

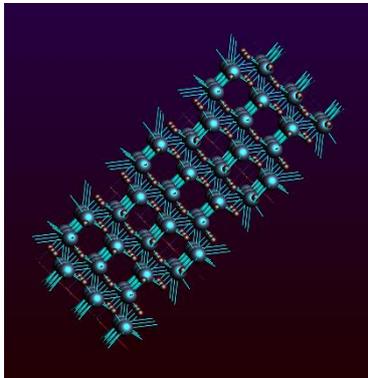
Benchmark of XC functionals for electronic calculations of periodic solids using BAND.

Internship Report

31/August/2017

Intern/Student:

Jetsabel Maria Figueroa-Tapia (TU-Delft)



Supervisors:

Msc. Mirko Franchini (SCM)

Stan van Gisbergen (SCM)

Dr. Ferdinand Grozema (TU-Delft)

Abstract

The goal of this project is to prove the ability of different Exchange Correlation (XC) Functionals to estimate the bandgap of crystalline inorganic semiconductors. For this we perform simulations in a total of 59 systems for which the bandgap is known. We used a selection of XC functionals which included: LDA, GGAs (PBE, PW91 and BLYP), Model Potentials (LB94, GLLB-SC and mBJ's) and one range separated hybrid (HSE06). All the simulations were done using BAND (from SCM). First arrived to a recipe in which we converged the basis set and K-space and empirically deduced which materials presented strong spin-orbit coupling effects (SOC). From the comparison of our results we concluded that (1) the different GGAs give very similar values and all of them outperform LDA in bandgap calculations. (2) Out of the different potentials the best results are obtained by TB-mBJ followed by GLLB-SC. (3) The HSE06 hybrid gives the best results in terms of numerical accuracy. However, this accuracy comes with a substantial computational cost. (4) the improvement in accuracy between HSE06 and TB-mBJ does not compensates for the large difference in computational time for which in general we would recommend to use TB-mBJ as the best balance between accuracy and computational time.

Contents

Abstract	2
1 Description of the problem	4
1.1 About SCM and BAND	4
1.2 Project goals:	5
2 Methodology	6
2.1 Choice of Exchange Correlation (XC) Functionals.....	6
2.2 Materials.....	6
2.2.1 Material sub-groups.	6
2.3 Basic Settings	7
2.4 Input Geometries.....	7
2.5 Statistical Analysis	9
2.6 Comments on experimental Data	10
3 Results	11
3.1 K-space and Basis Set convergence.....	11
3.1.1 Basis Set Convergence.....	11
3.1.2 K-space convergence.....	11
3.2 Reproducibility of PBE results.	13
3.3 Spin orbit contribution.	14
3.4 Comparison of XC functionals	15
3.4.1 LDA and GGA's.....	15
3.4.2 Model Potentials with SOC: mBJ's and LB94	16
3.4.3 Model Potential: GLLB-SC.....	17
3.5 Comparisons with HSE06.....	18
4 Summary and conclusions	20
A. Scripts	22
B. Geometric input.....	25
C. Tables of results.....	26
Table C1: Basic Set Convergence.....	26
Table C2: K-space Convergence.....	27
Table C3: PBE Results (CRYSTAL98 vs. BAND)	28
TABLE C4: BLYP and PW91 results.....	30
Table C5: LDA and LB94 results	31
Table C6: GLLB-SC results	33

1 Description of the problem

Density Functional Theory (DFT) is a computational tool based on quantum mechanics that is used to estimate the electronic distribution of a many-body system. Based on these calculations, we can estimate physical and chemical properties such as reaction energies, optical absorption, charge distribution, polarization, band structure, etc.

In the realm of solid-state physics, one of the main characteristics that DFT allows us to estimate is the distribution of the electrons in a crystalline solid, or electronic band structure. Understanding this band structure is one of the first steps to understand and predict the behavior of solid-state devices such as transistors, LEDs and solar cells. One of the key values derived from this band structure is the energy gap (or bandgap) between the conduction band and the valence band. This property determines whether a material is a metal, semiconductor or insulator and is therefore one of the first properties investigated for a given material.

Due to the importance of band structure and energy gap, a fair effort is being put into designing DFT functionals that can accurately predict these properties. One of the most basic functionals, the one based on the Perdew–Burke–Ernzerhof exchange model (PBE) is commonly used as a starting point for band gap calculations. However, as will be shown later in this report, there are other functionals that easily outcompete the accuracy obtained by PBE without resulting in a massive increase in computational time. Nevertheless, there is still a controversy as to which functional would be a better replacement for PBE. In order to answer this question some authors have performed comprehensive calculations spanning a wide variety of solids. Some of them, as in the case of Crowley and Goddard [1] have gone as far as declaring an absolute winner. According to them, the hybrid B3PW91 functional is capable of outperforming both the accuracy and computational time of much more strict calculations such as those based on the green function method (GW).

It is of interest for SCM, to test the validity of Goddard and Crowley' statement and therefore we decided to prepare simulations using different functionals that could give comparable results to those presented by these authors. For this, we choose some functionals already implemented in SCM' software, such as the GGAs: LB94, BLYP, PW91; those based on model potentials: GLLB-SC, TB-mBJ, KTB-mBJ, JTS-mTB-mBJ and one range-separated: HSE06. We performed all the calculations using SCM software for periodic materials: BAND and compare our results to those obtained by the Goddard and Crowley using CRYSTAL98.

1.1 About SCM and BAND

In view of the global interest for DFT is that SCM created its commercial software: Amsterdam Density Functional (ADF). The original software has been extended since the decade of 1970 and nowadays it includes a couple different software instances tailored for different types of systems and/or analysis. One of such instances is the Amsterdam Density Functional Band-structure program, BAND. This software uses numerical orbitals (NOs) and Slater-type orbitals to perform electronic calculations on periodic materials in 1D (polymers), 2D (slabs) or 3D (crystals).

1.2 Project goals:

To test relevant XC functionals to estimate the band gap of a large dataset of periodic materials, in this case all are inorganic semiconductors or insulators. Based on these calculations we expect to answer the following questions:

1. According to different statistical metrics, which functional(s) seems to be doing a better job at estimating the bandgap?
2. Which XC functional(s) results in a better balance between computational time and accuracy?
3. How does our results compare to similar studies reported in scientific literature?

On top of this, we would like to comment on the performance of BAND for the above-mentioned calculations.

2 Methodology

All the simulations were performed using BAND in three dimensional periodic systems (crystals). Based on PBE calculations we first set the values of the most basic settings: k-space, numerical quality, frozen cores, basis set and relativistic effects. Then, we tested the ability of our software, BAND, to reproduce the calculations obtained by Crowley & Goddard using CRYSTAL98. Once these preliminary steps were completed we proceed to perform calculations using a selection of XC functionals and evaluated their results.

2.1 Choice of Exchange Correlation (XC) Functionals

We did a comprehensive study including a wide selection of relevant functionals. This selection was of course limited to those functionals already implemented in BAND.

As a starting point, we run simulations using the most basic functional, based on the Local Density approximation (LDA). Then we proceed to evaluate the improvements brought up by different types of Generalized As mentioned before PBE is a common choice to calculate bandgap values, thus we wanted to know if there was a significative difference by using other type of Gradient Approximations (GGAs), for this we included PW91 and BLYP. A more empirical approach, the model potentials, has been proved to also give good bandgap results, given the low computational cost of these calculations we tested all the model potentials implemented in BAND, among which are GLLB-SC [2], LB94 [3] and the different types of modified Becke-Johnson potentials: TB-mBJ, kTB-mBJ and JTS-mTB-mBJ [4]. Finally, we included the final version of the HSE range-separated hybrid: HSE06 which also has been proven to give good bandgap results for a similar data-set that the one presented in this project [5].

2.2 Materials

In this project, we used the same crystals and geometries than Goddard and Crowley, their selection includes a wide variety of inorganic crystal formed by no more than three elements. Out of the 70 compounds that they used we made two exceptions and ended up with a list of 59 materials. We did not include the Bi_2Se_3 slabs or the so-called Mott insulators (NiO , MnO , FeO , CoO and VO_2). The first ones were eliminated due to the difficulty to reproduce the geometries. The second ones were eliminated because they resulted in underestimations and convergence errors in our calculations. These anomalies can be explained by the fact that most functionals cannot properly account for the strong correlations present in these materials which according to traditional band theory should be metals but are made insulators due to strong electron-electron repulsion effects. [6], [7]

2.2.1 Material sub-groups.

The materials used in this project can be classified in the following categories:

Group 4A	Diamond, Si, SiC polytypes: SiC-2H, SiC-3C (a.k.a. β -SiC), SiC-4H and SiC-6H (a.k.a α -SiC), and Ge.
Group 3A-5A	BN, BP, AlN (wurzite), AlP, AlAs, AlSb, GaN (wurzite & zincblende), GaP, GaAs, GaSb, InN, InP, InAs and InSb

Metal oxides	MgO, SiO ₂ (b-cristobalite and a-quartz), TiO ₂ (anatase & rutile), Cu ₂ O and ZnO.
Transition-metal chalcogenides	ZnS, ZnSe, ZnTe, CdS, CdSe, CdTe and MoS ₂ .
Transition-metal halides	AgCl, AgBr, AgI, CuCl, CuBr and CuI
Crystals containing Heavy Atoms (N>50)	SnSe, SnTe, PbSe, PbTe, Bi and BiVO ₄ .
Alkali halides	LiCl, LiF and NaCl
Topological insulators	HgTe, Bi ₂ Te ₃ , Sb ₂ Te ₃ and Bi ₂ Se ₃ .
Others	CuSCN and SrTiO ₃ .

2.3 Basic Settings

In this project, we performed only single-point calculations on experimental geometries. For this we used the BAND default settings and changed only the following values:

-The numerical quality of the calculation was set as follows:

-K-space and Basis set were converged as presented in section 3.1.

-Soft confinement was set to normal.

-Frozen cores were kept as “small” excepting in the case of HSE06 because the other options were disabled for this functional.

-The rest of the parameters were left in their default values. i.e. Integration = Becke Normal and Spline Zlm fit = normal

-All the calculations were spin “restricted”

-We perform calculations setting the relativistic contribution to either scalar or spin-orbit for those functionals that allowed it (LDA, GGAs and model potentials excepting GLLB-SC). The results were compared and from it, we decided on which systems seemed to be more affected by SOC.

2.4 Input Geometries

All the simulations shown in this project were run remotely using SCM cluster. In order to communicate with the cluster, we wrote python scripts using Python Library for Automating Molecular Simulations (PLAMS). The PLAMS script contains the atomic coordinates and all the settings required to run a job (or series of jobs) and to extract the relevant information, in our case the bandgap and computational time.

In order to ensure an easy comparison to their data, we choose to use the same materials and experimental coordinates that Crowley & Goddard had used in their paper. This meant that we needed to use exactly the same values that they did. The authors supplied the input files including

the geometries used in their simulations but these input files were given for another DFT program: CRYSTAL.

The input of CRYSTAL and that of BAND differ from each other. As shown in **Figure 1**, CRYSTAL's input includes the space group, lattice constants, angles and the basic atomic coordinates but some of this data, e.g. the atom type or angles, is codified in a not so explicit format. Furthermore, CRYSTAL performs additional calculations on the atomic coordinates using the entered space group.

```

PbSe

PbSe – Acta Crystallographica, Section B: Structural Science (1983) 39, p312-p317
crystal
0 0 0
225      <-- Space group
6.128    <-- Lattice constant & angles (if necessary)
2       <-- Number of atoms
Atom number
(Last 2 digits) 282 0.5 0.5 0.5      <--Atomic coordinates (fractional)
--> 234 0.00 0.00 0.00
printout
basisset
end
end
  
```

Figure 1. CRYSTAL' input for PbSe (only geometry section)

```

$cell_vectors
  6.1213000000      0.0000000000      0.0000000000
  0.0000000000      6.1213000000      0.0000000000
  0.0000000000      0.0000000000      6.1213000000
$coordinates
Pb      0.0000000000      0.0000000000      0.0000000000
Pb      0.0000000000      3.0606500000      3.0606500000
Pb      3.0606500000      0.0000000000      3.0606500000
Pb      3.0606500000      3.0606500000      0.0000000000
Se      3.0606500000      3.0606500000      3.0606500000
Se      3.0606500000      0.0000000000      0.0000000000
Se      0.0000000000      3.0606500000      0.0000000000
Se      0.0000000000      0.0000000000      3.0606500000
$end
  
```

Figure 2 BAND input for PbSe (only coordinates)

In contrast, as shown in **Figure 2**, the input of BAND contains all the information required using cell vectors and a larger number of atomic coordinates. BAND does not make any additional operation in these numbers. It is evident that the input required for both programs is quite different and requires a conversion step. We first tried to avoid converting the CRYSTAL inputs by directly downloading the CIF or XYZ coordinates from the database they cite, ICSD, or from the also well-known COD database. However, the data was missing or had some differences with the coordinates they cite. Therefore, we had to convert the CRYSTAL inputs ourselves. To do this we performed the following algorithm:

1. Copy the space group, lattice constants, angles and atomic coordinates into a hacked CIF file (see appendix B). The hacked CIF is a file containing only the very basic information required for a reader to generate the lattice.
2. Open the hacked CIF in the ADF-GUI and save the lattice as XYZ coordinates. The generated XYZ has a structure that differs from that required by PLAMS.
3. Use a script (see appendix A) to reorganize the XYZ files.

The final XYZ files were then loaded into the server and used to run the simulations with different BS. All these files together with the script used to perform the conversion and to run the simulations in the cluster are included in appendix A.

2.5 Statistical Analysis

The accuracy of the calculations can be assessed by a large number of statistical parameter. For example, in the paper of Garza & Scuseria paper [5] they compare hybrid functionals using 10 different statistical metrics. For the purpose of this report we selected only 4 of them: 3 different types of mean errors (ME, MAE and MAPE) and the Kendall correlation (τ). The three types of mean errors serve to quantify how close to the real values our calculations are. Together they give a picture of the type of deviations that are taking place. They are defined as follows.

Mean Error (ME)

$$ME = \frac{1}{n} \sum_{i=1}^n y_i - x_i$$

Mean Absolute Error (MAE)

$$MEA = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{y_i} \right|$$

Where y_i is the experimental bandgap, x_i is the calculated bandgap and n is the number of systems tested.

In addition to numerical values, we are interested in knowing if the trend in the calculated values follows the trend of the experimental one. This is achieved by calculating the kendall correlation (τ). In order to calculate τ we have to compare pairs of experimental data ($y_1, y_2, y_3, y_4, \dots, y_i$) with pairs of simulated ones ($x_1, x_2, x_3, x_4, \dots, x_i$). In doing so we say that a pair is concordant if $y_i < y_j$ and $x_i < x_j$ or if $y_i > y_j$ and $x_i > x_j$, tied if $x_i = x_j$ and $y_i = y_j$ and discordant otherwise. The Kendall correlation is then calculated as follows:

Kendall correlation (τ)

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T)(P + Q - T)}}$$

Where P is the number of concordant pairs, Q is the number of discordant pairs and T and U are the number of ties in the experimental and calculated pairs respectively [5].

It is important to highlight that it is inaccurate to draw conclusions from only one of these statistical tools because they are all biased towards a type of deviation. For example, ME can be very low if the different errors cancel each other, MAE has to be interpreted based on the magnitude of the values treated and MAPE is especially sensitive to low bandgap materials.

2.6 Comments on experimental Data

Lastly, we would like to remark that we used the experimental data that Crowley and Goddard [1]. This allows us easy comparison not only with their results but with other studies using very similar sets of systems [8], [5]. However, we must not forget that the data they provided cannot be regarded as an ultimate value. We must keep this in mind while interpreting the numerical results obtained.

3 Results

The goal of the project is to compare the accuracy of the bandgap calculations obtained using different functionals, but before doing this we needed to fix the remaining BAND settings. Out of all the options available, the most important was converging the k-space and basis set and deciding for which molecules did the spin-orbit coupling represented an important contribution. The simulations run to decide on these settings are presented in the next two sections. Then we moved on to compare the results obtained for different functionals.

3.1 K-space and Basis Set convergence

The numerical quality of the calculations depends on the choice of: integration accuracy, Spline Zlm fit, k-space, confinement, choice of basis set and frozen cores. Out of these parameters, it seems that the k-space and basis set had a larger impact in the technical accuracy and computational time of the results. Here we showed the results that allowed to fix the values for these two parameters using PBE calculations. The remaining options within numerical quality were set as mentioned in the methodology section.

Why converging the K-space and basis set?

In BAND, the Brillouin zone is sampled using a grid. The accuracy of the sampling depends on the density of points in this grid which is controlled by the k-space setting. We ran PBE simulations keeping the rest of the parameters constant and changing the k-space between good, very good and excellent. On the other hand, the need to use an adequate size of basis set is recognized for all DFT calculations despite the type of orbitals used. In order to have a good approximation to the complete basis set (CBS) limit but to minimize the computational time we ran the PBE calculations using DZ, TZP, TZ2P and QZ2P.

3.1.1 Basis Set Convergence

Figure 3 shows the results the result of the PBE calculations changing only the basis set and keeping the remaining parameters constant (FC=small, NQ=good and relativity=scalar). From this figure, it is clear that the results of DZ deviated from the values obtained for the other basis sets. In the other extreme, QZ4P had so many orbitals that it crashed for the heaviest atoms: Pb and Bi. This reduced our choice of the functional to further calculations to either TZP or TZ2P. The values of the bandgap calculated with these two functionals differ by a mean average value of 0.017 eV. Given that the uncertainty in the experimental data itself can be larger than this value, we decided to settle in for **TZP** because it is computationally cheaper.

3.1.2 K-space convergence

Once the basis set was fixed as TZP, we attempt to minimize the computational time by using the lowest K-space possible. The results of this step are shown in **Figure 4**. Analogously to the basis set example, we now observe that a normal K-space induces large deviation. Once again, we choose to fixed the setting at the minimal value for which the trend is conserved, this is **Kspace=Good**. In doing this we note that the mean absolute deviation between good K-space and an excellent one is only 0.021 eV. Once again, reasonable considering the uncertainty of experimental data. See Tables C1 and C2 for the numerical results of the both the basic set and k-space convergence.

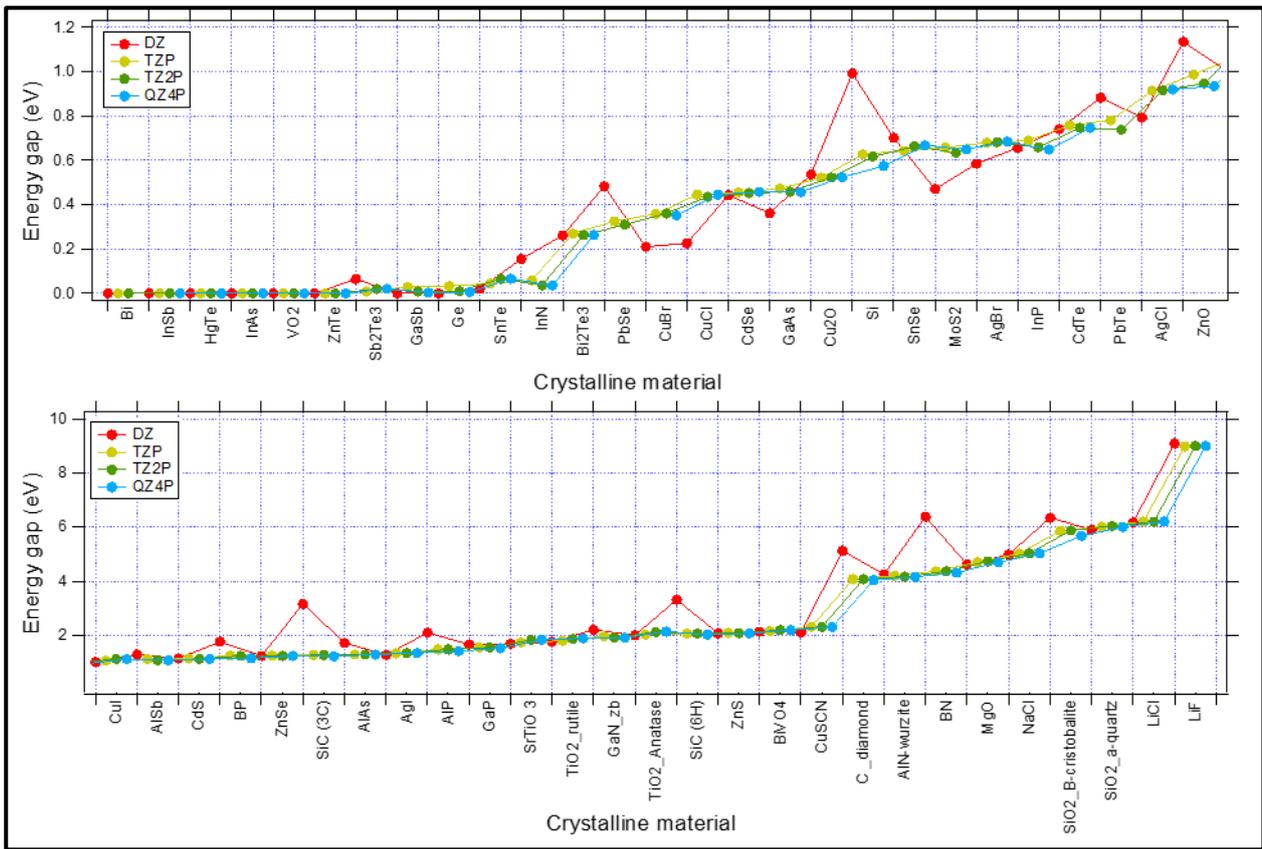


Figure 3 Simulations using different basis sets.

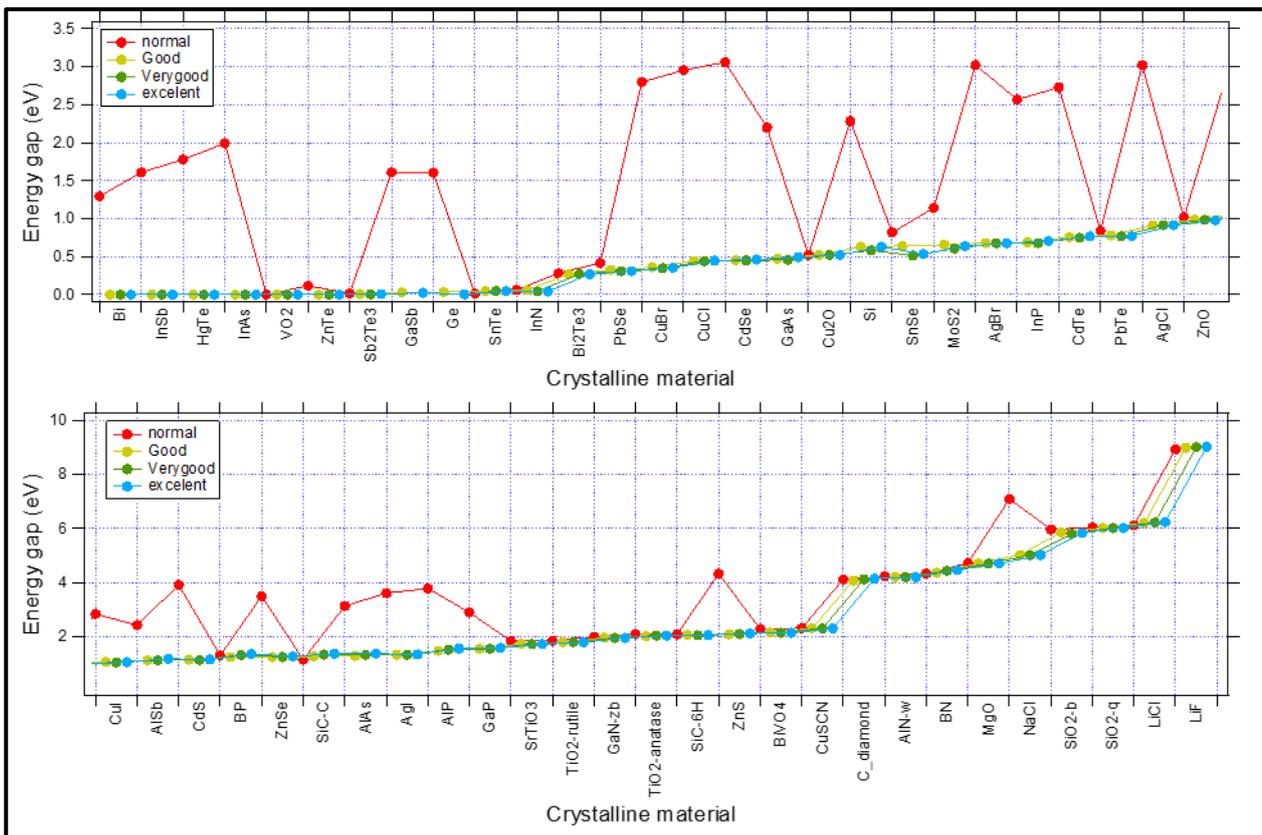


Figure 4. Simulations using different k-space.

3.2 Reproducibility of PBE results.

We compared our PBE results to those found calculated in ref. [1] using CRYSTAL98. Since we used the same input geometries, the only technical difference was that the authors of [1] used a “basis set recipe” to remove linear dependency errors. On contrast, we set TZP as the basis set for our BAND calculations.

A comparison between both PBE calculations is shown in **Figure 5**, and evaluated using the statistical metrics shown in Table 1. In general, we observe that the BAND-PBE results follow the same trend that the CRYSTAL-PBE ones. The changes observed are attributed to the basis set recipe used by our reference and possibly to differences within the working machinery of both programs.

Table 1 Statistical comparison of PBE results.

	Exp.	Crowley & Goddard (CRYSTAL98)	Our Calculation (BAND)
Mean absolute Error (MAE)	0	1.21	1.39
Mean error (ME)	0	1.19	1.40
Mean absolute Percental Error (MAPE)	0	53.3	52.8
Kendal coefficient (τ)	1	0.76	0.74

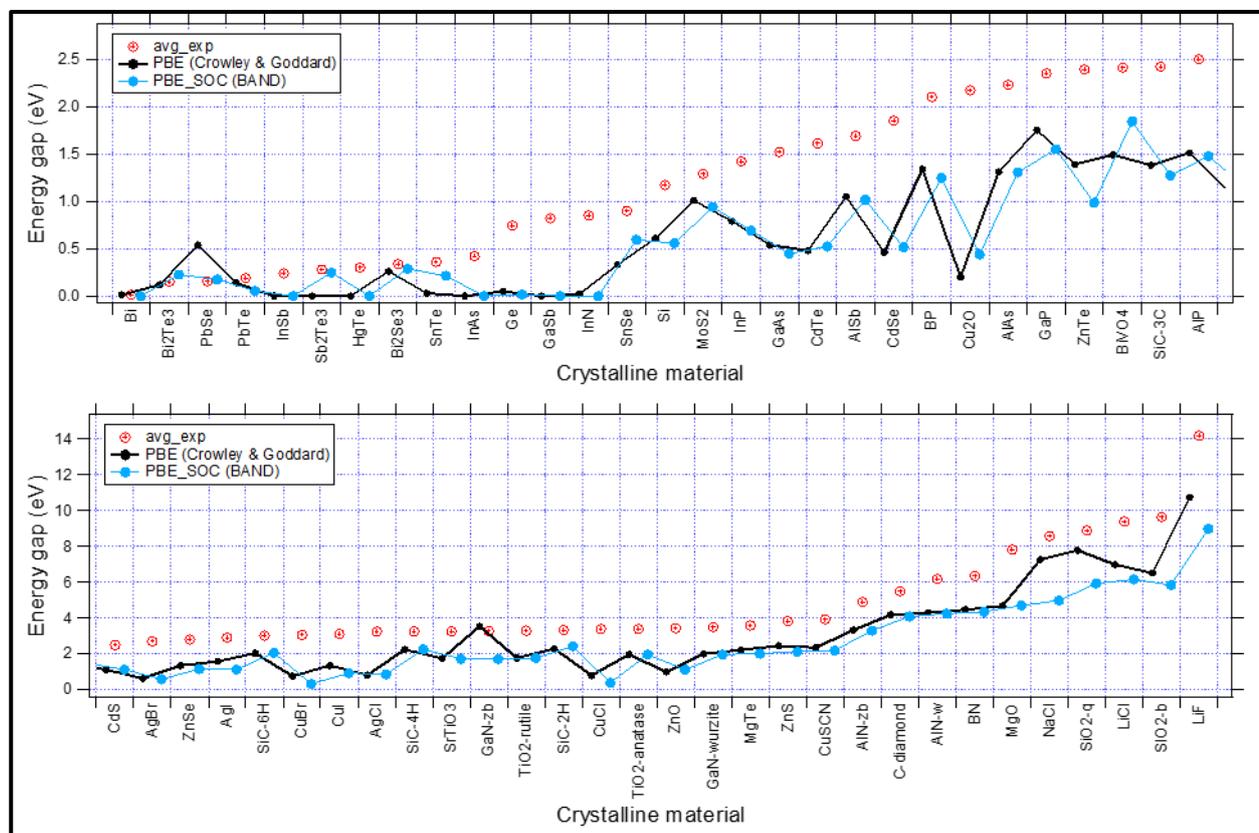


Figure 5 Comparison of PBE results obtained by BAND and CRYSTAL98

3.3 Spin orbit contribution.

Since not all the functionals can be simulated including Spin-Orbit effects, it is useful to have an idea of the systems for which SOC can represent a significant contribution and of which order of magnitude these values are. In order to do this, we run simulations with and without spin-orbit coupling (SOC) for our set of system and we compare the difference induced by SOC, Δ_{SOC} .

From our results (see table C3 in the appendix), we concluded that a consistent spin-orbit effect is observed for the same type of calculations, i.e. for all the GGA or model potentials. Furthermore, we identified that despite some variations in the number, the same systems were consistently identified as having a strong SOC contribution. The exception to this rule was only when a particular functional was not able to compute properly the bandgap. A summary of system for which SOC should be taken into account is listed in

Table 2 List of systems with substantial differences between Scalar and Spin-Orbit coupling

crystal	Exp E_g
AgBr	2.71
AgI	2.91
AlSb	1.69
Bi2Se3	0.335
Bi2Te3	0.1505
BiVO4	2.41
CdSe	1.85
CdTe	1.61
Cu2O	2.17
CuCl	3.4
CuI	3.12
GaSb	0.82
HgTe	0.3
InSb	0.24
MgTe	3.6
PbSe	0.155
PbTe	0.19
Sb2Te3	0.28
SnTe	0.36
ZnSe	2.82
ZnTe	2.39

Upon analysis of these results we observe that the bandgap is strongly affected for systems containing Bi, Pb and Te, and moderately affected by elements like Se, Sb, I, Br, Zn and Cu. The contributions calculated can be empirically taken into account to adjust the bandgap obtained by those functionals of the same type that have not been yet implemented spin-orbit coupling, for example correcting GLLB-SC using LB94 as shown in section 3.4.3.

3.4 Comparison of XC functionals

3.4.1 LDA and GGA's.

Finally, we arrive to the point of comparing the bandgaps obtained by the different functionals. First, we compared the most basic simulations, i.e. the LDA and different GGA functionals. All these simulations considered SOC effects. The results are shown in **Table 1** and **Figure 6**.

Table 3. Statistical analysis of the LDA and GGA calculations

Functional	Exp	LDA	PBE	PW91	BLYP
Kendall τ	1	0.71	0.74	0.74	0.75
ME	0	1.50	1.39	1.39	1.31
MAE	0	1.51	1.40	1.39	1.31
MAPE (%)	0	57.1	52.8	52.9	50.7

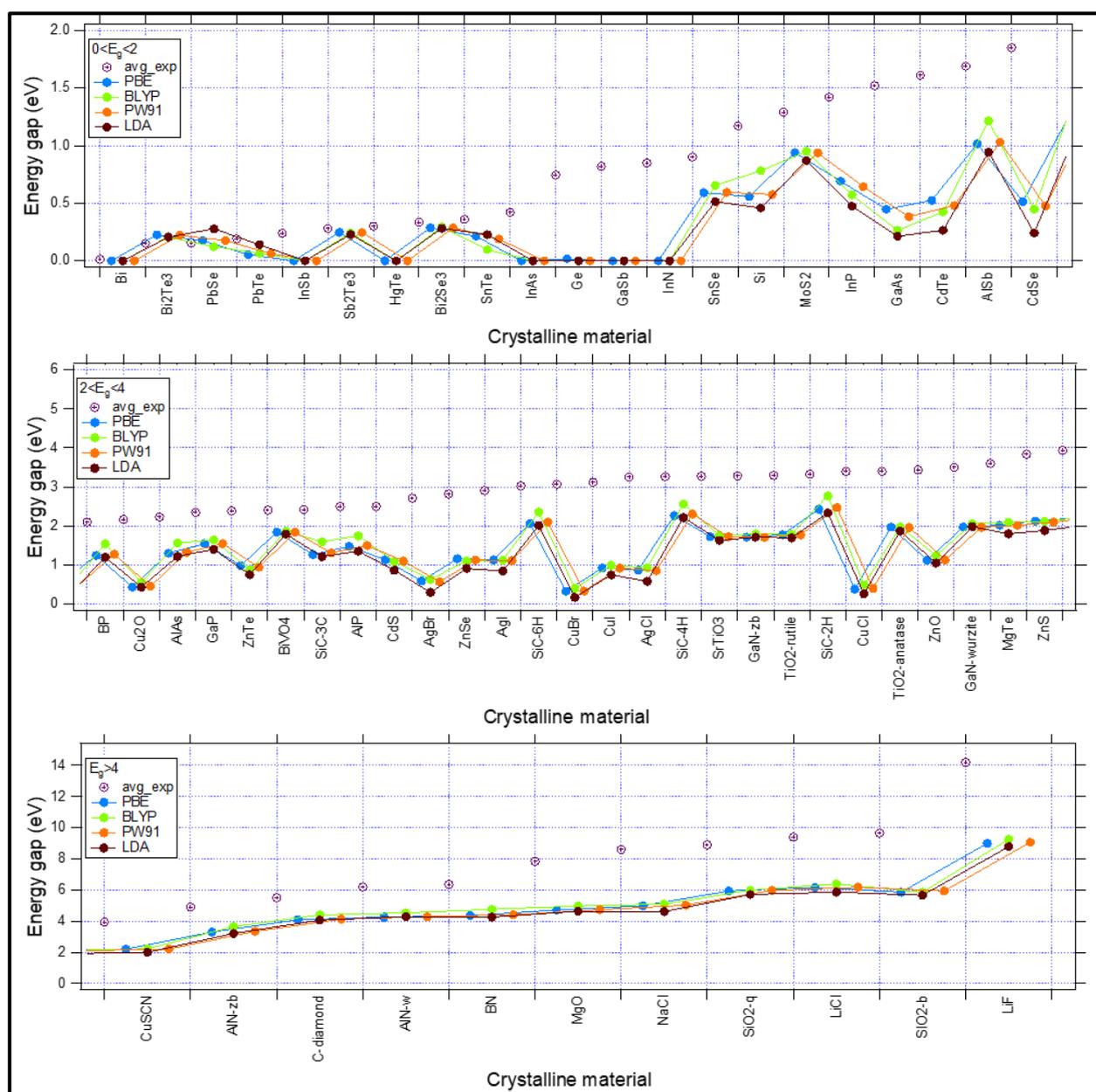


Figure 6 Graphic comparison of LDA and the different GGA functionals. Notes: (1) The ordering is based on the experimental value. (2) Data is subdivided for better visualization only.

From our results, it seems that LDA results in consistently higher errors while keeping a slightly better relation with the trend of the experimental data. On the other hand, the different gradient approximations give very similar values, with only a slight statistical advantage for BLYP functional. In conclusion, we note that there are very small differences between the different functionals and not a great improvement of the GGA over LDA.

3.4.2 Model Potentials with SOC: mBJ's and LB94

Next step was to compare the different model potentials available in BAND. For comparison purposes, we first consider those functionals for which the SOC contribution can be incorporated: the modified Becke-Johnson potentials (mBJs) and LB94. Next section builds up on these results to interpret the results for the remaining model potential: GLLB-SC for which SOC cannot be calculated using BAND.

From the analysis shown in **Table 4** it is evident that the JTS-mTB-mBJ gives really bad results (MAPE of 400!), which is why this this potential is immediately discarded. (This XC is not shown in **Figure 7** because it is out of the window of the rest of the results). It should be noted that the different mBJ potentials are in reality different parametrizations of the same potential [9], for which it is normal than one of them gives very good results while another gives really bad ones simply because it is not fitted for this type of calculation.

Table 4 Statistical analysis for the different model potentials.

Functional	Exp.	TB-mBJ	kTB-mBJ	JTS-mTB-mBJ	LB94
Kendall t	1	0.82	0.81	0.66	0.76
ME	0	0.03	-0.32	-2.17	1.28
MAE	0	0.39	0.60	3.36	1.28
MAPE (%)	0	26.0	27.8	446.3	52.6

Further analysis of **Table 4** and **Figure 7** shows that TB-mBJ and kTB-mBJ give comparable results, although the visual and numerical advantage of TB-mBJ is obvious. LB94, on the other hand seems to consistently underestimate the bandgap. In summary, the TB-mBJ outperforms the remaining model potentials in addition to the LDA and GGAs shown in the last section.

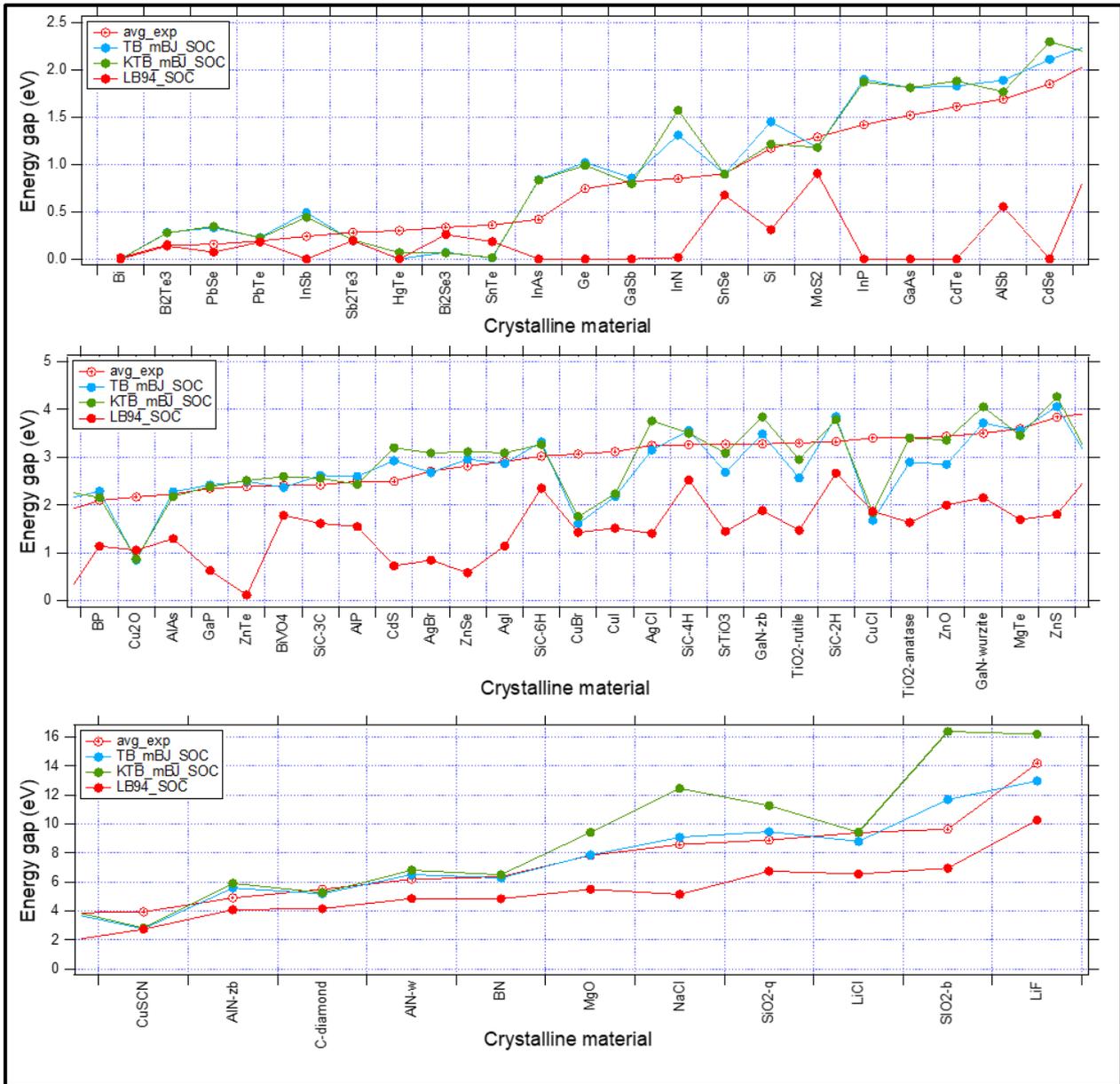


Figure 7. Graphic comparison of the different model potentials for which SOC is implemented.

3.4.3 Model Potential: GLLB-SC

Table 5 Statistical analysis for GLLB-SC.

	Scalar (59)	Scalar corrected*(59)	Scalar without SOC effects**(43)
Kendall τ	0.799	0.806	0.848
ME	0.783	0.845	0.860
MAE	0.847	0.859	0.875
MAPE	50.808	38.459	36.907

*Values of scalar calculations corrected taking the spin-orbit contributions from LB94.

** Systems for which the SOC effect in LB94 was seen to be larger than 0.1 eV were eliminated.

Direct comparison of GLLB-SC with the other Model potentials is, unfortunately, not possible because this function is not implemented in BAND. To go around this limitation two approaches

can be taken. The first one is to include the contribution of a similar XC for which the SOC contribution can be estimated. The second one is to simply eliminate the data for the system in which strong SOC effects have been observed (See section 3.3). We attempt both approaches and calculated the statistical values shown in **Table 5**.

Figure 8 shows the effect of applying the mentioned correction in our data. It is observed that this correction generally results on smaller values of SOC. This effect is beneficial on systems containing very heavy atoms such as Bi and Pb.

Finally, comparing the corrected statistical errors of GLLB-SC with the results of **Table 4** places GLLB-SC in second place in terms of numerical accuracy.

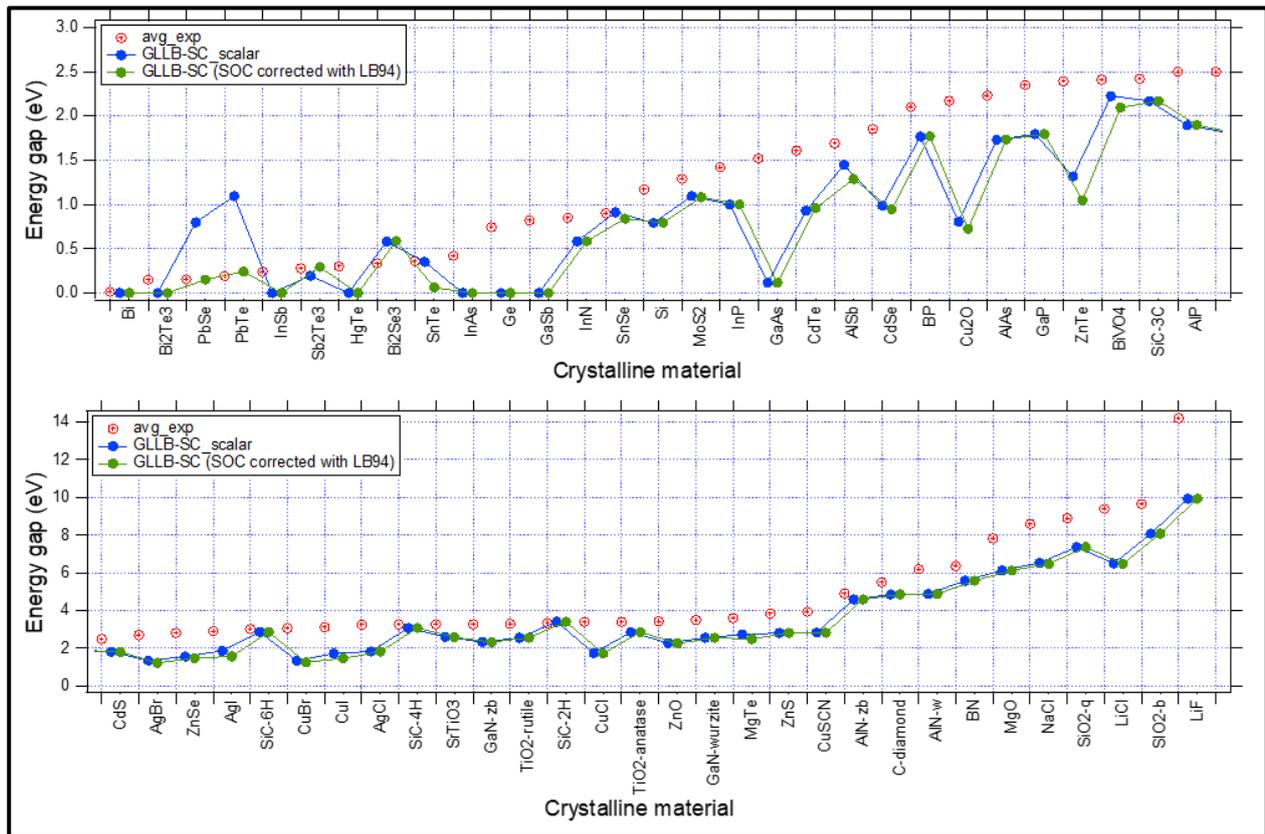


Figure 8. Graphic comparison of the results of GLLB-SC with and without the suggested correction for SOC.

3.5 Comparisons with HSE06

As a last experiment, we would like to compare the range separated hybrid functional HSE06 with the results previously obtained. This comparison is not straightforward due to technical reasons. On the first place, BAND does not implement a SOC functionality for this functional¹. On the second one, a part of the HSE06 did not converged or did so very slowly. In order to account for this effects, we eliminated the system for which either of these situation took place and recalculated the statistical metrics presented before. In total, we eliminate 12 systems that did not converge and/or 19 than had strong spin-orbit effect. This reduced our data set to 34 out of the 59 original systems and biased the metrics slightly towards HSE06. See **Table 6** and **Figure 9**.

¹ HSE calculations take the longest time and the inclusion of SOC effects multiplies this time by a factor of 10. So even if the SOC functionality was available it would have been too computationally expensive to compute within the timeframe of this project.

Table 6. Statistical analysis of the different XC functionals with the reduced data set.

	Exp.	LDA	PBE	BLYP	PW91	LB94	TB-mBJ	KTB-mBJ	GLLB-SC	HSE06
kendall τ	1	0.701	0.695	0.709	0.698	0.794	0.763	0.770	0.801	0.886
ME	0	-1.625	-1.493	-1.387	-1.487	-1.349	0.008	0.531	-0.842	-0.537
MAE	0	1.625	1.493	1.387	1.487	1.349	0.497	0.877	0.847	0.554
MAPE	0	57.990	53.137	49.858	53.354	52.473	25.025	28.325	33.806	18.238

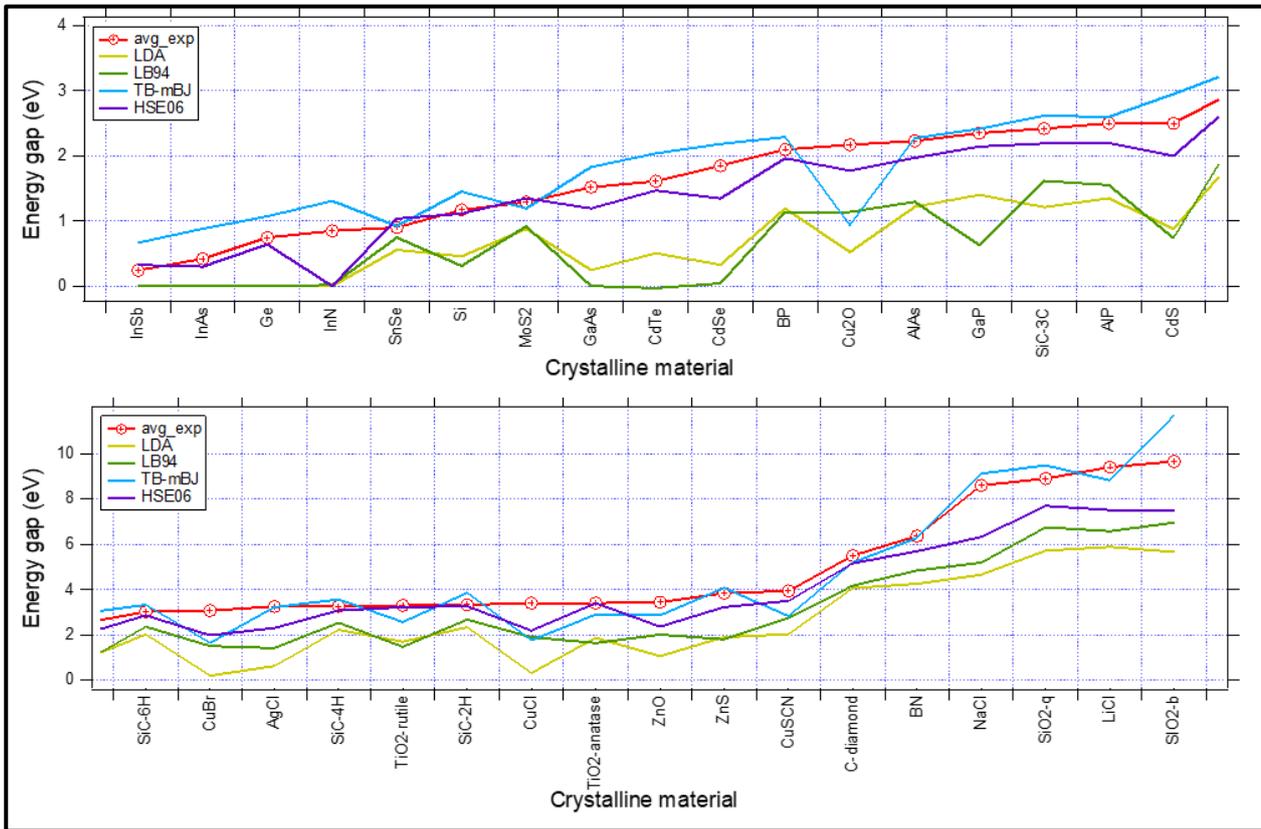


Figure 10 Graphic comparison of the different XC functionals with the reduced data set.

The above shown comparisons manifest the good properties of the HSE06 to calculate bandgaps. This functional was strongly favored by both the Kendall correlation and the MAPE and it shown still good values of the remaining two parameters.

Despite these good characteristics, there is a big disadvantage on HSE06, and it is that it that while the other functionals took comparable amounts of time, this one takes roughly 100 times more. This large difference somehow opaque the numerical advantage and makes it unpractical to use in case of limited resources.

4 Summary and conclusions

In summary, we perform simulations in a total of 59 systems for which the bandgap is known. We used a selection of XC functionals which included: LDA, GGAs (PBE, PW91 and BLYP), Model Potentials (LB94, GLLB-SC and mBJ's) and one range separated hybrid (HSE06). All the simulations were done using BAND (from SCM). We show the procedure used to arrive to the less computationally expensive settings (recipe) by converging the basis set and K-space.

We have empirically deduced which materials presented strong spin-orbit coupling effects (SOC) and that this effect is consistent for similar types of functionals.

From the comparison of our results we concluded that (1) the different GGAs give very similar values and all of them outperform LDA in bandgap calculations. (2) Out of the different potentials the best results are obtained by TB-mBJ followed by GLLB-SC. (3) The HSE06 hybrid gives the best results in terms of numerical accuracy. However, this accuracy comes with a substantial computational cost. (4) the improvement in accuracy between HSE06 and TB-mBJ does not compensate for the large difference in computational time for which in general we would recommend to use TB-mBJ as the best balance between accuracy and computational time.

In response to our initial questions...

1. According to different statistical metrics, which functional(s) seems to be doing a better job at estimating the bandgap?

Taking absolute numerical accuracy into account the best functionals out the ones studied seem to be the model potential TB-mBJ and the range-separated hybrid HSE06.

2. Which XC functional(s) results in a better balance between computational time and accuracy?

When the computational time comes into account the balance favours the functional TB-mBJ.

3. How does our results compare to similar studies reported in scientific literature?

We were able to reasonably reproduce the PBE results of Crowley and Goddard. Sadly, we could not reproduce their B3PW91 results because of BAND limitations. However, our analysis shows that drawing straight conclusions out of one single statistical value can be inaccurate, therefore we still doubt the validity of their conclusion that they have found the best functional for bandgap calculations.

On the other hand, in agreement to the findings of [5], HSE06 seems to give good bandgap estimations. Although we found that a very similar accuracy can be achieved by a model potential such as TB-mBJ.

References

- [1] J. M. Crowley, J. Tahir-Kheli, and W. A. Goddard, “Resolution of the Band Gap Prediction Problem for Materials Design,” *J. Phys. Chem. Lett.*, vol. 7, no. 7, pp. 1198–1203, 2016.
- [2] E. J. Baerends, “From the Kohn–Sham band gap to the fundamental gap in solids. An integer electron approach,” *Phys. Chem. Chem. Phys.*, vol. 19, no. 24, pp. 15639–15656, 2017.
- [3] R. van Leeuwen and E. J. Baerends, “Exchange-correlation potential with correct asymptotic behavior,” *Phys. Rev. A*, vol. 49, no. 4, pp. 2421–2431, Apr. 1994.
- [4] D. Koller, F. Tran, and P. Blaha, “Improving the modified Becke-Johnson exchange potential,” *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 85, no. 15, pp. 1–8, 2012.
- [5] A. J. Garza and G. E. Scuseria, “Predicting Band Gaps with Hybrid Density Functionals,” *J. Phys. Chem. Lett.*, vol. 7, no. 20, pp. 4165–4170, 2016.
- [6] S. Soman and R. Tang-Kong, “Mott Metal Insulator Transitions,” *Budker.Berkeley.Edu*. pp. 1–8, 2010.
- [7] V. I. Anisimov, M. A. Korotin, and E. Z. Kurmaev, “Band-structure description of Mott insulators (NiO, MnO, FeO, CoO),” *J. Phys. Condens. Matter*, vol. 2, no. 17, p. 3973, 1990.
- [8] M. J. Lucero, T. M. Henderson, and G. E. Scuseria, “Improved semiconductor lattice parameters and band gaps from a middle-range screened hybrid exchange functional,” *J. Phys. Condens. Matter*, vol. 24, no. 14, p. 145504, 2012.
- [9] SCM, “BAND Manual,” 2017.

A. Scripts

1. Bash: Convert CIF into XYZ using cif-reader within ADF

```
2. #!/bin/sh
3.
4. local_dir='/home/jetsa/scripts/crowley_coord:'
5.
6. for file in *.cif ; do
7.     echo $file
8.     b_name=${file%.*}
9.     echo $b_name
10.    $ADFBIN/cifreader $file -o $local_dir/$b_name.xyz
11. done
```

2. Bash: Run PLAMS script and save plams.* folder

```
1. #! /bin/sh
2.
3. dir=/home/figueroa/scripts
4.
5. $ADFBIN/plams $dir/runall3.py
6.
7. mv /scratch/plams.* $dir/tmpplams/
```

3. Python: script to reorganize XYZ files

```
1. #!/usr/bin/python
2.
3. local_dir='/home/jetsa/Desktop/crowley_coord'
4.
5.
6. def get_files_of_type (dir, type):
7.     import glob, os
8.     os.chdir(dir)
9.     return glob.glob("*. "+type)
10.
11.
12. def convert (list_of_files):
13.     for file_to_open in list_of_files:
14.         file_to_read = open(file_to_open, 'r')
15.         f = file_to_read.readlines()
16.
17.         outputVector = ''
18.         outputAtoms = ''
19.         counter = 0
20.         vecIndex = 1
21.         processCellVector = False
22.         processCoordinates = False
23.
24.         for index in range(len(f)):
25.             line = f[index].strip()
26.             if(line == '$end'):
27.                 break
28.
29.             if(line == "$cell_vectors" and processCellVector == False):
30.                 processCellVector = True
31.                 processCoordinates = False
32.                 continue
33.
34.             if(line == '$coordinates' and processCoordinates == False):
35.                 processCoordinates = True
36.                 processCellVector = False
```

```

37.         continue
38.
39.         if(processCellVector):
40.             outputVector += "VEC"+str(vecIndex) + '          ' + line + '\n'
41.             vecIndex += 1
42.
43.         if(processCoordinates):
44.             counter += 1
45.             outputAtoms += line + '\n'
46.
47.         # save output to an external file
48.         with open("converted/" + file_to_open , 'w') as fs:
49.             fs.write(str(counter)+'\n\n'+outputAtoms+outputVector)
50.
51.
52.
53. list_of_files = get_files_of_type(local_dir, 'xyz')
54. convert(list_of_files)

```

4. Python: main PLAMS script

```

5. import os
6. import sys
7.
8. # coord_dir = '/home/figueroa/converted_coord'
9. # coord_dir = '/home/jetsa/converted_coord'
10. coord_dir = '/home/figueroa/xyz_crowley'
11.
12. crystals = []
13. bandgaps = []
14. times = []
15.
16. common_settings = Settings()
17. common_settings.input.UNITS.length = 'Angstrom'
18. common_settings.input.UNITS.angle = 'Degree'
19.
20. sett = common_settings.copy()
21. sett.input.BasisDefaults.BasisType = 'TZP'
22. sett.input.BasisDefaults.core = 'None'
23. sett.input.NumericalQuality = 'Good'
24. sett.input.KSpace.Quality = 'Good'
25. sett.input.XC.libxc = 'hse06'
26. sett.input.Relativistic = 'Scalar ZORA'
27. sett.input.SCF.iterations = 60
28. sett.input.SoftConfinement.Quality = 'Normal'
29. #sett.input.UNRESTRICTED = '!'
30.
31. print(sett)
32.
33. for root, dirs, filenames in os.walk(coord_dir):
34.     d=sorted(filenames, key=str.lower)
35.     #d=['Sb2Te3.xyz']
36.     for f in d:
37.         filename = os.path.join(root, f)
38.         crystal = os.path.basename(f).rsplit('.')[0]
39.         print (crystal)
40.         mol=Molecule(filename)
41.         # Create and run the job:
42.         job = BANDJob(molecule=mol, settings=sett, name=crystal)
43.         # print(job.get_input())
44.         job.run()
45.
46.         if job.check():
47.             bandgap = job.results.readkf('BandStructure','BandGap')
48.             bandgap_ev = Units.convert(bandgap, 'Hartree', 'eV')
49.             time=job.results.grep_output('Elapsed')[-1].split()[-1]
50.             print(crystal, 'Bandgap: %f eV' % bandgap_ev)

```

```

51.         print('Time(s):', time)
52.         crystals.append(crystal)
53.         bandgaps.append(bandgap_ev)
54.         times.append(time)
55.     else:
56.         print("job crashed ;", crystal)
57.         crystals.append(crystal)
58.         bandgaps.append('error')
59.         print(job.results.grep_output('ERROR'))
60.         sys.stdout.flush()
61.
62. print ('Summary of results')
63. for x,y,z in zip(crystals,bandgaps,times):
64.     print (x,y,z)

```

5. Python: Calculate Kendell coefficient from data in excel table.

```

1. from scipy import stats as s
2. import xlrd
3. import os.path
4.
5. # read data from excel file
6. directory = "C:/Users/sakus/Desktop/internship_report"
7. fileName = "report.xlsx"
8.
9. workBook = xlrd.open_workbook(os.path.join(directory, fileName))
10. sh = workBook.sheet_by_index(12)
11.
12. i = 1
13. x1 = []
14. x2 = []
15. x3 = []
16.
17. while i<60:
18.     Load = round(sh.cell(i,1).value,2)
19.     Load2 = round(sh.cell(i,2).value,2)
20.     Load3 = round(sh.cell(i,3).value,2)
21.     x1.append(Load)
22.     x2.append(Load2)
23.     x3.append(Load3)
24.     i += 1
25.
26. print('Experimental\n', x1, '\n')
27.
28. print('BLYP calculations\n')
29. print('Scalar\n', x2, '\n')
30. tau, p_value = s.kendalltau(x1, x2)
31. print('kendall_coef:',tau, '\n')
32. print(p_value)
33.
34. print('SOC\n', x3, '\n')
35. tau2, p_value2 = s.kendalltau(x1, x3)
36. print('kendall_coef2:',tau2, '\n')
37. print(p_value2)

```

B. Geometric input

Example of “Hacked” CIF file

```
1. data_xxx
2. _chemical_formula_sum      'Na Cl'
3. _symmetry_Int_Tables_number 225
4. _cell_angle_alpha         90
5. _cell_angle_beta          90
6. _cell_angle_gamma         90
7. _cell_length_a            5.6573
8. _cell_length_b            5.6573
9. _cell_length_c            5.6573
10. loop_
11. _atom_site_label
12. _atom_site_fract_x
13. _atom_site_fract_y
14. _atom_site_fract_z
15. Na 0.00 0.00 0.00
16. Cl 0.5 0.5 0.5
```

C. Tables of results.

Table C1: Basic Set Convergence

	SP_PBE_DZ_small _scalar_good	SP_PBE_TZP_small l_scalar_good	SP_TZ2P_SMALL_G OOD_SCALAR	SP_QZ4P_SMALL_G OOD_SCALAR
Bi	0	0	0	
InSb	0	0	0	0
HgTe	0	0	0	0
InAs	0	0	0	0
VO2	0	0	0	0
ZnTe	0	0	0	0
FeO	0	0	0	0
MnO	0	0	0	0
CoO	0	0	0	0
NiO	0	0	0	0
Sb2Te3	0.064488	0.00827	0.019074299	0.020893418
GaSb	0	0.028651	0.007258636	0.003151231
Ge	0	0.032786	0.009770304	0.006127007
SnTe	0.021358	0.043874	0.06494913	0.066303944
InN	0.155136	0.058645	0.036718195	0.034674183
Bi2Te3	0.261819	0.269755	0.262241474	0.262856521
PbSe	0.483352	0.325546	0.310135274	
CuBr	0.210403	0.358636	0.360984777	0.351708084
CuCl	0.22623	0.445596	0.436517941	0.445730042
CdSe	0.442511	0.455318	0.452194549	0.457188342
GaAs	0.361375	0.472602	0.459382703	0.457118812
Cu2O	0.535762	0.522347	0.522248525	0.523167489
Si	0.993212	0.627505	0.617142327	0.574225522
SnSe	0.701062	0.642701	0.663062692	0.668584692
MoS2	0.4712	0.657845	0.634722071	0.649439551
AgBr	0.585254	0.679729	0.682856451	0.685240391
InP	0.656794	0.68958	0.659163007	0.648852281
CdTe	0.741007	0.759011	0.747317761	0.746571495
PbTe	0.883306	0.780458	0.738346818	
AgCl	0.793448	0.914799	0.915922712	0.920372065
ZnO	1.134493	0.987809	0.946698506	0.935345629
CuI	1.010257	1.065694	1.129594772	1.122490952
AlSb	1.294065	1.124749	1.082679344	1.075256664
CdS	1.139369	1.146549	1.12485151	1.12466074
BP	1.760029	1.24441	1.237868069	1.14872212
ZnSe	1.237479	1.245038	1.235419604	1.240642928
SiC (3C)	3.161356	1.271166	1.270789797	1.22366138
AlAs	1.720616	1.29762	1.297771399	1.28903636
AgI	1.270678	1.336654	1.341951953	1.34492156
AlP	2.101331	1.473435	1.473386987	1.4120361
GaP	1.65643	1.552316	1.550062241	1.519429889
SrTiO 3	1.685	1.736732	1.823340784	1.842858648

TiO2_rutile	1.753345	1.796451	1.8688065	1.890417053
GaN_zb	2.206259	1.961985	1.916130515	1.911749942
TiO2_Anatase	1.997811	2.024308	2.119979038	2.140271184
SiC (6H)	3.316069	2.063819	2.059941741	2.029751452
ZnS	2.068765	2.091962	2.075472084	2.072429267
BiVO4	2.130449	2.15337	2.185509131	2.186551357
CuSCN	2.105163	2.310314	2.30452677	2.3056931
C_diamond	5.126502	4.078494	4.075894515	4.040112078
AlN-wurzite	4.248402	4.203919	4.163845622	4.161312953
BN	6.3825	4.360008	4.358379322	4.314018636
MgO	4.622145	4.703865	4.730374	4.70033
NaCl	4.981979	5.019285	5.024442207	5.032860364
SiO2_B-cristobalite	6.346145	5.854323	5.877223856	5.666880464
SiO2_a-quartz	5.901942	6.015379	6.038653	6.000411788
LiCl	6.164661	6.197096	6.202152685	6.206116251
LiF	9.087668	8.988885	8.992682472	8.996421353

Table C2: K-space Convergence

Crystal	normal	Good	Verygood	excelent
Bi	1.294905	0	0	0
InSb	1.611346	0	0	0
HgTe	1.779276	0	0	0
InAs	1.994422	0	0	0
VO2	0	0	0	0
ZnTe	0.11493	0	0	0
Sb2Te3	0.015302	0.00827	0.006131	0.007416
GaSb	1.613953	0.028651	-0.01137	0.021241
Ge	1.607711	0.032786	-0.03412	0.004579
SnTe	0.011371	0.043874	0.048352	0.047741
InN	0.061345	0.058645	0.044713	0.039356
Bi2Te3	0.282548	0.269755	0.269577	0.270339
PbSe	0.416208	0.325546	0.309218	0.309304
CuBr	2.797449	0.358636	0.348468	0.354339
CuCl	2.956719	0.445596	0.441957	0.44797
CdSe	3.059616	0.455318	0.450139	0.463063
GaAs	2.200917	0.472602	0.457062	0.493473
Cu2O	0.521646	0.522347	0.522381	0.522387
Si	2.283525	0.627505	0.583302	0.629583
SnSe	0.816567	0.642701	0.51256	0.533
MoS2	1.144263	0.657845	0.606504	0.636259
AgBr	3.024243	0.679729	0.677082	0.677051
InP	2.569792	0.68958	0.680304	0.707448
CdTe	2.728419	0.759011	0.751168	0.76891
PbTe	0.838812	0.780458	0.769181	0.769434
AgCl	3.021971	0.914799	0.914413	0.914397
ZnO	1.01151	0.987809	0.985914	0.981357

CuI	2.838763	1.065694	1.049334	1.060183
AlSb	2.421583	1.124749	1.136422	1.179949
CdS	3.913636	1.146549	1.143423	1.15695
BP	1.306086	1.24441	1.323401	1.356428
ZnSe	3.495622	1.245038	1.250032	1.272602
SiC-C	1.15754	1.271166	1.343096	1.373798
AlAs	3.12966	1.29762	1.331511	1.37254
AgI	3.606916	1.336654	1.331465	1.340799
AlP	3.783546	1.473435	1.518684	1.560355
GaP	2.894289	1.552316	1.551323	1.588744
SrTiO3	1.840716	1.736732	1.729484	1.729013
TiO2-rutile	1.855685	1.796451	1.797393	1.797228
GaN-zb	1.995632	1.961985	1.958385	1.958028
TiO2-anatase	2.092228	2.024308	2.039937	2.038901
SiC-6H	2.090009	2.063819	2.048625	2.053254
ZnS	4.329464	2.091962	2.099978	2.120961
BiVO4	2.279724	2.15337	2.145543	2.14111
CuSCN	2.313718	2.310314	2.311062	2.31106
C_diamond	4.110221	4.078494	4.126799	4.148624
AlN-w	4.227094	4.203919	4.202369	4.202282
BN	4.332249	4.360008	4.432351	4.466143
MgO	4.71439	4.703865	4.705303	4.706722
NaCl	7.088977	5.019285	5.013964	5.015804
SiO2-b	5.965032	5.854323	5.797934	5.835989
SiO2-q	6.031945	6.015379	6.015928	6.015515
LiCl	6.117095	6.197096	6.222415	6.23213
LiF	8.920733	8.988885	9.009486	9.017307

Table C3: PBE Results (CRYSTAL98 vs. BAND)

crystal	avg exp	PBE_crowley	PBE_TZP_scalar_good_small	PBE_TZP_SOC_good_small
Bi	0.013	0.01	0.000	0.000
Bi2Te3	0.1505	0.12	0.269	0.226
PbSe	0.155	0.54	0.337	0.175
PbTe	0.19	0.14	0.788	0.054
InSb	0.24	0	0.000	0.000
Sb2Te3	0.28	0	0.008	0.247
HgTe	0.3	0	0.000	0.000
Bi2Se3	0.335	0.26	0.204	0.286
SnTe	0.36	0.03	0.047	0.213
InAs	0.42	0	0.000	0.000
Ge	0.744	0.05	0.077	0.018
GaSb	0.82	0	-0.030	0.000
InN	0.85	0.02	0.000	0.000
SnSe	0.9	0.33	0.636	0.592
Si	1.17	0.61	0.556	0.557

MoS2	1.29	1.01	0.949	0.939
InP	1.42	0.79	0.689	0.692
GaAs	1.52	0.54	0.479	0.447
CdTe	1.61	0.48	0.755	0.523
AlSb	1.69	1.05	1.130	1.015
CdSe	1.85	0.46	0.597	0.514
BP	2.1	1.34	1.244	1.247
Cu2O	2.17	0.2	0.524	0.438
AlAs	2.23	1.31	1.298	1.304
GaP	2.35	1.75	1.550	1.550
ZnTe	2.39	1.39	1.206	0.986
BiVO4	2.41	1.49	1.973	1.843
SiC-3C	2.42	1.38	1.271	1.272
AlP	2.5	1.51	1.475	1.477
CdS	2.5	1.12	1.150	1.135
AgBr	2.71	0.63	0.680	0.594
ZnSe	2.82	1.36	1.231	1.166
AgI	2.91	1.59	1.337	1.131
SiC-6H	3.023	2.04	2.064	2.060
CuBr	3.07	0.76	0.359	0.329
CuI	3.12	1.36	1.066	0.929
AgCl	3.25	0.81	0.915	0.873
SiC-4H	3.263	2.26	2.268	2.264
SrTiO3	3.275	1.75	1.743	1.727
GaN-zb	3.28	3.545	1.718	1.717
TiO2-rutile	3.3	1.78	1.777	1.777
SiC-2H	3.33	2.3	2.434	2.429
CuCl	3.4	0.8	0.446	0.387
TiO2-anatase	3.4	1.98	1.973	1.973
ZnO	3.44	1	1.135	1.124
GaN-wurzite	3.503	2.01	1.976	1.973
MgTe	3.6	2.23	2.215	2.020
ZnS	3.84	2.46	2.134	2.126
CuSCN	3.94	2.36	2.223	2.197
AlN-zb	4.9	3.33	3.306	3.306
C-diamond	5.5	4.19	4.096	4.096
AlN-w	6.19	4.31	4.254	4.253
BN	6.36	4.48	4.359	4.361
MgO	7.83	4.69	4.718	4.711
NaCl	8.595	7.27	5.032	4.990
SiO2-q	8.9	7.79	5.940	5.939
LiCl	9.4	7	6.187	6.175
SiO2-b	9.65	6.52	5.852	5.854
LiF	14.2	10.75	8.996	8.994
Kendall_coef		0.761	0.729	0.737

ME		1.188		1.396
MAE		1.210		1.392
MAPE		53.338		52.838

TABLE C4: BLYP and PW91 results

crystal	avg exp	BLYP_Scalar_TZP_s mall_good	BLYP_SOC_TZP_s mall_good	PW91_Scalar_TZP_s mall_good	PW91_SOC_TZP_s mall_good
Bi	0.01 3	0.000	0.0000	0.000	0.000
Bi2Te3	0.15 05	0.306	0.2042	0.263	0.221
PbSe	0.15 5	0.463	0.1222	0.343	0.176951
PbTe	0.19	0.894	0.0652	0.793	0.061885
InSb	0.24	0.000	0.0000	0.000	0
Sb2Te3	0.28	0.062	0.2433	0.008	0.247339997
HgTe	0.3	0.000	0.0000	0.000	0
Bi2Se3	0.33 5	0.258	0.2945	0.196	0.287
SnTe	0.36	0.124	0.0967	0.026	0.189831208
InAs	0.42	0.000	0.0000	0.000	0
Ge	0.74 4	0.000	0.0000	-0.002	0
GaSb	0.82	0.000	0.0000	0.000	0
InN	0.85	0.000	0.0000	0.000	0
SnSe	0.9	0.700	0.6556	0.635	0.594085433
Si	1.17	0.784	0.7817	0.573	0.575171713
MoS2	1.29	0.960	0.9492	0.946	0.93525
InP	1.42	0.570	0.5735	0.641	0.643804
GaAs	1.52	0.306	0.2638	0.417	0.383
CdTe	1.61	0.665	0.4243	0.714	0.480
AlSb	1.69	1.345	1.2158	1.148	1.030
CdSe	1.85	0.535	0.4471	0.561	0.476
BP	2.1	1.544	1.5468	1.275	1.278
Cu2O	2.17	0.646	0.5608	0.541	0.456
AlAs	2.23	1.560	1.5647	1.318	1.324
GaP	2.35	1.644	1.6445	1.549	1.548988
ZnTe	2.39	1.112	0.8817	1.165	0.942745295
BiVO4	2.41	1.994	1.8611	1.973	1.841
SiC-3C	2.42	1.597	1.5963	1.312	1.313904753
AlP	2.5	1.745	1.7473	1.494	1.497
CdS	2.5	1.108	1.0939	1.119	1.105
AgBr	2.71	0.725	0.6351	0.660	0.573
ZnSe	2.82	1.182	1.1105	1.195	1.128913826
AgI	2.91	1.338	1.1248	1.317	1.108
SiC-6H 3	3.02 3	2.362	2.3584	2.102	2.102060784

CuBr	3.07	0.435	0.4048	0.363	0.334
CuI	3.12	1.139	0.9941	1.061	0.922
AgCl	3.25	0.972	0.9323	0.899	0.857
SiC-4H 3	3.26	2.565	2.5619	2.306	2.304973326
SrTiO3 5	3.27	1.775	1.7593	1.740	1.724617805
GaN-zb	3.28	1.798	1.7974	1.713	1.711963
TiO2- rutile	3.3	1.783	1.7831	1.773	1.773185368
SiC-2H	3.33	2.774	2.7727	2.476	2.474842761
CuCl	3.4	0.551	0.4939	0.459	0.401
TiO2- anatase	3.4	1.976	1.9760	1.965	1.965736198
ZnO	3.44	1.258	1.2470	1.143	1.131948539
GaN- wurzite 3	3.50	2.066	2.0625	1.973	1.970
MgTe	3.6	2.299	2.0961	2.213	2.015642
ZnS	3.84	2.125	2.1183	2.106	2.098179545
CuSCN	3.94	2.244	2.2189	2.240	2.214
AlN-zb	4.9	3.654	3.6544	3.348	3.349
C- diamond	5.5	4.400	4.3997	4.131	4.131
AlN-w	6.19	4.536	4.5359	4.286	4.285
BN	6.36	4.761	4.7624	4.412	4.413
MgO	7.83	4.980	4.9725	4.741	4.733327
NaCl 5	8.59	5.138	5.0952	5.073	5.029952
SiO2-q	8.9	5.993	5.9947	5.967	5.967475113
LiCl	9.4	6.405	6.3922	6.195	6.182513
SiO2-b	9.65	5.838	5.8516	5.920	5.935177578
LiF	14.2	9.267	9.2645	9.060	9.057893
Kendall_coef		0.725	0.747		0.729
ME		1.256	1.308	1.348	1.388
MAE		1.296	1.310	1.379	1.391
MAPE			50.716		52.914

Table C5: LDA and LB94 results

crystal	avg exp	LB94_scalar	LB94_SOC	LDA	LDA-SOC
Bi	0.013	0.000	0	0.390144656	0.30096
Bi2Te3	0.1505	0.267	0.138420853	0.6230517	0.583178
PbSe	0.155	0.721	0.073162475	1.060900469	0.847011
PbTe	0.19	1.032	0.179315733	1.217255406	1.223496
InSb	0.24	0.000	0	4.292042794	4.291679
Sb2Te3	0.28	0.097	0.194415669	3.20697468	3.207861
HgTe	0.3	0.000	0	1.350966203	1.3536

Bi2Se3	0.335	0.250	0.256970485	1.068531149	0.944931
SnTe	0.36	0.474	0.18298019	0	0
InAs	0.42	0.000	0	0.086504273	0.282792
Ge	0.744	0.000	0	0.190112798	0.206267
GaSb	0.82	0.000	0	1.927522779	1.791031
InN	0.85	0.018	0.016704056	4.256081426	4.257376
SnSe	0.9	0.748	0.676449979	1.194154492	1.196273
Si	1.17	0.310	0.311326721	4.065279107	4.065257
MoS2	1.29	0.919	0.905533278	0.880636925	0.866345
InP	1.42	0.000	0	0.328047657	0.241406
GaAs	1.52	0.000	0	0.505660391	0.265501
CdTe	1.61	-0.030	0	0.521324351	0.436226
AlSb	1.69	0.712	0.553490334	0.203168946	0.174074
CdSe	1.85	0.041	0	0.317569452	0.263379
BP	2.1	1.139	1.141434693	0.900081737	0.751475
Cu2O	2.17	1.139	1.059828034	2.026898134	2.003229
AlAs	2.23	1.294	1.298900641	0.248007032	0.212882
GaP	2.35	0.629	0.631190026	1.984843788	1.981667
ZnTe	2.39	0.386	0.118635939	1.717001617	1.71641
BiVO4	2.41	1.914	1.784493725	1.400777957	1.401053
SiC-3C	2.42	1.615	1.615350916	0	0
AlP	2.5	1.551	1.553930593	0	0
CdS	2.5	0.740	0.732073578	0	0
AgBr	2.71	0.980	0.8452769	0	0
ZnSe	2.82	0.680	0.587382787	0	0
AgI	2.91	1.439	1.145286352	0.469828213	0.473731
SiC-6H	3.023	2.359	2.353495598	0	0
CuBr	3.07	1.510	1.429014731	5.879908943	5.867543
CuI	3.12	1.771	1.515904541	8.791799855	8.789262
AgCl	3.25	1.419	1.407892229	4.639378604	4.631996
SiC-4H	3.263	2.536	2.530660961	2.005435028	1.803974
SrTiO3	3.275	1.468	1.449853972	0.882010499	0.8712
GaN-zb	3.28	1.889	1.88919697	4.656238628	4.612663
TiO2- rutile	3.3	1.475	1.474796826	0.235047105	0.278665
SiC-2H	3.33	2.674	2.668621764	0.706645101	0.139361
CuCl	3.4	1.894	1.868718629	0.045533112	0.230127
TiO2- anatase	3.4	1.637	1.637504067	0.457721669	0.458704
ZnO	3.44	2.013	2.000295647	2.338548364	2.333536
GaN- wurzite	3.503	2.162	2.155429281	1.214280872	1.215132
MgTe	3.6	1.933	1.697508477	2.222247939	2.21772
ZnS	3.84	1.807	1.804954028	2.019801137	2.015149
CuSCN	3.94	2.744	2.739999952	5.657809615	5.658888
AlN-zb	4.9	4.087	4.087549092	5.717409225	5.716267
C- diamond	5.5	4.167	4.167186133	0.557637288	0.51389

AlN-w	6.19	4.854	4.853067185	0.064547353	0.227091
BN	6.36	4.841	4.842015035	1.647084256	1.631452
MgO	7.83	5.503	5.496054019	1.872481905	1.872559
NaCl	8.595	5.187	5.143549953	1.69827222	1.698402
SiO2-q	8.9	6.747	6.745375956	1.063069785	1.052232
LiCl	9.4	6.578	6.56313516	1.893258587	1.886299
SIO2-b	9.65	6.953	6.953989661	0.983607233	0.915265
LiF	14.2	10.274	10.27070107	0.983049017	0.754609
Kendall_coef		0.739	0.761	0.715	0.711
ME		1.218	1.282	1.470	1.502
MAE		1.273	1.282	1.491	1.509
MAPE			52.629		57.056

Table C6: GLLB-SC results

crystal	avg exp	GLLB-SC_scalar	GLLB-SC (SOC-corrected by LB94)
Bi	0.013	0	0.000
Bi2Te3	0.1505	0	0.000
PbSe	0.155	0.7989	0.151
PbTe	0.19	1.093912094	0.242
InSb	0.24	0	0.000
Sb2Te3	0.28	0.195758461	0.293
HgTe	0.3	0	0.000
Bi2Se3	0.335	0.579518	0.586
SnTe	0.36	0.351518391	0.061
InAs	0.42	0	0.000
Ge	0.744	0	0.000
GaSb	0.82	0	0.000
InN	0.85	0.584666849	0.584
SnSe	0.9	0.910054965	0.838
Si	1.17	0.791957733	0.793
MoS2	1.29	1.09504945	1.082
InP	1.42	0.998720045	0.999
GaAs	1.52	0.115131131	0.115
CdTe	1.61	0.92931087	0.959
AlSb	1.69	1.445425	1.287
CdSe	1.85	0.98710449	0.946
BP	2.1	1.766331404	1.769
Cu2O	2.17	0.803700675	0.724
AlAs	2.23	1.729417	1.735
GaP	2.35	1.794371666	1.796
ZnTe	2.39	1.314391384	1.047
BiVO4	2.41	2.224407445	2.095
SiC-3C	2.42	2.167155178	2.168
AlP	2.5	1.893953	1.897
CdS	2.5	1.814707487	1.807
AgBr	2.71	1.357263	1.223

ZnSe	2.82	1.561398332	1.469
AgI	2.91	1.866432	1.573
SiC-6H	3.023	2.863233269	2.858
CuBr	3.07	1.344416661	1.263
CuI	3.12	1.7221404	1.467
AgCl	3.25	1.836427	1.826
SiC-4H	3.263	3.07959593	3.075
SrTiO3	3.275	2.607617945	2.590
GaN-zb	3.28	2.334931293	2.335
TiO2-rutile	3.3	2.553908847	2.554
SiC-2H	3.33	3.407985734	3.403
CuCl	3.4	1.745438894	1.720
TiO2-anatase	3.4	2.85856926	2.859
ZnO	3.44	2.278931208	2.266
GaN-wurzite	3.503	2.56361295	2.557
MgTe	3.6	2.722	2.486
ZnS	3.84	2.815363432	2.814
CuSCN	3.94	2.833426196	2.829
AlN-zb	4.9	4.592643	4.594
C-diamond	5.5	4.855018344	4.856
AlN-w	6.19	4.880699	4.880
BN	6.36	5.576925353	5.578
MgO	7.83	6.128817076	6.122
NaCl	8.595	6.519960046	6.476
SiO2-q	8.9	7.369793388	7.369
LiCl	9.4	6.4931	6.478
SiO2-b	9.65	8.07765026	8.079
LiF	14.2	9.9339	9.930

Table C7: TB-mBJ and kTB-mBJ

crystal	avg exp	TB-mBJ_scalar	TB-mBJ_SOC	KTB-mBJ_scalar	KTB-mBJ_SOC
Bi	0.013	0.00	0.00	0	0
Bi2Te3	0.1505	0.64	0.28	0.618756372	0.275608
PbSe	0.155	0.99	0.33	1.005222386	0.346559892
PbTe	0.19	1.28	0.23	1.264140754	0.22247006
InSb	0.24	0.67	0.49	0.615864135	0.441515
Sb2Te3	0.28	0.25	0.19	0.257126729	0.200336872
HgTe	0.3	0.20	0.00	0.274995872	0.067705
Bi2Se3	0.335	0.80	0.07	0.813031011	0.064274
SnTe	0.36	0.28	0.01	0.26206523	0.015804416
InAs	0.42	0.88	0.84	0.878338124	0.834187
Ge	0.744	1.07	1.02	1.044657399	0.990195
GaSb	0.82	1.01	0.86	0.952922473	0.793787
InN	0.85	1.31	1.31	1.577683982	1.573922
SnSe	0.9	0.92	0.90	0.920863767	0.896460867

Si	1.17	1.45	1.45	1.214766792	1.215688493
MoS2	1.29	1.19	1.18	1.190307996	1.179866
InP	1.42	1.90	1.90	1.871685553	1.872416
GaAs	1.52	1.83	1.81	1.828945238	1.811055
CdTe	1.61	2.04	1.83	2.09009645	1.883354
AlSb	1.69	2.00	1.89	1.879639252	1.768121
CdSe	1.85	2.18	2.11	2.367667889	2.297131
BP	2.1	2.29	2.29	2.155450505	2.157871
Cu2O	2.17	0.93	0.85	0.958358077	0.871823
AlAs	2.23	2.27	2.28	2.179087385	2.18417
GaP	2.35	2.42	2.42	2.386165392	2.385542
ZnTe	2.39	2.69	2.50	2.712746	2.517057
BiVO4	2.41	2.48	2.37	2.701003706	2.597111
SiC-3C	2.42	2.62	2.62	2.561678381	2.562356875
AlP	2.5	2.60	2.60	2.432117154	2.434449
CdS	2.5	2.95	2.93	3.220360704	3.201035
AgBr	2.71	2.74	2.68	3.15084511	3.091796
ZnSe	2.82	3.01	2.96	3.165050417	3.115538699
AgI	2.91	3.03	2.87	3.250735677	3.09392
SiC-6H	3.023	3.33	3.32	3.269542167	3.265757093
CuBr	3.07	1.65	1.61	1.799927781	1.760813
CuI	3.12	2.24	2.18	2.294602578	2.235844
AgCl	3.25	3.22	3.15	3.839177586	3.762506
SiC-4H	3.263	3.57	3.56	3.509528431	3.505797708
SrTiO3	3.275	2.71	2.69	3.101786802	3.086190922
GaN-zb	3.28	3.49	3.49	3.848179844	3.845383
TiO2-rutile	3.3	2.57	2.57	2.953346422	2.953027376
SiC-2H	3.33	3.85	3.85	3.791775516	3.788070211
CuCl	3.4	1.76	1.68	1.93274897	1.845101
TiO2-anatase	3.4	2.90	2.90	3.407579131	3.407457164
ZnO	3.44	2.86	2.85	3.368302624	3.358570104
GaN-wurzite	3.503	3.72	3.72	4.057924254	4.05699
MgTe	3.6	3.75	3.56	3.643358769	3.456722
ZnS	3.84	4.07	4.06	4.281031751	4.271022573
CuSCN	3.94	2.83	2.78	2.877124551	2.829153
AlN-zb	4.9	5.59	5.59	5.911109044	5.911812
C-diamond	5.5	5.20	5.20	5.259953322	5.259946
AlN-w	6.19	6.51	6.51	6.815083969	6.814867
BN	6.36	6.28	6.28	6.50230633	6.503473
MgO	7.83	7.88	7.87	9.432308211	9.424517
NaCl	8.595	9.12	9.09	12.49699155	12.464967
SiO2-q	8.9	9.47	9.47	11.28095182	11.28015206
LiCl	9.4	8.82	8.80	9.451455969	9.437799
SiO2-b	9.65	11.69	11.69	20.9208	16.3926
LiF	14.2	12.98	12.98	16.21525096	16.210981

Kendall_coef	0.810	0.815	0.810	0.813
ME	-0.061	0.031	-0.486	-0.317
MAE	0.441	0.395	0.725	0.600
MAPE		25.995		27.752

Table C8: JTS-mTB-mBJ results

crystal	avg exp	JTS-mTB-mBJ-scalar	JTS-mTB-mBJ-SOC
AgBr	2.71	5.532976421	5.480401
AgCl	3.25	6.567544679	6.464038
AgI	2.91	6.195760005	6.085433
AlAs	2.23	4.463264489	4.465277
AlN-w	6.19	9.746698559	9.746537
AlN-zb	4.9	8.990341494	8.990947
AIP	2.5	5.657373969	5.659222
AlSb	1.69	3.944220249	3.811865
Bi	0.013	0	0
Bi2Se3	0.335	1.934465146	0.656946
Bi2Te3	0.1505	0.943982922	0.192093
BiVO4	2.41	3.294451065	3.200583
BN	6.36	8.953226951	8.95429
BP	2.1	4.788188134	4.789955
C-diamond	5.5	7.118373389	7.118888
CdS	2.5	6.701660993	6.675914
CdSe	1.85	5.45311409	5.382224
CdTe	1.61	4.898797061	4.697798
Cu2O	2.17	1.123443247	1.04294
CuBr	3.07	3.2090892	3.138199
CuCl	3.4	3.092113273	2.999444
CuI	3.12	3.448086801	3.387592
CuSCN	3.94	3.003047021	2.943142
GaAs	1.52	3.647454949	3.624261
GaN-wurzite	3.503	6.16825876	6.166432
GaN-zb	3.28	6.024209898	6.019647
GaP	2.35	4.621390689	4.6196
GaSb	0.82	2.865154657	2.714133
Ge	0.744	2.954288499	2.896606
HgTe	0.3	2.025849437	1.809455
InAs	0.42	3.254537993	3.21201
InN	0.85	3.837907906	3.831206
InP	1.42	4.4834677	4.478054
InSb	0.24	2.622991805	2.462781
LiCl	9.4	12.35591188	12.338618
LiF	14.2	16.86451998	16.859917
MgO	7.83	11.88127083	11.873616

MgTe	3.6	7.218767495	7.0147
MoS2	1.29	1.556890205	1.551485
NaCl	8.595	15.98559308	15.95562
PbSe	0.155	2.018055746	0.698518
PbTe	0.19	1.738994537	0.127412
Sb2Te3	0.28	0.163467209	4.980084
Si	1.17	4.979322633	6.500223
SiC-2H	3.33	6.503279862	5.030011
SiC-3C	2.42	5.029553772	5.927159
SiC-4H	3.263	5.930395093	5.665455
SiC-6H	3.023	5.668720789	16.899565
SiO2-b	9.65	16.90109009	13.077674
SiO2-q	8.9	13.07841964	0.983171
SnSe	0.9	1.002117344	0.101534
SnTe	0.36	0.049171157	3.90159
SrTiO3	3.275	3.93691085	4.223413
TiO2- anatase	3.4	4.223927632	3.562821
TiO2- rutile	3.3	3.563184019	3.562822
ZnO	3.44	4.88267715	4.869865
ZnS	3.84	7.203186265	7.186458
ZnSe	2.82	5.898323454	5.855713831
ZnTe	2.39	5.03566194	4.85756847
Kendall_coef		0.657	0.528
ME		-2.269	-2.170
MAE		2.361	3.361
MAPE			446.271

Table C9: Full HSE06 results, Scalar calculations only.

Notes:

Pink highlight means SOC effects might be present, yellow that the simulation did not fully converge)

Not-converged systems not shown.

Crystal	avg exp	hse06
AgBr	2.71	1.903
AgCl	3.25	2.301
AgI	2.91	2.387
AlAs	2.23	1.970
AlP	2.5	2.200
AlSb	1.69	1.596
BN	6.36	5.689
BP	2.1	1.963
C-diamond	5.5	5.154
CdS	2.5	1.999
CdSe	1.85	1.347
CdTe	1.61	1.467
Cu2O	2.17	1.775

CuBr	3.07	1.977
CuCl	3.4	2.187
CuI	3.12	2.486
CuSCN	3.94	3.503
FeO	2.4	2.734
GaAs	1.52	1.195
GaP	2.35	2.143
GaSb	0.82	0.558
Ge	0.744	0.641
HgTe	0.3	0.044
InAs	0.42	0.300
InN	0.85	0.000
InP	1.42	1.321
LiCl	9.4	7.513
MgTe	3.6	3.014
MnO	3.1	0.000
MoS2	1.29	1.349
NaCl	8.595	6.318
NiO	4	0.554
PbSe	0.155	0.628
PbTe	0.19	1.054
Si	1.17	1.113
SiC-2H	3.33	3.267
SiC-3C	2.42	2.193
SiC-4H	3.263	3.077
SiC-6H	3.023	2.856
SiO2-b	9.65	7.486
SiO2-q	8.9	7.696
SnSe	0.9	1.036
SnTe	0.36	0.104
TiO2-ana- tase	3.4	3.401
TiO2-rutile	3.3	3.223
ZnO	3.44	2.359
ZnS	3.84	3.232
ZnSe	2.82	2.190
ZnTe	2.39	2.081
Kendall_coef		0.743
ME		0.565
MAE		0.641
MAPE		38.062

Table C10. Final calculations reduced data set.

crystal	avg exp	LDA	PBE_TZP_scalar_good small	BLYP_Scalar_TZP_small good	PW91_Scalar_TZP_small good
InSb	0.24	0	0.000	0.000	0.000
InAs	0.42	0	0.000	0.000	0.000
Ge	0.744	0	0.077	0.000	-0.002
InN	0.85	0	0.000	0.000	0.000
SnSe	0.9	0.5576 37	0.636	0.700	0.635
Si	1.17	0.4577 22	0.556	0.784	0.573
MoS2	1.29	0.8820 1	0.949	0.960	0.946
GaAs	1.52	0.2480 07	0.479	0.306	0.417
CdTe	1.61	0.5056 6	0.755	0.665	0.714
CdSe	1.85	0.3280 48	0.597	0.535	0.561
BP	2.1	1.1941 54	1.244	1.544	1.275
Cu2O	2.17	0.5213 24	0.524	0.646	0.541
AlAs	2.23	1.2172 55	1.298	1.560	1.318
GaP	2.35	1.4007 78	1.550	1.644	1.549
SiC-3C	2.42	1.2142 81	1.271	1.597	1.312
AlP	2.5	1.3509 66	1.475	1.745	1.494
CdS	2.5	0.8806 37	1.150	1.108	1.119
SiC-6H	3.023	2.0198 01	2.064	2.362	2.102
CuBr	3.07	0.2031 69	0.359	0.435	0.363
AgCl	3.25	0.6230 52	0.915	0.972	0.899
SiC-4H	3.263	2.2222 48	2.268	2.565	2.306
TiO2-rutile	3.3	1.6982 72	1.777	1.783	1.773
SiC-2H	3.33	2.3385 48	2.434	2.774	2.476
CuCl	3.4	0.3175 69	0.446	0.551	0.459
TiO2-anatase	3.4	1.8724 82	1.973	1.976	1.965
ZnO	3.44	1.0630 7	1.135	1.258	1.143
ZnS	3.84	1.8932 59	2.134	2.125	2.106
CuSCN	3.94	2.0268 98	2.223	2.244	2.240
C-diamond	5.5	4.0652 79	4.096	4.400	4.131

BN	6.36	4.256081	4.359	4.761	4.412
NaCl	8.595	4.656239	5.032	5.138	5.073
SiO2-q	8.9	5.717409	5.940	5.993	5.967
LiCl	9.4	5.879909	6.187	6.405	6.195
SIO2-b	9.65	5.65781	5.852	5.838	5.920

crystal	avg exp	LB94_scalar	TB-mBJ_scalar	KTB-mBJ_scalar	GLLB-SC_scalar	hse06_scalar
InSb	0.24	0.000	0.666938	0.615864	0	0.3248
InAs	0.42	0.000	0.881876	0.878338	0	0.29981
Ge	0.744	0.000	1.071532	1.044657	0	0.641037
InN	0.85	0.018	1.30947	1.577684	0.584667	0
SnSe	0.9	0.748	0.923006	0.920864	0.910055	1.035927
Si	1.17	0.310	1.451097	1.214767	0.791958	1.112706
MoS2	1.29	0.919	1.193026	1.190308	1.095049	1.348919
GaAs	1.52	0.000	1.829148	1.828945	0.115131	1.194808
CdTe	1.61	-0.030	2.039658	2.090096	0.929311	1.467075
CdSe	1.85	0.041	2.181827	2.367668	0.987104	1.3472
BP	2.1	1.139	2.290075	2.155451	1.766331	1.963276
Cu2O	2.17	1.139	0.934677	0.958358	0.803701	1.775369
AlAs	2.23	1.294	2.271714	2.179087	1.729417	1.97
GaP	2.35	0.629	2.418537	2.386165	1.794372	2.143166
SiC-3C	2.42	1.615	2.617512	2.561678	2.167155	2.193
AlP	2.5	1.551	2.600385	2.432117	1.893953	2.1997
CdS	2.5	0.740	2.948961	3.220361	1.814707	1.999276
SiC-6H	3.023	2.359	3.325867	3.269542	2.863233	2.855633
CuBr	3.07	1.510	1.649687	1.799928	1.344417	1.977348
AgCl	3.25	1.419	3.220282	3.839178	1.836427	2.301
SiC-4H	3.263	2.536	3.566253	3.509528	3.079596	3.076595
TiO2-rutile	3.3	1.475	2.569894	2.953346	2.553909	3.2229
SiC-2H	3.33	2.674	3.854618	3.791776	3.407986	3.267152
CuCl	3.4	1.894	1.76255	1.932749	1.745439	2.186876
TiO2-anatase	3.4	1.637	2.897967	3.407579	2.858569	3.400976
ZnO	3.44	2.013	2.860723	3.368303	2.278931	2.358571
ZnS	3.84	1.807	4.069094	4.281032	2.815363	3.231509
CuSCN	3.94	2.744	2.828285	2.877125	2.833426	3.502696
C-diamond	5.5	4.167	5.197762	5.259953	4.855018	5.154022
BN	6.36	4.841	6.28045	6.502306	5.576925	5.689448
NaCl	8.595	5.187	9.121081	12.49699	6.51996	6.318443
SiO2-q	8.9	6.747	9.474021	11.28095	7.369793	7.695646
LiCl	9.4	6.578	8.815707	9.451456	6.4931	7.512806
SIO2-b	9.65	6.953	11.68566	20.9208	8.07765	7.485594