Supplemental material to "Universal Machine Learning Framework for Defect Predictions in Zinc Blende Semiconductors"

Arun Mannodi-Kanakkithodi,[1, 2, a)] Xiaofeng Xiang,[3] Laura Jacoby,[4] Robert Biegaj,[5], Scott T. Dunham,[6] Daniel R. Gamelin,[4] and Maria K. Y Chan [1, b)]

[1]*Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[2]*School of Materials Engineering, Purdue University, West Lafayette, Indiana 47907, USA*
[3]*Molecular Engineering & Sciences Institute, University of Washington, Seattle, Washington 98195, USA*
[4]*Department of Chemistry, University of Washington, Seattle, Washington 98195, USA*
[5]*Materials Science & Engineering, University of Washington, Seattle, Washington 98195, USA*
[6]*Department of Electrical and Computer Engineering, University of Washington, Seattle, Washington 98195, USA*

[a]amannodi@purdue.edu [b]mchan@anl.gov

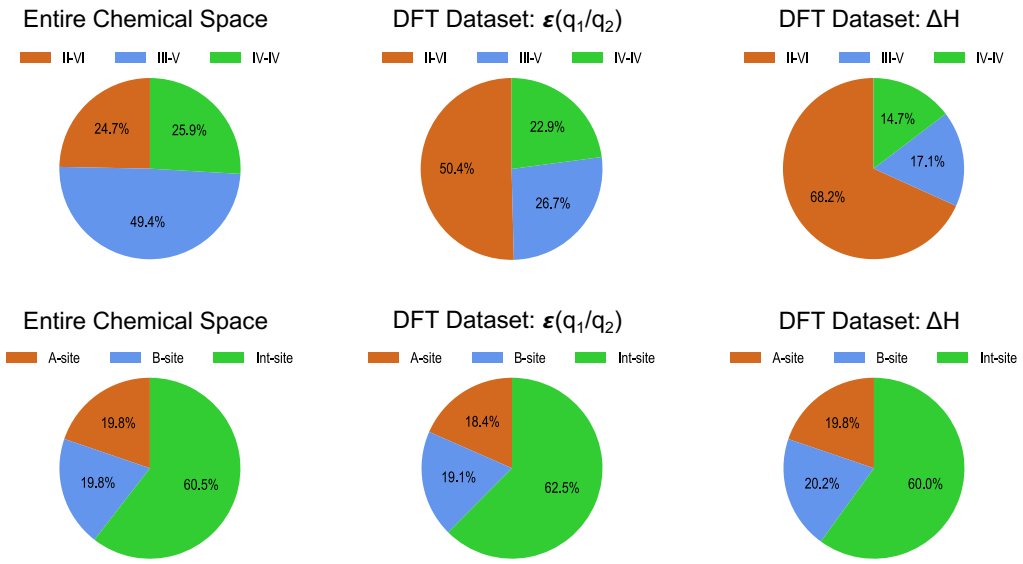

FIG. S1: Pie charts visualizing the distribution of different semiconductor and defect site types in the dataset.

TABLE SI: Reported experimental values and calculated results from PBE, HSE and HSE+SOC for the lattice constants and band gaps of all semiconductor compounds used in this work.

Type	**Compound**	**Expt. Lattice Constant (Å)**	**PBE Lattice Constant (Å)**	**HSE Lattice Constant (Å)**	**Expt. Band Gap (eV)**	**PBE Band Gap (eV)**	**HSE Band Gap (eV)**	**HSE+SOC Band Gap (eV)**
II-VI	CdO	5.10	5.15	5.10	2.30	0.47	0.99	0.97
II-VI	CdS	5.83	5.94	5.90	2.42	1.03	2.10	2.08
II-VI	CdSe	6.08	6.21	6.15	1.70	0.48	1.48	1.34
II-VI	CdTe	6.48	6.62	6.57	1.50	0.57	1.43	1.14
II-VI	ZnO	4.60	4.63	4.58	3.27	0.62	2.35	2.33
II-VI	ZnS	5.42	5.45	5.43	3.54	2.01	3.28	3.25
II-VI	ZnSe	5.67	5.74	5.70	2.70	1.14	2.33	2.19
II-VI	ZnTe	6.10	6.18	6.14	2.26	1.06	2.05	1.75
III-V	BN	3.62	3.63	3.60	6.10	4.44	5.81	-
III-V	BP	4.54	4.55	4.52	2.10	1.24	1.96	-
III-V	BAs	4.78	4.82	4.76	1.82	1.20	1.84	-
III-V	BSb	5.10	5.28	5.20	0.55	0.75	1.24	-
III-V	AlN	4.38	4.40	4.36	5.30	3.30	4.58	-
III-V	AlP	5.45	5.51	5.47	2.45	1.62	2.30	-
III-V	AlAs	5.66	5.73	5.69	2.15	1.49	2.13	-
III-V	AlSb	6.14	6.23	6.14	1.62	1.22	1.84	-
III-V	GaN	4.52	4.59	4.52	3.20	1.45	2.98	-
III-V	GaP	5.45	5.53	5.45	2.27	1.51	2.30	-
III-V	GaAs	5.65	5.77	5.67	1.42	0.13	1.34	-
III-V	GaSb	6.10	6.22	6.10	0.75	0.00	0.91	-
III-V	InN	4.98	5.10	5.03	0.70	0.00	0.46	-
III-V	InP	5.87	6.00	5.93	1.34	0.37	1.33	-
III-V	InAs	6.06	6.22	6.11	0.36	0.00	0.38	-
III-V	InSb	6.48	6.65	6.52	0.17	0.00	0.42	-
IV-IV	C	3.57	3.57	3.54	5.50	4.12	-	-
IV-IV	Si	5.43	5.47	5.45	1.17	0.61	-	-
IV-IV	Ge	5.65	5.79	5.68	0.66	0.00	0.70	-
IV-IV	Sn	6.49	6.65	6.54	0.12	0.00	0.00	-
IV-IV	SiC	4.36	4.38	4.35	2.42	1.36	2.25	-
IV-IV	GeC	4.60	4.63	4.57	1.42	1.63	2.39	-
IV-IV	SnC	5.02	5.10	5.03	1.22	0.59	1.65	-
IV-IV	SiGe	5.54	5.62	5.55	0.92	0.61	1.12	-
IV-IV	SiSn	-	6.09	6.01	-	0.34	1.09	-
IV-IV	GeSn	-	6.23	6.12	-	0.00	0.13	-
-								

TABLE SII: Experimentally measured defect levels (collected from references[42–57]) and their corresponding DFT-PBE computed values, as plotted in Fig. 2. All levels are relative to the semiconductor valence band maximum.

Compound	**Semiconductor Type**	**Defect**	**Defect Transition Label**	**Measured Value**	**DFT-PBE Value**
CdTe	II-VI	V_{Cd}	(-1/-2)	0.43	0.32
CdTe	II-VI	As_{Te}	(0/-1)	0.09	0.07
CdTe	II-VI	Cu_{Cd}	(0/-1)	0.35	0.15
CdTe	II-VI	Cu_i	(+1/0)	1.05	1.04
CdTe	II-VI	Te_{Cd}	(+2/+1)	0.32	0.35
CdTe	II-VI	Fe_{Cd}	(+1/0)	0.20	0.23
CdTe	II-VI	Cd_i	(+2/+1)	0.64	0.78
CdTe	II-VI	V^*_{Cd}	(+1/0)	0.65	0.65
CdTe	II-VI	Ge_i	(+2/+1)	0.65	0.91
CdTe	II-VI	Ni_{Cd}	(0/-1)	0.92	0.56
CdTe	II-VI	Co_{Cd}	(0/-1)	1.25	0.94
CdTe	II-VI	Cr_{Cd}	(0/-1)	1.34	0.83
CdTe	II-VI	Sn_i	(+2/+1)	0.75	1.05
CdTe	II-VI	Pb_{Cd}	(+1/0)	0.22	0.17
CdTe	II-VI	P_{Te}	(0/-1)	0.07	0.04
CdSe	II-VI	Cu_{Cd}	(0/-1)	0.64	0.32
CdSe	II-VI	Cu_{Cd}	(-1/-2)	1.65	1.75
CdSe	II-VI	Ag_i	(+1/0)	0.99	1.20
CdSe	II-VI	Ag_i	(0/-1)	1.64	1.79
CdSe	II-VI	V_{Cd}	(-1/-2)	0.60	0.42
CdSe	II-VI	Cd_i	(+1/0)	1.65	1.41
CdSe	II-VI	V_{Se}	(+1/0)	1.52	1.11
CdSe	II-VI	V_{Se}	(+2/+1)	1.02	0.99
CdSe	II-VI	Se_i	(+1/0)	0.75	0.70
CdSe	II-VI	Fe_i	(+2/+1)	0.45	0.42
CdSe	II-VI	Co_i	(+1/0)	1.30	1.24
CdS	II-VI	V_{Cd}	(+1/0)	1.33	1.29
CdS	II-VI	Li_i	(+1/0)	1.61	1.59
CdS	II-VI	V_{Cd}	(-1/-2)	0.35	0.57
CdS	II-VI	V_S	(+2/0)	0.84	1.13
CdS	II-VI	Te_i	(+3/+2)	0.20	0.37
CdS	II-VI	I_S	(+1/0)	2.44	1.70
ZnTe	II-VI	N_{Te}	(0/-1)	0.05	0.00
ZnTe	II-VI	Cu_{Zn}	(0/-1)	0.15	0.23
ZnTe	II-VI	Cr_i	(+2/+1)	1.45	1.58
ZnSe	II-VI	Al_{Zn}	(+2/+1)	0.03	0.00
ZnSe	II-VI	F_i	(0/-1)	0.03	0.16
ZnSe	II-VI	Cl_i	(0/-1)	0.03	0.20
ZnSe	II-VI	I_i	(0/-1)	0.03	0.18
ZnSe	II-VI	Cs_{Zn}	(0/-1)	0.07	0.12
ZnSe	II-VI	V_{Zn}	(0/-1)	0.22	0.16
ZnS	II-VI	V_{Zn}	(-1/-2)	0.80	0.88
ZnS	II-VI	V_{Zn}	(0/-1)	0.50	0.27
ZnS	II-VI	V_S	(+2/0)	1.80	1.20
ZnO	II-VI	In_i	D, +3/+2	0.02	0.20
ZnO	II-VI	Cu_i	D, +2/+1	0.19	0.00
ZnO	II-VI	V_{Zn}	(0/-2)	0.07	0.00
ZnO	II-VI	O_i	(0/-1)	0.66	0.80
ZnO	II-VI	V_O	(+1/0)	2.00	1.86
ZnO	II-VI	V_O	(+2/+1)	0.32	0.62
ZnO	II-VI	Cu_i	(+1/0)	0.85	1.38
AlP	III-V	V_{Al}	(0/-1)	0.37	0.47
AlSb	III-V	V_{Al}	(0/-1)	0.04	0.06
GaN	III-V	V_N	(+2/+1)	0.11	0.00
GaN	III-V	V_Ga	(0/-1)	0.23	0.00
GaP	III-V	Ti_i	Acceptor, 0/-1	1.77	1.72

GaP	III-V	Ti_i	Donor, +1/0	1.00	1.31
GaP	III-V	Ni_{Ga}	Acceptor, 0/-1	0.51	0.39
GaP	III-V	Ni_{Ga}	Acceptor, -2/-3	1.45	1.62
GaAs	III-V	V_{Ga}	(-2/-3)	0.58	0.34
GaAs	III-V	As_{Ga}	(0/-1)	0.77	1.01
Si	IV-IV	Cu_i	Donor, +1/0	0.41	0.50
Si	IV-IV	Mn_i	Donor, +2/+1	0.26	0.26
Si	IV-IV	Mn_i	Donor, +1/0	0.76	0.77
Si	IV-IV	Mn_i	Acceptor, 0/-1	1.06	1.18
Si	IV-IV	O_{Si}	Donor, +1/0	1.12	1.06
Si	IV-IV	O_{Si}	Donor, +2/+1	1.05	0.71
Si	IV-IV	Tl_{Si}	Acceptor, 0/-1	0.25	0.36
Si	IV-IV	Ag_i	Donor, +1/0	0.45	0.57
Si	IV-IV	Ag_i	Acceptor, 0/-1	0.83	1.01
Si	IV-IV	Ge_{Si}	Acceptor, 0/-1	0.98	0.78
Si	IV-IV	C_i	Donor, +1/0	0.25	0.20
Si	IV-IV	W_i	Donor, +1/0	0.31	0.12
Si	IV-IV	W_i	Acceptor, 0/-1	0.75	0.72
Si	IV-IV	Sr	Donor, +2/+1	0.50	0.37
Si	IV-IV	Sr	Donor, +1/0	0.84	0.76
Ge	IV-IV	Au	(+1/0)	0.16	0.12
Ge	IV-IV	Au	(0/-1)	0.47	0.37
Ge	IV-IV	Au	(-1/-2)	0.63	0.58
Ge	IV-IV	Pt		0.04	0.00
SiC	IV-IV	V_{Si}	(0/-1)	0.50	0.76
SiC	IV-IV	V_C	(+1/0)	1.20	1.54
SiC	IV-IV	V_C	(+2/+1)	1.28	0.99
SiC	IV-IV	V_{Si}	(0/-1)	0.50	0.76

TABLE SIII: Dominant native defects in each compound and the equilibrium defect formation energy, Fermi level and conductivity determined by them, at A-rich chemical potential conditions.

Semiconductor	**Dominant Defects**	**Eqm. E^f (eV)**	**Eqm. E_F (eV, from VBM)**	**Eqm. Conductivity**
CdO	V_O, q = -1 and Cd_i, q = +1	1.16	0.45	moderately p-type
CdS	V_{Cd}, q = -2 and V_S, q = +1	1.82	1.52	moderately n-type
CdSe	V_{Cd}, q = -2 and Cd_i, q = +1	1.61	1.18	moderately n-type
CdTe	V_{Cd}, q = -2 and Cd_i, q = +1	1.48	0.70	intrinsic
ZnO	V_O, q = -1 and Zn_O, q = +2	1.94	0.89	moderately p-type
ZnS	V_{Zn}, q = -2 and Zn_S, q = +1	2.48	1.61	intrinsic
ZnSe	V_{Se}, q = -1 and Zn_{Se}, q = +1	2.25	1.38	intrinsic
ZnTe	V_{Zn}, q = -2 and Zn_i, q = +1	1.58	0.83	moderately p-type
BN	V_B, q = -1 and V_N, q = +1	5.29	3.70	moderately n-type
BP	B_P, q = -1 and P_B, q = +2	2.57	-0.09	very p-type
BAs	V_B, q = +1 and B_{As}, q = -1	2.33	-0.13	very p-type
BSb	V_B, q = -1 and Sb_B, q = +3	1.30	-0.15	very p-type
AlN	V_{Al}, q = -2 and V_N, q = +1	4.15	4.44	moderately n-type
AlP	Al_P, q = -1 and Al_i, q = +2	3.27	1.12	intrinsic
AlAs	Al_{As}, q = -1 and Al_i, q = +2	2.80	0.74	moderately p-type
AlSb	V_{Al}, q = -3 and Sb_i, q = +3	1.18	0.97	moderately n-type
GaN	V_{Ga}, q = -3 and V_N, q = +1	3.82	3.07	very n-type
GaP	V_P, q = +1 and Ga_P, q = -1	2.16	0.26	moderately p-type
GaAs	Ga_{As}, q = -1 and Ga_i, q = +1	1.62	-0.16	very p-type
GaSb	Ga_{Sb}, q = -1 and Ga_i, q = +1	0.81	-0.15	very p-type
InN	V_N, q = -1 and In_i, q = +2	2.47	0.09	moderately p-type
InP	V_P, q = +1 and In_P, q = -1	2.14	0.65	intrinsic
InAs	V_{As}, q = -2 and In_i, q = +1	1.90	0.58	very n-type
InSb	In_{Sb}, q = -1 and In_i, q = +1	1.01	0.22	moderately p-type
C	V_C, q = -1 and C_i, q = +3	11.06	-2.60	very p-type
Si	V_{Si}, q = +1 and Si_i, q = -1	4.11	0.61	intrinsic
Ge	V_{Ge}, q = -1 and Ge_i, q = +3	2.21	-0.26	very p-type
Sn	V_{Sn}, q = -2 and Sn_i, q = +2	1.3	-0.11	very p-type
SiC	V_C, q = +2 and C_{Si}, q = -1	4.54	1.30	intrinsic
GeC	V_C, q = +1 and C_{Ge}, q = -1	3.36	0.69	intrinsic
SnC	V_C, q = +2 and C_{Sn}, q = -1	2.96	0.60	intrinsic
SiGe	Ge_{Si}, q = +1 and Si_{Ge}, q = -1	0.42	0.51	moderately n-type
SiSn	Sn_{Si}, q = +1 and Si_{Sn}, q = -1	0.29	0.75	very n-type
GeSn	Sn_{Ge}, q = +1 and Ge_{Sn}, q = -1	0.19	0.15	moderately p-type

TABLE SIV: Dominant native defects in each compound and the equilibrium defect formation energy, Fermi level and conductivity determined by them, at B-rich chemical potential conditions.

Semiconductor	**Dominant Defects**	**Eqm. E^f (eV)**	**Eqm. E_F (eV, from VBM)**	**Eqm. Conductivity**
CdO	V_{Cd}, q = -2 and V_O, q = +1	2.72	0.75	moderately p-type
CdS	V_{Cd}, q = -2 and V_S, q = +2	2.02	0.77	moderately p-type
CdSe	V_{Cd}, q = -2 and Cd_i, q = +2	1.73	0.48	moderately p-type
CdTe	V_{Cd}, q = -1 and Te_{Cd}, q = +1	1.54	0.14	very p-type
ZnO	V_{Zn}, q = -1 and O_i, q = +1	3.41	-0.06	very p-type
ZnS	V_{Zn}, q = -1 and Zn_i, q = +2	2.39	0.43	moderately p-type
ZnSe	V_{Zn}, q = -1 and Se_{Zn}, q = +2	1.85	0.19	very p-type
ZnTe	V_{Zn}, q = -1 and Zn_i, q = +2	1.49	-0.01	very p-type
BN	N_B, q = +1 and N_i, q = -1	0.99	5.66	very n-type
BP	B_P, q = -1 and P_B, q = +1	2.88	1.23	moderately n-type
BAs	B_{As}, q = -1 and As_B, q = +2	2.29	0.06	very p-type
BSb	V_B, q = +1 and B_{Sb}, q = -1	1.54	-0.36	very p-type
AlN	N_{Al}, q = -1 and N_i, q = +1	-1.46	0.40	very p-type
AlP	V_{Al}, q = -2 and P_{Al}, q = +2	2.78	1.29	intrinsic
AlAs	V_{Al}, q = -2 and As_{Al}, q = +1	1.73	1.17	intrinsic
AlSb	V_{Al}, q = -3 and Sb_i, q = +3	0.82	0.97	moderately n-type
GaN	V_{Ga}, q = -2 and N_{Ga}, q = +1	-1.00	2.38	moderately n-type
GaP	V_{Ga}, q = -3 and P_{Ga}, q = +1	2.20	1.08	intrinsic
GaAs	V_{Ga}, q = -2 and As_{Ga}, q = +2	1.73	0.32	moderately p-type
GaSb	Ga_{Sb}, q = -2 and Sb_{Ga}, q = +1	1.12	0.21	moderately p-type
InN	V_{In}, q = -3 and N_{In}, q = +1	-1.15	1.93	very n-type
InP	V_P, q = -2 and P_{In}, q = +1	2.26	1.04	moderately n-type
InAs	V_{In}, q = -2 and As_{In}, q = +1	1.52	0.67	very n-type
InSb	V_{In}, q = -1 and Sb_{In}, q = +1	1.32	0.56	very n-type
C	V_C, q = -1 and C_i, q = +1	12.95	-4.48	very p-type
Si	V_{Si}, q = +1 and Si_i, q = -1	4.11	0.61	intrinsic
Ge	V_{Ge}, q = -1 and Ge_i, q = +2	2.17	-0.21	very p-type
Sn	V_{Sn}, q = -1 and Sn_i, q = +2	1.35	-0.09	very p-type
SiC	V_C, q = +1 and C_{Si}, q = -1	4.17	0.83	moderately p-type
GeC	V_C, q = -1 and Ge_C, q = +1	2.90	1.33	very n-type
SnC	V_C, q = -1 and Sn_C, q = +2	1.55	0.41	moderately p-type
SiGe	Ge_{Si}, q = +1 and Si_{Ge}, q = -1	0.42	0.32	moderately p-type
SiSn	Sn_{Si}, q = +1 and Si_{Sn}, q = -1	0.29	0.07	moderately p-type
GeSn	Sn_{Ge}, q = +1 and Ge_{Sn}, q = -1	0.19	0.04	very p-type

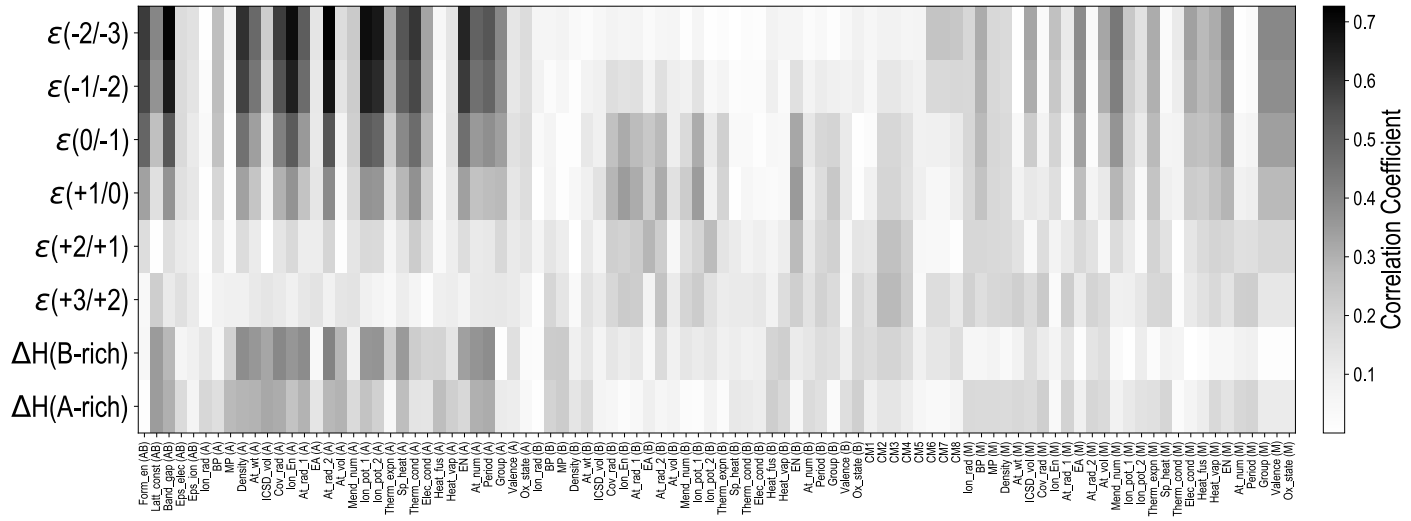

FIG. S2: Absolute values of coefficient of linear correlation between every unique descriptor and every property; darker boxes imply high correlation whereas white boxes mean there is zero correlation. The descriptor labels and the corresponding linear correlation values can also be found in Table SV.

TABLE SV: Tabulated absolute values of coefficient of linear correlation between every unique descriptor and every property.

Labels	**ΔH (A-rich)**	**ΔH (B-rich)**	**ε(+3/+2)**	**ε(+2/+1)**	**ε(+1/0)**	**ε(0/-1)**	**ε(-1/-2)**	**ε(-2/-3)**
Form en (AB)	0.01	0.05	0.04	0.17	0.34	0.49	0.56	0.59
Latt const (AB)	0.35	0.35	0.22	0.00	0.15	0.26	0.37	0.40
Band gap (AB)	0.30	0.28	0.12	0.16	0.37	0.53	0.66	0.72
Eps elec (AB)	0.17	0.06	0.16	0.11	0.15	0.17	0.17	0.17
Eps ion (AB)	0.07	0.09	0.11	0.10	0.12	0.12	0.14	0.15
Ion rad (A)	0.19	0.13	0.02	0.00	0.00	0.03	0.02	0.01
BP (A)	0.16	0.06	0.08	0.13	0.19	0.25	0.27	0.26
MP (A)	0.28	0.21	0.09	0.03	0.01	0.01	0.03	0.06
Density (A)	0.29	0.38	0.09	0.17	0.34	0.46	0.57	0.61
At wt (A)	0.29	0.36	0.12	0.11	0.25	0.34	0.44	0.48
ICSD vol (A)	0.32	0.33	0.14	0.00	0.07	0.12	0.19	0.22
Cov rad (A)	0.31	0.40	0.12	0.13	0.30	0.42	0.54	0.59
Ion En (A)	0.25	0.36	0.09	0.18	0.37	0.52	0.65	0.69
At rad 1 (A)	0.29	0.39	0.12	0.10	0.25	0.36	0.47	0.51
EA (A)	0.13	0.04	0.14	0.10	0.13	0.12	0.12	0.13
At rad 2 (A)	0.28	0.41	0.08	0.19	0.39	0.54	0.68	0.73
At vol (A)	0.29	0.29	0.13	0.01	0.04	0.07	0.14	0.17
Mend num (A)	0.17	0.08	0.14	0.14	0.18	0.22	0.22	0.23
Ion pot 1 (A)	0.25	0.36	0.09	0.18	0.37	0.52	0.65	0.69
Ion pot 2 (A)	0.21	0.37	0.07	0.18	0.36	0.49	0.62	0.67
Therm expn (A)	0.04	0.22	0.04	0.06	0.16	0.22	0.30	0.31
Sp heat (A)	0.27	0.35	0.10	0.13	0.28	0.40	0.50	0.53
Therm cond (A)	0.14	0.22	0.06	0.22	0.38	0.47	0.56	0.60
Elec cond (A)	0.13	0.19	0.02	0.14	0.22	0.29	0.32	0.33
Heat fus (A)	0.27	0.20	0.09	0.04	0.02	0.02	0.02	0.05
Heat vap (A)	0.21	0.13	0.10	0.10	0.12	0.16	0.15	0.13
EN (A)	0.18	0.34	0.07	0.16	0.34	0.46	0.59	0.63
At num (A)	0.30	0.37	0.12	0.11	0.25	0.36	0.46	0.50
Period (A)	0.31	0.38	0.12	0.12	0.27	0.38	0.49	0.53
Group (A)	0.11	0.01	0.12	0.18	0.28	0.34	0.38	0.40
Valence (A)	0.10	0.15	0.11	0.12	0.13	0.15	0.12	0.13
Ox state (A)	0.04	0.03	0.08	0.13	0.17	0.17	0.16	0.17
Ion rad (B)	0.01	0.02	0.01	0.06	0.01	0.03	0.08	0.06
BP (B)	0.20	0.22	0.19	0.13	0.07	0.07	0.02	0.06
MP (B)	0.22	0.23	0.14	0.10	0.01	0.01	0.05	0.07
Density (B)	0.14	0.09	0.04	0.00	0.05	0.01	0.01	0.03
At wt (B)	0.17	0.13	0.15	0.12	0.09	0.12	0.07	0.02
ICSD vol (B)	0.06	0.07	0.11	0.07	0.15	0.09	0.09	0.05
Cov rad (B)	0.05	0.13	0.19	0.21	0.29	0.26	0.16	0.07
Ion En (B)	0.02	0.11	0.22	0.20	0.35	0.31	0.15	0.06
At rad 1 (B)	0.04	0.12	0.22	0.22	0.31	0.27	0.15	0.07
EA (B)	0.10	0.04	0.07	0.29	0.22	0.24	0.08	0.03
At rad 2 (B)	0.06	0.17	0.25	0.22	0.31	0.28	0.15	0.05
At vol (B)	0.07	0.08	0.11	0.05	0.15	0.10	0.10	0.06
Mend num (B)	0.02	0.13	0.19	0.16	0.22	0.18	0.08	0.01
Ion pot 1 (B)	0.01	0.11	0.22	0.19	0.35	0.31	0.15	0.06
Ion pot 2 (B)	0.07	0.13	0.15	0.27	0.08	0.11	0.04	0.01
Therm expn (B)	0.02	0.09	0.11	0.14	0.20	0.21	0.07	0.02
Sp heat (B)	0.12	0.10	0.11	0.11	0.01	0.04	0.02	0.00
Therm cond (B)	0.11	0.08	0.12	0.07	0.05	0.04	0.02	0.02
Elec cond (B)	0.10	0.09	0.12	0.09	0.03	0.01	0.02	0.01
Heat fus (B)	0.22	0.20	0.08	0.06	0.03	0.05	0.10	0.08
Heat vap (B)	0.18	0.21	0.16	0.10	0.04	0.04	0.03	0.06
EN (B)	0.05	0.11	0.21	0.28	0.35	0.32	0.15	0.06
At num (B)	0.17	0.13	0.15	0.12	0.09	0.12	0.07	0.02
Period (B)	0.16	0.16	0.20	0.19	0.17	0.19	0.11	0.04
Group (B)	0.03	0.15	0.17	0.21	0.23	0.20	0.09	0.01
Valence (B)	0.16	0.10	0.04	0.04	0.11	0.11	0.08	0.04

Ox state (B)	0.21	0.18	0.19	0.13	0.01	0.01	0.10	0.08
CM1	0.09	0.17	0.15	0.13	0.08	0.02	0.04	0.05
CM2	0.13	0.20	0.28	0.26	0.19	0.19	0.12	0.07
CM3	0.13	0.20	0.28	0.26	0.19	0.19	0.12	0.07
CM4	0.10	0.17	0.22	0.21	0.15	0.16	0.10	0.06
CM5	0.05	0.09	0.03	0.05	0.09	0.08	0.08	0.06
CM6	0.03	0.13	0.17	0.04	0.05	0.09	0.18	0.25
CM7	0.03	0.13	0.17	0.04	0.05	0.09	0.18	0.25
CM8	0.06	0.16	0.12	0.01	0.10	0.12	0.19	0.24
Ion rad (M)	0.18	0.05	0.21	0.19	0.18	0.19	0.18	0.19
BP (M)	0.18	0.05	0.16	0.19	0.25	0.28	0.30	0.31
MP (M)	0.18	0.07	0.18	0.18	0.18	0.19	0.18	0.18
Density (M)	0.17	0.03	0.19	0.18	0.17	0.18	0.17	0.17
At wt (M)	0.18	0.07	0.21	0.15	0.07	0.04	0.00	0.01
ICSD vol (M)	0.17	0.16	0.17	0.04	0.12	0.22	0.31	0.33
Cov rad (M)	0.21	0.11	0.23	0.16	0.11	0.07	0.04	0.03
Ion En (M)	0.04	0.07	0.03	0.07	0.16	0.20	0.23	0.26
At rad 1 (M)	0.20	0.14	0.22	0.12	0.02	0.05	0.10	0.11
EA (M)	0.09	0.02	0.11	0.18	0.27	0.34	0.38	0.39
At rad 2 (M)	0.19	0.14	0.20	0.14	0.06	0.03	0.02	0.03
At vol (M)	0.17	0.16	0.17	0.04	0.12	0.22	0.31	0.32
Mend num (M)	0.08	0.03	0.09	0.17	0.29	0.37	0.42	0.44
Ion pot 1 (M)	0.12	0.01	0.09	0.11	0.17	0.19	0.21	0.24
Ion pot 2 (M)	0.14	0.10	0.13	0.03	0.04	0.11	0.15	0.15
Therm expn (M)	0.15	0.02	0.19	0.20	0.25	0.28	0.30	0.31
Sp heat (M)	0.19	0.07	0.19	0.15	0.09	0.06	0.03	0.03
Therm cond (M)	0.02	0.06	0.02	0.01	0.07	0.07	0.10	0.12
Elec cond (M)	0.07	0.00	0.11	0.13	0.22	0.26	0.32	0.34
Heat fus (M)	0.10	0.02	0.15	0.17	0.21	0.25	0.28	0.28
Heat vap (M)	0.18	0.05	0.18	0.20	0.25	0.28	0.30	0.31
EN (M)	0.14	0.04	0.14	0.19	0.29	0.35	0.39	0.40
At num (M)	0.19	0.08	0.21	0.15	0.08	0.05	0.01	0.00
Period (M)	0.20	0.09	0.22	0.16	0.10	0.07	0.03	0.03
Group (M)	0.11	0.01	0.12	0.18	0.28	0.34	0.38	0.40
Valence (M)	0.11	0.01	0.12	0.18	0.28	0.34	0.38	0.40
Ox state (M)	0.11	0.01	0.12	0.18	0.28	0.34	0.38	0.40

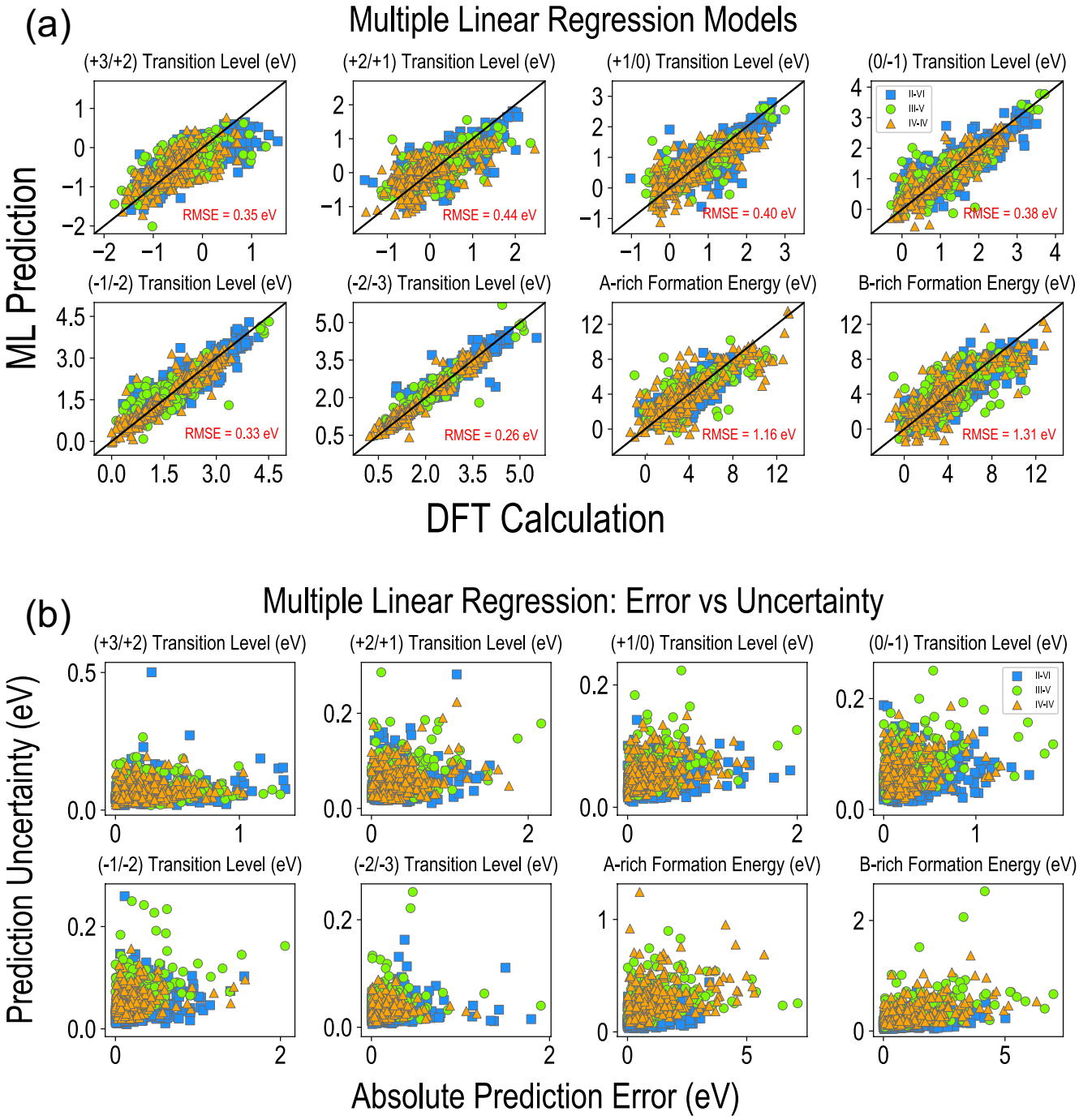

FIG. S3: Multiple Linear Regression results: (a) parity plots, and (b) prediction uncertainty as a function of absolute prediction error.

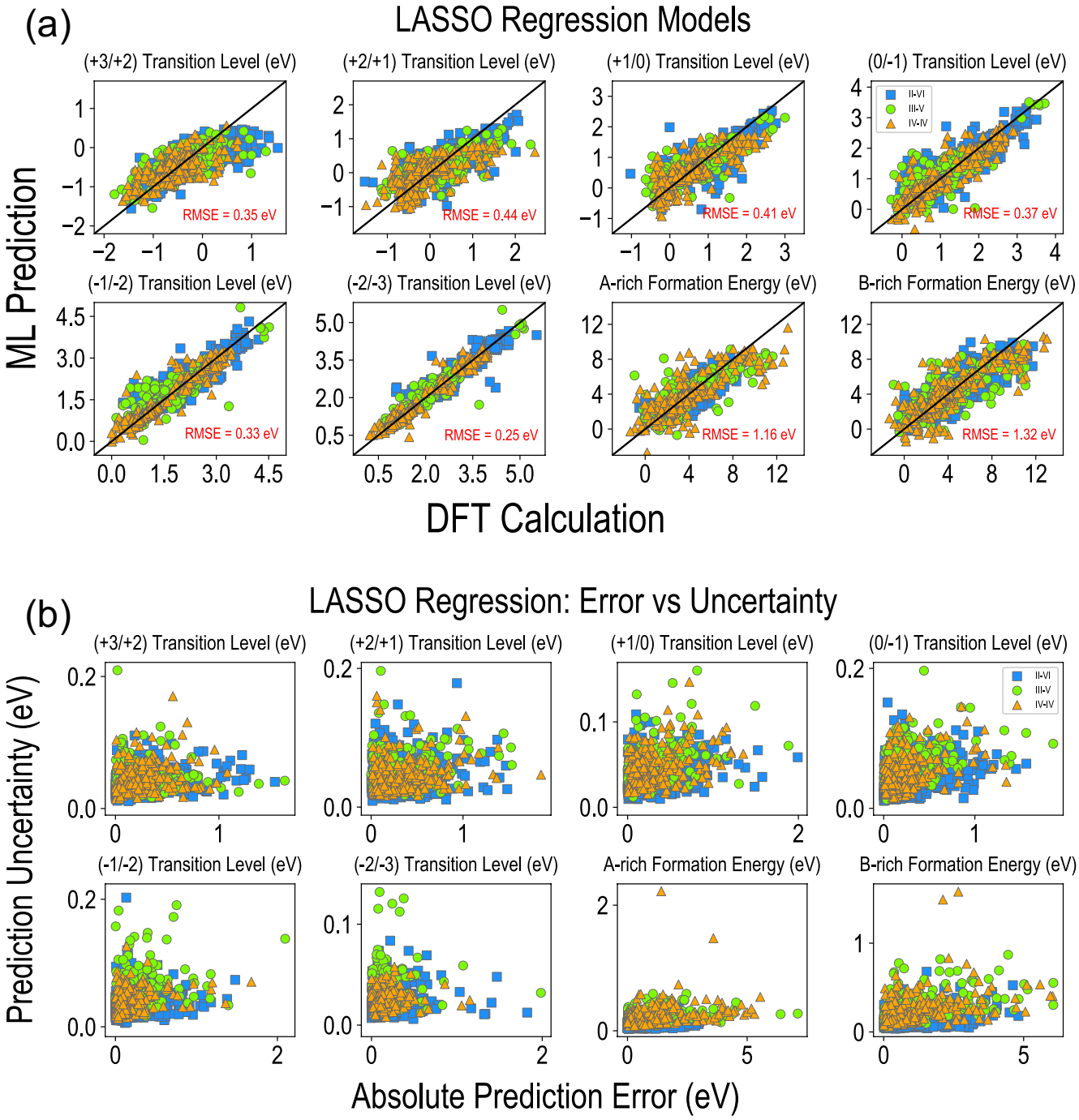

FIG. S4: LASSO Regression results: (a) parity plots, and (b) prediction uncertainty as a function of absolute prediction error.

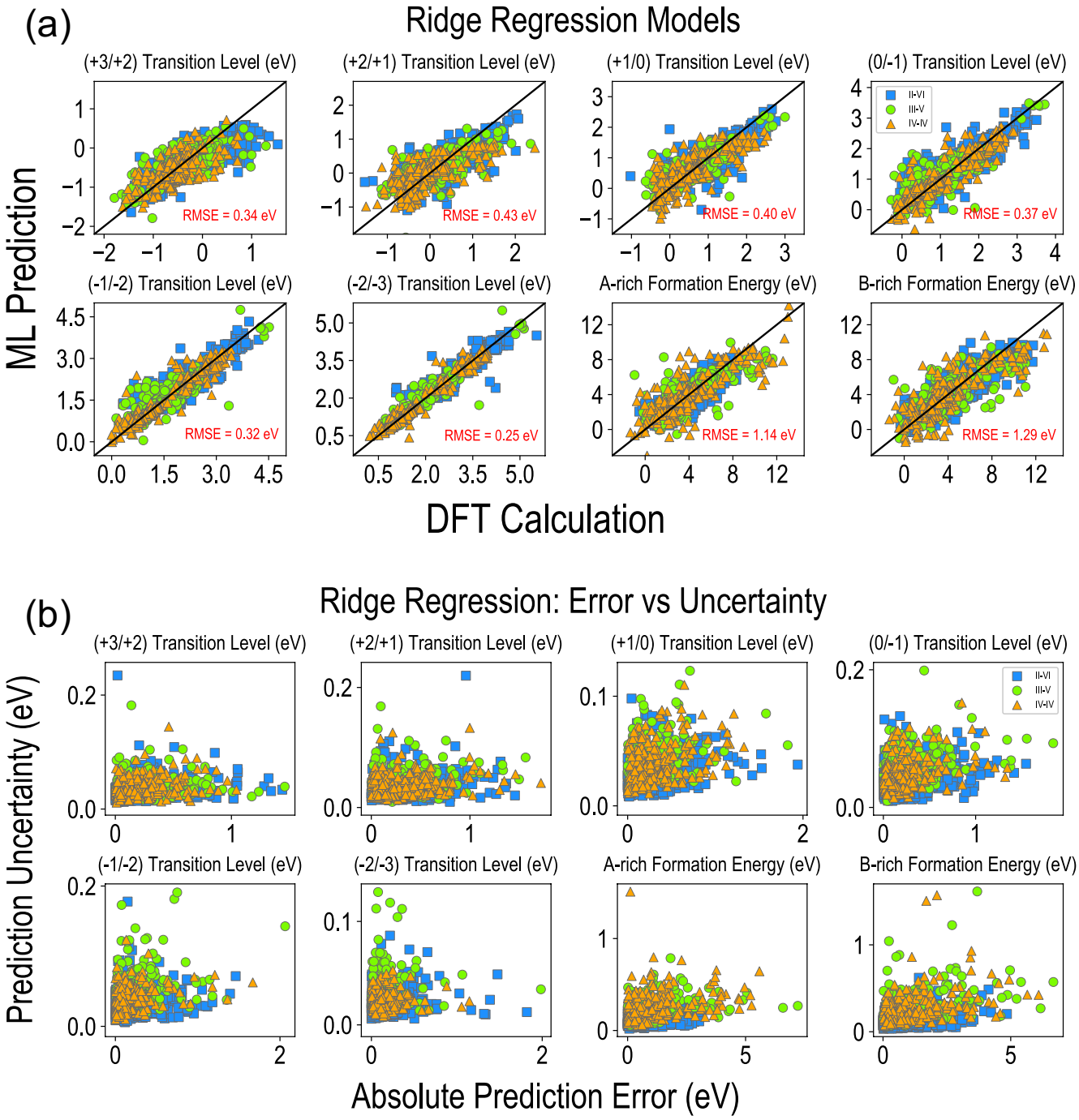

FIG. S5: Ridge Regression results: (a) parity plots, and (b) prediction uncertainty as a function of absolute prediction error.

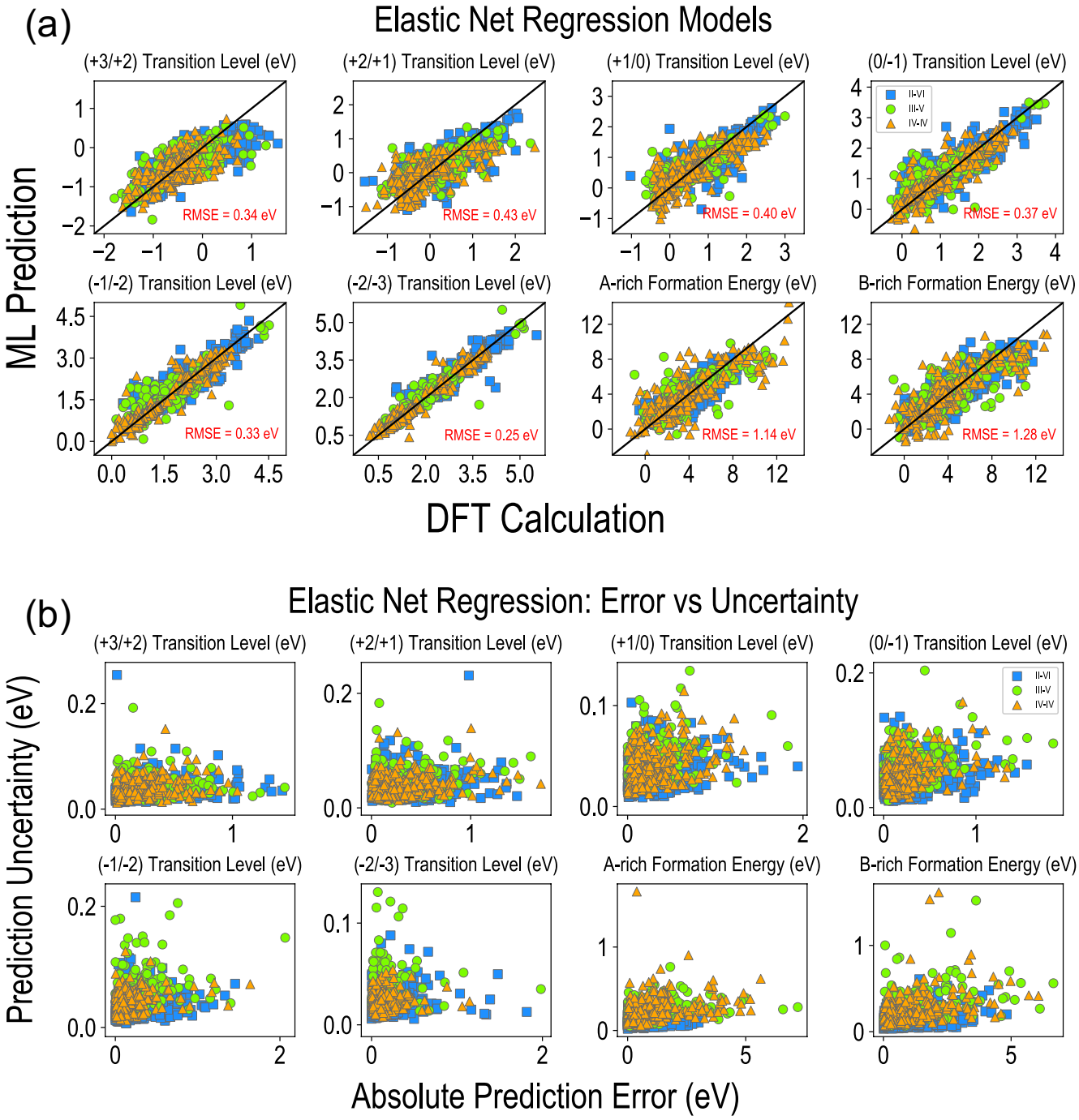

FIG. S6: Elastic Net Regression results: (a) parity plots, and (b) prediction uncertainty as a function of absolute prediction error.

TABLE SVI: ML training set prediction RMSE values for transition levels.

Property	**ML Method**	**II-VI Error (eV)**	**III-V Error (eV)**	**IV-IV Error (eV)**	**Total Error (eV)**
ε(+3/+2)	MLR	0.316	0.313	0.283	0.308
ε(+3/+2)	Ridge	0.324	0.319	0.284	0.314
ε(+3/+2)	LASSO	0.331	0.326	0.287	0.320
ε(+3/+2)	Elastic Net	0.323	0.319	0.284	0.313
ε(+3/+2)	RFR	0.213	0.255	0.203	0.222
ε(+3/+2)	KRR	0.231	0.259	0.200	0.231
ε(+3/+2)	GPR	0.237	0.262	0.209	0.237
ε(+3/+2)	NN	0.189	0.227	0.160	0.193
ε(+2/+1)	MLR	0.390	0.383	0.418	0.395
ε(+2/+1)	Ridge	0.403	0.389	0.422	0.404
ε(+2/+1)	LASSO	0.408	0.390	0.422	0.407
ε(+2/+1)	Elastic Net	0.401	0.387	0.420	0.403
ε(+2/+1)	RFR	0.243	0.290	0.275	0.262
ε(+2/+1)	KRR	0.218	0.217	0.231	0.221
ε(+2/+1)	GPR	0.177	0.172	0.178	0.176
ε(+2/+1)	NN	0.202	0.201	0.208	0.203
ε(+1/0)	MLR	0.371	0.329	0.377	0.363
ε(+1/0)	Ridge	0.380	0.343	0.390	0.374
ε(+1/0)	LASSO	0.386	0.347	0.393	0.379
ε(+1/0)	Elastic Net	0.378	0.342	0.389	0.373
ε(+1/0)	RFR	0.248	0.278	0.257	0.257
ε(+1/0)	KRR	0.191	0.168	0.202	0.189
ε(+1/0)	GPR	0.122	0.101	0.127	0.119
ε(+1/0)	NN	0.201	0.156	0.211	0.194
ε(0/-1)	MLR	0.333	0.324	0.279	0.319
ε(0/-1)	Ridge	0.342	0.337	0.303	0.332
ε(0/-1)	LASSO	0.341	0.338	0.301	0.331
ε(0/-1)	Elastic Net	0.341	0.336	0.302	0.331
ε(0/-1)	RFR	0.213	0.236	0.205	0.217
ε(0/-1)	KRR	0.205	0.179	0.171	0.192
ε(0/-1)	GPR	0.129	0.089	0.099	0.114
ε(0/-1)	NN	0.204	0.167	0.190	0.193
ε(-1/-2)	MLR	0.294	0.286	0.253	0.283
ε(-1/-2)	Ridge	0.301	0.306	0.258	0.293
ε(-1/-2)	LASSO	0.301	0.305	0.257	0.292
ε(-1/-2)	Elastic Net	0.299	0.300	0.255	0.289
ε(-1/-2)	RFR	0.221	0.262	0.216	0.230
ε(-1/-2)	KRR	0.178	0.157	0.145	0.166
ε(-1/-2)	GPR	0.134	0.106	0.101	0.121
ε(-1/-2)	NN	0.208	0.176	0.200	0.199
ε(-2/-3)	MLR	0.253	0.212	0.194	0.231
ε(-2/-3)	Ridge	0.257	0.224	0.198	0.237
ε(-2/-3)	LASSO	0.257	0.223	0.198	0.236
ε(-2/-3)	Elastic Net	0.257	0.223	0.197	0.236
ε(-2/-3)	RFR	0.202	0.182	0.167	0.190
ε(-2/-3)	KRR	0.232	0.161	0.156	0.201
ε(-2/-3)	GPR	0.182	0.133	0.130	0.160
ε(-2/-3)	NN	0.215	0.142	0.169	0.190

TABLE SVII: ML training set prediction RMSE values for formation energies.

Property	**ML Method**	**II-VI Error (eV)**	**III-V Error (eV)**	**IV-IV Error (eV)**	**Total Error (eV)**
ΔH (A-rich)	MLR	0.786	1.265	1.439	0.982
ΔH (A-rich)	Ridge	0.796	1.273	1.460	0.993
ΔH (A-rich)	LASSO	0.832	1.322	1.505	1.033
ΔH (A-rich)	Elastic Net	0.800	1.279	1.466	0.998
ΔH (A-rich)	RFR	0.440	0.860	1.045	0.632
ΔH (A-rich)	KRR	0.436	0.678	0.657	0.513
ΔH (A-rich)	GPR	0.318	0.482	0.519	0.380
ΔH (A-rich)	NN	0.489	0.645	0.765	0.560
ΔH (B-rich)	MLR	0.960	1.330	1.387	1.089
ΔH (B-rich)	Ridge	0.974	1.345	1.419	1.106
ΔH (B-rich)	LASSO	1.015	1.397	1.495	1.155
ΔH (B-rich)	Elastic Net	0.976	1.347	1.424	1.108
ΔH (B-rich)	RFR	0.556	0.918	1.074	0.711
ΔH (B-rich)	KRR	0.504	0.680	0.666	0.558
ΔH (B-rich)	GPR	0.527	0.773	0.793	0.611
ΔH (B-rich)	NN	0.634	0.713	0.723	0.659