q

Supplementary Materials
SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates

Runhai Ouyang,[1] Stefano Curtarolo,[1,2] Emre Ahmetcik,[1] Matthias Scheffler,[1] and Luca M. Ghiringhelli[1,*]

[1] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin-Dahlem, Germany*
[2] *Materials Science, Duke University, Durham, 27708, NC, USA*

(Dated: June 23, 2018)

1. CORRELATION BETWEEN FEATURES

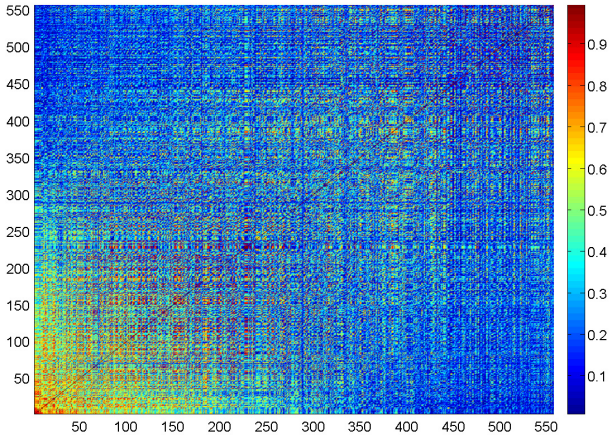

FIG. S1. The absolute Pearson correlation between features of Φ_1 with size 556. The maximal correlation between two different features is 0.99995.

2. MODELS FOR ΔE, THE ENERY DIFFERENCE BETWEEN ROCK-SALT AND ZINC BLENDE PROTOTYPES, OF OCTET BINARY MATERIALS

2.1 The 23 primary features [1]

- ionization potential: IP_A, IP_B;
- electron affinity: EA_A, EA_B;
- highest-occupied Kohn-Sham levels: H_A, H_B;
- lowest-unoccupied Kohn-Sham levels: L_A, L_B;
- radius at the maximum value of s, p, and d valence radial probability density: $r_{sA}, r_{sB}, r_{pA}, r_{pB}, r_{dA}, r_{dB}$;
- binding energy: $E_{bAA}, E_{bBB}, E_{bAB}$;
- HOMO-LUMO Kohn-Sham gap: $HL_{AA}, HL_{BB}, HL_{AB}$;
- equilibrium interatomic distance of dimer: d_{AA}, d_{BB}, d_{AB}.

2.2 Influence of subspace size on the training RMSE for a 2D model

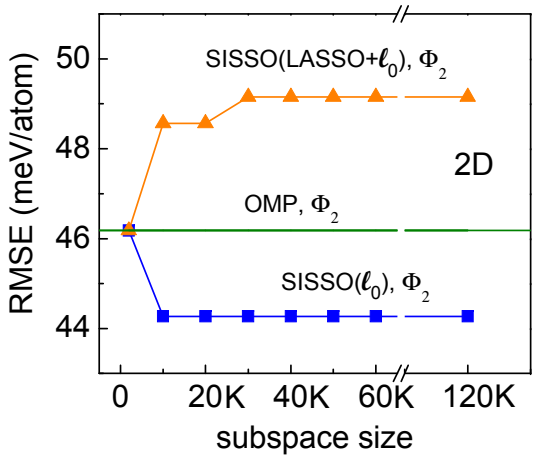

FIG. S2. Training RMSE versus subspace size in the SIS step to find a 2D descriptor by SISSO with the large features-space Φ_2. OMP results, with the same Φ_2, are shown for comparison. The 'K' in x-axis denotes 10^3

2.3 Models identified by SISSO(ℓ_0) with Φ_2

$$\text{1D:}\quad \Delta E = -0.304 + \frac{1.528}{d_{AB} r_{sA} r_{pA}}$$

$$\text{2D:}\quad \Delta E = -0.251 - 0.440\frac{IP_A}{r_{pA}(d_{AB}+d_{BB})} + 0.00465 H_A^3 |r_{sA} - r_{pB}|$$

$$\text{3D:}\quad \Delta E = -0.237 + 0.0667\frac{H_A}{d_{AA}(E_{bAB}+E_{bBB})} - 0.0302 H_A^2 |r_{sA} - r_{pB}| - 0.0970\frac{|HL_{AB} - HL_{AA}|}{H_A + L_B}$$

2.4 Models identified by SISSO(ℓ_0) with Φ_3

$$\text{1D:}\quad \Delta E = -0.285 - 3.726\frac{r_{sB}(r_{sB} - r_{dA})}{r_{dA}(r_{pA}^3 + r_{pB}^3)}$$

$$\text{2D:}\quad \Delta E = -0.590 + 5.035\frac{r_{sB}}{(r_{pA}+r_{pB})(r_{sA}r_{pA} + r_{pB}^2)} + 0.0486\,|IP_A r_{sB}(1 - E_{bAA}/E_{bBB}) - EA_B r_{sA}|$$

$$\text{3D:}\quad \Delta E = -0.500 + 4.742\frac{r_{sB}}{(r_{pA}+r_{pB})(r_{sA}r_{pA} + r_{pB}^2)} + 0.363\left|\frac{H_A}{H_B} - \frac{E_{bAA}}{E_{bBB}} - \frac{\log(r_{sA}) r_{sB}}{r_{pB}}\right| - 0.100\left|\frac{EA_A}{EA_B} + \frac{r_{pB}}{r_{sB}} - \frac{|r_{sB} - r_{dA}|}{|r_{pA} - r_{dA}|}\right|$$

2.5 Model identified by Eureqa, with totally 10^{12} functions evaluated

$$\Delta E = 0.769 + 0.648 H_A + 0.0677 E_{bAA} E_{bBB} + 0.0321 IP_A^2 + 0.00605 H_A L_A r_{dA} (0.769 + 0.0677 E_{bAA})^{L_A} - 0.00931 HL_{AB} HL_{BB}$$

3. METAL-NONMETAL CLASSIFICATION

3.1 Training materials and primary features

Atomic primary features are listed in the Table I below. Collective primary features (atoms embedded in the environment) and the 299 binary materials are listed in Table II.

TABLE I. List of the atomic primary features from experiments [2]: first ionization energy IE (eV), Pauling electronegativity χ, covalent radius $r_{\rm cov}$ (Å) [3], electron affinity EA (eV), valence υ (#valence electrons for A and (8-#valence electrons) for B). For all the A_xB_y binary materials, the "A" is the cation and "B" is the anion.

Atom	IE	χ	$r_{\rm cov}$	EA[a]	υ
H	13.59	2.20	0.31	0.755	1
Li	5.391	0.98	1.28	0.618	1
Na	5.138	0.93	1.66	0.547	1
K	4.340	0.82	2.03	0.502	1
Rb	4.176	0.82	2.20	0.486	1
Cs	3.893	0.79	2.44	0.472	1
Be	9.322	1.57	0.96	0.000	2
Mg	7.645	1.31	1.41	0.000	2
Ca	6.112	1.00	1.76	0.025	2
Sr	5.695	0.95	1.95	0.052	2
Ba	5.212	0.89	2.15	0.145	2
Ra	5.279	0.97	2.21	0.000	2
Sc	6.561	1.36	1.70	0.188	3
Y	6.218	1.22	1.90	0.307	3
Lu	5.425	1.27	1.87	0.342	3
Ti	6.827	1.54	1.60	0.079	4
Zr	6.634	1.33	1.75	0.426	4
Hf	6.824	1.30	1.75	0.104	4
V	6.746	1.63	1.53	0.524	5
Nb	6.758	1.60	1.64	0.892	5
Ta	7.887	1.50	1.70	0.321	5
Cr	6.766	1.66	1.39	0.666	6
Mo	7.092	2.16	1.54	0.745	6
W	7.980	2.36	1.62	0.815	6
Mn	7.434	1.55	1.39	0.000	7
Tc	7.275	1.90	1.47	0.549	7
Re	7.877	1.90	1.51	0.150	7
Fe	7.902	1.83	1.32	0.163	8
Ru	7.361	2.22	1.42	1.050	8
Os	8.706	2.20	1.44	1.100	8
Co	7.880	1.88	1.26	0.660	9
Rh	7.459	2.28	1.42	1.137	9
Ir	9.121	2.20	1.41	1.565	9
Ni	7.639	1.91	1.24	1.161	10
Pd	8.337	2.20	1.39	0.557	10
Pt	9.017	2.28	1.36	2.128	10
Cu	7.726	1.90	1.32	1.227	1
Ag	7.576	1.93	1.45	1.302	1
Au	9.225	2.54	1.36	2.309	1
Zn	9.394	1.65	1.22	0.000	2
Cd	8.994	1.69	1.44	0.000	2
Hg	10.43	2.00	1.32	0.000	2
B	8.297	2.04	0.84	0.277	5
Al	5.985	1.61	1.21	0.440	5
Ga	5.998	1.81	1.22	0.300	5
In	5.786	1.78	1.42	0.300	5
Tl	6.108	1.62	1.45	0.199	5
C	11.26	2.55	0.76	1.595	4
Si	8.151	1.90	1.11	1.385	4
Ge	7.897	2.01	1.20	1.233	4
Sn	7.344	1.96	1.39	1.112	4
Pb	7.416	2.33	1.46	0.364	4
N	14.53	3.04	0.71	0.073	3
P	10.48	2.19	1.07	0.746	3
As	9.814	2.18	1.19	0.808	3
Sb	8.643	2.05	1.39	1.070	3
Bi	7.286	2.02	1.48	0.945	3
O	13.61	3.44	0.66	1.461	2
S	10.36	2.58	1.05	2.073	2
Se	9.752	2.55	1.20	2.021	2
Te	9.009	2.10	1.38	1.971	2
Po	8.416	2.00	1.40	1.900	2
F	17.42	3.98	0.57	3.399	1
Cl	12.96	3.16	1.02	3.617	1
Br	11.81	2.96	1.20	3.364	1
I	10.45	2.66	1.39	3.060	1
La	5.577	1.10	2.07	0.497	3
Ce	5.538	1.12	2.04	0.518	4
Pr	5.461	1.13	2.03	0.518	5
Nd	5.525	1.14	2.01	0.518	6
Pm	5.597	1.15	1.99	0.518	7
Sm	5.643	1.17	1.98	0.518	8
Eu	5.670	1.18	1.98	0.518	9
Gd	6.150	1.20	1.96	0.518	10
Tb	5.864	1.21	1.94	0.518	11
Dy	5.938	1.22	1.92	0.518	12
Ho	6.021	1.23	1.92	0.518	13
Er	6.107	1.24	1.89	0.518	14
Tm	6.184	1.25	1.90	0.518	15
Yb	6.253	1.25	1.87	0.518	2
Ac	5.172	1.10	2.15	0.518	3
Th	6.083	1.30	2.06	0.518	4
Pa	5.887	1.50	2.00	0.518	5
U	6.193	1.38	1.96	0.518	6
Np	6.265	1.36	1.90	0.518	7
Pu	6.059	1.28	1.87	0.518	8
Am	5.991	1.30	1.80	0.518	9
Cm	6.022	1.30	1.69	0.518	10

[a] The EA of all Lanthanides (except La) and Actinides are simply set equal to that of Ce because of data unavalability [2]

TABLE II: Experimental data of the 299 binary training materials and the collective primary features extracted from the SpringerMaterials [4]. The three materials that are misclassified by the SISSO model described in the main text (the last three entries in the table) are in bold. $V_{\rm cell}/\sum V_{\rm atom}$: the ratio of the unit cell volume to the total atom volume in the unit cell of the crystal, with $V_{\rm atom} = 4\pi r_{\rm cov}^3/3$; d_{AB}: interatomic distance in crystal; CN: coordination number; x: atomic composition in crystal. Coordinates of all these materials in the classification map, $\boldsymbol{d}_x$ and $\boldsymbol{d}_y$, are shown in the rightmost two columns.

Mater.	prototype	category	$V_{\rm cell}/\sum V_{\rm atom}$	d_{AB}	CN_A	CN_B	x_A	x_B	$\boldsymbol{d}_x$	$\boldsymbol{d}_y$
Cr_3Ge	Cr_3Si	metal	1.2069	2.586	4	12	0.750	0.250	6.1050	0.4990
Cr_3Si	Cr_3Si	metal	1.2088	2.555	4	12	0.750	0.250	6.1361	0.3963
Nb_3Ga	Cr_3Si	metal	1.0967	2.891	4	12	0.750	0.250	4.1483	0.3490
V_3Ga	Cr_3Si	metal	1.0622	2.693	4	12	0.750	0.250	3.9439	0.3311
Mo_3Ga	Cr_3Si	metal	1.1212	2.757	4	12	0.750	0.250	3.1415	0.1336
Nb_3Ge	Cr_3Si	metal	1.0981	2.886	4	12	0.750	0.250	5.7629	0.5290
V_3Ge	Cr_3Si	metal	1.0485	2.675	4	12	0.750	0.250	5.4013	0.5145
Mo_3Si	Cr_3Si	metal	1.1325	2.734	4	12	0.750	0.250	4.4181	0.0243
Nb_3Sb	Cr_3Si	metal	1.0932	2.942	4	12	0.750	0.250	6.3414	0.5650
Nb_3Sn	Cr_3Si	metal	1.1082	2.956	4	12	0.750	0.250	5.3410	0.4840
V_3Sn	Cr_3Si	metal	1.0715	2.762	4	12	0.750	0.250	5.0690	0.4686
LaBi	NaCl	metal	1.4232	3.305	6	6	0.500	0.500	6.6990	0.5555
HfC	NaCl	metal	1.0283	2.320	6	6	0.500	0.500	7.1114	0.8287
ZrC	NaCl	metal	1.0673	2.349	6	6	0.500	0.500	7.2146	0.8479
LaSb	NaCl	metal	1.4119	3.245	6	6	0.500	0.500	7.9419	0.5637
YSb	NaCl	metal	1.4816	3.094	6	6	0.500	0.500	7.5142	0.6252
ThC	NaCl	metal	0.9788	2.660	6	6	0.500	0.500	6.7691	0.8287
YSe	NaCl	metal	1.2899	2.852	6	6	0.500	0.500	8.2325	0.7778
CeAs	NaCl	metal	1.3184	3.040	6	6	0.500	0.500	8.5285	0.6104
ErAs	NaCl	metal	1.3323	2.866	6	6	0.500	0.500	7.7844	0.6758
GdAs	NaCl	metal	1.3047	2.931	6	6	0.500	0.500	7.8772	0.6540
TbAs	NaCl	metal	1.3093	2.910	6	6	0.500	0.500	7.8397	0.6594
YbAs	NaCl	metal	1.3397	2.847	6	6	0.500	0.500	7.7650	0.6813
CeBi	NaCl	metal	1.4296	3.275	6	6	0.500	0.500	6.6089	0.5656
DyBi	NaCl	metal	1.4120	3.125	6	6	0.500	0.500	5.9925	0.6161
ErBi	NaCl	metal	1.4248	3.101	6	6	0.500	0.500	5.9493	0.6262
GdBi	NaCl	metal	1.3921	3.155	6	6	0.500	0.500	6.0065	0.6060
PrBi	NaCl	metal	1.3926	3.235	6	6	0.500	0.500	6.3809	0.5706
CrC	NaCl	metal	1.2502	2.015	6	6	0.500	0.500	6.7710	1.0582
FeC	NaCl	metal	2.0935	2.290	6	6	0.500	0.500	10.2850	1.1666
ScC	NaCl	metal	1.0230	2.255	6	6	0.500	0.500	6.7626	0.8670
TaC	NaCl	metal	0.9840	2.226	6	6	0.500	0.500	5.8977	0.9563
VC	NaCl	metal	1.0733	2.083	6	6	0.500	0.500	5.9199	1.0391
CeN	NaCl	metal	0.8554	2.512	6	6	0.500	0.500	9.6744	0.8512
CeP	NaCl	metal	1.2824	2.966	6	6	0.500	0.500	8.8789	0.6132
CeS	NaCl	metal	1.1872	2.884	6	6	0.500	0.500	8.8195	0.7224
CeSb	NaCl	metal	1.4198	3.215	6	6	0.500	0.500	7.8437	0.5740
CeSe	NaCl	metal	1.2504	2.991	6	6	0.500	0.500	8.6929	0.7140
CeTe	NaCl	metal	1.3823	3.181	6	6	0.500	0.500	8.0564	0.5880
DyS	NaCl	metal	1.1874	2.736	6	6	0.500	0.500	8.0980	0.7869
DySb	NaCl	metal	1.4289	3.080	6	6	0.500	0.500	7.2469	0.6252
ErP	NaCl	metal	1.3141	2.800	6	6	0.500	0.500	8.2179	0.6789
ErS	NaCl	metal	1.2082	2.715	6	6	0.500	0.500	8.1069	0.7998
ErSb	NaCl	metal	1.4355	3.050	6	6	0.500	0.500	7.1630	0.6355
ErSe	NaCl	metal	1.2736	2.828	6	6	0.500	0.500	7.9973	0.7905
GdP	NaCl	metal	1.2812	2.864	6	6	0.500	0.500	8.2792	0.6570
GdSb	NaCl	metal	1.3992	3.105	6	6	0.500	0.500	7.2146	0.6150
GdSe	NaCl	metal	1.2423	2.888	6	6	0.500	0.500	8.0608	0.7650
GdTe	NaCl	metal	1.3601	3.070	6	6	0.500	0.500	7.3985	0.6300
PdH	NaCl	metal	1.4386	2.015	6	6	0.500	0.500	6.5905	1.2100
RhH	NaCl	metal	1.3302	2.005	6	6	0.500	0.500	5.8801	1.2540
HfN	NaCl	metal	0.9576	2.255	6	6	0.500	0.500	9.3307	0.9880
HoP	NaCl	metal	1.2759	2.810	6	6	0.500	0.500	8.0439	0.6734

HoS	NaCl	metal	1.1835	2.733	6	6	0.500	0.500	8.0058	0.7934
HoSb	NaCl	metal	1.4081	3.065	6	6	0.500	0.500	7.0834	0.6304
LaSe	NaCl	metal	1.2434	3.022	6	6	0.500	0.500	8.8014	0.7013
LaTe	NaCl	metal	1.3838	3.218	6	6	0.500	0.500	8.2118	0.5775
LuS	NaCl	metal	1.2008	2.685	6	6	0.500	0.500	7.8670	0.8192
TcN	NaCl	metal	1.0646	1.990	6	6	0.500	0.500	7.0975	1.4440
YbN	NaCl	metal	0.9498	2.394	6	6	0.500	0.500	9.6249	0.9500
NdS	NaCl	metal	1.1913	2.850	6	6	0.500	0.500	8.6947	0.7353
NdSb	NaCl	metal	1.3942	3.160	6	6	0.500	0.500	7.5671	0.5843
NdSe	NaCl	metal	1.2446	2.950	6	6	0.500	0.500	8.5008	0.7268
NdTe	NaCl	metal	1.3530	3.123	6	6	0.500	0.500	7.7473	0.5985
SmO	NaCl	metal	0.9368	2.509	6	6	0.500	0.500	10.1060	1.0062
TbP	NaCl	metal	1.2895	2.845	6	6	0.500	0.500	8.2640	0.6625
YbP	NaCl	metal	1.3113	2.773	6	6	0.500	0.500	8.1348	0.6844
PbPo	NaCl	metal	2.9167	3.295	6	6	0.500	0.500	7.4495	1.1650
PrS	NaCl	metal	1.1902	2.874	6	6	0.500	0.500	8.7636	0.7289
PrSb	NaCl	metal	1.4025	3.190	6	6	0.500	0.500	7.6796	0.5791
PrTe	NaCl	metal	1.3679	3.158	6	6	0.500	0.500	7.9019	0.5933
ScS	NaCl	metal	1.3760	2.596	6	6	0.500	0.500	8.4182	0.8772
TbS	NaCl	metal	1.1841	2.758	6	6	0.500	0.500	8.1422	0.7804
TmS	NaCl	metal	1.1841	2.709	6	6	0.500	0.500	7.8817	0.8063
YS	NaCl	metal	1.2346	2.747	6	6	0.500	0.500	8.4199	0.7869
TbSb	NaCl	metal	1.4037	3.085	6	6	0.500	0.500	7.1779	0.6201
TmSb	NaCl	metal	1.4124	3.045	6	6	0.500	0.500	6.9913	0.6406
ThP	NaCl	metal	1.1878	2.916	6	6	0.500	0.500	7.0852	0.7117
ThS	NaCl	metal	1.1083	2.843	6	6	0.500	0.500	7.0934	0.8385
ThSb	NaCl	metal	1.3172	3.159	6	6	0.500	0.500	6.2693	0.6663
ThSe	NaCl	metal	1.1589	2.940	6	6	0.500	0.500	6.9412	0.8287
ThAs	NaCl	metal	1.2228	2.989	6	6	0.500	0.500	6.8149	0.7085
UC	NaCl	metal	0.9139	2.480	6	6	0.500	0.500	5.9539	0.8798
UN	NaCl	metal	0.8837	2.444	6	6	0.500	0.500	8.1114	1.0488
USe	NaCl	metal	1.2308	2.879	6	6	0.500	0.500	6.9445	0.8798
UTe	NaCl	metal	1.3761	3.082	6	6	0.500	0.500	6.5092	0.7245
NpAs	NaCl	metal	1.3884	2.918	6	6	0.500	0.500	7.3964	0.7412
NpC	NaCl	metal	1.0235	2.501	6	6	0.500	0.500	6.7659	0.8670
NpN	NaCl	metal	0.9849	2.460	6	6	0.500	0.500	9.1733	1.0336
NpP	NaCl	metal	1.3077	2.808	6	6	0.500	0.500	7.4563	0.7446
NpSb	NaCl	metal	1.5281	3.126	6	6	0.500	0.500	6.9522	0.6970
NpTe	NaCl	metal	1.5008	3.101	6	6	0.500	0.500	7.2034	0.7140
PuBi	NaCl	metal	1.5624	3.175	6	6	0.500	0.500	6.3200	0.6464
PuC	NaCl	metal	1.0525	2.487	6	6	0.500	0.500	7.3925	0.8160
PuN	NaCl	metal	1.0218	2.453	6	6	0.500	0.500	10.1120	0.9728
PuP	NaCl	metal	1.3141	2.775	6	6	0.500	0.500	7.9611	0.7008
PuS	NaCl	metal	1.3199	2.771	6	6	0.500	0.500	8.5797	0.8256
PuSb	NaCl	metal	1.5705	3.119	6	6	0.500	0.500	7.5917	0.6560
La_4As_3	Th_3P_4	metal	1.1477	3.185	6	8	0.571	0.429	8.6327	0.6100
La_4Sb_3	Th_3P_4	metal	1.2338	3.342	6	8	0.571	0.429	7.9256	0.5634
Sm_4As_3	Th_3P_4	metal	1.1404	3.058	6	8	0.571	0.429	8.0646	0.6372
Ce_4Bi_3	Th_3P_4	metal	1.2369	3.348	6	8	0.571	0.429	6.5301	0.5595
Eu_4Bi_3	Th_3P_4	metal	1.4637	3.461	6	8	0.571	0.429	7.3345	0.5794
Gd_4Bi_3	Th_3P_4	metal	1.2386	3.249	6	8	0.571	0.429	6.1031	0.5858
Nd_4Bi_3	Th_3P_4	metal	1.2316	3.305	6	8	0.571	0.429	6.3880	0.5663
Pr_4Bi_3	Th_3P_4	metal	1.2227	3.323	6	8	0.571	0.429	6.3980	0.5629
Tb_4Bi_3	Th_3P_4	metal	1.2419	3.227	6	8	0.571	0.429	6.0688	0.5890
Yb_4Bi_3	Th_3P_4	metal	1.4351	3.295	6	8	0.571	0.429	6.7885	0.6014
Ce_4Sb_3	Th_3P_4	metal	1.2221	3.292	6	8	0.571	0.429	7.7102	0.5705
Dy_4Sb_3	Th_3P_4	metal	1.2424	3.155	6	8	0.571	0.429	7.1958	0.6041
Gd_4Sb_3	Th_3P_4	metal	1.2254	3.191	6	8	0.571	0.429	7.2156	0.5976
Nd_4Sb_3	Th_3P_4	metal	1.2150	3.247	6	8	0.571	0.429	7.5309	0.5774
Pr_4Sb_3	Th_3P_4	metal	1.1788	3.240	6	8	0.571	0.429	7.3712	0.5739
Sm_4Sb_3	Th_3P_4	metal	1.2276	3.219	6	8	0.571	0.429	7.4139	0.5876
Tb_4Sb_3	Th_3P_4	metal	1.2351	3.174	6	8	0.571	0.429	7.2126	0.6008
U_3As_4	Th_3P_4	metal	1.2608	2.951	8	6	0.429	0.571	5.6794	0.2832
Th_3Bi_4	Th_3P_4	metal	1.3314	3.310	8	6	0.429	0.571	4.5497	0.2433

Gd_3Sb_4	Th_3P_4	metal	1.4071	3.194	8	6	0.429	0.571	6.2250	0.2483
La_3Se_4	Th_3P_4	metal	1.1167	2.963	8	6	0.429	0.571	6.7821	0.3444
Nd_3Se_4	Th_3P_4	metal	1.3359	3.073	8	6	0.429	0.571	7.8287	0.3505
U_3P_4	Th_3P_4	metal	1.2033	2.843	8	6	0.429	0.571	5.8014	0.2858
IrO_2	TiO_2	metal	2.2547	2.010	6	3	0.333	0.667	8.6148	0.7773
RuO_2	TiO_2	metal	2.1635	1.980	6	3	0.333	0.667	8.1919	0.7993
$PtAl_2$	CaF_2	metal	2.0335	2.559	8	4	0.333	0.667	2.2554	1.3292
AcH_2	CaF_2	metal	1.0882	2.455	8	4	0.333	0.667	6.6404	0.1358
$AuGa_2$	CaF_2	metal	2.1779	2.631	8	4	0.333	0.667	2.3041	1.6450
$AuIn_2$	CaF_2	metal	1.9942	2.817	8	4	0.333	0.667	2.0182	1.6535
$CoSi_2$	CaF_2	metal	1.9364	2.319	8	4	0.333	0.667	3.8536	0.7844
DyH_2	CaF_2	metal	1.1764	2.252	8	4	0.333	0.667	6.4725	0.1995
ErH_2	CaF_2	metal	1.1782	2.218	8	4	0.333	0.667	6.3778	0.2110
GdH_2	CaF_2	metal	1.1715	2.295	8	4	0.333	0.667	6.5529	0.1882
TiH_2	CaF_2	metal	1.2785	1.933	8	4	0.333	0.667	5.5726	0.4164
YH_2	CaF_2	metal	1.2165	2.254	8	4	0.333	0.667	6.6931	0.1995
SmH_2	CaF_2	metal	1.1816	2.325	8	4	0.333	0.667	6.7789	0.1718
ThH_2	CaF_2	metal	0.9790	2.272	8	4	0.333	0.667	5.0549	0.2473
HoH_2	CaF_2	metal	1.1522	2.237	8	4	0.333	0.667	6.2878	0.2052
LuH_2	CaF_2	metal	1.1531	2.179	8	4	0.333	0.667	6.0945	0.2289
NbH_2	CaF_2	metal	1.2692	1.976	8	4	0.333	0.667	5.3246	0.4647
NpH_2	CaF_2	metal	1.3195	2.316	8	4	0.333	0.667	6.5125	0.2860
PuH_2	CaF_2	metal	1.3920	2.321	8	4	0.333	0.667	7.2997	0.2350
TbH_2	CaF_2	metal	1.1706	2.272	8	4	0.333	0.667	6.4938	0.1938
TmH_2	CaF_2	metal	1.1376	2.204	8	4	0.333	0.667	6.1088	0.2169
$NiSi_2$	CaF_2	metal	2.0190	2.336	8	4	0.333	0.667	3.9549	0.8161
$PtSn_2$	CaF_2	metal	2.0146	2.786	8	4	0.333	0.667	3.0252	1.2407
PrTl	CsCl	metal	1.2359	3.373	8	8	0.500	0.500	4.2514	0.4577
ThBi	CsCl	metal	1.1899	3.385	8	8	0.500	0.500	4.7392	0.6565
CeTl	CsCl	metal	1.2160	3.367	8	8	0.500	0.500	4.2203	0.4536
DyIn	CsCl	metal	1.3084	3.283	8	8	0.500	0.500	4.1394	0.5429
DyTl	CsCl	metal	1.2363	3.242	8	8	0.500	0.500	3.9390	0.4941
ErIn	CsCl	metal	1.3042	3.243	8	8	0.500	0.500	4.0596	0.5518
ErTl	CsCl	metal	1.2490	3.217	8	8	0.500	0.500	3.9153	0.5022
EuTl	CsCl	metal	1.3817	3.438	8	8	0.500	0.500	4.5515	0.4779
HoTl	CsCl	metal	1.2284	3.235	8	8	0.500	0.500	3.8821	0.4981
YbIn	CsCl	metal	1.4020	3.298	8	8	0.500	0.500	4.3291	0.5563
LaTl	CsCl	metal	1.2214	3.409	8	8	0.500	0.500	4.3161	0.4455
MgTl	CsCl	metal	1.9481	3.142	8	8	0.500	0.500	5.7805	0.5305
NdTl	CsCl	metal	1.2178	3.333	8	8	0.500	0.500	4.1524	0.4617
SmTl	CsCl	metal	1.2213	3.300	8	8	0.500	0.500	4.0575	0.4738
TbTl	CsCl	metal	1.2261	3.256	8	8	0.500	0.500	3.9388	0.4900
TmTl	CsCl	metal	1.1911	3.178	8	8	0.500	0.500	3.7039	0.5062
FeAs	FeAs	metal	1.6468	2.450	6	6	0.500	0.500	6.5198	0.9973
AuGa	FeAs	metal	1.9101	2.660	6	6	0.500	0.500	3.0342	1.1494
CrP	FeAs	metal	1.5293	2.350	6	6	0.500	0.500	7.1439	0.9089
FeP	FeAs	metal	1.5861	2.300	6	6	0.500	0.500	6.7210	1.0019
MnP	FeAs	metal	1.5065	2.340	6	6	0.500	0.500	7.5369	0.8486
NiSi	FeAs	metal	1.7724	2.440	6	6	0.500	0.500	5.2130	0.9073
RuP	FeAs	metal	1.5633	2.400	6	6	0.500	0.500	5.4606	1.2155
PdSn	FeAs	metal	1.6563	2.770	6	6	0.500	0.500	3.8703	1.0780
PtSi	FeAs	metal	1.8227	2.550	6	6	0.500	0.500	4.4909	1.0830
RhSi	FeAs	metal	1.5554	2.410	6	6	0.500	0.500	3.8323	1.0830
VS	FeAs	metal	1.4168	2.500	6	6	0.500	0.500	7.2320	1.0514
TiSe	FeAs	metal	1.4371	2.670	6	6	0.500	0.500	7.2661	0.9818
MnAs	NiAs	metal	1.8656	2.578	6	6	0.500	0.500	8.7203	0.8448
NiAs	NiAs	metal	1.9009	2.440	6	6	0.500	0.500	7.2106	1.0410
AuSn	NiAs	metal	2.0501	2.852	6	6	0.500	0.500	4.1493	1.2446
MnBi	NiAs	metal	1.9575	2.908	6	6	0.500	0.500	6.5389	0.7827
PtBi	NiAs	metal	1.8312	2.842	6	6	0.500	0.500	4.1585	1.1514
CoSb	NiAs	metal	1.7778	2.621	6	6	0.500	0.500	5.8511	0.9635
CrSb	NiAs	metal	1.8719	2.777	6	6	0.500	0.500	6.9773	0.8508
VP	NiAs	metal	1.3536	2.407	6	6	0.500	0.500	6.4396	0.8924
PdSb	NiAs	metal	1.7883	2.737	6	6	0.500	0.500	5.0296	1.1275

PdTe	NiAs	metal	1.8785	2.773	6	6	0.500	0.500	5.5737	1.1550
PtSn	NiAs	metal	1.8188	2.731	6	6	0.500	0.500	4.1009	1.1172
VSe	NiAs	metal	1.5617	2.590	6	6	0.500	0.500	7.4601	1.0391
ReO_3	ReO3	metal	2.9194	1.874	6	2	0.250	0.750	9.6966	1.3965
ThH_2	ThH2	metal	1.1474	2.400	8	4	0.333	0.667	5.9244	0.2473
TiH_2	ThH2	metal	1.2380	1.930	8	4	0.333	0.667	5.3960	0.4164
ZrH_2	ThH2	metal	1.2093	2.080	8	4	0.333	0.667	6.1032	0.2663
AlAs	ZnS	nonmetal	3.1323	2.451	4	4	0.500	0.500	14.0960	0.8774
AlP	ZnS	nonmetal	3.2544	2.367	4	4	0.500	0.500	15.6750	0.8815
AlSb	ZnS	nonmetal	3.0394	2.641	4	4	0.500	0.500	11.6810	0.8251
GaAs	ZnS	nonmetal	3.0747	2.447	4	4	0.500	0.500	12.3070	0.9865
InAs	ZnS	nonmetal	2.9129	2.622	4	4	0.500	0.500	11.8560	0.9701
BP	ZnS	nonmetal	3.0684	1.965	4	4	0.500	0.500	11.6640	1.1169
BeS	ZnS	nonmetal	3.3607	2.106	4	4	0.500	0.500	17.8100	1.0127
BeSe	ZnS	nonmetal	3.1165	2.229	4	4	0.500	0.500	15.4560	1.0009
BeTe	ZnS	nonmetal	3.0271	2.437	4	4	0.500	0.500	12.5860	0.8243
CuBr	ZnS	nonmetal	2.7167	2.460	4	4	0.500	0.500	14.5260	1.4060
CdPo	ZnS	nonmetal	3.0839	2.886	4	4	0.500	0.500	10.8590	0.8450
CdTe	ZnS	nonmetal	2.8967	2.807	4	4	0.500	0.500	11.1890	0.8872
CuF	ZnS	nonmetal	1.8501	1.843	4	4	0.500	0.500	16.9200	1.8905
GaP	ZnS	nonmetal	3.1772	2.360	4	4	0.500	0.500	13.6120	0.9910
GaSb	ZnS	nonmetal	3.0021	2.639	4	4	0.500	0.500	10.2630	0.9276
HgSe	ZnS	nonmetal	3.3385	2.635	4	4	0.500	0.500	12.9970	1.2750
HgTe	ZnS	nonmetal	3.2649	2.797	4	4	0.500	0.500	10.6560	1.0500
MgTe	ZnS	nonmetal	3.0415	2.822	4	4	0.500	0.500	15.1560	0.6877
ZnPo	ZnS	nonmetal	3.2869	2.732	4	4	0.500	0.500	11.8550	0.8250
ZnS	ZnS	nonmetal	3.1606	2.338	4	4	0.500	0.500	15.9380	1.0642
Al_2O_3	Al_2O_3	nonmetal	2.3038	1.910	6	4	0.400	0.600	14.4480	0.3677
Lu_2S_3	Al_2O_3	nonmetal	1.7066	2.690	6	4	0.400	0.600	8.9445	0.2017
Rh_2O_3	Al_2O_3	nonmetal	1.9034	2.030	6	4	0.400	0.600	8.4293	0.2152
Yb_2S_3	Al_2O_3	nonmetal	1.6881	2.710	6	4	0.400	0.600	8.9891	0.2035
Ce_2O_3	La_2O_3	nonmetal	1.0627	2.507	7	4.667	0.400	0.600	9.5805	0.3656
La_2O_3	La_2O_3	nonmetal	1.0575	2.536	7	4.667	0.400	0.600	9.7070	0.3634
Nd_2O_3	La_2O_3	nonmetal	1.0636	2.469	7	4.667	0.400	0.600	9.4204	0.3675
AgBr	NaCl	nonmetal	2.5376	2.939	6	6	0.500	0.500	13.3580	1.4282
AgCl	NaCl	nonmetal	2.4826	2.775	6	6	0.500	0.500	14.8170	1.5247
AgF	NaCl	nonmetal	2.1980	2.460	6	6	0.500	0.500	19.7890	1.9204
BaO	NaCl	nonmetal	0.9924	2.770	6	6	0.500	0.500	14.0740	0.7654
BaS	NaCl	nonmetal	1.4021	3.194	6	6	0.500	0.500	13.1080	0.5740
BaSe	NaCl	nonmetal	1.4708	3.300	6	6	0.500	0.500	12.8680	0.5674
BaTe	NaCl	nonmetal	1.6332	3.503	6	6	0.500	0.500	11.9790	0.4673
KBr	NaCl	nonmetal	1.7015	3.301	6	6	0.500	0.500	21.0810	0.6068
LiBr	NaCl	nonmetal	2.5818	2.745	6	6	0.500	0.500	26.7650	0.7252
NaBr	NaCl	nonmetal	1.9608	2.958	6	6	0.500	0.500	21.4200	0.6882
RbBr	NaCl	nonmetal	1.5406	3.418	6	6	0.500	0.500	19.0870	0.6068
CaS	NaCl	nonmetal	1.5858	2.800	6	6	0.500	0.500	13.1940	0.6450
CaSe	NaCl	nonmetal	1.7177	2.956	6	6	0.500	0.500	13.3750	0.6375
CaTe	NaCl	nonmetal	1.6766	3.050	6	6	0.500	0.500	10.9440	0.5250
CdO	NaCl	nonmetal	1.8881	2.348	6	6	0.500	0.500	14.1010	1.4534
CsF	NaCl	nonmetal	0.8771	3.001	6	6	0.500	0.500	19.2920	0.7861
CsH	NaCl	nonmetal	1.0628	3.188	6	6	0.500	0.500	13.5590	0.4345
LiF	NaCl	nonmetal	1.6988	2.010	6	6	0.500	0.500	30.1210	0.9751
NaF	NaCl	nonmetal	1.3524	2.380	6	6	0.500	0.500	25.2690	0.9254
RbF	NaCl	nonmetal	1.0365	2.865	6	6	0.500	0.500	21.9640	0.8159
KH	NaCl	nonmetal	1.3166	2.850	6	6	0.500	0.500	16.1820	0.4510
LiH	NaCl	nonmetal	1.8835	2.032	6	6	0.500	0.500	19.3700	0.5390
NaH	NaCl	nonmetal	1.5270	2.451	6	6	0.500	0.500	16.5480	0.5115
KI	NaCl	nonmetal	1.9005	3.530	6	6	0.500	0.500	19.7510	0.5453
LiI	NaCl	nonmetal	2.7771	3.030	6	6	0.500	0.500	24.1490	0.6517
NaI	NaCl	nonmetal	2.2224	3.233	6	6	0.500	0.500	20.3640	0.6185
RbI	NaCl	nonmetal	1.7686	3.669	6	6	0.500	0.500	18.3800	0.5453
MgO	NaCl	nonmetal	1.4368	2.103	6	6	0.500	0.500	13.8430	1.1266
MgSe	NaCl	nonmetal	2.1463	2.731	6	6	0.500	0.500	12.7570	0.8351
SrO	NaCl	nonmetal	1.0745	2.588	6	6	0.500	0.500	14.2750	0.8170

SrSe	NaCl	nonmetal	1.5769	3.114	6	6	0.500	0.500	12.9250	0.6056
SrTe	NaCl	nonmetal	1.7555	3.330	6	6	0.500	0.500	12.0620	0.4988
KCl	NaCl	nonmetal	1.5711	3.142	6	6	0.500	0.500	22.0700	0.6478
CaO	NaCl	nonmetal	1.1573	2.405	6	6	0.500	0.500	14.6070	0.8600
NaCl	NaCl	nonmetal	1.8799	2.810	6	6	0.500	0.500	23.2850	0.7347
KF	NaCl	nonmetal	1.0869	2.690	6	6	0.500	0.500	23.0320	0.8159
TlBr	NaCl	nonmetal	3.5597	3.290	6	6	0.500	0.500	22.3240	1.1988
CoO	NaCl	nonmetal	2.0196	2.131	6	6	0.500	0.500	13.5590	1.6168
FeO	NaCl	nonmetal	1.8365	2.151	6	6	0.500	0.500	12.6660	1.5738
GdN	NaCl	nonmetal	0.9402	2.495	6	6	0.500	0.500	9.9246	0.9120
EuS	NaCl	nonmetal	1.4308	2.990	6	6	0.500	0.500	10.0890	0.7611
EuSe	NaCl	nonmetal	1.4959	3.098	6	6	0.500	0.500	9.8708	0.7522
MnO	NaCl	nonmetal	1.7571	2.220	6	6	0.500	0.500	14.3080	1.3330
MnSe	NaCl	nonmetal	2.1770	2.720	6	6	0.500	0.500	10.9360	0.9881
SiC	SiC	nonmetal	2.7320	1.890	4	6	0.500	0.500	12.9270	1.2112
Eu_4As_3	Th_3P_4	nonmetal	1.2931	3.189	6	8	0.571	0.429	9.0669	0.6410
Eu_3S_4	Th_3P_4	nonmetal	1.3278	2.953	8	6	0.429	0.571	8.0330	0.3626
CoF_2	TiO_2	nonmetal	3.5631	2.060	6	3	0.333	0.667	21.9330	0.3508
FeF_2	TiO_2	nonmetal	3.4804	2.130	6	3	0.333	0.667	22.0100	0.3109
MnF_2	TiO_2	nonmetal	3.0701	2.130	6	3	0.333	0.667	22.9220	0.1184
NiF_2	TiO_2	nonmetal	3.4912	2.000	6	3	0.333	0.667	21.1530	0.3755
VF_2	TiO_2	nonmetal	2.2564	2.090	6	3	0.333	0.667	16.0200	0.1680
MgF_2	TiO_2	nonmetal	2.4486	1.990	6	3	0.333	0.667	21.6310	0.0050
GeO_2	TiO_2	nonmetal	2.8633	1.870	6	3	0.333	0.667	11.9740	0.5827
SnO_2	TiO_2	nonmetal	2.6191	2.060	6	3	0.333	0.667	11.2330	0.5354
TiO_2	TiO_2	nonmetal	1.5936	1.950	6	3	0.333	0.667	8.6984	0.2047
AgI	ZnO	nonmetal	2.8516	2.790	4	4	0.500	0.500	12.5910	1.2834
AlN	ZnO	nonmetal	2.3408	1.890	4	4	0.500	0.500	18.4170	1.2236
BeO	ZnO	nonmetal	2.8158	1.640	4	4	0.500	0.500	22.6370	1.3502
CdSe	ZnO	nonmetal	2.8450	2.610	4	4	0.500	0.500	13.1080	1.0774
GaN	ZnO	nonmetal	2.4843	1.940	4	4	0.500	0.500	17.3860	1.3756
InN	ZnO	nonmetal	2.2947	2.150	4	4	0.500	0.500	16.3300	1.3528
ZnO	ZnO	nonmetal	2.7009	1.980	4	4	0.500	0.500	20.6600	1.4190
CdF_2	CaF_2	nonmetal	2.7845	2.333	8	4	0.333	0.667	19.0680	0.2081
CeO_2	CaF_2	nonmetal	1.0431	2.343	8	4	0.333	0.667	7.8287	0.0083
$SrCl_2$	CaF_2	nonmetal	2.1272	3.022	8	4	0.333	0.667	17.1780	0.0315
YbF_2	CaF_2	nonmetal	1.5161	2.424	8	4	0.333	0.667	14.0360	0.0298
ThO_2	CaF_2	nonmetal	1.1226	2.423	8	4	0.333	0.667	7.2588	0.0686
NpO_2	CaF_2	nonmetal	1.2882	2.353	8	4	0.333	0.667	7.9621	0.0990
PrO_2	CaF_2	nonmetal	1.0471	2.335	8	4	0.333	0.667	7.7892	0.0046
PuO_2	CaF_2	nonmetal	1.3188	2.337	8	4	0.333	0.667	8.6607	0.0590
TbO_2	CaF_2	nonmetal	1.1091	2.282	8	4	0.333	0.667	7.7049	0.0274
UO_2	CaF_2	nonmetal	1.2053	2.369	8	4	0.333	0.667	7.3417	0.1097
AmO_2	CaF_2	nonmetal	1.4571	2.333	8	4	0.333	0.667	9.4217	0.0686
CmO_2	CaF_2	nonmetal	1.7014	2.321	8	4	0.333	0.667	11.0010	0.0686
EuF_2	CaF_2	nonmetal	1.4587	2.527	8	4	0.333	0.667	14.3060	0.0557
PbF_2	CaF_2	nonmetal	3.5918	2.572	8	4	0.333	0.667	17.8400	0.7849
CsBr	CsCl	nonmetal	1.1580	3.714	8	8	0.500	0.500	14.8920	0.5846
CsCl	CsCl	nonmetal	1.0711	3.568	8	8	0.500	0.500	15.6180	0.6241
TlCl	CsCl	nonmetal	3.2943	3.327	8	8	0.500	0.500	23.4240	1.2798
Cu_3N	ReO_3	nonmetal	1.8149	1.904	2	6	0.750	0.250	18.1490	1.4440
YP	NaCl	nonmetal	1.4935	2.831	6	6	0.500	0.500	8.51786	0.6680
Th_3As_4	Th_3P_4	nonmetal	1.3775	3.060	8	6	0.429	0.571	6.00507	0.2793
La_3Te_4 [1]	Th_3P_4	nonmetal	1.5939	3.330	8	6	0.429	0.571	7.31331	0.2516

[1] A band gap of $\sim$ 0.9eV was reported for slightly off-stoichiometric $La_{3-x}Te_4$ in Ref. [5].

3.2 Isothermal bulk modulus for the prediction of insulator to metal transition

Links of the B_T from AFLOW.org:
- GaAs http://aflow.org/material.php?id=aflow:a9c6b29d70d7cbec
- CdTe http://aflow.org/material.php?id=aflow:db3c099649cd0199
- HgTe http://aflow.org/material.php?id=aflow:c2384ec3a6ca5619
- AgBr http://aflow.org/material.php?id=aflow:acc935168652beb5
- GaSb http://aflow.org/material.php?id=aflow:c35317ab558919f6
- CaTe http://aflow.org/material.php?id=aflow:f51506e74f2ec57e
- EuSe http://aflow.org/material.php?id=aflow:a5a07064deeee46a
- AgI http://aflow.org/material.php?id=aflow:d611e813a85efcb0
- AlSb http://aflow.org/material.php?id=aflow:8efc0f655f5d4309

3.3 Band gaps

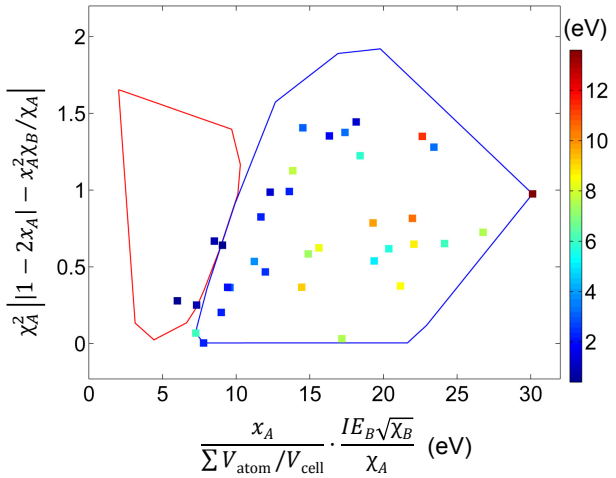

FIG. S3. Available experimental band gaps (at $\sim$ 300 K) for the training nonmetals. Corresponding materials of these points (coordinates) can be found from Table II.

* ghiringhelli@fhi-berlin.mpg.de

[1] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Phys. Rev. Lett. **114**, 105503 (2015).
[2] www.webelements.com/.
[3] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, Dalton Trans. , 2832 (2008).
[4] materials.springer.com/.
[5] A. F. May, J.-P. Fleurial, and G. J. Snyder, Phys. Rev. B **78**, 125205 (2008).