the electron correlation was applied using the generalized gradient approximation[2] of Perdew-Burke-Ernzerhof.[3,4] Hubbard-U corrections were considered to correct the strong on-site Coulomb interactions of localized electrons in transition metals.[5,6] For oxides containing Co, Cr, Fe, Mn, Mo, Ni, V, and W, the effective U values were chosen as 3.3, 3.5, 5.3, 3.9, 4.38, 6.4, 3.1, and 6.2 eV, respectively. All calculations were performed with a plane wave cutoff of 520 eV. The Brillouin zone was sampled by a Gamma centered k-mesh of 9 × 9 × 1, which is tested by varying the mesh size to ensure accurate results and the convergence of the k-grid. Geometry structures were fully relaxed until all the atom forces was lower than 0.03 eV/Å. Electronic structure analyses were performed with 11 × 11 × 1 Gamma centered k-mesh sampling. The delafossite crystal (ABX_2) examined in this study possesses a $R\bar{3}m$ space group, and its conventional cell is composed of 12 atoms. The formation energies (H_f) of delafossite crystals serve as an indicator of their stability. The H_f are calculated using the following formula:

$$H_f = E(ABX_2) - \mu_A - \mu_B - 2\mu_X$$

where $E(ABX_2)$ represents the total energy of the delafossite crystal, μ_A, μ_B, and μ_X are the chemical potentials of the elements A, B, and X, respectively.

Calculations of the phonon band structure were carried out using density-functional perturbation theory (DFPT)[7,8] as implemented in Phonopy[9]. Phonon calculations used a 4 × 4 × 1 supercell containing 192 atoms of the $R\bar{3}m$ conventional cell, while the k-point mesh used a Gamma point only. The calculations included symmetry points Γ (0, 0, 0), A (0, 0, 1/2), H (-1/3, 2/3, 1/2), K (-1/3, 2/3, 0) and M(0, 1/2, 0). Standard AIMD simulations were performed under isothermal-isobaric (NpT) conditions at temperatures of 300 K and 600 K. The simulations used a Langevin thermostat[10,11] in VASP. Based on the fully relaxed ground-state structure, the system was heated to 300 K and 600 K, respectively, within 1.5 ps by repeated velocity rescaling. After this, the constant temperature simulations run 20 ps with a 1 fs time step. The force convergence criteria and k-point mesh were reduced to 0.05 eV/Å and 2 × 2 × 1 to accelerate simulations with 2 × 2 × 1 supercells.

(2) Model building by SISSO code

In SISSO, the feature construction was performed by creating all the possible

mathematical combinations, within certain feature complexity, between the primary features using the operators { (+) (-) (*) (/) (exp) (exp-) (^-1) (^2) (^3) (sqrt) (cbrt) (log) (|-|) }. The SISSO approach enables us to identify the best low-dimensional descriptor in an immensity of offered candidates. The model is a linear combination of a few derived features, which are typically non-linear expressions created from mathematical operations on the primary features or between them. Due to the availability of explicit expressions, SISSO is more interpretable than some traditional machine learning methods. A list of the primary features for model training data sets can be found in the file "Supplementary file1". The code for SISSO and user guide are available at: https://github.com/rouyang2017/SISSO.

(3) Subgroup discovery analysis

The subgroup discovery was performed using RealKD package.[12] Each feature was split to 14 subsets using 14-means clustering algorithm.[13] The borders between adjacent data clusters (a1, a2, ...) are applied further for construction of inequalities (feature1 < a1), (feature2 ≥ a2), etc. While final result might depend on the number of considered clusters, in our previous study we found that relatively high numbers of considered clusters provide essentially the same result.[14] The candidate subgroups are built as conjunctions of obtained simple inequalities. The main idea of SGD is that the subgroups are unique if the distribution of the data in them is as different as possible from the data distribution in the whole sampling. The uniqueness is evaluated with a quality function.[15] In this study we used the following function:

$$Q(S) = \frac{s(S)}{s(P)}\left(\frac{med(P) - med(S)}{med(P) - min(P)}\right)\left(1 - \frac{amd(S)}{amd(P)}\right)$$

with S – subgroup, P – whole sampling, s – size, med and min – median and minimal values of a target property, amd – absolute average deviation of the data around the median of target property. With this function the algorithm is searching for subgroups with lower values of target properties. The search was done with an adapted for such purposes Monte-Carlo algorithm,[16] in which first a certain number of trial conjunctions (seeds) is generated. Afterwards, for each seed (accompanied with pruning of inequalities) the quality function is calculated. We have tested here several numbers of initial seeds: 10000, 30000, 50000, 70000 and 100000. The subgroups with the overall high quality function value were selected.

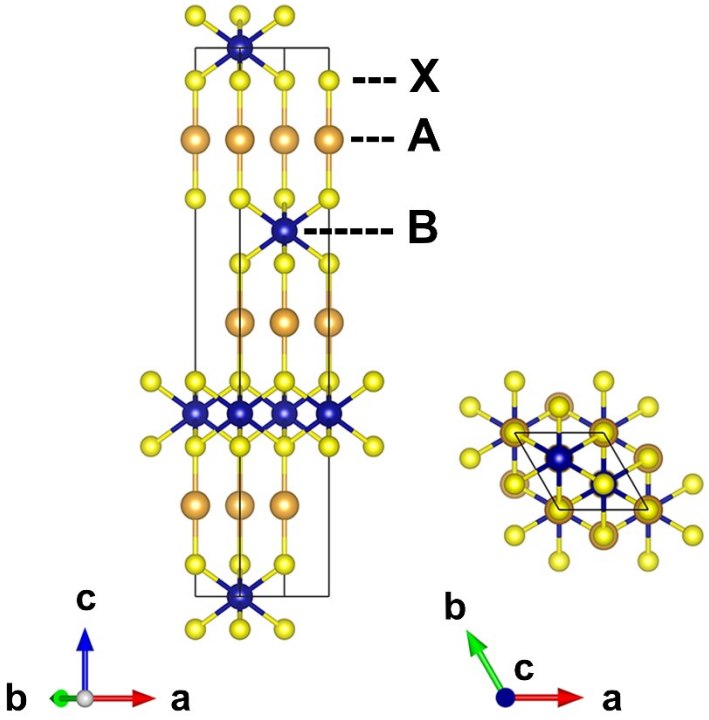

Fig. S1. Geometric structure of the delafossite crystal (ABX_2).

Table S1. Formation energies of 242 delafossite crystals, * denotes test set data.

Crystal	Formation energy (eV)	Crystal	Formation energy (eV)
$AgAlO_2$	-8.852	$AuCrSe_2$	-2.792
$AgAlS_2$	-3.874	$AuCrTe_2$	-1.137
$AgAlSe_2$	-3.206	$AuMoO_2$	-1.708
$AgAlTe_2$	-1.864	$AuMoS_2$	-3.182
$AgCrO_2$	-4.099	$AuMoSe_2$	-2.439
$AgCrS_2$	-3.065	$AuMoTe_2$	-1.280
$AgCrSe_2$	-2.585	$AuNiO_2$	-0.909
$AgCrTe_2$	-1.561	$AuNiS_2$	-2.023
$AgMoO_2$	-1.760	$AuNiSe_2$	-1.865
$AgMoS_2$	-2.925	$AuNiTe_2$	-1.565
$AgMoSe_2$	-2.268	$AuVO_2$	-5.015
$AgMoTe_2$	-1.188	$AuVS_2$	-3.947
$AgNiO_2$	-0.586	$AuVSe_2$	-3.150
$AgNiS_2$	-1.631	$AuVTe_2$	-2.097
$AgNiSe_2$	-1.544	$AuWO_2$	0.414

$AgNiTe_2$	-1.206	$AuWS_2$	-2.744
$AgVO_2$	-4.759	$AuWSe_2$	-1.907
$AgVS_2$	-3.620	$AuWTe_2$	-0.651
$AgVSe_2$	-2.757	$CuAlO_2$	-8.813
$AgVTe_2$	-1.776	$CuAlS_2$	-3.481
$AgWO_2$	0.526	$CuAlSe_2$	-2.657
$AgWS_2$	-2.461	$CuAlTe_2$	-1.182
$AgWSe_2$	-1.722	$CuCrO_2$	-4.333
$AgWTe_2$	-0.412	$CuCrS_2$	-2.816
$AuAlO_2$	-8.645	$CuCrSe_2$	-2.125
$AuAlS_2$	-4.104	$CuCrTe_2$	-0.917
$AuAlSe_2$	-3.362	$CuMoO_2$	-1.972
$AuAlTe_2$	-2.068	$CuMoS_2$	-2.576
$AuCrO_2$	-4.000	$CuMoSe_2$	-1.693
$AuCrS_2$	-3.355	$CuMoTe_2$	-0.481
$CuNiO_2$	-0.451	$PdWS_2$	-3.061
$CuNiS_2$	-1.185	$PdWSe_2$	-2.217
$CuNiSe_2$	-0.945	$PdWTe_2$	-0.959
$CuNiTe_2$	-0.481	$PtAlO_2$	-9.096
$CuVO_2$	-4.334	$PtAlS_2$	-3.488
$CuVS_2$	-3.117	$PtAlSe_2$	-2.465
$CuVSe_2$	-2.135	$PtAlTe_2$	-1.075
$CuVTe_2$	-1.019	$PtCrO_2$	-4.236
$CuWO_2$	0.743	$PtCrS_2$	-2.966
$CuWS_2$	-1.935	$PtCrSe_2$	-2.129
$CuWSe_2$	-1.011	$PtCrTe_2$	-0.800
$CuWTe_2$	0.376	$PtMoO_2$	-1.457
$PdAlO_2$	-9.196	$PtMoS_2$	-3.431
$PdAlS_2$	-3.684	$PtMoSe_2$	-2.498
$PdAlSe_2$	-2.828	$PtMoTe_2$	-1.343
$PdAlTe_2$	-1.539	$PtNiO_2$	-1.523
$PdCrO_2$	-4.292	$PtNiS_2$	-2.082
$PdCrS_2$	-2.936	$PtNiSe_2$	-1.610

$PdCrSe_2$	-2.452	$PtNiTe_2$	-1.124
$PdCrTe_2$	-1.310	$PtVO_2$	-5.077
$PdMoO_2$	-1.727	$PtVS_2$	-4.228
$PdMoS_2$	-3.492	$PtVSe_2$	-3.135
$PdMoSe_2$	-2.738	$PtVTe_2$	-1.829
$PdMoTe_2$	-1.693	$PtWO_2$	0.848
$PdNiO_2$	-1.241	$PtWS_2$	-3.011
$PdNiS_2$	-2.117	$PtWSe_2$	-1.968
$PdNiSe_2$	-1.847	$PtWTe_2$	-0.587
$PdNiTe_2$	-1.466	$AgPrO_2$	-9.476
$PdVO_2$	-4.953	$AuLuO_2$	-10.569
$PdVS_2$	-4.239	$CuCdO_2$	-2.802
$PdVSe_2$	-3.323	$CuHgO_2$	-1.186
$PdVTe_2$	-2.136	$IrHgO_2$	-1.445
$PdWO_2$	0.528	$IrLuO_2$	-8.700
$IrPrO_2$	-7.183	$IrHfTe_2$	-1.041
$IrZnO_2$	-4.262	$IrReTe_2$	1.079
$OsCdO_2$	-2.458	$IrTaTe_2$	-0.366
$OsHgO_2$	-1.911	$IrZrTe_2$	-1.290
$OsMoS_2$	3.124	$OsReTe_2$	1.878
$OsWO_2$	2.862	$OsTaTe_2$	0.537
$OsWS_2$	5.408	$PdLuO_2$	-10.351
$OsZnO_2$	-3.972	$PtGaO_2$	-5.809
$RhCeO_2$	-9.138	$RhBO_2$	-4.780
$RhHgO_2$	-1.044	$RhGaO_2$	-5.368
$RhNdO_2$	-8.100	$RuBO_2$	-4.568
$RuCeO_2$	-8.353	$RuCrTe_2$	0.481
$RuHgO_2$	-1.831	$RuGaO_2$	-5.031
$RuNdO_2$	-7.950	$RuMoTe_2$	-0.052
$RhHoSe_2$	-3.453	$RuNbTe_2$	-0.765
$IrCeO_2$	-8.153	$RuTaTe_2$	-0.220
$IrHgTe_2$	-0.142	$RuTcTe_2$	0.294
$IrNdS_2$	-2.843	$RuWTe_2$	0.728

$IrPrS_2$	-2.666	$CuGdO_2$	-0.472
$IrWO_2$	2.004	$OsGdTe_2$	9.190
$OsEuO_2$	1.798	$PdGdS_2$	3.932
$OsHgS_2$	0.646	$PtGdSe_2$	5.173
$OsHgSe_2$	0.336	$RhGdO_2$	0.506
$OsPrS_2$	-1.105	$RhGdSe_2$	5.998
$RhDySe_2$	-3.496	$RuGdO_2$	0.583
$RhHgSe_2$	0.551	$RuGdSe_2$	6.856
$RhNdS_2$	-3.772	$RuGdTe_2$	8.413
$RhWO_2$	1.563	$AgLuO_2$	-10.635
$RuHgS_2$	1.143	$IrGdS_2$	6.087
$RuPrS_2$	-2.673	$IrReS_2$	-1.527
$RuWO_2$	2.002	$IrWS_2$	-2.363
$IrGaO_2$	-5.127	$OsEuS_2$	6.849
$IrGdTe_2$	8.349	$OsEuTe_2$	8.284
$OsGdS_2$	7.448	$RuAlO_2$	-8.423
$*AgCeO_2$	-9.654	$*PdDySe_2$	-4.837
$*AgGaO_2$	-5.778	$*PdGaO_2$	-5.962
$*AgHgSe_2$	-0.582	$*PdHgSe_2$	-0.404
$*AgLaO_2$	-9.885	$*PdHoSe_2$	-4.794
$*AgLaS_2$	-6.264	$*PdYbO_2$	-8.022
$*AgNdS_2$	-5.939	$*PtDySe_2$	-4.267
$*AuCeO_2$	-9.615	$*PtHgSe_2$	-0.144
$*AuHgS_2$	-0.736	$*PtHgTe_2$	0.153
$*AuLaS_2$	-6.408	$*PtLuO_2$	-9.914
$*CuGaO_2$	-5.674	$*PtNdS_2$	-4.621
$*CuInTe_2$	-0.619	$*RhFeSe_2$	0.908
$*CuLaS_2$	-5.819	$*RhReTe_2$	0.481
$*CuLuO_2$	-10.270	$*RhTlSe_2$	0.443
$*CuYbTe_2$	-2.560	$*RuReTe_2$	1.103
$*IrAlO_2$	-8.565	$*RuYbSe_2$	-1.849
$*IrBO_2$	-4.600	$*RuZrSe_2$	-2.494
$*IrCrTe_2$	0.334	$*IrWTe_2$	0.617

*$IrInS_2$	-0.148	*$IrYSe_2$	-2.518
*$IrMoTe_2$	-0.156	*$OsGaO_2$	-4.612
*$IrNbTe_2$	-0.919	*$OsGdSe_2$	8.392
*$IrTaS_2$	-3.323	*$OsLaS_2$	-1.686
*$IrTcTe_2$	0.252	*$OsSmTe_2$	-0.073
*$IrTmTe_2$	-1.075	*$OsTcTe_2$	0.998
*$IrVTe_2$	-0.563	*$PdCeO_2$	-10.266

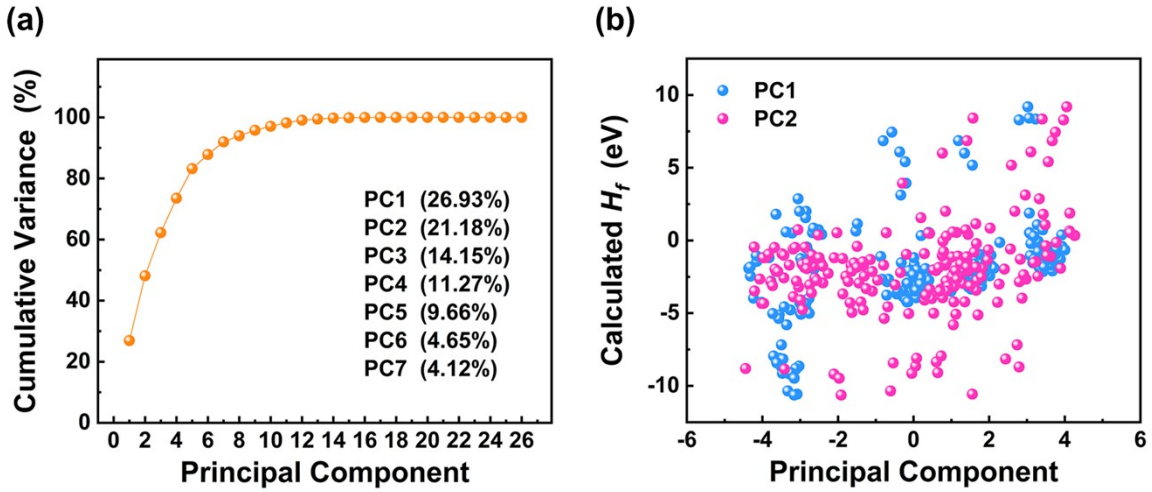

Fig. S2. (a) The cumulative explained variance varies with the inclusion of additional principal components, with a focus on the explained variance accounted for by the first six principal components. (b) The correlation between the calculated formation energies of the delafossite crystals and the first two principal components.

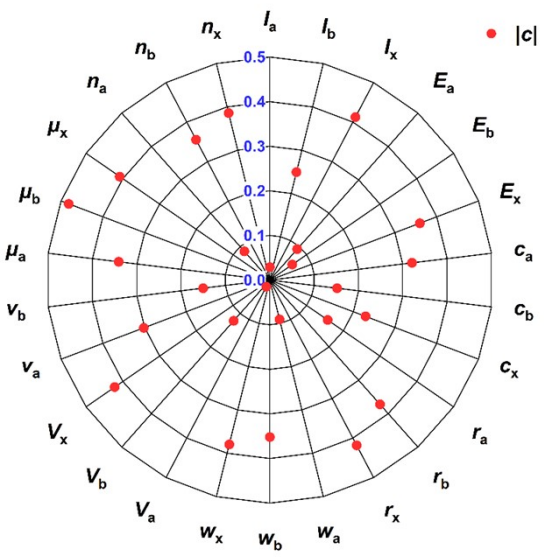

Fig. S3. Correlation between atomic features and the formation energies of delafossite crystals, |*c*| denotes the absolute value of the correlation coefficient.

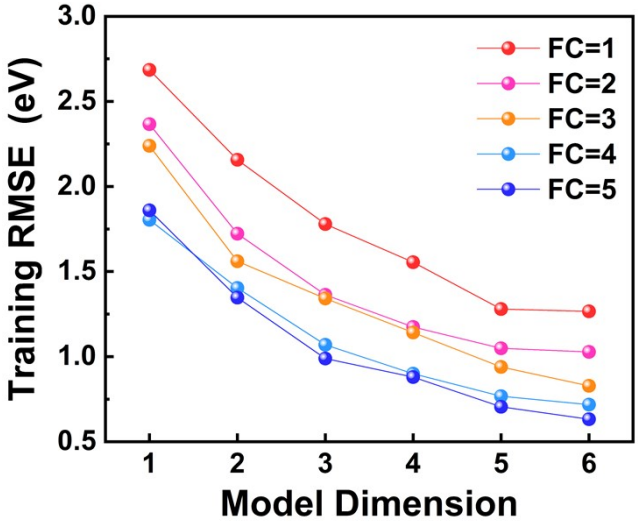

Fig. S4. Training error for best models of different dimension (D) and feature complexity (FC).

Table S2. RMSE (eV) of models for every iteration in CV10 (Feature complexity: 5; Dimension: from 1 to 5).

Iteration	1D	2D	3D	4D	5D
01	1.391	1.129	1.108	0.941	0.719
02	2.545	1.370	0.930	0.645	0.660
03	2.739	1.666	1.410	1.077	0.937

04	1.433	1.131	1.038	0.992	1.071
05	2.371	1.422	1.292	0.902	0.691
06	2.661	1.875	1.537	1.436	1.281
07	1.331	1.335	1.204	1.018	1.081
08	1.577	1.219	1.529	1.347	1.239
09	1.122	0.931	1.118	0.840	0.804
10	1.448	1.212	1.206	1.027	1.003

Table S3. Components and coefficients of the model identified by SISSO. The intercept value is -5.398 eV. Letters "a", "b", and "x" following the abbreviations denote A-site, B-site, and X-site elements in the delafossite crystals, respectively.

Symbol	**Descriptor**	**Coefficient**
d_1	$\exp(I_x/v_b)$-$\exp(-\mu_b/v_b)$	-0.077
d_2	$\mu_b(\mu_a V_b)^3/(c_b v_b)$	0.001
d_3	$\lvert(c_b-I_x)-(c_b+I_b)\log(r_x)\rvert$	0.322
d_4	$\lvert(c_x v_a)/c_b-2v_b-\mu_a\rvert$	0.177
d_5	$(V_b/r_a)^2/(E_b(E_b-E_x))$	-0.605

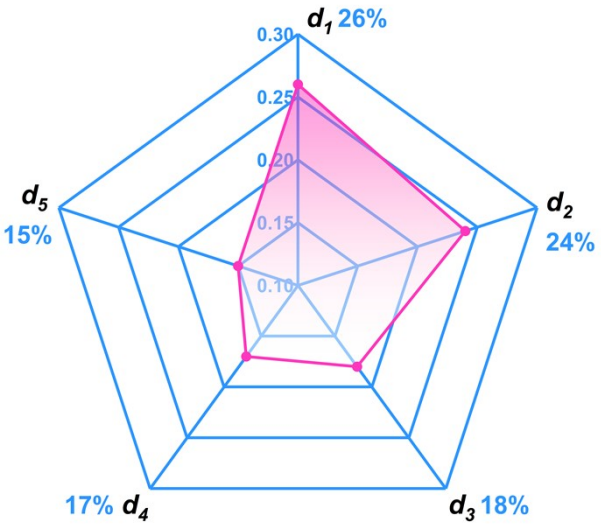

Fig. S5. The radar chart illustrates the relative importance of the five descriptors within the optimal SISSO model.

Table S4. Geometric structure information for three representative crystals.

Formula	$AgLaO_2$	$AgLaS_2$	$AgCeO_2$
Crystal system	$R\bar{3}m$	$R\bar{3}m$	$R\bar{3}m$
Space group	166	166	166
a(Å)	3.786	4.347	3.676
b(Å)	3.786	4.347	3.676
c(Å)	18.647	22.193	18.816
d_{A-B}(Å)	3.80	4.47	3.79
d_{B-X}(Å)	2.43	2.86	2.35
d_{A-X}(Å)	2.05	2.32	2.12

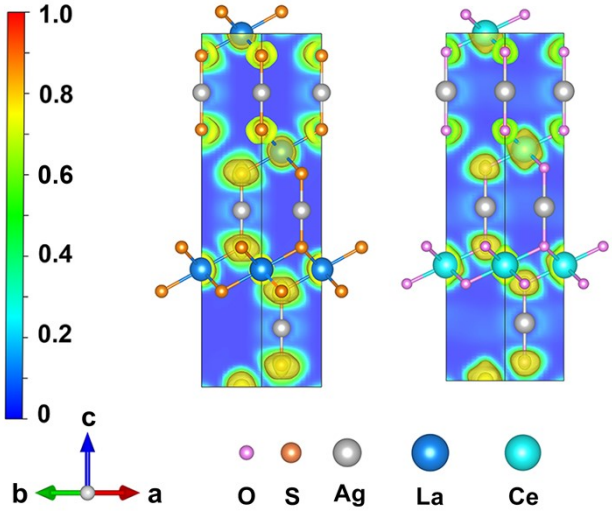

Fig. S6. Electron localization function (ELF) of $AgLaS_2$ and $AgCeO_2$.

Table S5. The integrated crystal orbital bonding index (ICOBI) for three representative crystals.

Formula	Bond	ICOBI
$AgLaO_2$	Ag-La	0.06
	La-O	0.21
	Ag-O	0.15
$AgLaS_2$	Ag-La	0.04
	La-S	0.23

	Ag-S	0.17
$AgCeO_2$	Ag-Ce	0.04
	Ce-O	0.19
	Ag-O	0.09

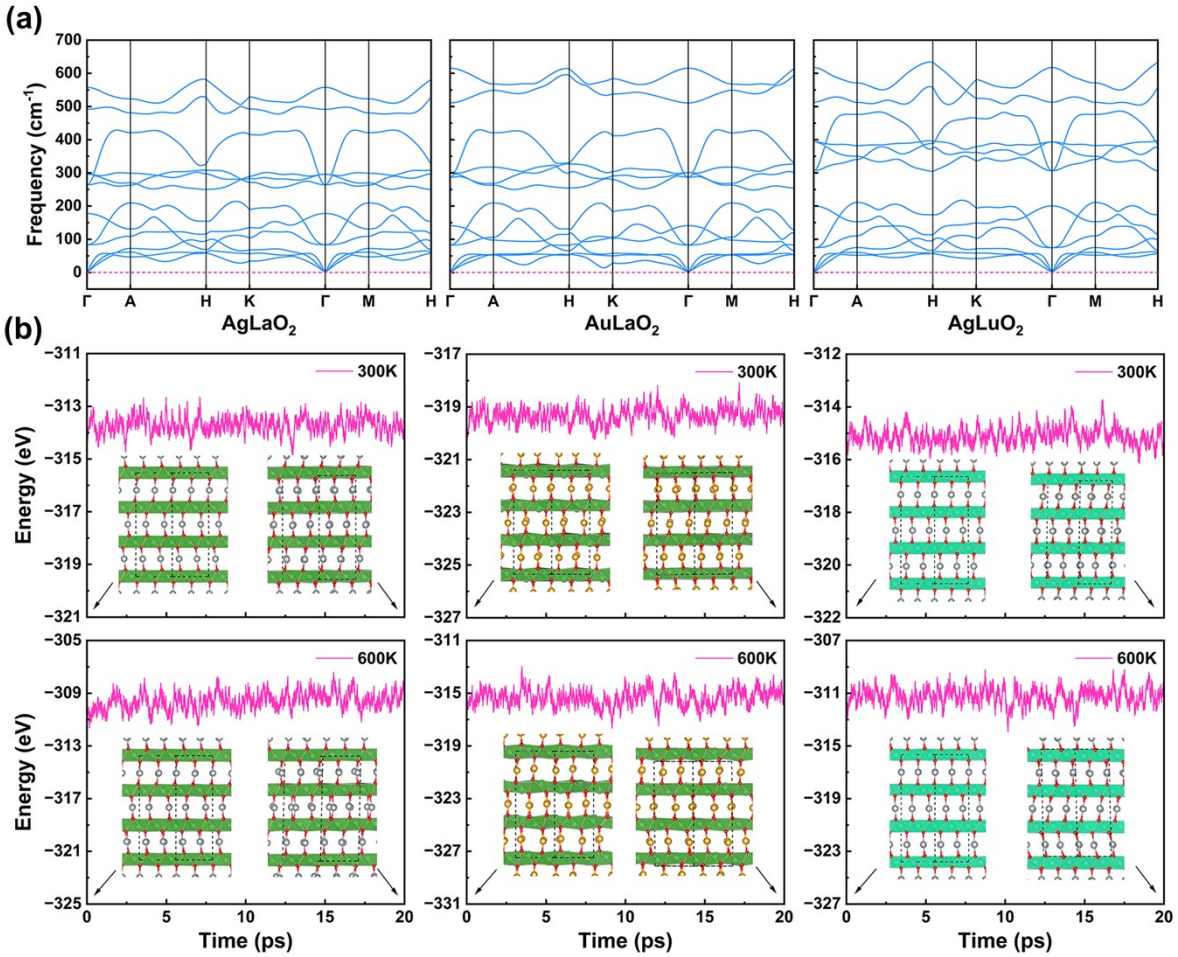

Fig. S7. Validation of dynamic stability for $AgLaO_2$, $AuLaO_2$, and $AgLuO_2$ systems. (a) The calculated phonon band structure along high-symmetry lines in the first Brillouin zone. (b) The energy fluctuations over time for $AgLaO_2$, $AuLaO_2$, and $AgLuO_2$ systems at 300 K and 600 K from AIMD simulations. Insets provide a view of the crystal configurations in the [110] direction before and after the 20 ps simulation.

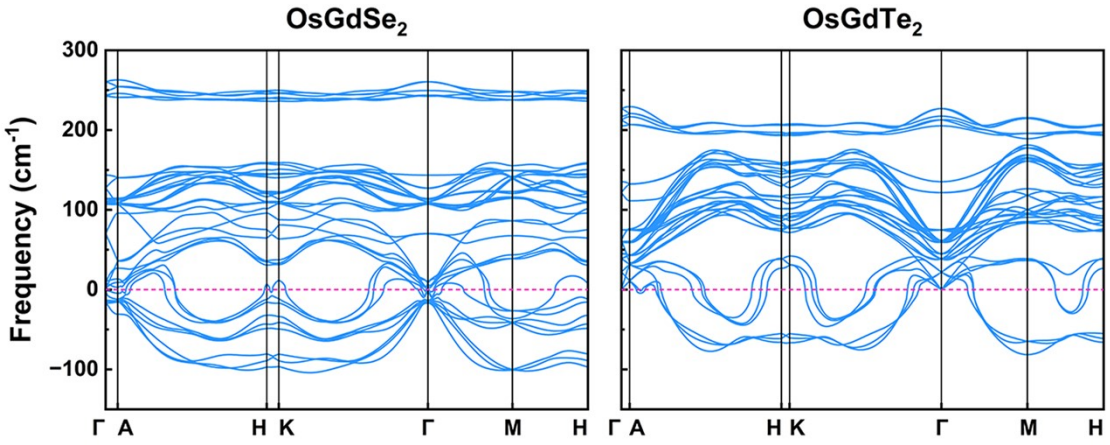

Fig. S8. The calculated phonon band structure of $OsGdSe_2$ and $OsGdTe_2$.

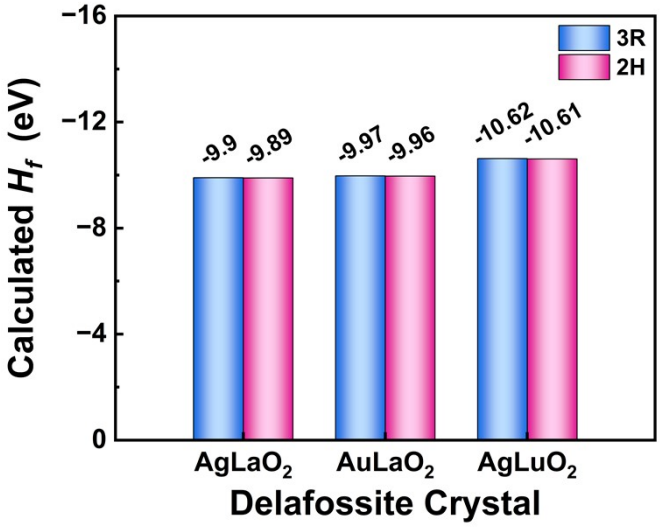

Fig. S9. Formation energy comparison for delafossite crystals of two polytypes for $AgLaO_2$, $AuLaO_2$, and $AgLuO_2$.

References:

[1] G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758.

[2] J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

[3] P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953.

[4] G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.

[5] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys and A. P. Sutton, *Phys. Rev. B*, 1998, **57**, 1505.

[6] M. Cococcioni and S. De Gironcoli, *Phys. Rev. B*, 2005, **71**, 035105.

[7] S. Baroni, S. De Gironcoli, A. Dal Corso and P. Giannozzi, *Rev. Mod. Phys.*, 2001, **73**, 515.

[8] X. Gonze and C. Lee, *Phys. Rev. B*, 1997, **55**, 10355.

[9] A. Togo, L. Chaput, T. Tadano and I. Tanaka, *J. Phys. Condens. Matter*, 2023, **35**, 353001.

[10] W. G. Hoover, A. J. C. Ladd and B. Moran, *Phys. Rev. Lett.*, 1982, **48**, 1818.

[11] B. L. Holian and D. J. Evans, *J. Chem. Phys*, 1983, **78**, 5147.

[12] C. Sutton, M. Boley, L. M. Ghiringhelli, M. Rupp, J. Vreeken and M. Scheffler, *Nat. Commun.*, 2020, **11**, 4428.

[13] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, *IEEE Trans. Pattern Anal. Machine Intell.*, 2002, **24**, 881.

[14] Z.-K. Han, D. Sarker, R. Ouyang, A. Mazheika, Y. Gao and S. V. Levchenko, *Nat. Commun.*, 2021, **12**, 1833.

[15] A. Mazheika, Y. Wang, R. Valero, L. M. Ghiringhelli, F. Vines, F. Illas, S. V. Levchenko and M. Scheffler, *Nat. Commun.*, 2022, **13**, 419.

[16] B. R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler and L. M. Ghiringhelli, *New J. Phys.*, 2017, **19**, 013031.