Downloaded from orbit.dtu.dk on: Apr 19, 2024

DTU Library

Bayesian error estimation in density-functional theory

Mortensen, Jens Jørgen; Kaasbjerg, Kristen; Frederiksen, Søren Lund; Nørskov, Jens Kehlet; Sethna, J.P.; Jacobsen, Karsten Wedel

Published in:
Physical Review Letters

Link to article, DOI:
10.1103/PhysRevLett.95.216401

Publication date:
2005

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Mortensen, J. J., Kaasbjerg, K., Frederiksen, S. L., Nørskov, J. K., Sethna, J. P., & Jacobsen, K. W. (2005). Bayesian error estimation in density-functional theory. *Physical Review Letters*, *95*(21), 216401. https://doi.org/10.1103/PhysRevLett.95.216401

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Bayesian Error Estimation in Density-Functional Theory

J. J. Mortensen,[1] K. Kaasbjerg,[1] S. L. Frederiksen,[1] J. K. Nørskov,[1] J. P. Sethna,[2] and K. W. Jacobsen[1]

[1]*CAMP and Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2]*Laboratory of Atomic and Solid State Physics (LASSP), Clark Hall, Cornell University, Ithaca, New York 14853-2501, USA*

(Received 9 June 2005; published 15 November 2005)

We present a practical scheme for performing error estimates for density-functional theory calculations. The approach, which is based on ideas from Bayesian statistics, involves creating an ensemble of exchange-correlation functionals by comparing with an experimental database of binding energies for molecules and solids. Fluctuations within the ensemble can then be used to estimate errors relative to experiment on calculated quantities such as binding energies, bond lengths, and vibrational frequencies. It is demonstrated that the error bars on energy differences may vary by orders of magnitude for different systems in good agreement with existing experience.

DOI: 10.1103/PhysRevLett.95.216401 PACS numbers: 71.15.Mb, 02.50.Ng, 31.15.Ew, 68.43.Bc

Over the past few decades the use of density-functional theory (DFT) [1] to predict structures, energetics, and other properties of atomic-scale systems has spread to many different fields and the number of applications has grown enormously. Today applications may vary from studies of chemical reactions in heterogeneous catalysis [2] through geophysical investigations of melting at the physical conditions of the Earth's core [3,4] to studies of biomolecular systems like DNA [5,6]. The general usefulness of the calculations lies in their unbiased "first principles" character and the relatively high degree of predictive power and reliability which has been established. With respect to the latter, it is, however, often difficult to assess directly to which extent a calculated quantity—this being a molecular bond length or some other property—is to be trusted. In practice the evaluation often falls back exclusively on the experience and acquired insight of the person performing the calculation.

In this Letter we present a scheme for systematic error evaluation within the generalized-gradient approximation (GGA) DFT. The scheme is simple to apply and intended to provide users of GGA-DFT with realistic error estimates for their calculations. The approach is based on ideas from Bayesian statistics [7] in which an ensemble of model parameters are assigned probabilities by comparing to a database of experimental results. The ensemble generated can then subsequently be used to estimate error bars on model predictions. In the end the scheme involves only a few additional non-self-consistent evaluations of exchange-correlation functionals. In spirit, it is in fact close to a rather common practice within the field of DFT-GGA calculations: To assess the validity of a calculated DFT-GGA result it is common to try out different versions of the GGA-functional or perhaps to compare with a local-density approximation (LDA) result. The scheme presented here provides a systematic framework for such an approach.

The statistical approach we use is inspired by Bayesian statistics [7] and was further developed in the context of modeling complex biomedical networks [8] and for construction of interatomic potentials [9]. The main ingredients are a model M with a set of parameters θ, and a database D. In our case the model will be a GGA-type exchange-correlation functional and the database will consist of experimental atomization or cohesive energies E_k^{exp} for a collection of molecules and solids (details below). We now define a probability distribution for the model parameters through

$$P(\theta|MD) \sim \exp(-C(\theta)/T), \tag{1}$$

where C denotes a standard least-squares cost function $C(\theta) = \frac{1}{2}\sum_k (E_k(\theta) - E_k^{\mathrm{exp}})^2$ with $E_k(\theta)$ being the atomization or cohesive energy of system k in the database calculated with the parameters θ. The "effective temperature" T determines the spread of the ensemble. In simple fitting procedures only the best-fit parameters θ_{bf}, which are obtained when the cost function takes on its minimal value C_{bf}, are considered. Here, in contrast, a whole ensemble of parameter sets are considered leading to a spread of values on model predictions. Following Ref. [9], we take the value of the effective temperature to be given by the minimal (best-fit) value of the cost function C_{bf} as $T = 2C_{\mathrm{bf}}/N_{\mathrm{p}}$, where N_{p} is the number of parameters. For a harmonic cost function each parameter will then, on average, contribute an additional cost of $T/2 = C_{\mathrm{bf}}/N_{\mathrm{p}}$. This choice was demonstrated to work well in the case of error estimation for interatomic potentials [9].

The model we shall consider is GGA-DFT [10] where the exchange functional is a local function of the density n and its dimensionless gradient $s = |\nabla n|/(2k_{\mathrm{F}}n)$ [$n = k_{\mathrm{F}}^3/(3\pi^2)$]. Several suggestions for different mathematical forms of the exchange functional within GGA exist [10–14]. A commonly used class of these including PW-91 [15], PBE [11], revPBE [16] and RPBE [17] differ by the choice of the so-called enhancement factor $F_{\mathrm{x}}(s)$ in the exchange energy E_{x}:

$$E_{\mathrm{x}}[n] = \int d\mathbf{r}\, n(\mathbf{r})\epsilon_{\mathrm{x}}^{\mathrm{LDA}}(n(\mathbf{r}))F_{\mathrm{x}}(s), \tag{2}$$

 © 2005 The American Physical Society

where $\epsilon_x^{LDA}(n) = -3k_F/(4\pi)$ [for a spin polarized density we have $E_x[n_\uparrow, n_\downarrow] = (E_x[2n_\uparrow] + E_x[2n_\downarrow])/2$]. The enhancement factors for PBE and RPBE are shown in Fig. 1. In the following we shall expand the enhancement factor as

$$F_x(s) = \sum_{i=1}^{N_p} \theta_i \left(\frac{s}{1+s}\right)^{2i-2}, \tag{3}$$

regarding the θ's as free parameters. We use three parameters ($N_p = 3$) which a train/test check for our database has shown to give the optimal fit without over-fitting. The model space could be extended in future work to include a fraction of exact exchange as for the B3LYP [18] or PBE0 [19] functionals.

The database we use consists of the experimental atomization energies of the molecules H_2, LiH, CH_4, NH_3, OH, H_2O, HF, Li_2, LiF, Be_2, C_2H_2, C_2H_4, HCN, CO, N_2, NO, O_2, F_2, P_2 and Cl_2 and the experimental cohesive energies (per atom) of the solids Na, Li, Si, C, SiC, AlP, MgO, NaCl, LiF, Cu, and Pt. In the cost function all systems in the database appear with the same unit weight.

All calculations are performed with a real-space multigrid DFT code [20] using the projector-augmented wave method [21] to describe the core regions. All calculated energy differences have been converged to an accuracy better than 50 meV with respect to the number of grid points and unit-cell size or number of k points. The electron density is calculated self-consistently using the PBE functional and the evaluation of the exchange-correlation energy for other enhancement factors are performed using the PBE density. This is a very good approximation due to the variational principle. Energies are calculated at the calculated equilibrium bond distances.

Since Eq. (3) is linear in the parameters θ, the total energy of a given system will also be linear in θ:

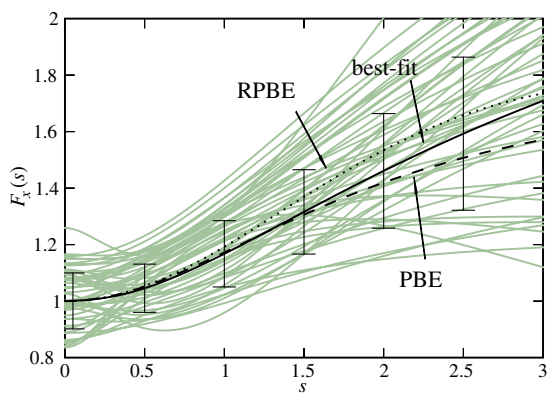

FIG. 1 (color online). Exchange enhancement factors as a function of dimensionless electron density gradient. The gray lines show enhancement factors drawn from the ensemble $\exp(-C(\theta)/T)$. The dashed, dotted, and full lines show enhancement factors for PBE, RPBE, and the best fit, respectively.

$$E(\theta) = E_0 + \sum_{i=1}^{3} \Delta E_i \theta_i, \tag{4}$$

where the coefficients E_0 and ΔE_i only have to be calculated once for each system. It is therefore very fast to calculate energy values for different enhancement factors in the ensemble.

Minimizing the cost function with respect to the three parameters leads to the best-fit enhancement factor shown in Fig. 1 corresponding to the parameters $\theta_{bf} = (1.0008, 0.1926, 1.8962)$. The function is seen to follow quite closely the PBE enhancement factor at low values of the gradient s. In particular, it is nearly one in the homogeneous limit ($s = 0$) which is exclusively a result of the fitting. For s values greater than 1.5 the best-fit enhancement factor increases more steeply than PBE being more similar to the RPBE factor. In Table I the resulting errors are shown for LDA, PBE, RPBE, and for the best fit. RPBE performs better on the molecules and PBE is better for the solids; the best fit represents a compromise between the two. We would like to stress that the main point of this Letter is not to derive an improved functional. Much experience has been acquired concerning how well different GGA functionals work for different systems [22,28,29] and we do not expect to obtain a large overall improvement within this simple GGA framework. But as we shall see in the following the ensemble construction allows for realistic evaluation of the error bars on calculated quantities.

The cost function appearing in the probability distribution Eq. (1) is very nearly quadratic in the model parameters in the relevant range of parameter space. We can therefore expand the exponent in the probability distribution as $C(\theta)/T = \text{const} + \frac{1}{2}\Delta\theta^T A \Delta\theta$, where A is a symmetric matrix. With U being the unitary matrix

TABLE I. Errors in DFT atomization energies and cohesive energies (in eV) relative to experiment. All calculations are based on self-consistent PBE densities. Experimental numbers are taken from Refs. [11,22–27].

Error	LDA	PBE	RPBE	Best fit
Molecules:				
Mean abs.	1.46	0.35	0.21	0.24
Mean	1.38	0.28	−0.01	0.12
Max. (−)	−0.35	−0.22	−0.32	−0.26
Max. (+)	3.07	0.89	0.46	0.71
Solids:				
Mean abs.	1.35	0.16	0.40	0.27
Mean	1.35	−0.09	−0.40	−0.24
Max. (−)		−0.72	−1.37	−0.94
Max. (+)	2.73	0.36		0.15
All:				
Mean abs.	1.42	0.28	0.28	0.25
Mean	1.37	0.14	−0.16	−0.02

that diagonalizes A ($AU = U\Lambda$), we can finally write the parameters of the enhancement factors in the ensemble as

$$\theta = \theta_{\rm bf} + U\Lambda^{-1/2}\alpha = \theta_{\rm bf} + \begin{pmatrix} 0.066 & 0.055 & -0.034 \\ -0.812 & 0.206 & 0.007 \\ 1.996 & 0.082 & 0.004 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \quad (5)$$

where α_1, α_2, and α_3 are stochastic variables which are Gaussian distributed with unit width: $\mathcal{P}(\alpha_i) \sim \exp(-\alpha_i^2/2)$. Using this formula it is very easy to generate a properly distributed ensemble of enhancement factors as shown in Fig. 1. The ensemble is seen to spread over a region including the PBE and PRBE functionals but not extending as far as LDA within 1 standard deviation.

The key suggestion of this Letter is that for a given calculated observable, say a bond length of a molecule, the variation of the calculated value of this observable within the ensemble of enhancement factors provides a useful estimate of how large the error of the best-fit value is compared to experiment. From an ensemble of parameters, $\theta^1, \theta^2, \ldots, \theta^N$, generated from Eq. (5), the standard deviation $\sigma_{\rm BEE}(O)$ which we shall refer to as the Bayesian error estimate (BEE) of the observable O can be determined. Considering O as a function of θ the BEE is evaluated as

$$\sigma_{\rm BEE}(O) = \sqrt{\frac{1}{N}\sum_{\mu=1}^{N}(O(\theta^\mu) - O_{\rm bf})^2}. \quad (6)$$

In the simple case where the observable is approximately linear in the parameters θ the BEE can be calculated without explicitly generating an ensemble through $\sigma_{\rm BEE}(O) = (\sum_{i=1}^{3}(\partial O/\partial\alpha_i)^2)^{1/2}$ and Eq. (5). Further details can be found in our implementation of the approach [30].

The ensemble in Eq. (6) is around the best-fit enhancement factor corresponding to $\alpha_i = 0$. However, considering that the ensemble of enhancement factors (Fig. 1) is quite wide compared with the difference between, for example, the PBE and the best-fit functional it seems reasonable to alternatively apply the fluctuations around either the PBE or RPBE functional.

It is useful to consider the ratio $(O_{\rm bf} - O_{\rm exp})/\sigma_{\rm BEE}(O)$ of the actual error relative to the estimate. For any given observable in a particular system this ratio is just a single number, so in order to assess whether our approach produces reliable error estimates from a statistical point of view we need to look at the distribution of ratios for several observables and systems.

In Fig. 2 we show histograms of the relative error for different observables for all the systems in the database. The upper panel shows the distribution in the case of the atomization or cohesive energies. As can be seen the distribution agrees quite well with a Gaussian distribution of unit width indicating that the error estimates are in fact reasonable for the binding energies. The individual standard deviations for the cohesive energies are in the range from 0.09 eV for Na to 0.75 eV for Al and for the atomization energies the range is from 0.07 eV for Li_2 to 0.60 eV for C_2H_4. As an example, the GGA estimate of the cohesive energy of Na (experimental value is 1.11 eV) is 1.02 ± 0.09 eV. The middle panel in Fig. 2 shows the relative error histogram for the bond lengths (for both molecules and solids) and relative errors for the molecular frequencies and solid bulk moduli is shown in the lower panel [31]. For both distributions, we see that the BEE's give reasonable estimates of the actual errors.

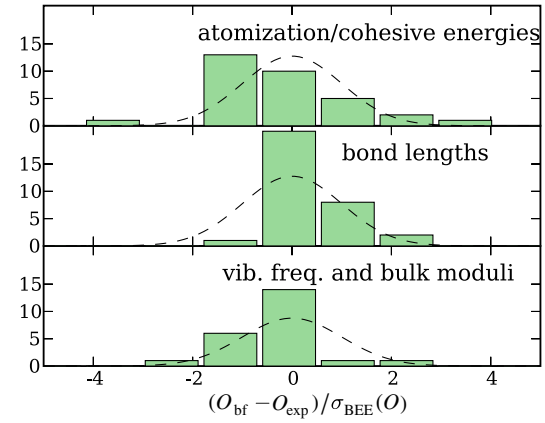

FIG. 2 (color online). Histograms of actual error relative to the BEE $(O_{\rm bf} - O_{\rm exp})/\sigma_{\rm BEE}(O)$ for different quantities. Top: atomization or cohesive energies. Middle: bond lengths and lattice constants. Bottom: vibrational frequencies and bulk moduli. The dashed lines show the expected normal distribution. Experimental numbers are taken from Refs. [22,27,28,34,35].

It is well-known for experienced users of DFT calculations that the reliability with which energy differences can be calculated vary dramatically depending on the particular system. The BEE catches this behavior as can be seen, for example, by comparing the cohesive energy and the bcc-fcc structural energy difference for bulk copper (Fig. 3). The BEE for the cohesive energy is 0.5 eV while the error bar on the structural energy difference is only 4 meV. The high reliability with which small structural energy differences can be calculated for bulk metals is confirmed by the fact that for almost all metals the correct equilibrium structures are predicted from the calculations [32].

For chemisorption systems the BEE for the energy difference between chemisorption at two different surface sites may also be somewhat smaller than the error bar for the total chemisorption energy as illustrated in the case of CO on Cu(100) in Fig. 3. However as can be seen from the figure the error bar on the site preference is so large that the preferred chemisorption site cannot be reliably determined. This is in good agreement with the fact that for a number of CO metal chemisorption systems DFT-GGA calculations do, in fact, predict a wrong chemisorption site [33].

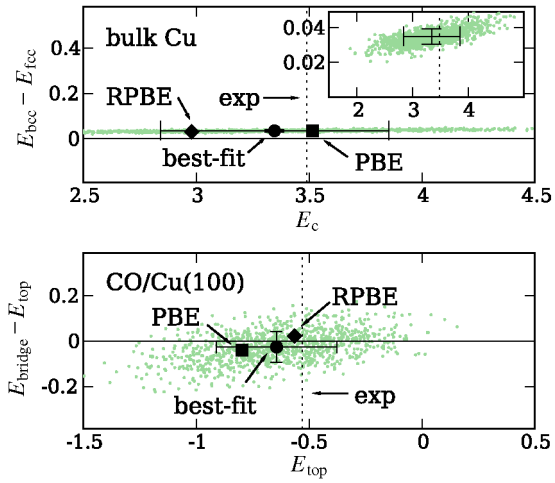

FIG. 3 (color online). Upper panel: Calculated ensemble for cohesive energy (x axis) and bcc-fcc energy difference (y axis) for a copper crystal. The BEE's are indicated by error bars. The inset uses rescaled axes. Lower panel: Calculated ensemble for binding energy (x axis) and bridge-top energy difference for CO on a Cu(100) surface. Values for the experimentally preferred states [23,36] (fcc and top) are indicated by vertical dotted lines. Units are in eV.

It should be noted that, in general, the error estimates depend on both the choice of database and the class of models (here the GGA's). This means that if some piece of physics is not present in the database or if the model is completely unable to describe a particular feature unrealistic estimates may occur. For example all GGA's are unable to properly describe the long-range van der Waals interactions and hence the BEE's will be unrealistic for that type of interactions. In practical calculations also other approximations like limited basis set size or the use of pseudopotentials may introduce additional errors which are not accounted for by the present approach which is based on well-converged calculations.

We further note that it may be possible to reduce the error bars in some cases by picking a database focused on a particular type of system. Using, for example, only the subset of molecular systems in our database a different best-fit functional is obtained where the mean absolute error is only 0.15 eV compared to the 0.24 eV in Table I. With this functional the fluctuations and hence the BEEs will be reduced by almost a factor of 2. However, applying this functional to bulk metals would lead to larger errors which would be underestimated.

Summarizing, we propose a simple way to estimate a large portion of the systematic error for DFT-GGA calculations, requiring only a few extra non-self-consistent energy calculations to calculate the error bars of any observable which is a function of energies.

We acknowledge support from the Carlsberg Foundation and the Danish Center for Scientific Computing through Grant No. HDW-1101-05.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] K. Honkala *et al.*, Science **307**, 555 (2005).
[3] A. Laio *et al.*, Science **287**, 1027 (2000).
[4] D. Alfe, M. J. Gillan, and G. D. Price, Nature (London) **401**, 462 (1999).
[5] C. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).
[6] E. Artacho *et al.*, Mol. Phys. **101**, 1587 (2003).
[7] E. T. Jaynes, *Probability Theory* (Cambridge University Press, Cambridge, United Kingdom, 2003).
[8] K. S. Brown and J. P. Sethna, Phys. Rev. E **68**, 021904 (2003).
[9] S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna, Phys. Rev. Lett. **93**, 165501 (2004).
[10] D. C. Langreth and J. P. Perdew, Phys. Rev. B **21**, 5469 (1980).
[11] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[12] A. D. Becke, J. Chem. Phys. **107**, 8554 (1997).
[13] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).
[14] A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**, 1670 (2000).
[15] J. P. Perdew *et al.*, Phys. Rev. B **46**, 6671 (1992); (The correlation part of PW91 is slightly different from the one of PBE).
[16] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).
[17] B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).
[18] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[19] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).
[20] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005); http://www.camp.dtu.dk/campos
[21] P. E. Blöchl, C. J. Forst, and J. Schimpl, Bull. Mater. Sci. **26**, 33 (2003).
[22] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, Phys. Rev. B **69**, 075102 (2004).
[23] C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, New York, 1996), 7th ed.
[24] I.-H. Lee and R. M. Martin, Phys. Rev. B **56**, 7197 (1997).
[25] B. Paulus, P. Fulde, and H. Stoll, Phys. Rev. B **54**, 2556 (1996).
[26] I. P. Daykov, T. A. Arias, and T. D. Engeness, Phys. Rev. Lett. **90**, 216402 (2003).
[27] R. N. Schmid *et al.*, Adv. Quantum Chem. **33**, 209 (1999).
[28] S. Kurth, J. P. Perdew, and P. Blaha, Int. J. Quantum Chem. **75**, 889 (1999).
[29] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, J. Chem. Phys. **119**, 12 129 (2003).
[30] http://www.camp.dtu.dk/campos/ASE
[31] Vibrational frequencies are only calculated for a subset of the molecules.
[32] H. L. Skriver, Phys. Rev. B **31**, 1909 (1985).
[33] P. J. Feibelman *et al.*, J. Phys. Chem. B **105**, 4018 (2001).
[34] B. G. Johnson, P. M. W. Gill, and J. A. Pople, J. Chem. Phys. **98**, 5612 (1993).
[35] P. Rodriguez-Hernández and A. Muñoz, Semicond. Sci. Technol. **7**, 1437 (1992).
[36] S. Vollmer, G. Witte, and C. Wöll, Catal. Lett. **77**, 97 (2001).