
On choosing the best density functional approximation

Bartolomeo Civalleri,^a Davide Presti,^a Roberto Dovesi^a and Andreas Savin^{*b}

DOI: 10.1039/9781849734790-00168

1 Introduction

1.1 Aim of the paper

Density functional approximations have been very successful for describing properties of molecules and crystals when these depended on their electronic structure. However, a multitude of approximations exists, the values for the properties produced by them can differ significantly, and some criterion must be used in order to choose a given functional to predict a given property. Most often, density functional approximations are evaluated against some benchmark data sets, and quantities like the mean error and mean absolute error are indicated in order to assess the validity of the approximation. Moreover, in many approximations, the exchange-correlation functional may contain parameters that are determined by optimizing some statistical measures like those mentioned above on a training set of data. In such a case, the approximation itself is evidently dependent on the choice of the measure used. The aim of the present paper is to show that the “best approximation” can be a different one, depending upon the criterion used for defining the closeness of the approximation to the reference data.

1.2 Systems and approximations

Much of the early success of density functional theory is due to the success in the description of the band structure of the crystals. However, a quantitative agreement is still missing: all existing density functional approximations produce errors which are larger than the (assumed) experimental error bars. In this paper, we analyze whether statistical descriptors of the errors in the calculated band gaps allow choosing the best approximation. It is shown that different measures recommend different approximations. We thus conclude that faced with such a dilemma, in order to make a decision, one is coerced into using some additional, external criterion.

The band gaps considered cover two orders of magnitude, between ≈ 0.2 and ≈ 20 eV. They were obtained for the following crystals (*Strukturbericht* designation¹ in parentheses):

- 34 semiconductors, also present in the SC40 data set,² namely: C(A4), Si(A4), Ge(A4), SiC(B3), BN(B3), BP(B3), BAs(B3), BSb(B3), AlN(B3),

^aDepartment of Chemistry and NIS Centre of Excellence, Via P. Giuria 7, 10129 University of Torino, Torino, Italy

^bLaboratoire de Chimie Théorique, CNRS and UPMC Sorbonne University, 4 place Jussieu, 75252 Paris, France. E-mail: andreas.savin@lct.jussieu.fr

AlN(B4), AlP(B3), AlAs(B3), AlSb(B3), GaN(B3), GaN(B4), GaP(B3), GaAs(B3), GaSb(B3), InN(B4), InP(B3), InAs(B3), InSb(B3), ZnS(B3), ZnSe(B3), ZnTe(B3), CdS(B3), CdSe(B3), CdTe(B3), MgS(B1), MgSe(B1), MgTe(B1), BaS(B1), BaSe(B1), BaTe(B1);

- LiH(B1) and alkali halides (*i.e.* LiF(B1), LiCl(B1), NaF(B1) and NaCl(B1));

- oxides, *i.e.*, MgO(B1), α -SiO₂(C8), ZnO(B4), SrTiO₃(E2₁);

- solid Ne(A1) and Ar(A1).

The values for the experimental (fundamental) band gaps are taken from references.²⁻¹⁰

The approximations used can be classified into the following groups:

- Hartree-Fock
- Local and semi-local density functionals
 - LDA (local density approximation, in the parametrization of Slater for exchange¹¹ and Vosko, Wilk and Nusair for correlation¹²),
 - PBE (Perdew, Burke and Ernzerhof¹³ exchange and correlation functional)
 - PBEsol (PBE functional revised for solids¹⁴)
- linear global hybrids (where the density functional exchange is mixed up linearly with the Hartree-Fock exchange)
 - B3LYP (Becke88¹⁵ for exchange and Lee, Yang and Parr¹⁶ for correlation, with mixing parameters from¹⁷; in contrast to the generally used B3LYP, we used the variant implemented in the CRYSTAL code, where the local functional is fitted to the accurate correlation energy of the uniform electron gas, *i.e.* VWN5,¹² and not to the random phase approximation of it, *i.e.*, VWN3 and not VWN3)
 - PBE0 (PBE functionals for exchange and correlation, with 25 per cent admixture of Hartree-Fock^{13,18})
 - PBEsol0 (as PBE0, replacing PBE by PBEsol)
 - B97 (Becke's 1997 hybrid functional for exchange and correlation,^{19,20} using 21 per cent of Hartree-Fock)
- long-range hybrids (where the long-range part of the exchange is described by Hartree-Fock)
 - RSHX LDA (range-separated hybrid with LDA exchange²¹⁻²³ and separation parameter $\omega = 0.40$ a.u. as proposed for solids²⁴)
 - LC- ω PBE (using PBE exchange and separation parameter $\omega = 0.40$ a.u.^{25,26})
 - LC- ω PBEsol (using PBEsol exchange and separation parameter $\omega = 0.60$ a.u.²⁵)
 - ω B97 (using B97 exchange and separation parameter $\omega = 0.40$ a.u.²⁷)
- mixture of long-range and linear hybrid
 - ω B97-X (Long-range and partly short-range ($c_{SR} = 0.157706$) corrected with Hartree-Fock, and using B97 exchange and separation parameter $\omega = 0.30$ a.u.²⁷)
- mixture of middle-range hybrid and linear hybrid
 - HISS (Henderson, Izmaylov, Scuseria, Savin,^{28,29} with range-separation parameters $\omega_{SR} = 0.84$ and $\omega_{LR} = 0.20$ a.u., and 2/5 of middle-range PBE, besides the short- and long-range part)

- mixture of short-range hybrid with linear hybrid
 - HSE06 (Heyd, Scuseria and Ernzerhof^{30,31} functional with range separation parameter $\omega = 0.11$ a.u. and linear parameter $c_{SR} = 0.25$)
 - HSEsol (as HSE06, with the PBE mode for the exchange hole³² replaced by that of PBEsol²⁵ functional and a range separation parameter $\omega = 0.11$ a.u.³³)

Some recent, successful approximations such as,³⁴ or³⁵ were not part of the test set.

The reference (experimental) and the calculated data are given in reference.³⁶

1.3 A visual test

In order to take a glimpse of the problematic, let us look at two typical plots of the calculated *vs* experimental band gaps. In Fig. 1 we see on the left graph that the method used has a systematic error for the whole range of band gaps. The graph on the right shows small errors for small band gaps and large errors for large gaps. If one is interested in small gaps only, one certainly should prefer the latter method. For the whole range of band gaps, however, it seems easier to correct for the systematic error, present in the former approximation. By the way, the left graph presents results obtained with the Hartree-Fock method, long time rejected for the computation of band gaps, while the right graph presents results obtained with the LDA method, considered useful for many years for the computation of band gaps.

The errors produced are shown in Fig. 2 for four different approximations (LDA, the gradient-corrected PBE, Hartree-Fock and the hybrid PBE0), for all systems considered in this paper, ordered according to the experimental band gaps. We can see that the errors of LDA become larger as the band gaps get bigger. The same trend is also present in PBE. It is somewhat reduced, but not eliminated, with the hybrid method, PBE0.

1.4 Overview of the paper

In the following, we will first present the results for some classical measures: the mean error, the variance, and the mean absolute error. Next we will

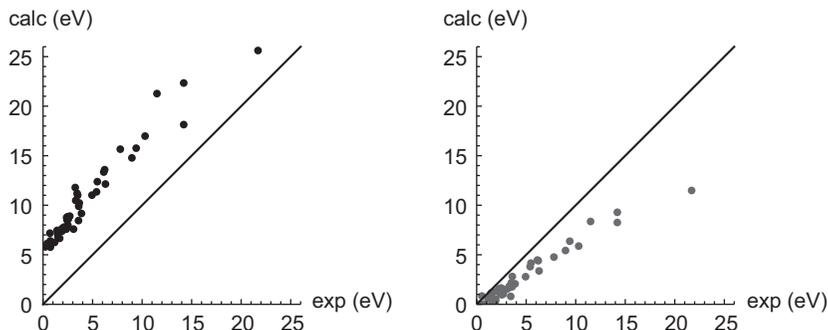


Fig. 1 Calculated *vs* experimental band gaps, for two different methods. The graph on the left shows a systematic error, that on the right seems good for small gaps, but becomes bad for large gaps.

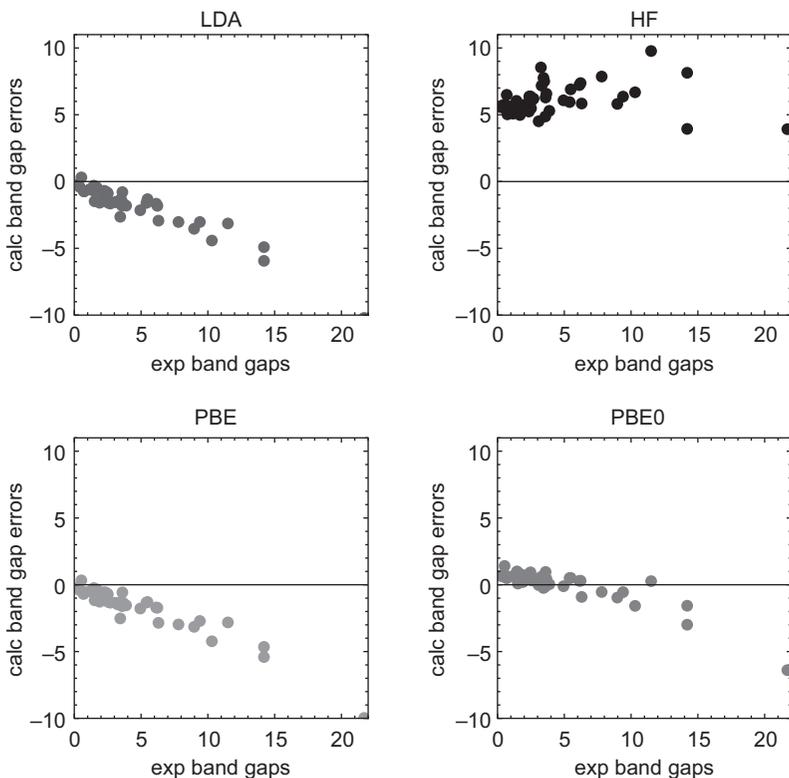


Fig. 2 Approximation errors vs experimental band gaps, for LDA (top left), the generalized gradient approximation PBE (bottom left), Hartree-Fock (top right), and the hybrid PBE0 (bottom right). Here and in the following figures all non adimensional data are in eV.

consider measures used in robust statistics, such as medians, or interquartile ranges. The following section deals with trends as given by rankings. Later, we will comment on the choice and on the reliability of experimental data, and ask whether the errors produced are due to the approximations, or to the inability of orbital energy differences to reproduce fundamental band gaps (even using exact density functionals).

2 Classical measures

2.1 Mean error

Let the measured data be $y_{i,exp}$ and the calculated ones be $y_{i,calc}$, the index i corresponding to one of $M = 45$ crystals under consideration. The mean error is defined as

$$\mu = \frac{1}{M} \sum_{i=1}^M (y_{i,calc} - y_{i,exp}) \quad (1)$$

The best method is that for which the mean error is closest to 0. The winner is PBEsol0, cf. Fig. 3, with the mean error of -0.04 eV. However, PBE0 and HISS are not far from it, with 0.06 and 0.07 eV respectively. The other methods have mean errors larger than 0.1 eV.

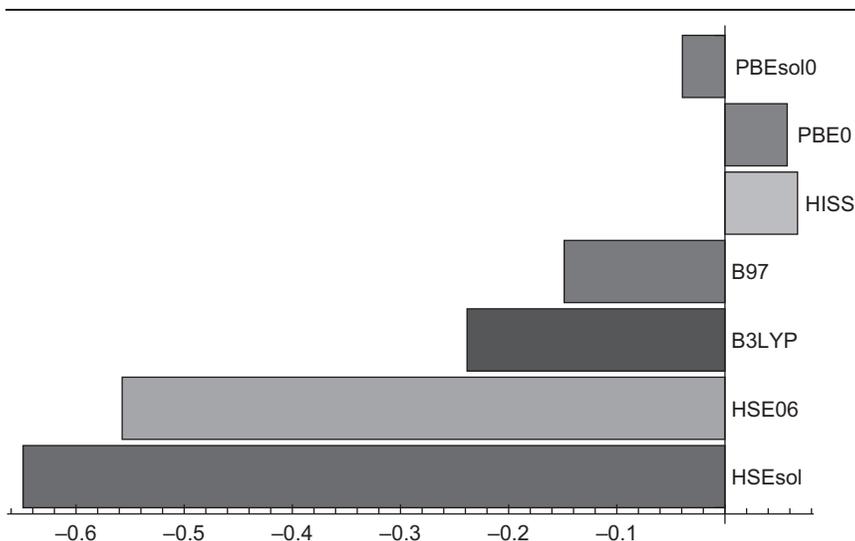


Fig. 3 Mean errors, μ , eq 1, for different methods; shown only for $|\mu| < 1$ eV.

Having a mean error equal to 0 is a necessary, but not a sufficient condition for having the “perfect” approximation. In Fig. 2 we can see in the panel corresponding to PBE0 that the mean close to zero is obtained by having errors of different signs. (Please notice that with our choice of data there are more systems with small band gaps, producing a shift towards positive errors not recognizable on the figure at the first glance.)

2.2 Variance

In order to have the “perfect” approximation, not only the mean should be zero, but the variance should be zero, too. We will use the variance defined by:

$$\sigma^2 = \frac{1}{M} \sum_{i=1}^M (\Delta y_i - \mu)^2 \quad (2)$$

where $\Delta y_i = y_{i,calc} - y_{i,exp}$, and μ is given by eq 1. In the following we will concentrate on its square root, σ .

We see (Fig. 4) that the winner of the “mean error contest” is not among the method having the smallest variance. At the same time, the winner of the “ σ contest” (LC- ω PBEsol) is not among the methods yielding the best mean error. Furthermore, we notice that even the method with the smallest σ still yields the quite large value of $\sigma \approx 1$ eV.

2.3 Mean absolute error

A compromise between the two preceding measures is the mean absolute error,

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |y_{i,calc} - y_{i,exp}| \quad (3)$$

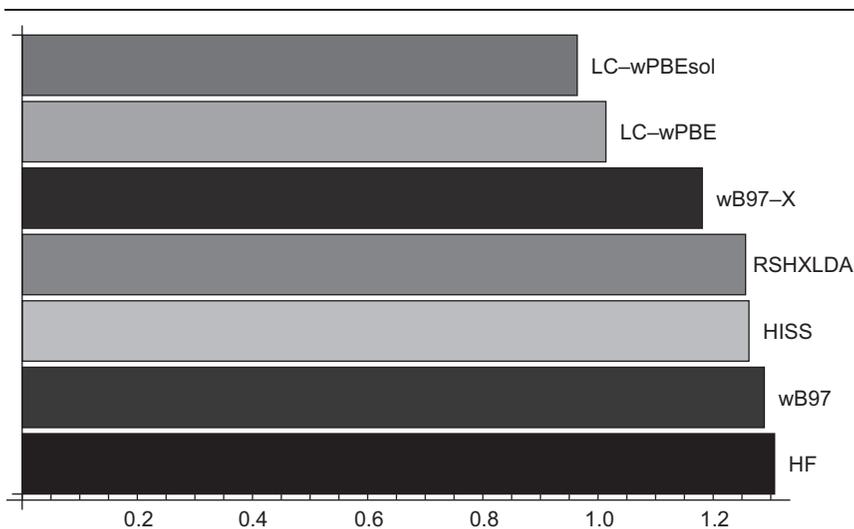


Fig. 4 σ , the square root of the variance, σ^2 , eq 2, for different methods; shown only when $\sigma < 1.5$ eV.

It is zero for the “perfect” approximation. However, a method producing a systematic error, *i.e.*, $y_{i,calc} = y_{i,exp} + constant$, is easy to correct by simply subtracting the constant. It has zero variance, but non-zero mean, and would produce a large MAE, despite of showing perfectly the trends.

The best method, based on the MAE is HISS, with a value of 0.7 eV, followed closely by B97, HSE06, PBEsol0, and B3LYP (see Fig. 5).

3 Robust statistics

3.1 Outliers

We can notice in Fig. 2 that there are only a few points producing large errors. In statistics they are called outliers, observations which deviate markedly from other observations.³⁷

Sometimes the existence of outliers can be simply due to some trivial (like copying) mistakes. They can also be due to mistakes which are more difficult to detect, *e.g.*, to situations where the technique used is not recommended. In these cases, the outliers should be eliminated. However, there are situations when outliers cannot be ignored: it is known in statistics that seldom events can happen. Outliers may denature our judgment on the data analyzed, by giving a too important weight to rare situations. Thus, it is not clear how to behave when outliers show up. Should they be weighted less, or even ignored? In *robust* statistics, some measures are used that are less sensitive to outliers, and some of them will be applied below.

3.2 Definitions from robust statistics

Let us order the data (our errors) in increasing order and split them into four quarters: the first contains the smallest 25 per cent of the data, the first

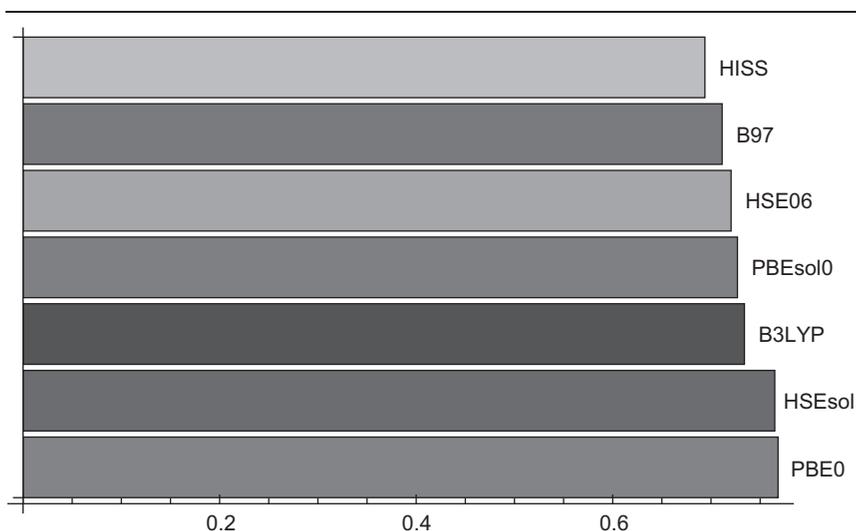


Fig. 5 Mean absolute errors, MAE, for different methods; shown only when smaller than 1 eV.

two contain 50 per cent of the data *etc.* The three values that produce that division are called quartiles, the second one being called median. The distance from the first to the third value is the interquartile range. For the normal distribution, the mean is equal to the median, and the interquartile range to $\approx 0.6745\sigma$; for a sufficiently large set of data following the normal distribution, the “best” method would be the same whether the mean or the median is chosen, the variance or the interquartile range is chosen.

That the robust measures are less sensitive to outliers than the classical measures discussed before can be seen in a simple example. Suppose that the data we consider are the integers from -6 to 6 , excluding 0 . (These numbers are chosen only for a simple illustration of the principle, and, in particular, do not follow a normal distribution.) The mean is 0 , and $\sigma \approx 3.89$. The quartiles are -3.5 , 0 , and 3.5 , yielding a median equal to 0 and the interquartile range of 7 . Imagine now that by mistake 6 is replaced by 16 . The mean becomes 0.83 , and $\sigma \approx 5.73$. However, median and interquartile range are unchanged.

3.3 Medians

We will consider below the medians of the errors $y_{i,calc} - y_{i,exp}$ obtained for different methods (cf. Fig. 6).

The best median defines a new winner, B3LYP. It is nearly two times better than that of the next candidate, B97.

3.4 Interquartile ranges

A zero median does not guarantee that the method is free of error, in the same way as when looking at the mean error only. Let us now look at the interquartile ranges (Fig. 7).

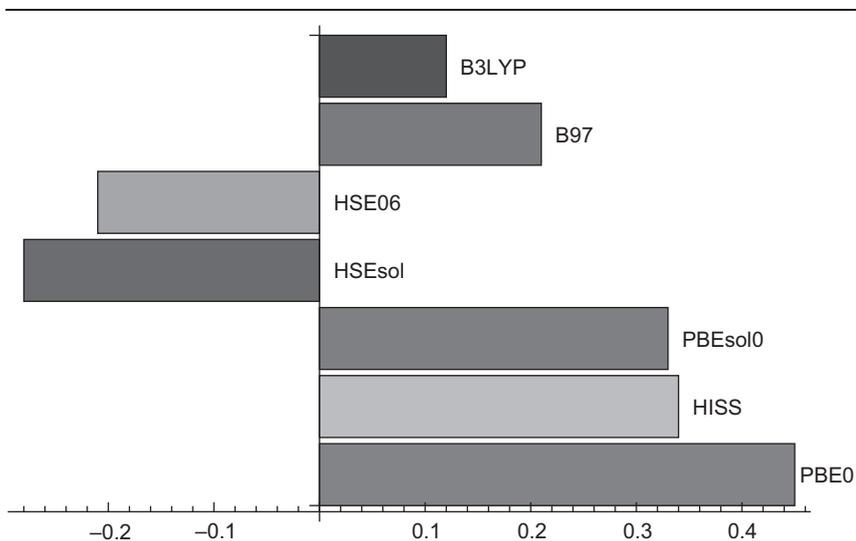


Fig. 6 Medians of the errors for different methods; shown only when smaller than 0.5 eV, in absolute value.

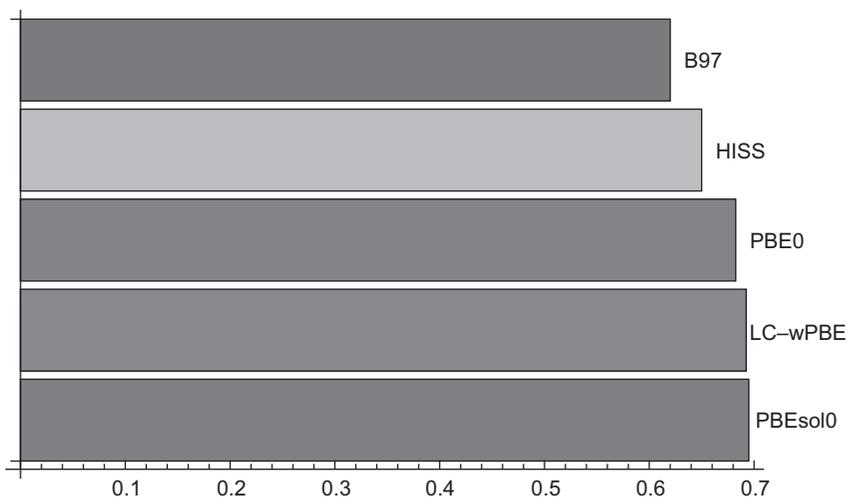


Fig. 7 Interquartile ranges of the errors for different methods; shown only when smaller than 0.7 eV.

Now, the best results are obtained for B97, although, the best value is relatively large (≈ 0.6 eV). HISS, PBE0, LC- ω PBE, PBE0, PBEsol0 are only less than 0.1 eV worse.

3.5 Median absolute deviations

Another robust measure is given by the median of the absolute deviations from the median. When all absolute deviations from the median are zero, and the median is zero we have a perfect fit. By now, it will not be surprising

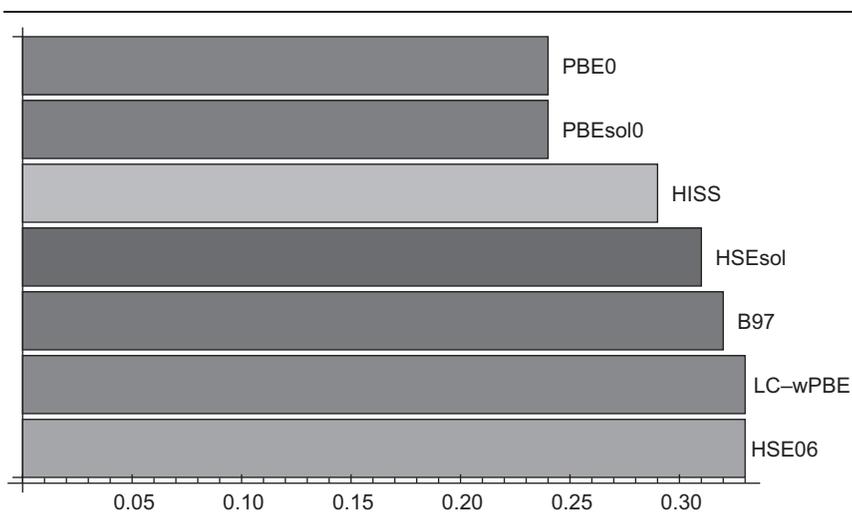


Fig. 8 Median absolute deviations for different methods; shown only when smaller than 0.4 eV.

that we do not have a perfect fit. The best median absolute deviation is obtained for PBE0 and PBEsol0 (0.24 eV, see Fig. 8).

4 Trends

4.1 Linear model fit

As none of the methods fits the experimental data (within some assumed experimental errors of the order of 0.1 eV), one can check whether trends are correctly followed. This might help to choose the “best” method. One may define a trend by using a linear fit

$$y_{i,model} = ay_{i,calc} + b \quad (4)$$

Ideally, the intercept b should be equal to zero, and the slope a equal to one.

All winners up to now were hybrid density functionals. This is not the case for the linear model fit. Figure 9 shows that the best intercept is obtained for a generalized gradient approximation, PBE. PBEsol and even LDA are not much worse, as they produce intercepts of ≈ 0.1 eV. The best slope is not obtained for a density functional: Hartree-Fock is the method providing the slope closest to unity (see Fig. 10).

It may be worth to notice that PBE, PBEsol and LDA are the methods having the largest median absolute deviation from the median, being 2–3 times worse than the winner, PBE0. Furthermore, Hartree-Fock is the method yielding the worst mean absolute error (6.1 eV, ≈ 1 eV larger than the worst method containing a density functional approximation).

Another way of looking at the data is to analyze the (Pearson’s) correlation coefficients, defined by:

$$r = \frac{\sum_i (y_{i,calc} - \mu_{calc})(y_{i,exp} - \mu_{exp})}{\sqrt{\sum_i (y_{i,calc} - \mu_{calc})^2} \sqrt{\sum_i (y_{i,exp} - \mu_{exp})^2}} \quad (5)$$

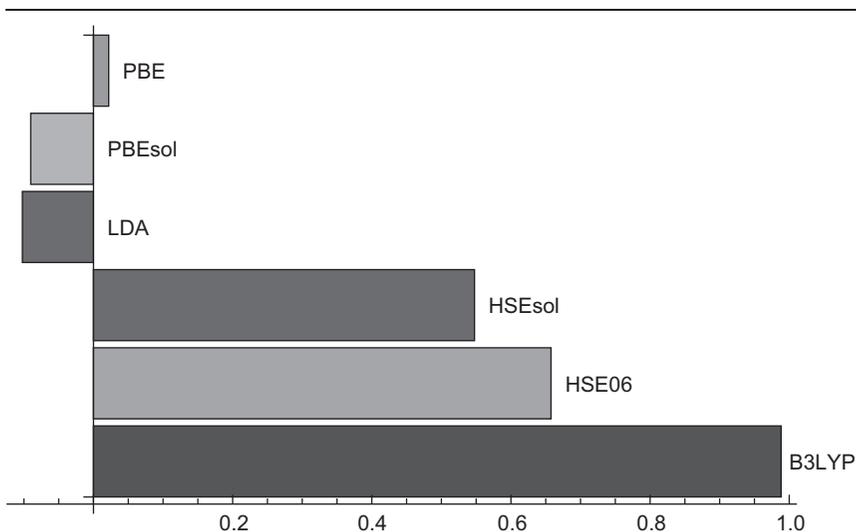


Fig. 9 Intercepts obtained from a linear model fit; shown only when smaller than 1 eV.

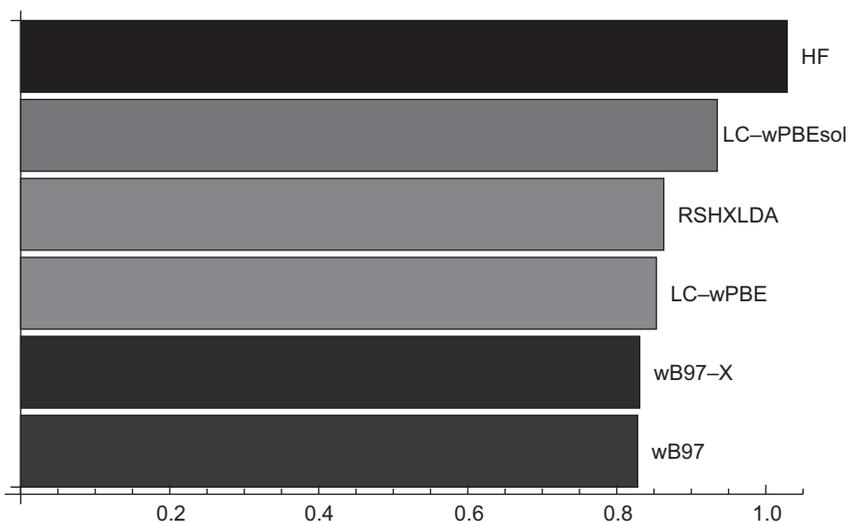


Fig. 10 Slopes obtained from a linear model fit; shown only when not deviating from 1 by more than 0.2.

where μ_{calc} is the mean value of the $y_{i,calc}$ and μ_{exp} the mean value of the $y_{i,exp}$. For the highest possible value for r , $r = 1$, the calculated band gaps increase linearly with the experimental band gaps. If the minimal value $r = -1$ were reached, a linear relationship exist, but the calculated band gaps would decrease when the experimental band gaps increase. We did not choose any “winner” for this criterion: r lies between ≈ 0.97 and ≈ 0.98 for all methods considered.

4.2 Ranking

Let us abandon the linear model, and measure how well the calculated band gaps are ordered. This can be seen as being closer to what an experienced

user of approximate methods would do, recognize whether a calculation shows the same trends as the exact data. The user will compare the results with similar systems, and, if the ordering is correct, will correctly know how to intercalate the calculated number between the already known ones.

To measure how well ordering is reproduced, we will use Kendall's rank correlation coefficient. In order to construct it, we first define the pairs $(y_{i,exp}, y_{i,calc})$. For two pairs, i and j , if $y_{i,calc} > y_{j,calc}$ when $y_{i,exp} > y_{j,exp}$ or if $y_{i,calc} < y_{j,calc}$ when $y_{i,exp} < y_{j,exp}$ the pairs are said to be concordant. If $y_{i,calc} > y_{j,calc}$ when $y_{i,exp} < y_{j,exp}$ or if $y_{i,calc} < y_{j,calc}$ when $y_{i,exp} > y_{j,exp}$ the pairs are said to be discordant. Let n_c be the number of concordant pairs, and n_d that of the discordant pairs. The total number of pairs (of systems considered) is $n_p = M(M - 1)/2$. Kendall's correlation coefficient is defined by:

$$\tau = \frac{n_c - n_d}{n_p} \quad (6)$$

Like r , equation 5, τ lies between 1 and -1 . When $\tau = 1$, all the pairs are concordant, the calculated band gaps follow the trend of the experimental band gaps. The spread of the τ is slightly larger than the spread of the r , yielding values between ≈ 0.90 and ≈ 0.83 . The value of τ decreases from HSE06, HSEsol, HISS, PBE0, LC- ω PBE, but is, for all these methods ≈ 0.90 .

We can extend the definition above, considering concordance only if the difference between the y_i and y_j exceeds what one might consider the experimental accuracy of 0.1 eV. In this case, only HISS keeps a correlation coefficient of 0.90, but several other methods follow with a correlation coefficient of 0.89. The worst correlation coefficient is 0.82, for LDA.

5 Reference data

5.1 Selecting data of interest

When calibrating or judging an approximation, something may be gained by choosing subsets of data for which the approximation is supposed to work. For example, one can see in Fig. 2 that the errors are much smaller for LDA, PBE and PBE0 when only systems with small band gaps are considered. In such a case, one may say that "density functional approximations should not be used when the band gaps are large". One might even argue that one does not even need more, as band gaps are of interest mostly in semiconductors. Let us suppose that we are only interested in solids having relatively large band gaps; we choose the range between 1.7 eV (visible light) to 6.5 eV (medium UV), and analyze the errors in this range.

The best mean absolute error (MAE) decreases to 0.34 eV, result obtained for HSE06 and not for HISS as for the full data set. However, other functionals follow closely, as seen in Fig. 11.

Splitting a data set into subsets may help to identify the cases when a method works, and when it does not. However, splitting may mislead in judging the accuracy. Imagine that two papers present the mean absolute errors for the band gaps for the same set of crystals. The authors of the two papers decide to split the systems into two subgroups, but the subgroups are

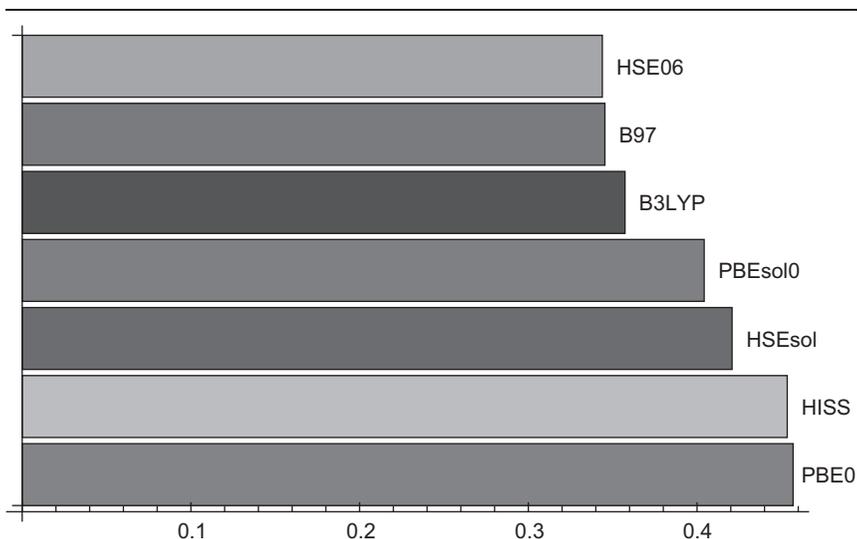


Fig. 11 Mean absolute errors for systems with band gaps between 1.7 and 6.5 eV. Results for methods having MAEs larger than 0.5 eV are not shown.

defined differently in the two papers. The errors are smaller for both subgroups in the second paper. Does this mean that the method used in the second paper is better? Not necessarily. Suppose that the first paper presents the band gaps divided into those with band gaps smaller and larger than 5 eV, and obtains the mean absolute errors of 0.54 and 1.34 eV, respectively. The second paper chooses one group to be constituted by the “useful cases” above (band gaps between 1.7 and 6.5 eV), the other being formed by the remaining systems. It obtains for both groups smaller mean absolute errors (0.46 and 1.19 eV, respectively). However, these figures were obtained from our data, with the *same* method, PBE0. To see how this can happen, consider the following simple example. For four systems the following errors were obtained (in arbitrary units): 1, 3, 8, 10. We now split the set into two sets, one with small errors, 1, 3, and one with large errors 8, 10. The MAEs are 2, and 9, respectively. Let us now split the data set into two different sets, 1, and 3, 8, 10, yielding the MAEs 1 and 7, respectively. (The largest error from the set with small error was moved into the set with large errors.) The second way of splitting produced lower MAEs, but we have not changed the data. One may find that a grouping into small and larger errors (1, 3, and 8, 10, respectively) is better than the one that puts one of the systems showing small errors into the group of systems presenting large errors.

5.2 Reliability of reference data

We have considered that the experimental have error bars too small to be taken into consideration for the comparison with calculated data. Even more importantly, we have considered that the calculation should reproduce some “ideal” measurement, corresponding to the fundamental band gap, which can be defined as the difference between the ionization potential

I and the electron affinity, A of the system (see, e.g.,^{38–40}). However, the experimental data do not always correspond to this. Let us take Si as an example. There are not many systems for which the purity of the material and the measurement of the band gap had the same importance as for Si. Let us consider the paper by R. Pässler⁴ which provides fundamental band gaps for a series of crystals. We see in his Fig. 3 that the experimental data seem to be very accurate, and that the model used to extrapolate to 0 K seems reliable to meV, *i.e.*, by orders of magnitude more accurate than the calculations discussed here. The experimental data used in ref.⁴ are taken from a different paper, by Bludau *et al.*⁴¹

In the latter paper it is stated: *The band-gap energy E_g of silicon has been re-evaluated with high precision between 2 and 300 K by the following method: the derivative of the absorption coefficient, resulting from free-exciton absorption, ...* Is this not the optical gap? In fact, Bludau *et al.* discuss this point. They write: *The subsequent analysis then involves determination of E_g the band-gap energy, from $E_g = E_{gx} + E_{xb} = hv_m + E_{xb} - \hbar\omega_{TO}$ where E_{xb} is the free-exciton binding energy. We have taken $E_{xb} = 14.7 \pm 0.4$ meV and $\omega_{TO} = 57.3 \pm 0.4$ meV from the work of Shaklee and Nahory.* They thus find that in order to obtain the fundamental gap, E_g , from the optical excitation energy, hv_m , one has to add two correcting terms. One is due to lattice vibrations (it is an indirect gap), $\hbar\omega_{TO}$. The second one, is called free-exciton binding energy, E_{xb} . Both corrections are taken from ref.⁴² where it is written: *We expect, on the basis of effective mass theory, that the exciton excited states should form a hydrogen series. ... We can then calculate, on the basis of a hydrogenic series, an exciton Rydberg of $R = 14.7 \pm 0.4$ meV.* The corrections are given to meV accuracy, but the experimental data do not correspond to the assumed difference between the ionization potential and the electron affinity. Thus, even a calculation yielding the exact $I - A$ would not coincide with the experimental result.

6 Conclusion

To summarize the situation, we show in Fig. 12 the number of calculated band gaps lying within a given interval Δ from the measured band gaps. We see that for most crystals the errors of the approximations are large, and outside of what one can expect as error bars (around 0.1 eV). The approximations are thus models in which we cannot be confident.

The ideal method should compute for the fundamental gap the energy difference between the ionization potential and the electron affinity of the system. However, we know that the exact Kohn-Sham method will not provide it.^{39,40}

We also know that the difference of Hartree-Fock orbital energies does not exactly yield the desired quantity. Hybrid methods could provide the fundamental gap, in principle^{43,44}: the exact functional ensures for the highest occupied orbitals to behave correctly asymptotically, and thus to yield the correct ionization potential, and the mixture with Hartree-Fock can be used to produce a value for the lowest unoccupied orbital energy corresponding to the electron affinity. However, at present, we do not know

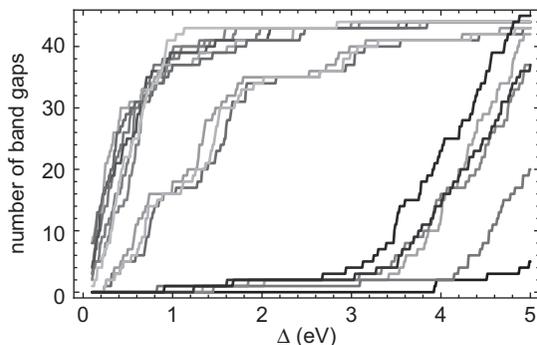


Fig. 12 Number of calculated band gaps within an interval Δ around the experimental band gaps, as a function of Δ , for different methods. The group of data making a jump at smaller Δ corresponds to linear, short-range, and middle range hybrids; the group in the middle is made by pure Kohn-Sham approximations, while that with largest errors corresponds to long-range hybrids and Hartree-Fock.

of any universal definition of a hybrid which will provide the exact fundamental gaps, even assuming the associated exact density functional known. Hybrids can be made system-dependent, and the hybrid tuned to give the correct gap^(45–47); for a recent review see⁴⁸⁾.

Even if we obtain the ideal method, we do not automatically get a good agreement with experimental data, because these, as the Si example shows, do not necessarily correspond to what the calculation aims to provide.

Applying measures from the statistical analysis of data does not permit to select one of the methods. The situation is summarized below. The best result is obtained for

- the mean error with PBEsol0,
- σ with LC- ω PBEsol,
- the mean absolute error with HISS,
- the median with B3LYP,
- the interquartile range with B97,
- the median absolute deviation with PBE0,
- the intercept (linear model fit): with PBE,
- the slope (linear model fit): Hartree-Fock,
- Pearson's correlation coefficient: HSEsol,
- Kendall's rank correlation coefficient: HSE06.

Thus, a method can be chosen only by introducing some supplementary, external criterion. For example, the user can decide whether his/her own experience with a given method is more valuable than some improvement obtained for one of the statistical measures.

One can try to reduce the range of application of a given approximation to a class of systems where it works well. However, even this enterprise can be risky, as it is difficult to define subsets for which the method is supposed to work.

Of course, one can choose some functional based on analyzing several criteria, weighting their quality, and choose one, *e.g.*, HSE06 (that yields small mean absolute errors for small band gaps), HISS (as recommended by

the recent paper of Lucero *et al.*⁴⁹), or another hybrid like PBE0, *etc.* Personal experience with a given method also plays a role, and may make some user decide to use B3LYP, although this method is not “better” than other hybrids in some statistical sense.

Appendix

The numbers used in this paper are compactly represented in Fig. 13. The directions correspond to different measures, *i.e.*,

- absolute value of the mean error ($|ME|$)
- square root of the variance (σ)
- mean absolute error (MAE)
- absolute value of the median of the errors ($|MED|$)
- interquartile range (IQR)
- median absolute deviation (MAD)

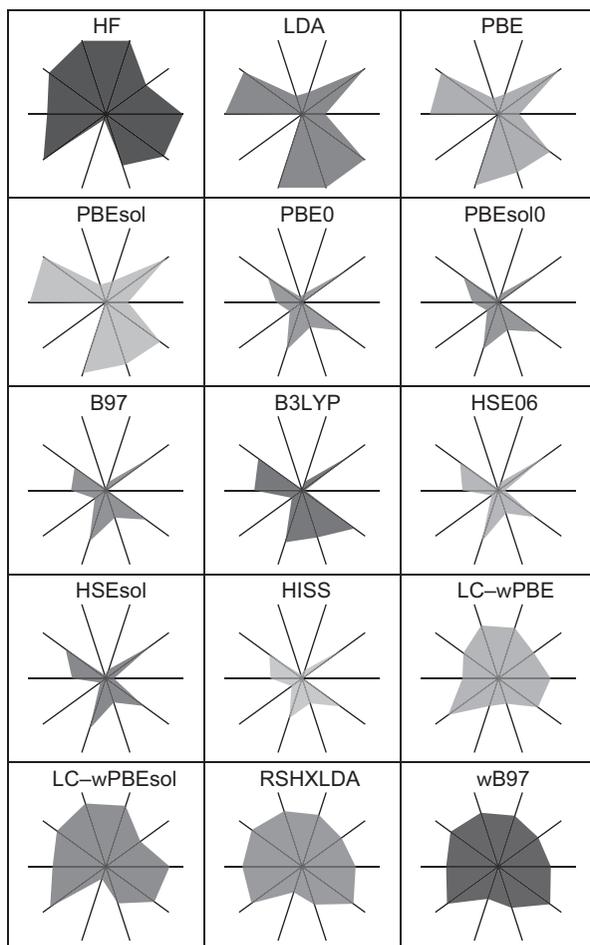


Fig. 13 Number of calculated band gaps within an interval Δ around the experimental band gaps, as a function of Δ , for different methods.

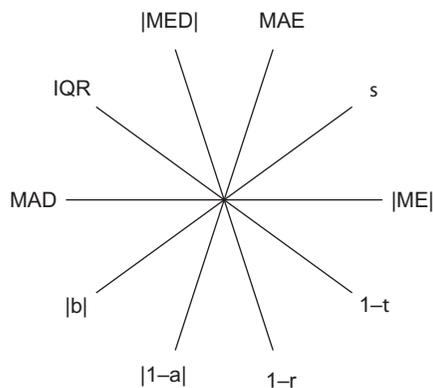


Fig. 14 Convention used for representing different measures in Fig. 13.

- absolute value of the intercept of the linear fit ($|b|$)
- the absolute value of one minus the slope of the linear fit ($|1 - a|$)
- one minus Pearson's correlation coefficient ($1 - r$)
- one minus Kendall's rank correlation coefficient ($1 - \tau$)

in trigonometric order (counterclockwise, cf. Fig. 14). Figure 13 shows the ratios between the measures (*e.g.*, MAE) obtained for a given method, *e.g.*, LDA, and the maximal value for that measure (*e.g.*, MAE) among all methods (Hartree-Fock, LDA, *etc.*). The latter are, for

- $|ME|$, 6.12 eV
- σ , 1.77 eV
- for MAE, 6.12 eV
- for $|MED|$, 5.95 eV
- for IQR, 1.24 eV
- for MAD, 0.71 eV
- for $|b|$, 6.00 eV
- for $|1 - a|$, 0.40
- for $1 - r$, 0.05
- for $1 - \tau$, 0.16

All the measures above would be zero for the ideal fit. The polygons ("blobs") shown in Fig. 13 are constructed by connecting the points on the stars corresponding to the ratios of the different measures. Thus, the larger the "blobs", the "worse" the method is. Notice, however, that often systematic errors can be easily dealt with, and that sometimes a large deviation from a zero error may be not impede the validity of the calculation.

Acknowledgment

The authors are grateful for enlightening discussions to Jerome and Ora Percus (NYU), and to Pascal Pernot (Université Paris-Sud). They also thank Gustavo Scuseria (Rice University) for comments on the manuscript.

References

- 1 *Strukturbericht* designation is taken from the Crystal Lattice Structures web page: <http://cst-www.nrl.navy.mil/lattice/struk/index.html>.
- 2 J. Heyd, J. E. Peralta, G. E. Scuseria and R. L. Martin, *J. Chem. Phys.*, 2005, **123**, 174101.
- 3 O. Madelung, *Semiconductors: Data Handbook*, Springer-Verlag Berlin Heidelberg, New York, 3rd ed., 2004.
- 4 R. Pässler, *Phys. Status Solidi B-Basic Res.*, 1999, **216**, 975.
- 5 D. M. Roessler and W. C. Walker, *Phys. Rev.*, 1968, **166**, 599.
- 6 A. M. Fox, *Optical Properties of Solids*, Oxford University Press, 2001.
- 7 J. Wrobel, K. J. Kurzydowski, K. Hummer, G. Kresse and J. Piechota, *Phys. Rev. B*, 2009, **80**, 155124.
- 8 F. El-Mellouhi, E. N. Brothers, M. J. Lucero and G. E. Scuseria, *Phys. Rev. B*, 2011, **84**, 155122.
- 9 M. A. L. Marques, J. Vidal, M. J. T. Oliveira, L. Reining and S. Botti, *Phys. Rev. B*, 2011, **83**, 035119.
- 10 F. Tran and P. Blaha, *Phys. Rev. Lett.*, 2009, **102**, 226401.
- 11 J. C. Slater, *Phys. Rev.*, 1951, **81**, 385.
- 12 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200.
- 13 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **38**, 3865.
- 14 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.*, 2008, **100**, 184109.
- 15 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.
- 16 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
- 17 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623.
- 18 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 5168.
- 19 A. D. Becke, *J. Chem. Phys.*, 1997, **107**, 8544.
- 20 F. A. Hamprecht, A. J. Cohen, D. J. Tozer and N. C. Handy, *J. Chem. Phys.*, 1998, **109**, 6264.
- 21 A. Savin, in *Recent development and applications of Density Functional Theory*, ed. J. Seminario, Elsevier, Amsterdam, 1996, p. 327.
- 22 P. M. W. Gill, R. D. Adamson and J. A. Pople, *Mol. Phys.*, 1996, **88**, 1005.
- 23 A. Savin, J. Toulouse and H. J. Flad, *Int. J. Quantum Chem.*, 2004, **100**, 1047.
- 24 I. C. Gerber, J. G. Angyan, M. Marsman and G. Kresse, *J. Chem. Phys.*, 2007, **127**, 054101.
- 25 E. Weintraub, T. M. Henderson and G. E. Scuseria, *J. Chem. Theory Comput.*, 2009, **5**, 754.
- 26 O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 234109.
- 27 S. Jeng-Da Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 084106.
- 28 T. M. Henderson, A. F. Izmaylov, G. E. Scuseria and A. Savin, *J. Chem. Phys.*, 2007, **127**, 221103.
- 29 T. M. Henderson, A. F. Izmaylov, G. E. Scuseria and A. Savin, *J. Chem. Theory Comput.*, 2008, **4**, 1254.
- 30 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207.
- 31 A. V. Krukau, O. A. Vydrov, A. F. Izmaylov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 224106.
- 32 T. M. Henderson, B. G. Janesko and G. E. Scuseria, *J. Chem. Phys.*, 2008, **128**, 194105.
- 33 L. Schimka, J. Harl and G. Kresse, *J. Chem. Phys.*, 2011, **134**, 024116.
- 34 J.-W. Song, K. Yamashita and K. Hirao, *J. Chem. Phys.*, 2011, **135**, 071103.
- 35 R. Peverati and D. G. Truhlar, *J. Chem. Phys.*, 2012, **136**, 134704.

-
- 36 D. Presti, B. Civalleri, M. Ferrabone, R. Dovesi and A. Savin, 2012, in preparation.
 - 37 NIST/SEMATECH e-Handbook of Statistical Methods: <http://www.itl.nist.gov/div898/handbook/>.
 - 38 N. F. Mott, *Rev. Mod. Phys.*, 1968, **40**, 677.
 - 39 J. P. Perdew and M. Levy, *Phys. Rev. Lett.*, 1983, **51**, 1884.
 - 40 L. J. Sham and M. Schluter, *Phys. Rev. Lett.*, 1983, **51**, 1888.
 - 41 W. Bludau, A. Onton and W. Heinke, *J. Appl. Phys.*, 1974, **45**, 1846.
 - 42 K. L. Shaklee and R. E. Nahory, *Phys. Rev. Lett.*, 1970, **24**, 942.
 - 43 A. J. Cohen, P. Mori-Sanchez and W. Yang, *Phys. Rev. B*, 2008, **77**, 115123.
 - 44 T. Tsuneda, J.-W. Song and K. Hirao, *J. Chem. Phys.*, 2010, **133**, 174101.
 - 45 E. Livshits and R. Baer, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2932.
 - 46 H. R. Eisenberg and R. Baer, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4674.
 - 47 T. Stein, H. Eisenberg, L. Kronik and R. Baer, *Phys. Rev. Lett.*, 2010, **105**, 266802.
 - 48 L. Kronik, T. Stein, S. Refaely-Abramson and R. Baer, *J. Chem. Theory Comput.*, 2012, **8**, 1515.
 - 49 M. J. Lucero, T. M. Henderson and G. Scuseria, *J. Phys.: Condens. Matter*, 2012, **24**, 145504.