Quantifying and understanding errors in molecular geometries

Stefan Vuckovic and Kieron Burke*

Departments of Chemistry and of Physics, University of California, Irvine, CA 92697, USA

July 29, 2020

Abstract

Electronic structure calculations are ubiquitous in most branches of chemistry, but all have errors in both energies and equilibrium geometries. Quantifying errors in possibly dozens of bond angles and bond lengths is a Herculean task. A single natural measure of geometric error is introduced, the geometry energy offset (GEO). GEO links many disparate aspects of geometry errors: a new ranking of different methods, quantitative insight into errors in specific geometric parameters, and insight into trends with different methods. GEO can also reduce the cost of high-level geometry optimizations and shows when geometric errors distort the overall error of a method. Results, including some surprises, are given for both covalent and weak interactions.

Whenever one runs an electronic structure calculation within the Born-Oppenheimer approximation, whether it is a density functional calculation, *ab initio* or semiempirical, of a molecule or a material, one must always answer the question: Which geometry should I use? Whatever the limitations of your method are, they will show up in giving an approximate energy at any given geometry which will minimize at some approximate geometry. Sometimes the differences between the true geometry and the approximation are so slight that it does not matter. Whenever it does matter, common sense often dictates a choice: When comparing different methods, the requirement of apples-to-apples comparison means comparing several methods with a fixed geometry. [1, 2, 3] Other times, when the cost of a single calculation is severe, geometry optimization is prohibitively expensive, and one resorts to using geometries from a cheaper method.

This problem is compounded when comparing geometric parameters computed with different methods. As a molecule grows in size, there are $3N-6$ distinct degrees of freedom for the equilibrium structure, with errors in bond lengths, angles, etc. Some are more accurate in one method, some are better in another (see, e.g., Ref. [4]). Should one average over all such parameters? But what if one method is better for bond lengths, and another for angles? And how do such errors in geometry correlate with other energetic errors?

We define the geometry energy offset of a given method as

$$E_{geo} = E\left(\tilde{\mathbf{G}}\right) - E\left(\mathbf{G}_0\right) \tag{1}$$

where $E(\mathbf{G})$ is the ground-state energy at geometry $\mathbf{G}$, $\mathbf{G}_0$ the exact geometry and $\tilde{\mathbf{G}}$ an approximate geometry. This simple definition leads to all the analysis and results contained in the paper. Figure 1 summarizes some of our most important results with GEO, with more details within this paper and supplementary information. On the left, we plot GEO energies averaged over a data set of small organic molecules (top). Accurate calculation of GEO using CCSD(T) is expensive (see methods), but using the approximations themselves is less expensive and yields nearly identical results (E'_{geo}). We can see that different approximations perform characteristically well or poorly. Thus every method can be ranked by its GEO value, and some perform much better for geometries than they do for, e.g., atomization energies. This is crucial information for understanding the accuracy of different methods for geometries. Directly below, we evaluate a much greater variety of methods for a data set of medium-sized organic molecules. Here, we use B2PLYP as the reference, since it is the winner in the top panel and CCSD(T) is already too expensive. This shows some surprises: lower level (and less costly) methods can outperform higher level methods because they have been trained empirically. For example the semiempirical GFN1-xTB method of Grimme and co-workers competes with DFT with the PBE functional[5], and outperforms DFT with BLYP[6, 7].

In the center, we show the very disparate behavior of two popular representative density functionals for single bonds and for double bonds. PBE, as a generalized gradient approximation (GGA) is far more accurate for double bonds than for singles. But a (global) hybrid, B3LYP, totally reverses this trend: more accurate for single bonds, but surprisingly far less accurate for double bonds. Explanations of the accuracy of hybrids[8, 9] typically center on atomization energies, not bond lengths, and do not explain these trends. The top right panel shows a trade-off between angle- and bond-length errors. There are clear behaviors of different levels of density functionals. The local density approximation has noticeable errors in both the bond length and angle (but far smaller than those of HF). One can clearly see how GGA's like PBE and BLYP and the meta-GGA TPSS [10] greatly reduce the angular error, but have almost no effect on the bond length. Finally, by mixing some fraction (about 1/4) of exact exchange, PBE0 and B3LYP lie along a line joining their parent GGA to HF, and the mixing fraction almost perfectly cancels the bond-length error,

*kieron@uci.edu

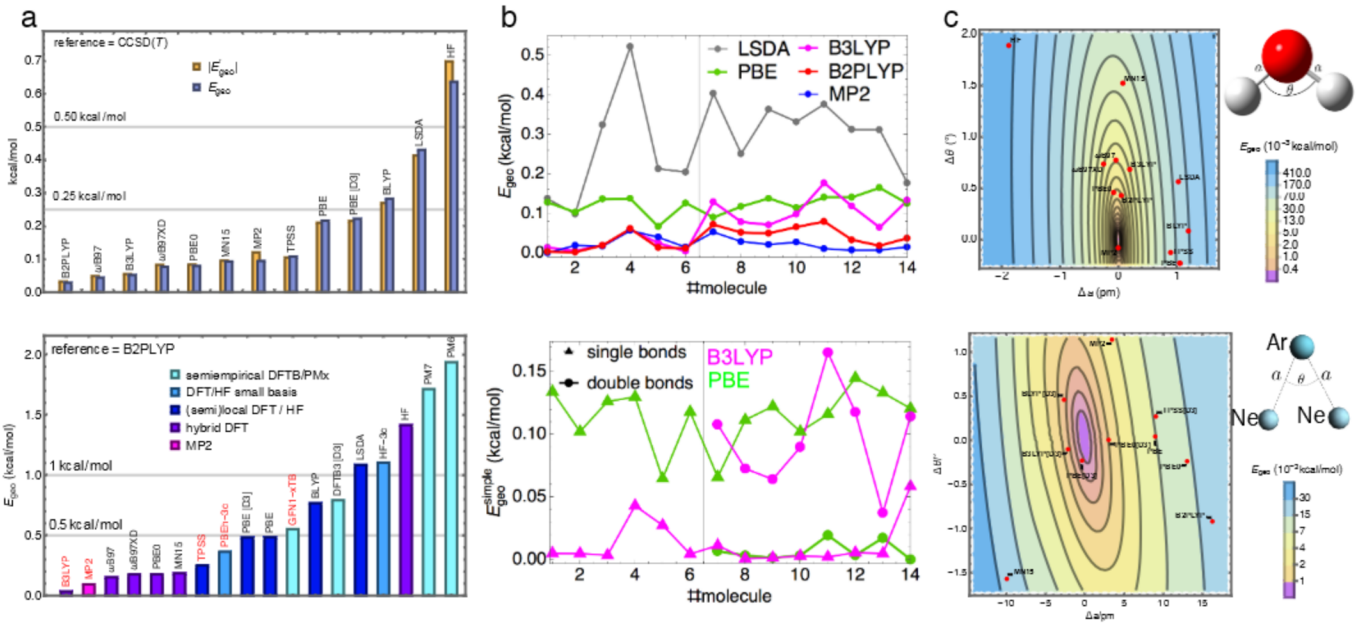

Figure 1: *Why GEO is useful. Left: Exact and approximate GEO rankings for quantum chemical methods on small molecules (top), and GEO of many different methods over medium-sized organic molecules (bottom) with B2PLYP as reference. For the lists of molecules and further details, see Figs. S2-S6, and Tables S1-S3. Center: E_{geo} for a few methods and a few molecules (top), single and double-bond contributions to $E_{geo}^{\mathrm{simple}}$ for two popular methods, showing huge difference in their accuracies (bottom). For the list of molecules, see Fig. 3. Right: GEO contours as a function of errors in the bond angle and length for the water molecule (top panel) and weakly bonded Ne_2Ar (lower panel), and positions of different approximations.*

while increasing the angle-error. Better functionals have about the same accuracy, while MP2 has almost perfect geometry.

All these results and trends are for strong covalent bonds. But GEO is even more important for weak interactions, where GEO energies can be comparable to the binding energy itself. To illustrate this, in the right, we contrast contours of GEO for two A_2B molecules, one covalently bonded and the other a non-covalent interaction: Water and the van der Waals trimer, Ne_2Ar, with the different methods from the left figures plotted as points in the plane. The non-covalent case is strikingly different. First, its binding energy is only 0.37 kcal/mol, so GEO errors are now more than relevant on this scale. This is accompanied by huge errors in bond length, related to the softness of the potential. Finally, the performance of different electronic structure methods is very different from the covalent case. For covalent bonds, MP2 is exceptionally good; for NCI's, approximate density functionals are much better. These effects seem to have largely been ignored when ranking functionals for such complexes, which is usually done at a fixed geometry. [1, 2, 3] We expect improved performance for weak interaction methods once GEO errors are accounted for.

The rest of this paper explains how GEO works and shows how useful it can be. We focus on just three immediate applications: (i) obtaining insight into geometric errors in molecular benchmark energy sets; (ii) establishing an energetic scale comparing the quality of geometries from different approximate quantum-mechanical (QM) solvers; (iii) how this scale can be used for choosing a geometry optimization solver that has a good accuracy to cost ratio. We apply our logic first to covalent bonds, where GEO is typically negligible relative to atomization energies, and then to non-covalent bonds, where GEO is often comparable to binding energies, and so is even more important.

Results and Discussion

Performance of approximations. One of the most valuable uses of GEO is to rank different electronic structure methods for their geometric accuracy, as illustrated in the upper left panel of Fig 1. We stress that this ranking is quite different from traditional rankings by purely energetic performance, such as for atomization energies (AE). In Fig. S2 (top panel), we give the errors for AE on our set of small organic molecules, and in Fig. S2 (lower panel), we show the correlation plot between GEO and MAE for AE. Roughly, GEO is typically about 1/80-th of the mean absolute error in AE. By such a measure, LSDA and HF are surprisingly good for geometries, because their MAE for atomization is so poor. Non-empirical functionals (PBE, TPSS) are close to this line, so improvements in AE's are reflected in improvements in GEO. Empirical functionals typically perform very well, especially the global hybrid B3LYP, and BLYP, as a GGA, yields surprisingly poor geometries. Also, the addition of D3 corrections to any functional makes little difference to its geometric performance for covalent bonds.

If we want to calculate GEO errors for larger main-group molecules, for which CCSD(T) is too expensive, we can use B2PLYP as a reference in place of CCSD(T). Multireference systems, such as transition metal dimers, are more delicate and would need a better reference than CCSD(T) [11, 12] to calculate GEO.

The bigger picture is shown in the lower-left panel of Fig. 1, which includes many different kinds of methods. Here, we had to use B2PLYP [13] as a reference (see above). Besides the QM solvers considered in Figure 1a(top), we also include semiempirical QM solvers, such as DFTB[14, 15] and PMx[16, 17], and the highly practical HF-3c[18] and PBEh-3c[19], both of which use a small basis set and contain empirical parameters. The results are summarized in the lower panel of Figure 1a, where different error bar colours indicate methods at different levels of theory, with the best method for each level of theory shown in red. Trends are similar to the panel above, but overall GEO's are larger as the molecules are bigger. This plot is not accurate below about 0.1 kcal/mol, because of the B2PLYP reference. Thus B3LYP does not really rank as No. 1, its errors are simply correlated with the reference. As we discuss later, since GFN1-xTB[15] has the best performance among all semiempirical methods shown, it can serve as an excellent starting point in optimization schemes.

Geometry optimization. Based on GEO, one can establish the following *sequence* composed of the best method for each level of complexity: GFN1-xTB $\rightarrow$ TPSS (or PBEh-3c) $\rightarrow$ B3LYP $\rightarrow$ B2PLYP. This sequence can be used in automated explorations of chemical space and molecular screenings assisted by QM solvers,[20, 21, 22] which are powerful tools for the discovery of new molecules with desired properties. In these procedures, based on energetic criteria (e.g., their binding energies with a specific enzyme) molecules are filtered out, and QM geometry optimizations of a large number of molecules make the procedure computationally demanding. As the number of molecules in the screening decreases, more expensive and more accurate methods are typically used. Thus, based on our sequence determined by the GEO criterion, in the first step of the screening one can employ GFN1-xTB for optimizing geometries of all initial molecular candidates. After the first cycle of filtering out molecules, TPSS[10] can be employed as an optimizer, and so on. In the last round, B2PLYP geometries can be confidently used, given they are energetically very close to the CCSD(T) ones ($\sim$0.03 kcal/mol for the testset considered in Figure 1a). Even if one only wants the CCSD(T) geometries for small molecules, one can use the same *sequence* to pre-optimize the molecular geometry, before the CCSD(T) optimizer is turned on, and thereby save computational time.

Simplifications and analysis tools. As mentioned above, a much less costly calculation is

$$E'_{geo} = \tilde{E}\left(\tilde{\mathbf{G}}\right) - \tilde{E}\left(\mathbf{G}_0\right), \tag{2}$$

where $\tilde{E}$ is the approximate energy (see Supplementary Section 1 for more mathematical details). This is typically an excellent approximation to GEO, because even inaccurate methods such as HF yield reasonable vibrational frequencies.[23] From Figure 1a, the mean E_{geo} error is in close agreement with its E'_{geo} counterpart for all approximations (see also Tables S1 and S2 for E_{geo} and E'_{geo} values for individual molecules).

The next simplification is to approximate GEO by expanding E_0 around its minimum to second order:

$$E^{\text{harm}}_{geo} = \frac{1}{2}\Delta\tilde{\mathbf{G}}^{\intercal}\mathbf{H}_0\Delta\tilde{\mathbf{G}}, \tag{3}$$

where $\mathbf{H}_0$ is the Hessian at the minimum, composed of force constants and $\Delta\tilde{\mathbf{G}} = \tilde{\mathbf{G}} - \mathbf{G}_0$ is the error in specific geometric parameters (degrees of freedom) that determine the relative positions of nuclei. These could be simple Cartesians or any other choice of coordinates. Again, this is extremely accurate when approximated with most electronic structure methods (see Figures S17- S22). Thus, we can use Eq. 3 for further analysis and decomposition of GEO. One can easily diagonalize $\mathbf{H}_0$ and obtain GEO modes, p. In these coordinates, Eq. 3 becomes:

$$E^{\text{harm}}_{geo} = \frac{1}{2}\sum_{i}^{N_g} f_i^p \left(\Delta p_i\right)^2, \tag{4}$$

where f_i^p are the underlying force constants ($\mathbf{H}_0$ eigenvalues) and $\Delta p_i = \tilde{p}_i - p_i$ represent $\Delta\tilde{\mathbf{G}}$ written in terms of the errors in the GEO normal modes ($\mathbf{H}_0$ eigenvectors) and $N_g \leq 3N-6$ is the number of GEO *active* modes. A highly appealing feature of Eq 4 is that each term contributes positively to E^{harm}_{geo}, which, in turn, allows us to obtain weights of each modes' contribution to the total GEO. In Figures S27- S29, for a set of small molecules we analyse how each of the N_g GEO-active modes contributes to the total E^{harm}_{geo} for different approximations and we also show the *GEO-inactive* modes, those that have no contribution to E^{harm}_{geo}. For example, by symmetry, no (sensible) electronic structure approximation gives unequal OH bond lengths in the water molecule, so the asymmetric stretch of the OH bond is GEO-inactive (see Fig. S5). The higher the symmetry of a molecule, the fewer modes are GEO active. For ethene, all modes that distort its D_{2h} symmetry (asymmetric and out-of-plane vibrations) are GEO-inactive, so only 3 of its 12 modes are GEO-active, as shown in Fig. S29.

Besides the GEO modes, a more chemically intuitive analysis in terms of bond lengths, angles, and torsion angles of E_{geo} can be obtained by considering Eq. 3 in internal coordinates. Considering only the underlying diagonal elements of $\mathbf{H}_0^q$ (the Hessian in internal coordinates), we find the following *simple* approximation to E^{harm}_{geo}:

$$E^{\text{harm}}_{geo} \approx E^{\text{simple}}_{geo} = \frac{1}{2}\sum_{i}^{3N-6} f^q_{i,i}\left(\Delta q_i\right)^2. \tag{5}$$

where $\Delta q_i = \tilde{q}_i - q$ are the errors in internal coordinates. While the r.h.s of Eq. 4 is exactly equal to E^{harm}_{geo}, this is not so for

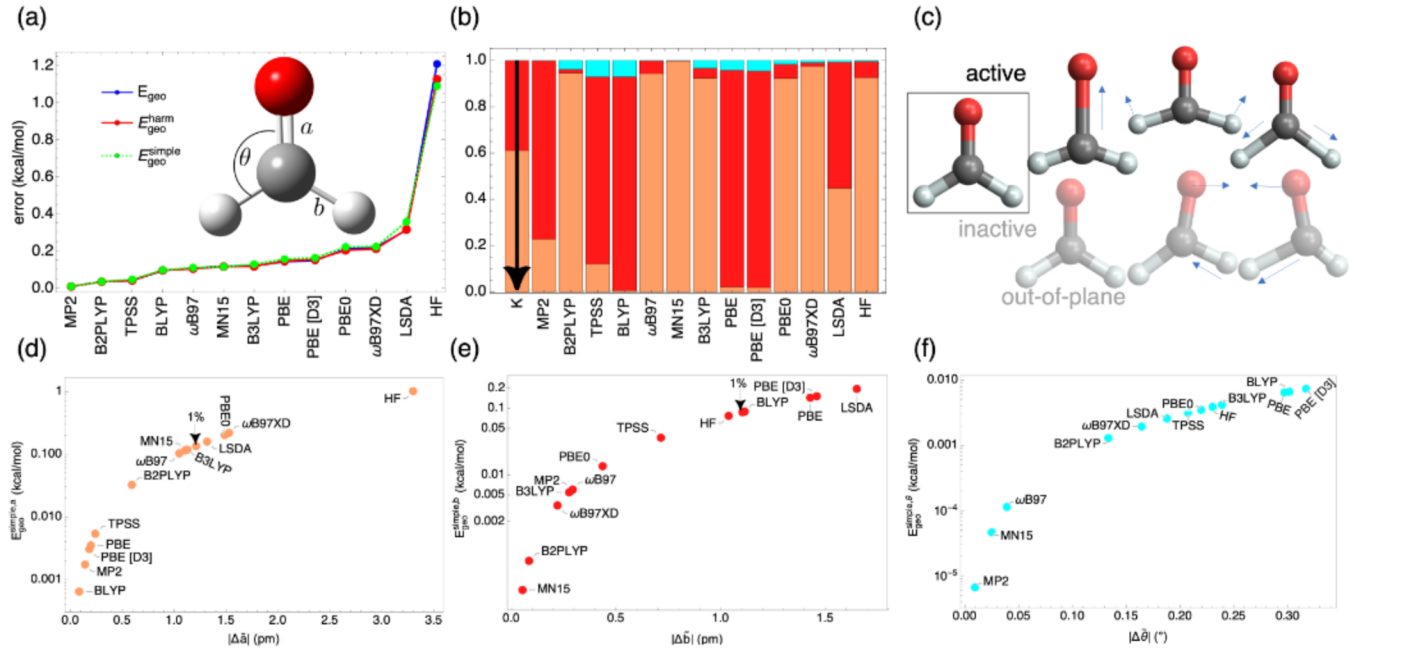

Figure 2: *GEO analysis for formaldehyde. Top: (a) GEO rankings of approximations. The plots also show that E_{geo} is accurately approximated by E_{geo}^{harm} (Eq. 3) and $E_{geo}^{\mathrm{simple}}$ (Eq. 5). (b) $E_{geo}^{\mathrm{simple}}$ weights, $E_{geo}^{\mathrm{simple,i}}$ / $E_{geo}^{\mathrm{simple}}$ (see lower panels for colour legends). When all bonds are stretched by the same γ factor, GEO becomes E_{geo}^{γ} (see the text), and the underlying weights are marked by the arrow. (c) GEO-active and GEO-inactive modes (those that have no contribution to the r.h.s. of Eq 4). For the E_{geo}^{harm} weights in normal modes, analogous to panel (b), see Fig. S28. (d)-(f) Plots showing errors in individual geometric parameters (x-axis), and how these errors translate to $E_{geo}^{\mathrm{simple}}$ terms by virtue of Eq. 5 (y-axis). The points marked by the arrows, show GEO when the bond lengths are stretched by 1%.*

$E_{geo}^{\mathrm{simple}}$, since the off-diagonal $\mathbf{H}_0^q$ are typically small but non-zero. For the organic molecules we consider here, $E_{geo}^{\mathrm{simple}}$ is typically in good agreement with both E_{geo}^{harm} and the "exact" E_{geo} (see Figures S17- S22). This, in turn, allows us to *safely* use Eq. 5 to decompose E_{geo}^{harm} into its positive contributions arising from errors in specific geometric parameters.

In Figure 2, we illustrate how a GEO analysis works for a simple case, formaldehyde. In panel (a), we give GEO rankings of different approximations for this molecule, which somewhat align with the database averages of Fig 1(a). As with all covalent cases we studied, E_{geo}^{harm} and $E_{geo}^{\mathrm{simple}}$ are in excellent agreement with GEO, which allows us to use Eq 5 to decompose contributions from different structural parameters. The fractional contributions of each coordinate are shown in panel (b). Angle errors give only a minor contribution to GEO for all methods. For the hybrids, GEO error comes nearly entirely from the error in the double bond, while in the case of semilocal functionals, nearly the entire GEO error comes from the error in the single bond lengths, consistent with the trends in Fig 1(b). The rankings for the single and double bond lengths and the bond angle are shown in the lower panels, and how they correlate with $E_{geo}^{\mathrm{simple}}$. In each case, the actual curve is parabolic (the GEO axis is logarithmic). The rankings differ substantially from those for the total GEO. The leftmost is the double bond, and here the semilocal functionals (no mixing of exact exchange) do best, outperforming even B2PLYP! The hybrids do no better than simple LSDA. But, the roles are reversed in the middle panel, showing that hybrids greatly improve single-bond length error. Finally, the angle-error is shown, with a variety of results, but no clear trends. The right panel of Fig S12 shows that MP2 and B2PLYP yield the best angles on average for the set of molecules considered in Fig. 1b. All three curves are monotonic, so rankings by a $E_{geo}^{\mathrm{simple}}$ contribution correspond directly to rankings by the error in the underlying geometric parameter.

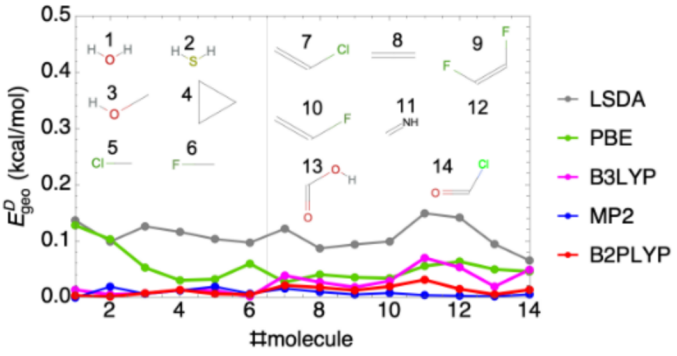

Figure 3: *The same plot as in the top panel of Fig. 1b but with geod (see the absolute GEO scale section) on the y-axis. For more plots comparing E_{geo} and E_{geo}^{D}, see S7.*

Absolute GEO scale. A problem that bedevils benchmarking of atomization energies is whether to consider total energy errors or errors per bond. Here we show that there exists a universal GEO scale, independent of any method, that overcomes this problem for geometric errors. Consider a small expansion of all coordinates, $\Delta\mathbf{G} = \gamma\mathbf{G}_0$, producing $E^{\gamma}_{geo} = \gamma^2 D/2$ where $D = \mathbf{G}_0^{\mathrm{T}}\mathbf{H}_0\mathbf{G}_0$. Thus, E^{γ}_{geo} is the GEO value for a very specific geometric error, that of expansion (or compression) of the exact geometry. For our small molecules, if $\gamma = 1\%$, E^{γ}_{geo} is a fraction of a kcal/mol. Thus, any calculation of GEO by any method for any molecule can be compared to this intrinsic property of the molecule. Moreover, E^{γ}_{geo} scales with the size of the molecule (compare, e.g., D values for small molecules shown in Table S7 with those for medium-sized molecules shown in Table S7), so that GEO's measured relative to it do not grow with molecular size. We can even decompose E^{γ}_{geo} in terms of Hessian eigenvectors or simple internal coordinates, giving an internally defined distribution of contributions. This only includes bond lengths, as no angle changes when molecules are uniformly expanded.

We define: $E^{D}_{geo} = E_{geo}\,(D_{\mathrm{H_2O}}/D)$, and in Figure 3, we repeat the plot in the top panel of Fig. 1b., but with E^{D}_{geo}, which varies much less with molecular size. In Table S6, we show that approximate calculations of D typically yield highly accurate estimates (even HF is not too bad). In Figure S30, we show the decomposition of D for each bond in each molecule, with double bonds being about 0.25 and most singles being about 0.1 times 10^4 kcal/mol if to an H atom, and about 0.16 if between heavier atoms. For rare gas dimers (bonded by weak interactions), D values are several orders of magnitude smaller (see Table S9).

Returning to Fig 2, the leftmost column of (b) is the D-decomposition of the single versus double bond, showing that under expansion, 60% of the energy cost is to stretch the double bond, 40% to stretch the single. Then BLYP clearly makes an unusually small error in the double, and a relatively large error in the single.

Non-covalent interactions. The rest of this paper is devoted to weak interactions. We re-examine all aspects of GEO for these cases, as GEO energies can be a more significant fraction of the binding energies here. Force constants for weak bonds are so much weaker that even very small GEO values can lead to large errors in bond lengths. For weak interactions, we include D3 corrections [26] to the approximations, which typically greatly improve energetic accuracy. We also only allow the weak bond length to vary in complexes, i.e., one degree of freedom.

The upper panel shows a prototypical van der Waals system, Ne_2, with the exact curve and various approximations. In most cases, the geometries are quite accurate (minima are marked by beads) so that GEO energies are very small. Surprisingly, ωB97XD [27] has a large geometric error, but nonetheless yields a highly accurate energy minimum, a cancellation of geometric and non-geometric errors (apparently missed by its

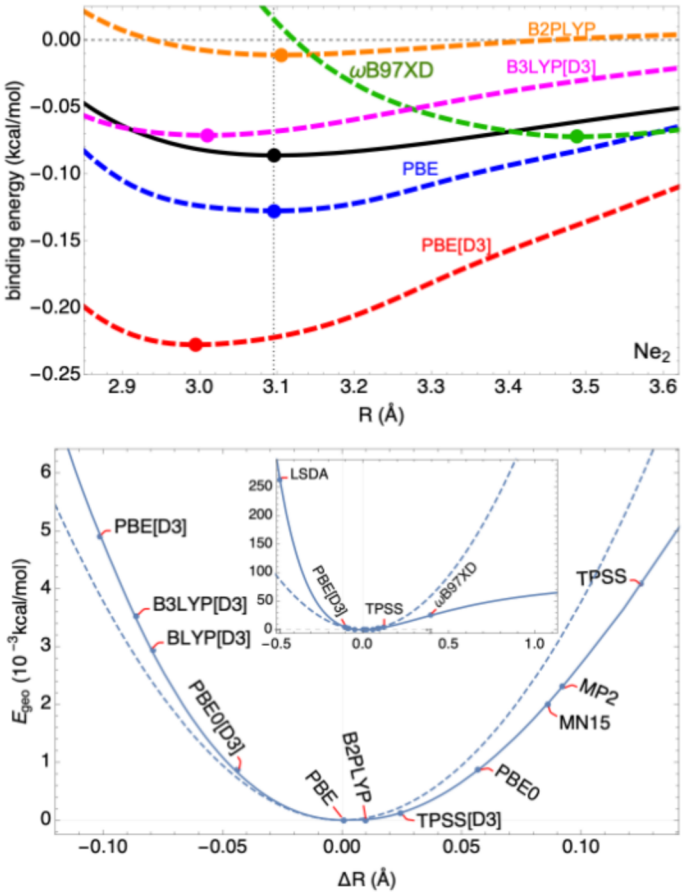

Figure 4: *GEO analysis for the Ne_2 binding energies. Top panel: Binding curves of Ne_2 with various methods. Lower panel: GEO for different approximations. For more details, see Figs. S35- S38.*