

Why Density-Gradient Corrections Improve Atomization Energies and Barrier Heights

John P. Perdew and Matthias Ernzerhof

Department of Physics and Quantum Theory Group, Tulane University, New Orleans, LA 70118

Alès Zupan

Department of Physical and Organic Chemistry, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Kieron Burke

Department of Chemistry, Rutgers University, Camden, NJ 08102

While the Hartree-Fock (HF) approximation typically underestimates the strength of the chemical bond, the local spin density (LSD) approximation overestimates it. Thus LSD overbinds atoms in molecules, and underestimates the heights of energy barriers when the transition state is more highly bonded than the initial state. Generalized gradient approximations (GGAs), which incorporate density-gradient corrections to LSD, improve the agreement between calculated and measured energetics. This has been previously understood as a consequence of the fact that gradient corrections favor density inhomogeneity, which increases when a bond is stretched or broken. We show that gradient corrections also favor high density, which increases when a bond is compressed or formed, but that the inhomogeneity effect usually prevails. To quantify the discussion, we present a thermodynamic-like inequality which is satisfied when gradient corrections favor a process.

I. ENERGY FUNCTIONALS AND CHEMICAL ENERGETICS

Quantum chemistry is most simply done with single-particle orbitals $\psi_{\alpha,\sigma}$ which are obtained from self-consistent single-particle equations. The exchange-correlation energy E_{xc} is then constructed from the orbitals, or from the spin densities n_\uparrow and n_\downarrow . The Hartree-Fock (HF) approximation neglects correlation but treats exchange exactly:

$$E_{xc}^{HF}[\{\psi_{\alpha,\sigma}\}] = -\frac{1}{2} \sum_{\sigma} \int d^3r \int d^3r' \sum_{\alpha} \psi_{\alpha,\sigma}^*(\mathbf{r}) \psi_{\alpha,\sigma}(\mathbf{r}') / |\mathbf{r}' - \mathbf{r}|. \quad (1)$$

The local spin density (LSD) approximation [1]

$$E_{xc}^{LSD}[n_\uparrow, n_\downarrow] = \int d^3r n(\mathbf{r}) \epsilon_{xc}(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})) \quad (2)$$

invokes $\epsilon_{xc}(n_\uparrow, n_\downarrow)$, the exchange-correlation energy per particle of an electron gas with uniform spin densities n_\uparrow and n_\downarrow [2]. The generalized gradient approximation (GGA) [3-8]

$$E_{xc}^{GGA}[n_\uparrow, n_\downarrow] = \int d^3r f(n_\uparrow, n_\downarrow, \nabla n_\uparrow, \nabla n_\downarrow) \quad (3)$$

makes use of the local density gradient as well as the local density. (Except where other units are specified, we use atomic units in which $\hbar = m = e^2 = 1$).

The ‘‘GGA made simple’’ [7], which we employ in this work, arises from two different non-empirical derivations [7,8]. All its parameters (other than those in LSD) are fundamental constants.

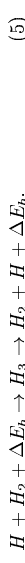
Both LSD and GGA predict realistic geometries and vibrational frequencies for molecules, but GGA gives a better account of atomization energies [9,10] and barrier heights [11-15]. The aim of this article is to explain why, using a density-parameter analysis we have recently developed [16,17], and to provide a pedagogical discussion of these issues. We will show how the chemical effects of gradient corrections follow from the gradient dependence of the functional E_{xc}^{GGA} . For an explanation of the physical origins of this non-locality, see Ref. [18]. For some reasons why LSD and GGA work as well as they do, see Ref. [19]. For methods which incorporate but go beyond GGA, see Refs. [20] and [21].

The atomization energy ΔE_a is the extra energy needed to break up a molecule into separate atoms. For the nitrogen dimer,



For this reaction, as for most others, HF underbinds ($\Delta E_a = 5.0$ eV), LSD overbinds ($\Delta E_a = 11.6$ eV), and GGA gives a more realistic binding energy ($\Delta E_a = 10.5$ eV). The experimental result is $\Delta E_a = 9.9$ eV. All values are taken from Ref. [7]. Other reaction exothermicities or endothermicities [22,23] have also been studied.

The rates of many chemical reactions are limited by an energy barrier ΔE_b which must be surmounted [24]. A well-studied case [14,15,17,25] is the hydrogen abstraction



In the initial state, a hydrogen atom approaches a hydrogen molecule along its bond axis from one side, with energy ΔE_b . The $H \cdots H_2$ distance will be denoted by x . Taking x to be the independent reaction coordinate, we determine y , the H_2 bond length ($y \leq x$), by minimizing the total energy for each x . In the transition state of H_3 , we find $y = x$. When the transition state H_3 fissions, it kicks out a hydrogen atom to the other side. The net effect of this one-dimensional process is a transfer of nuclei. In this reaction [17], as in others where the transition state is more highly bonded than the initial state, the HF barrier is too high ($\Delta E_b = 0.8$ eV), the LSD barrier is too low ($\Delta E_b = -0.1$ eV, making H_3 bound in LSD), and the GGA barrier is more realistic ($\Delta E_b = 0.2$ eV) in comparison with experiment ($\Delta E_b = 0.4$ eV, from Ref. [25]).

Sometimes the transition state has as many bonds as the initial state, e.g., in isomerizations or internal rotations. LSD gives realistic barrier heights

for isomerizations [26], and presumably so does GGA, LSD and GGA predict essentially the same (correct) barriers for internal rotations [27].

The transition state for an atom hopping on a metal surface [28] can have a *lower* coordination number than the initial state, and then GGA barrier heights are lower (and presumably again more correct) than LSD barriers.

II. QUALITATIVE EXPLANATIONS FOR FUNCTIONAL PERFORMANCE

All of the effects described in the preceding section are consistent with one neat statement [11,12]: LSD likes bonds too much, and GGA corrects this bias. Compelling further evidence for this statement can be found in a comparison of LSD and GGA energies for different structures of C_n clusters [29]: The fewer the bonds, the more the gradient corrections lower the total energy. GGA usually weakens and stretches bonds relative to LSD. When bonds stretch or break, the density gradient or inhomogeneity grows, and this lowers the GGA energy relative to LSD. Much can be explained by this statement [30]: GGA favors density inhomogeneity more than LSD does.

However, this cannot be the whole story, because gradient corrections usually and correctly *shrink* the lengths of covalent bonds to hydrogen [14,15,17], counter to expectations based upon the statement at the end of the previous paragraph.

Our current explanation can be summarized qualitatively in this statement [17]: GGA favors both inhomogeneity and high density more than LSD does. These effects usually compete; for example, inhomogeneity increases and density decreases as we stretch or break a bond. The inhomogeneity effect usually prevails but, for the bond length of H_2 , the density effect is large enough to carry the day. The final two sections of this article will quantify and test this statement.

Alternative explanations [10,13] appeal to the selfconsistent effect of gradient corrections on the electron density. For example, gradient corrections shrink the ionic cores of the atoms, producing a stronger screening of the nuclear potential and a more diffuse valence-electron density, hence a longer bond length between most atoms which have cores. We do not disagree with these alternative explanations, but we believe that ours is more robust, because it explains why gradient corrections produce all their characteristic energetic effects, even when applied to the LSD density [13,15].

III. DENSITY PARAMETERS AND THEIR AVERAGES

To quantify the explanation given at the end of the preceding section, let us define the local density parameter $r_s(\mathbf{r})$, the local relative spin polarization

$\zeta(\mathbf{r})$, and the local inhomogeneity parameter or reduced density gradient $s(\mathbf{r})$ by the equations

$$n = n_\uparrow + n_\downarrow, \quad (6)$$

$$\zeta = (n_\uparrow - n_\downarrow)/n, \quad (7)$$

$$s = |\nabla n|/2(3\pi^2)^{1/3}n^{4/3}. \quad (8)$$

The Seitz radius r_s is the radius of a sphere which on average contains one electron, so $r_s \rightarrow 0$ in the high-density limit; valence electrons typically have $1 \text{ bohr} \leq r_s \leq 6 \text{ bohr}$. The relative spin polarization ζ vanishes for a spin-unpolarized system, and equals +1 or -1 where all the electron spins are up or down, respectively. The inhomogeneity parameter s measures how fast the density varies on the scale of the local Fermi wavelength; valence electrons typically have $0 \leq s \leq 3$. Both r_s and s diverge in the low-density tail of an atom or molecule.

In previous work [16,17], we have studied distributions and averages of these parameters in atoms, molecules and solids, and how these change in various processes. Typically, the averages $\langle r_s \rangle$, $\langle |\zeta| \rangle$, and $\langle s \rangle$ all increase when we stretch or break a bond, as shown in Fig. 1 for the fission of the H_3 transition state.

Because the distributions of the density parameters p are broad, their averages must be defined carefully. We want true averages $\langle p \rangle$, which fall between the minimum and maximum values of p present in the system. But we also want averages that will give meaningful estimates of LSD and GGA exchange-correlation energies.

To a good approximation, any GGA can be written as

$$E_{xc}^{GGA} = \int d^3r \, n\epsilon_x(r_s)F_{xc}(r_s, \zeta, s), \quad (9)$$

where small $\nabla\zeta$ contributions have been neglected. (The barrier height in the reaction of Eq. (5) is 0.15 eV with and 0.21 eV without these contributions.) In Eq. (9),

$$\epsilon_x(r_s) = -\frac{3}{4\pi} \left(\frac{9\pi}{4}\right)^{1/3} \frac{1}{r_s} \quad (10)$$

is the exchange energy per electron of a spin-unpolarized uniform electron gas of density parameter r_s . The enhancement factor over spin-unpolarized local exchange, $F_{xc}(r_s, \zeta, s)$, is displayed in Fig. 1 of Ref. [7]. It can be decomposed into exchange and correlation components:

$$F_{xc}(r_s, \zeta, s) = F_x(\zeta, s) + F_c(r_s, \zeta, s) \quad (11)$$

where $F_c \rightarrow 0$ as $r_s \rightarrow 0$. The s -dependence of Eq. (11) is the nonlocality of GGA, which reduces to LSD when we set $s = 0$:

$$E_{\text{xc}}^{\text{LSD}} = \int d^3r n\epsilon_x(r_s)F_{\text{xc}}(r_s, \zeta, 0). \quad (12)$$

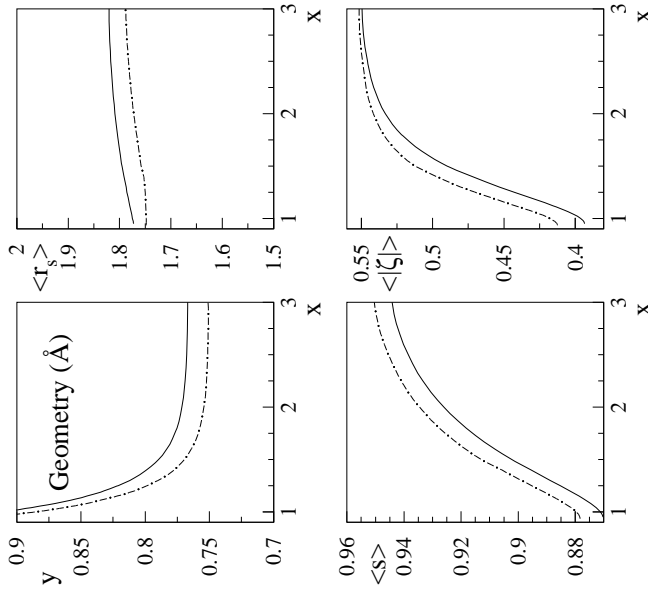


FIG. 1. Geometry and density-parameter averages for the fragmentation of H_2 transition state of Eq. (5). x is the longer of the two bond lengths, and increases from $x = y = 0.95$ Å (LSD) or 0.935 Å (GGA) at the transition state to infinity in the dissociated limit. The averages defined in Eqs. (13)-(15) have been evaluated selfconsistently in LSD (solid curve) and GGA (dashed curves). Note that gradient corrections increase the inhomogeneity parameter $\langle s \rangle$ and reduce $\langle r_s \rangle$, as expected.

To define our density-parameter averages, we use the exchange-only limit:

$$\epsilon_x(\langle r_s \rangle) = \int d^3r n\epsilon_x(r_s) / \int d^3r n, \quad (13)$$

$$F_x(\langle |\zeta| \rangle, 0) = \int d^3r n\epsilon_x(r_s)F_x(\zeta, 0) / \int d^3r n\epsilon_x(r_s), \quad (14)$$

$$F_x(0, \langle s \rangle) = \int d^3r n\epsilon_x(r_s)F_x(0, s) / \int d^3r n\epsilon_x(r_s). \quad (15)$$

The averages $\langle r_s \rangle$, $\langle |\zeta| \rangle$, and $\langle s \rangle$ are found as numerical roots of these equations. If the statement that “gradient corrections favor density inhomogeneity” were the whole story, we would expect that gradient corrections would favor any physical process in which

$$d\langle s \rangle > 0. \quad (16)$$

As we will see in the next section, Eq. (16) is *not* the right condition.

IV. QUANTITATIVE EXPLANATION OF FUNCTIONAL PERFORMANCE

The averages defined in Eqs. (13) - (15) can be used to estimate the GGA and LSD exchange-correlation energies as

$$E_{\text{xc}}^{\text{GGA}} \approx N\epsilon_x(\langle r_s \rangle)F_{\text{xc}}(\langle r_s \rangle, \langle |\zeta| \rangle, \langle s \rangle), \quad (17)$$

$$E_{\text{xc}}^{\text{LSD}} \approx N\epsilon_x(\langle r_s \rangle)F_{\text{xc}}(\langle r_s \rangle, \langle |\zeta| \rangle, 0), \quad (18)$$

where N is the number of electrons. Eqs. (17) and (18) become exact when spin-unpolarized exchange dominates all other effects, or when the distributions of the density variables are very narrow. However, these equations are surprisingly accurate for real atoms, molecules, and solids [17]. For very heavy atoms, one might want to eliminate the core-electron contributions to (17) and (18), but we have not found this to be necessary in our studies for atomic numbers up to 17.

Now the effect of gradient corrections upon the energy per electron is

$$\begin{aligned} \Delta &\equiv (E_{\text{xc}}^{\text{GGA}} - E_{\text{xc}}^{\text{LSD}})/N \\ &\approx \epsilon_x(\langle r_s \rangle)[F_{\text{xc}}(\langle r_s \rangle, \langle s \rangle) - F_{\text{xc}}(\langle r_s \rangle, \langle |\zeta| \rangle, 0)]. \end{aligned} \quad (19)$$

The exact evaluation of Δ is simpler and more accurate than evaluation of the last line of Eq. (19), but the latter expression permits us to make a quantitative test of our qualitative explanations.

Consider a physical process in which each density-parameter average $\langle p \rangle$ changes to $\langle p \rangle + d\langle p \rangle$ at fixed N . This process will be favored by gradient corrections if and only if

$$d\Delta < 0 \quad (20)$$

i.e., (for $\partial\Delta/\partial\langle s \rangle < 0$)

$$\frac{d\langle s \rangle}{\langle s \rangle} > P \frac{d\langle r_s \rangle}{2\langle r_s \rangle} + Qd\langle |\zeta| \rangle, \quad (21)$$

where

$$P = -\frac{2\langle r_s \rangle}{\langle s \rangle} \frac{\partial \Delta / \partial \langle r_s \rangle}{\partial \Delta / \partial \langle s \rangle}, \quad (22)$$

$$Q = -\frac{1}{\langle s \rangle} \frac{\partial \Delta / \partial \langle \zeta \rangle}{\partial \Delta / \partial \langle s \rangle}. \quad (23)$$

We have tested ([17,18]) the inequality (21) for about 20 atomization processes and a smaller number of bond stretchings and transition-state fragmentations. We find that the inequality (21) holds whenever gradient corrections favor a process.

A simple model in which $P = 1$ and $Q = 0$ is

$$F_{xc}(r_s, \zeta, s) = F_{xc}(r_s, \zeta, 0) + \mu s^2, \quad (24)$$

where μ is any positive constant. Although the true GGA nonlocality is more complicated than Eq. (24), it also yields [17] $P \approx 1$ and $Q \approx 0$. Thus we can simplify Eq. (21) for qualitative purposes to

$$\frac{d\langle s \rangle}{\langle s \rangle} > \frac{d\langle r_s \rangle}{2\langle r_s \rangle}, \quad (25)$$

which clearly differs from the naive condition (16) when $d\langle r_s \rangle \neq 0$. In bond stretching, for example, both the left- and right-hand sides of (25) are typically positive. The left-hand side is usually bigger, but in H_2 and H_3 it is smaller than the right-hand side. Thus gradient corrections stretch most bonds, but contract the H_2 and H_3 bonds, and affect atomization energies and barrier heights as described in the first section of this article.

Our inequalities (21) and (25), which follow from the fact that the GGA enhancement factor $F_{xc}(r_s, \zeta, s)$ increases with s , tell us when a process will be favored by gradient corrections. How strongly it will be favored then depends upon how strongly $F_{xc}(r_s, \zeta, s)$ increases with s . Fig. 1 of Ref. [7] shows that the exchange-only ($r_s = 0$) enhancement factor has an especially strong increase with s , which partly explains why the HF approximation underbinds atoms in molecules. For example, consider the atomization of NH_3 , in which the exchange-correlation energy increases by 8.9 eV. The corresponding increases in the exchange energy are 7.0 (LSD), 5.9 (GGA), and 4.7 (HF) eV [31]. Thus, neglecting the nonlocality of exchange has an effect somewhat like including correlation.

The inequalities (21) and (25) also explain the effects of gradient corrections upon structural transformations in solids [32] and upon surface energies [17].

V. ACKNOWLEDGMENTS

J.P.P. and M.E. acknowledge the support of the National Science Foundation under Grant No. DMR 95-21353. A.Z. acknowledges the support of the Slovenian Ministry for Science and Technology.

VI. REFERENCES

- [1] W. Kohn and L.J. Sham, Phys. Rev. 140, A1133 (1965).
- [2] J.P. Perdew and Y. Wang, Phys. Rev. B45, 13244 (1992).
- [3] D.C. Langreth and M. J. Mehl, Phys. Rev. B28, 1809 (1983).
- [4] J.P. Perdew and Y. Wang, Phys. Rev. B33, 8800 (1986); 40, 3399 (E) (1989).
- [5] J.P. Perdew, Phys. Rev. B33, 8822 (1986); 34, 7406 (E) (1986).
- [6] A.D. Becke, Phys. Rev. A38, 3098 (1988).
- [7] J.P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996); 78, 1396 (E) (1997).
- [8] J.P. Perdew, K. Burke, and Y. Wang, Phys. Rev. B 54, 16533 (1996).
- [9] L.A. Curtiss, K. Raghavachari, P.C. Redfern, and J.A. Pople, J. Chem. Phys. 106, 1063 (1997), and references therein.
- [10] D.C. Patton, D.V. Porezag, and M. R. Pederson, Phys. Rev. B55, 7454 (1997).
- [11] L. Fan and T. Ziegler, J. Am. Chem. Soc. 114, 10890 (1992).
- [12] L. Deng, T. Ziegler, and L. Fan, J. Chem. Phys. 99, 3823 (1993).
- [13] B. Hammer, K. W. Jacobsen, and J.K. Nørskov, Phys. Rev. Lett. 70, 3971 (1993).
- [14] B.G. Johnson, C.A. Gonzales, P.M.W. Gill, and J.A. Pople, Chem. Phys. Lett. 221, 100 (1994).
- [15] D. Porezag and M.R. Pederson, J. Chem. Phys. 102, 9345 (1995).
- [16] A. Zupan, J.P. Perdew, K. Burke, and M. Causá, Int. J. Quantum Chem. 61, 835 (1997).
- [17] A. Zupan, K. Burke, M. Ernzerhof, and J. P. Perdew, J. Chem. Phys. 106, 10184 (1997).

- [18] J.P. Perdew, M. Ernzerhof, A. Zupan, and K. Burke, *J. Chem. Phys.*, to appear.
- [19] K. Burke, M. Ernzerhof, and J.P. Perdew, unpublished.
- [20] M. Ernzerhof, *Chem. Phys. Lett.* **263**, 499 (1996); J.P. Perdew, M. Ernzerhof, and K. Burke, *J. Chem. Phys.* **105**, 9982 (1996); K. Burke, M. Ernzerhof, and J.P. Perdew, *Chem. Phys. Lett.* **265**, 115 (1997).
- [21] M. Noland, E.L. Cortiño, and D.G. Truhlar, *J. Phys. Chem.* **A101**, 1193 (1997).
- [22] R. Neumann and N.C. Handy, *Chem. Phys. Lett.* **246**, 381 (1995).
- [23] J.M. Martell, J.D. Goddard, and L.A. Eriksson, *J. Phys. Chem.* **A101**, 1927 (1997).
- [24] P.J. Feibelman and J. Harris, *Nature* **372**, 135 (1994).
- [25] W.R. Schultz and D.J. LeRoy, *J. Chem. Phys.* **42**, 3869 (1965).
- [26] J.M. Seminario, M. Grodzicki, and P. Politzer, in *Density Functional Methods in Chemistry*, edited by J. Labanowski and J. Andzelm (Springer-Verlag, Berlin, 1990).
- [27] A. Borrmann and R.O. Jones, *Chem. Phys. Lett.* **252**, 1 (1996).
- [28] B.D. Yu and M. Scheffler, *Phys. Rev. Lett.* **77**, 1095 (1996).
- [29] R.O. Jones and G. Seifert, *Phys. Rev. Lett.* **79**, 443 (1997).
- [30] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais, *Phys. Rev. B* **46**, 6671 (1992); **48**, 4978 (E) (1993).
- [31] M. Ernzerhof, J.P. Perdew, and K. Burke, *Int. J. Quantum Chem.* **64**, 285 (1997).
- [32] A. Zupan, P. Blaha, K. Schwarz, and J.P. Perdew, unpublished.