Approaching chemical accuracy with density functional calculations: Diatomic energy corrections

Scott Grindy, Bryce Meredig, Scott Kirklin, James E. Saal, and C. Wolverton

Phys. Rev. B **87**, 075150 — Published 28 February 2013

DOI: 10.1103/PhysRevB.87.075150

Approaching chemical accuracy with density functional calculations: Diatomic energy corrections

Scott Grindy, Bryce Meredig, Scott Kirklin, James Saal, C. Wolverton
Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208

Density functional theory is widely used to predict materials properties, but the local density approximation and generalized gradient approximation exchange-correlation functionals are known to poorly predict the energetics of reactions involving molecular species. In this paper, we obtain corrections for the O_2, H_2, N_2, F_2, and Cl_2 molecules within the Perdew-Burke-Enzerhof GGA, Perdew-Wang GGA, and Perdew-Zunger LDA exchange-correlation functionals by comparing DFT-calculated formation energies of oxides, hydrides, nitrides, fluorides, and chlorides to experimental values. We also show that the choice of compounds used to obtain the correction is significant, and we use a leave one out cross-validation approach to rigorously determine the proper fit set. We report confidence intervals with our correction values, which quantifies the variation caused by the choice of fit set after outlier removal. The remaining variation in the correction values is on the order of 1 kcal/mol, which indicates that chemical accuracy is a realistic goal for these systems.

PACS numbers:

I. INTRODUCTION

Density functional theory (DFT) is a ubiquitous and robust tool for computationally predicting materials properties. DFT is, in principle, an exact theory; in practice, however, one must approximate the exchange-correlation (XC) component of the DFT energy functional. Widespread XC approximations in the solid-state physics community, particularly varieties of the local-density approximation (LDA) and generalized gradient approximation (GGA), are typically very efficient and accurate for periodic solids. Unfortunately, these XC approximations often fail badly on molecular species, giving errors of up to 10 or 20% with respect to experimental bond lengths, bond angles, and formation energies[1–4]. However, we require simultaneously accurate energetics for solids and molecules when calculating the thermodynamics of many technologically relevant reactions, in application areas such as catalysis, hydrogen storage, fuel cells, and batteries[5–8]. Chemical accuracy (commonly taken as 1 kcal/mol $\approx$ 40 meV/formula unit), the accuracy of many thermochemical experiments, represents the ultimate goal for first-principles calculations, although it remains elusive in many systems. For example, the O_2 binding energy when calculated with GGA is in error by 2.40 eV/O_2 within LDA and 1.06 eV/O_2 within GGA[4]. Here, we attempt to approach chemical accuracy by correcting the large errors associated with the O_2, H_2, N_2, Cl_2, and F_2 molecules within three common XC functionals: the Perdew-Burke-Enzerhof GGA[9] (PBE-GGA), Perdew-Wang 1991 GGA[10,11] (PW91-GGA) and the Perdew-Zunger LDA[12] (PZ-LDA) as implemented within the VASP code.

Others have pointed out DFT's discrepancies with respect to experimental thermodynamics for certain classes of compounds, and have proposed various remedies[13–17]. For example, Stevanović *et al* fit one correction factor for each of fifty of the elements using experimental data from 252 binary compounds. The fitted correction factors were then used to improve agreement between their calculations and experimental values for formation energies of chalcogenides and pnictides[15]. Wang *et al*[14] and Lee *et al*[17] followed a more conservative approach, fitting a correction factor to the O_2 molecule using a set of seven binary oxides. Jain *et al*[16] proposed a mixed GGA/GGA+U framework to more accurately predict the formation energies of 49 ternary oxides.

While the practice of fitting a large number of correction factors across many elements raises the question of overfitting, we can say for certain that particular elemental reference states are problematic: the diatomic molecules. To identify errors associated with the diatomic molecules specifically, we take advantage of the fact that DFT's standard XC functionals are quite successful at calculating reaction energies involving only solid constituents[18,19] and use them to isolate the error associated with the diatomic molecules.

Following the approach of Wang, Maxisch, and Ceder[14], we fit fifteen diatomic energy corrections (five diatomic molecules times three XC functionals), which are a function only of the particular diatomic molecule and XC functional. Each diatomic energy correction is fit from a set of experimental and DFT-calculated reaction enthalpies from the general set of reactions

$$\frac{2a}{b}M + X_2 \longrightarrow \frac{2}{b}M_aX_b \tag{1}$$

where M is a metal and X = O, H, N, F, or Cl. All reactions studied here form binary oxides, hydrides, nitrides,

fluorides, or chlorides as their products. The reaction has been normalized per molecule X_2 for convenience. Since we expect M and M_aX_b to be relatively well-treated by DFT (with certain notable exceptions, such as transition metal oxides and rare earth compounds, which we omit from the present study), we can assume that the majority of the discrepancy between experimental and DFT-calculated reaction energies is due to the DFT calculation of X_2.

Each diatomic energy correction C is fitted to a set of data $\{\Delta H^{Expt.}(M_aX_b), \Delta H^{DFT}_{XC}(M_aX_b)\}$ with a fixed X and XC functional. If our hypothesis is correct and the majority of the discrepancy can be attributed to the diatomic molecule, a plot of $\Delta H^{Expt.}$ vs. ΔH^{DFT}_{XC} for a fixed X and XC functional will fall on a straight line, and the y-intercept will be C:

$$\Delta H^{Expt.}(M_aX_b) = \Delta H^{DFT}_{XC}(M_aX_b) + C(X_2, xc) \tag{2}$$

At zero temperature and pressure the enthalpy of a system is equal to its total energy, so we can calculate the enthalpy of each reaction within DFT by the equation

$$\Delta H^{DFT} = \frac{2}{b}E^{DFT}(M_aX_b) - \{\frac{2a}{b}E^{DFT}(M) + E^{DFT}(X_2)\} \tag{3}$$

Thus C corrects for the reaction energy error associated with the total energy calculation of the diatomic molecule. Wang *et al* calculated $C(O_2, PBE\text{-}GGA)$ to be -1.36 eV/O_2 molecule. We significantly extend this approach by applying it to other molecules and XC functionals, namely the O_2, H_2, N_2, F_2, and Cl_2 molecules and the PBE-GGA, PW91-GGA, and PZ-LDA parameterizations. Our study excludes all transition metals and all f-block metals, which are known to present additional problems for DFT calculations and could confound the error ascribed to the diatomic molecules (see the discussion about self-interaction errors in the introduction of Ref.[14]). It should be noted that this method will only quantify the error due to the diatomic molecule; any further error in the total energy can be ascribed to the other reactants or products. We also note, of course, that uncertainty exists in all experimental thermochemical measurements, which we do not consider here.

The fifteen diatomic energy corrections can be applied to significantly improve the predictive power of DFT energetics in any cases involving the aforementioned five diatomic molecules and three XC functionals. As we show below, our correction factors enable near-chemical accuracy for many of these practically important reactions within DFT.

II. METHODOLOGY

A. DFT calculations

We performed all calculations in this work with projector augmented wave (PAW) potentials[20,21] supplied with the Vienna *ab initio* Simulation Package (VASP)[22–25]. Our error correction values pertain explicitly to the set of PAW potentials we used, as specified in Table IV in the Appendix (see Section IV for further discussion on transferability to other sets of atomic potentials). All calculations used a cutoff energy of 520 eV for the plane-wave basis set. All calculations were not spin-polarized with the exception of the O_2, H_2, N_2, F_2, and Cl_2 molecules. A list of the crystal structures used for each compound is provided in the Appendix. Final energies were computed with a gamma-centered k-point mesh using approximately 8000 k-points per reciprocal atom.

B. Construction of binary compound fit set

Wang *et al* selected a set of six binary oxides (CaO, Li_2O, MgO, Al_2O_3, SiO_2, and Na_2O) with which to fit a value of C for O_2, based on the assumption that these s and p block oxides should be relatively well-behaved in DFT. We critically examine the choice of fit set, and as we show below, different choices of fit compounds can strongly influence the value of C, and further, the selection of the fit set can be put on more rigorous footing than relying on chemical intuition alone.

Leave one out cross-validation (LOOCV) is a technique commonly used to evaluate the statistical significance of each term in a proposed model or fit[26,27]. For our purposes, LOOCV allows us to quantitatively identify outliers for exclusion from our fit sets. During LOOCV, one data pair of ($\Delta H^{Expt.}(M_aX_b)$, $\Delta H^{DFT}_{XC}(M_aX_b)$) is removed from the data set, and the remaining set is least-squares fit to Eqn. 2. We then calculate the squared error from the removed data point to the fitted equation. The square root of the squared error is defined as the CV score for that data point, such that it has dimensions of energy.

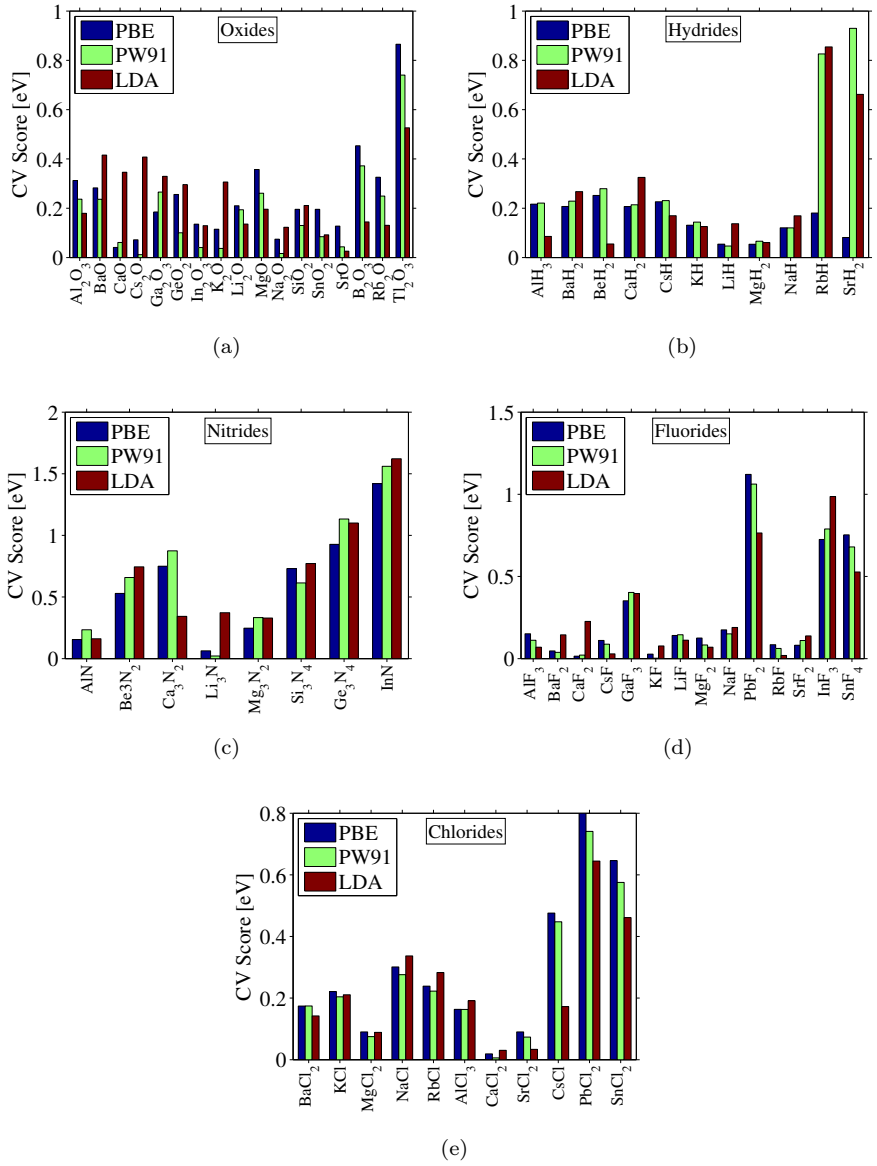

Figure 1: Cross-validation scores for each set of compounds, in units of eV. Higher CV scores indicate outlier compounds, which we then remove from the fitting sets via the LOOCV process described in Section II B. Tl_2O_3, RbH (PW91, LDA), SrH_2 (PW91, LDA), InN, InF_3 (PW91, LDA), PbF_2, and $PbCl_2$ were removed from the fitting sets.

The cross-validation process is repeated for each data point until each has been ascribed a CV score. We give the results of this CV approach for each X_2 molecule and XC functional in Figure 1. Much higher (i.e., outlier) CV scores indicate compounds for which our assumption of error dominated by X_2 molecules breaks down. Thus, the cross-validation approach shows us in an automated way where either (a) there is significant unexpected error in the DFT calculation of the M_aX_b compound itself; or (b) the experimental data may need to be re-examined. These compounds have been removed from the fitting set to increase the quality of the fit and the precision of the value for C. Compounds with CV scores greater than 1.5 standard deviations from the mean CV score for their respective XC functional and diatomic molecule were deemed outliers and removed from the fitting sets. In total, Tl_2O_3, RbH (PW91, LDA), SrH_2 (PW91, LDA), InN, InF_3 (PW91, LDA), PbF_2, and $PbCl_2$ were removed from the fitting sets.

Table I: Correction factors C fit from the equation $\Delta H^{Expt.} = \Delta H^{DFT}_{XC} + C$ (Eqn. 2) for each molecule and exchange-correlation potential. All units are in eV/X_2molecule, where X = O, H, N, F, or Cl. Ranges given are 95% confidence intervals, as described in Section III.

Molecule	PBE-GGA	PW91-GGA	PZ-LDA
O_2	-1.198 ± 0.007	-0.889 ± 0.006	0.254 ± 0.008
H_2	-0.284 ± 0.011	-0.182 ± 0.015	0.170 ± 0.014
N_2	-0.892 ± 0.063	-0.669 ± 0.072	0.439 ± 0.068
F_2	-0.884 ± 0.014	-0.703 ± 0.012	-0.084 ± 0.011
Cl_2	-0.966 ± 0.019	-0.834 ± 0.018	-0.370 ± 0.015

III. RESULTS AND DISCUSSION

Now that we have rigorously located and removed outliers from our initial lists of candidate binary compounds, we may now proceed to tabulating our numerical diatomic energy correction values. We graphically depict fitting results in Figure 2 (O_2), Figure 3 (H_2), Figure 4 (N_2), Figure 5 (F_2), and Figure 6 (Cl_2). In each plot, the solid line corresponds to a least-squares fit to Eqn. 2. Applying the diatomic energy correction C, indicated in blue, would improve agreement between our DFT calculations and experimental values. The dashed line indicates y = x, or perfect agreement.

First, we can see that our hypothesis is accurate, as the data appears to lie on lines of slope = 1, and the universally high R^2 values for the fits lend even more confidence to our approach. The oxides, hydrides, and nitrides (Figures 2-4) all follow a similar pattern, such that the two GGA functionals predict the compounds to be less stable than experiment, while the LDA functional predicts the compounds to be more stable than experiment. This is consistent with the work by Wolverton *et al*[13] and by Hector *et al*[28], suggesting that LDA functionals predict hydrides to be more stable and GGA functionals predict them as less stable than experiment. However, the fluorides (Figure 5) and chlorides (Figure 6) do not follow this trend, as all three functionals predict fluorides and chlorides to be less stable than experiment.

Table I summarizes the primary results of this paper: correction factors for each of the fifteen molecule-XC functional pairs, reported as 95% confidence intervals. If a fit set (after outlier removal) includes n compounds, we perform n fits to each possible subset of $n-1$ compounds, and use the n resulting C values to calculate a 95% confidence interval around their average value. The confidence intervals quantify exactly how much variation remains in C based on our choice of compounds for the fitting data set after the most extreme outliers have been removed. Overall, these are on the order of ±10 meV/X_2, a quite high standard of precision. However, the N_2 molecule shows a higher magnitude of variation, approximately ±70 meV/N_2. This could be due to inconsistencies in the DFT treatment of nitrides or experimental uncertainties. Nonetheless, the N_2 confidence intervals are of a much smaller magnitude than the correction factors themselves, indicating that they, like the other diatomic energy corrections, are still robust parameters for improving thermodynamic agreement between DFT and experiment.

We make a few comments comparing our work to that of Wang *et al*, who calculated $C(O_2, GGA-PBE)$. Numerically, they obtain a value of -1.36 eV, whereas our result in Table I is -1.20 eV. We determined that roughly half of the discrepancy in these quantities can be attributed to alternate choices of Ca and Mg PAW potentials, and slightly different experimental formation energies used in the fits. The other ≈80 meV/O_2 difference arises from our expanded oxide fit set as compared to the smaller selection of oxides in Wang *et al*, indicating that our value may be applicable to more diverse oxide chemistries.

Finally, it is worth noting the differences in behavior between the PBE-GGA and PW91-GGA functionals. The two GGA's may give qualitatively similar trends, but for these molecular systems and associated solids at least, their quantitative results differ significantly. Thus, despite the similarity between the two functionals, the choice of the proper functional for this application is non-trivial.

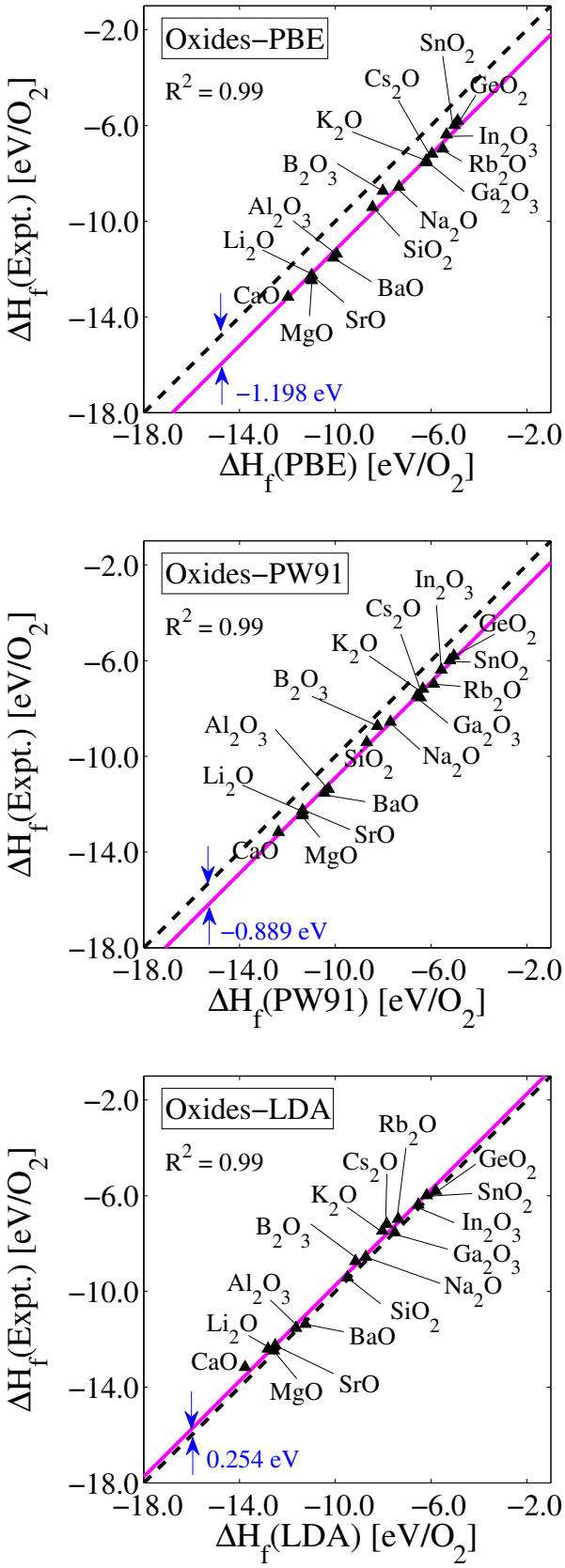

Figure 2: Plots for fitting diatomic energy corrections C for the O_2 molecule across the three exchange-correlation potentials, as indicated. The solid line represents a least-squares fit to Eqn. 2. Applying the diatomic energy correction, C in Eqn. 2, shown in blue (color online) would improve agreement between the DFT-calculated and experimental formation enthalpies.The dotted line represents y = x, or perfect agreement between DFT calculations and experimental values.

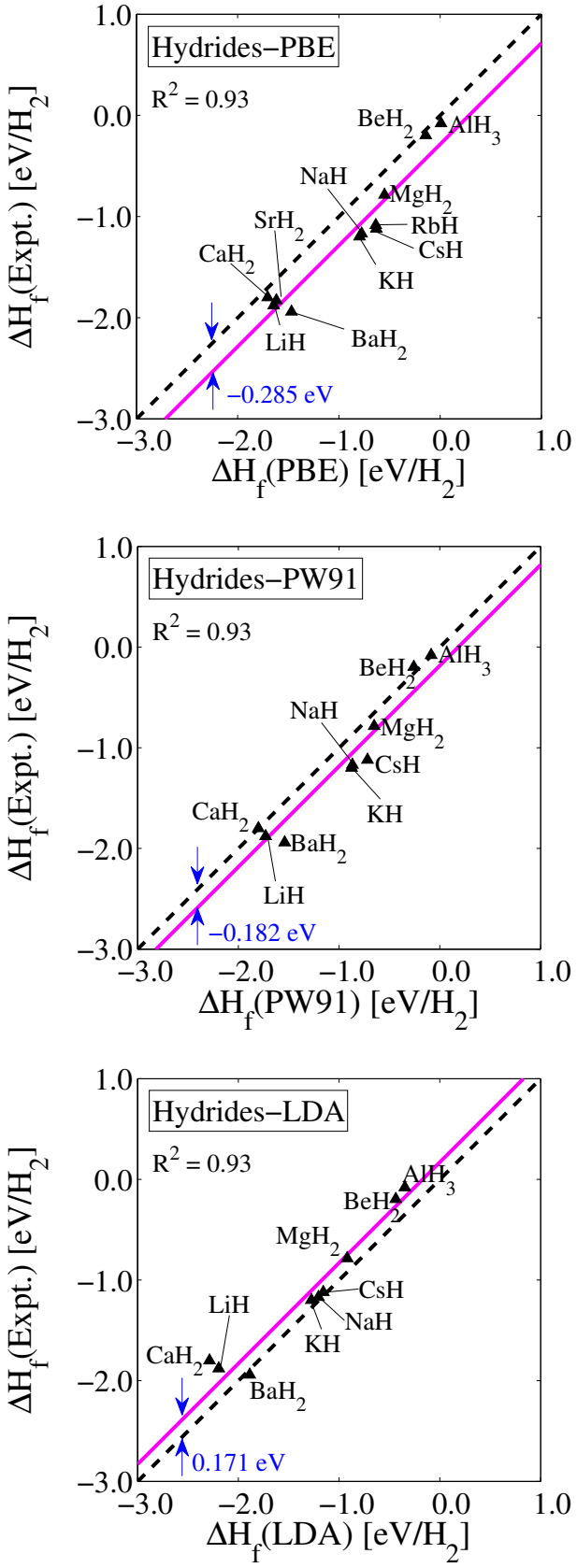

Figure 3: The diatomic energy corrections for the H_2 molecule across the three XC functionals. Contrasting the behavior between the two GGA functionals and the LDA functional is consistent with the results of Wolverton *et al*[13] and Hector *et al*[28].

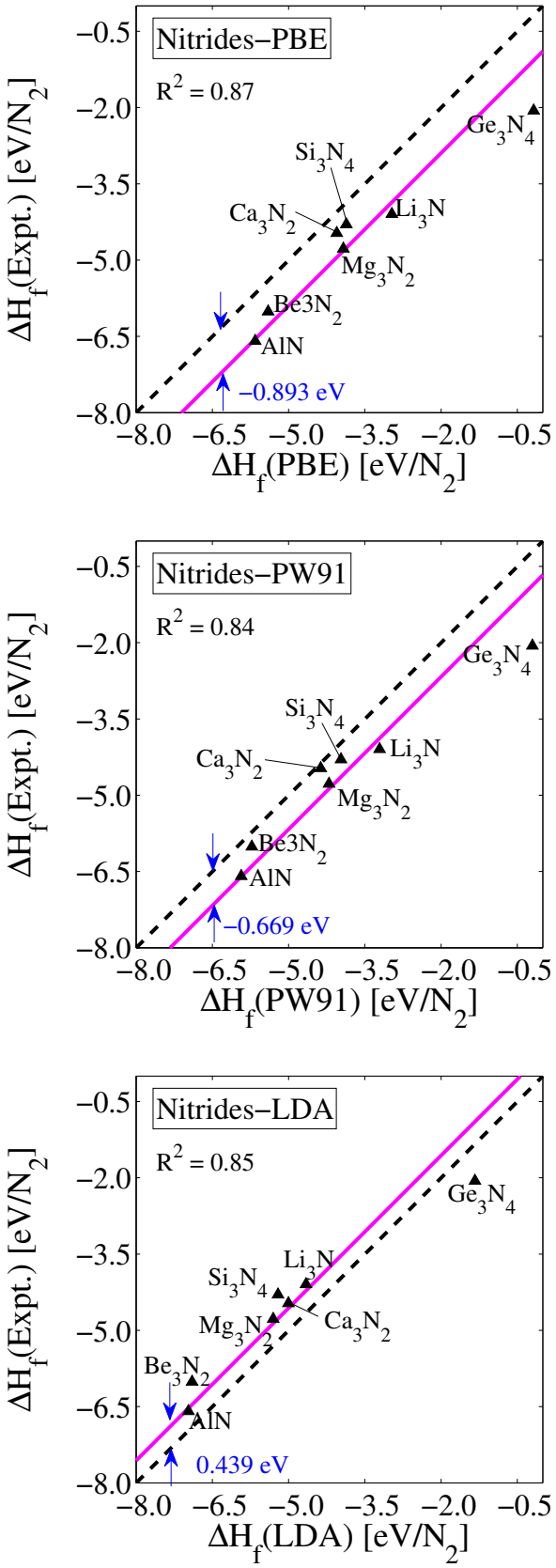

Figure 4: The diatomic energy corrections for the N_2 molecule across the three XC functionals. These figures show that the diatomic energy correction procedure may be less accurate for nitrides, which is also reflected in the 95% confidence intervals reported in Table I.

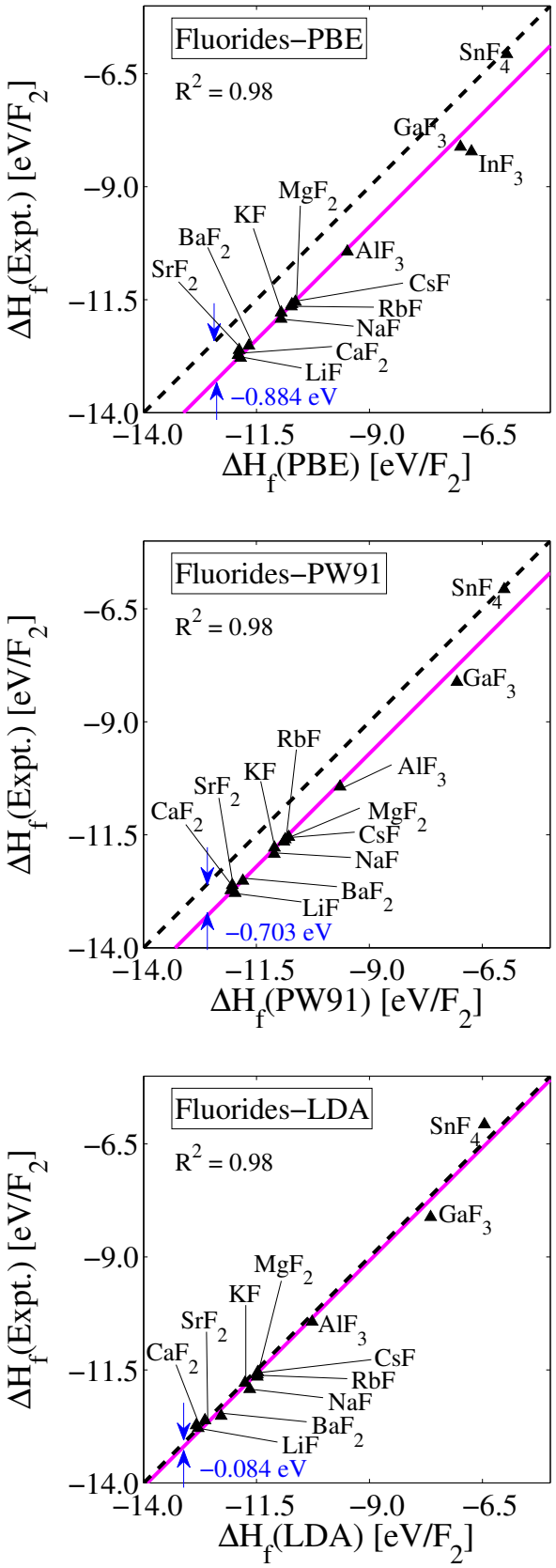

Figure 5: The diatomic energy corrections for the F_2 molecule across the three XC functionals. All three XC functionals predict fluorides to be less stable than experiment, in contrast to the predicted stability of oxides, hydrides.

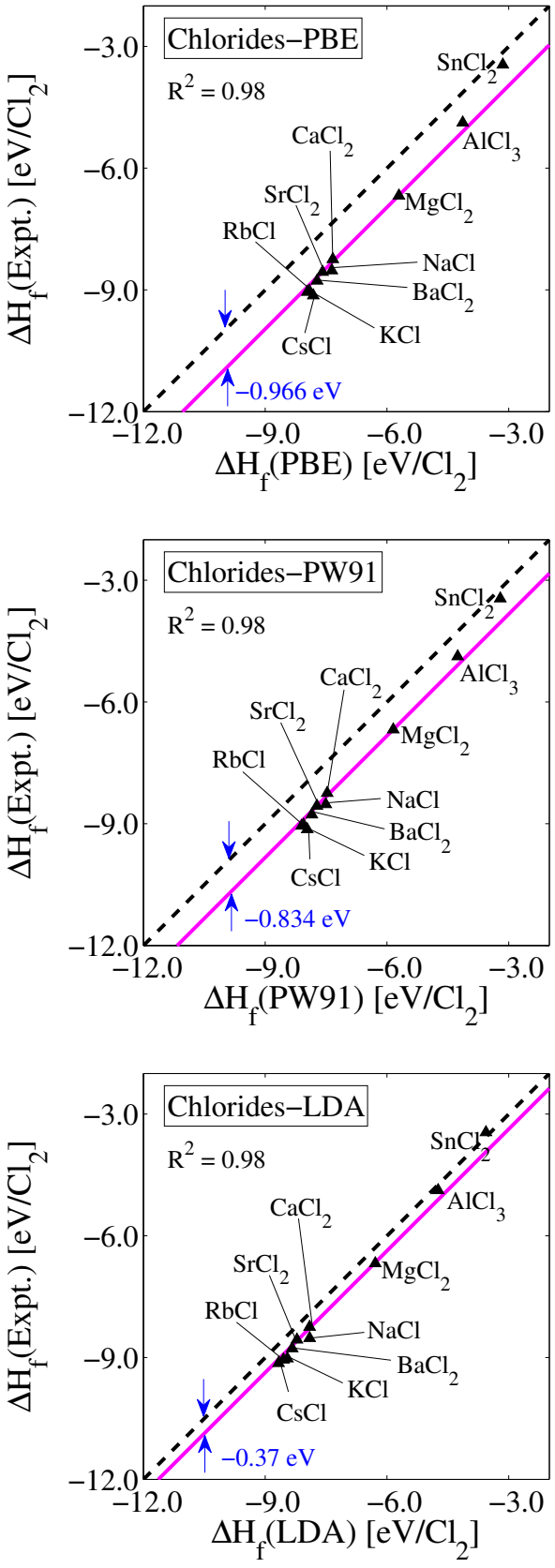

Figure 6: The diatomic energy corrections for the Cl_2 molecule across the three XC functionals. Similar to the fluorides, all three XC functionals predict chlorides to be less stable than experiment.

Table II: Comparison of the variation in formation enthalpies across different sets of atomic potentials. δ_i represents the difference between the DFT-calculated and experimental formation enthalpies for the indicated oxide; the difference between δ's gives a first-order approximation to how C from Eqn. 2 would vary from that reported in Table I for the indicated potential. All units are in eV/O_2. (a) Comparison between the O and O_h oxygen potentials using PBE-GGA. The right-hand column indicates that using the O_h potential instead of the O potential will change C by approximately 160 meV. (b) Comparison between PAW potentials and ultra-soft pseudopotentials (USPP) using the PW91-GGA XC functional. Here we have used the O (instead of O_h) potential as in the rest of this report. The right-hand column indicates that using USPP's would change C by approximately 200-250 meV.

	δ_O	δ_{O_h}	$\delta_O - \delta_{O\_h}$
Al_2O_3	1.441	1.601	-0.160
CaO	1.186	1.350	-0.164
Na_2O	1.217	1.384	-0.167

(a)

	$\delta_{PW91-PAW}$	$\delta_{PW91-USPP}$	$\delta_{PW91-PAW} - \delta_{PW91-USPP}$
Al_2O_3	1.068	1.319	-0.251
CaO	0.788	1.004	-0.216

(b)

IV. COMPARISON TO OTHER ATOMIC POTENTIALS

In order to assess the transferability of our approach to other sets of atomic potentials, we calculated the formation enthalpies of a few oxides using different oxygen potentials than were used in the rest of this work (see the list of those used in Table IV). In Table II, we investigate two specific potentials: (a) the O_h PBE-GGA potential, which we compare to the O PBE-GGA potential used above; (b) ultra-soft pseudopotentials (USPP) within the PW91-GGA XC functional, which we compare to the projector-augmented wave (PAW) method within the PW91-GGA XC functional used above. Table IIa indicates that C from Eqn. 2 would change by approximately 160 meV/O_2 were we to use the O_h potential, and Table IIb indicates that C from Eqn. 2 would change by approximately 200-250 meV/O_2 were we to use the USPP method. The magnitude of these deviations is non-trivial, and this serves to strongly reinforce the point that our correction factors are not independent of the choice of pseudopotential. We also direct the reader to the work of Lee *et al*[17], which made a comparison between the O PBE-GGA and O_s PW91-GGA PAW potentials, finding a similar result.

V. CONCLUSION

In summary, we have fit correction factors for our set of five molecules and three exchange-correlation functionals, applying a leave one out cross-validation approach to rigorously determine the precision of our fits. Our tabulated diatomic energy correction values greatly improve the quantitative agreement between DFT and experimental thermodynamics for reactions involving the common diatomic molecules O_2, H_2, N_2, F_2, and Cl_2. Our statistical approach uses 95% confidence intervals to indicate the precision of our fitted correction factors, and these confidence intervals are on the order of tens of meV. Thus, we show that - at least for diatomic molecules and some of their corresponding binary compounds - chemical accuracy is a realistic goal. However, we also showed here that our approach will not enable chemical accuracy for certain other compounds. For example, the outliers we identified in Figure 1—due to experimental uncertainty, pathological DFT shortcomings, or both—fall outside our diatomic energy correction framework.

Our tabulated diatomic energy corrections (Table I) should be applied in any thermodynamic calculations involving the five diatomic molecules considered here and our set of widely-used PAW potentials (Table IV). While the values of

C we determined will, of course, be most accurate for our set of fit compounds and any chemically similar materials, our obtained C values are so large compared to typical reaction energies that their use will surely increase the accuracy of thermodynamic predictions when applied to novel compounds and reactions.

VI. ACKNOWLEDGMENTS

Bryce Meredig was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program, and by DOE under Grant No. DE-FG02-07ER46433. Scott Kirklin was supported by the Center for Electrical Energy Storage: Tailored Interfaces, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science and Office of Basic Sciences. James Saal was supported by the US Department of Energy, Office of Basic Energy Sciences (Dr. John Vetrano, monitor) through grant DE-FG02-98ER45721.

APPENDIX

Table III: Binary compounds and formation energies used to fit the diatomic energy corrections in Table I; the values resulted from a least-squares fit to Eqn. 2 based on the energies in this table. All experimental values were taken from the SSUB database within the ThermoCalc software package[29] at T = 298K. While DFT-calculated formation enthalpies are at T = 0K, the difference between 0K and 298K formation enthalpies is minimal compared to the formation enthalpies themselves, follwing the procedure of[14]. All units are given in eV/X_2 molecule.

Compound	Space Group	ΔH^{PBE}	ΔH^{PW91}	ΔH^{LDA}	$\Delta H^{Expt.}$
Al_2O_3	R $\bar{3}$ c H	-10.085	-10.458	-11.642	-11.526
B_2O_3	P 31	-8.015	-8.239	-9.156	-8.735
BaO	F m $\bar{3}$ m	-9.946	-10.291	-11.253	-11.359
CaO	F m $\bar{3}$ m	-11.975	-12.373	-13.772	-13.161
Cs_2O	R $\bar{3}$ m H	-5.965	-6.346	-7.849	-7.180
Ga_2O_3	C 1 2/m 1	-6.217	-6.443	-7.513	-7.538
GeO_2	P 31 2 1	-4.890	-5.045	-5.803	-5.797
In_2O_3	I a $\bar{3}$	-5.357	-5.571	-6.541	-6.377
K_2O	F m $\bar{3}$ m	-6.205	-6.580	-8.034	-7.460
Li_2O	F m $\bar{3}$ m	-11.049	-11.365	-12.807	-12.393
MgO	F m $\bar{3}$ m	-10.986	-11.379	-12.571	-12.470
Na_2O	F m $\bar{3}$ m	-7.341	-7.696	-8.727	-8.557
Rb_2O	F m $\bar{3}$ m	-5.515	-5.889	-7.377	-6.969
SiO_2	P 41 21 2	-8.446	-8.686	-9.496	-9.409
SnO_2	P 42/m n m	-5.024	-5.221	-6.185	-5.987
SrO	F m $\bar{3}$ m	-10.983	-11.364	-12.511	-12.251
AlH_3	R $\bar{3}$ 2/c	0.008	-0.087	-0.347	-0.079
BaH_2	P n m a	-1.468	-1.540	-1.887	-1.940
BeH_2	I b a m	-0.142	-0.258	-0.437	-0.197
CaH_2	P n m a	-1.704	-1.801	-2.285	-1.800
CsH	F m $\bar{3}$ m	-0.631	-0.718	-1.156	-1.120
KH	F m $\bar{3}$ m	-0.796	-0.875	-1.274	-1.199
LiH	F m $\bar{3}$ m	-1.645	-1.729	-2.194	-1.879
MgH_2	P 42/m n m	-0.550	-0.652	-0.919	-0.785
NaH	F m $\bar{3}$ m	-0.776	-0.867	-1.205	-1.169
RbH	F m $\bar{3}$ m	-0.636	-	-	-1.084
SrH_2	P n m a	-1.618	-	-	-1.828
AlN	P 63 m cï¿œ	-5.657	-5.932	-6.969	-6.591
Be_3N_2	I a $\bar{3}$	-5.404	-5.723	-6.900	-6.011
Ca_3N_2	I a $\bar{3}$	-4.053	-4.368	-5.003	-4.467
Ge_3N_4	P 63/m	-0.178	-0.204	-1.334	-2.059
Li_3N	P 6/m m m	-2.968	-3.211	-4.657	-4.094
Mg_3N_2	I a $\bar{3}$	-3.923	-4.204	-5.301	-4.777
Si_3N_4	P 63	-3.864	-3.969	-5.207	-4.295
AlF_3	R $\bar{3}$ c	-9.492	-9.651	-10.270	-10.436
BaF_2	F m $\bar{3}$ m	-11.671	-11.802	-12.283	-12.519
CaF_2	F m $\bar{3}$ m	-11.909	-12.065	-12.836	-12.727
CsF	F m $\bar{3}$ m	-10.641	-10.784	-11.473	-11.548
GaF_3	R -3 c H	-6.987	-7.063	-7.650	-8.119
InF_3	R -3 c H	-6.744	-	-	-8.222
KF	F m $\bar{3}$ m	-10.956	-11.103	-11.757	-11.786
LiF	F m $\bar{3}$ m	-11.853	-11.971	-12.791	-12.788
MgF_2	P 42/m n mï¿œ	-10.731	-10.894	-11.486	-11.652
NaF	F m $\bar{3}$ m	-10.960	-11.106	-11.650	-11.927
RbF	F m $\bar{3}$ m	-10.718	-10.863	-11.483	-11.602
SnF_4	I 4/m m m	-5.962	-6.017	-6.456	-6.068
SrF_2	F m $\bar{3}$ m	-11.886	-12.035	-12.641	-12.615
$AlCl_3$	C 1 2/m 1	-4.132	-4.257	-4.738	-4.876
$BaCl_2$	P n a mï¿œ	-7.721	-7.847	-8.333	-8.773
$CaCl_2$	P n n mï¿œ	-7.333	-7.471	-7.905	-8.244
CsCl	P m -3 m	-7.813	-7.964	-8.670	-9.138
KCl	F m $\bar{3}$ m	-7.958	-8.099	-8.549	-9.052
$MgCl_2$	R $\bar{3}$ m	-5.703	-5.842	-6.285	-6.678
NaCl	F m $\bar{3}$ m	-7.355	-7.504	-7.904	-8.522
RbCl	F m $\bar{3}$ m	-7.911	-8.052	-8.453	-9.021
$SnCl_2$	P n a m	-3.146	-3.207	-3.559	-3.451
$SrCl_2$	F m $\bar{3}$ m	-7.333	-7.725	-8.217	-8.559

Table IV: PAW potentials used in all density functional calculations, as used with the VASP software package. The PAW potentials used are those recommended by G. Kresse[30].

Species	Atomic Potential Name
Al	Al
B	B
Ba	Ba_sv
Be	Be
Ca	Ca_pv
Cl	Cl
Cs	Cs_sv
F	F
Ga	Ga_d
Ge	Ge_d
H	H
In	In_d
K	K_sv
Li	Li_sv
Mg	Mg
N	N
Na	Na_pv
O	O
Pb	Pb_d
Rb	Rb_sv
Si	Si
Sn	Sn_d
Sr	Sr_sv
Tl	Tl_d

[1] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
[2] H. Smithson, C. A. Marianetti, D. Morgan, A. Van der Ven, A. Predith, and G. Ceder, Phys. Rev. B **66**, 144107 (2002).
[3] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, The Journal of Chemical Physics **106**, 1063 (1997), URL http://link.aip.org/link/?JCP/106/1063/1.
[4] D. C. Patton, D. V. Porezag, and M. R. Pederson, Phys. Rev. B **55**, 7454 (1997).
[5] C. Wolverton and V. Ozoliņš, Phys. Rev. B **75**, 064101 (2007), URL http://link.aps.org/doi/10.1103/PhysRevB.75.064101.
[6] K. L. Howard and D. J. Willock, Faraday Discuss. **152**, 135 (2011), URL http://dx.doi.org/10.1039/C1FD00026H.
[7] Y. A. Mastrikov, R. Merkle, E. Heifets, E. A. Kotomin, and J. Maier, The Journal of Physical Chemistry C **114**, 3017 (2010), http://pubs.acs.org/doi/pdf/10.1021/jp909401g, URL http://pubs.acs.org/doi/abs/10.1021/jp909401g.
[8] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jónsson, The Journal of Physical Chemistry B **108**, 17886 (2004), http://pubs.acs.org/doi/pdf/10.1021/jp047349j, URL http://pubs.acs.org/doi/abs/10.1021/jp047349j.
[9] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996), URL http://link.aps.org/doi/10.1103/PhysRevLett.77.3865.
[10] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992), URL http://link.aps.org/doi/10.1103/PhysRevB.46.6671.
[11] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **48**, 4978 (1993), URL http://link.aps.org/doi/10.1103/PhysRevB.48.4978.2.
[12] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981), URL http://link.aps.org/doi/10.1103/PhysRevB.23.5048.
[13] C. Wolverton, V. Ozoliņš, and M. Asta, Phys. Rev. B **69**, 144109 (2004).
[14] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B **73**, 195107 (2006).
[15] V. Stevanović, S. Lany, X. Zhang, and A. Zunger, Phys. Rev. B **85**, 115104 (2012), URL http://link.aps.org/doi/10.1103/PhysRevB.85.115104.
[16] A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson, and G. Ceder, Phys. Rev. B **84**, 045115 (2011), URL http://link.aps.org/doi/10.1103/PhysRevB.84.045115.
[17] Y.-L. Lee, J. Kleis, J. Rossmeisl, and D. Morgan, Phys. Rev. B **80**, 224101 (2009), URL http://link.aps.org/doi/10.1103/PhysRevB.80.224101.
[18] C. Wolverton and V. Ozoliņš, Phys. Rev. B **73**, 144104 (2006), URL http://link.aps.org/doi/10.1103/PhysRevB.73.144104.
[19] G. Hautier, S. P. Ong, A. Jain, C. J. Moore, and G. Ceder, Phys. Rev. B **85**, 155208 (2012), URL http://link.aps.org/doi/10.1103/PhysRevB.85.155208.
[20] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994), URL http://link.aps.org/doi/10.1103/PhysRevB.50.17953.
[21] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999), URL http://link.aps.org/doi/10.1103/PhysRevB.59.1758.
[22] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993), URL http://link.aps.org/doi/10.1103/PhysRevB.47.558.
[23] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994), URL http://link.aps.org/doi/10.1103/PhysRevB.49.14251.
[24] G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996), ISSN 0927-0256, URL http://www.sciencedirect.com/science/article/pii/0927025696000080.
[25] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996), URL http://link.aps.org/doi/10.1103/PhysRevB.54.11169.
[26] M. Stone, Journal of the Royal Statistical Society. Series B (Methodological) **36**, pp. 111 (1974), ISSN 00359246, URL http://www.jstor.org/stable/2984809.
[27] C. Ambroise and G. J. McLachlan, Proceedings of the National Academy of Sciences **99**, 6562 (2002), http://www.pnas.org/content/99/10/6562.full.pdf+html, URL http://www.pnas.org/content/99/10/6562.abstract.
[28] L. G. H. Jr and J. F. Herbst, Journal of Physics: Condensed Matter **20**, 064229 (2008), URL http://stacks.iop.org/0953-8984/20/i=6/a=064229.
[29] J.-O. Andersson, T. Helander, L. HÃ¶glund, P. Shi, and B. Sundman, Calphad **26**, 273 (2002), ISSN 0364-5916, URL http://www.sciencedirect.com/science/article/pii/S0364591602000378.
[30] G. Kresse (2003), hands-on tutorial course on VASP, URL http://www.vasp.at/vasp-workshop/slides/pseudoppdatabase.pdf.