



Targeting Real chemical accuracy at the EXascale

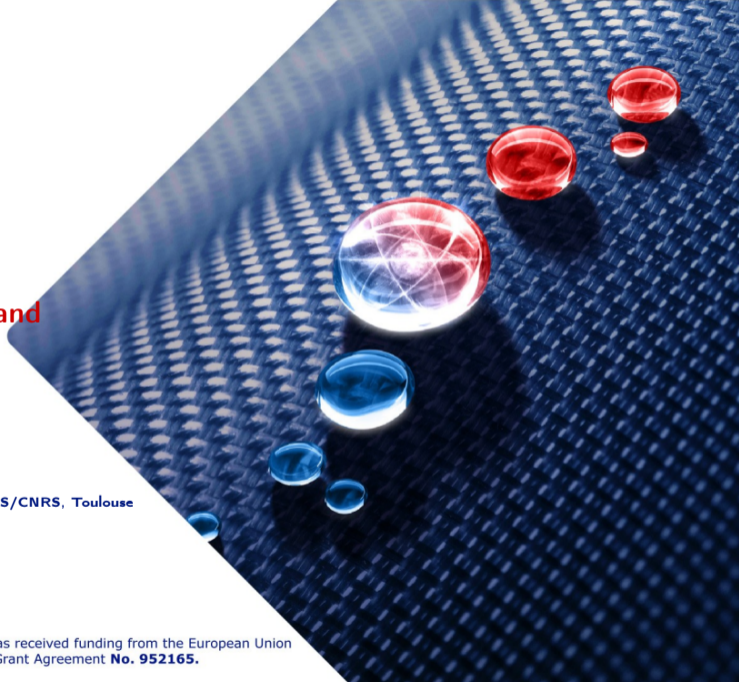
CIPSI: selected configuration interaction methods for ground and excited states

Abdallah Ammar, Emmanuel Giner,
Pierre-François Loos, Anthony Scemama

18th April 2023

Laboratoire de Chimie et Physique Quantiques, IRSAMC, UPS/CNRS, Toulouse

<https://lcpq.github.io/pterosor>



Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation programme under Grant Agreement No. 952165.

“Among the very large number of determinants contained in the FCI space, only a tiny fraction of them significantly contributes to the energy”

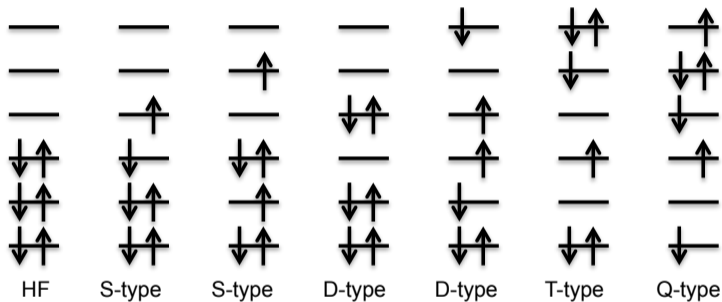
CIPSI = CI using a Perturbative Selection made Iteratively

- Developed in Toulouse many (many) years ago
Huron, Malrieu & Rancurel, JCP 58 (1973) 5745
- Based on old idea by Bender and Davidson, and Whitten and Hackmeyer
Bender & Davidson, Phys. Rev. 183 (1969) 23
Whitten & Hackmeyer, JCP 51 (1969) 5584
- CIPSI (and SCI methods in general) has been recently resurrected!
Giner, Scemama & Caffarel, CJC 91 (2013) 879
Giner, Scemama & Caffarel, JCP 142 (2015) 044115
- CIPSI \approx heat-bath CI (Umrigar) \approx adaptive sampling CI (Evangelista) \approx iterative CI (Liu) \approx incremental CI (Zimmerman) \approx FCIQMC (Alavi)

- This is the **oldest** and perhaps the **easiest method to understand**
- CI is based on the **variational principle** [like the Hartree-Fock (HF) approximation]
- The CI wave function is a **linear combination of determinants**
- CI methods use **excited determinants** to “improve” the reference (usually HF) wave function

$$|\Phi_{\text{CI}}\rangle = \underbrace{c_0 |\Psi_0\rangle}_{\text{reference}} + \underbrace{\sum_i^a c_i^a |\Psi_i^a\rangle}_{\text{singles}} + \underbrace{\sum_{\substack{i < j \\ a < b}} c_{ij}^{ab} |\Psi_{ij}^{ab}\rangle}_{\text{doubles}} + \underbrace{\sum_{\substack{i < j < k \\ a < b < c}} c_{ijk}^{abc} |\Psi_{ijk}^{abc}\rangle}_{\text{triples}} + \underbrace{\sum_{\substack{i < j < k < l \\ a < b < c < d}} c_{ijkl}^{abcd} |\Psi_{ijkl}^{abcd}\rangle}_{\text{quadruples}} + \dots$$

Excited determinants



CI wave function

$$|\Phi_{CI}\rangle = c_0 |0\rangle + c_S |S\rangle + c_D |D\rangle + c_T |T\rangle + c_Q |Q\rangle + \dots$$

- When $|S\rangle$ (**singles**) are taken into account: **CIS**

$$|\Phi_{\text{CIS}}\rangle = c_0 |0\rangle + c_S |S\rangle$$

NB: CIS is an **excited state method**

- When $|S\rangle$ and $|D\rangle$ are taken into account: **CISD**

$$|\Phi_{\text{CISD}}\rangle = c_0 |0\rangle + c_S |S\rangle + c_D |D\rangle$$

NB: CISD is the **most commonly-used** CI method

- When $|S\rangle$, $|D\rangle$ and $|T\rangle$ (**triples**) are taken into account: **CISDT**

$$|\Phi_{\text{CISDT}}\rangle = c_0 |0\rangle + c_S |S\rangle + c_D |D\rangle + c_T |T\rangle$$

- **CISDTQ**, etc.

- When all possible excitations are taken into account, this is called a Full CI calculation (FCI)

$$|\Phi_{\text{FCI}}\rangle = c_0 |0\rangle + c_S |S\rangle + c_D |D\rangle + c_T |T\rangle + c_Q |Q\rangle + \dots$$

- FCI gives the exact solution of the Schrödinger equation within a given basis
- FCI is becoming more and more fashionable these days (e.g. FCIQMC and CIPSI methods)
- So, why do we care about other methods?
- Because FCI is super computationally expensive!

"Assume we have 10 electrons in 38 spin MOs: 10 are occupied and 28 are empty"

- There is C_{10}^k possible ways of selecting k electrons out of the 10 occupied orbitals

$$C_n^k = \frac{n!}{k!(n-k)!}$$

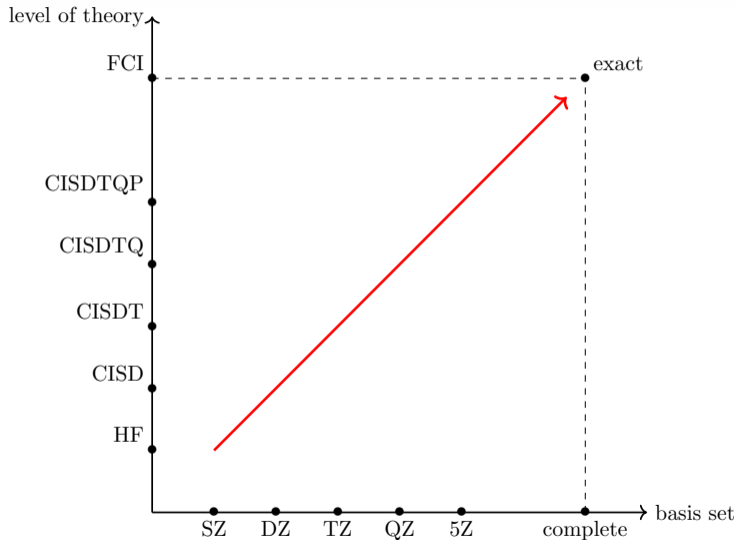
- There is C_{28}^k ways of distributing them out in the 28 virtual orbitals
- For a given excitation level k , there is $C_{10}^k C_{28}^k$ excited determinants
- The total number of possible excited determinant is

$$\sum_{k=0}^{10} C_{10}^k C_{28}^k = C_{38}^{10} = 472,733,756$$

- This is a lot...

For $n = 10$ and $N = 38$:

k	Num. of excitations
0	1
1	280
2	17,010
3	393,120
4	4,299,750
5	24,766,560
6	79,115,400
7	142,084,800
8	139,864,725
9	69,069,000
10	13,123,110
Tot.	472,733,756

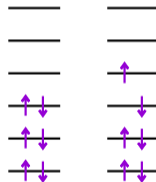


e	
0	
1	
2	
3	

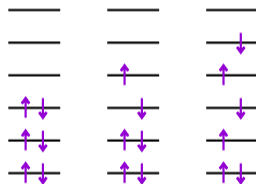
e	
0	HF
1	
2	
3	



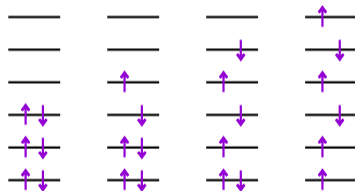
e	
0	
1	CIS
2	
3	



e	
0	
1	
2	CISD
3	



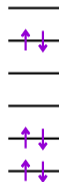
e	
0	
1	
2	
3	CISDT



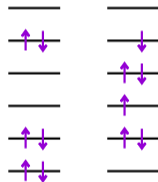


s	0	2	4	6

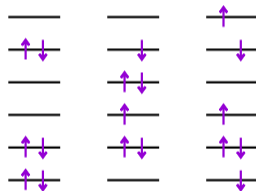
s	0	2	4	6
	sCI0			



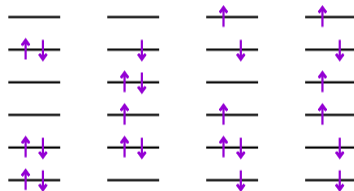
s	0	2	4	6
		sCI2		



s	0	2	4	6
			sCI4	



s	0	2	4	6
				sCI6



e/s	0	2	4	6	8
0					
1					
2					
3					
4					
5					
6					

e/s	0	2	4	6	8
0	HF				
1					
2					
3					
4					
5					
6					

e/s	0	2	4	6	8
0					
1		CIS			
2					
3					
4					
5					
6					

e/s	0	2	4	6	8
0					
1					
2			CISD		
3					
4					
5					
6					

e/s	0	2	4	6	8
0					
1					
2					
3				CISDT	
4					
5					
6					

e/s	0	2	4	6	8
0					
1					
2					
3					
4					
5					
6					

e/s	0	2	4	6	8
0	sCI0				
1					
2					
3					
4					
5					
6					

e/s	0	2	4	6	8
0					
1		sCI2			
2					
3					
4					
5					
6					

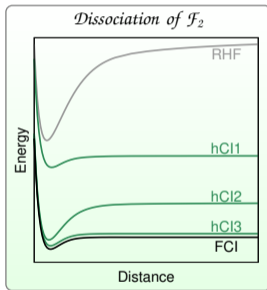
e/s	0	2	4	6	8
0					
1					
2			sCI4		
3					
4					
5					
6					

e/s	0	2	4	6	8
0					
1					
2					
3				sCI6	
4					
5					
6					

Hierarchy configuration interaction (hCI)

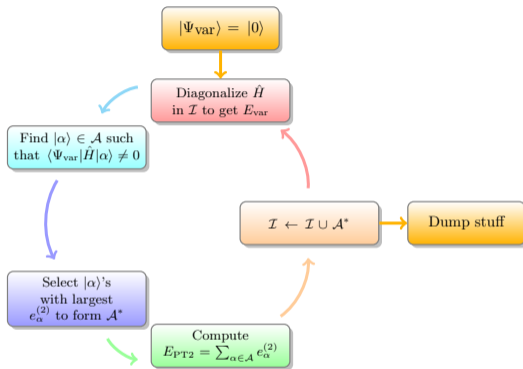
Excitation degree e
Seniority number s
Hierarchy parameter $h = \frac{e+s/2}{2}$

e/s	0	2	4	6
0	HF			
1		hCI1		
2		hCI1.5	hCI2	
3			hCI2.5	hCI3
4				
5				
6				

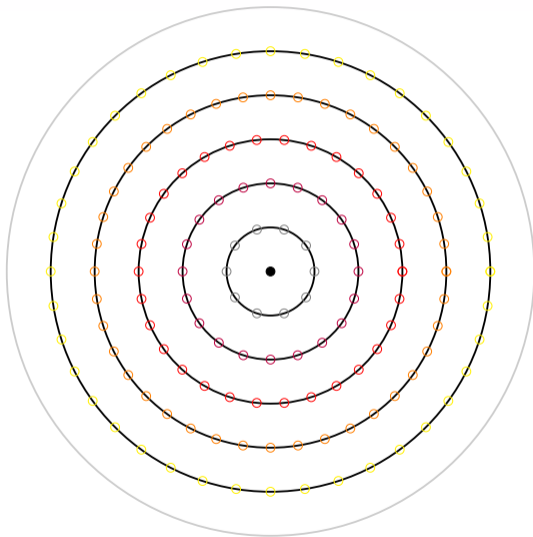


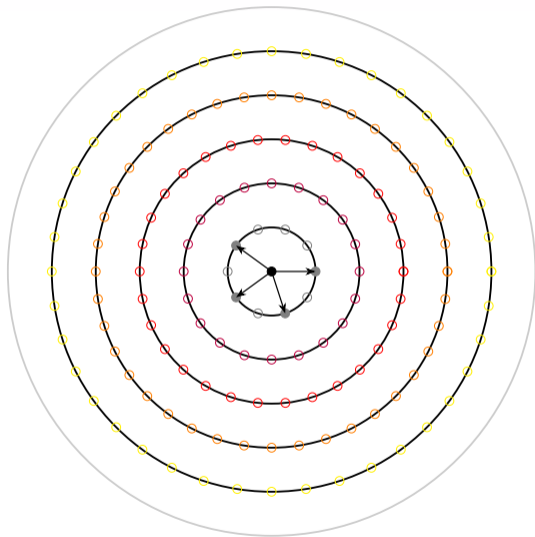
Fábris Kossoski

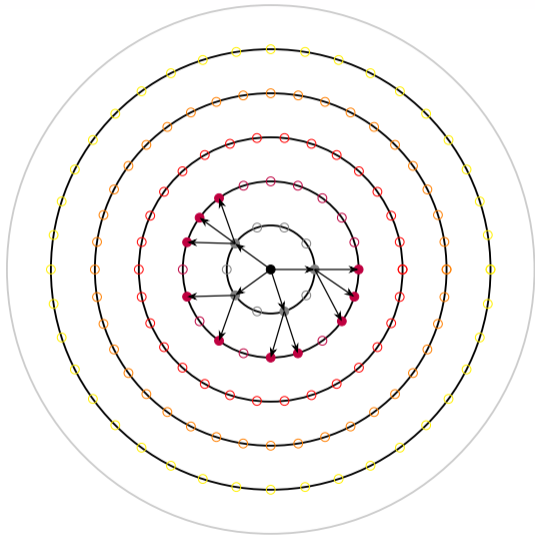
Kossoski, Damour & Loos, JPCL 13 (2022) 4342

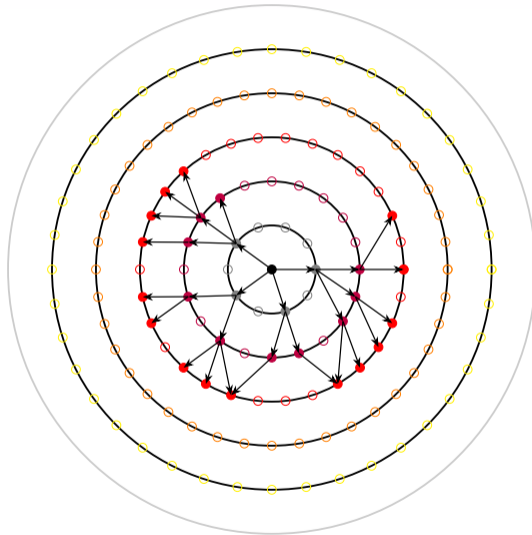


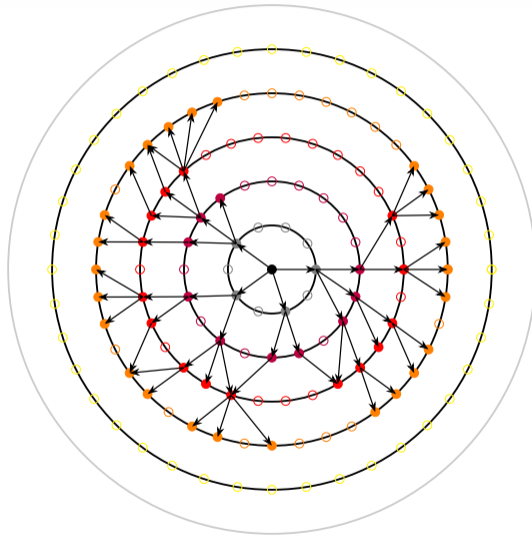
CIPSI is an algorithm, not a method...

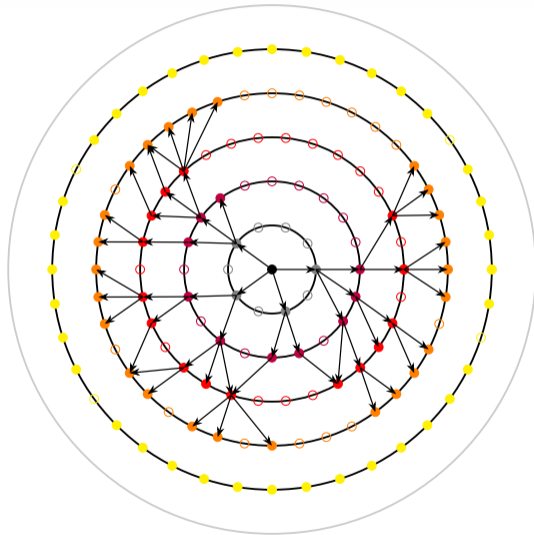












- **Green:** reference/variational/internal wave function (zeroth-order or model space)
- **Red:** perturbers or external wave function (first-order or perturbative space)

internal space $ I\rangle$	{	$\mathbf{H}^{(0)}$	\mathbf{h}^\dagger
perturbers $ \alpha\rangle$	{	\mathbf{h}	$\mathbf{H}^{(1)}$

- 1 Define a (zeroth-order) *reference* wave function:

$$|\Psi^{(0)}\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad E^{(0)} = \frac{\langle \Psi^{(0)} | \hat{H} | \Psi^{(0)} \rangle}{\langle \Psi^{(0)} | \Psi^{(0)} \rangle}$$

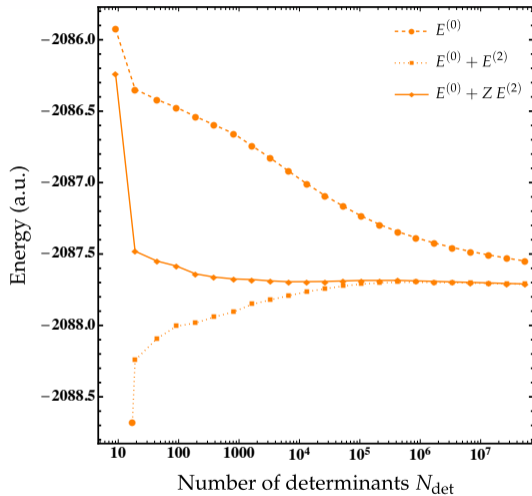
- 2 Generate *external determinants*:

$$\mathcal{A} = \left\{ (\forall I \in \mathcal{D}) \left(\forall \hat{T} \in \mathcal{T}_1 \cup \mathcal{T}_2 \right) : |\alpha\rangle = \hat{T} |I\rangle \right\}$$

- 3 Second-order perturbative contribution of each $|\alpha\rangle$:

$$\delta E(\alpha) = \frac{|\langle \Psi^{(0)} | \hat{H} | \alpha \rangle|^2}{E^{(0)} - \langle \alpha | \hat{H} | \alpha \rangle}$$

- 4 Select $|\alpha\rangle$ with largest $\delta E(\alpha)$ and add them to \mathcal{D}
- 5 Diagonalize \hat{H} in $\mathcal{D} \Rightarrow$ update $|\Psi^{(0)}\rangle$ and $E^{(0)}$
- 6 Iterate

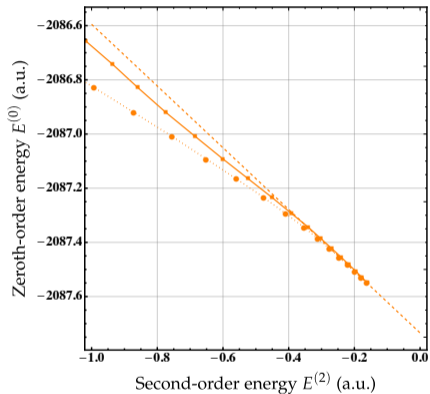
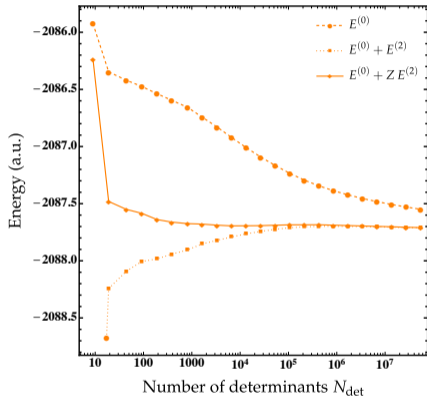


Garniron et al., JCTC 15 (2019) 3591

- Second-order Epstein-Nesbet correction:

$$E^{(2)} = \sum_{\alpha} \delta E(\alpha)$$

- $|\alpha\rangle$'s with largest $\delta E(\alpha)$ have been added to $\Psi^{(0)}$ previously
 ⇒ only small contributions remaining
- \mathcal{A} increases with \mathcal{D}
 ⇒ a *very large* number of *very small* contributions
- In practice, we use a semi-stochastic algorithm to compute $E^{(2)}$
 ⇒ *much faster!!*
 Garniron, Scemama, Loos & Caffarel, JCP 147 (2017) 034101
- We linearly extrapolate to $E^{(2)} = 0$ to reach the FCI limit (exFCI)



Garniron et al., JCTC 15 (2019) 3591

At a given CIPSI iteration, the SCI+PT2 energy is given by

$$E = E^{(0)} + E^{(2)}$$

Let us introduce the following energy-dependent second-order self-energy

$$\Sigma^{(2)}[E] = \sum_{\alpha} \frac{\langle \alpha | \hat{H} | \Psi^{(0)} \rangle^2}{E - \langle \alpha | \hat{H} | \alpha \rangle} \quad \text{with} \quad \Sigma^{(2)}[E^{(0)}] = E^{(2)}$$

Brillouin-Wigner perturbation theory tells us

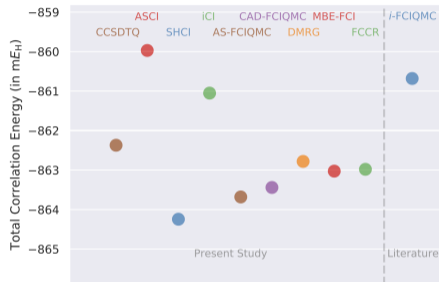
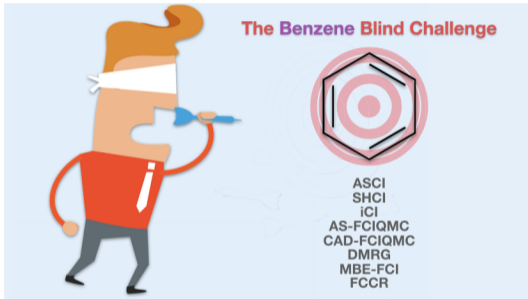
$$E = E^{(0)} + \Sigma^{(2)}[E]$$

Assuming that $\Sigma^{(2)}[E]$ behaves linearly for $E \approx E^{(0)}$

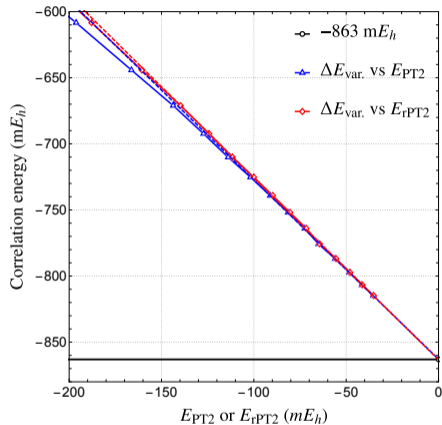
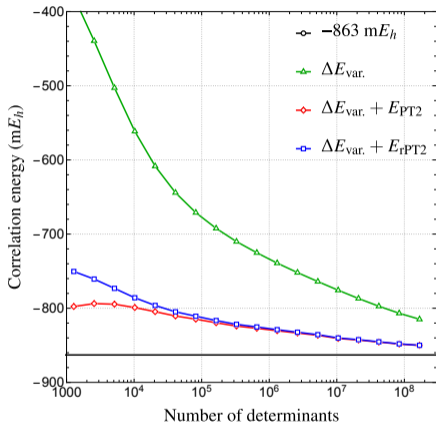
$$\Sigma^{(2)}[E] \approx \Sigma^{(2)}[E^{(0)}] + (E - E^{(0)}) \left. \frac{\partial \Sigma^{(2)}[E]}{\partial E} \right|_{E=E^{(0)}}$$

This yields

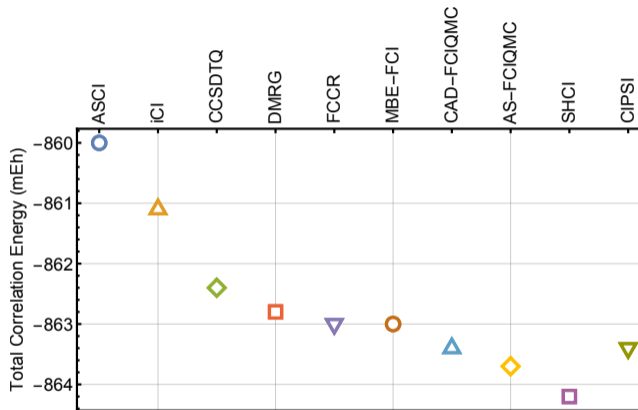
$$E = E^{(0)} + \Sigma^{(2)}[E^{(0)}] + (E - E^{(0)}) \left. \frac{\partial \Sigma^{(2)}[E]}{\partial E} \right|_{E=E^{(0)}} = E^{(0)} + Z E^{(2)} \quad \text{with} \quad Z = \left[1 - \left. \frac{\partial \Sigma^{(2)}[E]}{\partial E} \right|_{E=E^{(0)}} \right]^{-1}$$



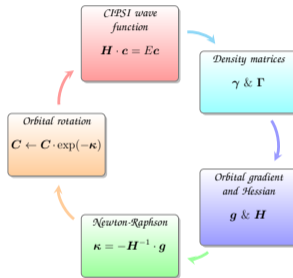
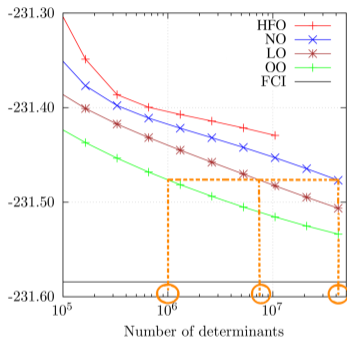
Eriksen et al. JPCL 11 (2020) 8922



Loos, Damour & Scemama, JCP 153 (2020) 176101



Loos, Damour & Scemama, JCP 153 (2020) 176101



- Orbital optimization largely accelerates the convergence of selected CI
- Trust-region Newton-Raphson algorithm



Yann Damour

Damour, V eril, Kossoski, Caffarel, Jacquemin, Scemama & Loos, JCP 155 (2020) 176101

Toward a systematic improvement of the fixed-node approximation in diffusion Monte Carlo for solids—A case study in diamond

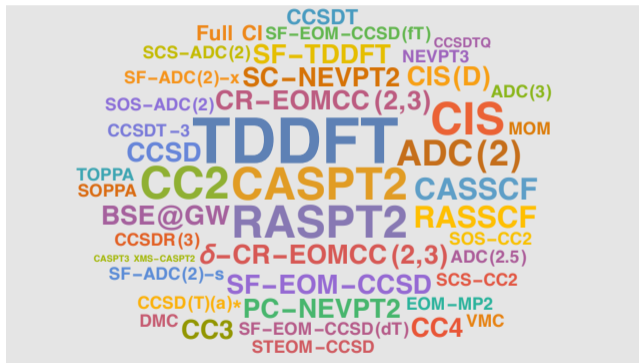
Cite as: J. Chem. Phys. **153**, 184111 (2020); <https://doi.org/10.1063/5.0021036>

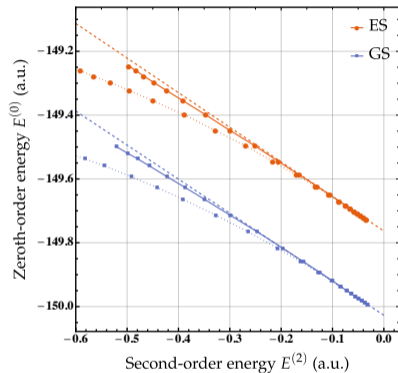
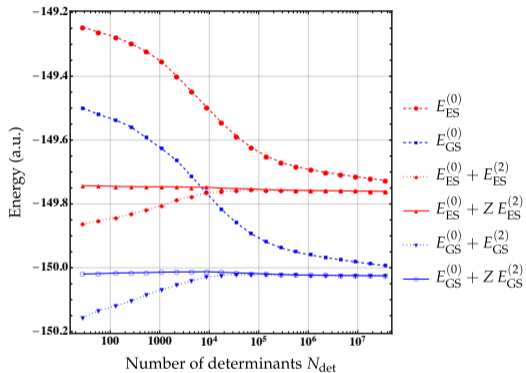
Submitted: 06 July 2020 . Accepted: 12 October 2020 . Published Online: 11 November 2020

 Anouar Benali,  Kevin Gasperich,  Kenneth D. Jordan, Thomas Applencourt,  Ye Luo,  M. Chandler Bennett,  Jaron T. Krogel,  Luke Shulenburger,  Paul R. C. Kent,  Pierre-François Loos,  Anthony Scemama, and  Michel Caffarel

See also [Scemama et al. JCP 153 \(2021\) 174107](#) for a range-separated approach in molecules

And this is just for excited states...





Garniron et al., JCTC 15 (2019) 3591

Table 1. Zeroth-Order Energy $E^{(0)}$, Second-Order Perturbative Correction $E^{(2)}$, and Its Renormalized Version $ZE^{(2)}$ (in hartree) of CN3 for Increasingly Large Wave Functions^a

N_{det}	$E^{(0)}$		$E^{(0)} + E^{(2)}$			$E^{(0)} + ZE^{(2)}$		
	GS (a.u.)	ES (a.u.)	GS (a.u.)	ES (a.u.)	$\Delta E(\text{eV})$	GS (a.u.)	ES (a.u.)	$\Delta E(\text{eV})$
28	-149.499574	-149.246268	-150.155(1)	-149.863(1)	7.95(5)	-150.020(1)	-149.743(1)	7.54(5)
58	-149.519908	-149.261390	-150.134(1)	-149.853(1)	7.67(5)	-150.018(1)	-149.744(1)	7.48(5)
131	-149.537424	-149.277496	-150.118(1)	-149.8427(9)	7.52(4)	-150.017(1)	-149.7449(9)	7.39(4)
268	-149.559465	-149.298484	-150.1035(9)	-149.8308(9)	7.42(4)	-150.0158(9)	-149.7457(9)	7.35(4)
541	-149.593434	-149.323302	-150.0845(8)	-149.8186(8)	7.24(4)	-150.0152(8)	-149.7463(8)	7.32(4)
1101	-149.627202	-149.354807	-150.0683(8)	-149.8045(8)	7.18(3)	-150.0137(8)	-149.7460(8)	7.28(3)
2207	-149.663850	-149.399522	-150.0549(7)	-149.7879(7)	7.26(3)	-150.0132(7)	-149.7462(7)	7.27(3)
4417	-149.714222	-149.448133	-150.0409(6)	-149.7762(6)	7.20(3)	-150.0130(6)	-149.7478(6)	7.22(3)
8838	-149.765886	-149.496401	-150.0296(5)	-149.7655(5)	7.19(2)	-150.0124(5)	-149.7473(5)	7.21(2)
17 680	-149.817301	-149.545048	-150.0239(4)	-149.7615(4)	7.14(2)	-150.0141(4)	-149.7505(4)	7.17(2)
35 380	-149.859737	-149.587668	-150.0216(3)	-149.7582(3)	7.17(1)	-150.0161(3)	-149.7518(3)	7.19(1)
70 764	-149.893273	-149.623235	-150.0207(2)	-149.7566(3)	7.18(1)	-150.0174(2)	-149.7530(3)	7.19(1)
141 545	-149.919463	-149.650109	-150.0214(2)	-149.7572(2)	7.189(8)	-150.0194(2)	-149.7550(2)	7.196(8)
283 108	-149.937839	-149.669735	-150.0224(2)	-149.7576(2)	7.206(7)	-150.0211(2)	-149.7562(2)	7.209(7)
566 226	-149.950918	-149.683278	-150.0233(1)	-149.7580(1)	7.217(6)	-150.0223(1)	-149.7570(1)	7.219(6)
1 132 520	-149.960276	-149.693053	-150.0238(1)	-149.7588(1)	7.212(5)	-150.0231(1)	-149.7580(1)	7.214(5)
2 264 948	-149.968203	-149.700907	-150.0240(1)	-149.7590(1)	7.211(4)	-150.0235(1)	-149.7584(1)	7.212(4)
4 529 574	-149.975230	-149.708061	-150.0245(1)	-149.7594(1)	7.215(4)	-150.0241(1)	-149.7589(1)	7.216(4)
9 057 914	-149.981770	-149.714526	-150.02463(9)	-149.75981(8)	7.206(3)	-150.02434(9)	-149.75948(8)	7.207(3)
18 110 742	-149.987928	-149.720648	-150.02495(7)	-149.76025(8)	7.203(3)	-150.02474(7)	-149.76000(8)	7.204(3)
36 146 730	-149.993593	-149.726253	-150.02527(6)	-149.76065(7)	7.198(3)	-150.02502(6)	-149.760 47(7)	7.198(3)

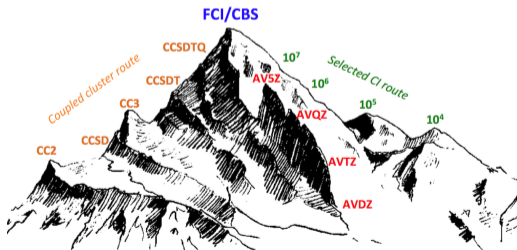
^aThe excitation energy ΔE (in eV) is the energy difference between the ground state (GS) and the excited state (ES). The statistical error, corresponding to one standard deviation, is reported in parentheses.

A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks

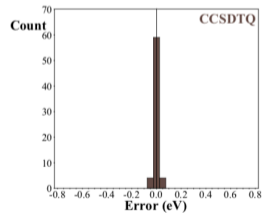
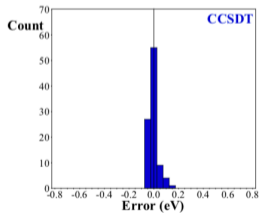
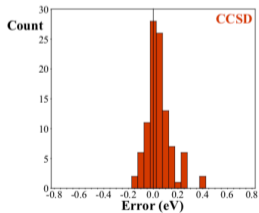
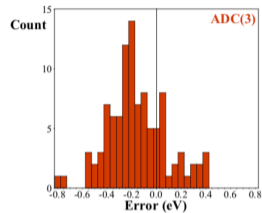
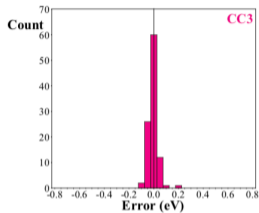
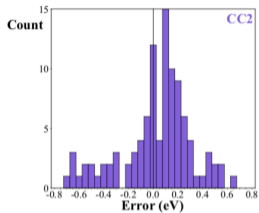
Pierre-François Loos,^{*,†} Anthony Scemama,[†] Aymeric Blondel,[‡] Yann Garniron,[†] Michel Caffarel,[†] and Denis Jacquemin^{*,‡}

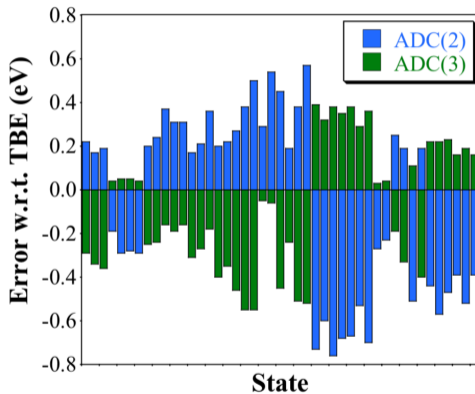
[†]Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, 31013 Toulouse Cedex 6, France

[‡]Laboratoire CEISAM - UMR CNRS 6230, Université de Nantes, 2 Rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France



- 110 vertical excitation energies (VTEs) and oscillator strengths
- 18 small molecules with 1 to 3 non-H atoms
- CC3/aug-cc-pVTZ geometries
- mostly singly-excited states and very few doubly-excited states
- rely on FCI to define “theoretical best estimates” (TBEs)
- aug-cc-pVTZ and CBS vertical energies
- benchmark popular excited-state methods accounting for double and triple excitations



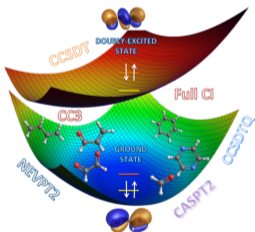


Reference Energies for Double Excitations

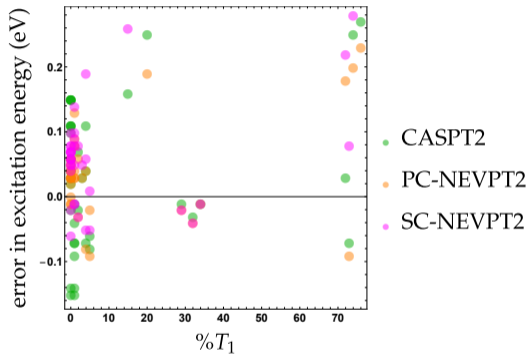
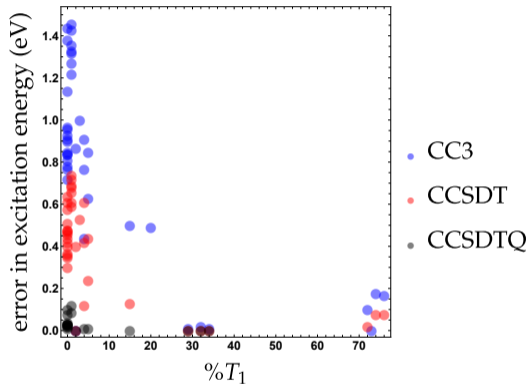
Pierre-François Loos,^{*,†} Martial Boggio-Pasqua,[†] Anthony Scemama,[†] Michel Caffarel,[†] and Denis Jacquemin[‡]

[†]Laboratoire de Chimie et Physique Quantiques (UMR 5626), Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

[‡]Laboratoire CEISAM (UMR 6230), CNRS, Université de Nantes, 44399 Cedex 3 Nantes, France

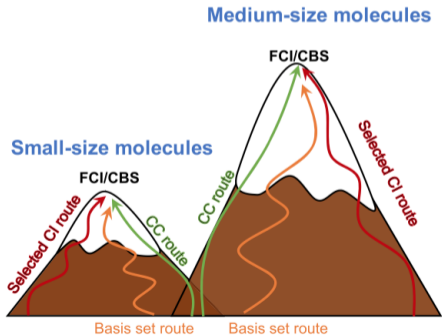


- 20 VTEs for doubly-excited states
- 14 small- and medium-sized molecules
- mostly rely on FCI to define TBEs (except for the largest molecules)
- aug-cc-pVTZ and CBS vertical energies
- benchmark excited-state methods including at least triple excitations
- additional benchmarks of multi-configurational methods



A Mountaineering Strategy to Excited States: Highly Accurate Energies and Benchmarks for Medium Sized Molecules

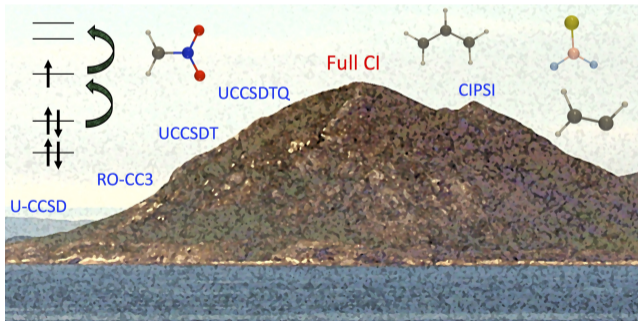
Pierre-François Loos,* Filippo Lipparini,* Martial Boggi-Pasqua, Anthony Scemama, and Denis Jacquemin*



- 238 VTEs (and oscillator strengths) with mostly singly-excited states and aug-cc-pVTZ basis
- 27 small- and medium-sized molecules with 4 to 6 non-H atoms
- rely mostly on **CCSDT** or CCSDTQ to define TBEs
- benchmark popular excited-state methods accounting for double and triple excitations
- recently improved TBEs with CC4 and CCSDTQ [JCP 154 (2021) 221103; JCTC 18 (2022) 4418]

Mountaineering Strategy to Excited States: Highly Accurate Energies and Benchmarks for Exotic Molecules and Radicals

Pierre-François Loos,* Anthony Scemama, Martial Boggio-Pasqua, and Denis Jacquemin*



- two subsets of excitations and oscillator strengths
- an “exotic” subset of 30 VTEs for closed-shell molecules containing F, Cl, P, and Si
- a “radical” subset of 51 doublet-doublet transitions in 24 small radicals
- total of 81 TBEs mostly obtained at the FCI/aug-cc-pVTZ level
- benchmark popular excited-state methods (U vs RO)

Received: 2 December 2020 | Revised: 5 January 2021 | Accepted: 7 January 2021

DOI: 10.1002/wcms.1517

FOCUS ARTICLE

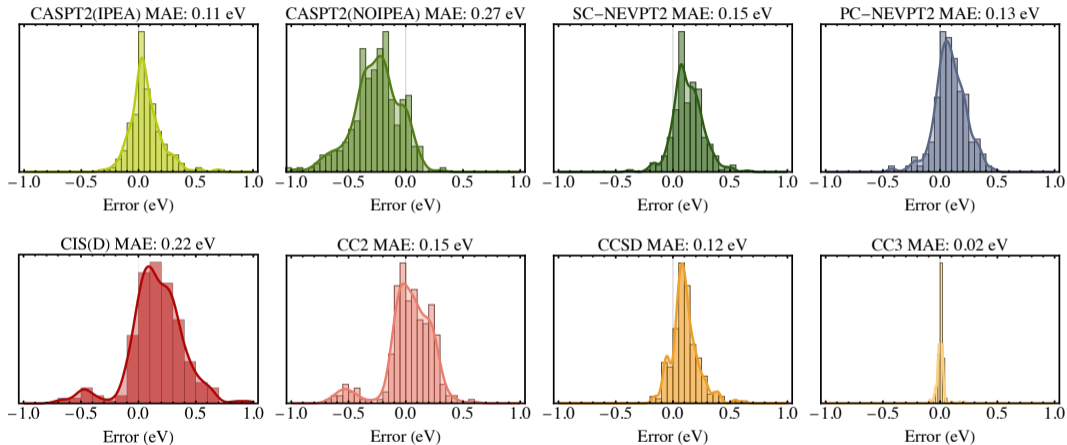


WILEY

QUESTDB: A database of highly accurate excitation energies for the electronic structure community

Mickaël Véral¹ | Anthony Scemama¹  | Michel Caffarel¹ | Filippo Lipparini² | Martial Boggio-Pasqua¹ | Denis Jacquemin³ | Pierre-François Loos¹ 

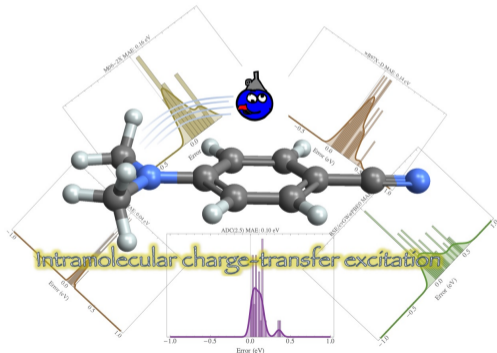
- 13 new systems composed by small molecules as well as **larger** molecules
- 80 new transitions the vast majority being of CCSDT quality
- benchmark popular excited-state methods **over the entire database**



JCTC 18 (2022) 2418; JCP 157 (2022) 014103

Reference Energies for Intramolecular Charge-Transfer Excitations

Pierre-François Loos,* Massimiliano Comin, Xavier Blase,* and Denis Jacquemin*

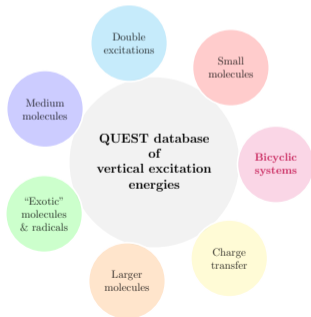


- intramolecular **charge-transfer** transitions in π -conjugated molecules
- 30 transitions of CCSDT quality in 17 systems with cc-pVTZ
- Basis-set correction up to aug-cc-pVQZ computed with CCSD and CC2
- benchmark popular excited-state methods
- **additional benchmarks** of BSE@GW and TD-DFT (hybrids and range-separated hybrids)

A Mountaineering Strategy to Excited States: Highly Accurate Energies and Benchmarks for Bicyclic Systems

Published as part of *The Journal of Physical Chemistry virtual special issue "Vincenzo Barone Festschrift"*.

Pierre-François Loos* and Denis Jacquemin*



- VTEs for 10 bicyclic molecules
- 91 new transitions of CCSDT quality for larger systems
- benchmark popular excited-state methods accounting for double and triple excitations

QUEST: a database of highly-accurate excitation energies

HOME DATASET SUBSETS REFERENCES

QUESTDB
 Quantum Excited States DataBase

AVTZ AVDZ NEVPT2 CASPT2 CCSDTQ CASPT2

Medium-size molecules

Small-size molecules

FCI/CBS Selected

UCCSDTQ Full CI CIPSI

UCCSDT

A mountaineering strategy to excited states



Mika Vénil

Vénil et al. WIREs Comput. Mol. Sci. 11 (2021) e1517

https://lcpq.github.io/QUESTDB_website

- Head-Gordon's group: orbital-optimized DFT for double excitations [JCTC 16 (2020) 1699; JPCL 12 (2021) 4517] and TD-DFT benchmark [JCTC 18 (2022) 3460]
- Kaupp's group: assessment of hybrid functionals [JCP 155 (2021) 124108]
- Kallay's and Goerigk's groups: double hybrids [JCTC 15 (2019) 4735; JCTC 17 (2021) 927; JCTC 17 (2021) 5165; JCTC 17 (2021) 4211]
- Truhlar/Gagliardi's group: p-DFT [JCTC 18 (2022) 6065]
- Bartlett's group: Variants of EOM-CC for doubly-excited states [JCP 156 (2022) 201102]
- Neuscamman's group: QMC for doubly-excited states [JCP 153 (2022) 234105]
- Filippi/Scemama's groups: QMC for excited states [JCTC 15 (2019) 4889; JCTC 17 (2021) 3426; JCTC 18 (2022) 1089]
- Gould's group: ensemble DFT [JPCL 13 (2022) 2452]

✗ Forget about large systems/basis sets:

JCTC 16 (2020) 1711

- 1-3 non-H atoms with triple- or quadruple- ζ basis
- 4-6 non-H atoms with double- ζ basis

✓ Open-shell systems are “easy” (no spin contamination and independent of starting orbitals)

JCTC 16 (2020) 3720

✓ Double excitations are easily accessible (especially if they have the same symmetry as the ground state)

JCTC 15 (2020) 1939

✓ You can post-process CIPSI wave functions!

- one- and two-body density matrices
- QMC trial wave functions