Data Review

Yi Xie

May 22, 2022

Table of contents

Intermolecular Energies

- Supermolecular approach

$$E_{int} = E_{AB} - E_A - E_B$$

  - Straightforward, but cannot separate different types of interactions
  - Can adopt to different electronic structure methods
  - DFT-D3 with proper functional can be both cheap and accurate
- Symmetry-Adapted Perturbation Theory
  - Can give details about interactions (electrostatic, induction/polarization, exchange repulsion, London dispersion components); important in understanding their nature
  - Not as cheap as DFT-D3
  - SAPT0 is somewhat cheap, but does not include intramonomer correlation

SAPT(DFT)

- Attempt to inlude intramonomer correlation in a cheap way
- Replaces HF orbitals with KS orbitals
- Needs to consider orbital response for dispersion terms
- Exchange-dispersion term needs to be estimated from scaling
- Investigate the accuracy and efficiency of SAPT(DFT)

Three-Body Interaction

- Crucial in computing high accuracy lattice energies (5% for benzene)
- DFT-D3 does not perform well for three-body interaction
- MP2.5 scales as $O(N^6)$, MP2 is $O(N^5)$ but lacks three-body dispersion
- Three-body dispersion can be implemented with SAPT(DFT) in $O(N^5)$
- Combine MP2 with SAPT(DFT) dispersion to model three-body interaction

Table of contents

Table of contents

Idea of SAPT(DFT)

- SAPT energy in orders of interaction and fluctuation potentials; n denotes order in V and k, l for W_A, W_B

$$H = F_A + F_B + V + W_A + W_B$$

$$E_{int} = \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left(E_{pol}^{(nkl)} + E_{exch}^{(nkl)} \right)$$

- SAPT0: $n = 2, k = l = 0$, no intramonomer correlation, $O(N^5)$ cost
- Many-body SAPT: $k + l \geq 2$, $O(N^6)$ or higher cost
- SAPT(DFT): Use Kohn-Sham operator $K_{A,B}$ instead of Fock operator $F_{A,B}$, $O(N^5)$ cost
- Primitive SAPT(DFT) works well on 1st-order terms, but not 2nd-order terms (especially dispersion). Needs orbital response for these terms

Coupled Dispersion Energy

- Uncoupled dispersion energy in terms of frequency-dependent density susceptibility (FDDS):

$$
\begin{aligned}
E^{(2)}_{disp,u} &= -4 \sum_{ia\in A, jb\in B} \frac{|(ia|jb)|^2}{\epsilon^{ab}_{ij}} \\
&= -\frac{1}{2\pi}\int_0^\infty d\omega \int d\mathbf{r}_A d\mathbf{r}'_A d\mathbf{r}_B d\mathbf{r}'_B \\
&\quad \frac{1}{|\mathbf{r}_A - \mathbf{r}_B|}\frac{1}{|\mathbf{r}'_A - \mathbf{r}'_B|}\chi^A_0\left(\mathbf{r}_A, \mathbf{r}'_A|i\omega\right)\chi^B_0\left(\mathbf{r}_B, \mathbf{r}'_B|i\omega\right)
\end{aligned}
$$

- Kohn-Sham DFT constructs a fictitious system of non-interacting particles, which reproduces the density and energy of the real electronic system
- Kohn-Sham FDDS does not reflect the correct response properties of the electronic system

Dispersion Term

- Coupled FDDS from solving TDDFT equations[1]:

$$\chi = \chi_0 + \chi_0 \mathbf{S}^{-1}\mathbf{W}\left(\mathbf{S} - \chi_0 \mathbf{S}^{-1}\mathbf{W}\right)^{-1}\chi_0$$

- $\mathbf{S}$ and $\mathbf{W}$ corresponds to Coulomb metric and xc kernel
- Pure ALDA kernel good for pure GGA functional, but not for hybrid functional
- Exact exchange in v_{xc} $\rightarrow$ increased ϵ_{ij}^{ab} $\rightarrow$ decreased $E_{disp}^{(2)}$

$$E_{disp,u}^{(2)} = -4 \sum_{ia \in A, jb \in B} \frac{\left|(ia|jb)\right|^2}{\epsilon_{ij}^{ab}}$$

- Hybrid ALDA kernel to compensate, or localized HF (LHF) exchange to avoid increase in ϵ_{ij}^{ab}
- Reformulating solution of hybrid TDDFT equations[2] in density-fitting basis on next slide

[1] M. Pitoňák and A. Hesselmann, J. Chem. Theory Comput. **6**, 168 (2010).

Coupled FDDS with hybrid kernel

$$\chi = \chi_0' + \left(\chi_0'\mathbf{S}^{-1}\mathbf{W} + \mathbf{K}'\right)\left[\mathbf{S} - \left(\chi_0'\mathbf{S}^{-1}\mathbf{W} + \mathbf{K}'\right)\right]^{-1}\chi_0'$$

$$\mathbf{K}' = \left[-\xi\mathbf{K}_1\left(\lambda d\right) - \xi\mathbf{K}_2\left(\lambda d\right) + \xi^2\mathbf{K}_{21}\left(\lambda\right)\right]\left(\mathbf{R}^T\right)^{-1}\mathbf{S}$$

$$[\mathbf{K}_1(\lambda d)]_{PQ} = (P|ar)\lambda_{ar}d_{ar}[(aa'|rr') + (ar'|a'r)](a'r'|\mathbf{Q}|Q)$$

$$[\mathbf{K}_2(\lambda d)]_{PQ} = (P|ar)\lambda_{ar}d_{ar}[(aa'|rr') - (ar'|a'r)](a'r'|\mathbf{Q}|Q)$$

$$\begin{aligned}[\mathbf{K}_{21}(\lambda)]_{PQ} \quad &= \quad (P|ar)\lambda_{ar}[(aa''|rr'') - (ar''|a''r)] \\ &\quad\quad [(a'a''|r'r'') - (a'r''|a''r')](a'r'|\mathbf{Q}|Q)\end{aligned}$$

$$[\mathbf{K}_2'(\lambda)]_{PQ} = (P|ar)\lambda_{ar}[(aa'|rr') - (ar'|a'r)](a'r'|Q)$$

$$\chi_0' = \chi_0 - \xi\mathbf{K}_2\left(\lambda\right)$$

$$(ar|Q) = (ar|\mathbf{Q}|P)(P|\mathbf{R}|Q)$$

Equation for $E^{(2)}_{disp}$

- Coupled Kohn-Sham (CKS) FDDS reflects correct response properties of electrons
- $O(N^5)$ scaling is limited to forming $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{K}_{21}$ and $\mathbf{K}'_2$. These quantities are frequency-dependent, but could store frequency-independent intermediates from the $O(N^5)$ contractions on disk, and the frequency-dependent contractions are only $O(N^4)$
- $E^{(2)}_{disp}$ from coupled FDDS: (Integration is approximated by Gauss-Legendre quadrature)

$$\begin{aligned} E^{(2)}_{disp,r} \quad &= \quad -\frac{1}{2\pi}\int_0^\infty d\omega \int d\mathbf{r}_A d\mathbf{r}'_A d\mathbf{r}_B d\mathbf{r}'_B \\ &\quad \frac{1}{|\mathbf{r}_A - \mathbf{r}_B|}\frac{1}{|\mathbf{r}'_A - \mathbf{r}'_B|}\chi^A\left(\mathbf{r}_A, \mathbf{r}'_A|i\omega\right)\chi^B\left(\mathbf{r}_B, \mathbf{r}'_B|i\omega\right) \end{aligned}$$

Exchange-Dispersion Term

- Explicit coupled exchange-dispersion not trivial to implement; currently working on this
- Estimate from scaling uncoupled exchange-dispersion; Scale with ratio in dispersion term or with pre-fitted (with S22×5) fixed factor

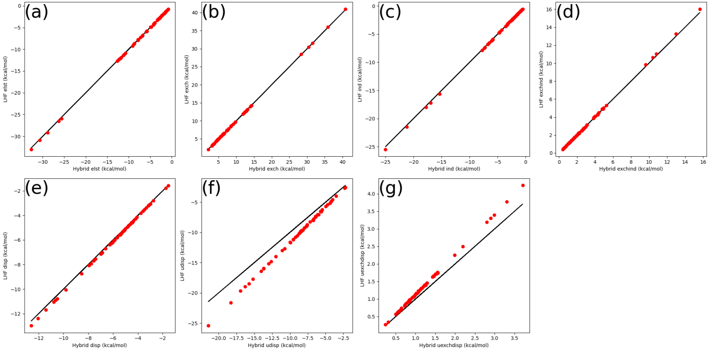
$$\tilde{E}^{(2)}_{exch-disp,r} = E^{(2)}_{exch-disp,u} \cdot \frac{E^{(2)}_{disp,r}}{E^{(2)}_{disp,u}}$$

$$\tilde{E}^{(2)}_{exch-disp,r} = \alpha \cdot E^{(2)}_{exch-disp,u}(\alpha = 0.686)$$

- The value above is fitted from $E^{(2)}_{exch-disp,u}$ with LHF orbitals
- Non-LHF orbitals have greater o-v gaps and smaller $E^{(2)}_{disp,u}$, needs to re-fit with non-LHF results

Georgia Tech

LHF vs non-LHF orbitals



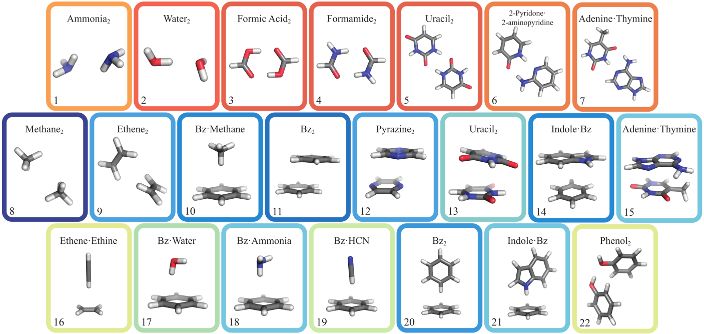
Hybrid vs. LHF values in kcal/mol for each term for S66 data set: (a) $E^{(1)}_{elst}$, (b) $E^{(1)}_{exch}$, (c) $E^{(2)}_{ind}$, (d) $E^{(2)}_{exch-ind}$, (e) $E^{(2)}_{disp,r}$, (f) $E^{(2)}_{disp,u}$, (g) $E^{(2)}_{exch-disp,u}$

Table of contents

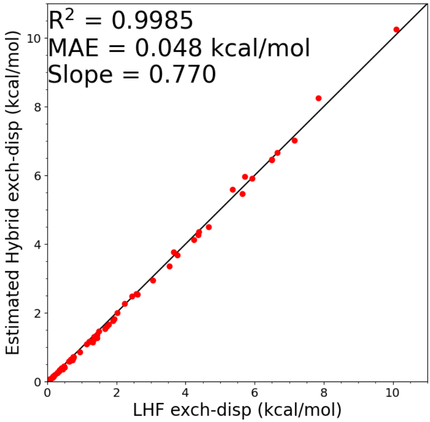
Exchange-Dispersion Refitting

- Need to fit the uncoupled $E^{(2)}_{exch-disp,r}$ with non-LHF orbitals on the coupled LHF orbital values (implemented in Molpro)
- Assuming coupled LHF and non-LHF orbital $E^{(2)}_{exch-disp,r}$ from the behavior of $E^{(2)}_{disp,r}$
- Exchange-related components depend heavily on distance between monomers, sets like S22$\times$5 and S66$\times$8 would be preferred (S22/S66 with various non-equilibrium intermolecular distances)
- Determine the scaling factor with S22$\times$5, validate with S66$\times$8

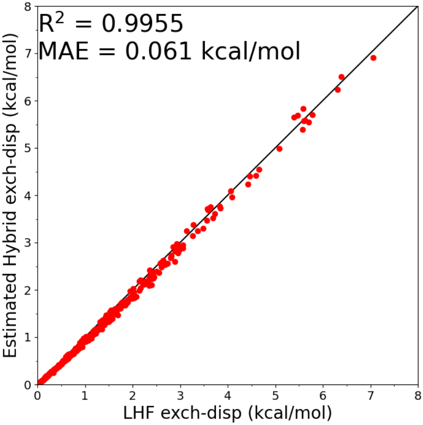
S22 dimer set



S22×5 Fitting Results



S66×8 Validating Results



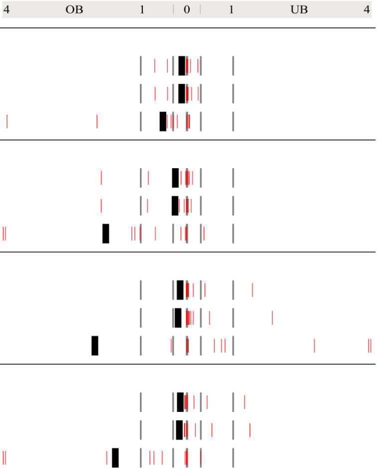
Termwise results

- Compared the SAPT(DFT)/aug-cc-pVTZ results of our code to SAPT(CCSD)/aug-cc-pVTZ results from Korona S2[3]
- Also comparing the results for S66 with SAPT2+3(CCD)δMP2/aug-cc-pVTZ as reference. Also added SAPT0/aug-cc-pVDZ, SAPT2+/aug-cc-pVDZ and SAPT2+(3)δMP2/aug-cc-pVTZ into comparison as side-reference.
- Errors of each system with respect to reference shown as vertical lines
- Mean absolute error (MAE) and mean unsigned relative error (MURE) listed for S2. MAE indicated by black box in the diagram
- Color scheme for S66 systems: Hydrogen-bonded (HB, red), mixed-influence (MX, green), dispersion-dominated (DD, blue)

[3] T. Korona, Mol. Phys. **111**, 3705 (2013).

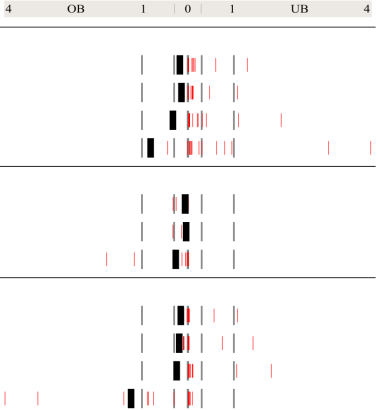
Korona S2 Results

Method	MAE	MURE	Error Distribution
Electrostatics			
SAPT(DFT) hybrid	0.112	2.39	
SAPT(DFT) LHF	0.114	3.68	
SAPT0	0.520	8.61	
Exchange			
SAPT(DFT) hybrid	0.251	3.38	
SAPT(DFT) LHF	0.258	3.09	
SAPT0	1.757	12.88	
Induction			
SAPT(DFT) hybrid	0.148	2.79	
SAPT(DFT) LHF	0.192	2.97	
SAPT0	1.993	16.83	
Exchange-Induction			
SAPT(DFT) hybrid	0.144	4.03	
SAPT(DFT) LHF	0.165	4.76	
SAPT0	1.551	26.80	



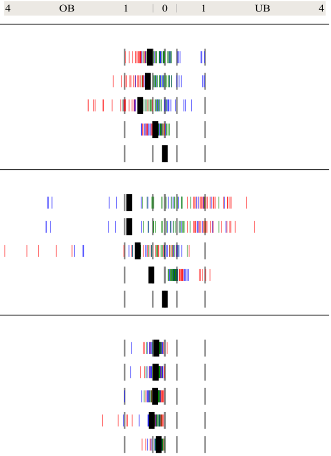
Korona S2 Results

Method	MAE	MURE	Error Distribution
Dispersion			
SAPT(DFT) hybrid	0.175	3.68	
SAPT(DFT) LHF	0.141	2.77	
SAPT(DFT) non-hybrid	0.326	9.58	
SAPT0	0.811	24.86	
Exchange-Dispersion			
SAPT(DFT) hybrid	0.062	12.47	
SAPT(DFT) LHF	0.039	3.25	
SAPT0	0.265	36.11	
Total			
SAPT(DFT) hybrid	0.155	4.98	
SAPT(DFT) LHF	0.189	4.17	
SAPT(DFT) hon-hybrid	0.244	10.64	
SAPT0	1.237	19.63	



S66 Results

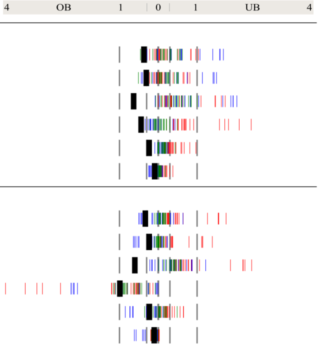
Method	Total	HB	MX	DD	Error Distribution
Electrostatics					
SAPT(DFT) hybrid	0.374	0.556	0.177	0.311	
SAPT(DFT) LHF	0.423	0.666	0.196	0.319	
SAPT0	0.613	1.034	0.439	0.297	
SAPT2+	0.236	0.270	0.136	0.263	
SAPT2+(3)δMP2	0.000	0.000	0.000	0.000	
Exchange					
SAPT(DFT) hybrid	0.886	1.127	0.426	0.926	
SAPT(DFT) LHF	0.886	1.121	0.431	0.928	
SAPT0	0.675	0.942	0.263	0.658	
SAPT2+	0.337	0.467	0.222	0.277	
SAPT2+(3)δMP2	0.000	0.000	0.000	0.000	
Induction					
SAPT(DFT) hybrid	0.211	0.201	0.212	0.220	
SAPT(DFT) LHF	0.224	0.223	0.223	0.225	
SAPT0	0.241	0.200	0.261	0.271	
SAPT2+	0.327	0.384	0.250	0.318	
SAPT2+(3)δMP2	0.152	0.179	0.121	0.145	



S66 Results

Method	Total	HB	MX	DD
Dispersion				
SAPT(DFT) hybrid	0.370	0.260	0.219	0.573
SAPT(DFT) LHF	0.308	0.200	0.173	0.499
SAPT(DFT) non-hybrid	0.635	0.581	0.419	0.822
SAPT0	0.443	0.862	0.162	0.195
SAPT2+	0.235	0.397	0.169	0.115
SAPT2+(3)δMP2	0.093	0.129	0.056	0.080
Total				
SAPT(DFT) hybrid	0.334	0.588	0.107	0.217
SAPT(DFT) LHF	0.234	0.382	0.046	0.199
SAPT(DFT) non-hybrid	0.604	0.955	0.389	0.385
SAPT0	0.990	1.197	0.692	0.965
SAPT2+	0.230	0.235	0.138	0.280
SAPT2+(3)δMP2	0.105	0.056	0.082	0.169

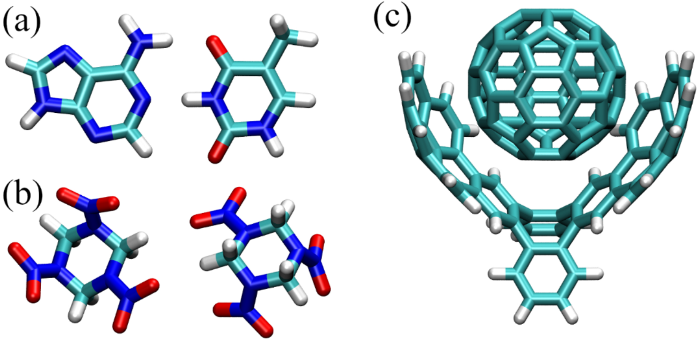
Error Distribution



Timing Performance

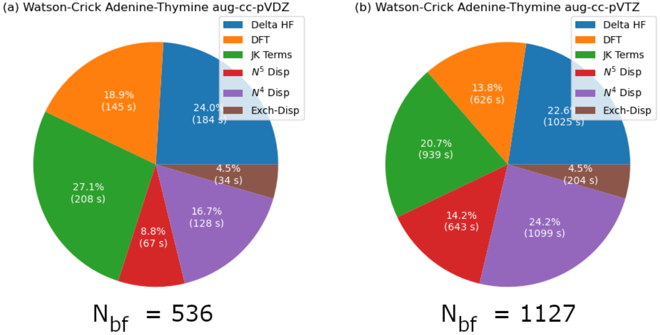
- Analyzing breakdown wall times for subroutines in SAPT(DFT) for a few systems with 500–3000 basis functions
- Using Intel Core i7-6800K processor with 6 cores for Watson-Crick adenine-thymine complex and RDX dimer
- Using Intel Core i9-10980XE processor with 18 cores for C_{60}–buckycatcher ($N_{bf} = 3012$), completed entire calculation in 4.03 days
- Contribution of $O(N^5)$ dispersion terms not dominant for smaller systems
- Cost of SCF (HF/DFT) calculations and induction terms are usually non-negligible

Timing Systems

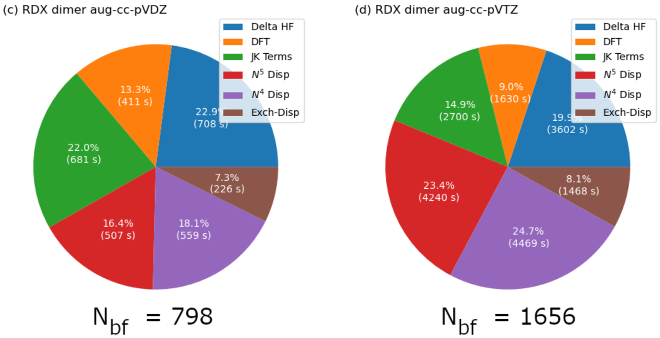


Dimer systems for timing: (a) Watson-Crick adenine-thymine complex, (b) RDX dimer, (c) C_{60}–buckycather complex.

Watson-Crick Adenine-Thymine



RDX Dimer

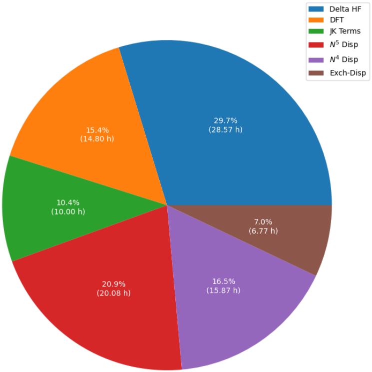

(c) RDX dimer aug-cc-pVDZ
Delta HF
DFT
JK Terms
N^5 Disp
N^4 Disp
Exch-Disp
13.3% (411 s)
22.9% (708 s)
22.0% (681 s)
7.3% (226 s)
16.4% (507 s)
18.1% (559 s)
N_{bf} = 798
(d) RDX dimer aug-cc-pVTZ
Delta HF
DFT
JK Terms
N^5 Disp
N^4 Disp
Exch-Disp
9.0% (1630 s)
14.9% (2700 s)
19.9% (3602 s)
23.4% (4240 s)
8.1% (1468 s)
24.7% (4469 s)
N_{bf} = 1656

Comparison with LHF Approach

- Comparison of subrountine wall times between hyybrid xc kernel approach implemented in Psi4 1.4 and LHF approach implemented in Molpro 2019.2
- Some subroutines does not exist or not included by default in the Molpro DFT-SAPT program

Subroutine	Psi4 hybrid time (h)	Molpro LHF time (h)
Delta HF	0.96	N/A
DFT	0.45	2.29
xc kernel	0.08	4.17
$O(N^5)$ objects formation	2.35	N/A
$E_{disp}^{(2)}$ time integration	0.37	3.59
$E_{exch-disp}^{(2)}$	0.41	1.99
Total	5.03	12.80

C_{60}–Buckycatcher Complex



$N_{bf} = 3012$

Table of contents

Table of contents

3B-69 Benchmark Set

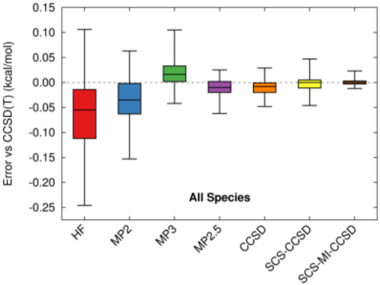
- Benchmark set for 3-body interaction energies[4]
- 69 trimers extracted from 23 different molecular crystal structures (3 each)
- Used focal point approach to obtain CCSD(T) (and other wavefunction method) energies

$$E = E^{\mathrm{HF}}(\mathrm{aQZ}) + \Delta\mathrm{E}^{\mathrm{MP2}}(\mathrm{aTZ/aQZ}) + \Delta\mathrm{E}^{\mathrm{CCSD(T)}}(\mathrm{aDZ})$$

- Assessing accuracy for various wavefunction and DFT methods
- Recommended MP2.5 and SCS-MI-CCSD, both $O(N^6)$
- We will extend this work to assess the performance of MP2 + FDDS dispersion for 3B-69 systems

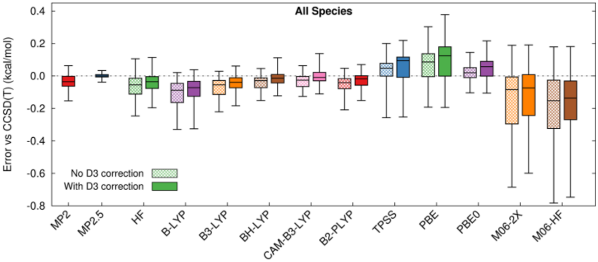
[4] J. Řezáč et al., J. Chem. Theory Comput. **11**, 3065 (2015).

3B-69 Wavefunctional Methods



- MP2.5: Non-iterative $O(N^6)$
- CCSD and variants: Iterative $O(N^6)$

3B-69 DFT Methods



- DFT-D3 accuracies comparable to MP2 at the best, in contrast to the two-body case where DFT-D3 models significantly outperform MP2
- Delocalization error leads to errors in many-body polarization and exchange

Table of contents

Three-Body FDDS Dispersion

- Three-body dispersion energy in terms of FDDS, analogous to the two-body dispersion:

$$
\begin{aligned}
E^{(3)}_{disp,r} \quad = \quad & -\frac{1}{\pi}\int_0^\infty d\omega \int d\mathbf{r}_A d\mathbf{r}'_A d\mathbf{r}_B d\mathbf{r}'_B d\mathbf{r}_C d\mathbf{r}'_C \\
& \frac{1}{|\mathbf{r}_A - \mathbf{r}_B|}\frac{1}{|\mathbf{r}'_A - \mathbf{r}_C|}\frac{1}{|\mathbf{r}'_B - \mathbf{r}'_C|} \\
& \chi^A\left(\mathbf{r}_A, \mathbf{r}'_A | i\omega\right)\chi^B\left(\mathbf{r}_B, \mathbf{r}'_B | i\omega\right)\chi^C\left(\mathbf{r}_B, \mathbf{r}'_B | i\omega\right)
\end{aligned}
$$

- Transform from position space into density-fitting auxiliary basis space:

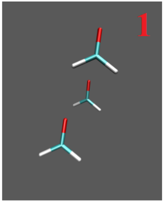
$$
E^{(3)}_{disp,r} = \int_0^\infty d\omega \,\mathrm{Tr}\left(\mathbf{S}^{-1}\boldsymbol{\chi}^A\mathbf{S}^{-1}\boldsymbol{\chi}^B\mathbf{S}^{-1}\boldsymbol{\chi}^C\right)
$$

New Set from X23

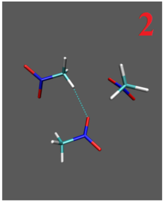
- Sampling trimer geometries from X23 crystal structures
- Trying to include trimer with different intermolecular distances and alignment
- Aiming to serve as a "three-body version" of S22$\times$5/S66$\times$8
- Distance: Geometry mean of 3 pairwise closest contact distance
- Alignment: Angles of the COM triangle; mainly looking at the greatest angle

Table of contents

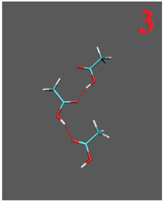
3B-69 Systems



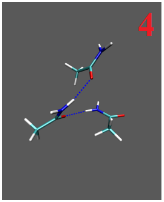
formaldehyde b



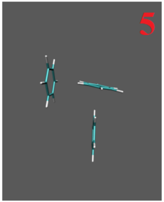
nitromethane a



acetic acid b



acetamide a



benzene c

3B-69 Results

- Total interaction energies in kcal/mol

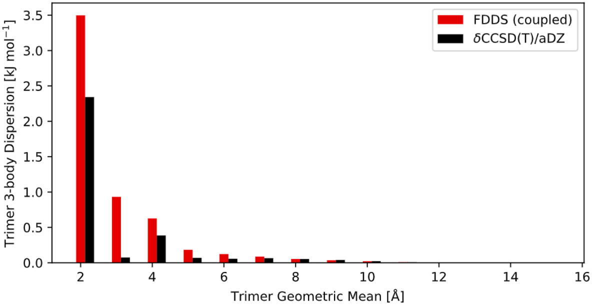
System	CCSD(T)	MP2 + FDDS/aDZ	MP2 + FDDS/aTZ	MP2	MP2.5
1	0.181	0.207	0.210	0.161	0.179
2	-0.122	-0.069	-0.065	-0.178	-0.143
3	-0.922	-0.905	-0.904	-0.937	-0.913
4	-0.089	-0.003	-0.003	-0.239	-0.151
5	-0.027	0.002	0.003	-0.061	-0.023

3B-69 Results

- Estimated 3-body dispersion energies comparison
- CCSD(T) corresponds to dispersion energy estimated by $E^{CCSD(T)} - E^{\text{MP2}}$

System	CCSD(T)	FDDS(aDZ)	FDDS(aTZ)
1	0.020	0.046	0.049
2	0.056	0.109	0.113
3	0.015	0.032	0.033
4	0.150	0.236	0.242
5	0.034	0.063	0.064

Distance Dependence



Distance Dependence