Supplemental Material for
“Quantifying and understanding errors in molecular geometries"

STEFLAN VUCKOVIC AND KIERON BURKE
Departments of Chemistry and of Physics, University of California, Irvine, CA 92697, USA
July 29, 2020

Contents

S1 Additional GEO mathematical details
S2 Additional details on the results in Figure 1a (top panel)
S3 Additional details on the results in Figure 1a (bottom panel)
S4 Details on the results of Figure 1b
S5 The GEO analysis for the water molecule
S6 GEO for small molecules in internal coordinates
S7 GEO analysis in normal modes (Hessian eigenvectors)
S8 Details on the $\gamma$ expansion and $D$ values for molecular sets
S9 Statistical measures for the quality of approximate geometries of the methylamine molecule
S10 GEO analysis for noble gas dimers
S11 Details on the results of Figure 5 [S66 results]
List of Tables

S1 \( E_{\text{geo}} \) values of Figure 1a (top panel) in kcal/mol for individual molecules. ................................................................. S4
S2 \( E_{\text{geo}} \) values of Figure 1a (top panel) in kcal/mol for individual molecules. ................................................................. S4
S3 \( E_{\text{geo}} \) values Figure 1b (bottom panel) for individual molecules and approximate methods. ............................................. S11
S4 \( E_{\text{geo}} \) values for molecules in Figs. S17–S22. ..................................................................................................................... S26
S5 Normal-mode GEO analysis for the H\(_2\)O molecule. ................................................................................................................ S28
S6 \( D \) values for molecules considered in Fig. 1b. ....................................................................................................................... S34
S7 \( D \) values for molecules considered in Fig. 1a(top). ................................................................................................................ S35
S8 \( D \) values for individual molecules considered in Fig. 1a (bottom). ..................................................................................... S36
S9 \( D \) values in 10\(^4\) kcal/mol for noble gas dimers. ..................................................................................................................... S36

List of Figures

S1 The decomposition of \( \Delta E \) error by Eqs. S1 and S3 for the atomization energy of the F\(_2\) molecules obtained from the B2PLYP (\( E_{\text{geo}} \) and \( E_{\text{geo}}' \) are negligible) and MN15 functionals (\( E_{\text{geo}} \) and \( E_{\text{geo}}' \) are non-negligible) ........................................ S3
S2 GEO error vs. atomization energy errors for the set of molecules and approximations shown in Fig. 1b (top panel).................. S5
S3 GEO error vs. atomization energy errors for the set of molecules and approximations shown in Fig. 1b (top panel).................. S6
S4 Plots showing data of Table S1. ................................................................................................................................................. S7
S5 Molecules used in error plots of Figure 1b (bottom panel). ...................................................................................................... S9
S6 \( E_{\text{geo}} \) errors for different approximate methods for the set of molecules shown in Figure S5. ............................................. S10
S7 \( E_{\text{geo}} \) and \( E_{\text{geo}}' \) for a set of molecules shown in Fig. 1b. ........................................................................................................ S12
S8 Bond vs. angle contribution to \( E_{\text{simple}} \) for a set of molecules shown in Fig. 1b. ................................................................. S13
S9 Total bond vs. double bond contribution to \( E_{\text{simple}} \) for a set of molecules shown in Fig. 1b. ............................................. S13
S10 \( E_{\text{geo}} \), \( E_{\text{harm}} \) and \( E_{\text{simple}} \) plots for a set of molecules shown in Fig. 1b. ................................................................. S13
S11 PBE and B3LYP errors in bond lengths for a set of molecules shown in Fig. 1b. ................................................................. S13
S12 Single bond, double bond, angles \( E_{\text{simple}} \) rankings for a set of molecules shown in Fig. 1b. .................................................... S14
S13 Bond length, bond angle, and the \( E_{\text{geo}} \) errors for the water molecule. ................................................................................ S15
S14 Different \( E_{\text{geo}} \) vs. \( E_{\text{harm}} \) vs \( E_{\text{simple}} \) plots for the water molecule. .................................................................................. S16
S15 The ellipses satisfying Eqs. S10 and S11 for the water molecule. ......................................................................................... S16
S17 GEO analysis for the water molecule in internal coordinates. .............................................................................................. S18
S18 The same plots as in Figure S17 but for the H\(_2\)S molecule. ................................................................................................. S19
S19 The same plots as in Figure S17 but for the H\(_2\)CO molecule. ............................................................................................ S20
S20 The same plots as in Figure S17 but for the ethene molecule. .............................................................................................. S21
S21 The same plots as in Figure S17 but for the methanol molecule. ........................................................................................... S22
S22 The same plots as in Figure S17 but for the cyclopropane molecule. .................................................................................. S23
S23 Summary of Figs. S17–S22. ................................................................................................................................................. S25
S24 Normal-mode vs. internal coordinates GEO analysis for the H\(_2\)O molecule. ................................................................. S30
S25 Normal-mode GEO analysis for the H\(_2\)S molecule. ................................................................................................................ S31
S26 Normal-mode GEO analysis for the H\(_2\)CO molecule. ........................................................................................................ S32
S27 Normal-mode GEO analysis for the ethene molecule. ........................................................................................................ S33
S28 Normal-mode GEO analysis for the ethene molecule. ........................................................................................................ S33
S29 Normal-mode GEO analysis for the ethene molecule. ........................................................................................................ S33
S30 \( D_{\text{simple}} \) bond-decomposition for the set of molecules considered in Fig. 1b. ................................................................. S35
S31 \( E_{\text{geo}} \) and other statistical measures for the quality of approximate methyamine geometries ........................................... S37
S32 Other statistical measures for the quality of approximate geometries of the methyamine molecule ................................ S38
S33 Atomization energy errors for the methyamine molecule ..................................................................................................... S38
S34 Geo analysis for the neon dimer ........................................................................................................................................ S40
S35 Geo analysis for the neon dimer ........................................................................................................................................ S40
S36 Counterpoise corrected and counterpoise uncorrected binding curves for the neon dimer ..................................................... S40
S37 \( E_{\text{geo}} \) (y-axis) vs. \( \Delta E \) (x-axis) errors in kcal/mol for different methods and for the S66 dataset. .......................................... S43
S38 \( E_{\text{geo}} \) (y-axis) vs. \( \Delta E \) (x-axis) errors in kcal/mol for different methods and for the S66 dataset. .......................................... S43
S39 \( E_{\text{geo}} \) (y-axis) vs. \( \Delta E \) (x-axis) errors in kcal/mol for different methods and for the S66 dataset. .......................................... S44
S40 Benzene-uracil binding curves (interpolated from the 8 datapoints along the dissociation curve). .......................................... S45
S1. ADDITIONAL GEO MATHEMATICAL DETAILS

For an approximate quantum-mechanical (QM) solver $\tilde{E}(\tilde{G})$, the total error is given by:

$$\Delta E = \tilde{E}(\tilde{G}) - E(G_0),$$

(S1)

where $E(G)$ is the exact ground-state energy at geometry $G$, $G_0$ is the exact and $\tilde{G}$ is the approximate geometry. Thus, $\Delta E$ contains errors both due to the approximate geometry and approximate energy. To decompose this error into the GEO and "non-GEO" part, we add and subtract $E(\tilde{G})$ to the r.h.s of Eq. S1:

$$\Delta E = E(\tilde{G}) - E(G_0) = E_{\text{geo}} + \tilde{E}(\tilde{G}) - E(G_0).$$

(S2)

We can also add and subtract $\tilde{E}(G_0)$ to the r.h.s of Eq. S1 to obtain an alternative form of Eq. S2:

$$\Delta E = \tilde{E}(\tilde{G}) - E(G_0) = E_{\prime\text{geo}} + \tilde{E}(G_0) - E(G_0).$$

(S3)

The signs of $\Delta E_{\text{geo}}$ and $\Delta \tilde{E}_{\text{geo}}$ also dictate the following chain of inequalities:

$$E_p \leq \Delta E \leq E_{\prime p}.$$

(S4)

For an illustration of the error decomposition by means of Eqs. S1 and S3, see Fig. S1. As discussed throughout the text, $E_{\text{geo}}$ is typically accurately approximated by $-E_{\prime\text{geo}}$, and thus : $E_{\text{geo}} \approx \frac{1}{2} (E_p - E_{\prime p})$ (see Fig. S1).

Figure S1: The decomposition of $\Delta E$ error by Eqs. S1 and S3 for the atomization energy of the $F_2$ molecules obtained from the B2PLYP ($E_{\text{geo}}$ and $E_{\prime\text{geo}}$ are negligible) and MN15 functionals ($E_{\text{geo}}$ and $E_{\prime\text{geo}}$ are non-negligible).

For equilibrium structures (minima of potential energy surfaces), $E_{\text{geo}}$ is always positive. On the other hand, in the case of transition states (the first order saddle points of potential energy surfaces), the sign of $E_{\text{geo}}$ is not definite. In that case, the $H_{0}$ Hessian will have one negative eigenvalue. Thus, one term on the r.h.s of Eq 4 of the underlying $E_{\text{harm}}$ is negative, and all other are positive.

S2. ADDITIONAL DETAILS ON THE RESULTS IN FIGURE 1A (TOP PANEL)

The CCSD(T)/A’V5Z (see Ref. 4 for the basis set description) $E(G)$ values and $G$ geometries have been taken from Ref. 4. All other results have been obtained from the Gaussian16 calculations by always employing the same (A’V5Z) basis set. The $E_{\prime\text{geo}}$ and $E_{\text{geo}}$ errors for individual molecules and employed approximate methods are given in Tables S2 and S1, respectively. In Figure S2 we show the MAE for the atomization energies for the same set of molecules. The plot in Figure S4 shows the data of Table S1.
<table>
<thead>
<tr>
<th></th>
<th>HF</th>
<th>PBE</th>
<th>BLYP</th>
<th>TPSS</th>
<th>B3LYP</th>
<th>MN15</th>
<th>ωB97</th>
<th>ωB97XD</th>
<th>LSDA</th>
<th>PBE [D3]</th>
<th>MP2</th>
<th>B2PLYP</th>
<th>PBE0</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.636</td>
<td>0.216</td>
<td>0.283</td>
<td>0.108</td>
<td>0.052</td>
<td>0.093</td>
<td>0.044</td>
<td>0.077</td>
<td>0.430</td>
<td>0.222</td>
<td>0.094</td>
<td>0.029</td>
<td>0.079</td>
</tr>
<tr>
<td>RMS</td>
<td>0.755</td>
<td>0.222</td>
<td>0.317</td>
<td>0.122</td>
<td>0.060</td>
<td>0.102</td>
<td>0.054</td>
<td>0.090</td>
<td>0.452</td>
<td>0.228</td>
<td>0.234</td>
<td>0.032</td>
<td>0.087</td>
</tr>
</tbody>
</table>

**Table S1:** $E_{\text{geo}}$ values of Figure 1a (top panel) in kcal/mol for individual molecules.

<table>
<thead>
<tr>
<th></th>
<th>HF</th>
<th>PBE</th>
<th>BLYP</th>
<th>TPSS</th>
<th>B3LYP</th>
<th>MN15</th>
<th>ωB97</th>
<th>ωB97XD</th>
<th>LSDA</th>
<th>PBE [D3]</th>
<th>MP2</th>
<th>B2PLYP</th>
<th>PBE0</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-1.055</td>
<td>-0.196</td>
<td>-0.196</td>
<td>-0.065</td>
<td>-0.050</td>
<td>-0.071</td>
<td>-0.042</td>
<td>-0.117</td>
<td>-0.393</td>
<td>-0.199</td>
<td>-0.032</td>
<td>-0.018</td>
<td>-0.126</td>
</tr>
<tr>
<td>RMS</td>
<td>-1.583</td>
<td>-0.311</td>
<td>-0.539</td>
<td>-0.219</td>
<td>-0.037</td>
<td>-0.095</td>
<td>-0.025</td>
<td>-0.094</td>
<td>-0.361</td>
<td>-0.314</td>
<td>-0.030</td>
<td>-0.016</td>
<td>-0.080</td>
</tr>
</tbody>
</table>

**Table S2:** $E'_{\text{geo}}$ values of Figure 1a (top panel) in kcal/mol for individual molecules.
Figure S2: Top panel: MAEs for the atomization energies for the set of molecules (listed in Tables S2 and S1) [Fig. 1a (top panel)] obtained from different methods. The LSDA and HF MAEs are divided by the factor of 4.5. Bottom panel: MAEs from the top panel vs. average $E_{geo}$ errors for the same set of molecules.
Figure S3: The inverted $\gamma$, $\gamma_{inv} = \sqrt{\frac{2E_{geo}}{D}}$, averaged over the molecular dataset considered in 1a (top panel). The underlying $E_{geo}$ values are given in Table S1 and the $D$ values are given in Table S7.
Figure S4: Plots showing data of Table S1. The order of molecules is the same as in Table S1.
The dataset of molecules used in Figure 1a (bottom panel) is shown in Figure S5. The PBEh-3c and HF-3c calculations have been performed with the ORCA package [5]. The GFN1-xTB and DFTB3 [D3] calculations have been performed with the ADF package [6]. All other calculations have been performed with the G16 package. We used aug-cc-pVTZ [1] basis set for all DFT and HF calculations, except for PBEh-3c and HF-3c, which come with their own basis set. For DFTB3 [D3] calculations, we have used the set of parameters of Ref. [7] in tandem with the D3 correction with the Becke-Johnson damping function [8]. The set of parameters for PM6 and PM7 are those implemented in G16 when the keywords "PM6" and "PM7", respectively, are used. The $E_{geo}$ errors for individual molecules (Figure S5) and employed approximate methods are shown in Figure S6 and the underlying data values are given in Table S4.
Figure S5: Molecules used in error plots of Figure 1b (bottom panel).
Figure S6: Plots showing $E_{\text{geo}}$ errors for different approximate methods for the set of molecules shown in Figure S5. Averaged error bars are shown in Figure S3 (bottom panel). Note the difference in the y-axes scales.
Table S3: $E_{\text{geo}}$ values Figure 1a (bottom panel) for individual molecules and approximate methods, all values are in kcal/mol. The molecules are ordered in Figure S5.
S4. Details on the results of Figure 1b

Figure S7: Plots comparing $E_{geo}$ and $E_{geo}^D$ for selected methods and for a set of molecules shown in Fig. 1b.
Figure S8: Plots comparing total bond length and angle errors contribution to $E_{\text{geo}}^{\text{simple}}$ for selected methods and for a set of molecules shown in Fig. 1b.

Figure S9: Plots comparing total bond and double bond length errors contribution to $E_{\text{geo}}^{\text{simple}}$ for selected methods and for a set of molecules shown in Fig. 1b.

Figure S10: Plots comparing $E_{\text{geo}}$, $E_{\text{geo}}^{\text{harm}}$ and $E_{\text{geo}}^{\text{simple}}$ for selected methods and for a set of molecules shown in Fig. 1b.

Figure S11: PBE and B3LYP errors in bond lengths for a set of molecules shown in Fig. 1b. Left panel: relative errors; middle panel: mean absolute errors.
Figure S12: Rankings of approximations for a set of molecules shown in Fig. 1b based on averaged $E_{\text{geo}}^{\text{simple}}$ (over a number of molecules) contributions from: single bonds (left panel), double bonds (middle panel), angles (right panel).
S5. The GEO analysis for the water molecule

We use CCSD(T)/A’V6Z [4] as a reference method in all calculations performed in this section. The same basis set is employed in calculations involving approximate electronic structure methods and all calculations have been performed with the Gaussian16 package.

To show how GEO depends on errors in individual bond lengths and angles, we consider the water molecule in internal coordinates. It has three degrees of freedom, but to analyse the $E_{\text{geo}}$ error we need just two: $a$ and $\theta$ (see Figure S13) since the lengths of the two O–H bonds are the same within any approximation. In Figure S13 we also show the errors of different methods in the bond length, angle, together with the $E_{\text{geo}}$ errors.

From a series of CCSD(T) calculations at fixed geometries, we can see how $E_{\text{geo}}$ depends on $\Delta a$ and $\Delta \theta$, the error in the bond length and bond angle, respectively. The resulting contour plot is shown in Figure S14a. From this plot, we can see that the $E_{\text{geo}}$ error increases much more quickly with the error in the bond length than in the bond angle. To better understand this observation, we consider the harmonic approximation to $E_{\text{geo}}$ (Eq. 3). For clarity, we write this equation again:

\[ E_{\text{geo}} \approx E_{\text{geo}}^{\text{harm}} = \frac{1}{2} \Delta \mathbf{G}^\top \mathbf{H}(\mathbf{G}) \Delta \mathbf{G}. \]  

(S5)

For the water molecule in internal coordinates, we can now write $E_{\text{geo}}^{\text{harm}}$ as a function of:

\[ \Delta \mathbf{G} = (2\Delta a, \Delta \theta). \]  

(S6)

The underlying $\mathbf{H}(\mathbf{G})$ matrix in internal coordinates is given by:

\[ \mathbf{H}(\mathbf{G}) = \begin{pmatrix} k_{aa} & k_{a\theta} \\ k_{\theta a} & k_{\theta\theta} \end{pmatrix} \sim \begin{pmatrix} 0.122 \text{pm}^{-2} & 0.006 \text{pm}^{-1} \\ 0.006 \text{pm}^{-1} & 0.031 \text{pm}^{-2} \end{pmatrix} \text{kcal/mol} \]  

(S7)

where $k$-s are the force constants at the highly accurate reference [CCSD(T)] geometry. Plugging Eqs. S6 and S7 into Eq. S5, we obtain:

\[ E_{\text{geo}}^{\text{harm}} (\Delta a, \Delta \theta) = k_{aa}\Delta a^2 + \frac{1}{2}k_{\theta\theta}\Delta \theta^2 + 2k_{a\theta}\Delta a\Delta \theta, \]  

(S8)

where we also used the Hessian symmetry: $k_{a\theta} = k_{\theta a}$. In Eq. S7 the units were chosen such that $\Delta a$ errors in pm and $\Delta \theta$ errors in degrees yield $E_{\text{geo}}^{\text{harm}}$ in kcal/mol. In Figure S14b, we show the $E_{\text{geo}}^{\text{harm}}$ contour plot as a function of $\Delta a$ and $\Delta \theta$. We can see that it is in a very close agreement with its exact counterpart (Figure S14a). A further simplification of $E_{\text{geo}}^{\text{harm}}$ (Eq. S8) obtained by setting $k_{a\theta}$ to 0:

\[ E_{\text{geo}}^{\text{harm}} (\Delta a, \Delta \theta) \approx E_{\text{geo}}^{\text{simple}} (\Delta a, \Delta \theta) = k_{aa}\Delta a^2 + \frac{1}{2}k_{\theta\theta}\Delta \theta^2, \]  

(S9)

gives a sensible approximation to $\Delta E_{\text{geo}} (\Delta a, \Delta \theta)$, as it can be seen from Figure S14c. In Figure S14d, we compare the size of the three errors from different methods and we can see that both $E_{\text{geo}}^{\text{harm}}$ and $E_{\text{geo}}^{\text{simple}}$ are in a close agreement with $E_{\text{geo}}$. Thus,

Figure S13: The error in the bond length, bond angle, and the $E_{\text{geo}}$ errors for the water molecule obtained from various approximations.
Figure S14: (a) Contour plot of GEO as a function of errors in the bond lengths for the water molecule (also shown in the top panel of Fig. 1); (b) same plot from the harmonic approximation (Eq. S8); (c) same plot from the simple quadratic approximation (Eq. S9); (d) Comparison of the $E_{\text{geo}}$, $E_{\text{harm}}$, and $E_{\text{simple}}$ obtained from different methods.

Figure S15: The ellipses satisfying Eqs. S10 and S11 for the water molecule.

Both Eq. S8 and S9 give us a simple yet very accurate representation of the GEO for the water molecule. From Eq. S8 and values of the force constants (Eq. S7), we can see that an error of 2pm in the bond length results in the GEO of $\sim 0.5$ kcal/mol. On the other hand, because of the smaller size of $k_{\theta\theta}$, we can see that the GEO for the water molecule is much more forgiving of the bond angle error, since an error of $2^\circ$, results in the GEO of only $\sim 0.06$ kcal/mol. Interestingly, Eq. S9 can be rearranged to be in a form of an ellipse equation:

$$\frac{k_{aa}}{E_{\text{harm}}^{\text{geo}}} \Delta a^2 + \frac{1}{2E_{\text{harm}}^{\text{geo}}} k_{\theta\theta} \Delta \theta^2 + 2 \frac{k_{a\theta}}{E_{\text{harm}}^{\text{geo}}} \Delta a \Delta \theta = 1.$$  \hspace{1cm} (S10)

Similarly, rearranging Eq. S9, we obtain an axis-aligned ellipse equation:

$$\frac{k_{aa}}{E_{\text{simple}}^{\text{geo}}} \Delta a^2 + \frac{k_{\theta\theta}}{2E_{\text{simple}}^{\text{geo}}} \Delta \theta^2 = 1.$$  \hspace{1cm} (S11)

Since $k_{a\theta}$ is small, the two ellipses of Eq. S10 and S11 are in a close agreement, as it can be seen from Figure S15.
In this section we show detailed GEO analysis for a set of small molecules. For all molecules, CCSD(T) has been used as a reference and all calculation have been obtained with the G16 package. The following basis set have been used: A’V6Z for the water molecule (Figure S17), aug-cc-pVQZ for H$_2$S (Figure S18) and ethene (Figure S20), and aug-cc-pVTZ for H$_2$CO (Figure S19), methanol (Figure S21), cyclopropane (Figure S21).
Figure S17: The GEO analysis for the water molecule. Panel (a): considered degrees of freedom; (b): Comparison of GEO with its harm and simple (labeled as ‘s’ in the plot for brevity) approximations obtained from various methods; (c): errors in specific i-th degree of freedom specified in (a), with insets showing the resulting $E_{\text{geo}}^{\text{simple},i}$ (see Eq. 5); (d) different $E_{\text{geo}}^{\text{simple},i}$ contributions to the total $E_{\text{geo}}^{\text{simple}}$; (e) the resulting weights, i.e. $E_{\text{geo}}^{\text{simple},i}/E_{\text{geo}}^{\text{simple}}$. 
Figure S18: The same plots as in Figure S17, but for the H$_2$S molecule.
Figure S19: The same plots as in Figure S17, but for the H$_2$CO molecule.
Figure S20: The same plots as in Figure S17 but for the ethene molecule.
\[
\begin{align*}
\theta_1 &= \angle 3, 1, 2 \\
\theta_2 &= \angle 4, 1, 2 \\
\theta_3 &= \angle 6, 1, 2 = 180^\circ
\end{align*}
\]

\[
\phi = \angle d 4, 1, 2, 3
\]

Figure S21: The same plots as in Figure S1 for the methanol molecule.
The same plots as in Figure S17, but for the cyclopropane molecule.
Figure S24: Summary of Figs. S17-S22

\[ \theta = 4, 1, 2 \]
\[ \phi = 4, 1, 2, 3 \]

\[ \theta = 4, 1, 2 \]
\[ \phi = 4, 1, 2, 3 \]
<table>
<thead>
<tr>
<th></th>
<th>HF</th>
<th>PBE</th>
<th>BLYP</th>
<th>TPSS</th>
<th>B3LYP</th>
<th>MN15</th>
<th>B97</th>
<th>B97XD</th>
<th>LSDA</th>
<th>PBE [D3]</th>
<th>MP2</th>
<th>B2PLYP</th>
<th>PBE0</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>0.450</td>
<td>0.129</td>
<td>0.173</td>
<td>0.094</td>
<td>0.014</td>
<td>0.038</td>
<td>0.009</td>
<td>0.013</td>
<td>0.138</td>
<td>0.129</td>
<td>0.000</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>H₂S</td>
<td>0.139</td>
<td>0.103</td>
<td>0.098</td>
<td>0.021</td>
<td>0.006</td>
<td>0.066</td>
<td>0.016</td>
<td>0.002</td>
<td>0.099</td>
<td>0.104</td>
<td>0.019</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>H₂CO</td>
<td>1.208</td>
<td>0.141</td>
<td>0.092</td>
<td>0.039</td>
<td>0.118</td>
<td>0.117</td>
<td>0.105</td>
<td>0.218</td>
<td>0.312</td>
<td>0.147</td>
<td>0.007</td>
<td>0.033</td>
<td>0.209</td>
</tr>
<tr>
<td>ethene</td>
<td>0.436</td>
<td>0.118</td>
<td>0.064</td>
<td>0.037</td>
<td>0.079</td>
<td>0.085</td>
<td>0.091</td>
<td>0.123</td>
<td>0.251</td>
<td>0.120</td>
<td>0.029</td>
<td>0.052</td>
<td>0.100</td>
</tr>
<tr>
<td>methanol</td>
<td>0.701</td>
<td>0.136</td>
<td>0.218</td>
<td>0.069</td>
<td>0.020</td>
<td>0.117</td>
<td>0.027</td>
<td>0.080</td>
<td>0.324</td>
<td>0.148</td>
<td>0.017</td>
<td>0.020</td>
<td>0.084</td>
</tr>
<tr>
<td>cyclopropane</td>
<td>0.408</td>
<td>0.137</td>
<td>0.118</td>
<td>0.032</td>
<td>0.059</td>
<td>0.172</td>
<td>0.118</td>
<td>0.116</td>
<td>0.522</td>
<td>0.143</td>
<td>0.057</td>
<td>0.061</td>
<td>0.159</td>
</tr>
<tr>
<td>mean</td>
<td>0.557</td>
<td>0.127</td>
<td>0.127</td>
<td>0.049</td>
<td>0.049</td>
<td>0.099</td>
<td>0.061</td>
<td>0.092</td>
<td>0.274</td>
<td>0.132</td>
<td>0.022</td>
<td>0.029</td>
<td>0.093</td>
</tr>
<tr>
<td>RMS</td>
<td>0.649</td>
<td>0.128</td>
<td>0.137</td>
<td>0.055</td>
<td>0.064</td>
<td>0.108</td>
<td>0.076</td>
<td>0.118</td>
<td>0.307</td>
<td>0.133</td>
<td>0.028</td>
<td>0.036</td>
<td>0.120</td>
</tr>
</tbody>
</table>

Table S4: The $E_{geo}$ values for molecules in Figs. S17-S22 in kcal/mol. CCSD(T) has been used as a reference in all calculations. The following basis set have been used: A’V6Z for the water molecule aug-cc-pVQZ for H₂S and ethene, and aug-cc-pVTZ for H₂CO, methanol, cyclopropane.
S7. GEO analysis in normal modes (Hessian eigenvectors)
Table S5: Normal-mode GEO analysis for the H$_2$O molecule. The "1%" label denotes the GEO errors (or weights) when γ = 1% (see the main text).
The water has 3 normal modes (note again that by normal modes we refer to the eigenvectors of $H_0$), and two of these are GEO-active. In the left panel of Fig. S26(1) we show the $E_{\text{geo}}$ weights in internal coordinates (see Eq. 5), i.e. the $E_{\text{geo}}$ contributions from the errors in $\theta$ (bond angle) and $a$ (OH bond length). In the right panel of the same figure, we show the $E_{\text{geo}}$ weights but given in terms of the errors in normal coordinates (modes), $p_1$ and $p_2$, which correspond to the eigenvectors of the Hessian (see Eq. 4). The third normal mode that leads to an asymmetric OH stretch is GEO-inactive (see Fig. S5). By the similarity of the patterns in the two panels of Fig. S26(1), we can expect that the changes in $p_1$ mostly affect $a$ and that the changes in $p_2$ mostly affect $\theta$ (see also the middle panel of Fig. S26(2), where we plot the normal modes.) But, to show explicitly how the errors in $p_1$ and $p_2$ translate into the errors in $a$ and $\theta$, we consider again the $\gamma$-scaling, and we scale the individual normal coordinates as: $p'_i = (1 + \gamma) p_i$. From the left panel of Fig. S26(2), we can see that upon scaling of $p_1$ by $1 + \gamma$, $a$ is scaled by the (almost) factor of $\gamma$. At the same time, scaling $p_2$ by $1 + \gamma$ affects very little $a$. Note that when all coordinates are scaled by $\gamma$: $G = (1 + \gamma) G_0$ all angles are unchanged and all bond lengths scaled by $\gamma$. (S12)

Thus, when $p_1$ and $p_2$ are scaled by $1 + \gamma$, the angle changes at the same speed in both cases, but in opposite directions, since the two changes in the angles need to cancel each other out [Eq S12]. This is shown in the right panel of Fig. S26(2). Overall, mode 1 contains contributions from both $R$ and $a$, whereas, mode 2 contains mostly contribution from $a$. That is why the weights in the two panels of Fig. S26(1) are similar, but not exactly the same.
**Figure S26:** Normal-mode vs. internal coordinates GEO analysis for the $H_2O$ molecule.

1. $\omega_{geo}$ simple weights for the water molecule in internal coordinates (left), GEO-active normal modes (right).

2. The change in the bond lengths (left) and bond angle (right) upon scaling GEO active normal modes: $p_i, \gamma = (1 + \gamma)p_i$ where $p_i$ are accurate [CCSD(T)] normal modes. The GEO active normal modes are shown in the middle panel.
Figure S27: Normal-mode GEO analysis for the H$_2$S molecule.
Figure S28: Normal-mode GEO analysis for the H$_2$CO molecule.
Figure S29: Normal-mode GEO analysis for the ethene molecule.
S8. Details on the $\gamma$ expansion and $D$ values for molecular sets

The $D$ molecule-specific constant (absolute GEO scale) is given by:

$$D = G^\top_0 H_0 G_0,$$

(S13)

and thereby can be easily calculated once the force constants (in e.g. Cartesian coordinates) are available. In analogy to approximating $E_{geo}$ by $E'_{geo}$, we can calculate $D$ more cheaply from approximate minima and force constants:

$$D' = \tilde{G}^\top \tilde{H} \tilde{G}.$$  

(S14)

Furthermore, in analogy to $E_{geo}^{\text{simple}}$ approximation to $E_{geo}$, we can also define $D^{\text{simple}}$ as:

$$D \approx D^{\text{simple}} = \sum_i K_{\text{simple},i} = \sum_i R_{i}^2 f_{ii},$$  

(S15)

where $R_i$-s are the bond lengths and $f_{ii}$ are the force constants for each of the bonds. Note that there is no angle contribution to the r.h.s. of Eq (S15) since by the $\gamma$-scaling (see the main text), angles remain unchanged.

In Table S6, we show the accurate $D$ and approximate $D'$ values. From this table, we can see that even HF $D'$ are decently accurate and already B3LYP ones are in an excellent agreement with $D$. For the same set of molecules, $E_{geo}^{\text{simple}}$ and contributions from individual bonds, we show in Fig. S30. For other organic molecules in the paper, we give the $D$ values in Tables S7 and S8. Finally, in Table S9, we show the $D$ values for rare gas dimers, and we can see that these $D$ values are by an order of magnitude smaller than those for covalently bonded molecules.

<table>
<thead>
<tr>
<th></th>
<th>CCSD(T)</th>
<th>MP2</th>
<th>B2PLYP</th>
<th>B3LYP</th>
<th>PBE0</th>
<th>TPSS</th>
<th>PBE</th>
<th>HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>h2o</td>
<td>0.221</td>
<td>0.222</td>
<td>0.220</td>
<td>0.219</td>
<td>0.224</td>
<td>0.211</td>
<td>0.211</td>
<td>0.247</td>
</tr>
<tr>
<td>h2s</td>
<td>0.220</td>
<td>0.228</td>
<td>0.221</td>
<td>0.215</td>
<td>0.221</td>
<td>0.213</td>
<td>0.209</td>
<td>0.239</td>
</tr>
<tr>
<td>methanol</td>
<td>0.566</td>
<td>0.572</td>
<td>0.567</td>
<td>0.563</td>
<td>0.579</td>
<td>0.547</td>
<td>0.545</td>
<td>0.629</td>
</tr>
<tr>
<td>cyclopropane</td>
<td>0.989</td>
<td>1.007</td>
<td>0.998</td>
<td>0.99</td>
<td>1.012</td>
<td>0.972</td>
<td>0.966</td>
<td>1.076</td>
</tr>
<tr>
<td>ch3cl</td>
<td>0.452</td>
<td>0.465</td>
<td>0.452</td>
<td>0.443</td>
<td>0.459</td>
<td>0.436</td>
<td>0.435</td>
<td>0.488</td>
</tr>
<tr>
<td>ch3f</td>
<td>0.463</td>
<td>0.469</td>
<td>0.463</td>
<td>0.458</td>
<td>0.47</td>
<td>0.447</td>
<td>0.443</td>
<td>0.511</td>
</tr>
<tr>
<td>c2h3cl</td>
<td>0.727</td>
<td>0.742</td>
<td>0.736</td>
<td>0.733</td>
<td>0.754</td>
<td>0.717</td>
<td>0.714</td>
<td>0.81</td>
</tr>
<tr>
<td>c2h4</td>
<td>0.634</td>
<td>0.643</td>
<td>0.643</td>
<td>0.642</td>
<td>0.651</td>
<td>0.629</td>
<td>0.623</td>
<td>0.699</td>
</tr>
<tr>
<td>c-c2h2f2</td>
<td>0.848</td>
<td>0.858</td>
<td>0.856</td>
<td>0.856</td>
<td>0.881</td>
<td>0.828</td>
<td>0.823</td>
<td>0.971</td>
</tr>
<tr>
<td>c2h3f</td>
<td>0.735</td>
<td>0.745</td>
<td>0.744</td>
<td>0.744</td>
<td>0.761</td>
<td>0.724</td>
<td>0.719</td>
<td>0.828</td>
</tr>
<tr>
<td>imine</td>
<td>0.555</td>
<td>0.557</td>
<td>0.564</td>
<td>0.569</td>
<td>0.58</td>
<td>0.551</td>
<td>0.547</td>
<td>0.636</td>
</tr>
<tr>
<td>h2co</td>
<td>0.485</td>
<td>0.485</td>
<td>0.489</td>
<td>0.497</td>
<td>0.508</td>
<td>0.48</td>
<td>0.476</td>
<td>0.563</td>
</tr>
<tr>
<td>formic</td>
<td>0.725</td>
<td>0.724</td>
<td>0.723</td>
<td>0.731</td>
<td>0.753</td>
<td>0.698</td>
<td>0.697</td>
<td>0.842</td>
</tr>
<tr>
<td>hcocl</td>
<td>0.594</td>
<td>0.595</td>
<td>0.592</td>
<td>0.599</td>
<td>0.623</td>
<td>0.573</td>
<td>0.573</td>
<td>0.696</td>
</tr>
<tr>
<td>MAE</td>
<td>0.007</td>
<td>0.005</td>
<td>0.007</td>
<td>0.019</td>
<td>0.013</td>
<td>0.017</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>RMS</td>
<td>0.009</td>
<td>0.006</td>
<td>0.008</td>
<td>0.021</td>
<td>0.015</td>
<td>0.018</td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>MARE (%)</td>
<td>1.3</td>
<td>0.7</td>
<td>1.3</td>
<td>3.2</td>
<td>2.5</td>
<td>3.1</td>
<td>12.2</td>
<td></td>
</tr>
</tbody>
</table>

Table S6: Accurate $D$ values [CCSD(T)] in $10^4$ kcal/mol, and approximate $D'$ values computed from Eqs. S13 and S14 respectively, for the set of molecules considered in Figs. 1b.
Figure S30: $D$ values in $10^4$ kcal/mol (a value in black in the top-right corner of each panel), $D_{\text{simple}}$ values (a value in blue the top-right corner of each panel), and $D_{\text{simple}}$ bond-decomposition of Eq. S15 for the set of molecules considered in Fig. 1b.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$D$ (10$^4$ kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetaldehyde</td>
<td>0.842</td>
</tr>
<tr>
<td>acetic</td>
<td>1.086</td>
</tr>
<tr>
<td>c2h3f</td>
<td>0.735</td>
</tr>
<tr>
<td>c2h5f</td>
<td>0.819</td>
</tr>
<tr>
<td>ch2ch</td>
<td>0.749</td>
</tr>
<tr>
<td>ch2nh2</td>
<td>0.598</td>
</tr>
<tr>
<td>ch3nh</td>
<td>0.543</td>
</tr>
<tr>
<td>ch3nh2</td>
<td>0.660</td>
</tr>
<tr>
<td>ethanol</td>
<td>0.917</td>
</tr>
<tr>
<td>formic</td>
<td>0.73</td>
</tr>
<tr>
<td>methanol</td>
<td>0.568</td>
</tr>
<tr>
<td>propane</td>
<td>1.066</td>
</tr>
<tr>
<td>propene</td>
<td>0.982</td>
</tr>
<tr>
<td>propyne</td>
<td>0.877</td>
</tr>
</tbody>
</table>

Table S7: $D$ values in $10^4$ kcal/mol for individual molecules for molecules considered in Fig. 1a(top) at the CCSD(T) level of theory. For sufficiently small $\gamma$, $D$ can be computed by finite difference from: $D = \frac{2E_{\gamma}}{\gamma}$, and we have used here $\gamma = 0.5\%$. We have found that this $\gamma$ value is small enough to stay within the harmonic approximation, but still not too small to lose numerical accuracy.
<table>
<thead>
<tr>
<th>mol. #</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.544</td>
</tr>
<tr>
<td>2</td>
<td>1.695</td>
</tr>
<tr>
<td>3</td>
<td>2.229</td>
</tr>
<tr>
<td>4</td>
<td>2.386</td>
</tr>
<tr>
<td>5</td>
<td>2.083</td>
</tr>
<tr>
<td>6</td>
<td>2.483</td>
</tr>
<tr>
<td>7</td>
<td>2.505</td>
</tr>
<tr>
<td>8</td>
<td>1.693</td>
</tr>
<tr>
<td>9</td>
<td>2.387</td>
</tr>
<tr>
<td>10</td>
<td>2.009</td>
</tr>
<tr>
<td>11</td>
<td>2.457</td>
</tr>
<tr>
<td>12</td>
<td>2.339</td>
</tr>
<tr>
<td>13</td>
<td>2.738</td>
</tr>
<tr>
<td>14</td>
<td>3.208</td>
</tr>
<tr>
<td>15</td>
<td>3.715</td>
</tr>
<tr>
<td>16</td>
<td>3.404</td>
</tr>
</tbody>
</table>

**Table S8:** $D$ values in $10^4$ kcal/mol obtained as described in the caption Table S7, but with B2PLYP/aug-cc-pVTZ as a reference. The order of molecules is shown in Fig. S3.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>He$_2$</td>
<td>0.000154</td>
</tr>
<tr>
<td>Ne$_2$</td>
<td>0.000724</td>
</tr>
<tr>
<td>Ar$_2$</td>
<td>0.002164</td>
</tr>
<tr>
<td>HeNe</td>
<td>0.000301</td>
</tr>
<tr>
<td>NeAr</td>
<td>0.001079</td>
</tr>
</tbody>
</table>

**Table S9:** $D$ values in $10^4$ kcal/mol computed from the CCSD(T) force constants for the noble gas dimers (Eq. S13). For the basis set informations, see Fig. S38.
S9. Statistical measures for the quality of approximate geometries of the methylamine molecule

Figure S31: Mean absolute errors (MAE) in (i) calculated bond lengths; (ii) bond angles; (iii) torsion angles (iv) GEO errors obtained from various approximate methods for the methylamine molecule.

In addition to the measures in Figure S31, in Figure S32, we show other approximate measures for the geometry of the methylamine molecule. For comparison, we also show in Figure S33 the errors in atomization energies obtained with the same methods.
Figure S32: Other statistical measures for the quality of the approximate geometries of the methylamine molecule (see Figure 3 of the main text): (i) means signed error (MSE) for the bond lengths; (ii): MSE for the bond angles; (iii) MSE for the torsion angles (iv) mean absolute error (MAE) for the distances between the C atom and the other atoms in the molecule (v) MAE for the distances between the N atom and the other atoms in the molecule (vi) MAE for the distances between the CN midbond point and the atoms in the molecule.

Figure S33: Absolute errors for the atomization energies of the methylamine molecule obtained from different methods. The LSDA and HF errors are divided by the factor of 4.5.
S10. GEO ANALYSIS FOR NOBLE GAS DIMERS
Figure S35: Top panel: Geo analysis for the neon dimer. For the basis set information, see Figure S38. Accurate binding energy of the dimer is shown in the square brackets in the subcaption. The aug-cc-pV5Z basis set has been used in all calculations. In the left top panel, the $\Delta R$ position of approximations are not accurate when $\Delta R > 1\,\text{Å}$. In these cases, the approximations typically give unphysical binding curves with no minimum and only infimum at $R \to \infty$. Bottom panel: Various dissociation curves for the neon dimer.

Figure S36: Counterpoise (CP) corrected (dashed lines) and counterpoise uncorrected binding curves for the neon dimer (solid lines) obtained from various approximations. Detailed analysis of Baerends and co-workers have shown that for small dimers bonded by dispersion, CP-uncorrected curves with a finite basis set are in a better agreement with the "exact" (i.e. no errors due to a finite basis set) curve curve than its CP-corrected counterpart (see Refs. [11, 12]). Thus, in the main text and in Figure S35, we use CP-uncorrected curves. Nonetheless, due to a large basis set used here [aug-cc-pV5Z], we can see from this figure that the energy difference between curves with and without the CP is small, even in the CCSD(T) case and nearly negligible in the case of DFT curves.
approximations are not accurate when \( \Delta R > 1 \text{Å} \). In these cases, the approximations typically give unphysical binding curves with no minimum and only infimum at \( R \to \infty \).

Figure S38: The GEO analysis for noble gas dimers. The aug-cc-pV5Z basis has been used for Ne2 and He2, and aug-cc-pVQZ has been used for all other dimers. CCSD(T) has been used as a reference for all dimers. In the left panels, the \( \Delta R \) position of approximations are not accurate when \( \Delta R > 1 \text{Å} \). In these cases, the approximations typically give unphysical binding curves with no minimum and only infimum at \( R \to \infty \).
S11. Details on the results of Figure 5 [S66 results]

The S66x8 set contains 66 dimers bonded by noncovalent interactions.[9] For each complex and each approximate method we do single point calculations to obtain 8 datapoints for different separations between fragments along the dissociation curve of a given complex. The geometries at different separations between fragments in a complex have been taken from the S66x8 dataset.[9] The CCSD(T)/CBS (used as a reference here), MP2/CBS, CCSD/CBS and SCS-CCSD/CBS binding energies have been taken from the original S66x8 dataset.[9, 10] All other binding energy we obtain from the G16, by performing counterpoise corrected calculations within the aug-cc-pVTZ basis set. From these 8 datapoints (for each complex and for each approximate method) we use a spline interpolation to construct a dissociation curve. The $E_{geo}$ error for a given approximate method is then calculated by the difference between the CCSD(T) binding energies at the approximate minimum and at the CCSD(T) minimum (see Eq. 1). In principle, in addition to finding a minimum intermolecular separation between fragments, we should have also optimized the fragments with a given approximate method. However, we assume here that this has a minor effect on the $E_{geo}$ contribution to the binding energies errors of the S66 complexes.

The same plots shown in Figure 5 in the main text are shown here with other approximate methods. Namely, in Figure S39 we show the $|E_{geo}|$ vs. $|\Delta E|$ plots (as in Figure 5 of the main text) for selected DFT functionals with the D3 empirical correction. The same plots are shown in Figure S40 for CCSD/CBS and SCS-CCSD/CBS. The same plots obtained from the PBE functional with and without the D3 correction is shown in Figure S41.
**Figure S39:** $E_{\text{geo}}$ (y-axis) vs. $|\Delta E|$ (x-axis) errors in kcal/mol for different methods and for the S66 dataset.

**Figure S40:** $E_{\text{geo}}$ (y-axis) vs. $|\Delta E|$ (x-axis) errors in kcal/mol for different methods and for the S66 dataset.
Table S10: MAE of different methods (rows) at different minima of S66x8 binding curves (columns) for the S66 dataset. The error is defined as: $E_X[R_Y] - E_0$, where $X$ is a method in rows and $Y$ is the method in columns, and $E_0$ is the binding energy at the CCSD(T) minimum. The S66x8 binding curves from CCSD(T)/CBS (used as a reference here) and MP2/CBS have been taken from the original S66x8 dataset. All other binding curves have been obtained from counterpoise corrected calculations within the aug-cc-pVTZ basis set. The minimum of each binding curve has been found numerically after the interpolation of 8 datapoints for each of the S66 complex (see details above).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MP2/CBS</td>
<td>0.51</td>
<td>0.26</td>
<td>0.45</td>
<td>0.43</td>
<td>0.38</td>
<td>0.47</td>
<td>0.46</td>
<td>0.28</td>
<td>0.47</td>
</tr>
<tr>
<td>B2PLYP</td>
<td>1.63</td>
<td>1.34</td>
<td>1.45</td>
<td>1.43</td>
<td>1.38</td>
<td>1.52</td>
<td>1.51</td>
<td>1.40</td>
<td>1.46</td>
</tr>
<tr>
<td>B3LYP[D3]</td>
<td>0.44</td>
<td>0.34</td>
<td>0.43</td>
<td>0.42</td>
<td>0.41</td>
<td>0.41</td>
<td>0.42</td>
<td>0.53</td>
<td>0.43</td>
</tr>
<tr>
<td>PBE0[D3]</td>
<td>0.51</td>
<td>0.43</td>
<td>0.48</td>
<td>0.49</td>
<td>0.48</td>
<td>0.45</td>
<td>0.47</td>
<td>0.54</td>
<td>0.48</td>
</tr>
<tr>
<td>PBE[D3]</td>
<td>0.48</td>
<td>0.4</td>
<td>0.44</td>
<td>0.45</td>
<td>0.46</td>
<td>0.40</td>
<td>0.41</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>MN15</td>
<td>0.56</td>
<td>0.34</td>
<td>0.59</td>
<td>0.58</td>
<td>0.53</td>
<td>0.6</td>
<td>0.60</td>
<td>0.34</td>
<td>0.59</td>
</tr>
<tr>
<td>wB97XD</td>
<td>0.42</td>
<td>0.25</td>
<td>0.47</td>
<td>0.45</td>
<td>0.41</td>
<td>0.47</td>
<td>0.48</td>
<td>0.38</td>
<td>0.46</td>
</tr>
</tbody>
</table>
| PBE      | 2.29    | 1.77   | 2.04      | 1.98     | 1.90    | 1.90 | 2.14   | 2.12  | 1.71    | 2.06

Figure S41: $E_{geo}$ (y-axis) vs. $|\Delta E|$ (x-axis) errors in kcal/mol for different methods and for the S66 dataset.
Figure S42: Benzene-uracil binding curves (interpolated from the 8 datapoints along the dissociation curve).
References


[10] We thank J. Rezác for giving us access to the S66x8 binding energies for these wavefunction methods.
