Dispersion, static correlation, and delocalisation errors in density functional theory: An electrostatic theorem perspective

Austin D. Dwyer and David J. Tozer

Department of Chemistry, Durham University, South Road, Durham, DH1 3LE, United Kingdom

(Received 12 July 2011; accepted 30 September 2011; published online 31 October 2011)

Dispersion, static correlation, and delocalisation errors in density functional theory are considered from the unconventional perspective of the force on a nucleus in a stretched diatomic molecule. The electrostatic theorem of Feynman is used to relate errors in the forces to errors in the electron density distortions, which in turn are related to erroneous terms in the Kohn-Sham equations. For H₂, the exact dispersion force arises from a subtle density distortion; the static correlation error leads to an overestimated force due to an exaggerated distortion. For H₂⁺, the exact force arises from a delicate balance between attractive and repulsive components; the delocalisation error leads to an underestimated force due to an underestimated distortion. The net force in H₂⁺ can become repulsive, giving the characteristic barrier in the potential energy curve. Increasing the fraction of long-range exact orbital exchange increases the distortion, reducing delocalisation error but increasing static correlation error. © 2011 American Institute of Physics. [doi:10.1063/1.3653980]

Commonly used exchange-correlation functionals in Kohn-Sham density functional theory (DFT) (Ref. 1) omit the physics of long-range dispersion,\(^2\)\(^-\)\(^4\) meaning they fail to provide a leading \(-C_6 R^{-6}\) interaction energy between non-overlapping fragments. They also suffer from so-called static correlation and delocalisation errors, which can be understood from the perspective of fractional spins and fractional charges, respectively.\(^5\)\(^-\)\(^7\) The paradigm for static correlation error is the overestimation of the energy of stretched H₂, comprising two local systems each containing half a spin-up and half a spin-down electron. The paradigm for delocalisation error is the underestimation of the energy of stretched H₂⁺, comprising two local half-electron systems. These three problems underline many of the well-known deficiencies of approximate DFT,\(^8\) including the description of weakly bound complexes,\(^2\)\(^,\)\(^3\) transition metal systems and molecular dissociation,\(^5\)\(^-\)\(^7\) bandgaps,\(^9\) and charge-transfer excitation energies.\(^10\) Traditionally, the problems have been considered from the perspective of the electronic energy. In the present study, we consider them from the unconventional perspective of the force on a nucleus. We use the electrostatic theorem of Feynman\(^11\) to relate the errors in the forces to errors in the electron density distortions, which in turn are related to erroneous terms in the Kohn-Sham orbital equations arising due to the approximate exchange-correlation functional. The analysis is elementary but insightful.

We consider H₂ and H₂⁺ at large internuclear separation, illustrating dispersion and static correlation error in the former and delocalisation error in the latter. Forces on the nuclei are determined using both the conventional analytic evaluation (negative of the energy derivative with respect to the nuclear coordinate), and using the electrostatic theorem.\(^11\) The electrostatic theorem—obtained by applying the differential Hellmann-Feynman theorem\(^11\)\(^,\)\(^12\) to a nuclear perturbation—states that the force on a nucleus is simply the classical electrostatic force arising due to the other nuclei and the electron density; it therefore has great physical appeal since it allows forces to be understood in terms of the electron density, in the true spirit of DFT. The theorem is formally exact, but breaks down for non-variational methodologies and/or finite basis sets, limiting its use in practical calculations. We avoid this breakdown by applying only variational methods with an extensive basis set (d-aug-cc-pV6Z, omitting g and h functions for technical reasons). In all cases, the forces that we determine using the electrostatic theorem are close to those determined using the conventional analytic evaluation, meaning the physical insight provided by the theorem can be fully exploited. The use of this large basis set also negates the need to include basis set superposition corrections, since they are significantly smaller than the precision quoted.

The nuclei in each diatomic are denoted by A and B, with the vector from B to A pointing along the positive z axis. The z-component of the force on nucleus A determined using the conventional analytic evaluation is

\[
F_A = -\frac{\partial E}{\partial z_A} = -\frac{\partial E}{\partial R}. \quad (1)
\]

where \(E\) is the electronic energy, \(z_A\) is the z coordinate of nucleus A, and \(R = z_A - z_B\) is the internuclear separation. A negative value of \(F_A\) corresponds to an attractive force on nucleus A, i.e., a force in the direction of B. A positive value corresponds to a repulsive force. The force on nucleus B is equal and opposite to \(F_A\) by symmetry/translational invariance.

For H₂ and H₂⁺, the force on nucleus A determined using the electrostatic theorem is

\[
F_A^{el} = \frac{1}{R^2} + \int \frac{\rho(r_1)}{r_{1A}^3} z_{1A} dr_1, \quad (2)
\]

where \(\rho(r_1)\) is the total electron density of the system, \(r_{1A}\) is the length of the vector from nucleus A to the electronic coordinate \(r_1\), and \(z_{1A}\) is the z-component of that vector. The first
term in Eq. (2) is the classical force due to nucleus B, whilst the second term is the classical force due to the electron density. Equations (1) and (2) are both determined analytically, which is a routine task in electronic structure programs.

At the large internuclear separations considered in the present study, it is natural to spatially partition the total electron density into two fragment densities, \( \rho(r) = \rho_A(r) + \rho_B(r) \). The fragment density \( \rho_A(r) \) is defined to be equal to the total density \( \rho(r) \) in regions of space closest to nucleus A, but zero in regions closest to nucleus B, and vice versa. Put another way, \( \rho_A(r) \) is the density of the molecule to the side of the bond mid-plane closest to nucleus A. The electrostatic force in Eq. (2) then becomes

\[
F_{el}^A = \frac{1}{R^2} + \int \frac{\rho_A(r_1)}{r_1^3} r_1 dz_1 + \int \frac{\rho_B(r_1)}{r_1^3} r_1 dz_1 = \frac{1}{R^2} + F_{el}^{\rho_A} + F_{el}^{\rho_B},
\]

where \( F_{el}^{\rho_A} \) and \( F_{el}^{\rho_B} \) are the classical forces on nucleus A, arising due to fragment densities \( \rho_A(r) \) and \( \rho_B(r) \), respectively. The forces \( F_{el}^{\rho_A} \) and \( F_{el}^{\rho_B} \) are determined numerically using standard DFT numerical integration schemes. Test calculations comparing numerical and analytic evaluations of the full electrostatic force, \( F_{el}^A \), were used to ensure that the numerical integration grid is sufficiently large; the error is always smaller than the precision quoted.

The electrostatic force is governed by the electron density and so insight into the forces from different methods could be obtained by comparing their electron densities. However, for reasons that will become clear, we shall instead compare electron density distortions, which measure the extent to which the density of the diatomic differs from the superposition of two spherical densities. We quantify this by using the density distortion function,

\[
\Delta \rho(r) = \rho(r) - \rho_A^0(r) - \rho_B^0(r),
\]

where \( \rho(r) \) is the total density of the diatomic AB from a given method and \( \rho_A^0(r) \) and \( \rho_B^0(r) \) are isolated spherical densities centred at the positions of nuclei A and B, respectively, obtained from a calculation on the diatomic at \( R = 10^7 \) \( a_0 \) using the same method. All calculations on H\(_2\) use a restricted formalism and so the resulting spherical densities each comprise half a spin-up and half a spin-down electron. Calculations on H\(_2^+\) use an unrestricted formalism and so the spherical densities each comprise half a spin-up electron. All density distortion plots presented in this study were determined using a bond length of \( R = 13 \) \( a_0 \), with the nuclei located at \( z = \pm 6.5 \) \( a_0 \).

For DFT, particular attention is paid to the amount of long-range (i.e., large inter-electron distance) exact orbital exchange in the exchange-correlation functional. For H\(_2\), we consider the Becke-Lee–Yang–Parr generalised gradient approximation\(^{15,14} \) (BLYP) containing no exact exchange; the Becke3-Lee–Yang–Parr hybrid functional\(^{14,15} \) (B3LYP) containing 20% exact exchange; the CAM-B3LYP Coulomb-attenuated functional\(^{17-19} \) where the fraction of exchange increases from 19% to 65% at long-range; and a modified version (denoted CAM-B3LYP-X), where the fraction of exchange increases to 100% at long-range; this functional was obtained by simply replacing the \( \beta \) parameter in the CAM-B3LYP functional definition\(^{19} \) by 0.81. For H\(_2^+\), the one-electron nature means only the respective exchange components of the DFT functionals are used, denoted B, B3, CAM-B3, and CAM-B3-X. The accuracy of the DFT calculations is quantified by comparing with reference CCSD and Hartree-Fock results for H\(_2\) and H\(_2^+\), respectively, which are exact within the basis set (i.e., equivalent to full CI). Calculations were performed using the CADPAC (Ref. 20) and DALTON (Ref. 21) programs.

Figure 1 presents the potential energy curves for H\(_2\) and H\(_2^+\), illustrating the static correlation error and delocalisation error, respectively, from the conventional perspective of the electronic energy. As the fraction of long-range exact exchange in the functional increases, the static correlation error (the discrepancy between DFT and CCSD at large \( R \) in Figure 1(a)) increases, whereas the delocalisation error (the discrepancy between DFT and Hartree-Fock at large \( R \) in Figure 1(b)) decreases.

First consider H\(_2\). Table I presents the forces \( F_A \), \( F_{el}^A \), and \( F_{el}^{\rho_A} \), evaluated at the stretched bond length, \( R = 13 \) \( a_0 \). The \( F_A \) values correspond exactly to the negative of the slopes of the potential energy curves in Figure 1(a).

### Table I. H\(_2\) forces (a.u.) at \( R = 13 \) \( a_0 \).

<table>
<thead>
<tr>
<th>Method</th>
<th>( F_A )</th>
<th>( F_{el}^A )</th>
<th>( F_{el}^{\rho_A} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSD (exact)</td>
<td>(-7.37 \times 10^{-7})</td>
<td>(-7.34 \times 10^{-7})</td>
<td>(-7.69 \times 10^{-7})</td>
</tr>
<tr>
<td>BLYP</td>
<td>(-7.65 \times 10^{-5})</td>
<td>(-7.64 \times 10^{-5})</td>
<td>(-7.24 \times 10^{-5})</td>
</tr>
<tr>
<td>B3LYP</td>
<td>(-7.00 \times 10^{-4})</td>
<td>(-6.99 \times 10^{-4})</td>
<td>(-6.89 \times 10^{-4})</td>
</tr>
<tr>
<td>CAM-B3LYP</td>
<td>(-2.07 \times 10^{-3})</td>
<td>(-2.07 \times 10^{-3})</td>
<td>(-2.05 \times 10^{-3})</td>
</tr>
<tr>
<td>CAM-B3LYP-X</td>
<td>(-3.17 \times 10^{-3})</td>
<td>(-3.17 \times 10^{-3})</td>
<td>(-3.13 \times 10^{-3})</td>
</tr>
</tbody>
</table>
at \( R = 13 \ a_0 \). In all cases, \( F_A \approx F_A^{el} \), indicating the validity of the electrostatic theorem. Furthermore, in all cases \( F_A^{el} \) is also close to \( F_A^{el,\rho_A} \), meaning the electrostatic force on nucleus A arises primarily due to the fragment density \( \rho_A(r) \). This is trivial to understand—the density \( \rho_B(r) \) contains one electron localised around nucleus B and so, at large \( R \), it exerts an attractive force that almost exactly cancels the repulsive force due to nucleus B,

\[
F_A^{el,\rho_A} \approx -\frac{1}{R^2},
\]

and so from Eq. (3),

\[
F_A \approx F_A^{el} \approx F_A^{el,\rho_A}.
\]

Given that a spherical fragment density \( \rho_A(r) \) would yield \( F_A^{el,\rho_A} = 0 \) by symmetry, a non-zero value of \( F_A^{el,\rho_A} \) must arise from a distortion of that density from spherical.

Consider the CCSD (exact) forces in Table I. \( F_A \) is just \(-7.37 \times 10^{-7} \) a.u., corresponding to a very weak attractive force. At this large value of \( R \), there is essentially no overlap between the atomic densities and so the only interaction is due to dispersion; this value of \( F_A \) is satisfyingly close to the asymptotic attractive dispersion force evaluated using experimental dispersion coefficients,\(^22\)

\[
F_A^{disp} = -6C_6 R^{-7} - 8C_6 R^{-9} - \ldots = -7.38 \times 10^{-7} \text{a.u.} \tag{7}
\]

Figure 2 presents \( \Delta \rho(r) \) for CCSD. There is a clear distortion around each nucleus, which leads to the non-zero \( F_A^{el,\rho_A} \) in Eq. (6). The situation is exactly as described by Feynman,\(^11\) the dispersion force arises because nucleus A is pulled towards B, primarily by its own distorted electron density, \( \rho_A(r) \). Indeed, Feynman conjectured that the distorted density \( \rho_A(r) \) is the sole source of the \( R^{-7} \) component of the dispersion force and this was explicitly verified by Hirschfelder and Eliason for \( \text{H}_2 \).\(^{23}\) See Refs. 24 and 25 for further discussion. Before the CCSD density distortion can be compared with those determined using DFT, it is necessary to change the scale of the plot. The revised plot is presented in Figure 3(a).

Next consider the DFT results. As the fraction of long-range exact exchange in the functional increases, the forces in Table I become increasingly too attractive, as reflected in the slopes of the potential energy curves in Figure 1(b) at \( R = 13 \ a_0 \). For CAM-B3LYP-X, the forces are more than three orders of magnitude too large. (Although not explicitly presented, Hartree-Fock theory yields values of \( F_A \), \( F_A^{el} \), and \( F_A^{el,\rho_A} \) that are essentially identical to those of CAM-B3LYP-X; the potential energy curves are also very similar, reflecting the key role of the fraction of long-range exact exchange). The reason for the increased forces is evident in Figure 3—the density distortions become increasingly pronounced. The static correlation error therefore manifests itself as an over-estimated attractive force due to an exaggerated density distortion. The origin of this exaggerated density distortion can, in turn, be understood in terms of the Kohn-Sham orbital equation. Consider the potential experienced by an electron in the vicinity of nucleus A. In the exact case, the potential must be close to \(-1/\alpha \), in order to ensure a hydrogen atom solution locally (albeit slightly distorted due to dispersion). The electron should therefore “see” no charge on the other atom. Within DFT, the potential near nucleus A arising due to the other atom comprises nuclear, Coulomb, and exchange-correlation components. The former two approximately cancel, whilst for the functionals considered, the latter gives a potential of approximately \(-\xi/2R_0 \), where \( \xi \) is the amount of long-range exact exchange; this dependence arises because the potential due to exact exchange in a two-electron system is minus half the coulomb potential. An electron near A therefore sees a fractional positive charge on the other atom of approximate magnitude \( \xi/2 \), which causes an unphysical distortion towards it. The problem becomes increasingly pronounced as \( \xi \) increases, with the extreme case (\( \xi = 1 \)) occurring for CAM-B3LYP-X and Hartree-Fock. In this case, the electron near A sees a positive charge of 0.5 on the other atom, in line with the standard interpretation that the Hartree-Fock wavefunction contains 50% ionic character. The CAM-B3LYP-X forces in Table I are close to \(-1/2R^2 \), consistent with 50% ionic character.

The static correlation error is so large for these DFT functionals that it conceals the fact that they completely omit the physics of long-range dispersion. In the absence of fractional spins, none of the functionals would reproduce the CCSD density distortion responsible for the dispersion force. For example, calculations on stretched \( \text{H}_2 \), which dissociates to atoms containing integer spin electrons, would yield a distortion much smaller than that of CCSD. For an explicit illustration of this within the context of Hartree-Fock-Kohn-Sham calculations, the reader is referred to Ref. 26.

Next consider \( \text{H}_2^+ \). Table II presents the forces evaluated at \( R = 13 \ a_0 \). The \( F_A \) values correspond exactly to the negative of the slopes in Figure 1(b) at \( R = 13 \ a_0 \). Once again, \( F_A \approx F_A^{el} \) in all cases, indicating the validity of the electrostatic theorem. In contrast to \( \text{H}_2 \), however, \( F_A^{el} \) is now very

<table>
<thead>
<tr>
<th>Method</th>
<th>( F_A )</th>
<th>( F_A^{el} )</th>
<th>( F_A^{el,\rho_A} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hartree-Fock (exact)</td>
<td>(-4.95 \times 10^{-5})</td>
<td>(-4.74 \times 10^{-5})</td>
<td>(-2.91 \times 10^{-5})</td>
</tr>
<tr>
<td>B</td>
<td>(+1.47 \times 10^{-3})</td>
<td>(+1.60 \times 10^{-3})</td>
<td>(+0.00 \times 10^{-3})</td>
</tr>
<tr>
<td>B3</td>
<td>(+1.17 \times 10^{-3})</td>
<td>(+1.17 \times 10^{-3})</td>
<td>(+0.00 \times 10^{-3})</td>
</tr>
<tr>
<td>CAM-B3</td>
<td>(+4.91 \times 10^{-4})</td>
<td>(+4.93 \times 10^{-4})</td>
<td>(+2.64 \times 10^{-5})</td>
</tr>
<tr>
<td>CAM-B3-X</td>
<td>(-4.09 \times 10^{-5})</td>
<td>(-3.86 \times 10^{-5})</td>
<td>(-2.98 \times 10^{-5})</td>
</tr>
</tbody>
</table>
different to $F_{A}^{\text{el,prot}}$, meaning the force on nucleus A is not primarily due to $\rho_{A}(r)$—the density $\rho_{B}(r)$ now only contains half of an electron localised around nucleus B and so the attractive force due to this density only approximately cancels half of the repulsive force due to nucleus B,

$$\tilde{F}_{A}^{\text{el,prot}} \approx \frac{1}{2R^2}.$$  

(8)

It follows from Eq. (3) that

$$F_{A} \approx F_{A}^{\text{el}} \approx \frac{1}{2R^2} + F_{A}^{\text{el,prot}},$$  

(9)

meaning that the net force on nucleus A comprises a repulsive force of approximately $1/2R^2$ due to the other proton/half-electron fragment plus the force due to $\rho_{A}(r)$. Explicitly, evaluating this at $R = 13a_0$ gives

$$F_{A} \approx F_{A}^{\text{el}} \approx 2.96 \times 10^{-3} \text{ a.u.} + F_{A}^{\text{el,prot}}.$$  

(10)

Consider the Hartree-Fock (exact) forces in Table II. $F_{A}$ is just $-4.59 \times 10^{-5}$ a.u., corresponding to a very weak attractive force; the value is in excellent agreement with the value determined numerically from energies in Ref. 27. It follows from Eqs. (9) and (10) that $F_{A}^{\text{el,prot}}$ is an attractive force, with a magnitude marginally larger than the repulsive force due to the proton/half-electron fragment; the value is $F_{A}^{\text{el,prot}} = -2.99 \times 10^{-3}$ a.u. The weak net interaction therefore arises from a delicate balance between two much larger forces (almost two orders of magnitude larger!). The significant value of $F_{A}^{\text{el,prot}}$ indicates a significant density distortion in the exact

FIG. 3. H$_2$ density distortions, $\Delta\rho(r)$. (a) CCSD (exact), (b) BLYP, (c) B3LYP, (d) CAM-B3LYP, and (e) CAM-B3LYP-X.
case, which is illustrated in Figure 4(a). At first sight, such a distortion at large $R$ is surprising, but it simply reflects the long-range inductive effect of the other proton/half-electron fragment.

Next consider the DFT results. For the B GGA functional (i.e., the exchange part of BLYP), the value of $F_{\text{el}}^{\rho_A}$ is only 50% of the exact Hartree-Fock value. The attractive force due to $\rho_A(r)$ is therefore smaller than the repulsive force due to the proton/half-electron fragment, leading to a net repulsive force. The value of $F_{A} = +1.47 \times 10^{-3}$ a.u. is essentially $1/4R^2$, as would arise from the Coulomb repulsion between two half-plus charges. The reason that $F_{\text{el}}^{\rho_A}$ is too small is evident in Figure 4(b)—the density distortion is too small. Interestingly, the use of the Dirac local density approximation exchange functional and other GGA exchange functionals yield essentially identical forces and density distortions to those from the B GGA functional, reflecting the fact that all such approximations are fundamentally local. This analysis provides a simple explanation for the characteristic barrier in the potential energy curves of $H_2^+$ (evident at intermediate $R$ in Figure 1(b) and earlier studies)—the force changes from attractive to repulsive as $R$ increases.

As the fraction of long-range exact exchange increases, the forces $F_{\text{el}}^{\rho_A}$ in Table II become increasingly attractive, eventually overcoming the repulsive force due to the proton/half-electron fragment, leading to a net attractive force as in the exact case. The CAM-B3LYP-X forces are in good agreement with the exact values. The increased values of $F_{\text{el}}^{\rho_A}$ arise because the density distortions become increas-
ingly pronounced, as shown in Figure 4; the variation with exchange-correlation functional is less than in Figure 3 because \( F^{\alpha,\rho_0}_A \) varies much less in \( H_2^+ \) than in \( H_2 \). The delocalisation error therefore manifests itself as an underestimated attractive force due to an underestimated density distortion, which can again be understood from a consideration of the Kohn-Sham orbital equation. In the exact case, the potential near nucleus A arising due to the other fragment comprises not only the correct nuclear potential, but also a self-interaction (Coulomb plus exchange-correlation) potential of approximately \(-1/r_A - 1/r_B\), i.e., it should just see the nuclei, and hence a unit positive charge on the other fragment, which causes the significant density distortion. Within DFT, the potential near nucleus \( A \) must experiences a potential of approximately \(-\xi/2\) on the other fragment. For small \( \xi \), this is smaller than the correct value (unity) and so the distortion is too small. As \( \xi \) increases to unity, the charge increases and the distortion and force become larger and more accurate.

In conclusion, we have considered the key problems of dispersion, static correlation, and delocalisation errors in DFT from the unconventional perspective of the force on a nucleus in a stretched diatomic molecule. We used the electrostatic theorem to relate errors in the forces to errors in the electron density distortions, which in turn were related to erroneous terms in the Kohn-Sham equations arising due to the approximate exchange-correlation functional. For \( H_2 \), we illustrated Feynman’s physically appealing—yet not widely appreciated—density distortion explanation of the dispersion force, whereby nucleus \( A \) is pulled primarily by its own distorted density towards \( B \). The static correlation error (associated with fractional spins) manifests itself as an underestimated attractive force due to an exaggerated density distortion. For \( H_2^+ \), we highlighted the delicate balance between the electrostatic force components. The delocalisation error (associated with fractional charges) manifests itself as an underestimated attractive force due to an underestimated density distortion; for functionals with insufficient long-range exact exchange, the distortion is too weak to overcome the repulsion from the other proton/half-electron fragment, leading to a net repulsive force and the characteristic barrier in the potential energy curves. Increasing the fraction of long-range exact exchange in the functional increases the density distortion, which reduces the delocalisation error but increases the static correlation error. We hope that the physical insight provided by this elementary analysis will stimulate further theoretical developments into these important problems.

We are grateful to Dr. Aron J. Cohen and Dr. Michael J. G. Peach for helpful discussions and to the Engineering and Physical Sciences Research Council (EPSRC) for studentship support.

21. DALTON, a molecular electronic structure program, release 2.0, 2005; see http://www.kjemi.uio.no/software/dalton/dalton.html.