



HAL
open science

Electronic transitions, statistics and digital tools for quantum chemistry

Mickaël Veril

► **To cite this version:**

Mickaël Veril. Electronic transitions, statistics and digital tools for quantum chemistry. Theoretical and/or physical chemistry. Université Paul Sabatier - Toulouse III, 2021. English. NNT : 2021TOU30254 . tel-03699990

HAL Id: tel-03699990

<https://theses.hal.science/tel-03699990v1>

Submitted on 20 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par
Mickaël VERIL

Le 9 décembre 2021

Transitions électroniques, statistiques et outils numériques pour la chimie quantique.

Ecole doctorale : **SDM - SCIENCES DE LA MATIERE - Toulouse**

Spécialité : **Physico-Chimie Théorique**

Unité de recherche :

LCPQ-IRSAMC - Laboratoire de Chimie et Physique Quantiques

Thèse dirigée par

Pierre-François LOOS

Jury

Mme Elise Dumont, Rapporteur

M. Nicolas Ferré, Rapporteur

Mme Adèle Laurent, Examinatrice

M. Romuald Poteau, Examineur

M. Anthony Scemama, Examineur

M. Pierre-François Loos, Directeur de thèse

UNIVERSITÉ TOULOUSE III-PAUL SABATIER

DOCTORAL THESIS

**Electronic transitions, statistics and digital tools
for quantum chemistry**

Author:
Mickaël VÉRIL

Supervisor:
Pierre-François LOOS

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy*

in the

Méthodes et outils de la chimie quantique
Laboratoire de Chimie et Physique Quantiques

December 9, 2021

“Un problème créé ne peut être résolu en réfléchissant de la même manière qu’il a été créé.”

Albert Einstein

UNIVERSITÉ TOULOUSE III-PAUL SABATIER

Abstract

Faculté Sciences et Ingénierie
Laboratoire de Chimie et Physique Quantiques

Doctor of Philosophy

Electronic transitions, statistics and digital tools for quantum chemistry

by Mickaël VÉRIL

After a general introduction on excited states in molecular systems and computational resources (Chapter 1) as well as a general overview of the various methods that are employed in this thesis (Chapter 2), we will talk about two main topics: i) molecular electronic excited states and ii) digital tools for theoretical quantum chemistry.

The present thesis starts with a study of the GW family of approximations (Chapter 3). These are known to be accurate and relatively cheap methodologies for the computation of charged excitations such as ionization potentials and electronic affinities in molecules and band gaps in solids. These methods can be perturbative, partially or fully self-consistent and are able to provide chemically accurate properties for weakly correlated systems. However, as discussed in details, discontinuities can be observed for several key properties, particularly in potential energy surfaces even in the weakly correlated regime. These discontinuities are due to poles in the self-energy — a key quantity of the GW formalism — as each of its branch is associated with a distinct solution of the quasiparticle equation. We show that, in diatomic molecules, multisolution behavior in frontier orbitals is more likely if the energy gap between these orbitals is small.

The central part of this thesis deals with the QUEST project (Chapter 4), a large database of more than 600 vertical excitations energies of various natures. The aim of this open-source and easily-modifiable database is to provide highly-accurate reference excitations energies for benchmarking and cross comparisons of computational models. These reference excitations energies are divided into six subsets depending on the size of the molecules and the types of excitations. In order to gather the huge amount of data of the QUEST project, we have created a website where one can easily test and compare the accuracy of a given method with respect to various variables such as the molecule size or its family, the nature of the excited states, the type of basis set, etc.

The next part is about digital tools and, in particular, the implementation of a web demo of QUANTUM PACKAGE (Chapter 5), an open-source software developed primarily by our research group, which has been designed to be easy to use by quantum chemists, whether they are users or developers. The possibility of testing the software without the need to install it is a real opportunity to make QUANTUM PACKAGE more popular in our community. The containerization technology provided by DOCKER and the LINUX kernel allow us to quickly start a clean demo environment for each user and stop it easily with interesting performance.

Finally, in Chapter 6, pursuing the recent work of our team on the benzene molecule, we report reference frozen-core correlation energies for the twelve five- and six-membered ring molecules of the QUEST project using the *Configuration Interaction using a Perturbative Selection made Iteratively* (CIPSI) algorithm in the standard correlation-consistent double- ζ Dunning basis set (cc-pVDZ). These correspond to Hilbert spaces with sizes ranging from 10^{28} to 10^{36} . The performance and convergence properties of several series of methods are investigated. In particular, we study the convergence properties of the Møller-Plesset perturbation series up to fifth-order as well as various single-reference coupled-cluster models which include up to quadruple excitations.

We draw our general conclusions in Chapter 7.

Remerciements

Cette thèse est l'aboutissement de mes 10 ans d'études à l'université Toulouse III-Paul Sabatier. Je tiens d'abord à remercier les enseignants que j'ai côtoyés toutes ces années. Je pense en particulier à Pascal Dufour mon enseignant référent qui m'a beaucoup aidé durant mes années universitaires. Je ne saurais oublier de remercier Jean-Louis Heully qui m'a accompagné durant mes années de stage au LCPQ avec sa gentillesse, sa bonne humeur et son savoir, sans oublier Fabienne Alary pour m'avoir donné cette opportunité. Je tiens à exprimer ma profonde reconnaissance à mon directeur de thèse Pierre-François Loos pour l'énorme soutien moral et scientifique ainsi que sa patience qu'il m'a généreusement accordé tout au long de ces trois ans. Mes sincères remerciements vont aussi à Anthony Scemama pour ses précieux conseils, ses connaissances. Avoir vu l'architecture des logiciels qu'il a conçus comme QUANTUM PACKAGE a été très inspirant pour moi. Merci aussi à Denis Jaquemin, Arjan Berger, Pina Romaniello, Michel Caffarel et tous ceux avec qui j'ai travaillé. Je voudrais surtout remercier l'ensemble du personnel de mon laboratoire en particulier Thierry Leininger le directeur, pour m'avoir accueilli depuis quelques années et de m'avoir autorisé dans le contexte de la Covid-19 à faire du télétravail par rapport à ma santé fragile. Merci à mon groupe de recherche PTEROSOR pour les réunions du vendredi par Zoom qui dans ce contexte me permettait de rompre mon isolement. Je remercie aussi les membres du groupe THEO pour les réunions du jeudi matin toujours dans la bonne humeur. Et je pense également à Lionel Calmels mon parrain de thèse pour son dévouement, sa bienveillance et sa disponibilité. Je garde un bon souvenir d'Eric Colledani pour nos discussions, sa gentillesse et nos repas passés ensemble le midi. Merci à mes rapporteurs Élise Dumont et Nicolas Ferré pour la lecture de ma thèse et également aux autres membres du Jury Romuald Poteau, Anthony Scemama et Adèle Laurent; et à Denis Jaquemin d'avoir accepté d'être chercheur invité. Je voudrais remercier grandement Kristina Jakobson qui en plus d'avoir été mon AVS au lycée, est également ma professeure à domicile d'anglais. Pour finir, un grand merci à mes parents de m'avoir toujours soutenu et accompagné dans cette aventure ainsi que dans ma vie de tous les jours. Et pardon à ceux que j'aurai oublié de mentionner.

Contents

Abstract	v
Remerciements	vii
1 Introduction	1
1.1 Generality	1
1.2 Excited states	1
1.3 Methods for excited states	2
1.4 Computational resources	5
1.4.1 Curie	5
1.4.2 Olympe	5
1.4.3 Joliot-Curie	6
1.5 Structure of the manuscript	7
2 Methods	9
2.1 Atomic units	9
2.2 Schrödinger equation	9
2.2.1 Born-Oppenheimer approximation	10
2.3 The Hartree-Fock approximation	10
2.3.1 Restricted Hartree-Fock	11
2.3.2 Roothaan-Hall equations	12
2.3.3 The self-consistent field procedure	13
2.4 Post-Hartree-Fock methods	14
2.4.1 Configuration interaction method	14
2.4.2 Selected CI methods	15
2.4.3 Coupled-cluster methods	16
2.4.4 Møller-Plesset perturbative methods	18
3 Unphysical discontinuities in GW methods	19
3.1 Many-body Green's function methods	19
3.2 Theory	20
3.3 Results	23
3.3.1 Virtual orbitals	23
G_0W_0	24
evGW	24
3.3.2 Occupied orbitals	25
3.3.3 Frontier orbitals	26
3.4 Concluding remarks	27
4 The QUEST database of vertical excitation energies	29
4.1 Introduction	29
4.2 Computational tools	32
4.2.1 Geometries	32
4.2.2 Basis sets	32
4.2.3 Computational methods	32

	Reference computational methods	32
	Benchmarked computational methods	33
	Estimating the extrapolation error	34
4.3	The QUEST database	37
4.3.1	Overview	37
4.3.2	QUEST#1	37
4.3.3	QUEST#2	41
4.3.4	QUEST#3	41
4.3.5	QUEST#4	42
4.3.6	QUEST#5	42
4.4	Theoretical best estimates	55
4.5	Benchmarks	61
4.6	The QUESTDB website	74
4.6.1	Website specifications	76
4.6.2	Architecture	76
	The static website	76
	Data generation tools	77
4.6.3	Usage	78
	Manipulation and scenarios	78
4.7	Concluding remarks	78
5	Quantum Package demo in a web browser	81
5.1	Introduction	81
5.2	Shellinabox	81
5.3	Docker	82
5.3.1	Container lifecycle	82
5.3.2	Linux kernel component used by libcontainer for containerization	83
	The container filesystem and union mounting	83
	Resources handling with cgroups	83
	Host separation with the namespaces	83
5.4	In practice	84
5.5	Conclusion and perspectives	86
6	Accurate full configuration interaction correlation energy estimates for five- and six-membered rings	87
6.1	Introduction	87
6.2	CIPSI with optimized orbitals	89
6.3	Computational details	97
6.4	Results and discussion	100
6.4.1	CIPSI estimates	100
6.4.2	Benchmark of CC and MP methods	101
6.5	Conclusion	102
7	General conclusion	103
A	Résumé en français	107
B	Résumé étendu en français	109
	Bibliography	131

À mes parents !

Chapter 1

Introduction

1.1 Generality

The aim of theoretical chemistry is to mathematically describe chemical and biochemical systems in order to predict their properties. Formally, in quantum chemistry, one must solve the Schrödinger equation but, except for the hydrogen atom and its isoelectronic ions (the so-called hydrogen-like ions), the many-body problem cannot be solved without approximation. One of the most common approximations is the basis set approximation where, roughly speaking, one “discretizes” the Hilbert space. Within this approximation, the full configuration interaction (FCI) wave function and energy represent the exact solution of the Schrödinger equation but, unfortunately, the computational cost of this method scales exponentially with system size and so can only be applied to very small systems.

The most basic approximation of quantum chemistry (and usually the starting point of any correlated treatment) is the Hartree-Fock (HF) method [1]. HF is what we call a mean-field method because the electron-electron interaction is treated approximately as an averaged interaction. More specifically, within the HF approximation, each electron “feels” the averaged interaction of the remaining electrons. The computational cost of such mean-field method is cubic with system size and can then be applied to large systems. Importantly, the energetic error associated with HF, *i.e.*, the energy difference between the exact solution of the Schrödinger equation and the HF energy is called the correlation energy [2, 3], a very central quantity in electronic structure theory.

Even though, since the advent of computer [4], one has access to more and more computational power, one can compute the FCI energy only for very small systems. Hence, one has to rely on approximations or alternative ansatz as FCI is usually too expensive. These approximations form a large family of models that range from expensive quantum chemistry methods to inexpensive molecular mechanics methods. Of course, the introduction of approximations induces errors on the properties or/and unphysical artifacts. Over the years, many different basis sets and many methods have been developed and used by theoretical chemists around the world, each of them having their own *pros* and *cons*. One of the main skills of quantum chemists is to be able to choose the right combination of method and basis for a given purpose.

1.2 Excited states

Within a given basis, the FCI Hamiltonian has many roots (or eigenstates), each corresponding to a different electronic state with a well-defined energy. The lowest-energy state is named the ground state while the higher-energy roots are excited states. Though it is already hard to obtain accurate energies and properties for the ground state, it is usually even more challenging for excited states. Hence, the difficulty of finding an appropriate method is really present in the context of excited states. Electronic excited states are especially relevant for photochemical reactions because one must calculate excitation energies defined as the energy difference between two electronic states.

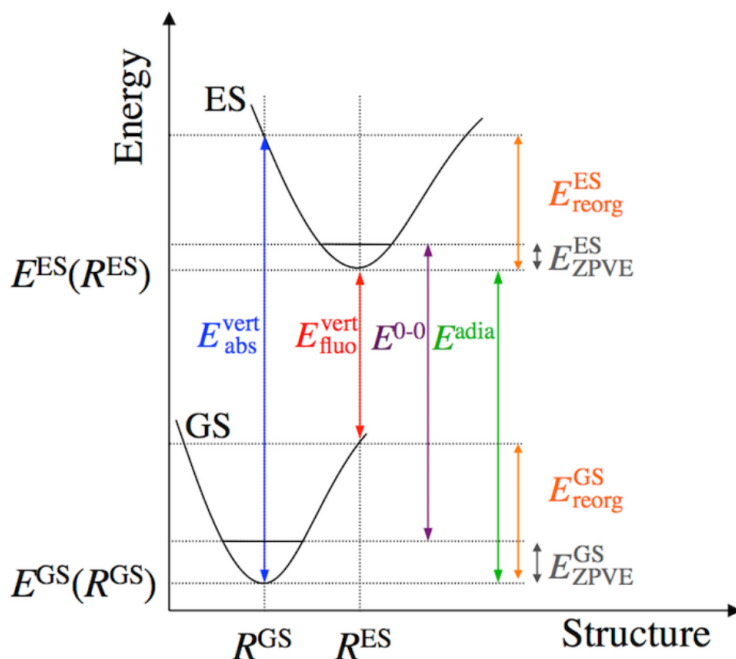


FIGURE 1.1: Representation of the different types of transition energies where E_{abs}^{vert} in red and E_{fluo}^{vert} in blue are respectively the absorption and fluorescence vertical transition energies. E^{adia} in green is the adiabatic transition energy and E^{0-0} in purple is the 0–0 transition energy. The E_{reorg}^{GS} and E_{reorg}^{ES} in orange are respectively the ground state and excited states (geometrical) reorganization energy. All these energies are defined as positive quantities [5].

1.3 Methods for excited states

As shown in Figure 1.1, there are different types of excitation energies. First, we have vertical transitions for which the electronic transition between the two states occurs without geometry relaxation. In absorption and fluorescence, one starts respectively from the geometry of the lowest-energy state (usually the ground state) and from the geometry of a given excited state. In adiabatic transitions, the geometry of the system relaxes to their corresponding minimum of the potential energy surface. In practice, we must therefore optimize not only the geometry of the lowest-energy state, but also the excited-state geometry. This can be a real technical and computational challenge for methods targeting excited states as one must have access to the first derivatives of the energy with respect to the nuclear displacements (Table 1.1). Last but not least, 0–0 energies are adiabatic energies where one takes into account zero-point vibrational energy (ZPVE) corrections for both states. However, the ZPVE correction represents a significant additional computational cost as one has to compute the second derivatives of the energy with respect to the nuclear displacements. Unfortunately, for the most expensive methods, this type of calculations is not possible with the current implementations and machines. Note that, although 0–0 transition energies are much more expensive from a computational point of view, they have the huge benefit to be directly comparable with experimental data (which is not the case of vertical excitation energies as commonly and wrongly done in the literature) [5].

The first mainstream *ab initio* method for excited states was probably configuration interaction with singles (CIS), which has been around since the 1970s [6]. Overall, this method overestimates significantly the energy of the excited states, and it is quite usual to have an error of ~ 1 eV. Because CIS often provides an incorrect energetic ordering of the excited states, it is usually dangerous to use CIS as a qualitative method. Twenty years later, a second-order perturbative correction was added to the CIS and gave birth to CIS(D) [7, 8]. This second-order correction greatly improves the accuracy with typical errors in the range 0.2–0.3 eV. In the early 90s, the complete active space self-consistent field (CASSCF) method [9, 10] and, especially, its CASPT2 extension [11] (where

one computes a second-order perturbative correction to take into account the missing dynamical correlation) were a real breakthrough. Even if it took more than 10 years to obtain analytical gradients [12], CASPT2 was probably the first method that could provide quantitative excitation energies for chemically-interesting molecular systems [13]. However, in its standard settings, it clearly underestimates vertical excitation energies in organic molecules. In the early 2000s, NEVPT2 (which fixes some of the issues of CASPT2) was developed by Angeli, Malrieu and coworkers [14]. CASPT2 and NEVPT2 permit to obtain more accurate results (0.1–0.2 eV) and access double excitations. The advent of time-dependent density-functional theory (TD-DFT) [15, 16] had a significant impact on theoretical chemistry community because it gave the possibility to obtain relatively accurate excitation energies (0.2–0.4 eV) with a (much) lower cost because TD-DFT scales, in its standard implementation, as $\mathcal{O}(N^4)$ (where N is proportional to the system size). It is currently one of the methods of choice for photochemistry though it suffers from many problems associated with, for example, charge-transfer states [17, 18], Rydberg states [19], and double excitations [20–22]. Another method also developed in the 2000s is equation-of-motion coupled cluster with singles and doubles (EOM-CCSD) [23] where it is not unusual to have errors as small as 0.1 eV and 0.2 eV for small and larger compounds, respectively. Another variant (EOM-CCSDT) includes triple excitations but with a higher cost [24]. The second- and third-order approximate coupled cluster methods, CC2 [25] and CC3 [26], where one skips the most expensive terms and avoids the storage of the higher-excitation amplitudes, were developed few years after. The CC3 method is particularly interesting because it provides almost systematically chemically-accurate excitation energies for single transitions with an error of ~ 0.04 eV at a reduced cost ($\mathcal{O}(N^7)$) compared to CCSDT ($\mathcal{O}(N^8)$). The recent and efficient reimplementations of the ADC(2) [27] and ADC(3) [28, 29] Green's function-based methods makes them competitive with coupled-cluster methods, especially ADC(2) that provides satisfactory results (0.1–0.2 eV) with a $\mathcal{O}(N^5)$ scaling only [30–33]. Finally, the recent development of highly parallel selected configuration interaction (SCI) methods (where one only selects the relevant determinants from the FCI space via a predefined criterion) [34, 35] provides near-FCI quality energies [30, 36–38] at a relatively cheap price compared to FCI. However, even if the prefactor of SCI methods is reduced, their overall scaling remains exponential with system size (see Table 1.1).

Everything that has been said in the previous paragraph is only correct in the case of single excitations (*i.e.*, excitations with a dominant contribution from the singly-excited determinants) but one can also encounter double excitations (*i.e.*, excitations with a dominant contribution from the doubly-excited determinants) that can be relevant in particular class of systems, such as polyenes [36]. However, from a theoretical point of view, one cannot strictly classify an excitation as single or double excitation as every transition is a complicated mixture of singles, doubles, triples, and higher excitations. Therefore, it is common to define $\%T_1$, the percentage of singly-excited determinants involved in the transition, to assess the extent to which this excitation can be considered as a genuine single excitation. For small $\%T_1$ values, it has been shown [36] that CASSCF-based models are better than methods based on response formalism (such as CC3). In the latter case, the introduction of quadruple excitations is mandatory to reach high accuracy in order to reach a balanced treatment of the ground and excited states. This drawback of linear-response and equation-of-motion formalisms has been referred to as the *sin of the ground state* by Burke and coworkers [40] due to their natural bias towards the ground state. That explains why CC3 and other similar methods describe correctly single excitations but badly double excitations.

In summary, during the last decades, theoretical chemists have continued to innovate and improve the accuracy of excited-state methods in order to faithfully describe the various types of electronic transitions present in chemical and biochemical systems. One can safely conclude from our discussion above that there is no perfect method for every situation as many factors must be taken into account, such as the type of excitations (single, double, ...) or its physical nature (charge transfer, Rydberg, ...). Therefore, it is challenging to find a systematic protocol to accurately compute excitation energies. Nonetheless, a tentative of such systematic procedure is actually the aim of the QUEST project that we are going to discuss in Chapter 4.

TABLE 1.1: Formal computational scaling of various excited-state methods with respect to the number of one-electron basis functions N and the accessibility of various key properties in popular computational software packages. For organic derivatives, the typical error range for single excitations is also provided as a qualitative indicator of the method accuracy [39].

Method	Formal scaling	Oscillator strength	Analytical gradients	Typical error (eV)
TD-DFT	$\mathcal{O}(N^4)$	✓	✓	0.2–0.4 ^a
BSE@GW	$\mathcal{O}(N^4)$	✓	✗	0.1–0.3 ^b
CIS	$\mathcal{O}(N^5)$	✓	✓	~1.0
CIS(D)	$\mathcal{O}(N^5)$	✗	✓	0.2–0.3
ADC(2)	$\mathcal{O}(N^5)$	✓	✓	0.1–0.2
CC2	$\mathcal{O}(N^5)$	✓	✓	0.1–0.2
ADC(3)	$\mathcal{O}(N^6)$	✓	✗	0.2
EOM-CCSD	$\mathcal{O}(N^6)$	✓	✓	0.1–0.3
CC3	$\mathcal{O}(N^7)$	✓	✗	~0.04
EOM-CCSDT	$\mathcal{O}(N^8)$	✗	✗	~0.03
EOM-CCSDTQ	$\mathcal{O}(N^{10})$	✗	✗	~0.01
CASPT2/NEVPT2	$\mathcal{O}(N!)$	✓	✓	0.1–0.2
SCI	$\mathcal{O}(N!)$	✗	✗	~0.03
FCI	$\mathcal{O}(N!)$	✓	✓	0.0

^a The error range is strongly functional and state dependent. The values reported here are for well-behaved cases;

^b Typical error bar for singlet transitions. Larger errors are often observed for triplet excitations.

TABLE 1.2: Curie CPUs specifications [41, 42]

	Intel® Xeon®					
	E5		E5		Gold	
	2670@2.60 GHz	2680 v2@2.80 GHz	2680 v3@2.50 GHz	2680 v4@2.40 GHz	6130@2.10 GHz	5218@2.30 GHz
Cores	8	10	12	14	16	16
Threads/core	2	2	2	2	2	2
Processor Base Frequency	2.60 GHz	2.80 GHz	2.50 GHz	2.40 GHz	2.10 GHz	2.30 GHz
Max Turbo Frequency	3.30 GHz	3.60 GHz	3.30 GHz	3.30 GHz	3.70 GHz	3.90 GHz
L3 Cache	20 MiB	25 MiB	30 MiB	35 MiB	22 MiB	22 MiB
AVX	AVX	AVX	AVX2	AVX2	AVX-512 (2)	AVX-512 (2)
FMA			FMA3	FMA3	FMA3	FMA3

TABLE 1.3: Curie nodes [41, 42]

Name	Model	Number of nodes	CPUs/node	Flops/node	Memory
xeonv1	Intel® Xeon® E5-2670	14	2	332.8 GFlops	64 GiB
xeonv2	Intel® Xeon® E5-2680 v2	7	2	448 GFlops	64 GiB
xeonv3	Intel® Xeon® E5-2680 v3	9	2	960 GFlops	64 GiB
xeonv4	Intel® Xeon® E5-2680 v4	4	2	1.075 TFlops	128 GiB
xeonv5	Intel® Xeon® Gold 6130	5	2	2.150 TFlops	128 GiB
xeonv6	Intel® Xeon® Gold 5218	1	2	1.075 TFlops	192 GiB

1.4 Computational resources

1.4.1 Curie

We rely on different computational strategies based on the size of the system that is studied and the method used. We have indeed access to several clusters with different computational power. The first cluster we have access to is Curie, the cluster of our laboratory, that is used for “small” calculations. Curie has many types of partitions with many types of processor. Here we will talk only about xeon nodes (see Table 1.3). With the set of Intel® Xeon® CPUs (Central Processing Units) (see Table 1.2) available in Curie, we can see that the evolution of processors is not currently performed by an improvement in terms of frequencies, but rather by an increase in parallelization through multi-core architectures with more and more cores and by a vectorization of calculations through the Advanced Vector Extensions (AVX) and its improvements. Unfortunately our quantum chemistry programs must be adapted to these new architectures by offering massively parallel implementation to our methods [43].

We can calculate theoretically the flops (floating point operations per second) per node using the equation below

$$Flops = n_{core} \times freq \times Flops_{Cycle} \quad (1.1)$$

where n_{core} is the number of processor cores, $freq$ is the processor frequency, and $Flops_{Cycle}$ is the number of double precision operation per cycle. This value depends on the possibility of a CPU to make multiple addition and multiplication operations per cycle for example with FMA and/or with the possibility to vectorize calculation using AVX. With the exception of xeonv6, we have an increase number of Flops/node (see Table 1.3).

1.4.2 Olympe

When we have to perform heavier calculations, we have access to the CALMIP computational resources. In 2018, CALMIP abandoned its 4-year old EOS cluster for a five times more powerful cluster named Olympe. Each CPU node of Olympe has a Skylake, Intel® Xeon® Gold 6140@2.30 GHz (see Table 1.4).

There are two types of nodes on Olympe, some with only CPUs and some hybrids with CPUs and a GPU (Graphics Processing Unit) also named accelerator card because the usage of this type of card is not limited to the graphics related operations but also for HPC (High Performance Computing) as it is the case here (see Table 1.5) We may notice that on Olympe, a calculation

TABLE 1.4: Olympe CPU (Intel® Xeon® Gold 6140@2.30 GHz) specifications [41, 42, 44]

Processor specifications	value
Cores	18
Threads/core (Hyper-threading)	2
Processor Base Frequency	2.30 GHz
Max Turbo Frequency	3.70 GHz
L3 Cache	24.75 MiB
AVX	AVX-512 (2)
FMA	FMA3

TABLE 1.5: Olympe nodes (see Table 1.4) [44]

Name	Number of nodes	CPUs/node	GPUs/node	Flops/node	Memory
Sequana X1120	360	2 ^a		2.26 TFlops	192 GiB
Sequana X1125	12	2 ^a	4 ^b	33.8 TFlops	384 GiB

^a Intel® Xeon® Gold 6140^b NVIDIA® Volta 100 SXM2

accelerated by GPU will be able to be 14 times faster. However, in the same way as for the multi-core architectures, to benefit from the computational power of the CPU/GPU hybrid nodes, one must change the implementation of the codes to adapt it to the GPU architecture. If we compare a CPU node of Olympe and a xeonv5 node which is the best type of Curie CPU node, we can see that there is no significant difference in term of computational power between a xeonv5 Curie node and a Sequana X1120 Olympe node, but on Curie, we have 180 times less xeonv5 nodes that Olympe has Sequana X1120 nodes.

1.4.3 Joliot-Curie

The TGCC Joliot-Curie of GENCI-CEA is another computational resource used for this thesis. This cluster has two Intel® CPU partition one called SKL Irene with a CPU Intel® skylake. Another called KNL Irene with a CPU Intel® Knight Landing. There is also the partition AMD Irene Rome with an AMD Epyc™ Rome processor (see Table 1.6). A CPU/GPU hybrid partition called Irene V100 is also available with an Intel® Cascade Lake CPU and an accelerator card NVIDIA® V100.

If we compare Olympe and Joliot-Curie, we can see that for example for an Olympe Sequana X1120 node and a SKL Irene node from Joliot-Curie that a SKL Irene node is 2 times more powerful than a Sequana X1120 Olympe node. And theoretically if we take all the nodes of each partition then we would have 9 times more computing power with the SKL Irene nodes than with Sequana X1120 nodes.

TABLE 1.6: Joliot-Curie CPU specifications [41, 42, 45]

	Intel® Xeon®		AMD Epyc™
	Platinum8168@2.70 GHz	Phi7250@2.30 GHz	Rome7H12@2.60 GHz
Cores	24	68	64
Threads/core	2	4	2
Processor Base Frequency	2.70 GHz	1.40 GHz	2.6 GHz
Max Turbo Frequency	3.70 GHz	1.60 GHz	3.30 GHz
Cache	33 MiB L3 Cache	32 MiB L2 Cache	250 MiB L3 Cache
AVX	AVX-512 (2)	AVX-512 (2)	AVX2
FMA	FMA3	FMA3	FMA3

TABLE 1.7: Joliot-Curie nodes [46] (see Table 1.6)

Name	Number of nodes	CPUs/node	GPUs/node	Flops/node	Memory
SKL Irene	1 656	2 ^a		4.4 TFlops	192 GiB
KNL Irene (Knight Landing)	828	1 ^b		2.4 TFlops	112 GiB
AMD Irene ROME	2 292	2 ^c		5.12 TFlops	256 GiB
Irene V100	32	2	4 ^d	35.3 TFlops	3 TB

^a Intel® Xeon® Platinum 8168@2.70 GHz

^b Intel® Xeon Phi™ 7250

^c AMD Epyc™ Rome@2.60 GHz

^d NVIDIA® V100

1.5 Structure of the manuscript

The main purpose of this thesis is the accurate calculation of excited states with various methods (ranging from relatively cheap Green’s function methods like GW to very expensive SCI methods) in order to create a database of excitation energies that allows to evaluate the accuracy of a given method thanks to the generation of a large number of statistical quantities. This database has been designed to be the most diverse possible and is divided into several subsets gathering from small, medium and large molecules, double excitations, singlet, doublet, and triplet transitions, more “exotic” molecules, etc. We are also interested in the accurate calculation of ground-state energies for cyclic molecules like furan or benzene. Another subject addressed in my thesis is the development of web tools for quantum chemistry. Indeed we have developed two web tools. The first one is a web terminal that offers to potential future users of the quantum chemistry software QUANTUM PACKAGE the possibility to test it without the burden of the installation. The second is a website created for the QUEST project in order to perform cross-comparisons between quantum chemistry methods thanks to statistics but also simply to gather the huge amount of data generated by this project.

More specifically this manuscript contains five parts. Chapter 2 introduces the methodology and the some of the mainstream methods of quantum chemistry. First, of course, we start with the Schrödinger equation and then the Hartree-Fock approximation. After that, we talk about the main methods used in this manuscript like CI. We describe CI methods in their full and truncated versions but also its selected variant called CIPSI which consists in selecting iteratively the most important determinants, allowing to move towards the FCI limit as quickly as possible. This method is very important for the third, fourth and fifth parts of the manuscript. We will also discuss about coupled cluster, another quantum chemistry method that relies on an exponential ansatz for the many-body wave function. This type of methods is considered in the fourth part of the manuscript.

In Chapter 3, we will take time to introduce GW approximations, a family of methods that can be applied as post-Kohn-Sham or post-Hartree-Fock corrections and that rely on the one-body Green’s function G . Electron correlation is explicitly incorporated via a sequence of self-consistent steps connected by Hedin’s equations [47] that transform an unsolvable many-electron problem into a set of non-linear one-electron equations thanks to the introduction of an effective potential Σ named the self-energy. Hedin’s approach uses a screened Coulomb interaction W instead of the standard bare Coulomb interaction. The GW methods are central in the second part of the manuscript, which is concerned with (unphysical) discontinuities in GW methods, a consequence of the mathematical structure of Σ . In this part, we describe our study of these discontinuities and the influence of the type of GW methods that one employs.

In Chapter 4, we describe the QUEST project which concerns transition energies and excited states in molecular systems. The aim of this research project is to create reference absorption and fluorescence energies with computational chemistry tools, leading to chemically accurate vertical transition energies for small-, medium- and large-sized molecules. To do this, we use a composite protocol that combines SCI and high-level coupled cluster calculations. Having at hand these reference energies, we then compare the respective accuracy of lower-level methods using various

statistical tools and quantities. All of these data are gathered in a website specifically designed for the present project that allows to view the data of the QUEST project and generate on-the-fly statistics to allow the user to compare methods specifically for his/her own purposes.

Chapter 5 describes the quantum chemistry software called QUANTUM PACKAGE developed in our group and our work to offer an interactive demo of this software directly in the web browser. QUANTUM PACKAGE is a determinant-driven open-source programming environment for quantum chemistry (and beyond) specifically designed for wave function methods. The flagship method of QUANTUM PACKAGE is the SCI algorithm named CIPSI (Configuration Interaction using a Perturbative Selection made Iteratively). The philosophy of QUANTUM PACKAGE is to create an easy-to-develop quantum chemistry software (using the IRPF90 preprocessor and a plug-in system) and easy to use (using QUANTUM PACKAGE SHELL). We use a LINUX container with an expiration time as a sandbox where the user can test QUANTUM PACKAGE. In Chapter 6, pursuing the recent work of our team on the benzene molecule, we undertook to compute accurate (with an error smaller than $1 mE_h$) FCI frozen-core correlation energy estimates for the twelve five- and six-membered ring molecules present in the QUEST project in the cc-pVDZ basis set. Our near-FCI estimates are based on energetically optimized-orbital CIPSI calculations. Based on these valuable reference correlation energies, we benchmarked several well-known families of methods and study their convergence properties.

Finally, in Chapter 7, we draw our conclusions and discuss prospects.

Chapter 2

Quantum chemistry methods

In this chapter we provide details about the main concepts and methods used in the present thesis.

2.1 Atomic units

Generally, in macroscopic physics, we use the international system of units (also called SI). In quantum mechanics, however, it is more convenient to use another unit system named atomic units (a.u.), where some constants are arbitrary set to 1 and form the base units (see Table 2.1). Widely-used atomic units are the Bohr radius (a_0) for the length and Hartree (E_h) for the energy:

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} = \frac{1}{k_e} \frac{\hbar^2}{m_e e^2} \quad (2.1)$$

$$E_h = \frac{\hbar^2}{m_e a_0^2} \quad (2.2)$$

Throughout this thesis, we consistently used atomic units unless otherwise stated.

2.2 Schrödinger equation

The Schrödinger equation, where one applies the Hamiltonian operator \hat{H} to the wave function Ψ , has two different forms, time-dependent and time-independent, that are used respectively for non-stationary and stationary states. The wave function depends on the vector $\mathbf{X} = (\mathbf{R}, \mathbf{S}) = (x_1, \dots, x_p)$ which is a composite coordinate vector gathering the spin coordinates $\mathbf{S} = (s_1, \dots, s_p)$ and the space coordinates $\mathbf{R} = (r_1, \dots, r_p)$ with p the number of particles.

In the case of the time-dependent Schrödinger equation, Ψ depends on the time t as additional argument

$$i \frac{\partial \Psi(\mathbf{X}, t)}{\partial t} = \hat{H} \Psi(\mathbf{X}, t). \quad (2.3)$$

The time-independent expression is an eigenvalue equation with the energy E as eigenvalue and with its corresponding wave function as eigenvector:

$$\hat{H} \Psi(\mathbf{X}) = E \Psi(\mathbf{X}). \quad (2.4)$$

TABLE 2.1: Base units of the atomic unit system

Name	Symbol	Value in SI
Electron rest mass	m_e	$9.109\,383\,56 \times 10^{-31} \text{ kg}$
Elementary charge	e	$1.602\,176\,565 \times 10^{-19} \text{ C}$
Reduced Planck's constant	$\hbar = h/(2\pi)$	$1.054\,571\,726 \times 10^{-34} \text{ J s}$
Coulomb force constant	$k_e = 1/(4\pi\epsilon_0)$	$8.987\,551\,787\,368\,1 \times 10^9 \text{ kg m}^2 \text{ s}^{-2} \text{ C}^{-2}$

The Hamiltonian operator \hat{H} is the operator that provides, via its expectation value, the energy of the system. In non-relativistic quantum mechanics, the molecular Hamiltonian reads

$$\begin{aligned}\hat{H} &= \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{T}_n + \hat{V}_{nn} \\ &= -\sum_i^N \frac{\nabla_i^2}{2} - \sum_i^N \sum_A^M \frac{Z_A}{r_{Ai}} + \sum_{i<j}^N \frac{1}{r_{ij}} - \sum_A^M \frac{\nabla_A^2}{2} + \sum_{A<B}^M \frac{Z_A Z_B}{r_{AB}},\end{aligned}\quad (2.5)$$

where r_{pq} is the distance between the particles p and q , N is the number of electrons, M is the number of nuclei, A and B are two nuclei and i and j are two electrons. Thus, the total energy corresponds to the sum of the kinetic energy of the electrons (\hat{T}_e), the kinetic energy of the nuclei (\hat{T}_n), the energy associated with the Coulomb attraction between electrons and nuclei (\hat{V}_{en}), the Coulomb repulsion between electrons (\hat{V}_{ee}), and the Coulomb repulsion between nuclei (\hat{V}_{nn}).

2.2.1 Born-Oppenheimer approximation

For any nucleus n we have $m_e \ll m_n$, therefore, one can assume that the nuclei are fixed (at least compared to the electrons). Hence, one can decouple the motions of electrons and nuclei and, consequently, separate the Hamiltonian operator as the sum of a nuclear Hamiltonian \hat{H}_n and an electronic Hamiltonian \hat{H}_e , such that

$$\hat{H} = \hat{H}_e + \hat{H}_n, \quad (2.6)$$

with

$$\hat{H}_e = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee}, \quad (2.7a)$$

$$\hat{H}_n = \hat{T}_n + \hat{V}_{nn}. \quad (2.7b)$$

Within the so-called Born-Oppenheimer approximation, the electronic Hamiltonian \hat{H}_e reads explicitly

$$\hat{H}_e = -\sum_i^N \frac{\nabla_i^2}{2} - \sum_i^N \sum_A^M \frac{Z_A}{r_{Ai}} + \sum_{i<j}^N \frac{1}{r_{ij}}. \quad (2.8)$$

2.3 The Hartree-Fock approximation

Within the Hartree-Fock (HF) approximation, the many-electron wave function is defined as a Slater determinant

$$\Psi_{\text{HF}}(\mathbf{X}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \dots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \dots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}, \quad (2.9)$$

made of N one-electron wave functions known as spinorbitals

$$\psi_i(\mathbf{x}) = \phi_i(\mathbf{r})\sigma(\mathbf{s}), \quad (2.10)$$

where $\sigma(\mathbf{s})$ is its spin part

$$\sigma(\mathbf{s}) = \begin{cases} \alpha(\mathbf{s}), & \text{for spin-up electrons,} \\ \beta(\mathbf{s}), & \text{for spin-down electrons,} \end{cases} \quad (2.11)$$

and $\phi_i(\mathbf{r})$ is its spatial part. A key property of the the spinorbitals is that they form an orthonormal basis, *i.e.*,

$$\langle \psi_i | \psi_j \rangle = \delta_{ij}, \quad (2.12)$$

where δ_{ij} is the Kronecker delta defined as

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \quad (2.13)$$

Now, one can define the HF energy, E_{HF} , as the expectation value of the Hamiltonian operator with respect to Ψ_{HF} :

$$E_{\text{HF}} = \frac{\langle \Psi_{\text{HF}} | \hat{H} | \Psi_{\text{HF}} \rangle}{\langle \Psi_{\text{HF}} | \Psi_{\text{HF}} \rangle}. \quad (2.14)$$

In the spinorbital basis, the HF energy reads

$$E_{\text{HF}} = \sum_i^N \langle \psi_i(\mathbf{x}_1) | \hat{H}^c | \psi_i(\mathbf{x}_1) \rangle + \sum_{i < j}^N \left[\langle \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) | \frac{1}{r_{12}} | \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) \rangle - \langle \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) | \frac{1}{r_{12}} | \psi_j(\mathbf{x}_1) \psi_i(\mathbf{x}_2) \rangle \right]. \quad (2.15)$$

where one usually defines three operators:

- The one-electron core Hamiltonian

$$\hat{H}^c = \hat{T}_e + \hat{V}_{en}. \quad (2.16)$$

- The two-electron Coulomb operator

$$\hat{\mathcal{J}}_i(\mathbf{x}_1) \psi_j(\mathbf{x}_1) = \psi_j(\mathbf{x}_1) \int \psi_i(\mathbf{x}_2) \frac{1}{r_{12}} \psi_i(\mathbf{x}_2) d\mathbf{x}_2. \quad (2.17)$$

- The two-electron exchange operator

$$\hat{\mathcal{K}}_i(\mathbf{x}_1) \psi_j(\mathbf{x}_1) = \psi_i(\mathbf{x}_1) \int \psi_i(\mathbf{x}_2) \frac{1}{r_{12}} \psi_j(\mathbf{x}_2) d\mathbf{x}_2. \quad (2.18)$$

Moreover, the stationarity of the HF energy with respect to the spinorbitals allows us to define the Fock operator $\hat{\mathcal{F}}$ as

$$\hat{\mathcal{F}} \psi_i(\mathbf{x}_1) = \varepsilon_i \psi_i(\mathbf{x}_1), \quad (2.19)$$

$$\hat{\mathcal{F}}(\mathbf{x}_1) = \hat{H}^c(\mathbf{x}_1) + \sum_i^N [\hat{\mathcal{J}}_i(\mathbf{x}_1) - \hat{\mathcal{K}}_i(\mathbf{x}_1)], \quad (2.20)$$

the spinorbitals $\psi_i(\mathbf{x})$ being defined as the eigenfunctions of $\hat{\mathcal{F}}$ with ε_i their corresponding one-electron energies.

2.3.1 Restricted Hartree-Fock

Within the restricted Hartree-Fock (RHF) approximation that applies to closed-shell systems (*i.e.*, a system where each spatial orbital is doubly-occupied by a pair of electrons), we assume that the spatial part of the spinorbitals is independent to the spin of the electrons:

$$\hat{F} \phi_i(\mathbf{r}_1) = \varepsilon_i \phi_i(\mathbf{r}_1), \quad (2.21)$$

where, in this case, the Fock operator becomes

$$\hat{F} = \hat{H}^c + \sum_i^{N/2} [2\hat{J}_i - \hat{K}_i], \quad (2.22)$$

with

- The Coulomb operator

$$\hat{J}_i(\mathbf{r}_1)\phi_j(\mathbf{r}_1) = \phi_j(\mathbf{r}_1) \int \phi_i(\mathbf{r}_2) \frac{1}{r_{12}} \phi_i(\mathbf{r}_2) d\mathbf{r}_2. \quad (2.23)$$

- The exchange operator

$$\hat{K}_i(\mathbf{r}_1)\phi_j(\mathbf{r}_1) = \phi_i(\mathbf{r}_1) \int \phi_i(\mathbf{r}_2) \frac{1}{r_{12}} \phi_j(\mathbf{r}_2) d\mathbf{r}_2. \quad (2.24)$$

2.3.2 Roothaan-Hall equations

To practically solve the HF equations, one must use the Roothaan-Hall equations where one starts by expanding the spatial orbitals [also known as molecular orbitals (MOs)] as linear combinations of basis functions $\chi_\mu(\mathbf{r})$ also known as atomic orbitals (AOs)

$$\phi_i(\mathbf{r}) = \sum_\mu^K c_{\mu i} \chi_\mu(\mathbf{r}).$$

In the AO basis, we have

$$F_{\mu\nu} = \langle \chi_\mu | \hat{F} | \chi_\nu \rangle \equiv \langle \mu | \hat{F} | \nu \rangle = H_{\mu\nu}^c + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\lambda | \nu\sigma \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \right], \quad (2.25)$$

where

$$H_{\mu\nu}^c = \langle \mu | \hat{H}^c | \nu \rangle \quad (2.26)$$

are the elements of the core Hamiltonian in the AO basis and

$$\langle \mu\lambda | \nu\sigma \rangle = (\mu\nu | \lambda\sigma) = \iint \chi_\mu(\mathbf{r}_1) \chi_\lambda(\mathbf{r}_2) \frac{1}{r_{12}} \chi_\nu(\mathbf{r}_1) \chi_\sigma(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.27)$$

are the two-electron integrals. Moreover, the density matrix is defined, for a closed-shell system, as

$$P_{\mu\nu} = 2 \sum_i^{N/2} c_{\mu i} c_{\nu i}. \quad (2.28)$$

The electronic HF energy of the system is then given by

$$E_{\text{HF}} = \sum_{\mu\nu}^K P_{\mu\nu} H_{\mu\nu}^c + \frac{1}{2} \sum_{\mu\nu\lambda\sigma}^K P_{\mu\nu} P_{\lambda\sigma} G_{\mu\nu\lambda\sigma}. \quad (2.29)$$

In matrix form, the Fock matrix F can be decomposed as

$$F = H^c + G, \quad (2.30)$$

where

$$G_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} G_{\mu\nu\lambda\sigma}, \quad (2.31)$$

$$G_{\mu\nu\lambda\sigma} = \langle \mu\lambda | \nu\sigma \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle. \quad (2.32)$$

is the two-electron part of the Fock operator.

This yields the Roothaan-Hall equations:

$$\sum_{\nu} F_{\mu\nu} c_{\nu i} = \sum_{\nu} S_{\mu\nu} c_{\nu i} \varepsilon_i, \quad (2.33)$$

or, in matrix form,

$$FC = SC\varepsilon \quad (2.34)$$

where the elements of the overlap matrix S are given by

$$S_{\mu\nu} = \langle \mu | \nu \rangle. \quad (2.35)$$

and C is the matrix of the MO coefficients and ε is a diagonal matrix gathering the MO eigenvalues. To switch between the non-orthogonal AO basis to an orthogonal basis, we define the orthogonalization matrix X such as

$$X^{\dagger}SX = I, \quad (2.36)$$

where I is the identity matrix. Two common methods are available to compute X :

- The Löwdin orthogonalization

$$X = S^{-1/2}, \quad (2.37)$$

- The canonical orthogonalization

$$X = Us^{-1/2}, \quad (2.38)$$

where U and s are respectively the eigenvectors and eigenvalues of the overlap matrix S .

Rotating the Fock matrix F into the orthogonal basis yields

$$F'C' = C'\varepsilon, \quad (2.39)$$

where

$$F' = X^{\dagger}FX. \quad (2.40)$$

The matrices C' and ε can be determined by the diagonalization of Eq. (2.39), and the matrix C is obtained by back-transforming the eigenvectors in the original basis

$$C = XC'. \quad (2.41)$$

2.3.3 The self-consistent field procedure

In order to obtain the MO coefficients C , we must diagonalize the Fock matrix F . But this matrix does depend on the MO coefficients itself. Therefore, one must employ an iterative procedure called self-consistent field (SCF) method. Before the procedure we need to define a convergence criterion. An appropriate choice for this purpose is the following error vector:

$$e = FPS - SPF. \quad (2.42)$$

which is known to be equal to zero at convergence as the Fock and density matrices must share a common set of eigenvectors. Then, in practice, at convergence, we have

$$\max |e| < \tau, \quad (2.43)$$

with τ a small value typically 10^{-6} .

The SCF algorithm is described below

1. Obtain an estimate of the density matrix P (from an extended Hückel calculation for example).
2. Build the Fock matrix: $F = H^c + G$.
3. Transform the Fock matrix in the orthogonal matrix: $F' = X^\dagger F X$.
4. Diagonalize F' to obtain C' and ε .
5. Back-transform the MOs in the original basis: $C = X C'$.
6. Compute the new density matrix $P = C C^\dagger$, as well as the HF energy:

$$E_{\text{HF}} = \frac{1}{2} \text{Tr}\{P(H^c + F)\}. \quad (2.44)$$

7. Convergence test using the error vector

$$\max |e| < \tau \quad (2.45)$$

8. If not satisfied, go to 2, otherwise, exit SCF procedure.

Even if the HF approximation can be seen as a useful first step, unfortunately, it cannot be used to obtain the exact energy due to its mean-field character which approximates the electron-electron interaction via an averaging over all the electrons of the system (*i.e.*, a given electron “feels” only the averaged repulsion of the $N - 1$ remaining electrons). As we shall see in the next section, one must go beyond the HF approximation in order to catch what is known as the correlation energy.

2.4 Post-Hartree-Fock methods

We define the correlation energy as the difference between the (non-relativistic) exact energy and the HF energy

$$E_c = E - E_{\text{HF}} \quad (2.46)$$

Moreover, we define Ψ_0 as the reference determinant, which may or may not be the HF ground-state determinant Ψ_{HF} . The aim of post-HF methods is to catch as much correlation energy as possible.

2.4.1 Configuration interaction method

In configuration interaction (CI) methods, one defines the correlated wave function as a linear combination of Slater determinants

$$|\Psi_{\text{CI}}\rangle = c_0 \Psi_0 + \sum_i^{\text{occ}} \sum_a^{\text{virt}} c_i^a |\Psi_i^a\rangle + \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} c_{ij}^{ab} |\Psi_{ij}^{ab}\rangle + \sum_{ijk}^{\text{occ}} \sum_{abc}^{\text{virt}} c_{ijk}^{abc} |\Psi_{ijk}^{abc}\rangle + \dots \quad (2.47)$$

where occ is the number of occupied orbitals, virt the number of virtual (*i.e.*, unoccupied) orbitals, and Ψ_i^a corresponds to a singly-excited determinant where an electron from the occupied orbital i has been excited to the vacant orbital a . Likewise, Ψ_{ij}^{ab} is a doubly-excited determinants where electrons from the occupied orbitals i and j have been excited to the vacant orbitals a and b . (Same rule applies for the higher excitations.) From here on, i, j, \dots are occupied spinorbitals, $a, b,$

... denote virtual (unoccupied) spinorbitals, and $p, q, r,$ and s indicate arbitrary (orthonormal) spinorbitals. If one considers all the possible excited determinants constructed by placing N electrons in $2K$ spinorbitals, one talks about full configuration interaction (FCI) which corresponds to the exact solution (hence energy and wave function) of the system for a given set of basis functions. Conversely, if one only considers all single excitations, one obtains CI with singles (CIS), a well-known excited-state method, while if one considers all single and double excitations, one obtains CI with singles and doubles (CISD).

At the CI level, the energy is computed as

$$E_{\text{CI}} = \frac{\langle \Psi_{\text{CI}} | \hat{H} | \Psi_{\text{CI}} \rangle}{\langle \Psi_{\text{CI}} | \Psi_{\text{CI}} \rangle} \quad (2.48)$$

and is an upper bound to the exact energy of the system thanks to the variational property of the above formula. Even if it recovers all the correlation energy, FCI is, in practice, very expensive and is limited to small Hilbert space.

2.4.2 Selected CI methods

In conventional CI methods (see above), one chooses the determinants to include in the so-called variational space based on their maximum excitation degree. However, within this predefined subspace of determinants, it is well known that only a small number of them significantly contributes to the total energy. Hence, the fundamental idea behind selected CI (SCI) methods is to select the determinants among the entire set of determinants (*i.e.*, the FCI space) based on their estimated contribution to the FCI energy and/or wave function. Although the SCI family of methods has many members, below, we only discuss the CIPSI (Configuration Interaction using a Perturbative Selection made Iteratively) algorithm [34] that is implemented in QUANTUM PACKAGE, a quantum chemistry software developed in our laboratory. However, the reader should keep in mind that the idea behind the CIPSI method is very generally and that SCI methods usually differ by the definition of the selection criterion.

In the CIPSI iterative process (which is described below) [34], a (multireference) second-order perturbation correction is employed to select the determinants. In practice, we start with a small set of determinants (typically the HF determinant or a determinant made of natural orbitals but a small set of determinants can also be chosen). Then, the following steps are iteratively performed:

1. Compute the variational wave function $|\Psi_{\text{var}}\rangle$ and its energy E_{var} for a given set of determinants $\{|I\rangle\}$ that defines, at a given iteration, the variational space:

$$|\Psi_{\text{var}}\rangle = \sum_I c_I |I\rangle \quad (2.49)$$

$$E_{\text{var}} = \frac{\langle \Psi_{\text{var}} | \hat{H} | \Psi_{\text{var}} \rangle}{\langle \Psi_{\text{var}} | \Psi_{\text{var}} \rangle} \geq E_{\text{FCI}} \quad (2.50)$$

2. For each determinant $|\alpha\rangle$ belonging to the external space, *i.e.*, $\langle \Psi_{\text{var}} | \hat{H} | \alpha \rangle \neq 0$, compute their (individual) second-order perturbative contribution:

$$e_\alpha = \frac{\langle \Psi_{\text{var}} | \hat{H} | \alpha \rangle^2}{E_{\text{var}} - \langle \alpha | \hat{H} | \alpha \rangle} \quad (2.51)$$

3. Estimate the missing correlation energy via the computation of the (total) second-order perturbative correction as follows:

$$E_{\text{PT2}} = \sum_\alpha e_\alpha \quad (2.52)$$

The sum of the variational energy and the PT2 correction is an approximation of the FCI energy, *i.e.*, $E_{\text{FCI}} \approx E_{\text{var}} + E_{\text{PT2}}$

4. Select a subset of external determinants $\{|\alpha\rangle^*\}$ with the largest contribution e_α , and add them to the variational space for the next iteration, *i.e.*,

$$\{I\} \leftarrow \{I\} \cup \{|\alpha\rangle^*\} \quad (2.53)$$

5. If convergence has not been reached, go back to 1

2.4.3 Coupled-cluster methods

The coupled cluster (CC) family of methods [3, 48–51] is one of the most successful wave function approaches for the description of chemical systems [52–54]. In particular, low-order truncated CC methods, such as CC with singles, doubles and perturbative triples CCSD(T) [52, 55], properly describe weak correlation, while inclusion of higher-order excitations is required for strongly correlated systems.

In CC theory, one represents the wave function as follows

$$\Psi_{\text{CC}} = e^{\hat{T}} \Psi_0. \quad (2.54)$$

where the cluster operator is

$$\hat{T} = \sum_k^N \hat{T}_k, \quad (2.55)$$

with

$$\begin{aligned} \hat{T}_1 &= \sum_i^{\text{occ}} \sum_a^{\text{virt}} t_i^a \hat{a}_i \hat{a}_a^\dagger \\ \hat{T}_2 &= \frac{1}{4} \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} t_{ij}^{ab} \hat{a}_i \hat{a}_j \hat{a}_a^\dagger \hat{a}_b^\dagger \\ &\vdots \end{aligned} \quad (2.56)$$

Here, \hat{a}_a^\dagger and \hat{a}_i are the usual annihilation and creation operators which annihilates an electron in the occupied spinorbital i and creates an electron in the vacant spinorbital a , respectively.

By performing a Taylor expansion of the exponentiated operator

$$e^{\hat{T}} = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!} = \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \frac{\hat{T}^4}{4!} + \dots \quad (2.57)$$

the CC wave function can be rewritten as

$$\begin{aligned} \Psi_{\text{CC}} = \Psi_0 \left[\hat{1} + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left(\hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{6} \hat{T}_1^3 \right) \right. \\ \left. + \left(\hat{T}_4 + \hat{T}_3 \hat{T}_1 + \frac{1}{2} \hat{T}_2^2 + \frac{1}{2} \hat{T}_2 \hat{T}_1^2 + \frac{1}{24} \hat{T}_1^4 \right) + \dots \right], \end{aligned} \quad (2.58)$$

where $\hat{1}$ is the identity operator. The first term in a parenthesis (of form $\hat{T}_2, \hat{T}_3 \dots$) generate the connected excitations while the others ($\hat{T}_1^2, \hat{T}_2 \hat{T}_1$) generate the disconnected excitations, *i.e.*, the excitations that are the product of two or more lower-order excitations. So physically, a connected type such as \hat{T}_4 corresponds to four electrons interacting simultaneous, while a disconnected term such as \hat{T}_2^2 corresponds to two non-interacting pairs of interacting electrons. We can now apply the

Schrödinger (2.4) to the CC wave function,

$$\hat{H}|\Psi_{CC}\rangle = E|\Psi_{CC}\rangle, \quad (2.59a)$$

$$\hat{H}e^{\hat{T}}|\Psi_0\rangle = Ee^{\hat{T}}|\Psi_0\rangle. \quad (2.59b)$$

At this point we could evaluate the energy as an expectation value of the CC wave function,

$$E_{CC}^{\text{var}} = \frac{\langle\Psi_{CC}|\hat{H}|\Psi_{CC}\rangle}{\langle\Psi_{CC}|\Psi_{CC}\rangle} = \frac{\langle e^{\hat{T}}\Psi_0|\hat{H}|e^{\hat{T}}\Psi_0\rangle}{\langle e^{\hat{T}}\Psi_0|e^{\hat{T}}\Psi_0\rangle}, \quad (2.60)$$

where by expanding $e^{\hat{T}}$, we get

$$E_{CC}^{\text{var}} = \frac{\langle[\hat{1} + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + \dots] \Psi_0 | \hat{H} | [\hat{1} + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + \dots] \Psi_0\rangle}{\langle[\hat{1} + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + \dots] \Psi_0 | [\hat{1} + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + \dots] \Psi_0\rangle}. \quad (2.61)$$

Unfortunately this expansion leads to a series that does not truncate before the number of electrons. The standard formulation of coupled cluster theory instead proceeds by projecting the coupled cluster Schrödinger equation (2.59) onto the reference wave function. Thus, multiplying to the left by $\langle\Psi_0|$ and integrating, one gets

$$\langle\Psi_0|\hat{H}e^{\hat{T}}|\Psi_0\rangle = E_{CC} \langle\Psi_0|e^{\hat{T}}\Psi_0\rangle, \quad (2.62)$$

$$\langle\Psi_0|\hat{H}e^{\hat{T}}|\Psi_0\rangle = E_{CC} \langle\Psi_0|[\hat{1} + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + \dots] \Psi_0\rangle, \quad (2.63)$$

$$E_{CC} = \langle\Psi_0|\hat{H}e^{\hat{T}}|\Psi_0\rangle. \quad (2.64)$$

Because the Hamiltonian operator given by Eq. (2.5) is a two-electron operator, we can limit ourselves to the first two levels of excitation, *i.e.*,

$$E_{CC} = \langle\Psi_0|\hat{H}\left(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right)|\Psi_0\rangle, \quad (2.65)$$

$$E_{CC} = \langle\Psi_0|\hat{H}|\Psi_0\rangle + \langle\Psi_0|\hat{H}|\hat{T}_1\Psi_0\rangle + \langle\Psi_0|\hat{H}|\hat{T}_2\Psi_0\rangle + \frac{1}{2}\langle\Psi_0|\hat{H}|\hat{T}_1^2\Psi_0\rangle, \quad (2.66)$$

$$E_{CC} = E_0 + \sum_i^{\text{occ}} \sum_a^{\text{virt}} t_i^a \langle\Psi_0|\hat{H}|\Psi_i^a\rangle + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} (t_{ij}^{ab} + t_i^a t_j^b - t_j^a t_i^b) \langle\Psi_0|\hat{H}|\Psi_{ij}^{ab}\rangle. \quad (2.67)$$

Thanks to Brillouin's theorem [1] that states that a singly-excited determinant $|\Psi_i^a\rangle$ does not interact directly with a reference Hartree-Fock determinant $|\Psi_0\rangle$ (*i.e.*, $\langle\Psi_0|\hat{H}|\Psi_i^a\rangle = 0$), Eq. (2.67) becomes

$$E_{CC} = E_{\text{HF}} + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} (t_{ij}^{ab} + t_i^a t_j^b - t_j^a t_i^b) \langle ij|ab\rangle. \quad (2.68)$$

From the previous equation, one can see that the CC correlation energy is determined completely by the singles and doubles amplitudes and the two-electron integrals (in the MO basis). To get the amplitudes, one must project the Schrödinger equation onto the space of the excited determinants. This can be done via a similarity transformation of the Hamiltonian operator $e^{-\hat{T}}\hat{H}e^{\hat{T}}$, where, by analogy with the excitation operator $e^{\hat{T}}$ we have defined the deexcitation operator $e^{-\hat{T}}$. More explicitly, this yields

$$e^{-\hat{T}}\hat{H}e^{\hat{T}}\Psi_0 = E_{CC}\Psi_0, \quad (2.69a)$$

$$E_{CC} = \langle\Psi_0|e^{-\hat{T}}\hat{H}e^{\hat{T}}|\Psi_0\rangle. \quad (2.69b)$$

which can be considered as the expectation value of a similarity transformed Hamiltonian. Following the same strategy, the so-called amplitude equations read

$$\begin{aligned}\langle \Psi_i^j | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0 \rangle &= 0 \\ \langle \Psi_{ij}^{ab} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0 \rangle &= 0 \\ \langle \Psi_{ijk}^{abc} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0 \rangle &= 0 \\ &\vdots\end{aligned}\tag{2.70}$$

The full coupled cluster (which is equivalent to FCI) is, of course, too expensive. Consequently, one generally truncates the coupled cluster amplitudes to a specific excitation level. For example, for CCSD (coupled cluster with singles and doubles), we have $\hat{T} = \hat{T}_1 + \hat{T}_2$ and

$$\begin{aligned}e^{\hat{T}} &= e^{\hat{T}_1 + \hat{T}_2}, \\ &= 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left(\hat{T}_1 \hat{T}_2 + \frac{1}{6} \hat{T}_1^3 \right) + \left(\frac{1}{2} \hat{T}_2^2 + \frac{1}{2} \hat{T}_2 \hat{T}_1^2 + \frac{1}{24} \hat{T}_1^4 \right) + \dots\end{aligned}\tag{2.71}$$

We can now compare coupled cluster and configuration interaction. Truncating the CI and CC ansatz to singles and doubles, we have from Eqs. (2.47) and (2.54)

$$\text{CI: } \hat{T} = 1 + \hat{T}_1 + \hat{T}_2,\tag{2.72}$$

$$\text{CC: } e^{\hat{T}} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left(\hat{T}_1 \hat{T}_2 + \frac{1}{6} \hat{T}_1^3 \right) + \left(\frac{1}{2} \hat{T}_2^2 + \frac{1}{2} \hat{T}_2 \hat{T}_1^2 + \frac{1}{24} \hat{T}_1^4 \right) + \dots\tag{2.73}$$

which clearly evidences that the CC ansatz produces parts of the triple, quadruple, and higher excitations as disconnected terms of the form $\frac{1}{2} \hat{T}_2^2$ or $\hat{T}_2 \hat{T}_1$. These terms are, of course, not present in its CI counterpart and is the cause of the non-size-extensivity of truncated CI methods.

2.4.4 Møller-Plesset perturbative methods

In Rayleigh-Schrödinger perturbative theory, one expands both the energy and the wave function as Taylor series with respect to the coupling parameter λ following the present partition of the Hamiltonian:

$$\hat{H} = \hat{H}_0 + \lambda \hat{V}\tag{2.74}$$

where \hat{H}_0 is the zeroth-order (reference) Hamiltonian and \hat{V} is the so-called perturbation. In the case of Møller-Plesset perturbation theory, we have

$$\hat{H}_0 = \sum_i^N \hat{\mathcal{F}}(i)\tag{2.75}$$

which gives, in the case of second-order Møller-Plesset perturbation theory (MP2), the well-known formulas

$$E^{(2)} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{virt}} \frac{\left| \langle \Psi_0 | \sum_{k < l}^n \frac{1}{r_{kl}} | \Psi_{ij}^{ab} \rangle \right|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} = \frac{1}{4} \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \langle ij || ab \rangle t_{ij}^{ab}\tag{2.76}$$

$$t_{ij}^{ab} = \frac{\langle ij || ab \rangle}{\Delta_{ij}^{ab}}\tag{2.77}$$

$$\Delta_{ij}^{ab} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b\tag{2.78}$$

Chapter 3

Unphysical discontinuities in GW methods

This chapter is based on

M. V eril et al., "Unphysical discontinuities in GW methods", *J. Chem. Theory. Comput.* **14**, 5220–5228 (2018)

3.1 Many-body Green's function methods

Many-body perturbation theory methods based on the one-body Green's function G are fascinating as they are able to transform an unsolvable many-electron problem into a set of non-linear one-electron equations, thanks to the introduction of an effective potential Σ , the self-energy. Electron correlation is explicitly incorporated via a sequence of self-consistent steps known as Hedin's equations [47], which connect G , the irreducible vertex function Γ , the irreducible polarizability P , the dynamically-screened Coulomb interaction W , and Σ through a set of five integro-differential equations (see Figure 3.1):

$$G(12) = G_H(12) + \int G_H(13)\Sigma(34)G(42)d(34), \quad (3.1a)$$

$$\Gamma(123) = \delta(12)\delta(13) + \int \frac{\delta\Sigma(12)}{\delta G(45)}G(46)G(75)\Gamma(673)d(4567), \quad (3.1b)$$

$$P(12) = -i \int G(13)\Gamma(324)G(41)d(34), \quad (3.1c)$$

$$W(12) = v(12) + \int v(13)P(34)W(42)d(34), \quad (3.1d)$$

$$\Sigma(12) = i \int G(13)W(14)\Gamma(324)d(34), \quad (3.1e)$$

where G_H is the one-body Hartree Green's function, v is the bare Coulomb interaction, $\delta(12)$ is Dirac's delta function and (1) is a composite coordinate gathering spin, space and time variables (σ_1, r_1, t_1) .

In particular, Hedin's approach uses a dynamically screened Coulomb interaction W instead of the standard bare Coulomb interaction v . Important experimental properties such as ionization potentials, electron affinities as well as spectral functions, which are related to direct and inverse photo-emission, can be obtained directly from the one-body Green's function [57]. A particularly successful and practical approximation to Hedin's equations is the so-called GW approximation [57–59] which bypasses the calculation of the most complicated part of Hedin's equations, the vertex function [47].

Although (perturbative) G_0W_0 is probably the simplest and most widely used GW variant [60–65], its starting point dependence has motivated the development of partially [66–74] and fully [75–83] self-consistent versions in order to reduce or remove this undesirable feature. Here, we will

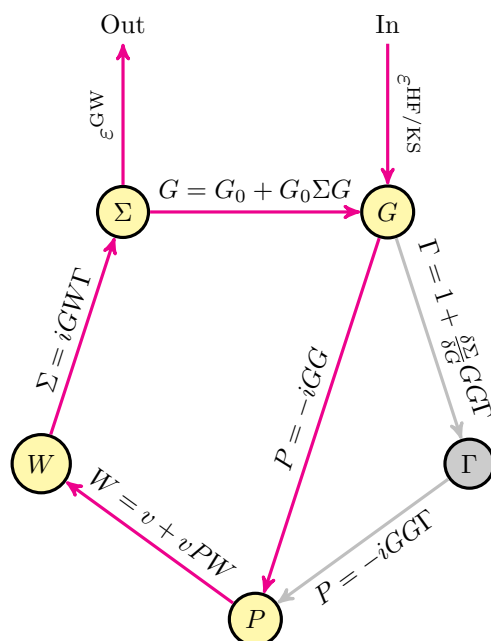


FIGURE 3.1: Hedin's pentagon [47]. The red path shows the self-consistent GW process which bypasses the computation of the vertex function Γ .

focus our attention on partially self-consistent schemes as they have demonstrated comparable accuracy and are computationally lighter than the fully self-consistent version [84]. Moreover, they are routinely employed for solid-state and molecular calculations and are available in various computational packages [61, 68, 74, 84–89]. Recently, an ever-increasing number of successful applications of partially self-consistent GW methods have sprung in the physics and chemistry literature for molecular systems [62–65, 73, 82, 85, 86, 90–96], as well as extensive and elaborate benchmark sets [64, 65, 89, 90, 97–101].

There exist two main types of partially self-consistent GW methods: i) “*eigenvalue-only quasiparticle*” GW (evGW) [66–69], where the quasiparticle (QP) energies are updated at each iteration, and ii) “*quasiparticle self-consistent*” GW (qsGW) [70–74], where one updates both the QP energies and the corresponding orbitals. Note that a starting point dependence remains in evGW as the orbitals are not self-consistently optimized in this case.

In an article [102], while studying a model two-electron system [103–108], we have observed that, within partially self-consistent GW (such as evGW and qsGW), one can observe, in the weakly correlated regime, (unphysical) discontinuities in the energy surfaces of several key quantities (ionization potential, electron affinity, HOMO-LUMO gap, total and correlation energies, as well as vertical excitation energies). In the present chapter, we provide further evidences and explanations of this undesirable feature in real molecular systems. For sake of simplicity, the present study is based on simple closed-shell diatomics (H_2 , F_2 and BeO). However, the same phenomenon can be observed in many other molecular systems, such as LiF , HeH^+ , LiH , BN , O_3 , etc. Although we mainly focus on G_0W_0 and evGW, similar observations can be made in the case of qsGW and second-order Green's function (GF2) methods [1, 102, 109–118]. Unless otherwise stated, all calculations have been performed with our locally-developed GW software, which closely follows the MOLGW implementation [86].

3.2 Theory

Here, we provide brief details about the main equations and quantities behind G_0W_0 and evGW considering a (restricted) Hartree-Fock (HF) starting point [1].

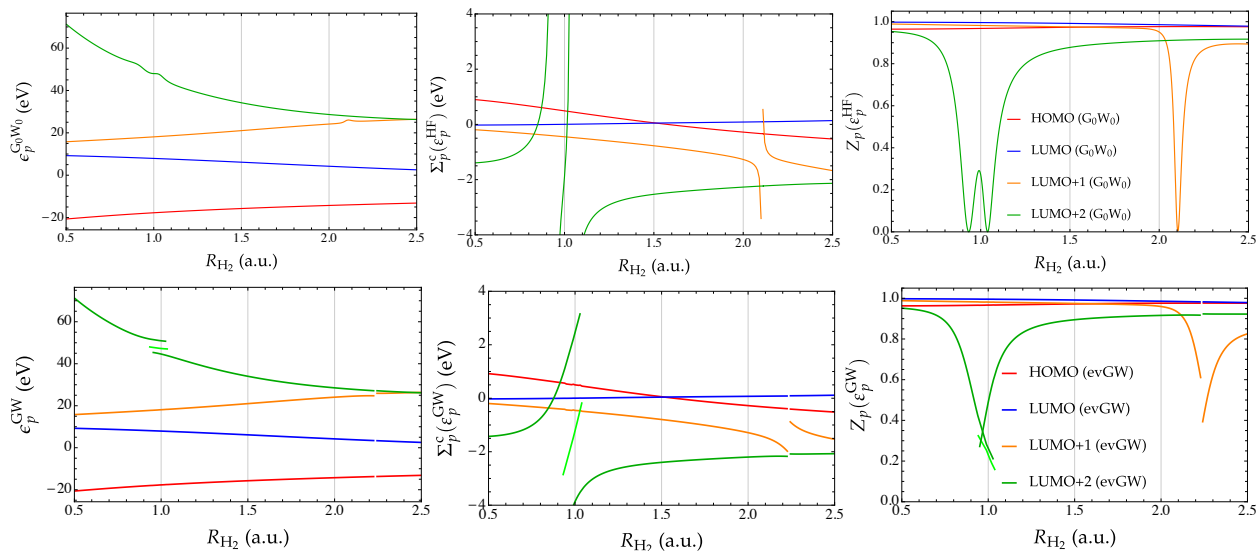


FIGURE 3.2: QP energies (left), correlation part of the self-energy (center) and renormalization factor (right) as functions of the internuclear distance R_{H_2} for various orbitals of H_2 at the $G_0W_0@HF/6-31G$ (top) and $evGW@HF/6-31G$ (bottom) levels. For convenience, the intermediate (center) branch is presented in lighter green for the LUMO+2.

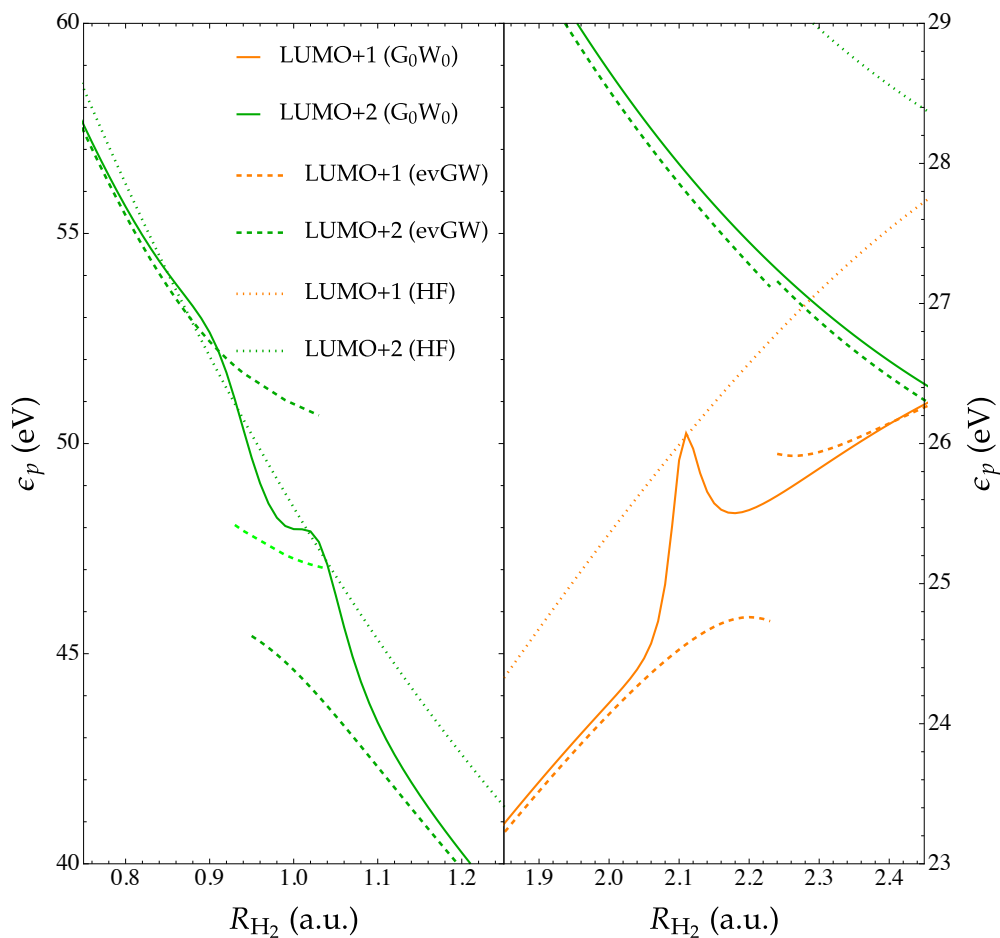


FIGURE 3.3: HF orbital energies (dotted lines) and QP energies as functions of the internuclear distance R_{H_2} for the LUMO+1 and LUMO+2 orbitals of H_2 at the $G_0W_0@HF/6-31G$ (solid lines) and $evGW@HF/6-31G$ (dashed lines) levels. For convenience, the intermediate (center) branch is presented in lighter green for the LUMO+2.

For a given (occupied or virtual) orbital p , the correlation part of the self-energy is conveniently split in its hole (h) and particle (p) contributions

$$\Sigma_p^c(\omega) = \Sigma_p^p(\omega) + \Sigma_p^h(\omega), \quad (3.2)$$

which, within the GW approximation, read

$$\Sigma_p^h(\omega) = 2 \sum_i^{\text{occ}} \sum_x \frac{[pi|x]^2}{\omega - \epsilon_i + \Omega_x - i\eta}, \quad (3.3a)$$

$$\Sigma_p^p(\omega) = 2 \sum_a^{\text{virt}} \sum_x \frac{[pa|x]^2}{\omega - \epsilon_a - \Omega_x + i\eta}, \quad (3.3b)$$

where η is a positive infinitesimal. The screened two-electron integrals

$$[pq|x] = \sum_{ia} (pq|ia)(\mathbf{X} + \mathbf{Y})_{ia}^x \quad (3.4)$$

are obtained via the contraction of the bare two-electron integrals [119] ($pq|rs$) and the transition densities $(\mathbf{X} + \mathbf{Y})_{ia}^x$ originating from a random phase approximation (RPA) calculation [120, 121]

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \mathbf{\Omega} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \quad (3.5)$$

with

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2(ia|jb), \quad B_{ia,jb} = 2(ia|bj), \quad (3.6)$$

and δ_{pq} is the Kronecker delta [122]. The one-electron energies ϵ_p in Eqs. (3.3a), (3.3b) and (3.6) are either the HF or the GW quasiparticle energies. Equation (3.5) also provides the neutral excitation energies Ω_x .

In practice, there exist two ways of determining the G_0W_0 QP energies [60, 61]. In its “graphical” version, they are provided by one of the many solutions of the (non-linear) QP equation

$$\omega = \epsilon_p^{\text{HF}} + \text{Re}[\Sigma_p^c(\omega)]. \quad (3.7)$$

In this case, special care has to be taken in order to select the “right” solution, known as the QP solution. In particular, it is usually worth calculating its renormalization weight (or factor), $Z_p(\epsilon_p^{\text{HF}})$, where

$$Z_p(\omega) = \left[1 - \frac{\partial \text{Re}[\Sigma_p^c(\omega)]}{\partial \omega} \right]^{-1}. \quad (3.8)$$

Because of sum rules [123–126], the other solutions, known as satellites, share the remaining weight. In a well-behaved case (belonging to the weakly correlated regime), the QP weight is much larger than the sum of the satellite weights, and of the order of 0.7–0.9.

Within the linearized version of G_0W_0 , one assumes that

$$\Sigma_p^c(\omega) \approx \Sigma_p^c(\epsilon_p^{\text{HF}}) + (\omega - \epsilon_p^{\text{HF}}) \left. \frac{\partial \Sigma_p^c(\omega)}{\partial \omega} \right|_{\omega=\epsilon_p^{\text{HF}}}, \quad (3.9)$$

that is, the self-energy behaves linearly in the vicinity of $\omega = \epsilon_p^{\text{HF}}$. Substituting (3.9) into (3.7) yields

$$\epsilon_p^{G_0W_0} = \epsilon_p^{\text{HF}} + Z_p(\epsilon_p^{\text{HF}}) \text{Re}[\Sigma_p^c(\epsilon_p^{\text{HF}})]. \quad (3.10)$$

Unless otherwise stated, in the remaining of this chapter, the G_0W_0 QP energies are determined via the linearized method.

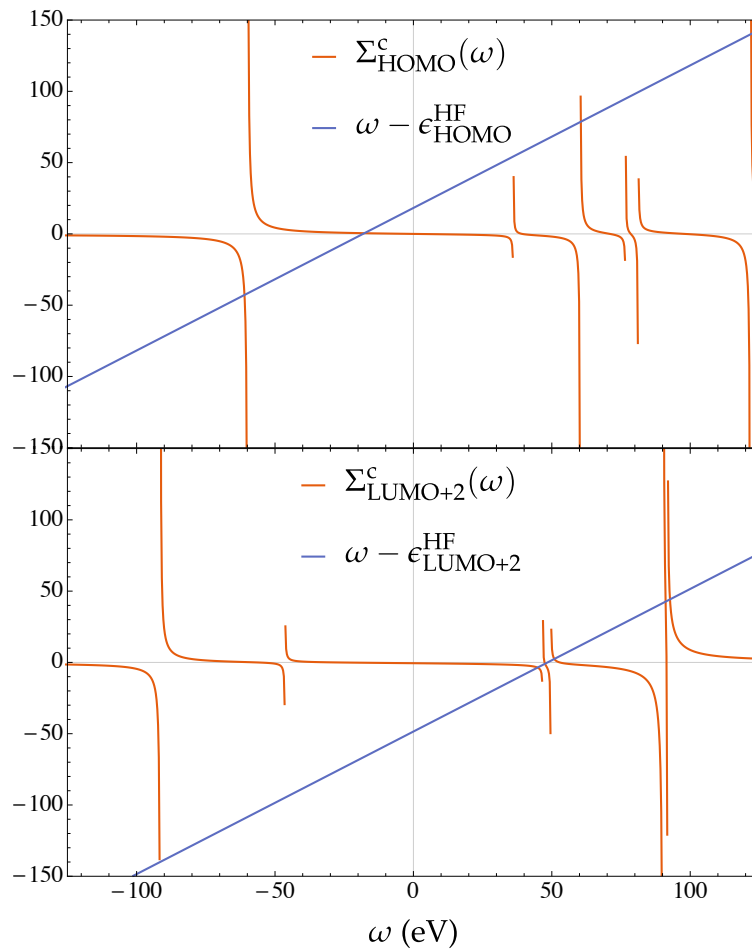


FIGURE 3.4: $\Sigma_{\text{HOMO}}^c(\omega)$ and $\Sigma_{\text{LUMO}+2}^c(\omega)$ (in eV) as functions of the frequency ω obtained at the evGW@HF/6-31G level for H_2 at $R_{\text{H}_2} = 1.0 a_0$. The solutions of the QP equation are given by the intersection of the orange and blue curves.

In the case of evGW, the QP energy, ϵ_p^{GW} , are obtained via Eq. (3.7), which has to be solved self-consistently due to the QP energy dependence of the self-energy [see Eq. (3.2)] [66–69]. At least in the weakly correlated regime where a clear QP solution exists, we believe that, within evGW, the self-consistent algorithm should select the solution of the QP equation (3.7) with the largest renormalization weight $Z_p(\epsilon_p^{\text{GW}})$. In order to avoid convergence issues, we have used the DIIS convergence accelerator technique proposed by Pulay [127, 128].

Moreover, throughout this chapter, we have set $\eta = 0$.

3.3 Results

3.3.1 Virtual orbitals

As a first example, we consider the hydrogen molecule H_2 in a relatively small gaussian basis set (6-31G) in order to be able to study easily the entire orbital energy spectrum. Although the number of irregularities/discontinuities as well as their locations may vary with the basis set, the conclusions we are going to draw here are general.

Figure 3.2 reports three key quantities as functions of the internuclear distance R_{H_2} for various orbitals at the G_0W_0 and the self-consistent evGW levels: i) the QP energies [$\epsilon_p^{G_0W_0}$ or ϵ_p^{GW}], ii) the correlation part of the self-energy [$\Sigma_p^c(\epsilon_p^{\text{HF}})$ or $\Sigma_p^c(\epsilon_p^{\text{GW}})$], and iii) the renormalization factor/weight [$Z_p(\epsilon_p^{\text{HF}})$ or $Z_p(\epsilon_p^{\text{GW}})$].

G_0W_0

Let us first consider the results of the G_0W_0 calculations reported in the top row of Figure 3.2. Looking at the curves of $\epsilon_p^{G_0W_0}$ as a function of R_{H_2} (top left graph of Figure 3.2), one notices obvious irregularities in the LUMO+2 around $R_{H_2} = 1.0 a_0$ and in the LUMO+1 around $R_{H_2} = 2.1 a_0$. For information, the experimental equilibrium geometry of H_2 is around $R_{H_2} = 1.4 a_0$ [129]. These irregularities are unphysical, and occur in correspondence with a series of poles in Σ_{LUMO+1}^c and Σ_{LUMO+2}^c (see top center graph of Figure 3.2). For example, one can notice two poles in Σ_{LUMO+2}^c just before and after $R_{H_2} = 1.0 a_0$, giving birth to three branches. The origin of the irregularities in ϵ_{LUMO+1} and ϵ_{LUMO+2} can, therefore, be traced back to the wrong assumption that $\Sigma_{LUMO+1}^c(\omega)$ and $\Sigma_{LUMO+2}^c(\omega)$ are linear functions of ω in the vicinity of, respectively, $\omega = \epsilon_{LUMO+1}^{HF}$ and $\omega = \epsilon_{LUMO+2}^{HF}$ [see Eq. (3.9)].

However, despite the divergencies in the self-energy, the QP energies $\epsilon_{LUMO+1}^{G_0W_0}$ and $\epsilon_{LUMO+2}^{G_0W_0}$ remain finite thanks to a rapid decrease of the renormalization factor at the R_{H_2} values for which the self-energy diverges [see Eq. (3.7) and top right graph of Figure 3.2]. For example, note that Z_{LUMO+2} reaches exactly zero at the pole locations. A very similar scenario unfolds for the LUMO+1, except that a single pole is present in Σ_{LUMO+1}^c .

Let us analyze this point further. Since the self-energy behaves as $\Sigma_p^c \sim \delta^{-1}$ (with $\delta \rightarrow 0$) in the vicinity of a singularity, one can easily show that $Z_p \sim (1 + \delta^{-2})^{-1} \sim \delta^2$, which yields $\epsilon_p^{G_0W_0} \sim \epsilon_p^{HF} + \delta$. In plain words, $\epsilon_p^{G_0W_0}$ remains finite near the poles of the self-energy thanks to the linearization of the QP equation [see Eq. (3.7)]. It also evidences that, at the pole locations (*i.e.* $\delta = 0$), we have $\epsilon_p^{G_0W_0} = \epsilon_p^{HF}$, *i.e.*, by construction the QP energy is forced to remain equal to the zeroth-order energy. This is nicely illustrated in Figure 3.3, where we have plotted the HF orbital energies (dotted lines) as well as the G_0W_0 QP energies (solid lines) around the two “problematic” internuclear distances. The behavior of $\epsilon_{LUMO+1}^{G_0W_0}$ (solid orange line on the right panel of Figure 3.3) is particularly instructive and shows that the G_0W_0 QP energies can have an erratic behavior near the poles of the self-energy.

It is interesting to investigate further the origin of these poles. As evidenced by Eq. (3.2), for a calculation involving $2n$ electrons and N basis functions, the self-energy has exactly $nN(N - n)$ poles originating from the combination of the N poles of the Green’s function G (at frequencies ϵ_p) and the $n(N - n)$ poles of the screened Coulomb interaction W (at the RPA singlet excitations Ω_x). For example, at $R_{H_2} = 2.11 a_0$, the combination of $\epsilon_{LUMO}^{HF} = 3.83$ eV and the HOMO-LUMO-dominated first neutral excitation energy $\Omega_1 = 22.24$ eV are equal to the LUMO+1 energy $\epsilon_{LUMO+1}^{G_0W_0} = 26.07$ eV. Around $R_{H_2} = 1.0 a_0$, the two poles of Σ_{LUMO+1}^c are due to the following accidental equalities: $\epsilon_{LUMO+1}^{G_0W_0} = \epsilon_{LUMO}^{HF} + \Omega_2$, and $\epsilon_{LUMO+1}^{G_0W_0} = \epsilon_{LUMO+1}^{HF} + \Omega_1$. Because the number of poles in G and W (at the non-interacting or HF level) are both proportional to N , these spurious poles in the self-energy become more and more frequent for larger gaussian basis sets. For virtual orbitals, the higher in energy the orbital is, the earlier the singularities seem to appear.

Finally, the irregularities in the G_0W_0 QP energies as a function of R_{H_2} can also be understood as follows. Since within G_0W_0 only one pole of G is calculated, *i.e.*, the QP energy, all the satellite poles are discarded. Mixing between QP and satellites poles, which is important when they are close to each other, hence, is not considered. This situation can be compared to the lack of mixing between single and double excitations in adiabatic time-dependent density-functional theory and the Bethe-Salpeter equation [20, 130–132]

evGW

Within partially self-consistent schemes, the presence of poles in the self-energy at a frequency similar to a QP energy has more dramatic consequences. The results for H_2 at the evGW@HF/6-31G level are reported in the bottom row of Figure 3.2. Around $R_{H_2} = 1.0 a_0$, we observe that, for the LUMO+2, one can fall onto three distinct solutions depending on the algorithm one relies on to solve self-consistently the QP equation (see bottom left graph of Figure 3.2). In order to obtain each

of the three possible solutions in the vicinity of $R_{\text{H}_2} = 1.0 a_0$, we have run various sets of calculations using different starting values for the QP energies and sizes of the DIIS space. In particular, we clearly see that each of these solutions yield a distinct energy separated by several electron volts (see zoom in Figure 3.3), and each of them is associated with a well-defined branch of the self-energy, as shown by the center graph in the bottom row of Figure 3.2. For convenience, the intermediate (center) branch is presented in lighter green in Figures 3.2 and 3.3, while the left and right branches are depicted in darker green. Interestingly, the evGW iterations are able to “push” the QP solution away from the poles of the self-energy, which explains why the renormalization factor is never exactly equal to zero (see bottom right graph of Figure 3.2). However, one cannot go smoothly from one branch to another, and each switch between solutions implies a significant energetic discontinuity. Moreover, we observe “ripple” effects in other virtual orbitals: a discontinuity in one of the QP energies induces (smaller) discontinuities in the others. This is a direct consequence of the global energy dependence of the self-energy [see Eq. (3.2)], and is evidenced on the left graph in the bottom row of Figure 3.2 around $R_{\text{H}_2} = 2.1 a_0$.

The main observation of the present study is that each branch of the self-energy is associated with a distinct QP solution. We clearly see that, when one goes from one branch to another, there is a transfer of weight between the QP and one of the satellites, which becomes the QP on the new branch [102]. As opposed to the strongly correlated regime where the QP picture breaks down, *i.e.*, there is no clear QP, here there is always a clear QP except at the vicinity of the poles where the weight transfer occurs. As for G_0W_0 , this sudden transfer is caused by the artificial removal of the satellite poles. However, in the evGW results the problem is amplified by the self-consistency. We expect that keeping the full frequency dependence of the self-energy would solve this problem.

It is also important to mention that the self-consistent algorithm is fairly robust as it rarely selects a solution with a renormalization weight lower than 0.5, as shown by the center graph in the bottom row of Figure 3.2. In other words, when the renormalization factor of the QP solution becomes too small, the self-consistent algorithm switches naturally to a different solution. From a technical point of view, around the poles of the self-energy, it is particularly challenging to converge self-consistent calculations, and we heavily relied on DIIS to avoid such difficulties. We note that an alternative *ad hoc* approach to stabilize such self-consistent calculations is to increase the value of the positive infinitesimal η .

Figure 3.4 shows the correlation part of the self-energy for the HOMO and LUMO+2 orbitals as a function of ω (orange curves) obtained at the self-consistent evGW@HF/6-31G level for H_2 with $R_{\text{H}_2} = 1.0 a_0$. The solutions of the QP equation (3.7) are given by the intersections of the orange and blue curves. On the one hand, in the case of the HOMO, we have an unambiguous QP solution (at $\omega \approx -20$ eV) which is well separated from the other solutions. In this case, one can anticipate a large value of the renormalization factor Z_{HOMO} as the self-energy is flat around the intersection of the two curves. On the other hand, for the LUMO+2, we see three solutions of the QP equation very close in energy from each other around $\omega = 50$ eV. In this particular case, there is no well-defined QP peak as each solution has a fairly small weight. Therefore, one may anticipate multiple solution issues when a solution of the QP equation is close to a pole of the self-energy.

Finally, we note that the multiple solutions discussed here are those of the QP equation, *i.e.*, multiple QP poles associated to a *single* Green’s function.

3.3.2 Occupied orbitals

So far, we have seen that multiple solutions seem to only appear for virtual orbitals (LUMO excluded). However, we will show here that it can also happen in occupied orbitals. We take as an example the fluorine molecule (F_2) in a minimal basis set (STO-3G), and perform evGW@HF calculations within the frozen-core approximation, that is, we do not update the orbital energies associated with the core orbitals. Figure 3.5 shows the behavior (as a function of the distance between the two fluorine atoms R_{F_2}) of the same quantities as in Figure 3.2 but for some of the occupied orbitals of F_2 (HOMO-6, HOMO-5 and HOMO-4). Similarly to the case of H_2 discussed in the previous section, we see discontinuities in the QP energies around $R_{\text{F}_2} = 2.3 a_0$ (for the

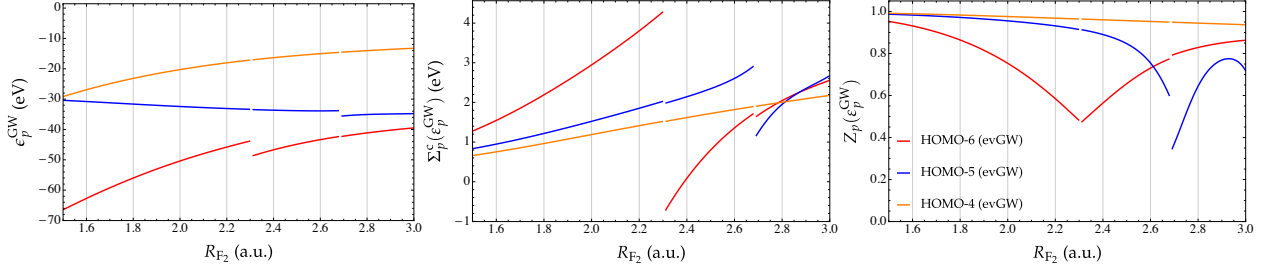


FIGURE 3.5: QP energies (left), correlation part of the self-energy (center) and renormalization factor (right) as functions of the internuclear distance R_{F_2} for various occupied orbitals of F_2 at the evGW@HF/STO-3G level.

HOMO-6) and $R_{F_2} = 2.7 a_0$ (for the HOMO-5). For information, the experimental equilibrium geometry of F_2 is $R_{F_2} = 2.668 a_0$, which evidences that the second discontinuity is extremely close to the experimental geometry. Let us mention here that we have not found any discontinuity in the HOMO orbital. The case of the frontier orbitals will be discussed below. For F_2 , here again, we clearly observe ripple effects on other occupied orbitals. Similarly to virtual orbitals, we have found that the lower in energy the orbital is, the earlier the singularities seem to appear.

3.3.3 Frontier orbitals

Before concluding, we would like to know, whether or not, this multisolution behavior can potentially appear in frontier orbitals. This is an important point to discuss as these orbitals are directly related to the ionization potential and the electron affinity, hence to the gap.

Let us take the HOMO orbital as an example. A similar rationale holds for the LUMO orbital. According to the expression of the hole and particle parts of the self-energy given in Eqs. (3.3a) and (3.3b) respectively, $\Sigma_{\text{HOMO}}^c(\omega)$ has poles at $\omega = \epsilon_i - \Omega_x$ and $\omega = \epsilon_a + \Omega_x$ with $\Omega_x > 0$. Evaluating the self-energy at $\omega = \epsilon_{\text{HOMO}}$ would yield $\epsilon_{\text{HOMO}} - \epsilon_i = -\Omega_x$ and $\epsilon_{\text{HOMO}} - \epsilon_a = +\Omega_x$, which is in clear contradiction with the assumption that $\Omega_x > 0$. Therefore, the self-energy is never singular at $\omega = \epsilon_{\text{HOMO}}$ and $\omega = \epsilon_{\text{LUMO}}$ and the linearized G_0W_0 equations can be solved without any problem for the frontier orbitals. This is true for any G_0 , that is, it does not depend on the starting point. As can be seen from Eqs. (3.3a) and (3.3b), the two poles of the self-energy closest to the Fermi level are located at $\omega = \epsilon_{\text{HOMO}} - \Omega_1$ and $\omega = \epsilon_{\text{LUMO}} + \Omega_1$. As a consequence, there is a region equal to $\epsilon_{\text{HOMO}} - \epsilon_{\text{LUMO}} + 2\Omega_1$ around the Fermi level in which the self-energy does not have poles. Because $\Omega_1 \approx \epsilon_{\text{HOMO}} - \epsilon_{\text{LUMO}} = E_{\text{gap}}$, this region is approximately equal to $3E_{\text{gap}}$.

For “graphical” G_0W_0 , the solution might lie outside this range, even for the frontier orbitals. This can happen when E_{gap} is much smaller than the true GW gap. In particular, this could occur for graphical G_0W_0 on top of a Kohn-Sham starting point, which is known to yield gaps that are (much) smaller than GW gaps. Within graphical G_0W_0 , multiple solution issues for the HOMO have been reported by van Setten and coworkers [64, 89] in several systems (LiH, BN, BeO and O_3). In their calculations, they employed PBE orbital energies [133] as starting point, and this type of functionals is well known to drastically underestimate E_{gap} [134].

As an example, we have computed, within the frozen-core approximation, $\Sigma_{\text{HOMO}}^c(\omega)$ and $\Sigma_{\text{LUMO}}^c(\omega)$ as functions of ω at the G_0W_0 @PBE/cc-pVDZ level for beryllium monoxide (BeO) at its experimental geometry (*i.e.* $R_{\text{BeO}} = 2.515 a_0$) [129]. These calculations have been performed with MOLGW [86]. The results are gathered in Figure 3.6, where one clearly sees that multiple solutions appear for both the HOMO and LUMO orbitals. Note that performing the same set of calculations with a HF starting point yields a perfectly unambiguous single QP solution. For this system, PBE is a particularly bad starting point for a GW calculation with a HOMO-LUMO gap equal to 1.35 eV. Using the same basis set, HF yields a gap of 8.96 eV, while G_0W_0 @HF and G_0W_0 @PBE yields 7.54 and 5.60 eV. The same observations can be made for the other systems reported as problematic by van Setten and coworkers [64, 89]. As a general rule, it is known that HF is usually a better starting point for GW in small molecular systems [63, 68, 102, 135]. For larger systems, hybrid

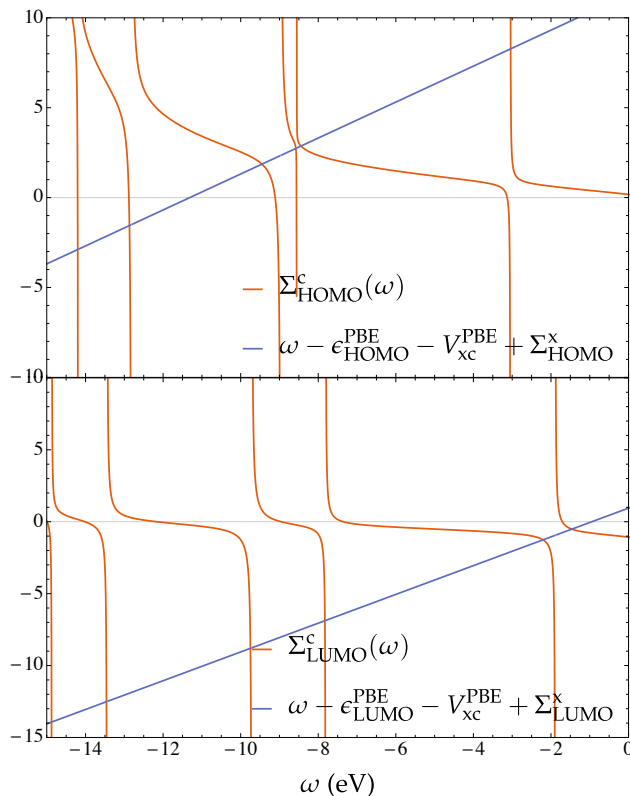


FIGURE 3.6: $\Sigma_{\text{HOMO}}^c(\omega)$ and $\Sigma_{\text{LUMO}}^c(\omega)$ (in eV) as functions of the frequency ω obtained at the G_0W_0 @PBE/cc-pVDZ level for BeO at its experimental geometry [129]. The solutions of the QP equations are given by the intersection of the orange and blue curves.

functionals [136] might be the ideal compromise, thanks to the increase of the HOMO-LUMO gap via the addition of (exact) HF exchange [63, 90, 93, 101, 137].

3.4 Concluding remarks

The GW approximation of many-body perturbation theory has been highly successful at predicting the electronic properties of solids and molecules [57–59]. However, it is also known to be inadequate to model strongly correlated systems [138–142]. Here, we have found severe shortcomings of two widely-used variants of GW in the weakly correlated regime. We have evidenced that one can hit multiple solution issues within G_0W_0 and evGW due to the location of the QP solution near poles of the self-energy. Within linearized G_0W_0 , this implies irregularities in key experimentally-measurable quantities of simple diatomics, while, at the partially self-consistent evGW level, discontinuities arise. Because the RPA correlation energy [86, 120, 143, 144] and the Bethe-Salpeter excitation energies [85, 145, 146] directly dependent on the QP energies, these types of discontinuities are also present in these quantities, hence in the energy surfaces of ground and excited states. Illustrative examples can be found in our previous study [102]. We believe that such discontinuities would not exist within a fully self-consistent scheme where one does not iterate the QP energies but the one-body Green’s function and therefore takes into account each QP peak as well as its satellites at every iteration. Obviously, this latter point deserves further investigations. However, if confirmed, this would be a strong argument in favor of fully self-consistent schemes. Also, for extended systems, these issues might be mitigated by the plasmon modes that dominate the high-energy spectrum of the screened Coulomb interaction. The results of this work will be useful for self-consistent GW calculations of dynamical phenomena, *i.e.*, with nuclear motion.

We are currently exploring different routes in order to remove these unphysical features. Padé resummation technique could be of great interest [147] for such purpose. However, other techniques

might be successful at alleviating this issue. For example, one could i) impose a larger offset from the real axis (*i.e.* increasing the value of η), ii) favor, in the case of small systems, a HF starting point in order to avoid small HOMO-LUMO gaps, or iii) rely, for larger systems, on hybrid functionals including a significant fraction of HF exchange. Also, regularization techniques, such as the one developed for orbital-optimized second-order Møller-Plesset perturbation theory, could be a pragmatic and efficient way of removing such discontinuities [148].

Chapter 4

The QUEST database of vertical excitation energies

This chapter is based on

M. V eril et al., "QUESTDB : a database of highly accurate excitation energies for the electronic structure community", *WIREs Comput. Mol. Sci.* **11**, e1517 (2021)

In this chapter, my main contribution was to build the QUEST website and the different tools around it. I have also imported all the data on the website.

4.1 Introduction

Nowadays, there exist a very large number of electronic structure computational approaches, more or less expensive depending on their overall accuracy, able to quantitatively predict the absolute and/or relative energies of electronic states in molecular systems [1–3, 150]. One important aspect of some of these theoretical methods is their ability to access the energies of electronic excited states, *i.e.*, states that have higher total energies than the so-called ground (that is, lowest-energy) state [13, 39, 121, 151–158]. The faithful description of excited states is particularly challenging from a theoretical point of view but is key to a deeper understanding of photochemical and photophysical processes like absorption, fluorescence, phosphorescence, chemoluminescence, and others [159–165]. For a given level of theory, ground-state methods are usually more accurate than their excited-state analogs. The reasons behind this are (at least) threefold: i) accurately modeling the electronic structure of excited states usually requires larger one-electron basis sets (including diffuse functions most of the times) than their ground-state counterpart, ii) excited states can be governed by different amounts of dynamic/static correlations, present very different physical natures ($\pi \rightarrow \pi^*$, $n \rightarrow \pi^*$, charge transfer, double excitation, valence, Rydberg, singlet, doublet, triplet, etc), yet be very close in energy from one another, and iii) one usually has to rely on response theory formalisms [26, 166–172], which inherently introduce a ground-state "bias". Hence, designing excited-state methods able to tackle simultaneously and on an equal footing all these types of excited states at an affordable cost remains an open challenge in theoretical computational chemistry as evidenced by the large number of review articles on this particular subject [13, 39, 121, 151–158, 173].

When designing a new theoretical model, the first feature that one might want to test is its overall accuracy, *i.e.*, its ability to reproduce reference (or benchmark) values for a given system with a well-defined setup (same geometry, basis set, etc). These values can be absolute and/or relative energies, geometrical parameters, physical or chemical spectroscopic properties extracted from experiments, high-level theoretical calculations, or any combination of these. To this end, the electronic structure community has designed along the years benchmark sets, *i.e.*, sets of molecules for which one can (very) accurately compute theoretical estimates and/or access solid experimental data for given properties. Regarding ground-states properties, two of the oldest and most employed sets are probably the Gaussian-1 and Gaussian-2 benchmark sets [174–176] developed by the group of Pople in the 1990's. For example, the Gaussian-2 set gathers atomization energies, ionization energies, electron affinities, proton affinities, bond dissociation energies, and reaction barriers. This

set was subsequently extended and refined [177, 178]. Another very useful set for the design of methods able to catch dispersion effects [179] is the S22 benchmark set [180] (and its extended S66 version [181]) of Hobza and collaborators which provides benchmark interaction energies for weakly-interacting (non covalent) systems. One could also mention the GW100 set [64, 182, 183] (and its GW5000 extension [184]) of ionization energies which has helped enormously the community to compare the implementation of GW-type methods for molecular systems [61, 84, 86, 185]. The extrapolated *ab initio* thermochemistry (HEAT) set designed to achieve high accuracy for enthalpies of formation of atoms and small molecules (without experimental data) is yet another successful example of benchmark set [186–188]. More recently, let us mention the benchmark datasets of the *Simons Collaboration on the Many-Electron Problem* providing, for example, highly-accurate ground-state energies for hydrogen chains [189] as well as transition metal atoms and their ions and monoxides [190]. Let us also mention the set of Zhao and Truhlar for small transition metal complexes employed to compare the accuracy of density-functional methods [134] for 3d transition-metal chemistry [191], and finally the popular GMTKN24 [192], GMTKN30 [193, 194] and GMTKN55 [195] databases for general main group thermochemistry, kinetics, and non-covalent interactions developed by Goerigk, Grimme and their coworkers.

The examples of benchmark sets presented above are all designed for ground-state properties, and there exist specific protocols tailored to accurately model excited-state energies and properties as well. Indeed, benchmark datasets of excited-state energies and/or properties are less numerous than their ground-state counterparts but their number has been growing at a consistent pace in the past few years. Below, we provide a short description for some of them. One of the most characteristic examples is the benchmark set of vertical excitation energies proposed by Thiel and coworkers [196–199]. The so-called Thiel (or Mülheim) set of excitation energies gathers a large number of excitation energies determined in 28 medium-sized organic CNOH molecules with a total of 223 valence excited states (152 singlet and 71 triplet states) for which theoretical best estimates (TBEs) were defined. In their first study, Thiel and collaborators performed CC2 [25, 200], CCSD [23, 168, 201, 202], CC3 [26, 203], and CASPT2 [10, 11, 13, 204] calculations (with the TZVP basis) on MP2/6-31G(d) geometries in order to provide (based on additional high-quality literature data) TBEs for these transitions. These TBEs were quickly refined with the larger *aug-cc-pVTZ* basis set [198, 199]. In the same spirit, it is also worth mentioning Gordon's set of vertical transitions (based on experimental values) [205] used to benchmark the performance of time-dependent density-functional theory (TD-DFT) [15, 16, 206, 207], as well as its extended version by Goerigk and coworkers who decided to replace the experimental reference values by CC3 excitation energies [208–210]. For comparisons with experimental values, there also exist various sets of measured 0–0 energies used in various benchmarks, notably by the Furche [211, 212], Hättig [213] and our [5, 102, 214] groups for gas-phase compounds and by Grimme [215, 216] and one of us [217, 218] for solvated dyes. Let us also mention the new benchmark set of charge-transfer excited states recently introduced by Szalay and coworkers [based on equation-of-motion coupled cluster (EOM-CC) methods] [219] as well as the Gagliardi-Truhlar set employed to compare the accuracy of multiconfiguration pair-density functional theory [157] against the well-established CASPT2 method [220].

Following a similar philosophy and striving for chemical accuracy, we have recently reported in several studies highly-accurate vertical excitations for small- and medium-sized molecules [30, 32, 36, 39, 221]. The so-called QUEST dataset of vertical excitations which we will describe in detail in the present chapter is composed by 5 subsets (see Figure 4.1): i) a subset of excitations in small molecules containing from 1 to 3 non-hydrogen atoms known as QUEST#1, ii) a subset of double excitations in molecules of small and medium sizes known as QUEST#2, iii) a subset of excitation energies for medium-sized molecules containing from 4 to 6 non-hydrogen atoms known as QUEST#3, iv) a subset composed by more "exotic" molecules and radicals labeled as QUEST#4, and v) a subset known as QUEST#5, specifically designed for the present chapter, gathering excitation energies in larger molecules as well as additional smaller molecules. One of the key aspects of the QUEST dataset is that it does not rely on any experimental values, avoiding potential biases inherently linked to experiments and facilitating in the process theoretical comparisons.

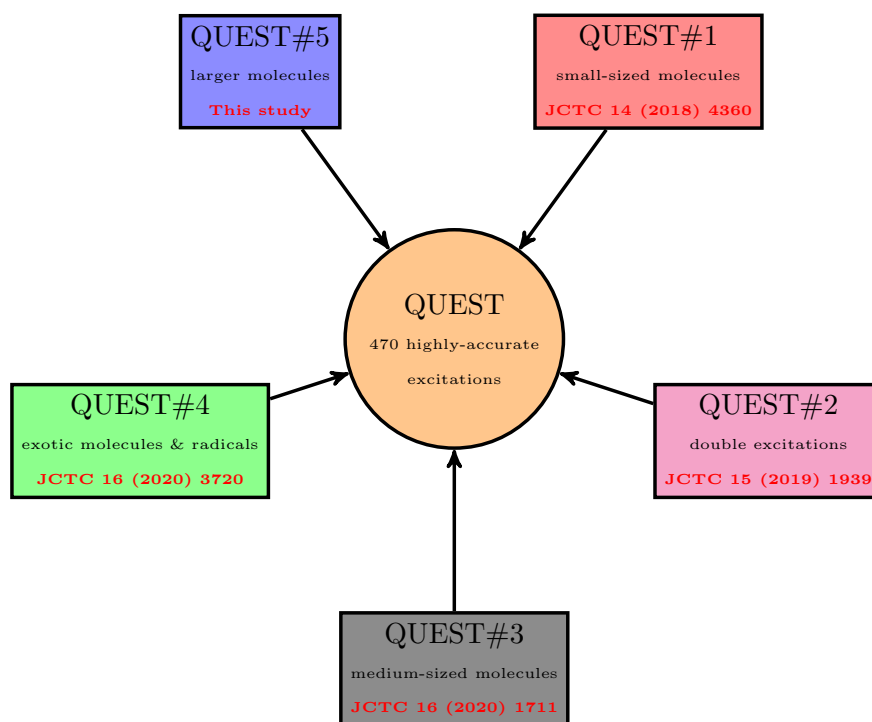


FIGURE 4.1: Composition of each of the five subsets making up the present QUEST dataset of highly-accurate vertical excitation energies.

Moreover, our protocol has been designed to be as uniform as possible, which means that we have designed a very systematic procedure for all excited states in order to make cross-comparison as straightforward as possible. Importantly, it allowed us to benchmark, in a very systematic and balanced way, a series of popular excited-state wave function methods partially or fully accounting for double and triple excitations as well as multiconfigurational methods (see below). In the same vein, as evoked above, we have also produced chemically-accurate theoretical 0–0 energies [5, 102, 214] which can be more straightforwardly compared to experimental data [211–213, 215–218, 222–224]. We refer the interested reader to Ref. [5] for a review of the generic benchmark studies devoted to adiabatic and 0–0 energies performed in the past two decades.

The QUEST dataset has the particularity to be based to a large extent on selected configuration interaction (SCI) reference excitation energies as well as high-order linear-response (LR) CC methods such as LR-CCSDT and LR-CCSDTQ [24, 168, 169, 171, 225–230]. Recently, SCI methods have been a force to reckon with for the computation of highly-accurate energies in small- and medium-sized molecules as they yield near full configuration interaction (FCI) quality energies for only a very tiny fraction of the computational cost of a genuine FCI calculation [30, 32, 36–39, 221, 231–250]. Due to the fairly natural idea underlying these methods, the SCI family is composed of numerous members [34, 36–38, 240, 251–276]. Their fundamental philosophy consists, roughly speaking, in retaining only the most relevant determinants of the FCI space following a given criterion to slow down the exponential increase of the size of the CI expansion. Originally developed in the late 1960’s by Bender and Davidson [251] as well as Whitten and Hackmeyer [252], new efficient SCI algorithms have resurfaced recently. Three examples are iCI [272, 277–279], semistochastic heat-bath CI (SHCI) [37, 240–242, 262, 263], and *Configuration Interaction using a Perturbative Selection made Iteratively* (CIPSI) [34, 35, 256, 258]. These flavors of SCI include a second-order perturbative (PT2) correction which is key to estimate the “distance” to the FCI solution (see below). The SCI calculations performed for the QUEST set of excitation energies relies on the CIPSI algorithm, which is, from a historical point of view, one of the oldest SCI algorithms. It was developed in 1973 by Huron, Rancurel, and Malrieu [34] (see also Refs. [280–284]). Recently, the determinant-driven CIPSI algorithm has been efficiently implemented [35] in the open-source programming environment QUANTUM PACKAGE by the Toulouse group enabling to perform massively parallel

computations [35, 249, 259, 266]. CIPSI is also frequently employed to provide accurate trial wave functions for quantum Monte Carlo calculations in molecules [256–258, 260, 261, 264, 265, 285–290] and more recently for periodic solids [291]. We refer the interested reader to Ref. [35] where one can find additional details regarding the implementation of the CIPSI algorithm.

The present chapter is organized as follows. In section 4.2, we detail the specificities of our protocol by providing computational details regarding geometries, basis sets, (reference and benchmarked) computational methods, and a new way of estimating rigorously the extrapolation error in SCI calculations which is tested by computing additional FCI values for five- and six-membered rings. We then describe in section 4.3 the content of our five QUEST subsets providing for each of them the number of reference excitation energies, the nature and size of the molecules, the list of benchmarked methods, as well as other specificities. A special emphasis is placed on the add-on published during my thesis, QUEST#5, specifically designed for the present part of this thesis where we have considered, in particular but not only, larger molecules. Section 4.4 discusses the generation of the TBEs, while section 4.5 proposes a comprehensive benchmark of various methods on the entire QUEST set which is composed by more than 400 excitations with, in addition, a specific analysis for each type of excited states. Section 4.6 describes the feature of the website that we have specifically designed to gather the entire data generated during these last few years. Thanks to this website, one can easily test and compare the accuracy of a given method with respect to various variables such as the molecule size or its family, the nature of the excited states, the size of the basis set, etc. Finally, we draw our conclusions in section 4.7 where we discuss, in particular, future projects aiming at expanding and improving the usability and accuracy of the QUEST database.

4.2 Computational tools

4.2.1 Geometries

The ground-state structures of the molecules included in the QUEST dataset have been systematically optimized at the CC3/*aug-cc-pVTZ* level of theory, except for a very few cases. As shown in Refs. [172, 292], CC3 provides extremely accurate ground- and excited-state geometries. These optimizations have been performed using DALTON 2017 [293] and CFOUR 2.1 [294] applying default parameters. For the open-shell derivatives belonging to QUEST#4 [221], the geometries are optimized at the UCCSD(T)/*aug-cc-pVTZ* level using the GAUSSIAN16 program [295] and applying the “tight” convergence threshold. For the purpose of the present chapter, one can find all the geometries in the supporting information of the associated review article.

4.2.2 Basis sets

For the entire set, we rely on the 6-31+G(d) Pople basis set [296–302], the augmented family of Dunning basis sets *aug-cc-pVXZ* (where X = D, T, Q, and 5) [303–307], and sometimes its doubly- and triply-augmented variants, *d-aug-cc-pVXZ* and *t-aug-cc-pVXZ* respectively. Doubly- and triply-augmented basis sets are usually employed for Rydberg states where it is not uncommon to observe a strong basis set dependence due to the very diffuse nature of these excited states. These basis sets are available from the [basis set exchange](#) website [308–310].

4.2.3 Computational methods

Reference computational methods

In order to compute reference vertical energies, we have designed different strategies depending on the actual nature of the transition and the size of the system. For small molecules (typically 1–3 non-hydrogen atoms), we mainly resort to SCI methods which can provide near-FCI excitation energies for compact basis sets. Obviously, the smaller the molecule, the larger the basis we can afford. For larger systems (*i.e.*, 4–6 non-hydrogen atom), one cannot afford SCI calculations

anymore except in a few special occasions, and we then rely on LR-CC theory (LR-CCSDT and LR-CCSDTQ typically [171, 226, 228–230]) to obtain accurate transition energies. In the following, we will omit the prefix LR for the sake of clarity, as equivalent values would be obtained with the equation-of-motion (EOM) formalism [23, 201].

The CC calculations are performed with several codes. For closed-shell molecules, CC3 [26, 203] calculations are achieved with DALTON [293] and CFOUR [294]. CCSDT and CCSDTQ calculations are performed with CFOUR [294] and MRCC 2017 [311, 312], the latter code being also used for CCSDTQP. The reported oscillator strengths have been computed in the LR-CC3 formalism only. For open-shell molecules, the CCSDT, CCSDTQ, and CCSDTQP calculations performed with MRCC [311, 312] do consider an unrestricted Hartree-Fock wave function as reference but for a few exceptions. All excited-state calculations are performed, except when explicitly mentioned, in the frozen-core (FC) approximation using large cores for the third-row atoms.

All the SCI calculations are performed within the frozen-core approximation using QUANTUM PACKAGE [35] where the CIPSI algorithm [34] is implemented. Details regarding this specific CIPSI implementation can be found in Refs. [35] and [287]. A state-averaged formalism is employed, *i.e.*, the ground and excited states are described with the same set of determinants and orbitals, but different CI coefficients. Our usual protocol [30, 32, 36, 39, 221, 264, 265, 287] consists of performing a preliminary CIPSI calculation using Hartree-Fock orbitals in order to generate a CIPSI wave function with at least 10^7 determinants. Natural orbitals are then computed based on this wave function, and a new, larger CIPSI calculation is performed with this new set of orbitals. This has the advantage to produce a smoother and faster convergence of the SCI energy toward the FCI limit. The CIPSI energy E_{CIPSI} is defined as the sum of the variational energy E_{var} (computed via diagonalization of the CI matrix in the reference space) and a PT2 correction E_{PT2} which estimates the contribution of the determinants not included in the CI space [259]. By linearly extrapolating this second-order correction to zero, one can efficiently estimate the FCI limit for the total energies. These extrapolated total energies (simply labeled as E_{FCI} in the remainder of the chapter) are then used to compute vertical excitation energies. Depending on the set, we estimated the extrapolation error via different techniques. For example, in Ref. [32], we estimated the extrapolation error by the difference between the transition energies obtained with the largest SCI wave function and the FCI extrapolated value. This definitely cannot be viewed as a true error bar, but it provides an idea of the quality of the FCI extrapolation and estimate. Below, we provide a much cleaner way of estimating the extrapolation error in SCI methods, and we adopt this scheme for the five- and six-membered rings considered in the QUEST#3 subset. The particularity of the current implementation is that the selection step and the PT2 correction are computed *simultaneously* via a hybrid semistochastic algorithm [35, 259]. Moreover, a renormalized version of the PT2 correction (dubbed rPT2) has been recently implemented for a more efficient extrapolation to the FCI limit [35]. We refer the interested reader to Ref. [35] where one can find all the details regarding the implementation of the CIPSI algorithm. Note that all our SCI wave functions are eigenfunctions of the \hat{S}^2 spin operator which is, unlike ground-state calculations, paramount in the case of excited states [313].

Benchmarked computational methods

Using a large variety of codes, our benchmark effort consists in evaluating the accuracy of vertical transition energies obtained at lower levels of theory. For example, we rely on GAUSSIAN [295] and TURBOMOLE 7.3 [314] for CIS(D) [7, 315]; Q-CHEM 5.2 [316] for EOM-MP2 [CCSD(2)] [317] and ADC(3) [29, 31, 173]; Q-CHEM [316] and TURBOMOLE [314] for ADC(2) [173, 318]; DALTON [293] and TURBOMOLE [314] for CC2 [25, 200]; DALTON [293] and GAUSSIAN [295] for CCSD [23, 168, 202]; DALTON [293] for CCSDR(3) [319]; CFOUR [294] for CCSDT-3 [320, 321]; and ORCA [322] for similarity-transformed EOM-CCSD (STEOM-CCSD) [323, 324]. In addition, we evaluate the spin-opposite scaling (SOS) variants of ADC(2), SOS-ADC(2), as implemented in both Q-CHEM [325] and TURBOMOLE [326]. Note that these two codes have distinct SOS implementations, as explained in Ref. [325]. We also test the SOS and spin-component scaled (SCS) versions of CC2, as implemented in TURBOMOLE [314, 326]. Discussion of various spin-scaling schemes can be

found elsewhere [216]. For the STEOM-CCSD calculations, it was checked that the active character percentage was, at least, 98 %. For radicals, we applied both the U (unrestricted) and RO (restricted open-shell) versions of CCSD and CC3 as implemented in the PSI4 code [327] to perform our benchmarks. Finally, the composite approach, ADC(2.5), which follows the spirit of Grimme’s and Hobza’s MP2.5 approach [328] by averaging the ADC(2) and ADC(3) excitation energies, is also tested in the following [33].

For the double excitations composing the QUEST database, we have performed additional calculations using various multiconfigurational methods. In particular, state-averaged (SA) CASSCF and CASPT2 [10, 204] have been performed with MOLPRO (RS2 contraction level) [329]. Concerning the NEVPT2 calculations (which are also performed with MOLPRO), the partially-contracted (PC) and strongly-contracted (SC) variants have been tested [14, 330, 331]. From a strict theoretical point of view, we point out that PC-NEVPT2 is supposed to be more accurate than SC-NEVPT2 given that it has a larger number of perturbers and greater flexibility. PC-NEVPT2 calculations were also systematically performed for the QUEST#3. In the case of double excitations [36], we have also performed calculations with multi-state (MS) CASPT2 (MS-MR formalism) [332], and its extended variant (XMS-CASPT2) [333] when there is a strong mixing between states with same spin and spatial symmetries. The CASPT2 calculations have been performed with level shift and IPEA parameters set to the standard values of 0.3 and 0.25 a.u., respectively. Large active spaces carefully chosen and tailored for the desired transitions have been selected. The definition of the active space considered for each system as well as the number of states in the state-averaged calculation is provided in their corresponding publication.

Estimating the extrapolation error

In this section, we present our scheme to estimate the extrapolation error in SCI calculations. This new protocol is then applied to five- and six-membered ring molecules for which SCI calculations are particularly challenging even for small basis sets. Note that the present method does only apply to *state-averaged* SCI calculations where ground- and excited-state energies are produced during the same calculation with the same set of molecular orbitals, not to *state-specific* calculations where one computes solely the energy of a single state (like conventional ground-state calculations).

For the m th excited state (where $m = 0$ corresponds to the ground state), we usually estimate its FCI energy $E_{\text{FCI}}^{(m)}$ by performing a linear extrapolation of its variational energy $E_{\text{var}}^{(m)}$ as a function of its rPT2 correction $E_{\text{rPT2}}^{(m)}$ [35, 37] using

$$E_{\text{var}}^{(m)} \approx E_{\text{FCI}}^{(m)} - \alpha^{(m)} E_{\text{rPT2}}^{(m)}, \quad (4.1)$$

where $E_{\text{var}}^{(m)}$ and $E_{\text{rPT2}}^{(m)}$ are calculated with CIPSI and $E_{\text{FCI}}^{(m)}$ is the FCI energy to be extrapolated. This relation is valid in the regime of a sufficiently large number of determinants where the second-order perturbational correction largely dominates. However, in practice, due to the residual higher-order terms, the coefficient $\alpha^{(m)}$ deviates slightly from unity.

Using Eq. (4.1) the estimated error on the CIPSI energy is calculated as

$$E_{\text{CIPSI}}^{(m)} - E_{\text{FCI}}^{(m)} = \left(E_{\text{var}}^{(m)} + E_{\text{rPT2}}^{(m)} \right) - E_{\text{FCI}}^{(m)} = \left(1 - \alpha^{(m)} \right) E_{\text{rPT2}}^{(m)}, \quad (4.2)$$

and thus the extrapolated excitation energy associated with the m th state is given by

$$\Delta E_{\text{FCI}}^{(m)} = \left[E_{\text{var}}^{(m)} + E_{\text{rPT2}} + \left(\alpha^{(m)} - 1 \right) E_{\text{rPT2}} \right] - \left[E_{\text{var}}^{(0)} + E_{\text{rPT2}} + \left(\alpha^{(0)} - 1 \right) E_{\text{rPT2}} \right] + \mathcal{O} \left[E_{\text{rPT2}}^2 \right], \quad (4.3)$$

which evidences that the error in $\Delta E_{\text{FCI}}^{(m)}$ can be expressed as $\left(\alpha^{(m)} - \alpha^{(0)} \right) E_{\text{rPT2}} + \mathcal{O} \left[E_{\text{rPT2}}^2 \right]$.

Now, for the largest systems considered here, $|E_{\text{rPT2}}|$ can be as large as ≈ 2 eV and, thus, the accuracy of the excitation energy estimates strongly depends on our ability to compensate the errors in the calculations. Here, we greatly enhance the compensation of errors by making use of

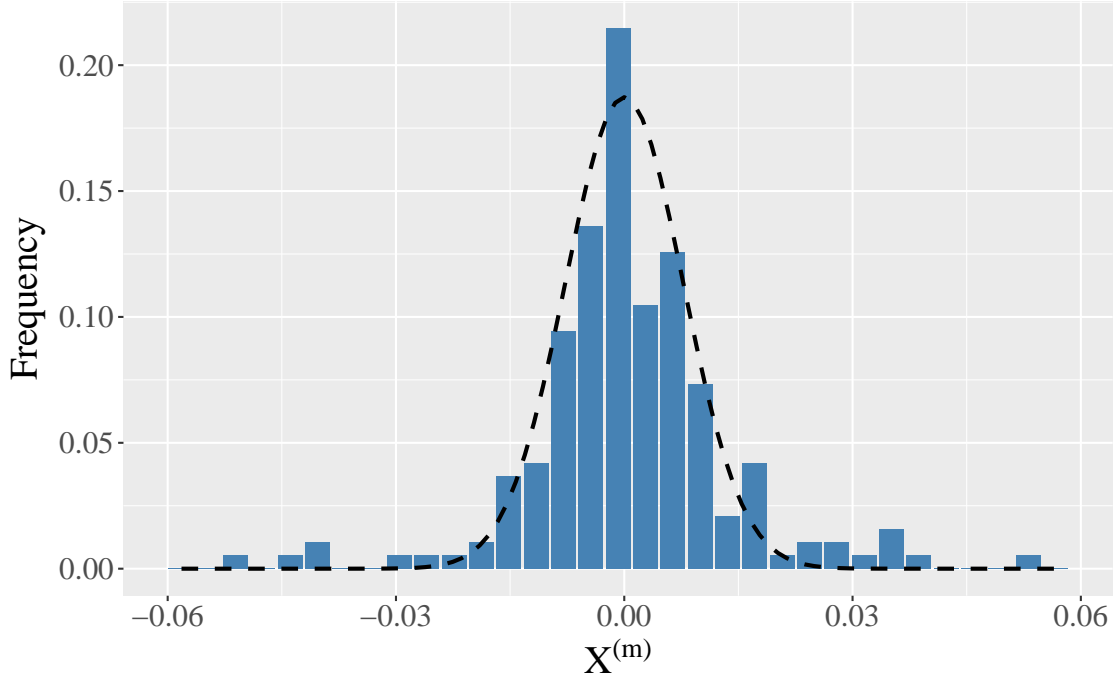


FIGURE 4.2: Histogram of the random variable $X^{(m)}$ (see, text). About 200 values of the transition energies for the 13 five- and six-membered ring molecules, both for the singlet and triplet transitions and for a number of CIPSI iterations, are used. The number M of iterations kept is chosen according to the statistical test presented in the text.

our selection procedure ensuring that the PT2 values of both states match as well as possible (a trick known as PT2 matching [288, 289]), *i.e.* $E_{\text{rPT2}} = E_{\text{rPT2}}^{(0)} \approx E_{\text{rPT2}}^{(m)}$, and by using a common set of state-averaged natural orbitals with equal weights for the ground and excited states. This last feature tends to make the values of $\alpha^{(0)}$ and $\alpha^{(m)}$ very close to each other, such that the error on the energy difference is decreased. In the ideal case where we would be able to fully correlate the CIPSI calculations associated with the ground and excited states, the fluctuations of $\Delta E_{\text{CIPSI}}^{(m)}(n)$ as a function of n would completely vanish and the exact excitation energy would be obtained from the first CIPSI iterations. Quite remarkably, in practice, numerical experience shows that the fluctuations with respect to the extrapolated value $\Delta E_{\text{FCI}}^{(m)}$ are small, zero-centered, almost independent of n when not too close iteration numbers are considered, and display a Gaussian-like distribution. In addition, as stated just above, the fluctuations are found to be (very weakly) dependent on the iteration number n (see Figure 4.2), so this dependence will not significantly alter our results and will not be considered here. We thus introduce the following random variable

$$X^{(m)} = \frac{\Delta E_{\text{CIPSI}}^{(m)}(n) - \Delta E_{\text{FCI}}^{(m)}}{\sigma(n)} \quad (4.4)$$

where

$$\Delta E_{\text{CIPSI}}^{(m)}(n) = \left[E_{\text{var}}^{(m)}(n) + E_{\text{rPT2}}^{(m)}(n) \right] - \left[E_{\text{var}}^{(0)}(n) + E_{\text{rPT2}}^{(0)}(n) \right], \quad (4.5)$$

and $\sigma(n)$ is a quantity proportional to the average fluctuations of $\Delta E_{\text{CIPSI}}^{(m)}$. A natural choice for $\sigma^2(n)$, playing here the role of a variance, is

$$\sigma^2(n) \propto \left[E_{\text{rPT2}}^{(m)}(n) \right]^2 + \left[E_{\text{rPT2}}^{(0)}(n) \right]^2, \quad (4.6)$$

which vanishes in the large- n limit as it should.

The histogram of $X^{(m)}$ resulting from the excitation energies obtained at different values of the CIPSI iterations n and for the 13 five- and six-membered ring molecules, both for the singlet and triplet transitions, is shown in Figure 4.2. To avoid transient effects, only excitation energies at sufficiently large n are retained in the data set. The criterion used to decide from which precise value of n the data should be kept will be presented below. In our application, the total number of values employed to make the histogram is about 200. The dashed line of Figure 4.2 represents the best Gaussian fit (in the sense of least-squares) reproducing the data. As seen, the distribution can be described by the Gaussian probability

$$P[X^{(m)}] \propto \exp\left[-\frac{X^{(m)2}}{2\sigma^{*2}}\right] \quad (4.7)$$

where σ^{*2} is some “universal” variance depending only on the way the correlated selection of both states is done, not on the molecule considered in our set.

An estimate of $\Delta E_{\text{FCI}}^{(m)}$ as the average excitation energy of $\Delta E_{\text{CIPSI}}^{(m)}$ is thus

$$\Delta E_{\text{FCI}}^{(m)} = \frac{\sum_{n=1}^M \frac{\Delta E_{\text{CIPSI}}^{(m)}(n)}{\sigma(n)}}{\sum_{n=1}^M \frac{1}{\sigma(n)}},$$

where M is the number of data kept. Now, regarding the estimate of the error on $\Delta E_{\text{FCI}}^{(m)}$ some caution is required since, although the distribution is globally Gaussian-like (see Figure 4.2) there exists some significant departure from it and we need to take this feature into account.

More precisely, we search for a confidence interval \mathcal{I} such that the true value of the excitation energy $\Delta E_{\text{FCI}}^{(m)}$ lies within one standard deviation of $\Delta E_{\text{CIPSI}}^{(m)}$, i.e., $P(\Delta E_{\text{FCI}}^{(m)} \in [\Delta E_{\text{CIPSI}}^{(m)} \pm \sigma] \mid \mathcal{G}) = 0.6827$. In a Bayesian framework, the probability that $\Delta E_{\text{FCI}}^{(m)}$ is in an interval \mathcal{I} is

$$P(\Delta E_{\text{FCI}}^{(m)} \in \mathcal{I}) = P(\Delta E_{\text{FCI}}^{(m)} \in \mathcal{I} \mid \mathcal{G}) \times P(\mathcal{G}) \quad (4.8)$$

where $P(\mathcal{G})$ is the probability that the random variables considered in the latest CIPSI iterations are normally distributed. A common test in statistics of the normality of a distribution is the Jarque-Bera test J and we have

$$P(\mathcal{G}) = 1 - \chi_{\text{CDF}}^2(J, 2) \quad (4.9)$$

where $\chi_{\text{CDF}}^2(x, k)$ is the cumulative distribution function (CDF) of the χ^2 -distribution with k degrees of freedom. As the number of samples M is usually small, we use Student’s t -distribution to estimate the statistical error. The inverse of the cumulative distribution function of the t -distribution, t_{CDF}^{-1} , allows us to find how to scale the interval by a parameter

$$\beta = t_{\text{CDF}}^{-1}\left[\frac{1}{2}\left(1 + \frac{0.6827}{P(\mathcal{G})}\right), M\right] \quad (4.10)$$

such that $P(\Delta E_{\text{FCI}}^{(m)} \in [\Delta E_{\text{CIPSI}}^{(m)} \pm \beta\sigma]) = p = 0.6827$. Only the last $M > 2$ computed transition energies are considered. M is chosen such that $P(\mathcal{G}) > 0.8$ and such that the error bar is minimal. If all the values of $P(\mathcal{G})$ are below 0.8, M is chosen such that $P(\mathcal{G})$ is maximal.

The singlet and triplet FCI/6-31+G(d) excitation energies and their corresponding error bars estimated with the method presented above based on Gaussian random variables are reported in Table 4.1. For the sake of comparison, we also report the CC3 and CCSDT vertical energies from Ref. [32] computed in the same basis. We note that there is for the vast majority of considered states a very good agreement between the CC3 and CCSDT values, indicating that the CC values can be trusted. The estimated values of the excitation energies obtained via a three-point linear extrapolation considering the three largest CIPSI wave functions are also gathered in Table 4.1. In

this case, the error bar is estimated via the extrapolation distance, *i.e.*, the difference in excitation energies obtained with the three-point linear extrapolation and the largest CIPSI wave function. This strategy has been considered in some of our previous works [32, 221, 249]. The deviation from the CCSDT excitation energies for the same set of excitations are depicted in Figure 4.3, where the red dots correspond to the excitation energies and error bars estimated via the present method, and the blue dots correspond to the excitation energies obtained via a three-point linear fit and error bars estimated via the extrapolation distance. These results contain a good balance between well-behaved and ill-behaved cases. For example, cyclopentadiene and furan correspond to well-behaved scenarios where the two flavors of extrapolations yield nearly identical estimates and the error bars associated with these two methods nicely overlap. In these cases, one can observe that our method based on Gaussian random variables provides almost systematically smaller error bars. Even in less idealistic situations (like in imidazole, pyrrole, and thiophene), the results are very satisfactory and stable. The six-membered rings represent much more challenging cases for SCI methods, and even for these systems the newly-developed method provides realistic error bars, and allows to easily detect problematic events (like pyridine for instance). The present scheme has also been tested on smaller systems when one can tightly converge the CIPSI calculations. In such cases, the agreement is nearly perfect in every scenario that we have encountered.

4.3 The QUEST database

4.3.1 Overview

The QUEST database gathers more than 500 highly-accurate excitation energies of various natures (valence, Rydberg, $n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$, singlet, doublet, triplet, and double excitations) for molecules ranging from diatomics to molecules as large as naphthalene (see Figure 4.4). This set is also chemically diverse, with organic and inorganic systems, open- and closed-shell compounds, acyclic and cyclic systems, pure hydrocarbons and various heteroatomic structures, etc. Each of the five subsets making up the QUEST dataset is detailed below. Throughout the present review, we report several statistical indicators: the mean signed error (MSE), mean absolute error (MAE), root-mean square error (RMSE), and standard deviation of the errors (SDE), as well as the maximum positive [Max(+)] and maximum negative [Max(-)] errors.

4.3.2 QUEST#1

The QUEST#1 benchmark set [30] consists of 110 vertical excitation energies (as well as oscillator strengths) from 18 molecules with sizes ranging from one to three non-hydrogen atoms (water, hydrogen sulfide, ammonia, hydrogen chloride, dinitrogen, carbon monoxide, acetylene, ethylene, formaldehyde, methanimine, thioformaldehyde, acetaldehyde, cyclopropene, diazomethane, formamide, ketene, nitrosomethane, and the smallest streptocyanine). For this set, we provided two sets of TBEs: i) one obtained within the frozen-core approximation and the *aug-cc-pVTZ* basis set, and ii) another one including further corrections for basis set incompleteness and “all electron” effects. For the former set, we systematically employed FCI/*aug-cc-pVTZ* values to define our TBEs, except for a few cases. For the latter set, both the “all electron” correlation and the basis set corrections were systematically obtained at the CC3 level of theory and with the *d-aug-cc-pV5Z* basis for the nine smallest molecules, and slightly more compact basis sets for the larger compounds. Our TBE/*aug-cc-pVTZ* reference excitation energies were employed to benchmark a series of popular excited-state wave function methods partially or fully accounting for double and triple excitations, namely CIS(D), CC2, CCSD, STEOM-CCSD, CCSDR(3), CCSDT-3, CC3, ADC(2), and ADC(3). Our main conclusions were that i) ADC(2) and CC2 show strong similarities in terms of accuracy, ii) STEOM-CCSD is, on average, as accurate as CCSD, the latter overestimating transition energies, iii) CC3 is extremely accurate (with a mean absolute error of only ~ 0.03 eV) and that although slightly less accurate than CC3, CCSDT-3 could be used as a reliable reference for benchmark studies,

TABLE 4.1: Singlet and triplet excitation energies (in eV) obtained at the CC3, CCSDT, and CIPSI levels of theory with the 6-31+G(d) basis set for various five- and six-membered rings.

Molecule	Transition	CC3	CCSDT	CIPSI (Gaussian) ^a	CIPSI (3-point) ^b
Five-membered rings					
Cyclopentadiene	$^1B_2(\pi \rightarrow \pi^*)$	5.79	5.80	5.80(2)	5.79(2)
	$^3B_2(\pi \rightarrow \pi^*)$	3.33	3.33	3.32(4)	3.29(7)
Furan	$^1A_2(\pi \rightarrow 3s)$	6.26	6.28	6.31(5)	6.37(1)
	$^3B_2(\pi \rightarrow \pi^*)$	4.28	4.28	4.26(4)	4.22(7)
Imidazole	$^1A''(\pi \rightarrow 3s)$	5.77	5.77	5.78(5)	5.96(14)
	$^3A'(\pi \rightarrow \pi^*)$	4.83	4.81	4.82(7)	4.65(22)
Pyrrole	$^1A_2(\pi \rightarrow 3s)$	5.25	5.25	5.23(7)	5.31(1)
	$^3B_2(\pi \rightarrow \pi^*)$	4.59	4.58	4.54(7)	4.37(23)
Thiophene	$^1A_1(\pi \rightarrow \pi^*)$	5.79	5.77	5.75(8)	5.73(9)
	$^3B_2(\pi \rightarrow \pi^*)$	3.95	3.94	3.98(1)	3.99(2)
Six-membered rings					
Benzene	$^1B_{2u}(\pi \rightarrow \pi^*)$	5.13	5.10	5.06(9)	5.21(7)
	$^3B_{1u}(\pi \rightarrow \pi^*)$	4.18	4.16	4.28(6)	4.17(7)
Cyclopentadienone	$^1A_2(n \rightarrow \pi^*)$	3.03	3.03	3.08(2)	3.13(3)
	$^3B_2(\pi \rightarrow \pi^*)$	2.30	2.32	2.37(5)	2.10(25)
Pyrazine	$^1B_{3u}(n \rightarrow \pi^*)$	4.28	4.28	4.26(9)	4.10(25)
	$^3B_{3u}(n \rightarrow \pi^*)$	3.68	3.68	3.70(3)	3.70(1)
Tetrazine	$^1B_{3u}(n \rightarrow \pi^*)$	2.53	2.54	2.56(5)	5.07(16)
	$^3B_{3u}(n \rightarrow \pi^*)$	1.87	1.88	1.91(3)	4.04(49)
Pyridazine	$^1B_1(n \rightarrow \pi^*)$	3.95	3.95	3.97(10)	3.60(43)
	$^3B_1(n \rightarrow \pi^*)$	3.27	3.26	3.27(15)	3.46(14)
Pyridine	$^1B_1(n \rightarrow \pi^*)$	5.12	5.10	5.15(12)	4.90(24)
	$^3A_1(\pi \rightarrow \pi^*)$	4.33	4.31	4.42(85)	3.68(105)
Pyrimidine	$^1B_1(n \rightarrow \pi^*)$	4.58	4.57	4.64(11)	2.54(5)
	$^3B_1(n \rightarrow \pi^*)$	4.20	4.20	4.55(37)	2.18(27)
Triazine	$^1A''_1(n \rightarrow \pi^*)$	4.85	4.84	4.77(13)	5.12(51)
	$^3A''_2(n \rightarrow \pi^*)$	4.40	4.40	4.45(39)	4.73(6)

^a Excitation energies and error bars estimated via the novel statistical method based on Gaussian random variables (see section 4.2.3). The error bars reported in parenthesis correspond to one standard deviation.

^b Excitation energies obtained via a three-point linear fit using the three largest CIPSI variational wave functions, and error bars estimated via the extrapolation distance, *i.e.*, the difference in excitation energies obtained with the three-point linear extrapolation and the largest CIPSI wave function.

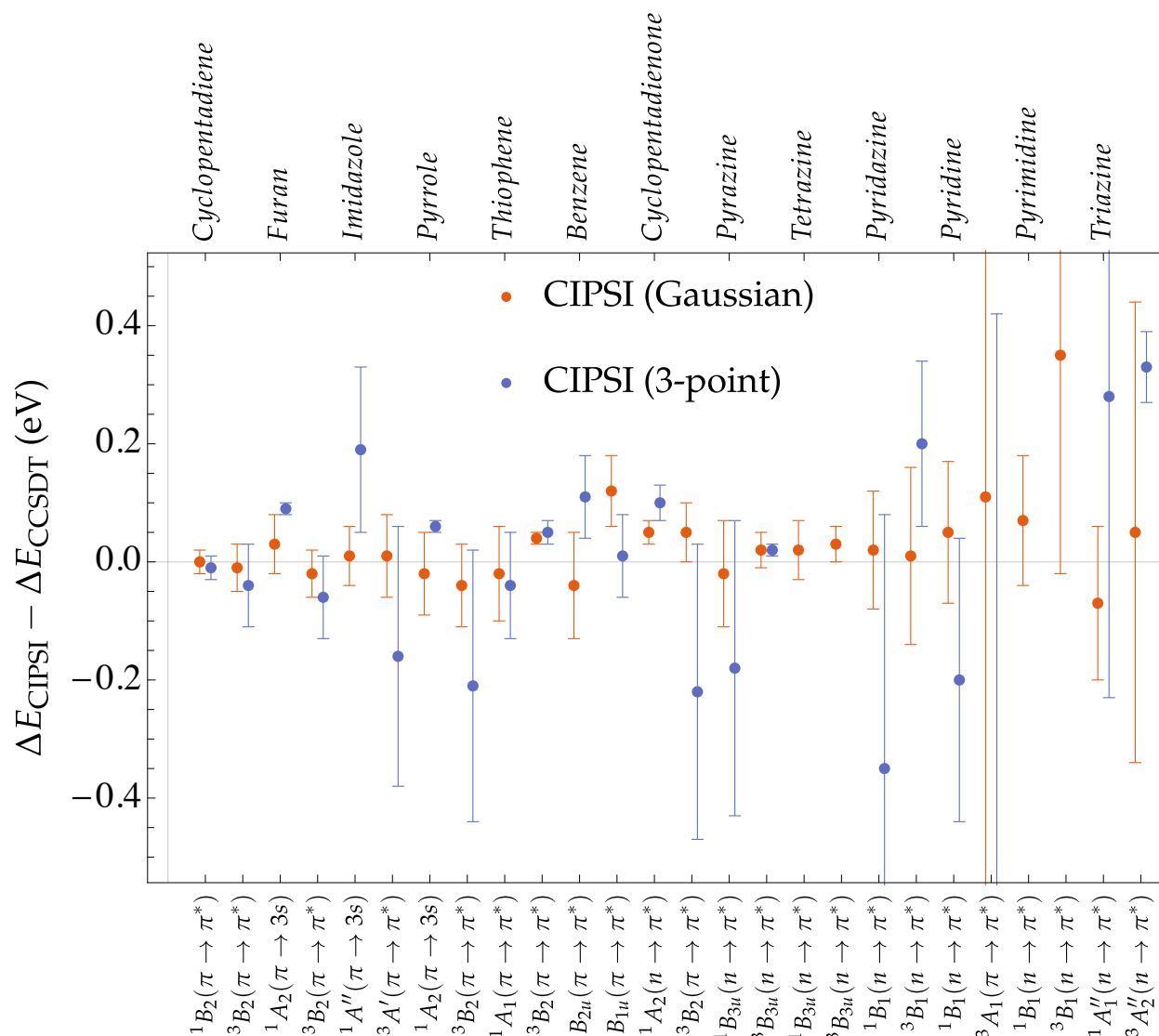
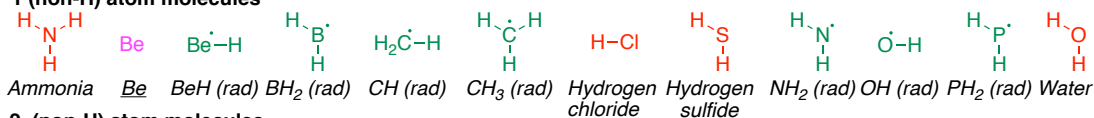
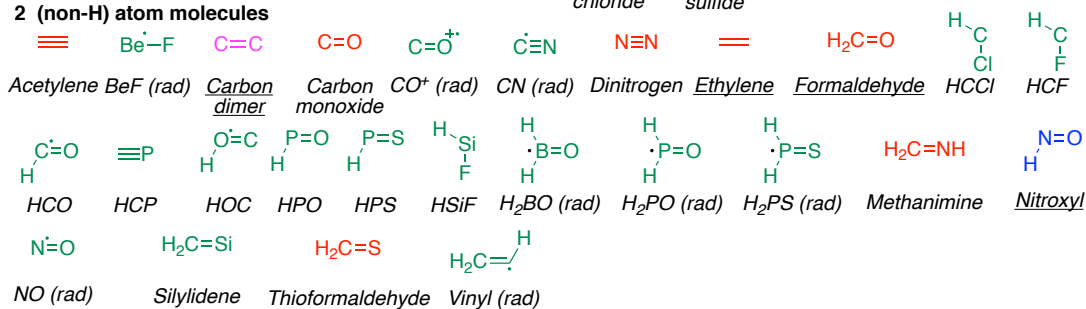


FIGURE 4.3: Deviation from the CCSDT excitation energies for the lowest singlet and triplet excitation energies (in eV) of five- and six-membered rings obtained at the CIPSI/6-31+G(d) level of theory. Red dots: excitation energies and error bars estimated via the present method (see section 4.2.3). Blue dots: excitation energies obtained via a three-point linear fit using the three largest CIPSI wave functions, and error bars estimated via the extrapolation distance, *i.e.*, the difference in excitation energies obtained with the three-point linear extrapolation and the largest CIPSI wave function.

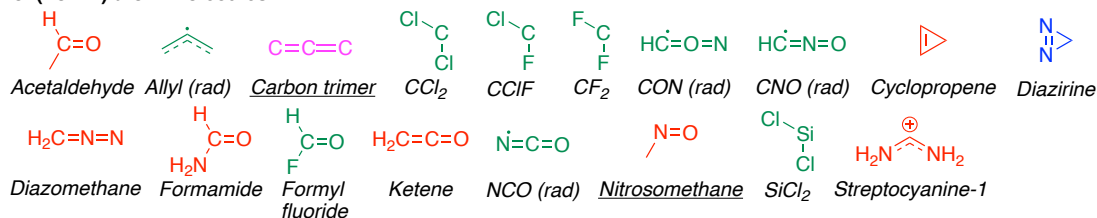
1 (non-H) atom molecules



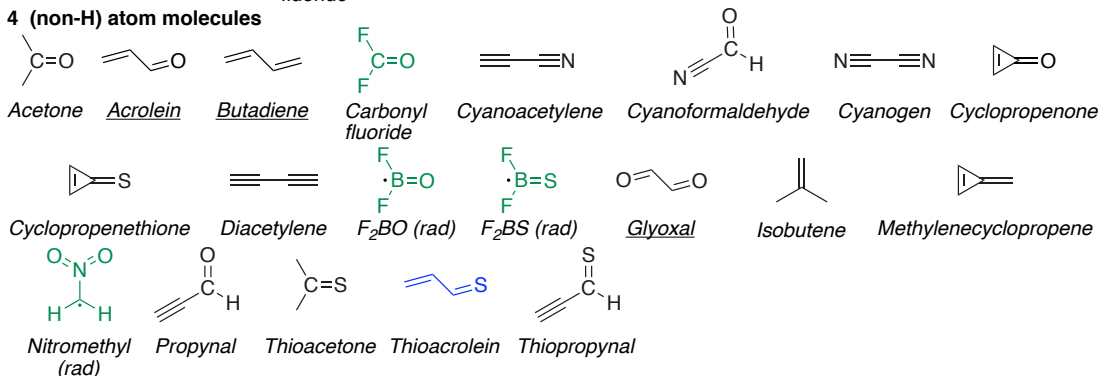
2 (non-H) atom molecules



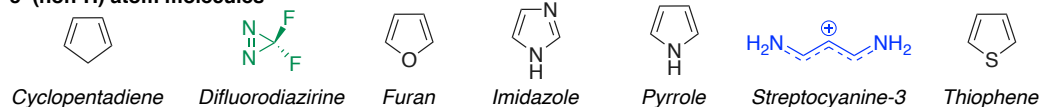
3 (non-H) atom molecules



4 (non-H) atom molecules



5 (non-H) atom molecules



6-10 (non-H) atom molecules

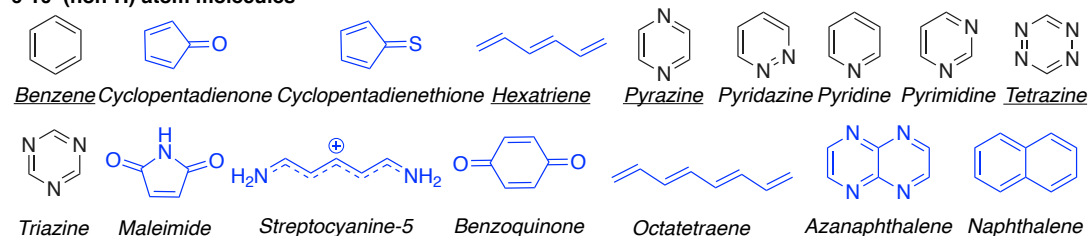


FIGURE 4.4: Molecules from each of the five subsets making up the present QUEST dataset of highly-accurate vertical excitation energies: QUEST#1 (red), QUEST#2 (magenta and/or underlined), QUEST#3 (black), QUEST#4 (green), and QUEST#5 (blue).

and iv) ADC(3) was found to be significantly less accurate than CC3 by overcorrecting ADC(2) excitation energies.

4.3.3 QUEST#2

The QUEST#2 benchmark set [36] reports reference energies for double excitations. This set gathers 20 vertical transitions from 14 small- and medium-sized molecules (acrolein, benzene, beryllium atom, butadiene, carbon dimer and trimer, ethylene, formaldehyde, glyoxal, hexatriene, nitrosomethane, nitroxyl, pyrazine, and tetrazine). The TBEs of the QUEST#2 set are obtained with SCI and/or multiconfigurational [CASSCF, CASPT2, (X)MS-CASPT2, and NEVPT2] calculations depending on the size of the molecules and the level of theory that we could afford. An important addition to this second study was also the inclusion of various flavors of multiconfigurational methods (CASSCF, CASPT2, and NEVPT2) in addition to high-order CC methods including, at least, perturbative triples (CC3, CCSDT, CCSDTQ, etc). Our results demonstrated that the error of CC methods is intimately linked to the amount of double-excitation character in the vertical transition. For “pure” double excitations (*i.e.*, for transitions which do not mix with single excitations), the error in CC3 and CCSDT can easily reach 1 and 0.5 eV, respectively, while it goes down to a few tenths of an eV for more common transitions involving a significant amount of single excitations (such as the well-known A_g transition in butadiene or the E_{2g} excitation in benzene). The quality of the excitation energies obtained with CASPT2 and NEVPT2 was harder to predict as the overall accuracy of these methods is highly dependent on both the system and the selected active space. Nevertheless, these two methods were found to be more accurate for transitions with a very small percentage of single excitations (error usually below 0.1 eV) than for excitations dominated by single excitations where the error is closer to 0.1–0.2 eV.

4.3.4 QUEST#3

The QUEST#3 benchmark set [32] is, by far, our largest set, and consists of highly accurate vertical transition energies and oscillator strengths obtained for 27 molecules encompassing 4, 5 and 6 non-hydrogen atoms (acetone, acrolein, benzene, butadiene, cyanoacetylene, cyanoformaldehyde, cyanogen, cyclopentadiene, cyclopropenone, cyclopropenethione, diacetylene, furan, glyoxal, imidazole, isobutene, methylenecyclopropene, propynal, pyrazine, pyridazine, pyridine, pyrimidine, pyrrole, tetrazine, thioacetone, thiophene, thiopropynal, and triazine) for a total of 238 vertical transition energies and 90 oscillator strengths with a reasonably good balance between singlet, triplet, valence, and Rydberg excited states. For these 238 transitions, we have estimated that 224 are chemically accurate for the *aug-cc-pVTZ* basis and for the considered geometry. To define the TBEs of the QUEST#3 set, we employed CC methods up to the highest technically possible order (CC3, CCSDT, and CCSDTQ), and, when affordable SCI calculations with very large reference spaces (up to hundred million determinants in certain cases), as well as one of the most reliable multiconfigurational methods, NEVPT2, for double excitations. Most of our TBEs are based on CCSDTQ (4 non-hydrogen atoms) or CCSDT (5 and 6 non-hydrogen atoms) excitation energies. For all the transitions of the QUEST#3 set, we reported at least CCSDT/*aug-cc-pVTZ* (sometimes with basis set extrapolation) and CC3/*aug-cc-pVQZ* transition energies as well as CC3/*aug-cc-pVTZ* oscillator strengths for each dipole-allowed transition. Pursuing our previous benchmarking efforts, we confirmed that CC3 almost systematically delivers transition energies in agreement with higher-level theoretical models (± 0.04 eV) except for transitions presenting a dominant double-excitation character where multiconfigurational methods like NEVPT2 have logically the edge. This settles down, at least for now, the debate by demonstrating the superiority of CC3 (in terms of accuracy) compared to methods like CCSDT-3 or ADC(3). For the latter model, this was further demonstrated in a recent study by two of the present authors [33].

4.3.5 QUEST#4

The QUEST#4 benchmark set [221] consists of two subsets of excitations and oscillator strengths. An “exotic” subset of 30 excited states for closed-shell molecules containing F, Cl, P, and Si atoms (carbonyl fluoride, CCl_2 , CClF , CF_2 , difluorodiazirine, formyl fluoride, HCCl , HCF , HCP , HPO , HPS , HSiF , SiCl_2 , and silylidene) and a “radical” subset of 51 doublet-doublet transitions in 24 small radicals (allyl, BeF , BeH , BH_2 , CH , CH_3 , CN , CNO , CON , CO^+ , F_2BO , F_2BS , H_2BO , HCO , HOC , H_2PO , H_2PS , NCO , NH_2 , nitromethyl, NO , OH , PH_2 , and vinyl) characterized by open-shell electronic configurations and an unpaired electron. This represents a total of 81 high-quality TBEs, the vast majority being obtained at the FCI level with at least the *aug-cc-pVTZ* basis set. We additionally performed high-order CC calculations to ascertain these estimates. For the exotic set, these TBEs have been used to assess the performances of 15 “lower-order” wave function approaches, including several CC and ADC variants. Consistent with our previous works, we found that CC3 is very accurate, whereas the trends for the other methods are similar to that obtained on more standard CNOSH organic compounds. In contrast, for the radical set, even the refined ROCC3 method yields a comparatively large MAE of 0.05 eV. Likewise, the excitation energies obtained with CCSD are much less satisfying for open-shell derivatives (MAE of 0.20 eV with UCCSD and 0.15 eV with ROCCSD) than for closed-shell systems of similar size (MAE of 0.07 eV).

4.3.6 QUEST#5

The QUEST#5 subset is composed of additional accurate excitation energies that we have produced for the present chapter. This new set gathers 13 new systems composed by small molecules as well as larger molecules (see blue molecules in Figure 4.4): aza-naphthalene, benzoquinone, cyclopentadienone, cyclopentadienethione, diazirine, hexatriene, maleimide, naphthalene, nitroxyl, octatetraene, streptocyanine-C3, streptocyanine-C5, and thioacrolein. For these new transitions, we report again quality vertical transition energies, the vast majority being of CCSDT quality, and we consider that, out of these 80 new transitions, 55 of them can be labeled as “safe”, *i.e.*, considered as chemically accurate or within 0.05 eV of the FCI limit for the given geometry and basis set.

Aza-naphthalene In contrast to naphthalene (see below), its tetraaza counterpart (1,4,5,8-tetraazanaphthalene) has not been much investigated although it also has a D_{2h} symmetry. The vibronic couplings of one low-lying state were nevertheless well characterized theoretically by Dierksen and Grimme with TD-DFT [334] and compared to the experimental spectrum [335]. The latter work also contains some CIS and CNDO calculations and a few assignments for higher-lying states. Our CC results collected in Table 4.2 are therefore clearly the most advanced to date. For the singlet transitions, no $\%T_1$ is smaller than 80%, and we have obtained consistent CC3 and CCSDT values with the Pople basis set. Indeed, the two models yield values within ± 0.03 eV of each other, the two exceptions (the second B_{1u} and the ${}^1B_{3u}$ states) being considered as “unsafe” in the database. Comparisons to experimental 0–0 energies in condensed medium and CNDO calculations can be found in the Table, but are not very helpful to assess our TBEs. To the best of our knowledge, the present work is the first to report triplet excited states, and we list in Table 4.2 eight valence transitions obtained at the CC3/*aug-cc-pVTZ* level. As we were not able to perform CCSDT calculations for the triplets, all these transition energies are labeled “unsafe” in the QUEST database. Nevertheless, given the large $\%T_1$ values, one can likely consider them accurate (for the basis set used at least).

Benzoquinone Benzoquinone, the simplest quinoidic dye, has been treated at several levels of theory previously, *e.g.*, CASPT2 [196, 198, 336, 337], SAC-CI [338–340] and various CC levels up to CC3 [196, 199], ADC(2) and ADC(3) [29, 33] as well as TD-DFT [197, 339–343]. Our results and comparisons with a selection of the existing literature can be found in Table 4.3.

For the singlet transitions, we could obtain CCSDT/*aug-cc-pVDZ* values for the 10 considered excited states. For the two lowest transitions of $n \rightarrow \pi^*$ character, these agree well with the

TABLE 4.2: Transition energies (in eV) determined in aza-naphthalene (1,4,5,8-tetraazanaphthalene). For all transitions, we provide the single-excitation character $\%T_1$ obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	<i>f</i>	$\%T_1$	CC3			CCSDT		Litt.	
				6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	6-31+G(d)	Th. ^a	Exp. ^b	
Aza-naphthalene	Nature									
	¹ B _{3g} (<i>n</i> → π^*)		88.5	3.26	3.15	3.11	3.29	2.64	2.61	
	¹ B _{2u} (π → π^*)	0.190	86.5	4.37	4.32	4.28	4.37	4.05	3.86	
	¹ B _{1u} (<i>n</i> → π^*)	n.d.	88.5	4.47	4.39	4.34	4.47	3.32		
	¹ B _{2g} (<i>n</i> → π^*)		87.3	4.62	4.55	4.53	4.64	4.48		
	¹ B _{2g} (<i>n</i> → π^*)		84.1	5.00	4.91	4.86	5.03	4.76		
	¹ B _{1u} (<i>n</i> → π^*)	n.d.	82.6	5.31	5.17	5.13	5.42			
	¹ A _u (<i>n</i> → π^*)		83.1	5.47	5.40	5.34	5.47	4.51		
	¹ B _{3u} (π → π^*)	0.028	88.5	5.86	5.69	5.63	5.91	5.26	5.03	
	¹ A _g (π → π^*)		85.3	5.96	5.87	5.81	5.95			
	¹ A _u (<i>n</i> → π^*)		84.8	5.97	5.90	5.89	6.00			
	¹ A _g (<i>n</i> → 3s)		90.5	6.56	6.39	6.49	6.57			
	³ B _{3g} (<i>n</i> → π^*)		96.5	2.93	2.84	2.82				
	³ B _{2u} (π → π^*)		97.2	3.86	3.73	3.67				
	³ B _{3u} (π → π^*)		97.7	3.78	3.76	3.75				
	³ B _{1u} (<i>n</i> → π^*)		97.1	3.85	3.79	3.77				
	³ B _{2g} (<i>n</i> → π^*)		96.2	4.43	4.37	4.34				
	³ B _{2g} (<i>n</i> → π^*)		95.3	4.71	4.63	4.61				
	³ B _{3u} (π → π^*)		96.6	4.82	4.78	4.75				
	³ A _u (<i>n</i> → π^*)		96.6	4.96	4.91	4.87				

^a CNDO/S values from Ref. [335].

^b 0–0 energy in solution estimated in Ref. [335]. There are other tentative assignments in that work.

corresponding CC3 values and we can define safe TBE/*aug-cc-pVTZ* of 2.82 and 2.96 eV. These values are within 0.10 eV of the most recent estimates of Thiel [199] and are clearly larger than older CASPT2 estimates [336, 337], which appear too low. The experimentally available data return 0–0 energies of roughly 2.5 eV for both transitions [344, 345]. These values are located 0.3–0.4 eV below our vertical estimates which is a quite reasonable difference between vertical and 0–0 energies. The next transition of A_g symmetry has a pure double excitation character so that unsurprisingly both CC3 and CCSDT yield values that are much too large. Indeed, at the NEVPT2/*aug-cc-pVTZ* level, we obtain a transition energy of 4.57 eV for this excitation, more than 1 eV below the CC3 estimate, yet again slightly above earlier CASPT2 estimates. Although one cannot consider this 4.57 eV estimate as chemically accurate (the error bar of NEVPT2 is typically ± 0.10 eV for transitions of pure double character [36]), it is likely the most accurate value available at this stage. The next transition of ¹B_{3g} symmetry is the first of π → π^* character. Although its associated $\%T_1$ value is rather large, there is substantial difference between CC3 and CCSDT, making our TBE of 4.58 eV falling in the “unsafe” category, though it is obviously more accurate than previous CASPT2 values that are too close from the experimental 0–0 energies to be trustworthy. For the next strongly-allowed transition (¹B_{1u}), one also notices small yet non-negligible differences between CC3 and CCSDT. Our TBE of 5.62 eV is slightly larger than Thiel’s one (5.47 eV obtained using CC3) [198], It is objectively hard to determined which one is the most accurate, and if the difference in the ground-state geometries (CC3 here *vs* MP2 in Thiel’s work) also plays a significant role in this discrepancy. The situation is similar for the B_{3u} transition, although in that case the present TBE of 5.79 eV is close to Thiel’s CC3 value of 5.71 eV. Note that Thiel selected a CASPT2 value of 5.55 eV as TBE due to the rather small $\%T_1$ value for that state. However, this value is likely slightly too low as the agreement between CC3 and CCSDT is rather good. For the higher-lying singlet transitions, we note that: i) we could produce “safe” TBEs for the two B_{2g} excited states that both show minimal changes between CC3 and CCSDT, although $\%T_1$ is small for the lowest transition of that symmetry; ii) for the (second) ¹A_u transition, the differences are too large between CC3 and CCSDT to provide trustworthy estimates; iii) for the (second) ¹B_{1g} excitation, these differences are less marked and although we rated our TBE (6.38 eV) as “unsafe”, it is likely the best estimate proposed to date in the literature.

For the triplets, we considered the four lowest transitions, two of *n* → π^* character (³B_{1g} and

3A_u) and two of $\pi \rightarrow \pi^*$ character ($^3B_{1u}$ and $^3B_{3g}$), see Table 4.3. For all four excited states, there is a very nice match between the CC3 and CCSDT transition energies obtained with Pople's basis set, and the single-excitation characters are very large, so that we are confident that our TBEs are trustworthy. For the $^3B_{1g}$ and 3A_u transitions, our energies are very slightly larger than Thiel's extrapolated CC3 ones, and again ca. 0.3–0.4 eV above the experimental 0–0 energies. As for the singlet transitions, the early CASPT2 values are too low. For the two $\pi \rightarrow \pi^*$ excitations, exactly the same trends are found, but no experimental measurements exist to our knowledge.

TABLE 4.3: Transition energies (in eV) determined in *p*-benzoquinone. For all transitions, we provide the single-excitation character % T_1 obtained at LR-CC3/*aug*-cc-pVTZ level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	CC3			CCSDT			Litt.								
		<i>f</i>	% T_1	6-31+G(d)	<i>aug</i> -cc-pVDZ	<i>aug</i> -cc-pVTZ	6-31+G(d)	<i>aug</i> -cc-pVDZ	Th. ^a	Th. ^b	Th. ^c	Exp. ^d	Exp. ^e	Exp. ^f	Exp. ^g	
Benzoquinone	Nature															
	$^1B_{1g}(n \rightarrow \pi^*)$		85.3	2.85	2.81	2.79	2.87	2.84	2.50	2.39	2.74			2.52	2.49	
	$^1A_u(n \rightarrow \pi^*)$		84.1	2.99	2.95	2.94	3.01	2.97	2.50	2.43	2.86			2.49	2.48	
	$^1A_g(n, n \rightarrow \pi^*, \pi^*)$		0.0	5.92	5.94	6.02	5.79	5.84	4.41	4.36						
	$^1B_{3g}(\pi \rightarrow \pi^*)$		88.6	4.66	4.58	4.53	4.71	4.63	4.19	4.01	4.44	4.3	4.09	4.07		
	$^1B_{1u}(\pi \rightarrow \pi^*)$	0.471	88.4	5.71	5.63	5.58	5.75	5.67	5.15	5.09	5.47	5.38–5.70	5.08	5.12		
	$^1B_{3u}(n \rightarrow \pi^*)$	0.001	79.8	5.95	5.77	5.75	5.96	5.81	5.15	4.91	5.71					
	$^1B_{2g}(n \rightarrow \pi^*)$		76.2	6.11	5.96	5.94	6.10	5.97	4.80	4.99						
	$^1A_u(n \rightarrow \pi^*)$		74.8	6.41	6.29	6.27	6.46	6.37	5.79	5.47						
	$^1B_{1g}(n \rightarrow \pi^*)$		83.5	6.48	6.37	6.34	6.51	6.41	5.76	5.68						
	$^1B_{2g}(n \rightarrow \pi^*)$		86.6	7.33	7.28	7.20	7.33	7.30	5.49	5.62						
	$^3B_{1g}(n \rightarrow \pi^*)$		96.0	2.61	2.56	2.56	2.63		2.17	2.16	2.50			2.31	2.31	
	$^3A_u(n \rightarrow \pi^*)$		95.6	2.76	2.71	2.71	2.77		2.27	2.22	2.61			2.35	2.31	
	$^3B_{1u}(\pi \rightarrow \pi^*)$		97.7	3.13	3.16	3.16	3.14	3.11	2.91	2.57	3.02					
	$^3B_{3g}(\pi \rightarrow \pi^*)$		97.9	3.46	3.46	3.46	3.44	3.48	3.19	3.09	3.37					

^a CASPT2 values from Ref. [336].

^b CASPT2 values from Ref. [337].

^c CC3 (extrapolated to *aug*-cc-pVTZ) values from Ref. [198].

^d EELS from Ref. [343].

^e Absorption spectroscopy (0–0 energies) from Ref. [346].

^f Absorption spectroscopy (0–0 energies) from Ref. [344] (gas-phase and pure crystals).

^g Absorption spectroscopy (0–0 energies) from Ref. [345].

Cyclopentadienone and cyclopentadienethione This two five-membered rings with an external (thio)ketone moiety have been investigated theoretically by Serrano-Andrés and coworkers in 2002 [347] who used the best method available at the time, namely CASPT2 (without IPEA). Our results are compared to these earlier estimates in Table 4.4. For both structures and both spin symmetries, the two lowest singlet transitions are of $A_2(n \rightarrow \pi^*)$ and $B_2(\pi \rightarrow \pi^*)$ spatial symmetries. There is a good to excellent agreement between the CC3 and CCSDT values for both the 6-31+G(d) and the *aug-cc-pVDZ* basis sets for these eight excited states, consistent with the large single excitation character. Thus, one can likely trust the obtained TBEs as these transitions are unproblematic. Experimentally, a *t*Bu substituted cyclopentadienone shows a weakly-allowed band peaking at 3.22 eV in vapour [348], which is likely the 1B_2 state. For the lowest triplet state of cyclopentadienone, the experimental triplet energy was (indirectly) estimated experimentally to be 1.50(2) eV [349], but this value corresponds to the triplet lowest-energy geometry, so that direct comparisons with our data is unreasonable. This 2014 work also contains CCSD(T) estimates of the adiabatic electron affinities for the two lowest triplet states.

In the singlet manifold of this molecule, one next finds one dark transition of purely double $(n, \pi) \rightarrow (\pi^*, \pi^*)$ character, for which CC theory is not well suited, as clearly illustrated by the huge difference between CC3 and CCSDT. Using a minimal active space (π space and lone pairs), we obtained with NEVPT2 value of 5.02 eV, *i.e.*, 0.7 eV below the CCSDT estimate and roughly half an eV higher than the earlier CASPT2 values. This 5.02 eV estimate although not chemically accurate is likely the best available today. The fourth singlet state is an interesting yet challenging $(\pi, \pi) \rightarrow (\pi^*, \pi^*)$ showing a % T_1 of 49.9%. For this particular state, the NEVPT2 value is 6.02 eV, which is likely again the most realistic value available. The fifth transition shows a butadiene- A_g -like character, that is a totally-symmetric $\pi \rightarrow \pi^*$ transition with a significant double excitation character around 25%. For this transition, we based our TBE on the CCSDT estimate, but it might be too large by roughly 0.10 eV. The experimental spectrum of the related compounds shows a strong peak at 5.93 eV [348], likely corresponding to the overlap between these two 1A_1 transitions. In cyclopentadienone, the third triplet is an unproblematic $\pi \rightarrow \pi^*$ transition, for which there is a remarkable consistency between our CC estimates, again significantly above the previous CASPT2 data [347]. Finally, the highest triplet considered has originally a highly dominant multi-excitation character and our best estimate of 4.91 eV was obtained with NEVPT2, the CC values being too large.

For the singlet manifold of cyclopentadienethione, one finds again a $^1B_1(n, \pi \rightarrow \pi^*, \pi^*)$ purely double transition and a 50/50 single/double $^1A_1(\pi, \pi) \rightarrow (\pi^*, \pi^*)$ transition. The methodological trends are similar to those noted for the oxygen-derivatives, and the TBE listed in the QUEST database are obtained with NEVPT2: 3.16 eV and 5.43 eV. For the latter, it should be noted that there are several transitions of mixed character close in energy, so that definitive attribution is challenging. In contrast to cyclopentadienone, the $^1A_1(\pi \rightarrow \pi^*)$ transition shows a very large % T_1 value and small differences between CC3 and CCSDT, so that we have been able to define a “safe” TBE of 4.96 eV. For the records, our NEVPT2 estimate for that state is consistent, 4.90 eV. For the triplets of cyclopentadienethione, the trends are totally similar to those of the oxygen structure. For the fourth triplet of double character, our TBE is 3.13 eV, a value again obtained with NEVPT2.

TABLE 4.4: Transition energies (in eV) determined in cyclopentadienone and cyclopentadienethione. For all transitions, we provide the single-excitation character $\%T_1$ obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	<i>f</i>	$\%T_1$	CC3			CCSDT			Litt. ^a
				6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	<i>aug-cc-pVQZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	
Cyclopentadienone	Nature									
	$^1A_2(n \rightarrow \pi^*)$		88.5	3.03	2.95	2.94	2.95	3.03	2.95	2.48
	$^1B_2(\pi \rightarrow \pi^*)$	0.004	91.2	3.69	3.57	3.54	3.53	3.72	3.61	3.00
	$^1B_1(n, \pi \rightarrow \pi^*, \pi^*)$	0.000	3.1	6.06	6.07	6.12	6.12	5.67	5.69	4.49
	$^1A_1(\pi, \pi \rightarrow \pi^*, \pi^*)$	0.131	49.9	7.20	7.12	7.10	7.08	7.07	6.95	5.42
	$^1A_1(\pi \rightarrow \pi^*)$	0.090	73.6	6.26	6.23	6.21	6.20	6.12	6.11	5.98
	$^3B_2(\pi \rightarrow \pi^*)$		98.0	2.30	2.29	2.28	2.28	2.32	2.30	1.97
	$^3A_2(n \rightarrow \pi^*)$		96.9	2.72	2.63	2.64	2.65	2.72	2.64	2.51
	$^3A_1(\pi \rightarrow \pi^*)$		98.2	4.21	4.20	4.19	4.20	4.20	4.20	3.78
Cyclopentadienethione	$^3B_1(n, \pi \rightarrow \pi^*, \pi^*)$		10.0	5.98	5.99	6.05	6.04	5.55		4.46
	$^1A_2(n \rightarrow \pi^*)$		87.2	1.76	1.74	1.71	1.72	1.74	1.73	1.43
	$^1B_2(\pi \rightarrow \pi^*)$	0.000	85.3	2.71	2.66	2.62	2.61	2.71	2.67	1.99
	$^1B_1(n, \pi \rightarrow \pi^*, \pi^*)$	0.000	1.1	4.27	4.35	4.40	4.39	3.84	3.93	2.89
	$^1A_1(\pi \rightarrow \pi^*)$	0.378	89.2	5.13	5.01	4.94	4.93	5.14	5.03	4.42
	$^1A_1(\pi, \pi \rightarrow \pi^*, \pi^*)$	0.003	51.7	5.86	5.90	5.89	5.88	5.68		4.84
	$^3A_2(n \rightarrow \pi^*)$		97.0	1.50	1.47	1.47	1.48	1.49	1.47	1.26
	$^3B_2(\pi \rightarrow \pi^*)$		97.1	1.91	1.90	1.88	1.88	1.91	1.90	1.61
	$^3A_1(\pi \rightarrow \pi^*)$		98.1	2.51	2.55	2.52	2.53	2.50	2.54	2.41
$^3B_1(n, \pi \rightarrow \pi^*, \pi^*)$		4.2	4.25	4.34	4.39	4.37	3.81	3.90	2.88	

^a CASPT2 values from Ref. [347].

Diazirine This compact molecule is the diazo equivalent of cyclopropene, and it has been introduced in our latest work [350]. It is a rather elusive compound experimentally so that the most complete study of its transitions energies are theoretical and have been performed at the MCQDPT2 [351] and EOM-CCSD [352] levels. For the eight considered transitions (Table 4.5), the % T_1 values are larger than 90% and the differences between the CC3, CCSDT, and CCSDTQ values are at most 0.02 eV. In addition, except possibly for the 1B_2 transition, the basis set effects are rather limited. In short, one can be very confident that the TBEs/*aug-cc-pVTZ* given in the database are chemically accurate at least for the selected geometry and basis set. If we compare to the previously published values, one notes that the MCQDPT2 results are likely slightly off target, whereas there is quite a good agreement with the EOM-CCSD values of Krylov and coworkers [352]. This latter work used a very diffuse basis set, so that one cannot be definitive that the remaining differences are purely related to the level of theory. Finally, we are aware of only one experimental data: the 0–0 energy of the lowest singlet state at 3.87 eV [353], a value slightly smaller than the computed vertical transition energy, as it should.

TABLE 4.5: Transition energies (in eV) determined in diazirine with CC3, CCSDT, and CCSDTQ. For all transitions, we provide the single-excitation character $\%T_1$ obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

Transition	CC3			CCSDT			CCSDTQ			Litt.				
	<i>f</i>	$\%T_1$	6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	<i>aug-cc-pVQZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	Th. ^a	Th. ^b	
Diazirine	Nature													
	$^1B_1(n \rightarrow \pi^*)$	0.002	92.5	4.20	4.16	4.11	4.11	4.18	4.15	4.10	4.17	4.14	4.01	4.27
	$^1A_2(\sigma \rightarrow \pi^*)$		90.9	7.43	7.40	7.31	7.29	7.40	7.37	7.28	7.39	7.36	7.79	7.61
	$^1B_2(n \rightarrow 3s)$	0.000	93.5	7.62	7.36	7.45	7.48	7.62	7.35	7.44		7.35	7.16	7.32
	$^1A_1(n \rightarrow 3p)$	0.132	93.8	8.05	7.97	8.04	8.05	8.03	7.95	8.03	8.02	7.95	9.91	7.86
	$^3B_1(n \rightarrow \pi^*)$		98.2	3.56	3.53	3.51	3.51	3.55	3.53	3.50	3.54		3.37	
	$^3B_2(\pi \rightarrow \pi^*)$		98.8	5.08	5.06	5.05	5.05	5.08	5.06	5.05	5.09		5.46	
	$^3A_2(n \rightarrow \pi^*)$		98.3	6.20	6.15	6.13	6.13	6.18	6.14	6.12	6.18		5.97	
	$^3A_1(n \rightarrow 3p)$		98.4	6.87	6.82	6.83	6.85	6.85	6.80	6.81	6.85		6.57	

^a MCQDPT2 values from Ref. [351].

^b EOM-CCSD energies from Ref. [352].

Hexatriene and octatetraene Ethylene and butadiene, the two shortest members of the polyenes, have been treated in QUEST#1 and QUEST#3, respectively [30, 32]. The evolution of excited state energies in longer polyenic chains is obviously of interest and this is why we considered hexatriene and octatetraene here. It should be noted that the transition energies are rather sensitive to the bond length alternation in polyenes, so that we trust that our choice of CC3 geometries is an asset as compared to previous estimates. Our results are collected in Table 4.6. Let us start by the famous singlet B_u and A_g valence states. The transition to B_u is bright and has a strong single-excitation character, so that one expects CC to be an adequate methodology, and one indeed finds small differences between CC3 and CCSDT values (ca. 0.03 eV). Our TBE/*aug-cc-pVTZ* are 5.37 and 4.78 eV for hexatriene and octatetraene, respectively. These values are compared to selected previous theoretical estimates in Table 4.6 and one clearly notices quite a large spread. Experimentally, the 0–0 transition has been measured to be 4.95 eV by electron impact [354] and 4.93 eV by optical spectroscopy [355] for hexatriene, and 4.41 eV with the latter technique for the longer oligomer [356], values logically smaller than our vertical estimates. However, one clearly notices a decrease of 0.53 eV when increasing the chain length, as compared to 0.59 eV with our TBE/*aug-cc-pVTZ* highlighting the consistency of quantum and measured trends. The A_g transitions are known to be much more challenging: the states are dark in one-photon absorption, and it has a very significant multi-excitation character (% T_1 of roughly 65 % for both compounds). On a positive note, the basis set effects are very limited for the A_g state, 6-31+G(d) being apparently sufficient. In contrast, as expected for such transition, there is a significant drop of the theoretical estimate in going from CC3 to CCSDT. From the analysis performed for double excitations in Ref. [36], it is unclear if NEVPT2 or CASPT2 would in fact outperform CCSDT for such “mixed-character” state, so that we cannot define a trustworthy TBE on this basis. However, based on our experience for butadiene [32], one can widely estimate the transition energy to be in the range 5.55–5.60 eV for hexatriene and in the range 4.80–4.85 eV for octatetraene. Interestingly the FCI value of Chien *et al.* with a small basis set for hexatriene (5.59 eV) is compatible with such an estimate. Experimentally, for hexatriene, multiphoton experiments estimate the A_g state to be slightly above the B_u transition [357], an outcome that theory reproduces.

For the two Rydberg transitions of hexatriene, the differences between CC3 and CCSDT estimates are very small, % T_1 values are large, so that CC estimates can likely be trusted. However, the basis set effects are rather large, and *aug-cc-pVTZ* might be insufficient to reach basis set convergence. Our values are reasonably similar to those obtained with CASPT2 almost thirty years ago [358]. The experimental 0–0 energies are 5.68 eV and 6.06 eV [359], but the assignments of Rydberg transitions is a matter of discussion [358], so that we prefer again not to use measured data as reference.

For the triplet excited states, given their very large % T_1 values, we logically trust the CC estimates. We note that, to the best of our knowledge, this work is the first to report true CC3/*aug-cc-pVTZ* values for these two systems. Indeed, the previous CC3 estimates provided by Thiel [198, 199] were obtained by correcting CC3/TZVP values thanks to CC2 calculations using a larger basis set. These authors nevertheless provided very close estimates to ours: 2.71, 4.33, 2.32 and 3.69 eV (going down the list of triplet excited states in Table 4.6). These data are within 0.04 eV of the current values. Comparatively, the previous MRMP and CASPT2 results [198, 358, 360] are therefore slightly too low for the triplet transition energies. For hexatriene (octatetraene), electron impact studies return maxima at 2.61(210) and 4.11(355) eV for the two lowest transitions [354] ([361]), in reasonable agreement with the values listed in the QUEST database.

Maleimide Maleimide was quite surprisingly much less studied theoretically than other similar compounds. We are only aware of the 2003 CASPT2 analysis of Climent and coworkers [363], a refined 2020 joint theory/experiment study using CASPT2, ADC(3), and EOM-CCSD [364], as well as two quite recent investigations focussed on the geometries of specific states [292, 365] rather than on the transition energies. Our results are listed in Table 4.7. For all considered singlet (triplet) transitions, we obtained % T_1 values larger than 85 % (95 %), and one indeed notices very consistent estimates with CC3 and CCSDT, the largest difference being 0.03 eV. All transitions can therefore

TABLE 4.6: Transition energies (in eV) determined in hexatriene and octatetraene. For all transitions, we provide the single-excitation character % T_1 obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	f	% T_1	CC3			CCSDT		Litt.			
				6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	Th. ^a	Th. ^b	Th. ^c	Th. ^d
Hexatriene	Nature	1.115	92.2	5.54	5.37	5.34	5.56	5.40	5.01	5.37	5.74	5.59
	$^1B_u(\pi \rightarrow \pi^*)$		65.3	5.76	5.76	5.75	5.62	5.63	5.19	5.34	5.73	5.58
	$^1A_g(\pi \rightarrow \pi^*)$	0.009	93.6	6.04	5.71	5.78	6.05	5.72	5.84			
	$^1B_g(\pi \rightarrow 3p)$		93.5	6.05	5.84	5.92	6.07	5.86	6.12			
	$^3B_u(\pi \rightarrow \pi^*)$		97.9	2.73	2.73	2.73	2.73		2.55	2.40		
	$^3A_g(\pi \rightarrow \pi^*)$		98.3	4.37	4.37	4.36	4.37		4.12	4.15		
Octatetraene	$^1B_u(\pi \rightarrow \pi^*)$	n.d.	91.5	4.95	4.78	4.75	4.98		4.35	4.66		
	$^1A_g(\pi \rightarrow \pi^*)$		63.7	5.05	5.05	5.04	4.91		4.53	4.47		
	$^3B_u(\pi \rightarrow \pi^*)$		97.5	2.35	2.36	2.36			2.27	2.20		
	$^3A_g(\pi \rightarrow \pi^*)$		98.0	3.73	3.73	3.73			3.61	3.55		

^a CASPT2 values from Ref. [358] (hexatriene) and Ref. [198] (octatetraene).

^b MRMP values from Ref. [360].

^c Cumulant values from Ref. [362].

^d FCI/double- ζ values from Ref. [38].

be considered as rather “safe”. Comparing the 2003 and 2020 CASPT2 values Table 4.7, one notices large differences between the two, and our present estimates are (much) closer from the most recent values. Nevertheless, even the 2020 CASPT2 results seem rather too low as compared to the values provided here. The experimental data are limited. Interestingly, Climent and coworkers attributed the experimental 0–0 absorption at 3.33 eV (see footnotes in Table 4.7) to the second transition, but given our data, we believe that it is more likely the B_1 transition, an assignment consistent with the fact that this band shows non-zero experimental intensities. The A_2 transition seems indeed significantly too high with CCSDT to be attributed to the 3.33 eV measurement. For this assignment, we therefore agree with the analysis of Ref. [364]. Globally, our CCSDT values are typically bracketed by the EOM-CCSD and CASPT2 values of this recent study, which we consider a good hint of accuracy.

TABLE 4.7: Transition energies (in eV) determined in maleimide. For all transitions, we provide the single-excitation character % T_1 obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	f	% T_1	CC3			CCSDT		Litt.			
				6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	Th. ^a	Th. ^b	Exp ^c	Exp ^d
Maleimide	Nature											
	$^1B_1(n \rightarrow \pi^*)$	0.000	87.6	3.86	3.80	3.78	3.87	3.82	2.48	3.37	3.33	
	$^1A_2(n \rightarrow \pi^*)$		85.9	4.58	4.54	4.51	4.59	4.55	3.29	3.96		
	$^1B_2(\pi \rightarrow \pi^*)$	0.025	88.2	4.93	4.87	4.86	4.96	4.90	4.44	4.62	~4.4	4.72
	$^1B_2(\pi \rightarrow \pi^*)$	0.373	89.1	6.32	6.23	6.18	6.35	6.26	5.59	5.80	5.53	4.95
	$^1B_2(n \rightarrow 3s)$	0.034	89.1	7.25	7.08	7.19	7.27	7.09	5.98			
	$^3B_1(n \rightarrow \pi^*)$		96.3	3.63	3.57	3.56	3.64		2.31	3.61		
	$^3A_2(n \rightarrow \pi^*)$		98.4	3.72	3.75	3.73	3.73		3.49	3.32		
	$^3B_2(\pi \rightarrow \pi^*)$		96.9	4.28	4.25	4.25	4.27		3.84	3.83		
	$^3A_2(n \rightarrow \pi^*)$		96.1	4.37	4.33	4.31	4.38		3.14	4.32		

^a CASPT2 values from Ref. [363].

^b CASPT2 (singlet) or ADC(3) (triplet) values from Ref. [364].

^c From Ref. [366]: the 3.33 eV value is a 0–0 energy at low temperature in (frozen) EPA, the 4.4 eV value is the lowest (close from 0–0) peak observed in vapour for *N*-Me-maleimide, and the 5.53 eV value is a 0–0 energy in vapour for the *N*-Me-maleimide.

^d Vapour measurements at 315 K from Ref. [364].

Naphthalene Naphthalene, due to its high-symmetry and significance for organic electronics, is also a popular benchmark molecule [29, 196, 198, 199, 367–370], although some studies are focussed on the lowest-energy states “only” [371, 372]. Our results are listed in Table 4.8. We believe that the convincing work of Fliegl and Sundholm remains the most complete analysis to date [370].

For the singlet transitions, we could obtain CCSDT, albeit only with Pople’s basis set. However, these remain quite significant as none of the considered excited state (even those with Rydberg character) seems to be strongly affected by the basis set. Indeed, the mean absolute deviation between CC3/6-31+G(d) and CC3/*aug-cc-pVTZ* is 0.16 eV, the maximal discrepancy being 0.23 eV.

When one compares the CCSDT and CC3 values, we note that there are only two transitions (the lowest 1A_g and the second $^1B_{3u}$) for which changes exceeding 0.03 eV can be found between the two bases. It is also striking that for the higher-lying A_g state, both CC3 and CCSDT transition energies are the same despite % T_1 being 72 % only. It therefore appears that naphthalene provides a series of well-behaved transitions for which CC theory is well suited. For the valence transition, the TBEs that we obtained are very close from the previous Thiel's CC3 values. Let us now discuss the valence transitions in more details. For the lowest $^1B_{3u}$ and $^1B_{2u}$ transitions, our TBEs/*aug-cc-pVTZ* are 4.27 eV and 4.90 eV. On the theoretical side, these can be compared to Thiel's 4.25 eV and 4.82 eV values [198], or Fliegl and Sundholm 4.16 and 4.80 eV estimates (obtained with larger basis sets) [370]. On the experimental side, we are aware of vapour phase energy loss values of 4.0 and 4.45 eV [373] and (optical) 0–0 energies of 3.97 and 4.45 eV [374], as well as a 3.93 and 4.35 eV 0–0 energy measurement in cyclohexane [375]. The next transition is the lowest Rydberg state [$^1A_u(\pi \rightarrow 3s)$], and our TBE/*aug-cc-pVTZ* is 5.65 eV, which fits very well the old CASPT2 estimates of Rubio and coworkers (5.54 eV [367]), the more recent CC-derived value (5.56 eV [370]) and the only experimental value we are aware of (5.60 eV by energy loss [373]). It is also likely that the use of even large basis set than *aug-cc-pVTZ* would decrease a bit our estimate. Next come the valence $^1B_{1g}$ and A_g states for which our TBE/*aug-cc-pVTZ* are 5.84 and 5.89 eV (we recall that we labeled the second one as "unsafe"). On the theoretical side, previous calculations led 5.75 and 5.90 eV with exCC3 [198], 5.87 and 6.00 eV with RASPT2 [369], 5.64 and 5.77 eV with exCC3 [370]. On the experimental side, the measured 0–0 energies are 5.22 and 5.52 eV [376] and 5.28 and 5.50 eV, both in solution [375]. All these estimates are rather consistent with one another. Next come two $\pi \rightarrow 3p$ Rydberg transitions, for which our TBE of 6.07 and 6.09 eV are slightly larger than the CASPT2 values of Ref. [367] and the extrapolated CC2 estimates of 5.94 and 5.96 eV. To our knowledge, no experimental measurement exists for these two transitions. We estimate the next valence $^1B_{3u}$ excited state to be close to 6.19 eV with the *aug-cc-pVTZ* basis set. This value is consistent with recent estimates of 6.11 [198], 6.20 [369] and 6.06 eV [370]. For this bright state, there are several available experimental values: 5.55 eV (crystal) [377], 5.62/5.63 eV (solution) [377, 378] and 5.89 eV (vapour) [374], the latter value being also found by energy loss [373]. The last of the four Rydberg states considered herein, $^1B_{1u}(\pi \rightarrow 3s)$, is located by us at 6.33 eV, a value 0.3 eV above the CASPT2 value [367] but consistent with a recent CC2 estimate (6.26 eV [370]). For the second $^1B_{2u}(\pi \rightarrow \pi^*)$ transition, our vertical best estimate is 6.42 eV, fitting Thiel's (6.36 eV) [198] and Sundholm's (6.30 eV) results [370]. The experimental values are around 6.0 eV (energy loss [373]) and 6.14 eV (optical spectroscopy) [374]. Finally, for the higher-lying $^1B_{1g}(\pi \rightarrow \pi^*)$ and $^1A_g(\pi \rightarrow \pi^*)$ states, our TBE/*aug-cc-pVTZ* values of 6.48 and 6.87 eV appear too high as compared to the estimates of Ref. [370] (6.19 and 6.40 eV), which is likely due to the strong basis set effects for these two excited states. The 1A_g transition was estimated at 6.05 eV [376] by two-photon spectroscopy.

For the triplet transitions, we have investigated almost the same valence transitions as in Thiel's set [196, 198, 199]. Let us note that for all states, CC3 returns very large % T_1 , and that the differences between *aug-cc-pVDZ* and *aug-cc-pVTZ* estimates is at most 0.04 eV. This clearly hints that the present estimates are trustworthy, but as we have been unable to perform CCSDT calculations, we nevertheless rate all of them as "unsafe" in the QUEST database, which is a conservative choice. The present values are also very similar to those obtained by basis set extrapolation thanks to the work of Thiel [198], except for the highest triplet transition considered here in which a significant difference of 0.2 eV is found. For most transitions, one also find a good consistency with earlier RASPT2 calculations [369]. Experimentally, the available data are typically T-T absorption, and this includes values of +2.25 eV, +2.93 eV for the two 3A_g states [379] and 3.12 eV for the intermediate $^3B_{1g}$ transition [380], that we can compare to our +2.32, +3.22 and +3.00 eV values, respectively.

Nitroxl (HNO) In QUEST#2 [36], we treated only one excited state, of pure double-excitation character, of this compact molecule. We have used a large panel of high-level methods here, considering five excited states Table 4.9. For all transitions (except the Rydberg one), we could nicely converge FCI/*aug-cc-pVTZ* values that can be used as solid references. For the Rydberg

TABLE 4.8: Transition energies (in eV) determined in naphthalene. For all transitions, we provide the single-excitation character $\%T_1$ obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	CC3			CCSDT			Litt.			
		<i>f</i>	$\%T_1$	6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	6-31+G(d)	Th. ^a	Th. ^b	Th. ^c	
Naphthalene	Nature										
	¹ B _{3u} ($\pi \rightarrow \pi^*$)	0.000	85.8	4.36	4.33	4.30	4.33	4.03	4.25	4.23	
	¹ B _{2u} ($\pi \rightarrow \pi^*$)	0.067	90.3	5.10	4.91	4.87	5.13	4.56	4.82	4.61	
	¹ A _u ($\pi \rightarrow 3s$)		92.7	5.85	5.57	5.63	5.87	5.54			
	¹ B _{1g} ($\pi \rightarrow \pi^*$)		84.7	5.99	5.85	5.83	6.00	5.53	5.75	5.87	
	¹ A _g ($\pi \rightarrow \pi^*$)		83.8	6.03	5.97	5.94	5.98	5.39	5.90	6.00	
	¹ B _{3g} ($\pi \rightarrow 3p$)		92.8	6.12	5.98	6.04	6.15	5.98			
	¹ B _{2g} ($\pi \rightarrow 3p$)		92.5	6.24	6.00	6.07	6.26	5.94			
	¹ B _{3u} ($\pi \rightarrow \pi^*$)	n.d.	90.6	6.30	6.19	6.15	6.34	5.54	6.11	6.20	
	¹ B _{1u} ($\pi \rightarrow 3s$)	n.d.	91.9	6.55	6.27	6.32	6.56	6.03			
	¹ B _{2u} ($\pi \rightarrow \pi^*$)	n.d.	90.2	6.61	6.45	6.39	6.64	5.93	6.36	6.12	
	¹ B _{1g} ($\pi \rightarrow \pi^*$)		87.5	6.64	6.52	6.46	6.66	5.87	6.46	6.35	
	¹ A _g ($\pi \rightarrow \pi^*$)		71.5	6.99	6.91	6.87	6.99	6.04	6.87	6.66	
	³ B _{2u} ($\pi \rightarrow \pi^*$)		97.7	3.19	3.18	3.17			3.09	3.21	
	³ B _{3u} ($\pi \rightarrow \pi^*$)		96.6	4.25	4.19	4.16			4.09	4.11	
	³ B _{1g} ($\pi \rightarrow \pi^*$)		97.8	4.53	4.49	4.48			4.42	4.44	
	³ B _{2u} ($\pi \rightarrow \pi^*$)		96.8	4.71	4.67	4.64			4.56	4.62	
	³ B _{3u} ($\pi \rightarrow \pi^*$)		97.5	5.17	4.99	4.95			4.92	4.66	
	³ A _g ($\pi \rightarrow \pi^*$)		97.3	5.56	5.52	5.49			5.42	5.46	
	³ B _{1g} ($\pi \rightarrow \pi^*$)		95.6	6.37	6.21	6.17			6.12	5.95	
³ A _g ($\pi \rightarrow \pi^*$)		95.2	6.52	6.42	6.39			6.17	6.25		

^a CASPT2 values from Ref. [367].

^b exCC3 values from Ref. [198].

^c RASPT2 values from Ref. [369].

transition, which is naturally more sensitive to the basis set effect, the CCSDTQ/*aug-cc-pVDZ* and FCI/*aug-cc-pVDZ* also perfectly match. A previous CASPT2 work [381], reported transition energies to the lowest singlet and triplet of 0.67 and 1.53 eV, two values that now appear rather too low, a usual trend for CASPT2 when no IPEA shift is applied.

TABLE 4.9: Transition energies (in eV) determined in nitroxyl with CC3, CCSDT, CCSDTQ and FCI. For all transitions, we provide the single-excitation character $\%T_1$ obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	CC3			CCSDT			CCSDTQ		FCI
		<i>f</i>	$\%T_1$	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>
HNO	Nature									
	¹ A''(<i>n</i> → π^*)	0.000	93.2	1.78	1.75	1.77	1.74	1.77	1.78(1)	1.74(2)
	¹ A'(n, n → π^* , π^*)	0.000	0.3	5.25	5.26	4.76	4.79	4.42	4.41(1)	4.33(0)
	¹ A'(Ryd)	0.038	92.4	6.12	6.26	6.12	6.25	6.14	6.15(1)	
	³ A''(<i>n</i> → π^*)		99.2	0.87	0.88	0.87	0.88	0.87	0.87(1)	0.88(2)
	³ A'($\pi \rightarrow \pi^*$)		98.5	5.62	5.59	5.62	5.59	5.64		5.61(1)

Streptocyanines In addition to the smallest streptocyanine treated in our earliest work [30], we have investigated here the properties of the two next members of the cationic series which contain 3 and 5 carbon atoms respectively bracketed by NH₂ groups (see Table 4.10). These systems are of specific interest because it is well-known that there are challenging for TD-DFT [382]. We report in Table 4.10 the transition energies to the lowest excited states of both spin symmetries with (valence) $\pi \rightarrow \pi^*$ character. For the smallest of the two compounds treated here, we have been able to converge a CIPSI calculation with the 6-31+G(d) basis set, and it clearly gives us confidence that both CC3 and CCSDT values are accurate, the former method being actually even closer to the FCI extrapolation for that specific molecules. The detailed investigation of these compounds likely remains the one of Send, Valsson, and Filippi [383]. These authors reported for the singlet states of these two cyanines: i) exCC3 values of 4.84 and 3.65 eV; ii) DMC values of 5.03(2) and 3.83(2) eV; and iii) CASPT2 estimates of 4.69 and 3.53 eV. The present estimates better match the previous CC estimates, the DMC (CASPT2) transition energies appearing slightly too large (too low). Once more,

given the results in Table 4.10, we believe that our TBEs are the most accurate to date, at least for the considered geometries.

TABLE 4.10: Transition energies (in eV) determined in two cyanines with CC3, CCSDT and FCI. For all transitions, we provide the single-excitation character % T_1 obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	f	% T_1	CC3			CCSDT			FCI
				6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	<i>aug-cc-pVQZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	6-31+G(d)
Streptocyanine-3	Nature									
	$^1B_2(\pi \rightarrow \pi^*)$	0.755	87.2	4.83	4.83	4.82	4.82	4.80	4.81	4.83(1)
	$^3B_2(\pi \rightarrow \pi^*)$		98.0	3.45	3.45	3.44	3.44	3.44		3.45(1)
Streptocyanine-5	Nature									
	$^1B_2(\pi \rightarrow \pi^*)$	1.182	85.8	3.63	3.66	3.66		3.60	3.64	
	$^3B_2(\pi \rightarrow \pi^*)$		97.7	2.49	2.49	2.48		2.48		

Thioacrolein This heavier analog of acrolein was not much studied theoretically before, except for calculations of its 0–0 energies [102] and rather old TD-DFT calculations [384]. We report in Table 4.11 the transition energies to the lowest excited states of both spin symmetries, of clear valence $n \rightarrow \pi^*$ character. As one can see, there is for both excited states, a remarkable insensibility to the basis set size, and also very similar CC3 and CCSDT estimates. The experimental 0–0 energies are 1.88 eV (singlet) and 1.74 eV (triplet) [385], both slightly below our vertical estimates as it should.

TABLE 4.11: Transition energies (in eV) determined in thioacrolein with CC3 and CCSDT. For all transitions, we provide the single-excitation character % T_1 obtained at LR-CC3/*aug-cc-pVTZ* level. For the dipole-allowed transitions, we provide the corresponding values of the oscillator strength at the same level of theory.

	Transition	f	% T_1	CC3			CCSDT			
				6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>	<i>aug-cc-pVQZ</i>	6-31+G(d)	<i>aug-cc-pVDZ</i>	<i>aug-cc-pVTZ</i>
Thioacrolein	Nature									
	$^1A''(n \rightarrow \pi^*)$	0.000	86.4	2.17	2.17	2.14	2.15	2.14	2.15	2.11
	$^3A''(n \rightarrow \pi^*)$		96.9	1.96	1.95	1.93	1.94	1.94	1.93	

4.4 Theoretical best estimates

We discuss in this section the generation of the TBEs obtained with the *aug-cc-pVTZ* basis. For the closed-shell compounds, the exhaustive list of TBEs can be found in Table 4.12 alongside various specifications: the molecule’s name, the excitation, its nature (valence, Rydberg, or charge transfer), its oscillator strength (when symmetry- and spin-allowed), and its percentage of single excitations $\%T_1$ (computed at the LR-CC3 level). All these quantities are computed with the same *aug-cc-pVTZ* basis. Importantly, we also report the composite approach considered to compute the TBEs (see column “Method”). Following an ONIOM-like strategy [386, 387], the TBEs are computed as “A/SB + [B/TB - B/SB]”, where A/SB is the excitation energy computed with a method A in a smaller basis (SB), and B/SB and B/TB are excitation energies computed with a method B in the small basis and target basis TB, respectively. Table 4.13 reports the TBEs for the open-shell molecules belonging to the QUEST#4 subset.

Talking about numbers, the QUEST database is composed of 551 excitation energies, including 302 singlet, 197 triplet, 51 doublet, 412 valence, and 176 Rydberg excited states. Amongst the valence transitions in closed-shell compounds, 135 transitions correspond to $n \rightarrow \pi^*$ excitations, 200 to $\pi \rightarrow \pi^*$ excitations, and 23 are doubly-excited states. In terms of molecular sizes, 146 excitations are obtained in molecules having in-between 1 and 3 non-hydrogen atoms, 97 excitations from 4 non-hydrogen atom compounds, 177 from molecules composed by 5 and 6 non-hydrogen atoms, and, finally, 68 excitations are obtained from systems with 7 to 10 non-hydrogen atoms. In addition, QUEST is composed by 24 open-shell molecules with a single unpaired electron. Amongst these excited states, 485 of them are considered as “safe”, *i.e.*, chemically-accurate for the considered basis set and geometry. Besides this energetic criterion, we consider as “safe” transitions that are either: i) computed with FCI or CCSDTQ, or ii) in which the difference between CC3 and CCSDT excitation energies is small (*i.e.*, around 0.03–0.04 eV) with a large $\%T_1$ value.

TABLE 4.12: Theoretical best estimates TBEs (in eV), oscillator strengths f , percentage of single excitations $\%T_1$ involved in the transition (computed at the CC3 level) for the full set of closed-shell compounds of the QUEST database. “Method” provides the protocol employed to compute the TBEs. The nature of the excitation is also provided: V, R, and CT stands for valence, Rydberg, and charge transfer, respectively. [F] indicates a fluorescence transition, *i.e.*, a vertical transition energy computed from an excited-state geometry. AVXZ stands for *aug-cc-pVXZ*.

#	Molecule	Excitation	Nature	$\%T_1$	f	TBE	Method	Safe?	
1	Acetaldehyde	$^1A''(n \rightarrow \pi^*)$	V	91	0.000	4.31	FCI/AVTZ	Y	
2		$^3A''(n \rightarrow \pi^*)$	V	97		3.97	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y	
3		$^1A_2(n \rightarrow \pi^*)$	V	91		4.47	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y	
4	Acetone	$^1B_2(n \rightarrow 3s)$	R	90	0.000	6.46	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y	
5		$^1A_2(n \rightarrow 3p)$	R	90		7.47	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y	
6		$^1A_1(n \rightarrow 3p)$	R	90	0.004	7.51	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y	
7		$^1B_2(n \rightarrow 3p)$	R	91	0.029	7.62	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y	
8		$^3A_2(n \rightarrow \pi^*)$	V	97		4.13	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y	
9	Acetylene	$^3A_1(\pi \rightarrow \pi^*)$	V	98		6.25	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y	
10		$^1\Sigma_u^+(\pi \rightarrow \pi^*)$	V	96		7.10	FCI/AVTZ	Y	
11		$^1\Delta_u(\pi \rightarrow \pi^*)$	V	93		7.44	FCI/AVTZ	Y	
12		$^3\Sigma_u^+(\pi \rightarrow \pi^*)$	V	99		5.53	FCI/AVTZ	Y	
13		$^3\Delta_u(\pi \rightarrow \pi^*)$	V	99		6.40	FCI/AVTZ	Y	
14		$^3\Sigma_u^-(\pi \rightarrow \pi^*)$	V	98		7.08	FCI/AVTZ	Y	
15		$^1A_u[F](\pi \rightarrow \pi^*)$	V	95		3.64	FCI/AVTZ	Y	
16		$^1A_2[F](\pi \rightarrow \pi^*)$	V	95		3.85	FCI/AVTZ	Y	
17		Acrolein	$^1A''(n \rightarrow \pi^*)$	V	87	0.000	3.78	FCI/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
18			$^1A'(\pi \rightarrow \pi^*)$	V	91	0.344	6.69	CCSDT/AVTZ	Y
19	$^1A''(n \rightarrow \pi^*)$		V	79	0.000	6.72	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N	
20	$^1A'(n \rightarrow 3s)$		R	89	0.109	7.08	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y	
21	$^1A'$ (double)		V	75	(n.d.)	7.87	FCI/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	
22	$^3A''(n \rightarrow \pi^*)$		V	97		3.51	FCI/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	
23	$^3A'(\pi \rightarrow \pi^*)$		V	98		3.94	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y	
24	$^3A'(\pi \rightarrow \pi^*)$		V	98		6.18	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y	
25	$^3A''(n \rightarrow \pi^*)$		V	92		6.54	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N	
26	Ammonia		$^1A_2(n \rightarrow 3s)$	R	93	0.086	6.59	FCI/AVTZ	Y
27		$^1E(n \rightarrow 3p)$	R	93	0.006	8.16	FCI/AVTZ	Y	
28		$^1A_1(n \rightarrow 3p)$	R	94	0.003	9.33	FCI/AVTZ	Y	
29		$^1A_2(n \rightarrow 3s)$	R	93	0.008	9.96	FCI/AVTZ	Y	
30	Aza-naphthalene	$^3A_2(n \rightarrow 3s)$	R	98		6.31	FCI/AVTZ	Y	
31		$^1B_{3g}(n \rightarrow \pi^*)$	V	88		3.14	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	
32		$^1B_{2u}(\pi \rightarrow \pi^*)$	V	86	0.190	4.28	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	
33		$^1B_{1u}(n \rightarrow \pi^*)$	V	88	(n.d.)	4.34	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	
34		$^1B_{2g}(n \rightarrow \pi^*)$	V	87		4.55	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	
35		$^1B_{2g}(n \rightarrow \pi^*)$	V	84		4.89	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y	

Continued on next page

Table 4.12 – Continued from previous page

#	Molecule	Excitation	Nature	%T ₁	f	TBE	Method	Safe?
36		¹ B _{1u} (n → π*)	V	82	(n.d.)	5.24	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N
37		¹ A _u (n → π*)	V	83		5.34	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
38		¹ B _{3u} (π → π*)	V	88	0.028	5.68	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N
39		¹ A _g (π → π*)	V	85		5.80	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
40		¹ A _u (n → π*)	V	84		5.92	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
41		¹ A _g (n → 3s)	R	90		6.50	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
42		³ B _{3g} (n → π*)	V	96		2.82	CC3/AVTZ	N
43		³ B _{2u} (π → π*)	V	97		3.67	CC3/AVTZ	N
44		³ B _{3u} (π → π*)	V	97		3.75	CC3/AVTZ	N
45		³ B _{1u} (n → π*)	V	97		3.77	CC3/AVTZ	N
46		³ B _{2g} (n → π*)	V	96		4.34	CC3/AVTZ	N
47		³ B _{2g} (n → π*)	V	95		4.61	CC3/AVTZ	N
48		³ B _{3u} (π → π*)	V	96		4.75	CC3/AVTZ	N
49		³ A _u (n → π*)	V	96		4.87	CC3/AVTZ	N
50	Beryllium	¹ D(double)	R	32		7.15	FCI/AVTZ	Y
51	Benzene	¹ B _{2u} (π → π*)	V	86		5.06	CCSDT/AVTZ	Y
52		¹ B _{1u} (π → π*)	V	92		6.45	CCSDT/AVTZ	Y
53		¹ E _{1g} (π → 3s)	R	92		6.52	CCSDT/AVTZ	Y
54		¹ A _{2u} (π → 3p)	R	93	0.066	7.08	CCSDT/AVTZ	Y
55		¹ E _{2u} (π → 3p)	R	92		7.15	CCSDT/AVTZ	Y
56		¹ E _{2g} (π → π*)	V	73		8.28	FCI/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
57		¹ A _{1g} (double)	V	n.d.		10.55	XMS-CASPT2/AVTZ	N
58		³ B _{1u} (π → π*)	V	98		4.16	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
59		³ E _{1u} (π → π*)	V	97		4.85	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
60		³ B _{2u} (π → π*)	V	98		5.81	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
61	Benzoquinone	¹ B _{1g} (n → π*)	V	85		2.82	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
62		¹ A _u (n → π*)	V	84		2.96	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
63		¹ A _g (double)	V	0		4.57	NEVPT2/AVTZ	N
64		¹ B _{3g} (π → π*)	V	88		4.58	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
65		¹ B _{1u} (π → π*)	V	88	0.471	5.62	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
66		¹ B _{3u} (n → π*)	V	79	0.001	5.79	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
67		¹ B _{2g} (n → π*)	V	76		5.95	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
68		¹ A _u (n → π*)	V	74		6.35	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
69		¹ B _{1g} (n → π*)	V	83		6.38	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
70		¹ B _{2g} (n → π*)	V	86		7.22	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
71		³ B _{1g} (n → π*)	V	96		2.58	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
72		³ A _u (n → π*)	V	95		2.72	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
73		³ B _{1u} (π → π*)	V	97		3.12	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
74		³ B _{3g} (π → π*)	V	97		3.46	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
75	Butadiene	¹ B _{1u} (π → π*)	V	93	0.664	6.22	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
76		¹ B _g (π → 3s)	R	94		6.33	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
77		¹ A _g (π → π*)	V	75		6.50	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
78		¹ A _u (π → 3p)	R	94	0.001	6.64	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
79		¹ A _u (π → 3p)	R	94	0.049	6.80	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
80		¹ B _u (π → 3p)	R	93	0.055	7.68	CCSDTQ/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
81		³ B _u (π → π*)	V	98		3.36	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
82		³ A _g (π → π*)	V	98		5.20	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
83		³ B _g (π → 3s)	R	97		6.29	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
84	Carbon Dimer	¹ Δ _g (double)	R	0		2.09	FCI/AVTZ	Y
85		¹ Σ _g ⁺ (double)	R	0		2.42	FCI/AVTZ	Y
86	Carbon monoxide	¹ Π(n → π*)	V	93	0.168	8.49	FCI/AVTZ	Y
87		¹ Σ ⁻ (π → π*)	V	93		9.92	FCI/AVTZ	Y
88		¹ Δ(π → π*)	V	91		10.06	FCI/AVTZ	Y
89		¹ Σ ⁺ (n.d.)	R	91	0.003	10.95	FCI/AVTZ	Y
90		¹ Σ ⁺ (n.d.)	R	92	0.200	11.52	FCI/AVTZ	Y
91		¹ Π(n.d.)	R	92	0.106	11.72	FCI/AVTZ	Y
92		³ Π(n → π*)	V	98		6.28	FCI/AVTZ	Y
93		³ Σ ⁺ (π → π*)	V	98		8.45	FCI/AVTZ	Y
94		³ Δ(π → π*)	V	98		9.27	FCI/AVTZ	Y
95		³ Σ ⁻ (π → π*)	V	97		9.80	FCI/AVTZ	Y
96		³ Σ ⁺ (n.d.)	R	98		10.47	FCI/AVTZ	Y
97	Carbon Dimer	¹ Δ _g (double)	R	1		5.22	FCI/AVTZ	Y
98		¹ Σ _g ⁺ (double)	R	1		5.91	FCI/AVTZ	Y
99	Carbonylfluoride	¹ A ₂ (n → π*)	V	91		7.31	FCI/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
100		³ A ₂ (n → π*)	V	97		7.06	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
101	CCl2	¹ B ₁ (σ → π*)	V	93	0.002	2.59	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
102		¹ A ₂ (n.d.)	V	88		4.40	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
103		³ B ₁ (σ → π*)	V	98		1.22	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
104		³ A ₂ (n.d.)	V	96		4.31	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
105	CClF	¹ A''(σ → π*)	V	93	0.007	3.57	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
106	CF2	¹ B ₁ (σ → π*)	V	94	0.034	5.09	FCI/AVTZ	Y
107		³ B ₁ (σ → π*)	V	99		2.77	FCI/AVTZ	Y
108	Cyanoacetylene	¹ Σ ⁻ (π → π*)	V	94		5.80	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
109		¹ Δ(π → π*)	V	94		6.07	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
110		³ Σ ⁺ (π → π*)	V	98		4.44	CCSDT/AVTZ	Y
111		³ Δ(π → π*)	V	98		5.21	CCSDT/AVTZ	Y
112		¹ A''[F](π → π*)	V	93	0.004	3.54	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
113	Cyanoformaldehyde	¹ A''(n → π*)	V	89	0.001	3.81	CCSDT/AVTZ	Y
114		¹ A''(π → π*)	V	91	0.000	6.46	CCSDT/AVTZ	Y
115		³ A''(n → π*)	V	97		3.44	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
116		³ A''(π → π*)	V	98		5.01	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
117	Cyanogen	¹ Σ ⁻ (π → π*)	V	94		6.39	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
118		¹ Δ _u (π → π*)	V	93		6.66	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
119		³ Σ _u ⁺ (π → π*)	V	98		4.91	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
120		¹ Σ _u ⁻ [F](π → π*)	V	93		5.05	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
121	Cyclopentadiene	¹ B ₂ (π → π*)	V	93	0.084	5.54	FCI/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
122		¹ A ₂ (π → 3s)	R	94		5.78	CCSDT/AVTZ	Y
123		¹ B ₁ (π → 3p)	R	94	0.037	6.41	CCSDT/AVTZ	Y
124		¹ A ₂ (π → 3p)	R	93		6.46	CCSDT/AVTZ	Y

Continued on next page

Table 4.12 – Continued from previous page

#	Molecule	Excitation	Nature	%T ₁	f	TBE	Method	Safe?
125		¹ B ₂ (π → 3p)	R	94	0.046	6.56	CCSDT/AVTZ	Y
126		¹ A ₁ (π → π*)	V	78	0.010	6.52	CCSDT/AVTZ	N
127		³ B ₂ (π → π*)	V	98		3.31	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
128		³ A ₁ (π → π*)	V	98		5.11	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
129		³ A ₂ (π → 3s)	R	97		5.73	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
130		³ B ₁ (π → 3p)	R	97		6.36	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
131	Cyclopentadienone	¹ A ₂ (n → π*)	V	88		2.94	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
132		¹ B ₂ (π → π*)	V	91	0.004	3.58	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
133		¹ B ₁ (double)	V	3	0.000	5.02	NEVPT2/AVTZ	N
134		¹ A ₁ (double)	V	49	0.131	6.00	NEVPT2/AVTZ	N
135		¹ A ₁ (π → π*)	V	73	0.090	6.09	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
136		³ B ₂ (π → π*)	V	98		2.29	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
137		³ A ₂ (n → π*)	V	96		2.65	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
138		³ A ₁ (π → π*)	V	98		4.19	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
139		³ B ₁ (double)	V	10		4.91	NEVPT2/AVTZ	N
140	Cyclopentadienethione	¹ A ₂ (n → π*)	V	87		1.70	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
141		¹ B ₂ (π → π*)	V	85	0.000	2.63	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
142		¹ B ₁ (double)	V	1	0.000	3.16	NEVPT2/AVTZ	N
143		¹ A ₁ (π → π*)	V	89	0.378	4.96	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
144		¹ A ₁ (double)	V	51	0.003	5.43	NEVPT2/AVTZ	N
145		³ A ₂ (n → π*)	V	97		1.47	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
146		³ B ₂ (π → π*)	V	97		1.88	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
147		³ A ₁ (π → π*)	V	98		2.51	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
148		³ B ₁ (double)	V	4		3.13	NEVPT2/AVTZ	N
149	Cyclopropene	¹ B ₁ (σ → π*)	V	92	0.001	6.68	CCSDT/AVTZ	Y
150		¹ B ₂ (π → π*)	V	95	0.071	6.79	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
151		³ B ₂ (π → π*)	V	98		4.38	FCI/AVTZ	Y
152		³ B ₁ (σ → π*)	V	98		6.45	FCI/AVTZ	Y
153	Cyclopropenone	¹ B ₁ (n → π*)	V	87	0.000	4.26	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
154		¹ A ₂ (n → π*)	V	91		5.55	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
155		¹ B ₂ (n → 3s)	R	90	0.003	6.34	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
156		¹ B ₂ (π → π*)	V	86	0.047	6.54	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
157		¹ B ₂ (n → 3p)	R	91	0.018	6.98	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
158		¹ A ₁ (n → 3p)	R	91	0.003	7.02	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
159		¹ A ₁ (π → π*)	V	90	0.320	8.28	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
160		³ B ₁ (n → π*)	V	96		3.93	CCSDT/AVTZ	Y
161		³ B ₂ (π → π*)	V	97		4.88	CCSDT/AVTZ	Y
162		³ A ₂ (n → π*)	V	97		5.35	CCSDT/AVTZ	Y
163		³ A ₁ (π → π*)	V	98		6.79	CCSDT/AVTZ	Y
164	Cyclopropenethione	¹ A ₂ (n → π*)	V	89		3.41	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
165		¹ B ₁ (n → π*)	V	84	0.000	3.45	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
166		¹ B ₂ (π → π*)	V	83	0.007	4.60	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
167		¹ B ₂ (n → 3s)	R	91	0.048	5.34	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
168		¹ A ₁ (π → π*)	V	89	0.228	5.46	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
169		¹ B ₂ (n → 3p)	R	91	0.084	5.92	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
170		³ A ₂ (n → π*)	V	97		3.28	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
171		³ B ₁ (n → π*)	V	94		3.32	CCSDT/AVTZ	Y
172		³ B ₂ (π → π*)	V	96		4.01	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
173		³ A ₁ (π → π*)	V	98		4.01	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
174	Diacetylene	¹ Σ _g ⁺ (π → π*)	V	94		5.33	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
175		¹ Δ _u (π → π*)	V	94		5.61	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
176		³ Σ _g ⁺ (π → π*)	V	98		4.10	CCSDTQ/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
177		³ Δ _u (π → π*)	V	98		4.78	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
178	Diazirine	¹ B ₁ (n → π*)	V	92	0.002	4.09	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
179		¹ B ₂ (σ → π*)	V	90		7.27	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
180		¹ A ₂ (n → 3s)	R	93	0.000	7.44	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
181		¹ A ₁ (n → 3p)	R	93	0.132	8.03	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
182		³ B ₁ (n → π*)	V	98		3.49	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
183		³ B ₂ (π → π*)	V	98		5.06	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
184		³ A ₂ (n → π*)	V	98		6.12	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
185		³ A ₁ (n → 3p)	R	98		6.81	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
186	Diazomethane	¹ A ₂ (π → π*)	V	90		3.14	FCI/AVTZ	Y
187		¹ B ₁ (π → 3s)	R	93	0.016	5.54	FCI/AVTZ	Y
188		¹ A ₁ (π → π*)	V	91	0.234	5.90	FCI/AVTZ	Y
189		³ A ₂ (π → π*)	V	97		2.79	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
190		³ A ₁ (π → π*)	V	98		4.05	FCI/AVTZ	Y
191		³ B ₁ (π → 3s)	R	98		5.35	FCI/AVTZ	Y
192		³ A ₁ (π → 3p)	R	98		6.82	FCI/AVTZ	Y
193		¹ A ^u [F](π → π*)	V	87	0.000	0.71	FCI/AVTZ	Y
194	Difluorodiazirine	¹ B ₁ (n → π*)	V	93	0.002	3.74	CCSDT/AVTZ	Y
195		¹ A ₂ (π → π*)	V	91		7.00	CCSDT/AVTZ	Y
196		¹ B ₂ (π → π*)	V	93	0.026	8.52	CCSDT/AVTZ	Y
197		³ B ₁ (n → π*)	V	98		3.03	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
198		³ B ₂ (π → π*)	V	98		5.44	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
199		³ A ₂ (π → π*)	V	98		5.80	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
200	Dinitrogen	¹ Π _g ⁺ (n → π*)	V	92		9.34	FCI/AVTZ	Y
201		¹ Σ _g ⁺ (π → π*)	V	97		9.88	FCI/AVTZ	Y
202		¹ Δ _u (π → π*)	V	95	0.000	10.29	FCI/AVTZ	Y
203		¹ Σ _g ⁺ (n.d.)	R	92		12.98	FCI/AVTZ	Y
204		¹ Π _u (n.d.)	R	82	0.458	13.03	FCI/AVTZ	Y
205		¹ Σ _g ⁺ (n.d.)	R	92	0.296	13.09	FCI/AVTZ	Y
206		¹ Π _u (n.d.)	R	87	0.000	13.46	FCI/AVTZ	Y
207		³ Σ _g ⁺ (π → π*)	V	99		7.70	FCI/AVTZ	Y
208		³ Π _g ⁺ (n → π*)	V	98		8.01	FCI/AVTZ	Y
209		³ Δ _u (π → π*)	V	99		8.87	FCI/AVTZ	Y
210		³ Σ _u ⁺ (π → π*)	V	98		9.66	FCI/AVTZ	Y
211	Ethylene	¹ B _{3u} β3s	R	95	0.078	7.39	FCI/AVTZ	Y
212		¹ B _{1u} (π → π*)	V	95	0.346	7.93	FCI/AVTZ	Y
213		¹ B _{1g} (π → 3p)	R	95		8.08	FCI/AVTZ	Y
214		¹ A _g (double)	V	20		12.92	FCI/AVTZ	Y

Continued on next page

Table 4.12 – Continued from previous page

#	Molecule	Excitation	Nature	%T ₁	f	TBE	Method	Safe?
215		³ B _{1g} (π → π*)	V	99		4.54	FCI/AVTZ	Y
216		³ B _{3u} p3s	R	98		7.23	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
217		³ B _{1g} (π → 3p)	R	98		7.98	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
218	Formaldehyde	¹ A ₂ (n → π*)	V	91		3.98	FCI/AVTZ	Y
219		¹ B ₂ (n → 3s)	R	91	0.021	7.23	FCI/AVTZ	Y
220		¹ B ₂ (n → 3p)	R	92	0.037	8.13	FCI/AVTZ	Y
221		¹ A ₁ (n → 3p)	R	91	0.052	8.23	FCI/AVTZ	Y
222		¹ A ₂ (n → 3p)	R	91		8.67	FCI/AVTZ	Y
223		¹ B ₁ (n.d.)	V	90	0.001	9.22	FCI/AVTZ	Y
224		¹ A ₁ (π → π*)	V	90	0.135	9.43	FCI/AVTZ	Y
225		¹ A ₁ (double)	V	5	(n.d.)	10.35	FCI/AVTZ	Y
226		³ A ₂ (n → π*)	V	98		3.58	FCI/AVTZ	Y
227		³ A ₁ (π → π*)	V	99		6.06	FCI/AVTZ	Y
228		³ B ₂ (n → 3s)	R	97		7.06	FCI/AVTZ	Y
229		³ B ₂ (n → 3p)	R	97		7.94	FCI/AVTZ	Y
230		³ A ₁ (n → 3p)	R	97		8.10	FCI/AVTZ	Y
231		³ B ₁ (n.d.)	R	97		8.42	FCI/AVTZ	Y
232		¹ A''[F](n → π*)	V	87	0.000	2.80	FCI/AVTZ	Y
233	Formamide	¹ A''(n → π*)	V	90	0.000	5.65	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
234		¹ A'(n → 3s)	R	88	0.001	6.77	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	N
235		¹ A'(n → 3p)	R	89	0.111	7.38	CCSDT/AVTZ	N
236		¹ A'(π → π*)	V	89	0.251	7.63	FCI/AVTZ	N
237		³ A''(n → π*)	V	97		5.38	FCI/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
238		³ A'(π → π*)	V	98		5.81	FCI/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
239	Formylfluoride	¹ A''(n → π*)	V	91		5.96	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
240		³ A''(n → π*)	V	98	0.001	5.63	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
241	Furan	¹ A ₂ (π → 3s)	R	93		6.09	CCSDT/AVTZ	Y
242		¹ B ₂ (π → π*)	V	93	0.163	6.37	CCSDT/AVTZ	Y
243		¹ A ₁ (π → π*)	V	92	0.000	6.56	CCSDT/AVTZ	Y
244		¹ B ₁ (π → 3p)	R	93	0.038	6.64	CCSDT/AVTZ	Y
245		¹ A ₂ (π → 3p)	R	93		6.81	CCSDT/AVTZ	Y
246		¹ B ₂ (π → 3p)	R	93	0.007	7.24	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
247		³ B ₂ (π → π*)	V	98		4.20	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
248		³ A ₁ (π → π*)	V	98		5.46	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
249		³ A ₂ (π → 3s)	R	97		6.02	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
250		³ B ₁ (π → 3p)	R	97		6.59	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
251	Glyoxal	¹ A _u (n → π*)	V	91	0.000	2.88	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
252		¹ B _g (n → π*)	V	88		4.24	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
253		¹ A _g (double)	V	0	0.000	5.61	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
254		¹ B _g (n → π*)	V	83		6.57	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
255		¹ B _u (n → 3p)	R	91	0.095	7.71	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
256		³ A _u (n → π*)	V	97		2.49	CCSDT/AVTZ	Y
257		³ B _g (n → π*)	V	97		3.89	CCSDT/AVTZ	Y
258		³ B _u (π → π*)	V	98		5.15	CCSDT/AVTZ	Y
259		³ A _g (π → π*)	V	98		6.30	CCSDT/AVTZ	Y
260	HCCI	¹ A''(σ → π*)	V	94	0.003	1.98	FCI/AVTZ	Y
261	HCF	¹ A''(σ → π*)	V	95	0.006	2.49	FCI/AVTZ	Y
262	HCP	¹ Σ ⁻ (π → π*)	V	94		4.84	FCI/AVTZ	Y
263		¹ Δ(π → π*)	V	94		5.15	FCI/AVTZ	Y
264		³ Σ ⁺ (π → π*)	V	98		3.47	FCI/AVTZ	Y
265		³ Δ(π → π*)	V	98		4.22	FCI/AVTZ	Y
266	Hexatriene	¹ B _u (π → π*)	V	92	1.115	5.37	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
267		¹ A _g (π → π*)	V	65		5.62	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
268		¹ A _u (π → 3s)	R	93	0.009	5.79	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
269		¹ B _g (π → 3p)	R	93		5.94	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
270		³ B _u (π → π*)	V	97		2.73	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
271		³ A _g (π → π*)	V	98		4.36	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
272	HPO	¹ A''(n → π*)	V	90	0.003	2.47	FCI/AVTZ	Y
273	HPS	¹ A''(n → π*)	V	90	0.001	1.59	FCI/AVTZ	Y
274	HSiF	¹ A''(σ → π*)	V	93	0.024	3.05	FCI/AVTZ	Y
275	Hydrogen chloride	¹ Π	CT	94	0.056	7.84	FCI/AVTZ	Y
276	Hydrogen sulfide	¹ A ₂ (n → 3p)	R	94		6.18	FCI/AVTZ	Y
277		¹ B ₁ (n → 3p)	R	94	0.063	6.24	FCI/AVTZ	Y
278		³ A ₂ (n → 3p)	R	98		5.81	FCI/AVTZ	Y
279		³ B ₁ (n → 3p)	R	98		5.88	FCI/AVTZ	Y
280	Imidazole	¹ A''(π → 3s)	R	93	0.001	5.71	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
281		¹ A'(π → π*)	V	89	0.124	6.41	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
282		¹ A''(n → π*)	V	93	0.028	6.50	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
283		¹ A'(π → 3p)	R	88	0.035	6.83	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
284		³ A'(π → π*)	V	98		4.73	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
285		³ A''(π → 3s)	R	97		5.66	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
286		³ A'(π → π*)	V	97		5.74	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
287		³ A''(n → π*)	V	97		6.31	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
288	Isobutene	¹ B ₁ (π → 3s)	R	94	0.006	6.46	CCSDT/AVTZ	Y
289		¹ A ₁ (π → 3p)	R	94	0.228	7.01	CCSDT/AVTZ	Y
290		³ A ₁ (π → π*)	V	98		4.53	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
291	Ketene	¹ A ₂ (π → π*)	V	91		3.85	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
292		¹ B ₁ (n → 3s)	R	93	0.035	6.01	FCI/AVTZ	Y
293		¹ A ₁ (π → π*)	V	92	0.154	7.25	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
294		¹ A ₂ (π → 3p)	R	94		7.18	FCI/AVTZ	Y
295		³ A ₂ (n → π*)	V	91		3.77	FCI/AVTZ	Y
296		³ A ₁ (π → π*)	V	98		5.61	FCI/AVTZ	Y
297		³ B ₁ (n → 3p)	R	98		5.79	FCI/AVTZ	Y
298		³ A ₂ (π → 3p)	R	94		7.12	FCI/AVTZ	Y
299		¹ A''[F](π → π*)	V	87	0.000	1.00	FCI/AVTZ	Y
300	Maleimide	¹ B ₁ (n → π*)	V	87	0.000	3.80	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
301		¹ A ₂ (n → π*)	V	85		4.52	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
302		¹ B ₂ (π → π*)	V	88	0.025	4.89	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
303		¹ B ₂ (π → π*)	V	89	0.373	6.21	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
304		¹ B ₂ (n → 3s)	R	89	0.034	7.20	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y

Continued on next page

Table 4.12 – Continued from previous page

#	Molecule	Excitation	Nature	%T ₁	f	TBE	Method	Safe?
305		³ B ₁ (n → π*)	V	96		3.57	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
306		³ B ₂ (π → π*)	V	98		3.74	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
307		³ B ₂ (π → π*)	V	96		4.24	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
308		³ A ₂ (n → π*)	V	96		4.32	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
309	Methanimine	¹ A ₁ (n → π*)	V	90	0.003	5.23	FCI/AVTZ	Y
310		³ A ₁ (n → π*)	V	98		4.65	FCI/AVTZ	Y
311	Methylenecyclopropene	¹ B ₂ (π → π*)	V	85	0.011	4.28	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
312		¹ B ₁ (π → 3s)	R	93	0.005	5.44	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
313		¹ A ₂ (π → 3p)	R	93		5.96	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
314		¹ A ₁ (π → π*)	V	92	0.224	6.12	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	N
315		³ B ₂ (π → π*)	V	97		3.49	CCSDT/AVTZ	Y
316		³ A ₁ (π → π*)	V	98		4.74	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
317	Naphthalene	¹ B _{3u} (π → π*)	V	85	0.000	4.27	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
318		¹ B _{2u} (π → π*)	V	90	0.067	4.90	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
319		¹ A _u (π → 3s)	R	92		5.65	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
320		¹ B _{1g} (π → π*)	V	84		5.84	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
321		¹ A _g (π → π*)	V	83		5.89	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N
322		¹ B _{3g} (π → 3p)	R	92		6.07	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
323		¹ B _{2g} (π → 3p)	R	92		6.09	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
324		¹ B _{3u} (π → π*)	V	90	(n.d.)	6.19	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N
325		¹ B _{1u} (π → 3s)	R	91	(n.d.)	6.33	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
326		¹ B _{2u} (π → π*)	V	90	(n.d.)	6.42	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
327		¹ B _{1g} (π → π*)	V	87		6.48	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
328		¹ A _g (π → π*)	V	71		6.87	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
329		³ B _{2u} (π → π*)	V	97		3.17	CC3/AVTZ	N
330		³ B _{3u} (π → π*)	V	96		4.16	CC3/AVTZ	N
331		³ B _{1g} (π → π*)	V	97		4.48	CC3/AVTZ	N
332		³ B _{2u} (π → π*)	V	96		4.64	CC3/AVTZ	N
333		³ B _{3u} (π → π*)	V	97		4.95	CC3/AVTZ	N
334		³ A _g (π → π*)	V	97		5.49	CC3/AVTZ	N
335		³ B _{1g} (π → π*)	V	95		6.17	CC3/AVTZ	N
336		³ A _g (π → π*)	V	95		6.39	CC3/AVTZ	N
337	Nitrosomethane	¹ A ^{''} (n → π*)	V	93	0.000	1.96	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
338		¹ A ['] (double)	V	2	0.000	4.76	FCI/AVTZ	Y
339		¹ A ['] (n.d.)	R	90	0.006	6.29	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
340		³ A ^{''} (n → π*)	V	98		1.16	FCI/AVTZ	Y
341		³ A ['] (π → π*)	V	98		5.60	FCI/AVTZ	Y
342		¹ A ^{''} [F](n → π*)	V	92	0.000	1.67	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
343	Nitroxyl (HNO)	¹ A ^{''} (n → π*)	V	93	0.000	1.74	FCI/AVTZ	Y
344		¹ A ['] (double)	V	0	0.000	4.33	FCI/AVTZ	Y
345		¹ A ['] (n.d.)	R	92	0.038	6.27	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
346		³ A ^{''} (n → π*)	V	99		0.88	FCI/AVTZ	Y
347		³ A ['] (π → π*)	V	98		5.61	FCI/AVTZ	Y
348	Octatetraene	¹ B _u (π → π*)	V	91	(n.d.)	4.78	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
349		¹ A _g (π → π*)	V	63		4.90	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N
350		³ B _u (π → π*)	V	97		2.36	CC3/AVTZ	N
351		³ A _g (π → π*)	V	98		3.73	CC3/AVTZ	N
352	Propynal	¹ A ^{''} (n → π*)	V	89	0.000	3.80	CCSDT/AVTZ	Y
353		¹ A ^{''} (π → π*)	V	92	0.000	5.54	CCSDT/AVTZ	Y
354		³ A ^{''} (n → π*)	V	97		3.47	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
355		³ A ['] (π → π*)	V	98		4.47	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
356	Pyrazine	¹ B _{3u} (n → π*)	V	90	0.006	4.15	CCSDT/AVTZ	Y
357		¹ A _u (n → π*)	V	88		4.98	CCSDT/AVTZ	Y
358		¹ B _{2u} (π → π*)	V	86	0.078	5.02	CCSDT/AVTZ	Y
359		¹ B _{2g} (n → π*)	V	85		5.71	CCSDT/AVTZ	Y
360		¹ A _g (n → 3s)	R	91		6.65	CCSDT/AVTZ	Y
361		¹ B _{1g} (n → π*)	V	84		6.74	CCSDT/AVTZ	Y
362		¹ B _{1u} (π → π*)	V	92	0.063	6.88	CCSDT/AVTZ	Y
363		¹ B _{1g} (π → 3s)	R	93		7.21	CCSDT/AVTZ	Y
364		¹ B _{2u} (n → 3p)	R	90	0.037	7.24	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
365		¹ B _{1u} (n → 3p)	R	91	0.128	7.44	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
366		¹ B _{1u} (π → π*)	V	90	0.285	7.98	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
367		¹ A _g (double)	V	12		8.04	NEVPT2/AVTZ	N
368		¹ A _g (π → π*)	V	71		8.69	CC3/AVTZ	N
369		³ B _{3u} (n → π*)	V	97		3.59	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
370		³ B _{1u} (π → π*)	V	98		4.35	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
371		³ B _{2u} (π → π*)	V	97		4.39	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
372		³ A _u (n → π*)	V	96		4.93	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
373		³ B _{2g} (n → π*)	V	97		5.08	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
374		³ B _{1u} (π → π*)	V	97		5.28	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
375	Pyridazine	¹ B ₁ (n → π*)	V	89	0.005	3.83	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
376		¹ A ₂ (n → π*)	V	86		4.37	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
377		¹ A ₁ (π → π*)	V	85	0.016	5.26	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
378		¹ A ₂ (n → π*)	V	86		5.72	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
379		¹ B ₂ (n → 3s)	R	88	0.001	6.17	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
380		¹ B ₁ (n → π*)	V	87	0.004	6.37	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
381		¹ B ₂ (π → π*)	V	90	0.010	6.75	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
382		³ B ₁ (n → π*)	V	97		3.19	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
383		³ A ₂ (n → π*)	V	96		4.11	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
384		³ B ₂ (π → π*)	V	98		4.34	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
385		³ A ₁ (π → π*)	V	97		4.82	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
386	Pyridine	¹ B ₁ (n → π*)	V	88	0.004	4.95	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
387		¹ B ₂ (π → π*)	V	86	0.028	5.14	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
388		¹ A ₂ (n → π*)	V	87		5.40	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
389		¹ A ₁ (π → π*)	V	92	0.010	6.62	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
390		¹ A ₁ (n → 3s)	R	89	0.011	6.76	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
391		¹ A ₂ (π → 3s)	R	93		6.82	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
392		¹ B ₁ (π → 3p)	R	93	0.045	7.38	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
393		¹ A ₁ (π → π*)	V	90	0.291	7.39	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y

Continued on next page

Table 4.12 – Continued from previous page

#	Molecule	Excitation	Nature	%T ₁	f	TBE	Method	Safe?
394		¹ B ₂ (π → π*)	V	90	0.319	7.40	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
395		³ A ₁ (π → π*)	V	98		4.30	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
396		³ B ₁ (n → π*)	V	97		4.46	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
397		³ B ₂ (π → π*)	V	97		4.79	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
398		³ A ₁ (π → π*)	V	97		5.04	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
399		³ A ₂ (n → π*)	V	95		5.36	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
400		³ B ₂ (π → π*)	V	97		6.24	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
401	Pyrimidine	¹ B ₁ (n → π*)	V	88	0.005	4.44	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
402		¹ A ₂ (n → π*)	V	88		4.85	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
403		¹ B ₂ (π → π*)	V	86	0.028	5.38	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
404		¹ A ₂ (n → π*)	V	86		5.92	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
405		¹ B ₁ (n → π*)	V	86	0.005	6.26	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
406		¹ B ₂ (n → 3s)	R	90	0.005	6.70	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
407		¹ A ₁ (π → π*)	V	91	0.036	6.88	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
408		³ B ₁ (n → π*)	V	96		4.09	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
409		³ A ₁ (π → π*)	V	98		4.51	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
410		³ A ₂ (n → π*)	V	96		4.66	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
411		³ B ₂ (π → π*)	V	97		4.96	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
412	Pyrrrole	¹ A ₂ (π → 3s)	R	92		5.24	CCSDT/AVTZ	Y
413		¹ B ₁ (π → 3p)	R	92	0.015	6.00	CCSDT/AVTZ	Y
414		¹ A ₂ (π → 3p)	R	93		6.00	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
415		¹ B ₂ (π → π*)	V	92	0.164	6.26	CCSDT/AVTZ	Y
416		¹ A ₁ (π → π*)	V	86	0.001	6.30	CCSDT/AVTZ	Y
417		¹ B ₂ (π → 3p)	R	92	0.003	6.83	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
418		³ B ₂ (π → π*)	V	98		4.51	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
419		³ A ₂ (π → 3s)	R	97		5.21	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
420		³ A ₁ (π → π*)	V	97		5.45	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
421		³ B ₁ (π → 3p)	R	97		5.91	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
422	SiCl ₂	¹ B ₁ (σ → π*)	V	92	0.031	3.91	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
423		³ B ₁ (σ → π*)	V	98		2.48	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
424	Silylidene	¹ A ₂ (n.d.)	R	92		2.11	FCI/AVTZ	Y
425		¹ B ₂ (n.d.)	R	88	0.033	3.78	FCI/AVTZ	Y
426	Streptocyanine-1	¹ B ₂ (π → π*)	V	88	0.347	7.13	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
427		³ B ₂ (π → π*)	V	98		5.52	FCI/AVTZ	Y
428	Streptocyanine-3	¹ B ₂ (π → π*)	V	87	0.755	4.82	FCI/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
429		³ B ₂ (π → π*)	V	98		3.44	FCI/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
430	Streptocyanine-5	¹ B ₂ (π → π*)	V	85	1.182	3.64	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
431		³ B ₂ (π → π*)	V	97		2.47	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
432	Tetrazine	¹ B _{3u} (n → π*)	V	89	0.006	2.47	CCSDT/AVTZ	Y
433		¹ A _u (n → π*)	V	87		3.69	CCSDT/AVTZ	Y
434		¹ A _g (double)	V	0		4.61	NEVPT2/AVTZ	N
435		¹ B _{1g} (n → π*)	V	83		4.93	CCSDT/AVTZ	Y
436		¹ B _{2u} (π → π*)	V	85	0.055	5.21	CCSDT/AVTZ	Y
437		¹ B _{2g} (n → π*)	V	81		5.45	CCSDT/AVTZ	Y
438		¹ A _u (n → π*)	V	87		5.53	CCSDT/AVTZ	Y
439		¹ B _{3g} (double)	V	0		6.15	NEVPT2/AVTZ	N
440		¹ B _{2g} (n → π*)	V	80		6.12	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
441		¹ B _{1g} (n → π*)	V	85		6.91	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
442		³ B _{3u} (n → π*)	V	97		1.85	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
443		³ A _u (n → π*)	V	96		3.45	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
444		³ B _{1g} (n → π*)	V	97		4.20	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
445		¹ B _{1u} (π → π*)	V	98		4.49	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	N
446		³ B _{2u} (π → π*)	V	97		4.52	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
447		³ B _{2g} (n → π*)	V	96		5.04	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
448		³ A _u (n → π*)	V	96		5.11	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
449		³ B _{3g} (double)	V	5		5.51	NEVPT2/AVTZ	N
450		³ B _{1u} (π → π*)	V	96		5.42	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
451	Thioacetone	¹ A ₂ (n → π*)	V	88		2.53	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
452		¹ B ₂ (n → 3s)	R	91	0.052	5.56	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
453		¹ A ₁ (π → π*)	V	90	0.242	5.88	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
454		¹ B ₂ (n → 3p)	R	92	0.028	6.51	CCSDTQ/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
455		¹ A ₁ (n → 3p)	R	91	0.023	6.61	CCSDTQ/6-31+G(d) + [CCSDT/AVTZ - CCSDT/6-31+G(d)]	Y
456		³ A ₂ (n → π*)	V	97		2.33	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
457		³ A ₁ (π → π*)	V	98		3.45	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
458	Thioacrolein	¹ A ^u (n → π*)	V	86	0.000	2.11	CCSDT/AVTZ	Y
459		³ A ^u (n → π*)	V	96		1.91	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
460	Thioformaldehyde	¹ A ₂ (n → π*)	V	89		2.22	FCI/AVTZ	Y
461		¹ B ₂ (n → 3s)	R	92	0.012	5.96	FCI/AVTZ	Y
462		¹ A ₁ (π → π*)	V	90	0.178	6.38	CCSDTQ/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
463		³ A ₂ (n → π*)	V	97		1.94	FCI/AVTZ	Y
464		³ A ₁ (π → π*)	V	98		3.43	FCI/AVTZ	Y
465		³ B ₂ (n → 3s)	R	97		5.72	FCI/AVDZ + [CCSDT/AVTZ - CCSDT/AVDZ]	Y
466		¹ A ₂ [F](n → π*)	V	87		1.95	FCI/AVTZ	Y
467	Thiophene	¹ A ₁ (π → π*)	V	87	0.070	5.64	CCSDT/AVTZ	Y
468		¹ B ₂ (π → π*)	V	91	0.079	5.98	CCSDT/AVTZ	Y
469		¹ A ₂ (π → 3s)	R	92		6.14	CCSDT/AVTZ	Y
470		¹ B ₁ (π → 3p)	R	90	0.010	6.14	CCSDT/AVTZ	Y
471		¹ A ₂ (π → 3p)	R	91		6.21	CCSDT/AVTZ	Y
472		¹ B ₁ (π → 3s)	R	92	0.000	6.49	CCSDT/AVTZ	Y
473		¹ B ₂ (π → 3p)	R	92	0.082	7.29	CCSDT/AVTZ	Y
474		¹ A ₁ (π → π*)	V	86	0.314	7.31	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	N
475		³ B ₂ (π → π*)	V	98		3.97	FCI/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
476		³ A ₁ (π → π*)	V	97		4.76	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
477		³ B ₁ (π → 3p)	R	96		5.93	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
478		³ A ₂ (π → 3s)	R	97		6.08	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
479	Thiopropynal	¹ A ^u (n → π*)	V	87	0.000	2.03	CCSDT/AVTZ	Y
480		³ A ^u (n → π*)	V	97		1.80	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
481	Triazine	¹ A ^u (n → π*)	V	88		4.72	CCSDT/AVTZ	Y
482		¹ A ^u (n → π*)	V	88	0.014	4.75	CCSDT/AVTZ	Y
483		¹ E ^u (n → π*)	V	88		4.78	CCSDT/AVTZ	Y

Continued on next page

Table 4.12 – Continued from previous page

#	Molecule	Excitation	Nature	%T ₁	f	TBE	Method	Safe?
484		¹ A ₂ '(π → π*)	V	85		5.75	CCSDT/AVTZ	Y
485		¹ A ₁ '(π → π*)	V	90		7.24	CCSDT/AVTZ	Y
486		¹ E'(n → 3s)	R	90	0.016	7.32	CCSDT/AVTZ	Y
487		¹ E''(n → π*)	V	82		7.78	CCSDT/AVTZ	Y
488		¹ E'(π → π*)	V	90	0.451	7.94	CCSDT/AVTZ	Y
489		³ A ₂ ''(n → π*)	V	96		4.33	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
490		³ E''(n → π*)	V	96		4.51	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
491		³ A ₁ ''(n → π*)	V	96		4.73	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
492		³ A ₁ '(π → π*)	V	98		4.85	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
493		³ E'(π → π*)	V	96		5.59	CCSDT/6-31+G(d) + [CC3/AVTZ - CC3/6-31+G(d)]	Y
494		³ A ₂ '(π → π*)	V	97		6.62	CCSDT/AVDZ + [CC3/AVTZ - CC3/AVDZ]	Y
495	Water	¹ B ₁ (n → 3s)	R	93	0.054	7.62	FCI/AVTZ	Y
496		¹ A ₂ (n → 3p)	R	93		9.41	FCI/AVTZ	Y
497		¹ A ₁ (n → 3s)	R	93	0.100	9.99	FCI/AVTZ	Y
498		³ B ₁ (n → 3s)	R	98		7.25	FCI/AVTZ	Y
499		³ A ₂ (n → 3p)	R	98		9.24	FCI/AVTZ	Y
500		³ A ₁ (n → 3s)	R	98		9.54	FCI/AVTZ	Y

4.5 Benchmarks

In this section, we report a comprehensive benchmark of various lower-order methods on the entire set of closed-shell compounds belonging to the QUEST database. Statistical quantities are reported in Table 4.14. Additionally, we also provide a specific analysis for each type of excited states. Hence, the statistical values are reported for various types of excited states and molecular sizes for the MSE and MAE. The distribution of the errors in vertical excitation energies (with respect to the TBE/*aug-cc-pVTZ* reference values) are represented in Figure 4.15 for all the “safe” excitations having a dominant single excitation character (*i.e.*, the double excitations are discarded).

TABLE 4.13: Theoretical best estimates TBEs (in eV) for the doublet-doublet transitions of the open-shell molecules belonging to QUEST#4. These TBEs are obtained with the *aug-cc-pVTZ* basis set, and “Method” indicates the protocol employed to compute them.

#	Molecule	Transition	TBE/ <i>aug-cc-pVTZ</i>	Method
1	Allyl	2B_1	3.39	FCI/6-31+G(d) + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/6-31+G(d)]
2		2A_1	4.99	FCI/6-31+G(d) + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/6-31+G(d)]
3	BeF	$^2\Pi$	4.14	FCI/ <i>aug-cc-pVTZ</i>
4		$^2\Sigma^+$	6.21	FCI/ <i>aug-cc-pVTZ</i>
5	BeH	$^2\Pi$	2.49	FCI/ <i>aug-cc-pVTZ</i>
6		$^2\Pi$	6.46	FCI/ <i>aug-cc-pVTZ</i>
7	BH ₂	2B_1	1.18	FCI/ <i>aug-cc-pVTZ</i>
8	CH	$^2\Delta$	2.91	FCI/ <i>aug-cc-pVTZ</i>
9		$^2\Sigma^-$	3.29	FCI/ <i>aug-cc-pVTZ</i>
10		$^2\Sigma^+$	3.98	FCI/ <i>aug-cc-pVTZ</i>
11	CH ₃	$^2A'_1$	5.85	FCI/ <i>aug-cc-pVTZ</i>
12		$^2E'$	6.96	FCI/ <i>aug-cc-pVTZ</i>
13		$^2E'$	7.18	FCI/ <i>aug-cc-pVTZ</i>
14		$^2A''_2$	7.65	FCI/ <i>aug-cc-pVTZ</i>
15	CN	$^2\Pi$	1.34	FCI/ <i>aug-cc-pVTZ</i>
16		$^2\Sigma^+$	3.22	FCI/ <i>aug-cc-pVTZ</i>
17	CNO	$^2\Sigma^+$	1.61	FCI/ <i>aug-cc-pVTZ</i>
18		$^2\Pi$	5.49	FCI/6-31+G(d) + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/6-31+G(d)]
19	CON	$^2\Pi$	3.53	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
20		$^2\Sigma^+$	3.86	CCSDTQ/6-31+G(d) + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/6-31+G(d)]
21	CO ⁺	$^2\Pi$	3.28	FCI/ <i>aug-cc-pVTZ</i>
22		$^2\Sigma^+$	5.81	FCI/ <i>aug-cc-pVDZ</i>
23	F ₂ BO	2B_1	0.73	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
24		2A_1	2.80	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
25	F ₂ BS	2B_1	0.51	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
26		2A_1	2.99	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
27	H ₂ BO	2B_1	2.15	FCI/ <i>aug-cc-pVTZ</i>
28		2A_1	3.49	FCI/ <i>aug-cc-pVTZ</i>
29	HCO	$^2A''$	2.09	FCI/ <i>aug-cc-pVTZ</i>
30		$^2A'$	5.45	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
31	HOC	$^2A''$	0.92	FCI/ <i>aug-cc-pVTZ</i>
32	H ₂ PO	$^2A''$	2.80	FCI/ <i>aug-cc-pVTZ</i>
33		$^2A'$	4.21	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
34	H ₂ PS	$^2A''$	1.16	FCI/ <i>aug-cc-pVTZ</i>
35		$^2A'$	2.72	FCI/ <i>aug-cc-pVTZ</i>
36	NCO	$^2\Sigma^+$	2.89	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
37		$^2\Pi$	4.73	FCI/ <i>aug-cc-pVDZ</i> + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/ <i>aug-cc-pVDZ</i>]
38	NH ₂	2A_1	2.12	FCI/ <i>aug-cc-pVTZ</i>
39	Nitromethyl	2B_2	2.05	CCSDT/ <i>aug-cc-pVTZ</i>
40		2A_2	2.38	CCSDT/ <i>aug-cc-pVTZ</i>
41		2A_1	2.56	CCSDT/ <i>aug-cc-pVTZ</i>
42		2B_1	5.35	CCSDT/ <i>aug-cc-pVTZ</i>
43	NO	$^2\Sigma^+$	6.13	FCI/ <i>aug-cc-pVTZ</i>
44		$^2\Sigma^+$	7.29	CCSDTQ/ <i>aug-cc-pVTZ</i>
45	OH	$^2\Sigma^+$	4.10	FCI/ <i>aug-cc-pVTZ</i>
46		$^2\Sigma^-$	8.02	FCI/ <i>aug-cc-pVTZ</i>
47	PH ₂	2A_1	2.77	FCI/ <i>aug-cc-pVTZ</i>
48	Vinyl	$^2A''$	3.26	FCI/ <i>aug-cc-pVTZ</i>
49		$^2A''$	4.69	FCI/ <i>aug-cc-pVTZ</i>
50		$^2A'$	5.60	FCI/ <i>aug-cc-pVTZ</i>
51		$^2A'$	6.20	FCI/6-31+G(d) + [CCSDT/ <i>aug-cc-pVTZ</i> - CCSDT/6-31+G(d)]

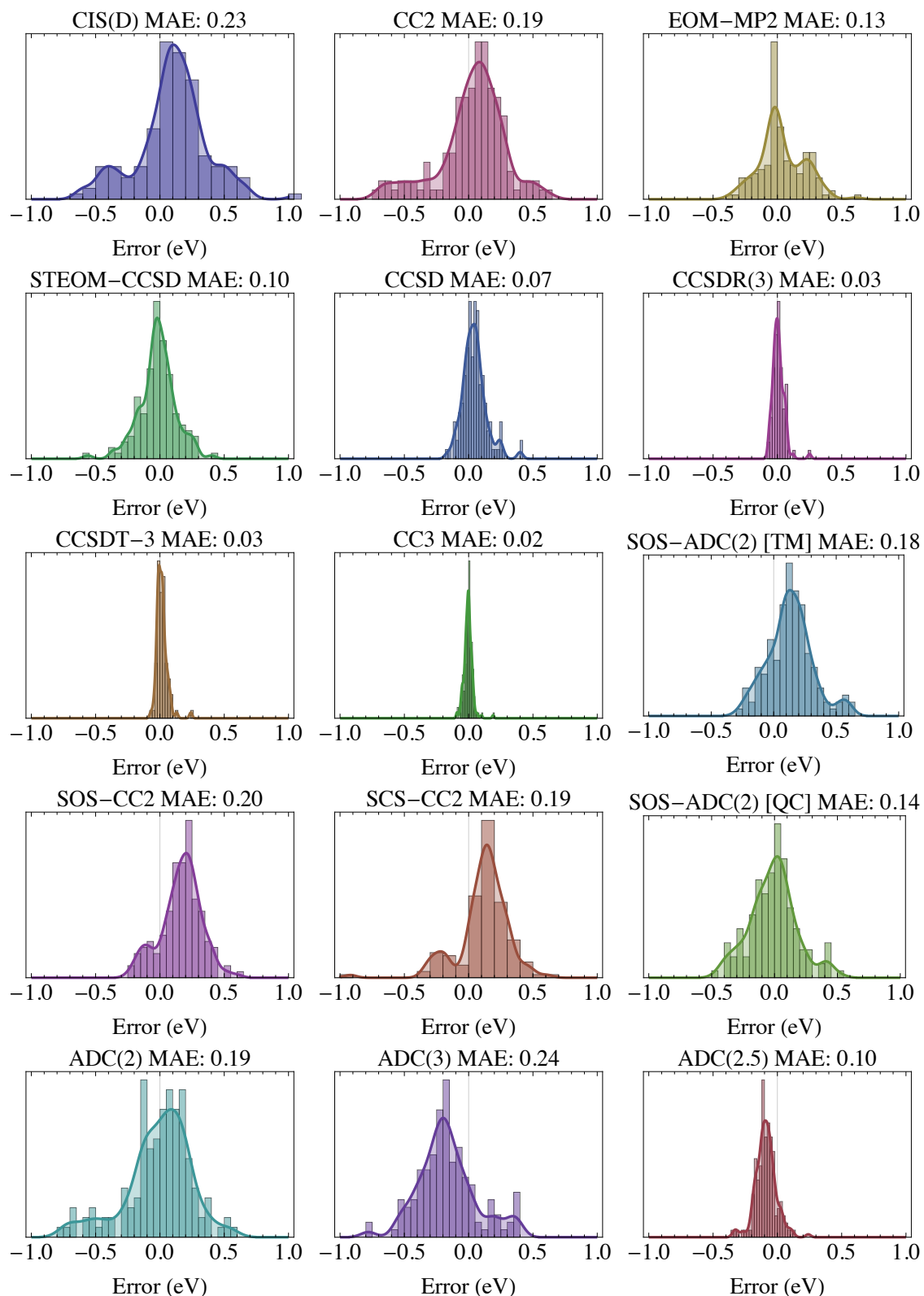


FIGURE 4.5: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the molecules of the QUEST database containing from one to three non-hydrogen atoms (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

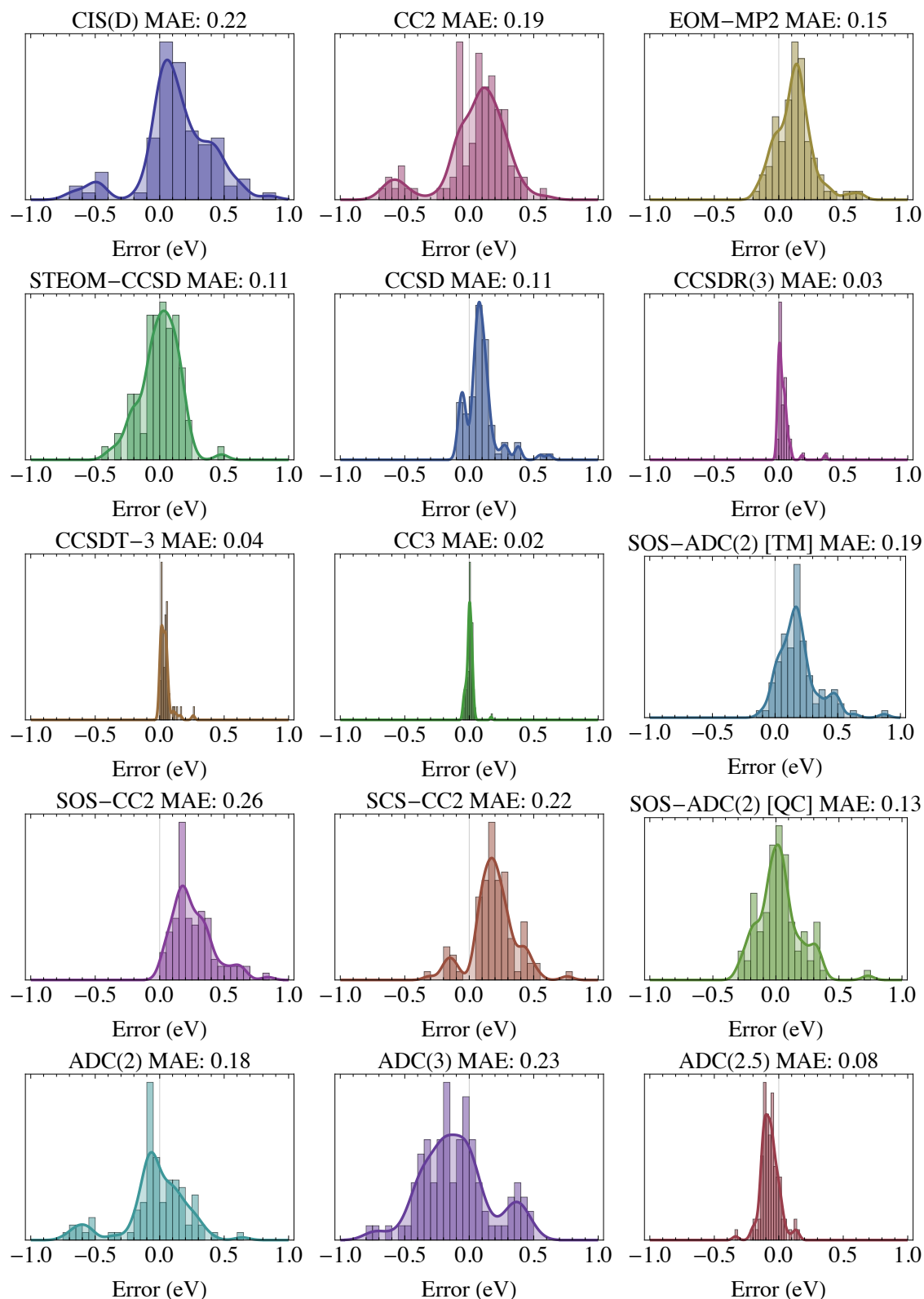


FIGURE 4.6: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the molecules of the QUEST database containing four non-hydrogen atoms (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

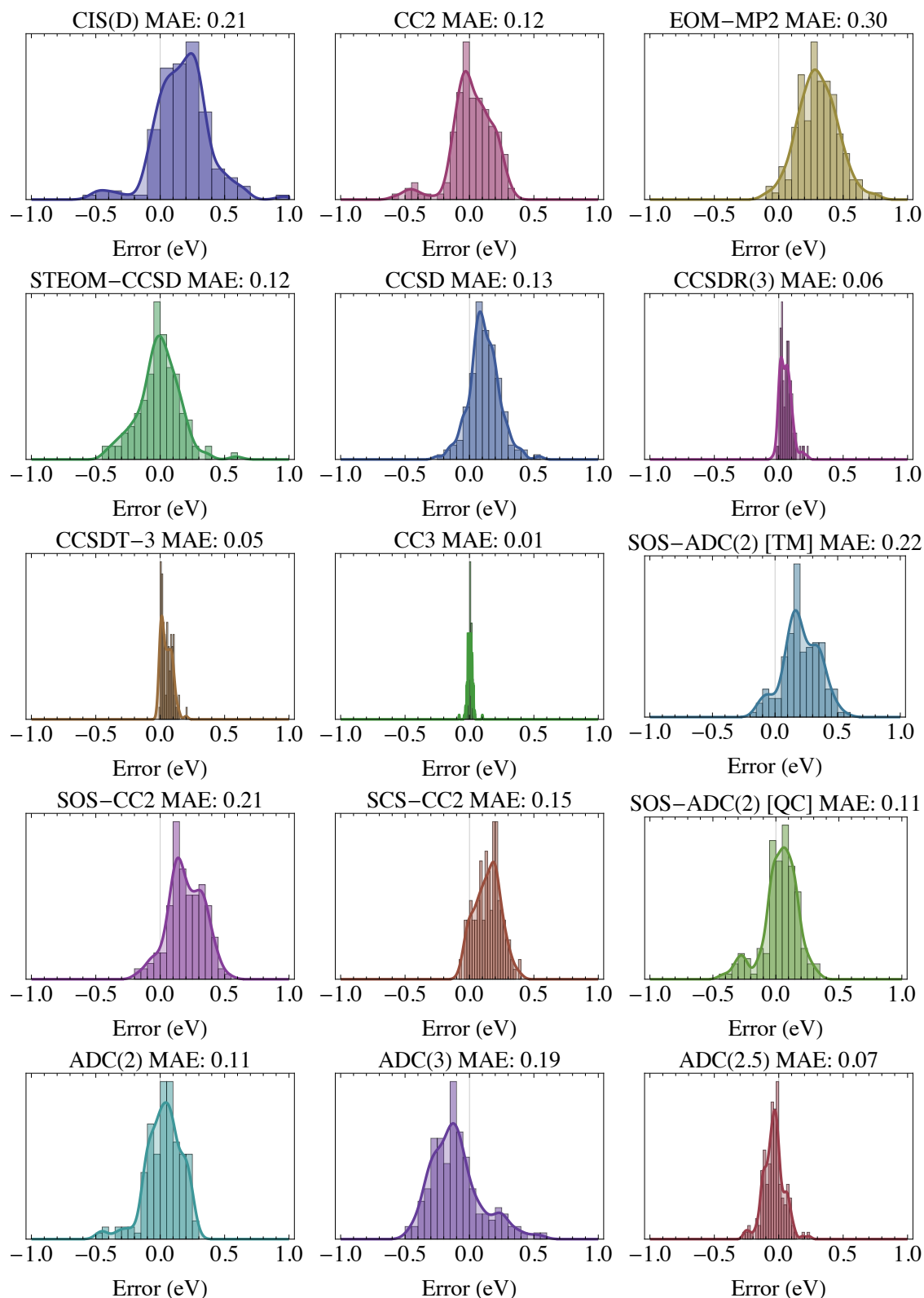


FIGURE 4.7: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the molecules of the QUEST database containing five or six non-hydrogen atoms (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

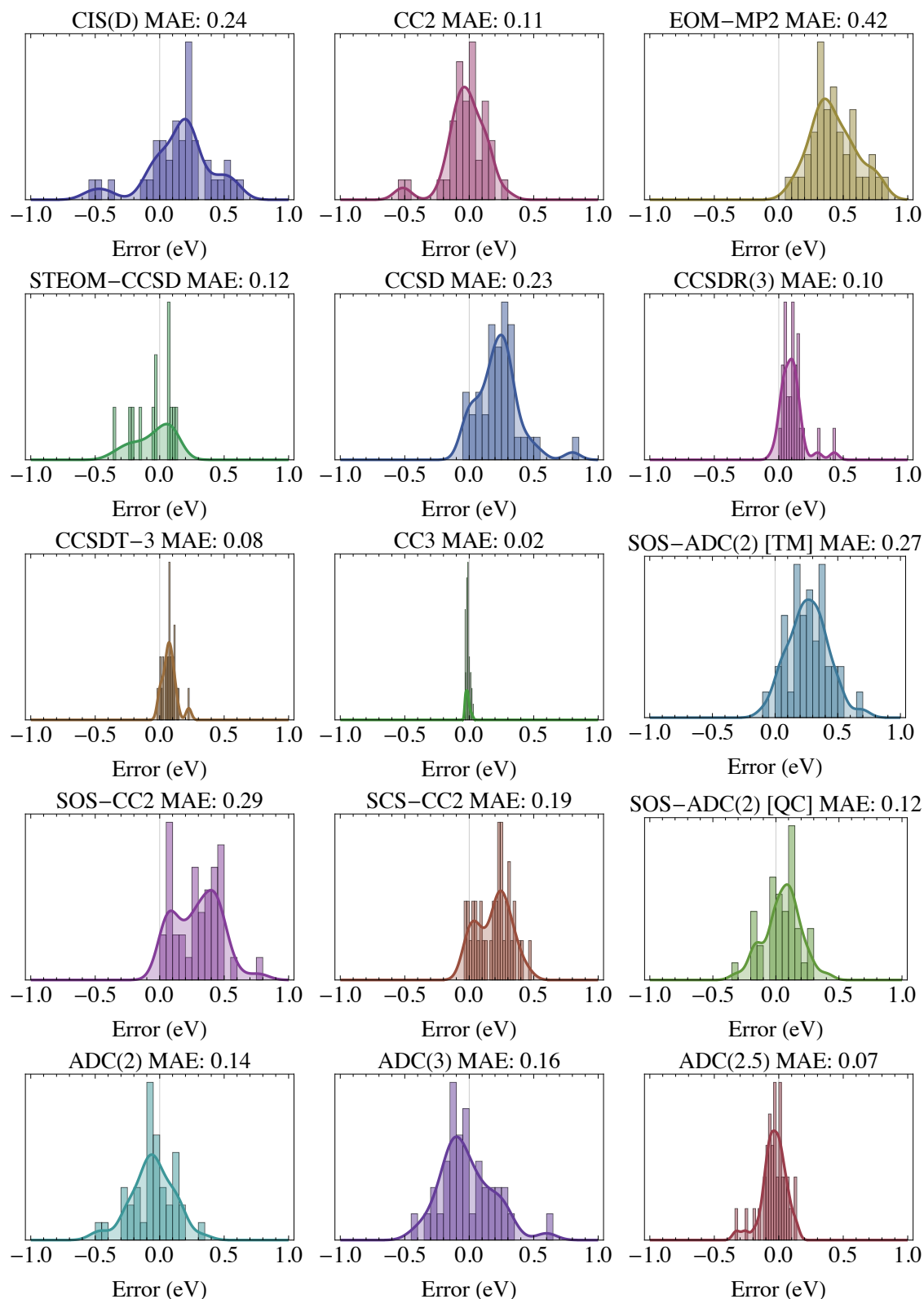


FIGURE 4.8: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the molecules of the QUEST database containing from 7 and 10 non-hydrogen atoms (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

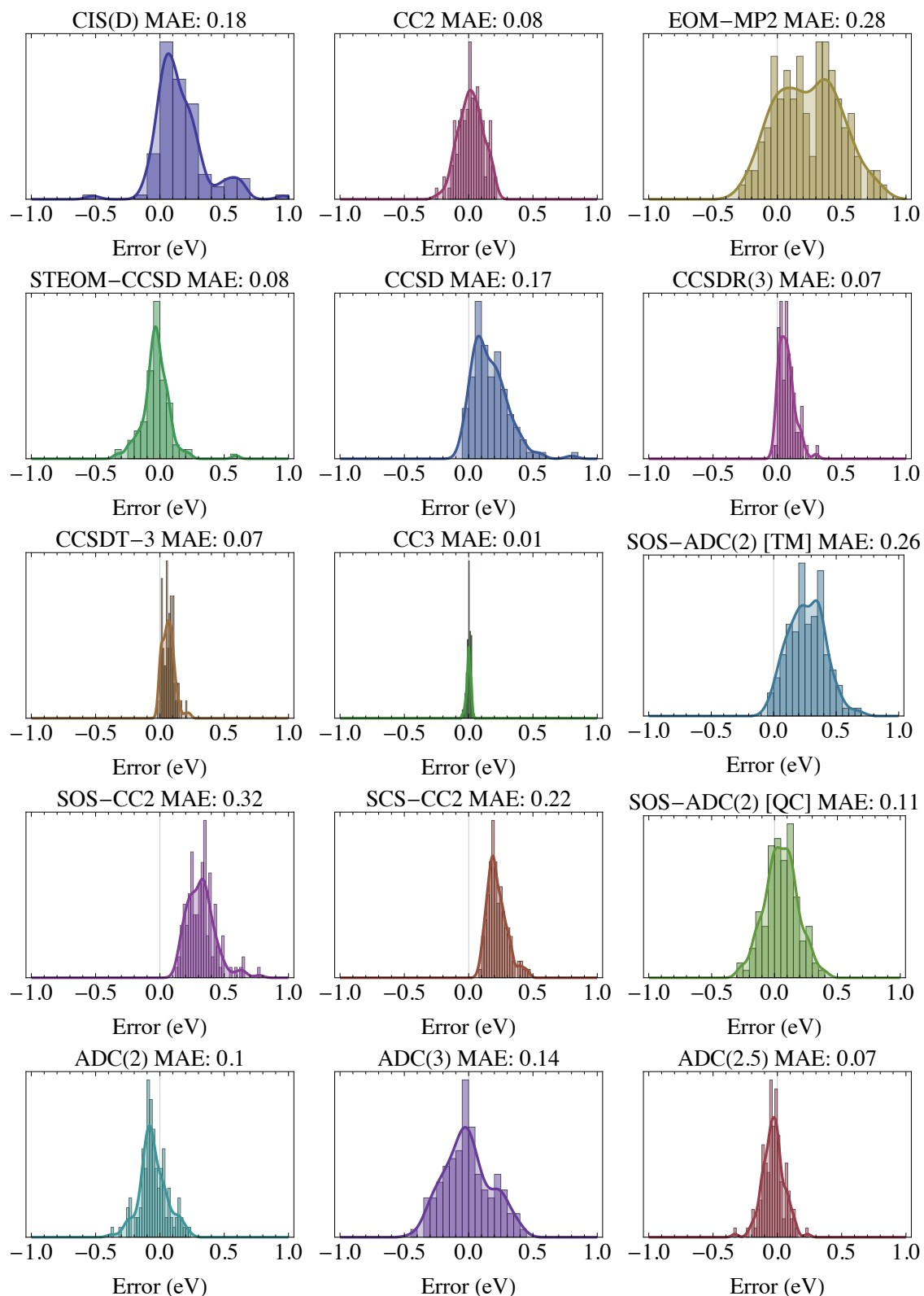


FIGURE 4.9: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the $n \rightarrow \pi^*$ excitations of the QUEST database (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

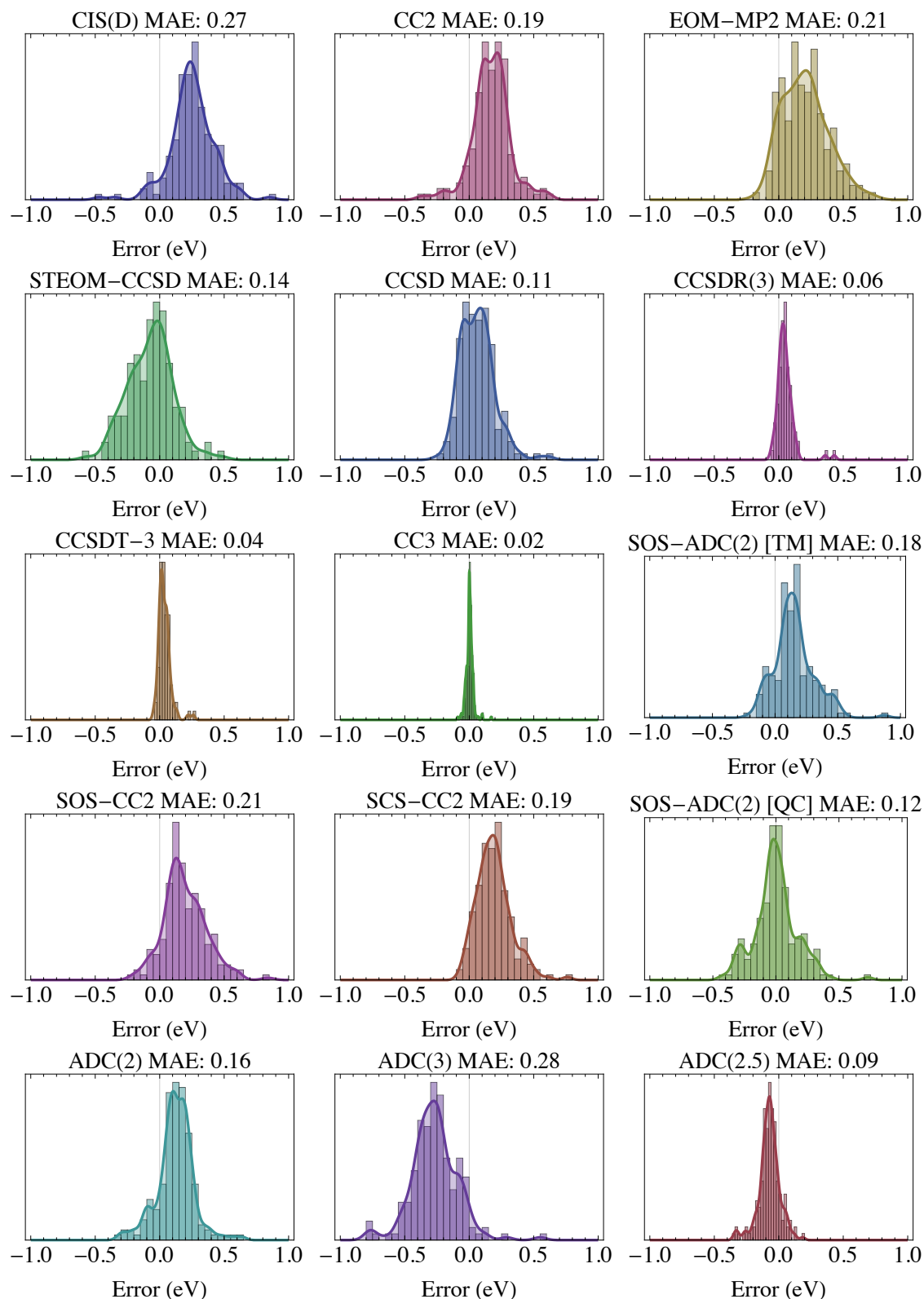


FIGURE 4.10: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the $\pi \rightarrow \pi^*$ excitations of the QUEST database (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

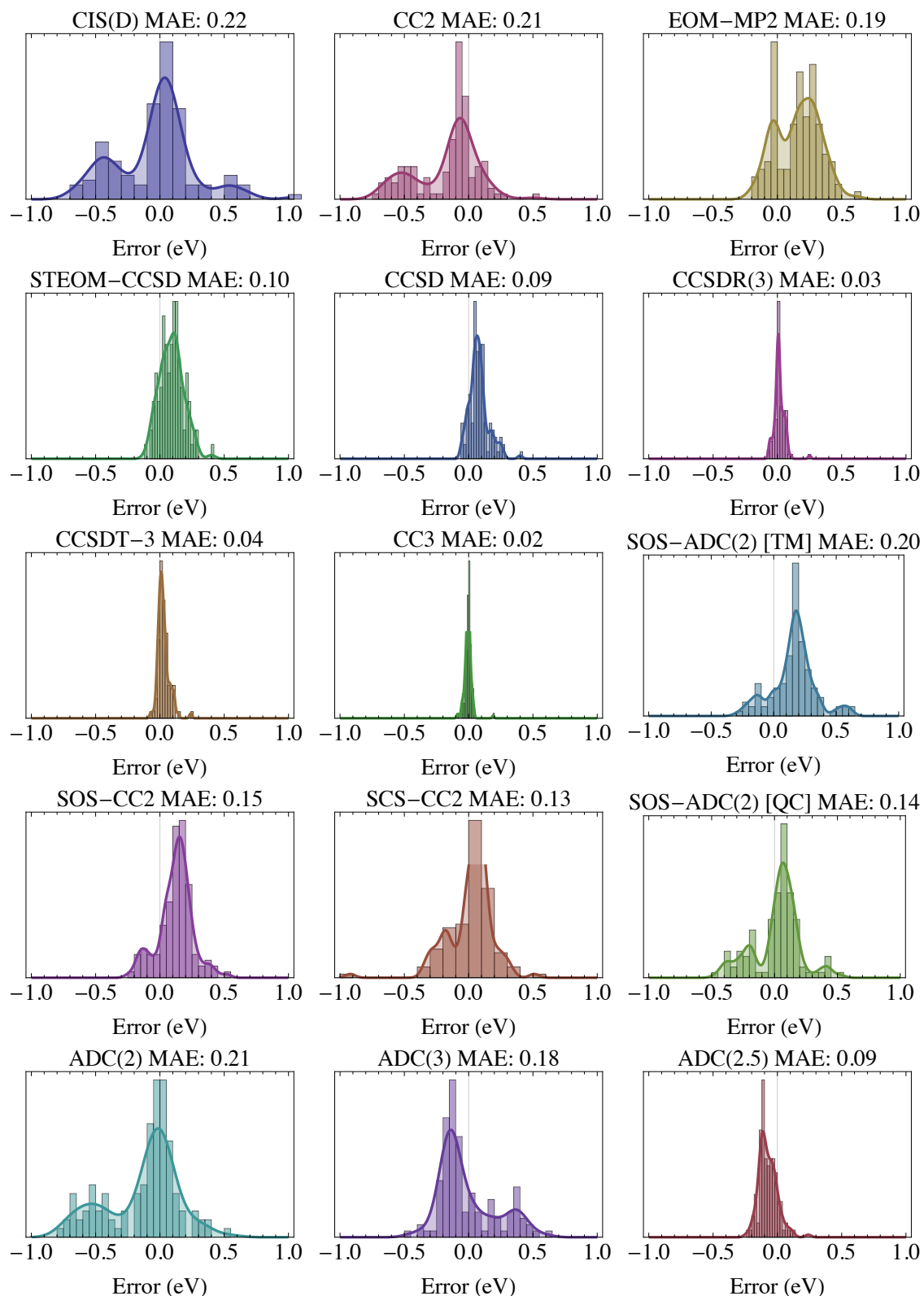


FIGURE 4.11: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the Rydberg excitations of the QUEST database (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

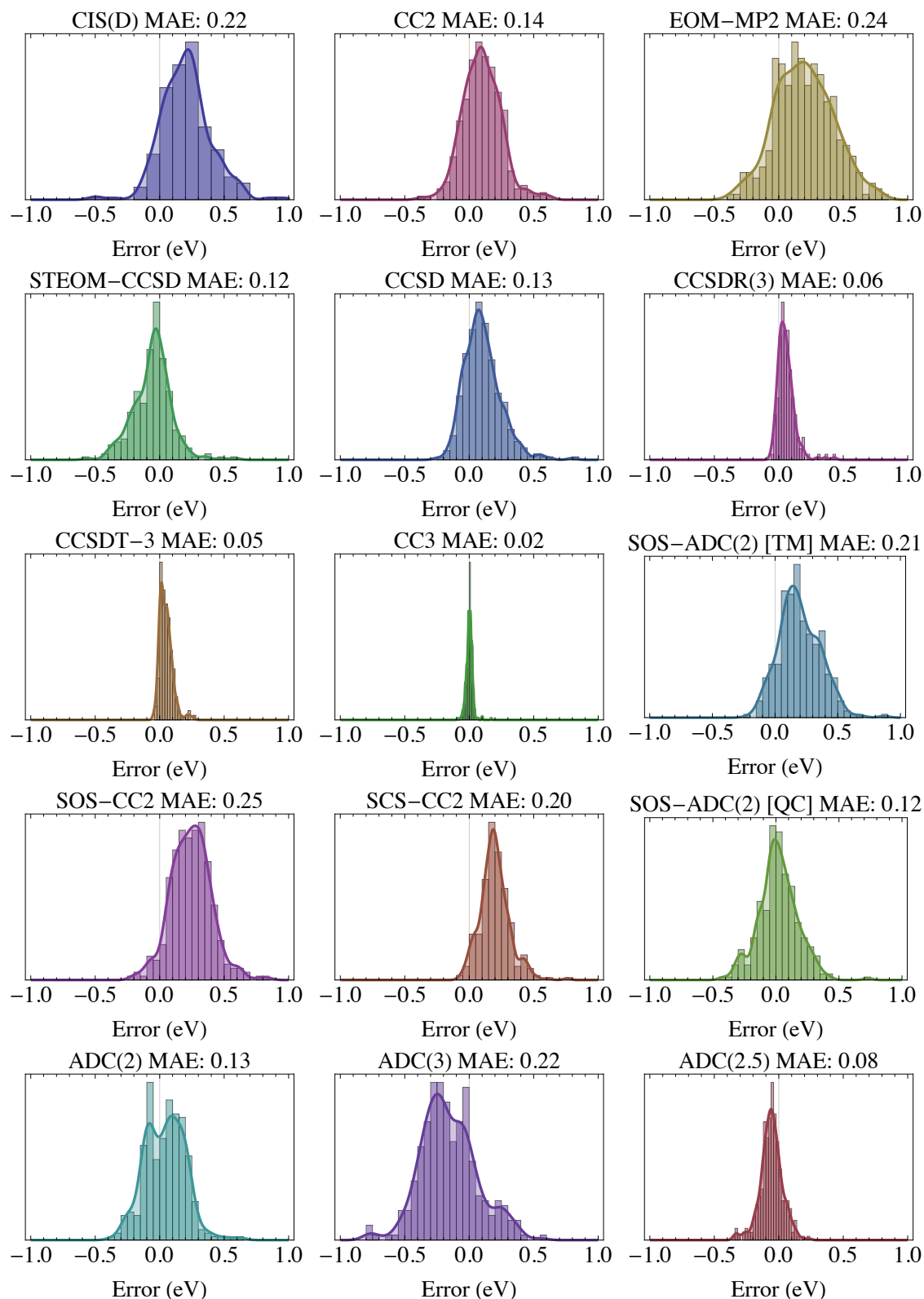


FIGURE 4.12: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the valence excitations of the QUEST database (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

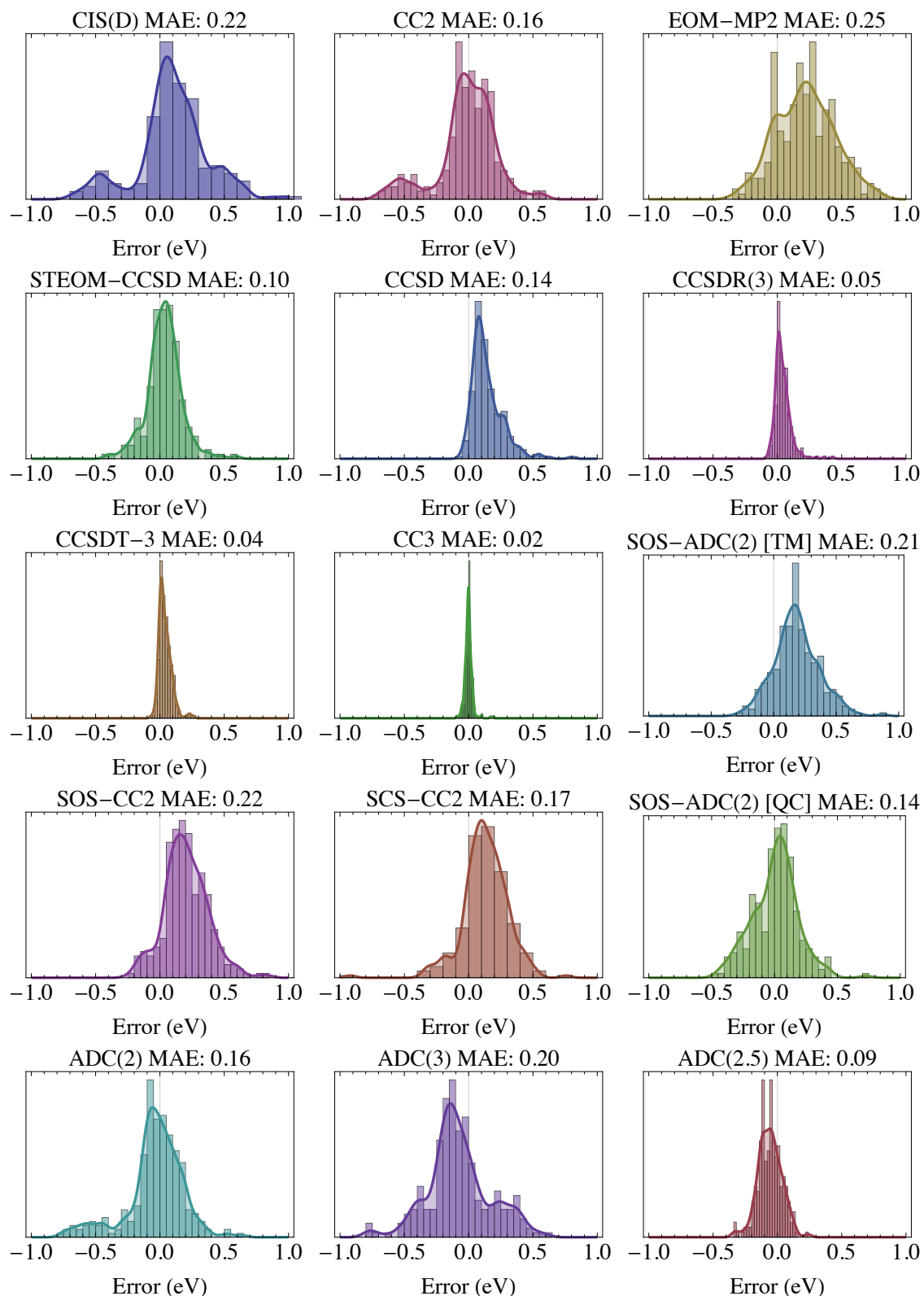


FIGURE 4.13: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the singlet excitations of the QUEST database (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

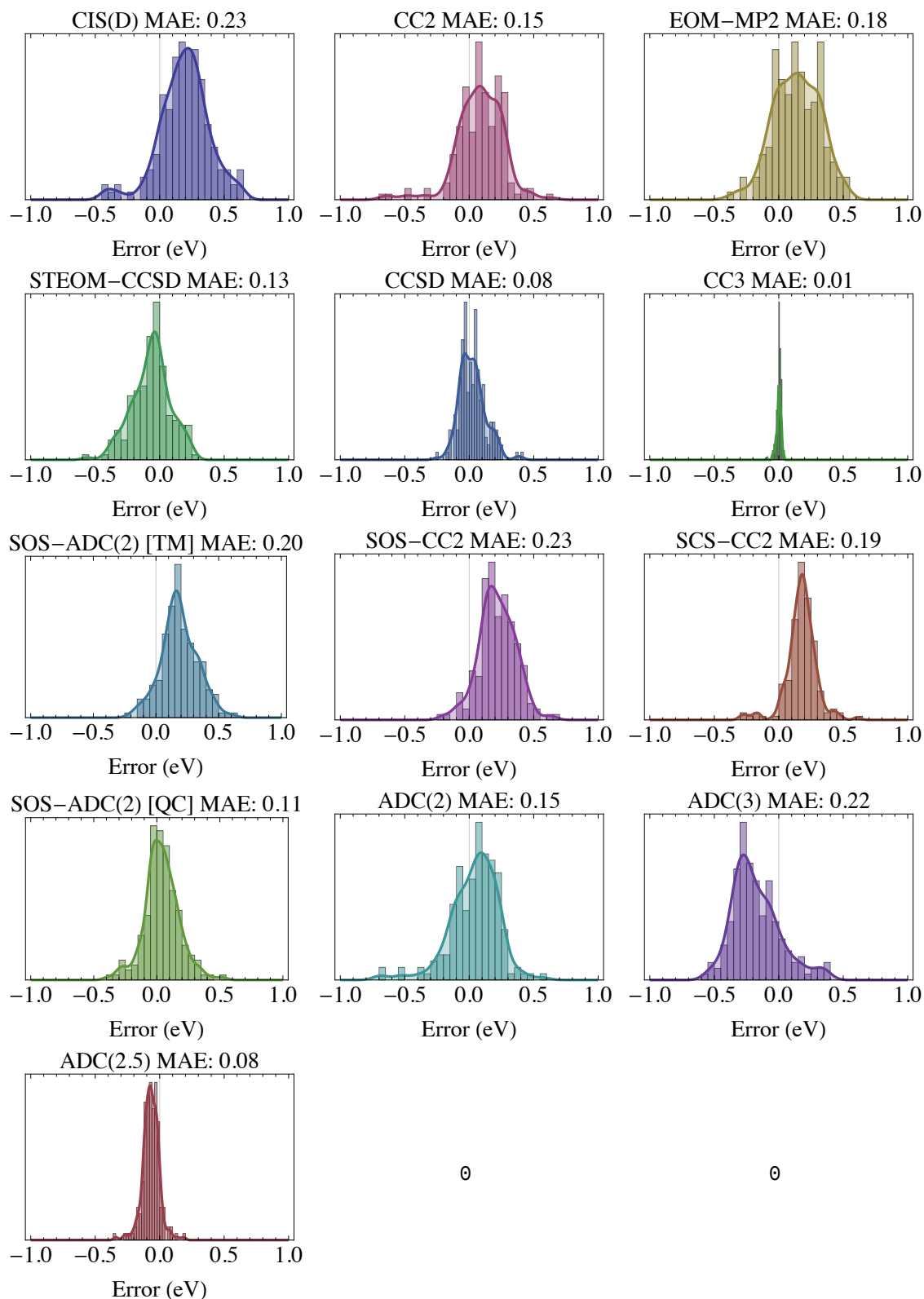


FIGURE 4.14: Distribution of the error (in eV) in excitation energies (with respect to the TBE/aug-cc-pVTZ values) for various methods for the triplet excitations of the QUEST database (closed-shell compounds only). Only the “safe” TBEs are considered Table 4.12. See Table in the main text for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

TABLE 4.14: Mean signed error (MSE), mean absolute error (MAE), root-mean-square error (RMSE), standard deviation of the errors (SDE), as well as the maximum positive error [Max(+)] and negative error [Max(-)] with respect to the TBE/*aug-cc-pVTZ* for the entire QUEST database. Only the “safe” TBEs are considered Table 4.12. For the MSE and MAE, the statistical values are reported for various types of excited states and molecular sizes. All quantities are given in eV. “Count” refers to the number of transitions considered for each method.

	CIS(D)	CC2	EOM-MP2	STEOM-CCSD	CCSD	CCSDR(3)	CCCSDT-3	CC3	SOS-ADC(2) ^a	SOS-CC2 ^a	SCS-CC2 ^a	SOS-ADC(2) ^b	ADC(2)	ADC(3)	ADC(2.5)
Count	429	431	427	360	431	259	251	431	430	430	430	430	426	423	423
Max(+)	1.06	0.63	0.80	0.59	0.80	0.43	0.26	0.19	0.87	0.84	0.76	0.73	0.64	0.60	0.24
Max(-)	-0.69	-0.71	-0.38	-0.56	-0.25	-0.07	-0.07	-0.09	-0.29	-0.24	-0.92	-0.46	-0.76	-0.79	-0.34
MSE	0.13	0.02	0.18	-0.01	0.10	0.04	0.04	0.00	0.18	0.21	0.15	0.02	-0.01	-0.12	-0.06
singlet	0.10	-0.02	0.22	0.03	0.14	0.04	0.04	0.00	0.18	0.20	0.13	0.00	-0.04	-0.08	-0.06
triplet	0.19	0.08	0.14	-0.07	0.03			0.00	0.19	0.22	0.17	0.04	0.04	-0.18	-0.07
valence	0.20	0.10	0.20	-0.06	0.10	0.06	0.05	0.00	0.19	0.24	0.20	0.02	0.04	-0.16	-0.06
Rydberg	-0.04	-0.17	0.15	0.09	0.08	0.01	0.03	-0.01	0.16	0.12	0.01	0.02	-0.13	-0.02	-0.07
$n \rightarrow \pi^*$	0.16	0.02	0.24	-0.03	0.17	0.07	0.07	0.00	0.26	0.32	0.22	0.05	-0.05	-0.01	-0.03
$\pi \rightarrow \pi^*$	0.25	0.17	0.20	-0.07	0.06	0.05	0.04	0.00	0.15	0.19	0.19	0.00	0.12	-0.27	-0.07
1-3 non-H	0.10	0.03	0.03	-0.02	0.04	0.01	0.01	0.00	0.13	0.16	0.11	-0.01	-0.01	-0.17	-0.09
4 non-H	0.13	0.04	0.12	0.00	0.09	0.03	0.04	0.00	0.19	0.26	0.19	0.03	-0.04	-0.10	-0.07
5-6 non-H	0.17	0.02	0.30	-0.01	0.11	0.05	0.05	0.00	0.21	0.20	0.14	0.03	0.03	-0.10	-0.04
7-10 non-H	0.15	-0.03	0.42	-0.05	0.22	0.10	0.08	-0.01	0.26	0.29	0.19	0.05	-0.06	-0.02	-0.04
SDE	0.24	0.20	0.21	0.13	0.12	0.05	0.04	0.02	0.17	0.16	0.16	0.15	0.20	0.22	0.08
RMSE	0.29	0.22	0.28	0.15	0.16	0.07	0.06	0.03	0.25	0.26	0.22	0.17	0.21	0.26	0.10
MAE	0.22	0.16	0.22	0.11	0.12	0.05	0.04	0.02	0.20	0.22	0.18	0.13	0.15	0.21	0.08
singlet	0.22	0.16	0.25	0.10	0.14	0.05	0.04	0.02	0.21	0.22	0.17	0.14	0.16	0.20	0.09
triplet	0.23	0.15	0.18	0.12	0.08			0.01	0.20	0.23	0.19	0.11	0.15	0.22	0.08
valence	0.22	0.14	0.24	0.12	0.13	0.06	0.05	0.02	0.21	0.25	0.20	0.12	0.13	0.22	0.08
Rydberg	0.22	0.21	0.19	0.10	0.08	0.03	0.03	0.02	0.20	0.15	0.13	0.14	0.21	0.18	0.09
$n \rightarrow \pi^*$	0.18	0.08	0.28	0.08	0.17	0.07	0.07	0.01	0.26	0.32	0.22	0.11	0.10	0.14	0.07
$\pi \rightarrow \pi^*$	0.27	0.19	0.21	0.14	0.11	0.06	0.04	0.02	0.18	0.21	0.20	0.12	0.16	0.28	0.09
1-3 non-H	0.23	0.19	0.13	0.10	0.07	0.03	0.03	0.02	0.18	0.20	0.19	0.14	0.19	0.24	0.10
4 non-H	0.22	0.19	0.15	0.11	0.11	0.03	0.04	0.02	0.19	0.26	0.22	0.13	0.18	0.23	0.08
5-6 non-H	0.21	0.12	0.30	0.12	0.13	0.06	0.05	0.01	0.22	0.21	0.15	0.11	0.11	0.19	0.07
7-10 non-H	0.24	0.11	0.42	0.12	0.23	0.10	0.08	0.02	0.27	0.29	0.19	0.12	0.14	0.16	0.07

^a Excitation energies computed with TURBOMOLE.

^b Excitation energies computed with Q-CHEM.

The most striking feature from the statistical indicators gathered in Table 4.14 is the overall accuracy of CC3 with MAEs and MSEs systematically below the chemical accuracy threshold (errors < 0.043 eV or 1 kcal mol⁻¹), irrespective of the nature of the transition and the size of the molecule. CCSDR(3) and CCSDT-3 can also be regarded as excellent performers with overall MAEs below 0.05 eV, though one would notice a slight degradation of their performances for the $n \rightarrow \pi^*$ excitations and the largest molecules of the database. The other third-order method, ADC(3), which enjoys a lower computational cost, is significantly less accurate and does not really improve upon its second-order analog, even for the largest systems considered here, an observation in line with a previous analysis by some of the authors [33]. Nonetheless, ADC(3)'s accuracy improves in larger compounds, with a MAE of 0.24 eV (0.16 eV) for the subsets of the most compact (extended) compounds considered herein. The ADC(2.5) composite method introduced in Ref. [33], which corresponds to grossly average the ADC(2) and ADC(3) values, yields an appreciable accuracy improvement, as shown in Figure 4.15. Indeed, we note that the MAE of 0.07 eV obtained for "large" compounds is comparable to the one obtained with CCSDR(3) and CCSDT-3 for these molecules. All these third-order methods are rather equally efficient for valence and Rydberg transitions.

Concerning the second-order methods (which have the indisputable advantage to be applicable to larger molecules than the ones considered here), we have the following ranking in terms of MAEs: EOM-MP2 \approx CIS(D) $<$ CC2 \approx ADC(2) $<$ CCSD \approx STEOM-CCSD, which fits our previous conclusions on the specific subsets [30, 32, 33, 36, 221]. A very similar ranking is obtained when one looks at the MSEs. It is noteworthy that the performances of EOM-MP2 and CCSD are getting notably worse when the system size increases, while CIS(D) and STEOM-CCSD have a very stable behavior with respect to system size. Indeed, the EOM-MP2 MAE attains 0.42 eV for molecules containing between 7 and 10 non-hydrogen atoms, whereas the CCSD tendency to overshoot the transition energies yield a MSE of 0.22 eV for the same set (a rather large error). For CCSD, this conclusion fits benchmark studies published by other groups [196, 324, 388–391]. For example, Kánnár and Szalay obtained a MAE of 0.18 eV on Thiel's set for the states exhibiting a dominant single excitation character. The CCSD degradation with system size might partially explain the similar (though less pronounced) trend obtained for CCSDR(3). Regarding the apparently better performances of STEOM-CCSD as compared to CCSD, we recall that several challenging states have been naturally removed from the STEOM-CCSD statistics because the active character percentage was lower than 98 % (see above). In contrast to EOM-MP2 and CCSD, the overall accuracy of CC2 and ADC(2) does significantly improve for larger molecules, the performances of the two methods being, as expected, similar [29]. Let us note that these two methods show similar accuracies for singlet and triplet transitions, but are significantly less accurate for Rydberg transitions, as already pointed out previously [391]. Therefore, both CC2 and ADC(2) offer an appealing cost-to-accuracy ratio for large compounds, which explains their popularity in realistic chemical scenarios [172, 212, 213, 216, 218, 224]. For the scaled methods [SOS-ADC(2), SOS-CC2, and SCS-CC2], the TURBOMOLE scaling factors do not seem to improve things upon the unscaled versions, while the Q-CHEM scaling factors for ADC(2) provide a small, yet significant improvement for this set of molecules. Of course, one of the remaining open questions regarding all these methods is their accuracy for even larger systems.

4.6 The QUESTDB website

Quite a large number of calculations were required for each of the QUEST articles [30, 32, 33, 36, 221]. Up to now, all the curated data was shared as supplementary information presented as a file in portable document format (pdf). This way of sharing data does not require too much effort for the authors, but it is obviously not optimal from the user's point of view. We have now addressed this problem by creating a database which contains all the vertical and fluorescence transition energies as well the corresponding molecular geometries. This data can be manipulated via a web application which allows to plot the statistical indicators (generated with the Plotly library) computed on selected subsets of molecules, methods and basis sets. The application also gives the

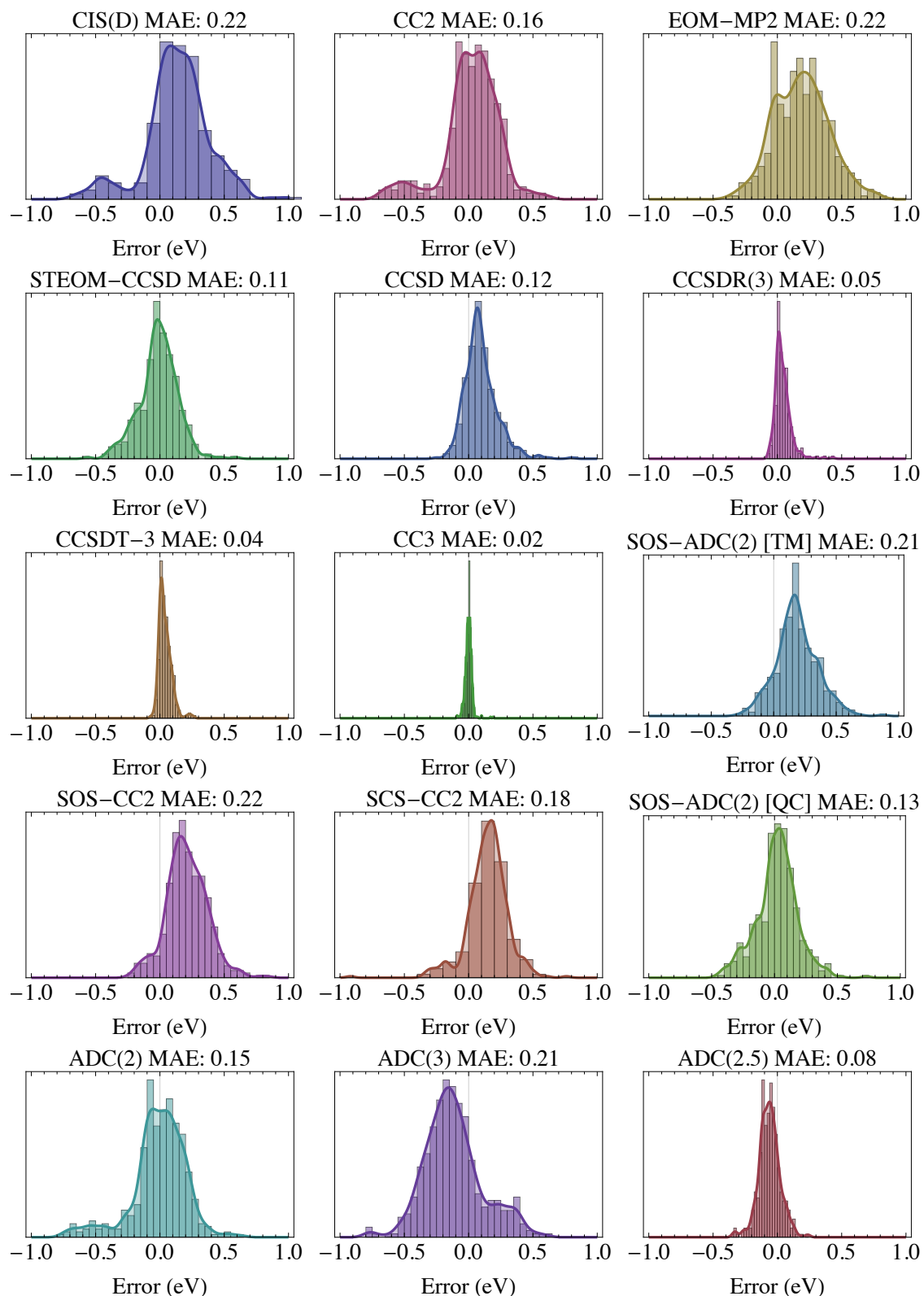


FIGURE 4.15: Distribution of the error (in eV) in excitation energies (with respect to the TBE/*aug-cc-pVTZ* values) for various methods for the entire QUEST database considering only closed-shell compounds. Only the “safe” TBEs are considered Table 4.12. See Table 4.14 for the values of the corresponding statistical quantities. QC and TM indicate that Q-CHEM and TURBOMOLE scaling factors are considered, respectively. The SOS-CC2 and SCS-CC2 approaches are obtained with the latter code.

possibility to the user to import external data files, in order to compare the performance of methods that are not in our database. Both the web application and the data are hosted in a single GitHub repository (https://github.com/LCPQ/QUESTDB_website) and available at the following address: https://lcpq.github.io/QUESTDB_website. In this way, extending the database is as simple as adding new data files to the repository, together with the corresponding bibliographic references, and we strongly encourage users to contribute to enlarge this database via GitHub pull requests.

4.6.1 Website specifications

The specifications of the website are as follows.

- Display and export excitation energy values of the QUEST project as tables;
- Download of various data such as geometries and excitation energies;
- Import local files from the user's computer;
- Data filtering using various physical and chemical parameters;
- Calculate the statistics based on these parameters;
- Display box plots to easily visualize the accuracy of the methods and the corresponding statistics.

4.6.2 Architecture

The architecture of the website is designed to be simple and facilitate the integration of new data. It is composed of two parts.

- A static website designed to view data and generate statistics;
- A series of PYTHON tools used to generate readable data through the website.

The static website

The static website is the main component of this project. All statistical calculations are performed locally on the Dataset page. The server is only used to serve the pages and the data of the QUEST project to the client. If one wants to work with data from the QUEST project, one must access the Dataset page. Firstly, the website offers the possibility to import new data (see section 4.6.2) but also to export some of them (*e.g.*, geometries) as simple text files. These imported data are temporarily added to the current session (and deleted after leaving the page). There are four multi-selection list. Each list depends on the previous ones. These lists allow to filter information from the QUEST subsets (see Figure 4.1), such as the molecules (see Figure 4.4), the methods, or the basis sets (see section 4.2.3). The next step is to choose the type of excitations to include using the dedicated filters. We also provide the possibility of filtering with respect to the size of the molecules or by the percentage of single excitations involved in a given transition ($%T_1$). After that, one must define a reference method among previously selected methods (default TBE). We also provide a flag to remove all values declared as "unsafe". As defined previously, a value is defined as "unsafe" if we consider that it is not chemically accurate.

Statistical calculations Our aim is to assess the accuracy of each method for a given basis set with respect to the reference (usually TBEs). For each method, we define a vector containing all the energies of the vertical transitions selected by the user. Defining as "meth" a couple method/basis, E_{meth}^X the vertical excitation energy associated with the transition X computed at the meth level of

```

% Commands to apply in tabular cells
\newcommand{\command}{text}
\newcommand{\command2}{text2}
% Options for the importation script
\begin{dfbOptions}
  % Choose the handler
  \format{mySuperTabularFormat}
  % Other options
  \initialState{\sim A_1}
  \initialState[Acetaldehyde]{\sim A^\prime}
  \initialState[Acetylene]{\sim \Sigma_g}
\end{dfbOptions}

\begin{tabular}
  % Data to import
\end{tabular}

```

FIGURE 4.16: L^AT_EX input file skeleton

theory and $e_{\text{meth}}^X = E_{\text{meth}}^X - E_{\text{ref}}^X$ the error of the method meth for the transition X with respect to the reference excitation energy E_{ref}^X , we have

$$\mathbf{E}_{\text{meth}} = \{E_{\text{meth}'}^1, \dots, E_{\text{meth}'}^X, \dots\} \quad (4.11)$$

$$\mathbf{e}_{\text{meth}} = \{e_{\text{meth}'}^1, \dots, e_{\text{meth}'}^X, \dots\} \quad (4.12)$$

Hence, the relevant statistical data that inform us about the accuracy of a given method relative to the reference are

$$\text{MSE}_{\text{meth}} = \overline{e_{\text{meth}}} = \frac{1}{\mathcal{X}} \sum_{X=1}^{\mathcal{X}} e_{\text{meth}}^X \quad (4.13)$$

$$\text{MAE}_{\text{meth}} = \overline{|e_{\text{meth}}|} \quad (4.14)$$

$$\text{RMSE}_{\text{meth}} = \sqrt{\overline{e_{\text{meth}}^2}} \quad (4.15)$$

$$\text{SDE}_{\text{meth}} = \sqrt{\frac{1}{\mathcal{X}} \sum_{X=1}^{\mathcal{X}} (e_{\text{meth}}^X - \text{MAE}_{\text{meth}})^2} \quad (4.16)$$

where \mathcal{X} is the total number of excitation energies considered in the selected set. On the website, the statistics are displaying using a table and a box plot.

Data generation tools

We use several tools to generate the data. These tools are also available for the user (see scenario 2). The main tool is `datafileBuilder` used to generate data files from a L^AT_EX tabular. The tabular is associated with certain options and L^AT_EX `\newcommand` analyzed by the main script and the tabular environment is converted to a NumPy 2D array. Therefore, the options, *i.e.*, the L^AT_EX `\newcommand` to be applied and the 2D array that represents the tabular environment, are passed to the appropriate tabular parser module chosen using the `\formatName` option in the input file. The module declares its format name itself using the `@formatName` (see Figure 4.17) decorator. Each module is responsible to parse the tabular and return all the corresponding data files as object. Next, the main script serializes these objects into the corresponding files. These files can be used on the website by

```
@formatName("mySuperTabularFormat")
class mySuperFoTabularFormatHandler(formatHandlerBase):
    def readFromTable(self, table):
        texCommand=self.Commands
        texOptions=self.TexOps
        : : :
    return fileList
```

FIGURE 4.17: PYTHON format module skeleton

temporarily importing it or by making a pull request for new data. The modular aspect of this tool gives us enough flexibility to easily convert many types of \LaTeX tabular to a standard file format.

4.6.3 Usage

Manipulation and scenarios

First, the user can add his own absorption and fluorescence data if he/she wants to analyze a custom dataset. In the Dataset tab, the user can select the data of interest by selecting the sets, molecules, methods and basis. After that, the user can customize the excitations he/she wants to take into account.

One can anticipate that two different types of users may be interested by the QUEST website:

Scenario 1

The user wants to choose a method for a calculation or a series of calculations on a given system of a given type of systems. Of course, he/she looks for a compromise between accuracy and cost. In this case, he/she wishes to compare the accuracy of each method with a subset of excitation data corresponding to his/her target. Hence, the user can optimize the filters to match its target (molecule, molecule size, or excitation type). If he/she is particularly lucky, he/she can only select the target molecule when this molecule is available in QUEST dataset.

Scenario 2

The user has created a new very promising method and wants to assess its accuracy with respect to other methods of the same family or highly accurate reference data (or both) provided by the QUEST dataset. First of all, the user must create an input file for PYTHON tools (see section 4.6.2) formatting the calculated results as \LaTeX tabular. After generating the data, the same PYTHON tools are employed to import them in the QUEST database. A dedicated button is provided on the website for such purposes. The new data are therefore used in the same way as the other data to generate statistics and for comparison purposes.

4.7 Concluding remarks

In this part, we have presented and extended the QUEST database of highly-accurate excitation energies for molecular systems [30, 32, 36, 39, 221] that we started building in 2018 and that is now composed by more than 500 vertical excitations, many of which can be reasonably considered as within 1 kcal mol⁻¹ (or less) of the FCI limit for the considered CC3/*aug-cc-pVTZ* geometry and basis set (*aug-cc-pVTZ*). In particular, we have detailed the specificities of our protocol by providing computational details regarding geometries, basis sets, as well as reference and benchmarked computational methods. The content of our five QUEST subsets has been presented in detail, and for each of them, we have provided the number of reference excitation energies, the nature and size of the molecules, the list of benchmarked methods, as well as other useful specificities. Importantly, we have proposed a new statistical method that produces much safer estimates of the extrapolation

error in SCI calculations. This new method based on Gaussian random variables has been tested by computing additional FCI values for five- and six-membered rings. After having discussed the generation of our TBEs, we have reported a comprehensive benchmark for a significant number of methods on the entire QUEST set with, in addition, a specific analysis for each type of excited states. Finally, the main features of the website specifically designed to gather the entire data generated during these past few years have been presented and discussed.

Paraphrasing Thiel's conclusions [196], we hope that not only the QUEST database will be used for further benchmarking and testing, but that other research groups will also improve it, providing not only corrections (inevitable in such a large data set), but more importantly extensions with both improved estimates for some compounds and states, or new molecules.

Regarding future improvements and extensions, we would like to mention that although our present goal is to produce chemically accurate vertical excitation energies, we are currently devoting great efforts to obtain highly-accurate excited-state properties [392, 393] such as dipoles and oscillator strengths for molecules of small and medium sizes [350, 394], so as to complete previous efforts aiming at determining accurate excited-state geometries [292, 395]. Reference ground-state properties (such as correlation energies and atomization energies) are also being currently produced [249, 290]. Besides this, because computing 500 (or so) excitation energies can be a costly exercise even with cheap computational methods, we are planning on developing a "diet set" (*i.e.*, a much smaller set of excitation energies which can reproduce key results of the full QUEST database, including ranking of approximations) following the philosophy of the "diet GMTKN55" set proposed recently by Gould [396]. We hope to report on this in the near future. A new article was recently published by some of my coworkers in order to complete the QUEST project with intermolecular charge-transfer (CT) excitations where a significant chunk of the charge density spatially migrates upon electronic excitation [397]. This type of excitations are ubiquitous in many technologies like OLEDs or photovoltaics. CT transitions are characterized by a large change in dipole moment as well as a small overlap between the starting and final molecular orbitals (MOs), or electron densities, involved in the transitions. For this set which includes much larger molecule, one cannot afford SCI calculations. Hence, the TBEs were calculated by correcting EOM-CCSDT/cc-pVDZ values by the difference between CC3/cc-pVTZ and CC3/cc-pVDZ and extended to *aug-cc-pVTZ* and *aug-cc-pVQZ* by applying a basis set correction. Globally, for the lowest CT characters, as one might expect, we reach a similar accuracy as the previous sets, except for transitions with a more pronounced CT nature.

One other aspect that we could improve in the future concerns the the data generation tools. Indeed, the present tools depend on several PYTHON packages like TEXSOUP, NUMPY, CROSS-REF_COMMONS, or GITPYTHON. Therefore, in order to prevent the user from manually installing dependencies, it might be useful to dissociate the tools from the website and gather them in a PYTHON package. Besides, it would also be interesting to create a real plug-in system to allow anyone to easily create table formats for DATAFILEBUILDER.

Chapter 5

Quantum Package demo in a web browser

5.1 Introduction

QUANTUM PACKAGE [35] is an open source quantum chemistry program developed mainly in our group, at the Laboratoire de Chimie Théorique (LCT) in Sorbonne University and by Anouar Benali and his coworkers in Argonne National Lab [291]. It is also used by several groups and researchers in the world like Hugh Burton from Oxford [398], Joshua Hollett from Winnipeg in Canada [399], Eric Neuscammen from Berkeley [400] or Claudia Filippi from the University of Twente in the Netherlands [289].

The main idea behind the creation of QUANTUM PACKAGE is to provide a platform for an easy implementation and experimentation of new *ab initio* methods, while making parallel computation as simple and efficient as possible. Following this philosophy, it contains a standalone easy-to-use library for developers (plug-in system), its own shell QUANTUM PACKAGE SHELL (QPSH) and the IRPF90 fortran preprocessor. However, even if QUANTUM PACKAGE is easy to use, it takes a while to install and this can limit its democratisation in our community. To allow quantum chemists to test QUANTUM PACKAGE quickly without the burden of its installation, we decided to create a demo which can be directly accessible via a web browser. This allows anyone to run QUANTUM PACKAGE simply from a web page without any configuration or installation effort. The purpose of this chapter is to explain how we have integrated the demonstration of QUANTUM PACKAGE in its website.

The specifications were

- An interactive terminal with all standard UNIX and QUANTUM PACKAGE commands;
- A limited number of users;
- A non-persistent and isolated environment without privileges and identical for each user;
- A fixed time limit;

To enforce these specifications, we have used the two tools described below.

5.2 Shellinabox

The first challenge of this work was to create a UNIX compatible terminal emulator in a web page connected to a remote machine. We found a tool called SHELLINABOX [401] that does exactly what we wanted. SHELLINABOX is a web server that creates a terminal web application and exports all commands to the host. This tool is fully configurable and one can easily choose the command to run at startup. On the client side, this tool requires only a web browser compatible with JAVASCRIPT and CSS.

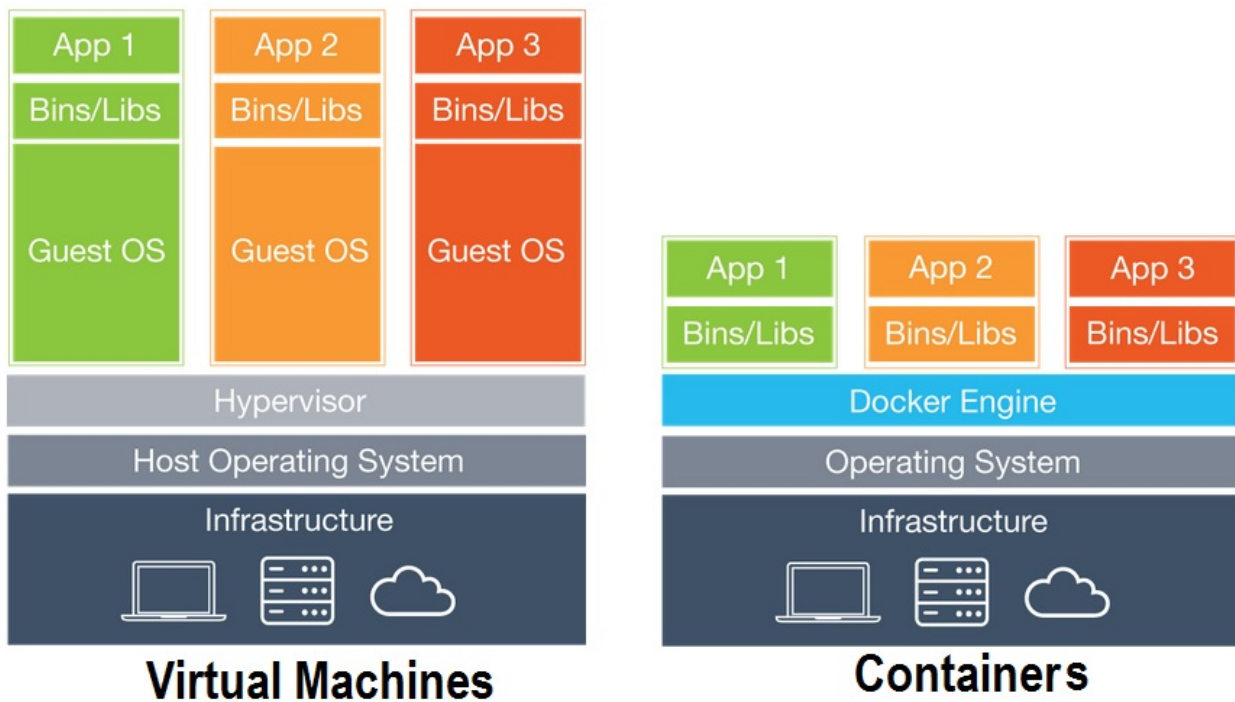


FIGURE 5.1: DOCKER container compared to virtualization technology [403]

5.3 Docker

DOCKER [402] is a well-known software in the community of system administrators. This software looks like a conventional virtualization software but it is not really the case. This is a containerization (also called operating-system-level virtualization) software. The host shares its kernel with containers and this drastically improves performance compared to conventional virtualization technologies because there is no guest operating system and not hypervisor (see Figure 5.1) and no virtualized hardware. In addition, this containerization technology is more advanced with respect to isolation than a simple `chroot`. To create DOCKER container DOCKER uses a library called LIBCONTAINER now wrapped in RUNC since this tool is managed by the Open Container Initiative [404]. This library is responsible for the container lifecycle and uses LINUX kernel components to create containers. In the coming sections, we will discuss the lifecycle of the container and the LINUX kernel components used by LIBCONTAINER for the containerization. Because it is LINUX Specific features, this description is not applicable to MACOS and WINDOWS (to the best of our knowledge DOCKER supports natively WINDOWS containers and the WINDOWS kernel since WINDOWS 10 version 1607 but nevertheless, we chose to talk about LINUX containers and LINUX kernel only because QUANTUM PACKAGE (and, of course, the QUANTUM PACKAGE container) is only LINUX compatible). In the case of a LINUX container on WINDOWS or MACOS the DOCKER engine is running on a LINUX kernel inside a lightweight virtual machine (usually shortened to VM) as backend [or WINDOWS SUBSYSTEM FOR LINUX 2 (WSL 2) on WINDOWS since DOCKER DESKTOP 2.3].

5.3.1 Container lifecycle

By definition, a container is designed to be ephemeral. It can be in different states (see Table 5.1). When a container is created, its content is based on a specific prebuild image and when a container is deleted, all its content is deleted.

TABLE 5.1: Different states of the DOCKER container [405]

Name	Description
Created	The container is created but no started yet
Restarting	The container is restarting
Running	The container is running
Removing	The container is removing
Paused	The container is in pause
Exited	The process of the container is terminated or the container is stopped
Dead	The container is broken

TABLE 5.2: List of LINUX namespaces types (the following list is indicative and depends on the kernel version) [406]

Name	Description
Cgroup	Cgroup root directory
IPC	System V IPC, POSIX message queues
Network	Network devices, stacks, ports, etc.
Mount	Mount points
PID	Process IDs
User	User and group IDs
UTS	Hostname and NIS domain name

5.3.2 Linux kernel component used by libcontainer for containerization

The container filesystem and union mounting

The filesystem of a container is constructed by successive layers using a type of filesystem that supports union mounting (`overlay2` for a modern LINUX system), a mount type used to merge multiple filesystems into a virtual filesystem. When we build a DOCKER image, we base this image on another image downloaded from the internet. We perform operations on the image and these operations are saved in another layer on top of the base image. In addition, when we use this image, all changes are combined with a union mounting to create the filesystem. The full container filesystem is based on an image built as described above. A union mounting is made between all parts of the image as read-only and a read-write layer is added for the modifications inside the container [405]. With this union mounting technology, the size of the image can be very light, the build time can be very short and the modifications inside a container are independent of the other containers even if they are based on the same image. In addition, the modifications made in this container are not persistent and respects the container's lifecycle.

Resources handling with cgroups

In the LINUX kernel, there is a feature called `cgroup` that limits some processes to a certain amount of resources. DOCKER is suitable for resource management because it uses `cgroups`. One can therefore choose the amount of memory, the number of CPUs and so on for each container.

Host separation with the namespaces

The LINUX namespaces provide to DOCKER the ability to create separate environments. Indeed the process environment depends on the namespaces associated with it. There are different types of namespaces (see Table 5.2).

For each process, the kernel associates a namespace of each type. One can easily watch the namespace IDs of a process using the `Proc` filesystem. (see Figure 5.2).

```
# For example, if we want to show the namespace IDs
# for the process with the PID 42
# One can type the following command
ls -l /proc/42/ns
```

FIGURE 5.2: List namespaces for a pid

```
<iframe id="shellframe" src="http://host:port"
width="100%" height="500"
frameBorder="0" scrolling="no"/>
```

FIGURE 5.3: HTML code to integrate SHELLINABOX into another website

5.4 In practice

In practice, for better security, we installed DOCKER and SHELLINABOX in a VM. We use SHELLINABOX by integrating it to the QUANTUM PACKAGE official website on the following page <https://quantumpackage.github.io/qp2/page/try> thanks to an HTML inline frame (iframe) (see Figure 5.3).

SHELLINABOX is configured to start a SHELL script in the DOCKER host. We wrote this script to handle the lifecycle of the demo container. We have built the image that we use to create demo containers on the official DOCKER image of the well known LINUX distribution UBUNTU in its long-term support (LTS) version (currently 20.04) in which we have added useful software like HTOP, VIM and TMUX. We used a complete and popular [407] distribution as a base image to ensure that the user has a complete LINUX environment with all the commands he/she can expect to have. Of course, we have also added a new standard user. Then at the root of this user's directory, we added QUANTUM PACKAGE alongside some examples. When the user opens the demo web page containing the SHELLINABOX terminal, it runs the script that creates and start a QUANTUM PACKAGE container in interactive mode (see Figure 5.4). The containers are configured to start the QUANTUM PACKAGE SHELL as a standard user. The script handles the closing of the terminal or the end of the allocated time to kill the container; it is configured to automatically self-destruct on exit and any changes made by the user are therefore destroyed.

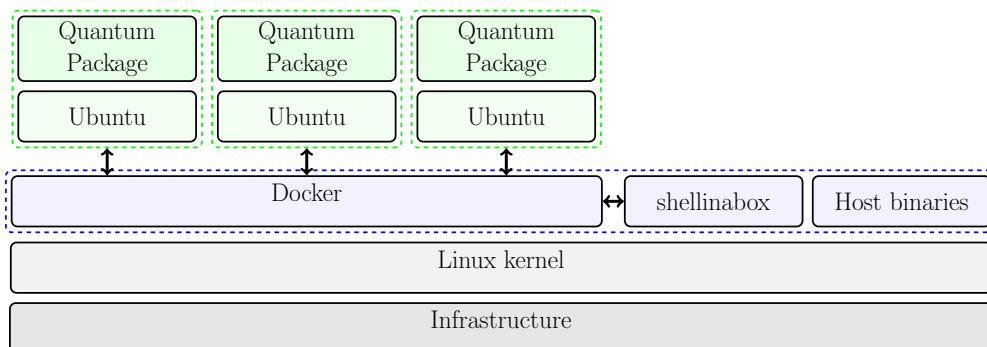
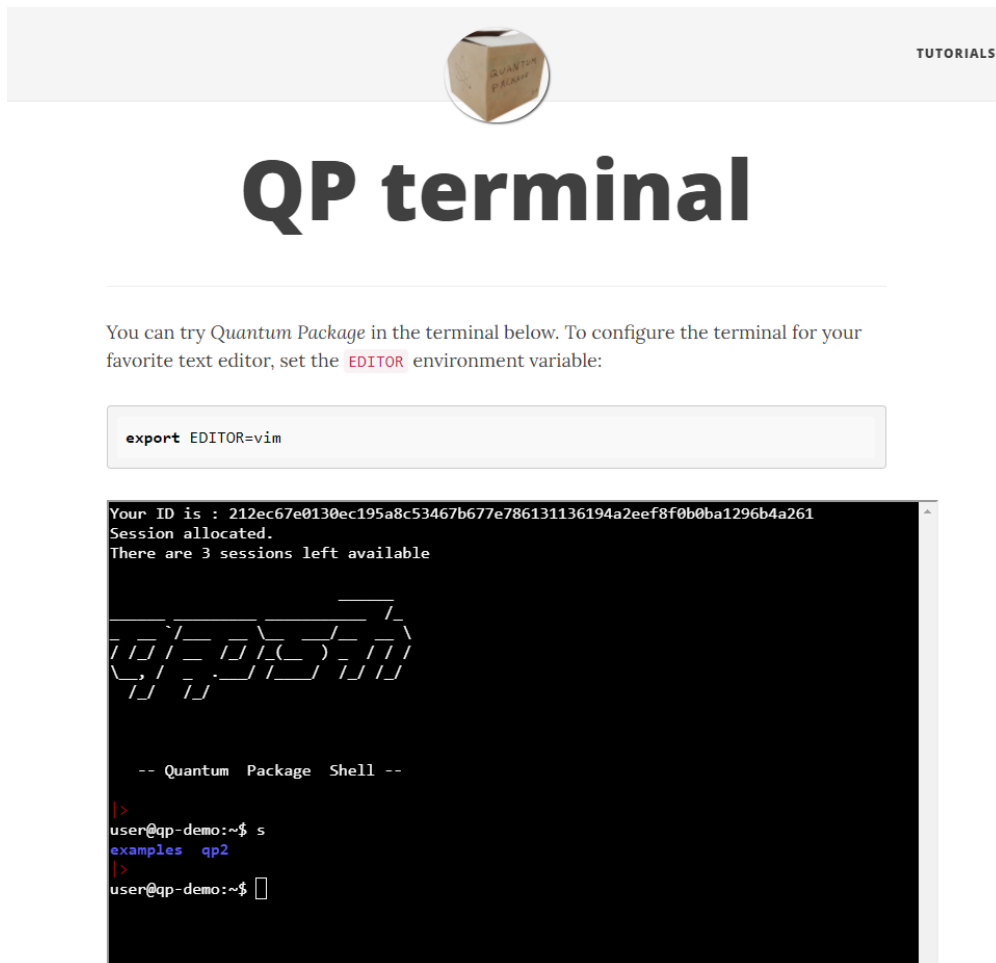


FIGURE 5.4: Representation of the QUANTUM PACKAGE demo architecture. In dashed line, the different environments isolated from each other. In blue the OS level, in green the DOCKER containers level. The arrows represent the communications between the demo components.



The screenshot shows a web page titled "QP terminal" under a "TUTORIALS" header. The page contains a code block with the command `export EDITOR=vim`. Below this is a terminal window with the following output:

```
Your ID is : 212ec67e0130ec195a8c53467b677e786131136194a2eef8f0b0ba1296b4a261
Session allocated.
There are 3 sessions left available

  Quantum Package

-- Quantum Package Shell --

|>
user@qp-demo:~$ s
examples qp2
|>
user@qp-demo:~$
```

FIGURE 5.5: Screenshot of the terminal giving access to the QUANTUM PACKAGE demo directly integrated on the QUANTUM PACKAGE website

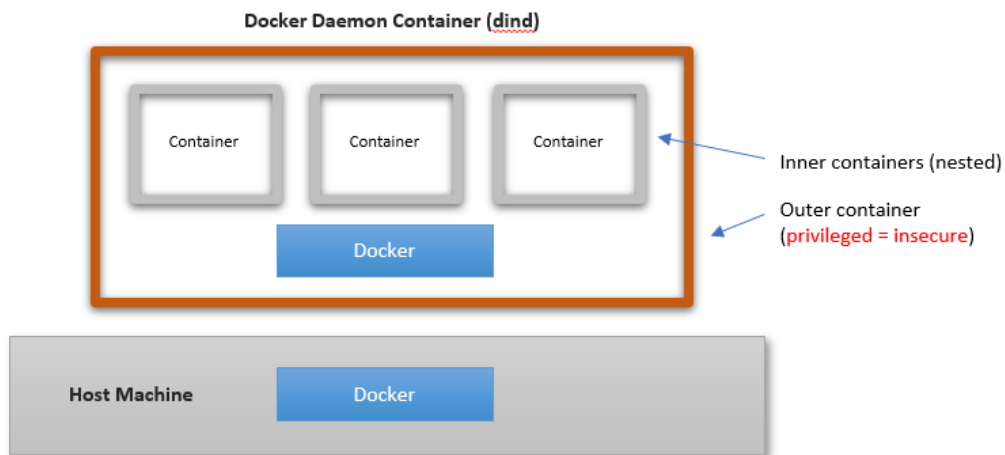


FIGURE 5.6: DOCKER in DOCKER using privileged container

5.5 Conclusion and perspectives

Following the QUANTUM PACKAGE philosophy and in order to make it easy to test, we designed a demo environment accessible via a terminal emulator on a web page. The demo is built using two software: (i) SHELLINABOX is the tool that provides the terminal emulator web elements and the connection to the remote machine; (ii) DOCKER provides the isolated and ephemeral environment for the demo. The demo's lifecycle is handled by a BASH script started by SHELLINABOX on the remote machine. One can test QUANTUM PACKAGE at <https://quantumpackage.github.io/qp2/page/try> and one can download the ready to run DOCKER image used for the demo on the Docker Hub at https://hub.docker.com/r/mveril/qp_demo.

Normally to use DOCKER in a DOCKER container, one needs to use a privileged container which is somehow insecure (see Figure 5.6). This is why we chose a VM to host SHELLINABOX and DOCKER even if it affects performance. Recently, we discovered [408] the ability to create a system-level container in order to use DOCKER in a DOCKER container without security issue. This can be a solution for transferring the host containing DOCKER and SHELLINABOX from a VM to a DOCKER container and thus improve performance without creating a security issue for the host machine.

We also recently heard about a containerization software called SINGULARITY. The main interest of SINGULARITY compared to DOCKER is its features that allow it to be adapted to the HPC world. We plan on investigating the compatibility of QUANTUM PACKAGE with the latter in the future.

Chapter 6

Accurate full configuration interaction correlation energy estimates for five- and six-membered rings

This chapter is based on

Y. Damour et al., "Accurate full configuration interaction correlation energy estimates for five- and six-membered rings", *The Journal of Chemical Physics* **155**, 134104 (2021)

In this chapter, my main contribution was to perform the CIPSI calculations using natural orbitals for the five- and six-membered cyclic molecules considered in this study.

6.1 Introduction

Electronic structure theory relies heavily on approximations [1–3]. Loosely speaking, to make any method practical, three main approximations are typically enforced. The first fundamental approximation, known as the Born-Oppenheimer (or clamped-nuclei) approximation, consists in assuming that the motion of nuclei and electrons are decoupled [410]. The nuclei coordinates can then be treated as parameters in the electronic Hamiltonian. The second central approximation which makes calculations computationally achievable is the basis set approximation where one introduces a set of pre-defined basis functions to represent the many-electron wave function of the system. In most molecular calculations, a set of one-electron, atom-centered Gaussian basis functions are introduced to expand the so-called one-electron molecular orbitals which are then used to build the many-electron Slater determinant(s). The third and most relevant approximation in the present context is the ansatz (or form) of the electronic wave function Ψ . For example, in configuration interaction (CI) methods, the wave function is expanded as a linear combination of Slater determinants, while in (single-reference) coupled-cluster (CC) theory [48–51, 151, 411], a reference Slater determinant Ψ_0 [usually taken as the Hartree-Fock (HF) wave function] is multiplied by a wave operator defined as the exponentiated excitation operator $\hat{T} = \sum_{k=1}^N \hat{T}_k$ (where N is the number of electrons and \hat{T}_k is k th-degree excitation operator).

The truncation of \hat{T} allows to define a hierarchy of non-variational and size-extensive methods with increasing levels of accuracy: CC with singles and doubles (CCSD) [48, 52], CC with singles, doubles, and triples (CCSDT) [53, 225], CC with singles, doubles, triples, and quadruples (CCSDTQ) [412, 413], with corresponding formal computational scalings of $\mathcal{O}(N^6)$, $\mathcal{O}(N^8)$, and $\mathcal{O}(N^{10})$, respectively (where N denotes the number of orbitals). Parallel to the "complete" CC series presented above, an alternative family of approximate iterative CC models has been developed by the Aarhus group in the context of CC response theory [169] where one skips the most expensive terms and avoids the storage of the higher-excitation amplitudes: CC2 [25], CC3 [26, 203], and CC4 [414, 415]. These iterative methods scale as $\mathcal{O}(N^5)$, $\mathcal{O}(N^7)$, and $\mathcal{O}(N^9)$, respectively, and can be seen as cheaper approximations of CCSD, CCSDT, and CCSDTQ. Coupled-cluster methods have been particularly successful at computing accurately ground- and excited-state properties for small- and medium-sized molecules. [228, 350, 394, 416–419]

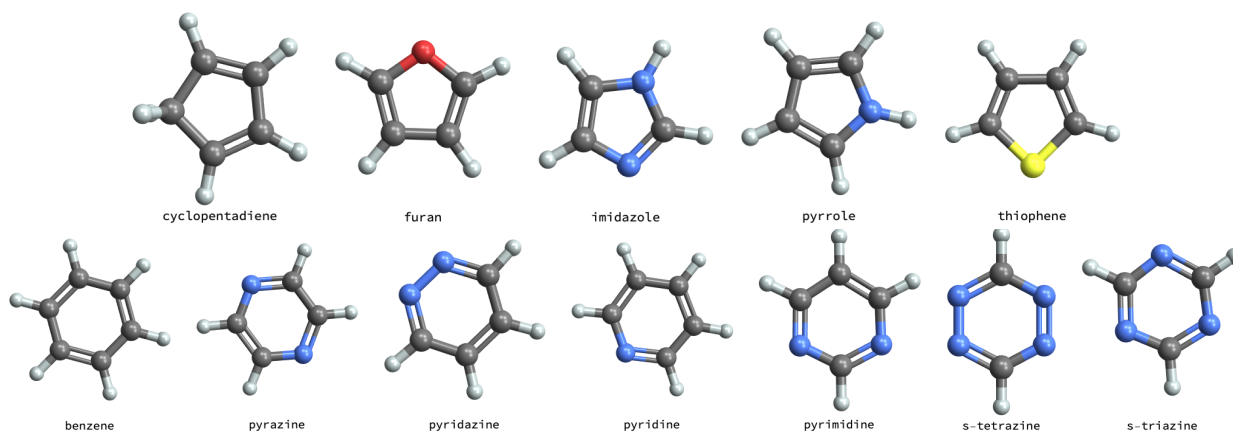


FIGURE 6.1: Five-membered rings (top) and six-membered rings (bottom) considered in this study.

A similar systematic truncation strategy can be applied to CI methods leading to the well-established family of methods known as CISD, CISDT, CISDTQ, ... where one systematically increases the maximum excitation degree of the determinants taken into account. Except for full CI (FCI) where all determinants from the Hilbert space (*i.e.*, with excitation degree up to N) are considered, truncated CI methods are variational but lack size-consistency. The non-variationality of truncated CC methods being, in practice, less of an issue than the size-inconsistency of the truncated CI methods, the formers have naturally overshadowed the latters in the electronic structure landscape. However, a different strategy recently came back in the limelight in the context of CI methods [34, 251–254, 280–284, 420, 421]. Indeed, selected CI (SCI) methods [231, 256, 258, 261, 262, 266–268, 272, 274–276], where one iteratively selects the important determinants from the FCI space (usually) based on a perturbative criterion, has been recently shown to be highly successful in order to produce reference energies for both ground and excited states in small- and medium-sized molecules [30, 32, 35–38, 149, 190, 221, 240–242, 249, 257, 260, 264, 265, 286, 397, 422] thanks to efficient deterministic, stochastic, or hybrid algorithms well suited for massive parallelization. We refer the interested reader to Refs. [39, 250] for recent reviews. SCI methods are based on a well-known fact: amongst the very large number of determinants contained in the FCI space, only a tiny fraction of them significantly contributes to the energy (see, for example, Ref. [423]). Accordingly, the SCI+PT2 family of methods performs a sparse exploration of the FCI space by selecting iteratively only the most energetically relevant determinants of the variational space and supplementing it with a second-order perturbative correction (PT2) [34, 35, 259, 263, 266]. Although the formal scaling of such algorithms remains exponential, the prefactor is greatly reduced which explains their current attractiveness in the electronic structure community thanks to their much wider applicability than their standard FCI parent. Note that, very recently, several groups [424–426] have coupled CC and SCI methods via the externally-corrected CC methodology [427], showing promising performances for weakly and strongly correlated systems.

A rather different strategy in order to reach the holy grail FCI limit is to resort to Møller-Plesset (MP) perturbation theory [428], whose popularity originates from its black-box nature, size-extensivity, and relatively low computational requirement, making it easily applied to a broad range of molecular systems. Again, at least in theory, one can obtain the exact energy of the system by ramping up the degree of the perturbative series [429]. The second-order Møller-Plesset (MP2) method [428] [which scales as $\mathcal{O}(N^5)$] has been broadly adopted in quantum chemistry for several decades, and is now included in the increasingly popular double-hybrid functionals [430] alongside exact exchange. Its higher-order variants [MP3 [431], MP4 [432], MP5 [433], and MP6 [434, 435] which scale as $\mathcal{O}(N^6)$, $\mathcal{O}(N^7)$, $\mathcal{O}(N^8)$, and $\mathcal{O}(N^9)$ respectively] have been investigated much more scarcely. However, it is now widely recognized that the series of MP approximations might show erratic, slowly convergent, or divergent behavior that limits its applicability and systematic improvability [429, 436–445]. Again, MP perturbation theory and CC methods can be coupled.

The most iconic example of such coupling, namely the CCSD(T) method [55], includes iteratively the single and double excitations and perturbatively (from MP4 and partially MP5) the triple excitations, leading to the so-called “gold-standard” of quantum chemistry for weakly correlated systems thanks to its excellent accuracy/cost ratio.

Motivated by the recent blind test of Eriksen *et al.* [422] reporting the performance of a large panel of emerging electronic structure methods [the many-body expansion FCI (MBE-FCI) [243–246], adaptive sampling CI (ASCI) [268, 269, 446], iterative CI (iCI) [272, 277–279], semistochastic heat-bath CI (SHCI) [37, 262, 263], the full coupled-cluster reduction (FCCR) [247, 248], density-matrix renormalization group (DMRG) [447–449], adaptive-shift FCI quantum Monte Carlo (AS-FCIQMC) [231, 233, 237], and cluster-analysis-driven FCIQMC (CAD-FCIQMC) [238, 239]] on the non-relativistic frozen-core correlation energy of the benzene molecule in the standard correlation-consistent double- ζ Dunning basis set (cc-pVDZ), some of us have recently investigated the performance of the SCI method known as *Configuration Interaction using a Perturbative Selection made Iteratively* (CIPSI) [34, 35, 256, 258, 266]. on the very same system [249] [see also Ref. [450] for a study of the performance of phaseless auxiliary-field quantum Monte Carlo (ph-AFQMC) [451]]. In the continuity of this recent work, we report here a large extension by accurately estimating the (frozen-core) FCI/cc-pVDZ correlation energy of twelve cyclic molecules (cyclopentadiene, furan, imidazole, pyrrole, thiophene, benzene, pyrazine, pyridazine, pyridine, pyrimidine, s-tetrazine, and s-triazine) with the help of CIPSI employing energetically-optimized orbitals at the same level of theory [241, 452]. These systems are depicted in Figure 6.1. This set of molecular systems corresponds to Hilbert spaces with sizes ranging from 10^{29} to 10^{36} . In addition to CIPSI, the performance and convergence properties of several series of methods are investigated. In particular, we study i) the MP perturbation series up to fifth-order (MP2, MP3, MP4, and MP5), ii) the CC2, CC3, and CC4 approximate series, and iii) the “complete” CC series up to quadruples (*i.e.*, CCSD, CCSDT, and CCSDTQ). The performance of the ground-state gold standard CCSD(T) as well as the completely renormalized (CR) CC model, CR-CC(2,3) [151, 453–456], are also investigated.

The present chapter is organized as follows. In section 6.2, we provide theoretical details about the CIPSI algorithm and the orbital optimization procedure employed here. Section 6.3 deals with computational details concerning geometries, basis sets, and methods. In section 6.4, we report our reference FCI correlation energies for the five-membered and six-membered cyclic molecules obtained thanks to extrapolated orbital-optimized CIPSI calculations (section 6.4.1). These reference correlation energies are then used to benchmark and study the convergence properties of various perturbative and CC methods (section 6.4.2). Finally, we draw our conclusions in section 6.5.

6.2 CIPSI with optimized orbitals

Here, we provide key details about the CIPSI method [34, 35] as well as the orbital optimization procedure which has been shown to be highly effective in the context of SHCI by Umrigar and coworkers [241, 422, 452]. Although we focus on the ground state, the present discussion can be easily extended to excited states [149, 287].

At the k th iteration, the total CIPSI energy $E_{\text{CIPSI}}^{(k)}$ is defined as the sum of the variational energy

$$E_{\text{var}}^{(k)} = \frac{\langle \Psi_{\text{var}}^{(k)} | \hat{H} | \Psi_{\text{var}}^{(k)} \rangle}{\langle \Psi_{\text{var}}^{(k)} | \Psi_{\text{var}}^{(k)} \rangle} \quad (6.1)$$

and a second-order perturbative energy correction

$$E_{\text{PT2}}^{(k)} = \sum_{\alpha \in \mathcal{A}_k} e_{\alpha}^{(k)} = \sum_{\alpha \in \mathcal{A}_k} \frac{|\langle \Psi_{\text{var}}^{(k)} | \hat{H} | \alpha \rangle|^2}{E_{\text{var}}^{(k)} - \langle \alpha | \hat{H} | \alpha \rangle}, \quad (6.2)$$

where \hat{H} is the (non-relativistic) electronic Hamiltonian,

$$\Psi_{\text{var}}^{(k)} = \sum_{I \in \mathcal{I}_k} c_I^{(k)} |I\rangle \quad (6.3)$$

is the variational wave function, \mathcal{I}_k is the set of internal determinants $|I\rangle$ and \mathcal{A}_k is the set of external determinants (or perturbers) $|\alpha\rangle$ which do not belong to the variational space at the k th iteration but are linked to it via a nonzero matrix element, *i.e.*, $\langle \Psi_{\text{var}}^{(k)} | \hat{H} | \alpha \rangle \neq 0$. The sets \mathcal{I}_k and \mathcal{A}_k define, at the k th iteration, the internal and external spaces, respectively. In the selection step, the perturbers corresponding to the largest $|e_{\alpha}^{(k)}|$ values are then added to the variational space at the next iteration. In our implementation, the size of the variational space is roughly doubled at each iteration. Hereafter, we label these iterations over k which consist in enlarging the variational space as *macroiterations*. In practice, $E_{\text{var}}^{(k)}$ is the lowest eigenvalue of the $N_{\text{det}}^{(k)} \times N_{\text{det}}^{(k)}$ CI matrix with elements $\langle I | \hat{H} | J \rangle$ obtained via Davidson's algorithm [457]. The magnitude of $E_{\text{PT2}}^{(k)}$ provides, at iteration k , a qualitative idea of the distance to the FCI limit [266]. We then linearly extrapolate, using large variational wave functions, the CIPSI energy to $E_{\text{PT2}} = 0$ (which effectively corresponds to the FCI limit). Further details concerning the extrapolation procedure are provided below (see section 6.4).

Orbital optimization techniques at the SCI level are theoretically straightforward, but practically challenging. Some of the technology presented here has been borrowed from complete-active-space self-consistent-field (CASSCF) methods [458–462] but one of the strength of SCI methods is that one does not need to select an active space and to classify orbitals as active, inactive, and virtual orbitals. Here, we detail our orbital optimization procedure within the CIPSI algorithm and we assume that the variational wave function is normalized, *i.e.*, $\langle \Psi_{\text{var}} | \Psi_{\text{var}} \rangle = 1$.

As stated in section 6.1, E_{var} depends on both the CI coefficients $\{c_I\}_{1 \leq I \leq N_{\text{det}}}$ [see Eq. (6.3)] but also on the orbital rotation parameters $\{\kappa_{pq}\}_{1 \leq p, q \leq N}$. Motivated by cost saving arguments, we have chosen to optimize separately the CI and orbital coefficients by alternatively diagonalizing the CI matrix after each selection step and then rotating the orbitals until the variational energy, for a given number of determinants, is minimal. We refer the interested reader to the recent work of Yao and Umrigar for a detailed comparison of coupled, uncoupled, and partially-coupled optimizations within SCI methods [452]. Following the standard procedure [2], we conveniently rewrite the variational energy as

$$E_{\text{var}}(\mathbf{c}, \boldsymbol{\kappa}) = \langle \Psi_{\text{var}} | e^{\hat{\kappa}} \hat{H} e^{-\hat{\kappa}} | \Psi_{\text{var}} \rangle, \quad (6.4)$$

where \mathbf{c} gathers the CI coefficients, $\boldsymbol{\kappa}$ the orbital rotation parameters, and

$$\hat{\kappa} = \sum_{p < q} \sum_{\sigma} \kappa_{pq} \left(\hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma} - \hat{a}_{q\sigma}^{\dagger} \hat{a}_{p\sigma} \right) \quad (6.5)$$

is a real-valued one-electron antisymmetric operator, which creates an orthogonal transformation of the orbital coefficients when exponentiated, $\hat{a}_{p\sigma}$ ($\hat{a}_{p\sigma}^{\dagger}$) being the second quantization annihilation (creation) operator which annihilates (creates) a spin- σ electron in the real-valued spatial orbital $\phi_p(\mathbf{r})$ [2].

Applying the Newton-Raphson method by Taylor-expanding the variational energy to second order around $\boldsymbol{\kappa} = \mathbf{0}$, *i.e.*,

$$E_{\text{var}}(\mathbf{c}, \boldsymbol{\kappa}) \approx E_{\text{var}}(\mathbf{c}, \mathbf{0}) + \mathbf{g} \cdot \boldsymbol{\kappa} + \frac{1}{2} \boldsymbol{\kappa}^{\dagger} \cdot \mathbf{H} \cdot \boldsymbol{\kappa}, \quad (6.6)$$

one can iteratively minimize the variational energy with respect to the parameters κ_{pq} by setting

$$\boldsymbol{\kappa} = -\mathbf{H}^{-1} \cdot \mathbf{g}, \quad (6.7)$$

where \mathbf{g} and \mathbf{H} are the orbital gradient and Hessian matrices, respectively, both evaluated at $\boldsymbol{\kappa} = \mathbf{0}$. Their elements are explicitly given by the following expressions [463, 464]:

$$\begin{aligned} g_{pq} &= \left. \frac{\partial E_{\text{var}}(\mathbf{c}, \boldsymbol{\kappa})}{\partial \kappa_{pq}} \right|_{\boldsymbol{\kappa}=\mathbf{0}} \\ &= \sum_{\sigma} \langle \Psi_{\text{var}} | [\hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma} - \hat{a}_{q\sigma}^{\dagger} \hat{a}_{p\sigma}, \hat{H}] | \Psi_{\text{var}} \rangle \\ &= \mathcal{P}_{pq} \left[\sum_r (h_p^r \gamma_r^q - h_r^q \gamma_p^r) + \sum_{rst} (v_{pt}^{rs} \Gamma_{rs}^{qt} - v_{rs}^{qt} \Gamma_{pt}^{rs}) \right], \end{aligned} \quad (6.8)$$

and

$$\begin{aligned} H_{pq,rs} &= \left. \frac{\partial^2 E_{\text{var}}(\mathbf{c}, \boldsymbol{\kappa})}{\partial \kappa_{pq} \partial \kappa_{rs}} \right|_{\boldsymbol{\kappa}=\mathbf{0}} \\ &= \mathcal{P}_{pq} \mathcal{P}_{rs} \left\{ \frac{1}{2} \sum_{\sigma\sigma'} \langle \Psi_{\text{var}} | [\hat{a}_{r\sigma'}^{\dagger} \hat{a}_{s\sigma'}, [\hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma}, \hat{H}]] | \Psi_{\text{var}} \rangle \right. \\ &\quad \left. + \frac{1}{2} \sum_{\sigma\sigma'} \langle \Psi_{\text{var}} | [\hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma}, [\hat{a}_{r\sigma'}^{\dagger} \hat{a}_{s\sigma'}, \hat{H}]] | \Psi_{\text{var}} \rangle \right\} \\ &= \mathcal{P}_{pq} \mathcal{P}_{rs} \left\{ \frac{1}{2} \sum_u [\delta_{qr} (h_p^u \gamma_u^s + h_u^s \gamma_p^u) + \delta_{ps} (h_r^u \gamma_u^q + h_u^q \gamma_r^u)] \right. \\ &\quad - (h_p^s \gamma_r^q + h_r^q \gamma_p^s) \\ &\quad + \frac{1}{2} \sum_{tuv} \delta_{qr} (v_{pt}^{uv} \Gamma_{uv}^{st} + v_{uv}^{st} \Gamma_{pt}^{uv}) \\ &\quad + \frac{1}{2} \sum_{tuv} \delta_{ps} (v_{uv}^{qt} \Gamma_{rt}^{uv} + v_{rt}^{uv} \Gamma_{uv}^{qt}) \\ &\quad + \sum_{uv} (v_{pr}^{uv} \Gamma_{uv}^{qs} + v_{uv}^{qs} \Gamma_{pr}^{uv}) \\ &\quad \left. - \sum_{tu} (v_{pu}^{st} \Gamma_{rt}^{qu} + v_{pu}^{tr} \Gamma_{tr}^{qu} + v_{rt}^{qu} \Gamma_{pu}^{st} + v_{tr}^{qu} \Gamma_{pu}^{st}) \right\}, \end{aligned} \quad (6.9)$$

where δ_{pq} is the Kronecker delta, $\mathcal{P}_{pq} = 1 - (p \leftrightarrow q)$ is a permutation operator,

$$\gamma_p^q = \sum_{\sigma} \langle \Psi_{\text{var}} | \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma} | \Psi_{\text{var}} \rangle, \quad (6.10a)$$

$$\Gamma_{pq}^{rs} = \sum_{\sigma\sigma'} \langle \Psi_{\text{var}} | \hat{a}_{p\sigma}^{\dagger} \hat{a}_{r\sigma'}^{\dagger} \hat{a}_{s\sigma'} \hat{a}_{q\sigma} | \Psi_{\text{var}} \rangle \quad (6.10b)$$

are the elements of the one- and two-electron density matrices, and

$$h_p^q = \int \phi_p(\mathbf{r}) \hat{h}(\mathbf{r}) \phi_q(\mathbf{r}) d\mathbf{r}, \quad (6.11a)$$

$$v_{pq}^{rs} = \iint \phi_p(\mathbf{r}_1) \phi_q(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_r(\mathbf{r}_1) \phi_s(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (6.11b)$$

are the one- and two-electron integrals, respectively.

Because the size of the CI space is much larger than the orbital space, for each macroiteration, we perform multiple *microiterations* which consist in iteratively minimizing the variational energy

(6.4) with respect to the $N(N - 1)/2$ independent orbital rotation parameters for a fixed set of determinants. After each microiteration (*i.e.*, orbital rotation), the one- and two-electron integrals [see Eqs. (6.11a) and (6.11b)] have to be updated. Moreover, the CI matrix must be re-diagonalized and new one- and two-electron density matrices [see Eqs. (6.10a) and (6.10b)] have to be computed. Microiterations are stopped when a stationary point is found, *i.e.*, $\|g\|_\infty < \tau$, where τ is a user-defined threshold which has been set to 10^{-4} a.u. in the present study, and a new CIPSI selection step is performed. Note that a tight convergence is not critical here as a new set of microiterations is performed at each macroiteration and a new production CIPSI run is performed from scratch using the final set of orbitals (see section 6.3). This procedure might sound computationally expensive but one has to realize that the microiterations are usually performed only for relatively compact variational spaces. Therefore, the computational bottleneck of this approach remains the diagonalization of the CI matrix for very large variational spaces.

To enhance the convergence of the microiteration process, we employ an adaptation of the Newton-Raphson method known as “trust region” [465]. This popular variant defines a region where the quadratic approximation (6.6) is an adequate representation of the objective energy function (6.4) and it evolves during the optimization process in order to preserve the adequacy via a constraint on the step size preventing it from overstepping, *i.e.*, $\|\kappa\| \leq \Delta$, where Δ is the trust radius. By introducing a Lagrange multiplier λ to control the trust-region size, one replaces Eq. (6.7) by $\kappa = -(\mathbf{H} + \lambda \mathbf{1})^{-1} \cdot \mathbf{g}$. The addition of the level shift $\lambda \geq 0$ removes the negative eigenvalues and ensures the positive definiteness of the Hessian matrix by reducing the step size. By choosing the right value of λ , $\|\kappa\|$ is constrained within a hypersphere of radius Δ and is able to evolve from the Newton direction at $\lambda = 0$ to the steepest descent direction as λ grows. The evolution of the trust radius during the optimization and the use of a condition to reject the step when the energy rises ensure the convergence of the algorithm. More details can be found in Ref. [465].

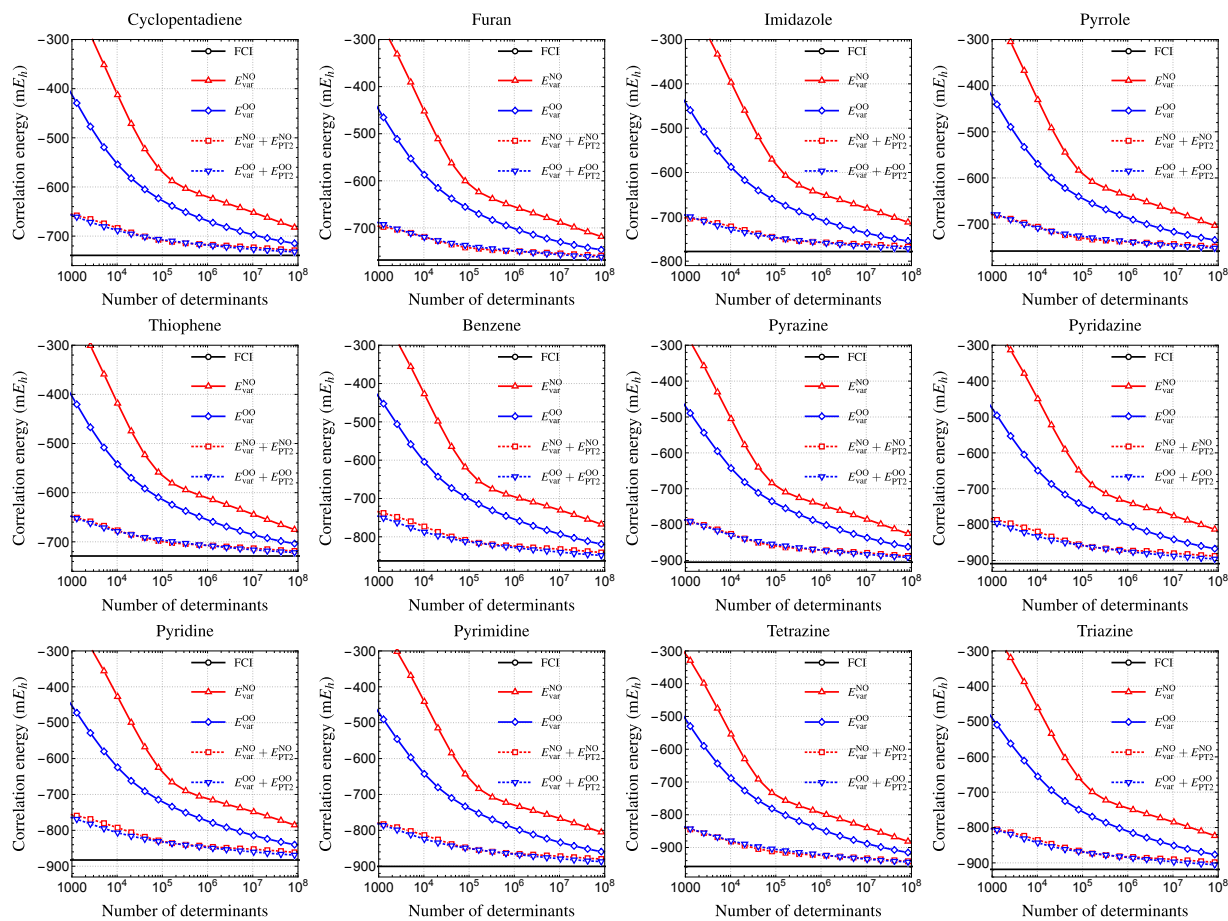


FIGURE 6.2: ΔE_{var} (solid) and $\Delta E_{\text{var}} + E_{\text{PT2}}$ (dashed) computed in the cc-pVDZ basis as functions of the number of determinants N_{det} in the variational space for the twelve cyclic molecules represented in Figure 6.1. Two sets of orbitals are considered: natural orbitals (NOs, in red) and optimized orbitals (OOs, in blue). The FCI estimate of the correlation energy is represented as a thick black line.

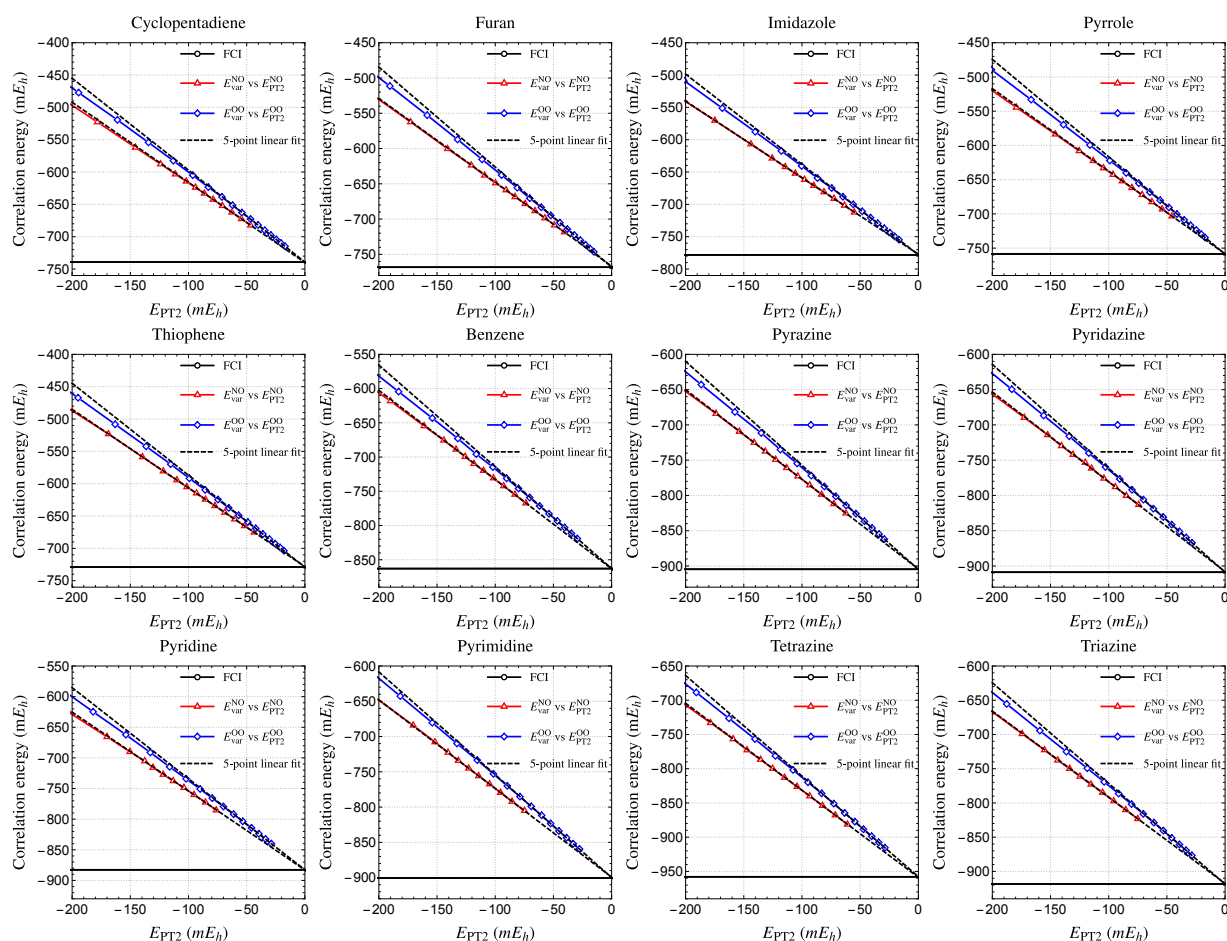


FIGURE 6.3: ΔE_{var} as a function of E_{PT2} computed in the cc-pVDZ basis for the twelve cyclic molecules represented in Figure 6.1. Two sets of orbitals are considered: natural orbitals (NOs, in red) and optimized orbitals (OOs, in blue). The five-point weighted linear fit using the five largest variational wave functions for each set is depicted as a dashed black line. The weights are taken as the inverse square of the perturbative corrections. The FCI estimate of the correlation energy is represented as a thick black line.

TABLE 6.1: Total energy E (in E_h) and correlation energy ΔE (in mE_h) for the frozen-core ground state of five-membered rings in the cc-pVDZ basis set. For the CIPSI estimates of the FCI correlation energy, the fitting error associated with the weighted five-point linear fit is reported in parenthesis.

Method	Cyclopentadiene		Furan		Imidazole		Pyrrole		Thiophene	
	E	ΔE	E	ΔE	E	ΔE	E	ΔE	E	ΔE
HF	-192.8083		-228.6433		-224.8354		-208.8286		-551.3210	
MP2	-193.4717	-663.4	-229.3508	-707.5	-225.5558	-720.4	-209.5243	-695.7	-551.9825	-661.5
MP3	-193.5094	-701.0	-229.3711	-727.8	-225.5732	-737.8	-209.5492	-720.6	-552.0104	-689.4
MP4	-193.5428	-734.5	-229.4099	-766.6	-225.6126	-777.2	-209.5851	-756.5	-552.0476	-726.6
MP5	-193.5418	-733.4	-229.4032	-759.9	-225.6061	-770.8	-209.5809	-752.3	-552.0426	-721.6
CC2	-193.4782	-669.9	-229.3605	-717.2	-225.5644	-729.0	-209.5311	-702.5	-551.9905	-669.5
CC3	-193.5449	-736.6	-229.4090	-765.7	-225.6115	-776.1	-209.5849	-756.3	-552.0473	-726.3
CC4	-193.5467	-738.4	-229.4102	-766.9	-225.6126	-777.2	-209.5862	-757.6	-552.0487	-727.7
CCSD	-193.5156	-707.2	-229.3783	-735.0	-225.5796	-744.2	-209.5543	-725.7	-552.0155	-694.5
CCSDT	-193.5446	-736.2	-229.4076	-764.3	-225.6099	-774.6	-209.5838	-755.2	-552.0461	-725.1
CCSDTQ	-193.5465	-738.2	-229.4100	-766.7	-225.6123	-776.9	-209.5860	-757.4	-552.0485	-727.5
CCSD(T)	-193.5439	-735.6	-229.4073	-764.0	-225.6099	-774.5	-209.5836	-754.9	-552.0458	-724.8
CR-CC(2,3)	-193.5439	-735.6	-229.4075	-764.2	-225.6098	-774.5	-209.5835	-754.9	-552.0459	-724.9
FCI		-739.2(1)		-768.2(1)		-778.2(1)		-758.5(1)		-728.9(3)

TABLE 6.2: Total energy E (in E_h) and correlation energy ΔE (in mE_h) for the frozen-core ground state of six-membered rings in the cc-pVDZ basis set. For the CIPSI estimates of the FCI correlation energy, the fitting error associated with the weighted five-point linear fit is reported in parenthesis.

Method	Benzene		Pyrazine		Pyridazine		Pyridine		Pyrimidine		s-Tetrazine		s-Triazine	
	E	ΔE	E	ΔE	E	ΔE	E	ΔE	E	ΔE	E	ΔE	E	ΔE
HF	-230.7222		-262.7030		-262.6699		-246.7152		-262.7137		-294.6157		-278.7173	
MP2	-231.5046	-782.3	-263.5376	-834.6	-263.5086	-838.7	-247.5227	-807.5	-263.5437	-830.1	-295.5117	-895.9	-279.5678	-850.5
MP3	-231.5386	-816.4	-263.5567	-853.7	-263.5271	-857.3	-247.5492	-834.0	-263.5633	-849.6	-295.5152	-899.5	-279.5809	-863.6
MP4	-231.5808	-858.5	-263.6059	-902.9	-263.5778	-907.9	-247.5951	-879.9	-263.6129	-899.3	-295.5743	-958.6	-279.6340	-916.7
MP5	-231.5760	-853.8	-263.5968	-893.8	-263.5681	-898.3	-247.5881	-872.9	-263.6036	-890.0	-295.5600	-944.3	-279.6228	-905.4
CC2	-231.5117	-789.4	-263.5475	-844.5	-263.5188	-848.9	-247.5315	-816.3	-263.5550	-841.3	-295.5247	-909.0	-279.5817	-864.4
CC3	-231.5814	-859.1	-263.6045	-901.5	-263.5761	-906.2	-247.5948	-879.6	-263.6120	-898.4	-295.5706	-954.9	-279.6329	-915.6
CC4	-231.5828	-860.6	-263.6056	-902.6	-263.5773	-907.5	-247.5960	-880.8	-263.6129	-899.3	-295.5716	-955.9	-279.6334	-916.1
CCSD	-231.5440	-821.8	-263.5640	-861.0	-263.5347	-864.9	-247.5559	-840.7	-263.5716	-858.0	-295.5248	-909.1	-279.5911	-873.8
CCSDT	-231.5802	-857.9	-263.6024	-899.4	-263.5739	-904.0	-247.5931	-877.9	-263.6097	-896.1	-295.5673	-951.6	-279.6300	-912.7
CCSDTQ	-231.5826	-860.4	-263.6053	-902.3	-263.5770	-907.1	-247.5960	-880.8	-263.6126	-899.0	-295.5712	-955.4	-279.6331	-915.8
CCSD(T)	-231.5798	-857.5	-263.6024	-899.4	-263.5740	-904.1	-247.5929	-877.7	-263.6099	-896.2	-295.5680	-952.2	-279.6305	-913.1
CR-CC(2,3)	-231.5792	-856.9	-263.6017	-898.7	-263.5732	-903.3	-247.5922	-877.1	-263.6091	-895.5	-295.5670	-951.3	-279.6298	-912.5
FCI		-862.9(3)		-904.6(4)		-908.8(1)		-882.7(4)		-900.5(2)		-957.9(4)		-918.4(3)

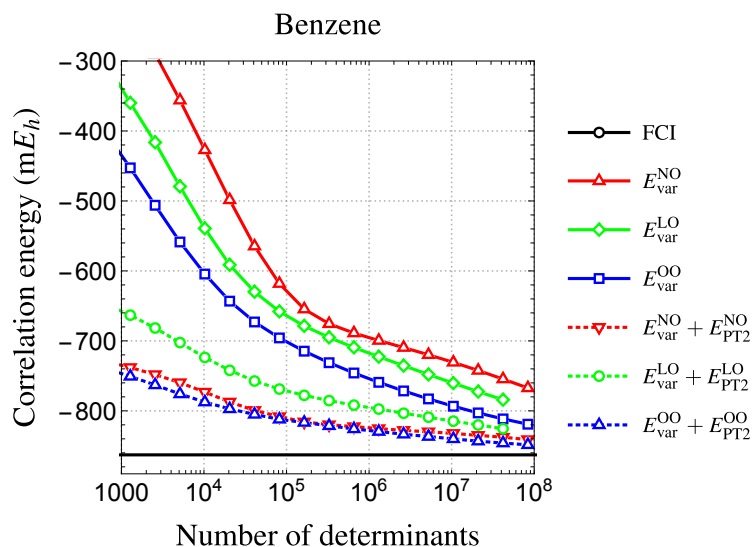


FIGURE 6.4: ΔE_{var} (solid) and $\Delta E_{\text{var}} + E_{\text{PT2}}$ (dashed) computed in the cc-pVDZ basis as functions of the number of determinants N_{det} in the variational space for the benzene molecule. Three sets of orbitals are considered: natural orbitals (NOs, in red), localized orbitals (LOs, in green), and optimized orbitals (OOs, in blue). The FCI estimate of the correlation energy is represented as a thick black line.

6.3 Computational details

The geometries of the twelve systems considered in the present study were all obtained at the CC3/*aug*-cc-pVTZ level of theory and were extracted from a previous study [32]. Note that, for the sake of consistency, the geometry of benzene considered here is different from the one of Ref. [249] which was obtained at a lower level of theory [MP2/6-31G(d)] [196]. The MP2, MP3, MP4, CC2, CC3, CC4, CCSD, CCSDT, and CCSDTQ calculations were performed with CFOUR [466], the CR-CC(2,3) calculations were made with GAMESS 2014R1 [467], and MP5 and CCSD(T) calculations were computed with GAUSSIAN 09 [468]. The CIPSI calculations were performed with QUANTUM PACKAGE [35]. In the current implementation, the selection step and the PT2 correction are computed simultaneously via a hybrid semistochastic algorithm.[35, 259] Here, we employ the renormalized version of the PT2 correction which was recently implemented and tested for a more efficient extrapolation to the FCI limit thanks to a partial resummation of the higher orders of perturbation [35]. We refer the interested reader to Ref. [35] for further details. For all these calculations, Dunning’s correlation-consistent double- ζ basis (cc-pVDZ) has been employed.

Although the FCI energy has the enjoyable property of being independent of the set of one-electron orbitals used to construct the many-electron Slater determinants, as a truncated CI method, the convergence properties of CIPSI strongly depend on this orbital choice. In the present study, we investigate, in particular, the convergence behavior of the CIPSI energy for two sets of orbitals: natural orbitals (NOs) and optimized orbitals (OOs). Following our usual procedure [30, 32, 36, 39, 221, 249, 264, 265, 287], we perform first a preliminary SCI calculation using HF orbitals in order to generate a SCI wave function with at least 10^7 determinants. Natural orbitals are computed based on this wave function and they are used to perform a new CIPSI run up to 8×10^7 determinants. Successive orbital optimizations are then performed, which consist in minimizing the variational CIPSI energy at each macroiteration up to approximately 2×10^5 determinants. When convergence is achieved in terms of orbital optimization, as our production run, we perform a new CIPSI calculation from scratch using this set of optimized orbitals to 8×10^7 determinants. Using optimized orbitals has the undeniable advantage to produce, for a given variational energy, more compact CI expansions (see section 6.4). For the benzene molecule, we have also explored the use of localized orbitals (LOs) which are produced with the Boys-Foster localization procedure [469] that we apply to the natural orbitals in several orbital windows in order to preserve a strict

TABLE 6.3: Extrapolated correlation energies ΔE_{extrap} (in mE_h) computed in the cc-pVDZ basis for the twelve cyclic molecules represented in Figure 6.1 and their associated fitting errors (in mE_h) obtained via weighted linear fits with a varying number of points. Two sets of orbitals are considered: natural orbitals and optimized orbitals. The weights are taken as the inverse square of the perturbative corrections.

For a m -point fit, the m largest variational wave functions are used.

Molecule	Number of fitting points	Natural orbitals		Optimized orbitals	
		ΔE_{extrap}	Fitting error	ΔE_{extrap}	Fitting error
Cyclopentadiene	3	-740.639	0.273	-739.295	0.199
	4	-740.243	0.306	-739.309	0.088
	5	-740.047	0.242	-739.230	0.074
	6	-739.952	0.187	-739.304	0.072
	7	-739.761	0.204	-739.292	0.055
Furan	3	-766.090	0.729	-767.790	0.064
	4	-766.445	0.459	-768.104	0.196
	5	-766.582	0.318	-768.194	0.135
	6	-766.366	0.288	-768.060	0.131
	7	-766.507	0.254	-768.086	0.101
Imidazole	3	-778.148	2.197	-778.295	0.356
	4	-777.436	1.107	-778.270	0.150
	5	-776.300	0.996	-778.178	0.105
	6	-776.104	0.712	-778.174	0.072
	7	-776.098	0.541	-778.051	0.099
Pyrrole	3	-758.309	0.447	-758.650	0.321
	4	-758.749	0.393	-758.389	0.174
	5	-758.405	0.359	-758.460	0.110
	6	-758.136	0.334	-758.352	0.100
	7	-757.990	0.283	-758.347	0.075
Thiophene	3	-728.054	0.134	-728.744	0.691
	4	-728.240	0.139	-729.052	0.331
	5	-728.243	0.087	-728.948	0.203
	6	-728.242	0.062	-728.987	0.140
	7	-728.420	0.144	-729.067	0.117
Benzene	3	-860.350	0.496	-862.325	0.279
	4	-861.949	0.811	-863.024	0.424
	5	-861.807	0.474	-862.890	0.266
	6	-861.110	0.539	-862.360	0.383
	7	-861.410	0.444	-862.083	0.339
Pyrazine	3	-904.148	0.035	-904.867	1.420
	4	-904.726	0.377	-904.588	0.650
	5	-904.274	0.383	-904.550	0.385
	6	-903.980	0.341	-903.982	0.439
	7	-903.621	0.370	-903.746	0.359
Pyridazine	3	-910.856	3.053	-909.292	0.024
	4	-908.222	1.834	-908.808	0.230
	5	-909.282	1.191	-908.820	0.133
	6	-912.566	1.727	-908.342	0.303
	7	-910.694	2.210	-908.368	0.224
Pyridine	3	-883.025	3.919	-883.363	0.047
	4	-883.862	1.869	-883.413	0.029
	5	-881.664	1.760	-882.700	0.405
	6	-880.422	1.456	-882.361	0.341
	7	-880.191	1.084	-882.023	0.330
Pyrimidine	3	-900.386	1.884	-900.817	0.726
	4	-901.441	0.991	-900.383	0.356
	5	-900.354	0.865	-900.496	0.214
	6	-900.240	0.594	-900.698	0.190
	7	-899.689	0.565	-900.464	0.206
s-Tetrazine	3	-958.736	0.320	-957.559	0.246
	4	-958.727	0.148	-957.299	0.160
	5	-958.500	0.172	-957.869	0.349
	6	-958.162	0.260	-957.744	0.247
	7	-958.161	0.198	-957.709	0.183
s-Triazine	3	-917.221	0.693	-919.596	0.105
	4	-918.723	0.913	-918.457	0.538
	5	-917.402	0.956	-918.355	0.312
	6	-916.517	0.862	-918.206	0.226
	7	-916.544	0.643	-917.876	0.267

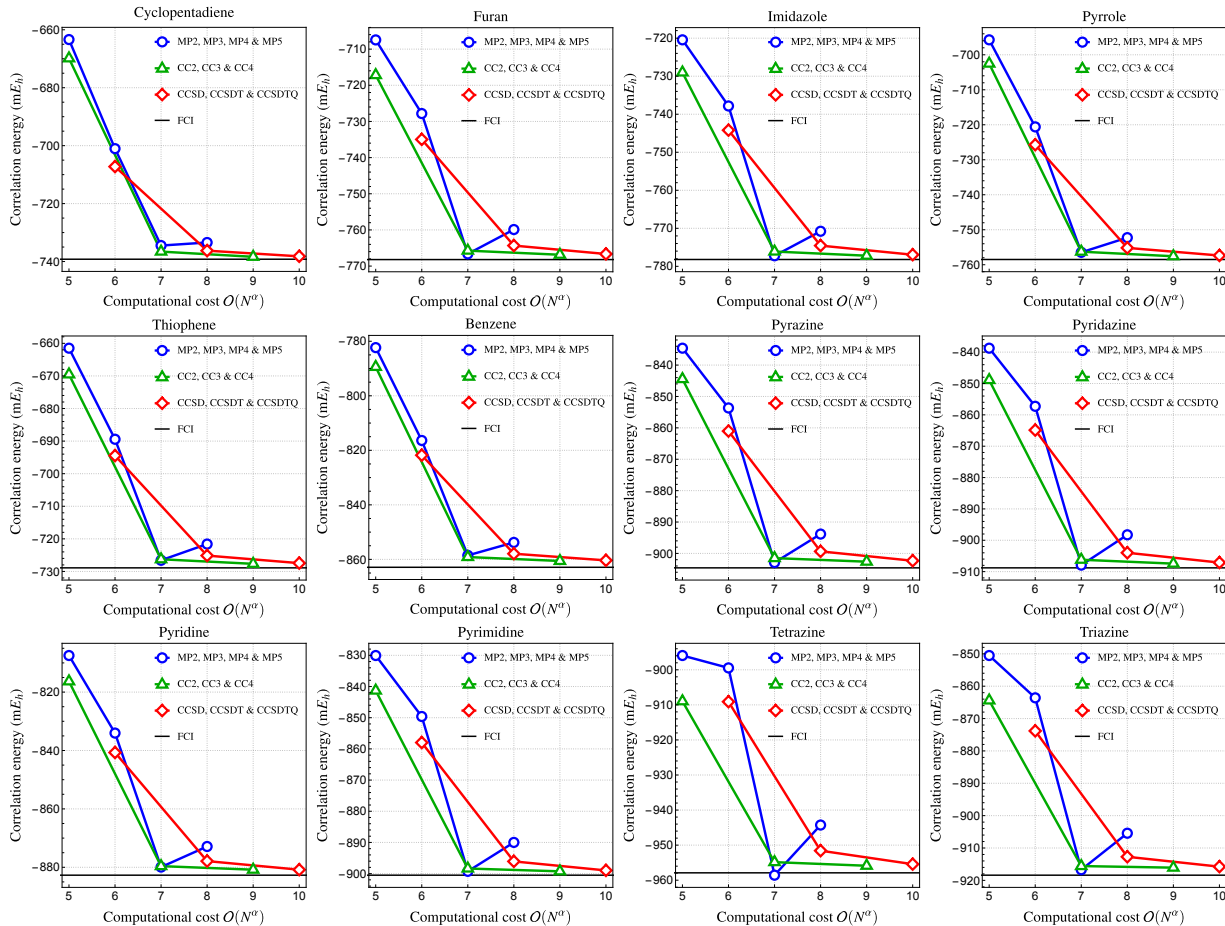


FIGURE 6.5: Convergence of the correlation energy (in mE_h) computed in the cc-pVDZ basis as a function of the formal computational scaling for the twelve cyclic molecules represented in Figure 6.1. Three series of methods are considered: i) MP2, MP3, MP4, and MP5 (blue), ii) CC2, CC3, and CC4 (green), and iii) CCSD, CCSDT, CCSDTQ (red). The FCI estimate of the correlation energy is represented as a black line.

TABLE 6.4: Mean absolute error (MAE), mean signed error (MSE), and minimum (Min) and maximum (Max) absolute errors (in mE_h) with respect to the FCI correlation energy computed in the cc-pVDZ basis for various methods. The formal computational scaling of each method is also reported.

Method	Scaling	MAE	MSE	Max	Min
MP2	$\mathcal{O}(N^5)$	68.4	68.4	80.6	57.8
MP3	$\mathcal{O}(N^6)$	46.5	46.5	58.4	37.9
MP4	$\mathcal{O}(N^7)$	2.1	2.0	4.7	0.7
MP5	$\mathcal{O}(N^8)$	9.4	9.4	13.6	5.8
CC2	$\mathcal{O}(N^5)$	58.9	58.9	73.5	48.9
CC3	$\mathcal{O}(N^7)$	2.7	2.7	3.8	2.1
CC4	$\mathcal{O}(N^9)$	1.5	1.5	2.3	0.8
CCSD	$\mathcal{O}(N^6)$	39.4	39.4	48.8	32.0
CCSDT	$\mathcal{O}(N^8)$	4.5	4.5	6.3	3.0
CCSDTQ	$\mathcal{O}(N^{10})$	1.8	1.8	2.6	1.0
CCSD(T)	$\mathcal{O}(N^7)$	4.5	4.5	5.7	3.6
CR-CC(2,3)	$\mathcal{O}(N^7)$	5.0	5.0	6.6	3.6

σ - π separation in the planar systems considered here [249]. Because they take advantage of the local character of electron correlation, localized orbitals have been shown to provide faster convergence towards the FCI limit compared to natural orbitals [38, 249, 422, 470–473]. As we shall see below, employing optimized orbitals has the advantage to produce an even smoother and faster convergence of the SCI energy toward the FCI limit. Note that both localized and optimized orbitals do break the spatial symmetry. Unlike excited-state calculations where it is important to enforce that the wave functions are eigenfunctions of the \hat{S}^2 spin operator [474], the present wave functions do not fulfill this property as we aim for the lowest possible energy of a closed-shell singlet state. We have found that $\langle \hat{S}^2 \rangle$ is, nonetheless, very close to zero ($\sim 10^{-3}$) for each system.

The present CIPSI calculations have been performed on the AMD partition of GENCI's Irene supercomputer. Each Irene's AMD node is a dual-socket AMD Rome (Epyc™) CPU at 2.60 GHz with 256 GiB of RAM, with a total of 64 physical cores per socket. These nodes are connected via Infiniband HDR100. In total, the present calculations have required around 3 million core hours.

All the data (geometries, energies, etc) and supplementary material associated with the present chapter are openly available in Zenodo at <http://doi.org/10.5281/zenodo.5150663>.

6.4 Results and discussion

6.4.1 CIPSI estimates

We first study the convergence of the CIPSI energy as a function of the number of determinants. Our motivation here is to generate FCI-quality reference correlation energies for the twelve cyclic molecules represented in Figure 6.1 in order to benchmark the performances of various mainstream MP and CC methods (see section 6.4.2). For the natural and optimized orbital sets, we report, in Figure 6.2, the evolution of the variational correlation energy $\Delta E_{\text{var}} = E_{\text{var}} - E_{\text{HF}}$ (where E_{HF} is the HF energy) and its perturbatively corrected value $\Delta E_{\text{var}} + E_{\text{PT2}}$ with respect to the number of determinants N_{det} for each cyclic molecule. As compared to natural orbitals (solid red lines), one can see that, for a given number of determinants, the use of optimized orbitals greatly lowers ΔE_{var} (solid blue lines). Adding the perturbative correction E_{PT2} yields very similar curves for both sets of orbitals (dashed lines). This indicates that, for a given number of determinants, E_{PT2} (which, we recall, provides a qualitative idea to the distance to the FCI limit) is much smaller for optimized orbitals than for natural orbitals. This is further evidenced in Figure 6.3 where we show the behavior of ΔE_{var} as a function of E_{PT2} for both sets of orbitals. From Figure 6.3, it is clear that the behavior of ΔE_{var} is much more linear and produces smaller E_{PT2} values when optimized orbitals are selected, hence facilitating the extrapolation procedure to the FCI limit (see below). The five-point weighted linear fit using the five largest variational wave functions are also represented (dashed black lines), while the FCI estimate of the correlation energy (solid black line) is reported for reference in Figures 6.2 and 6.3.

Figure 6.4 compares the convergence of ΔE_{var} for natural, localized, and optimized orbitals for benzene. As mentioned in section 6.3, although both the localized and optimized orbitals break the spatial symmetry to take advantage of the local nature of electron correlation, the latter set further improves on the use of former set. More quantitatively, optimized orbitals produce the same variational energy as localized orbitals with, roughly, a ten-fold reduction in the number of determinants. A similar improvement is observed going from natural to localized orbitals. According to these observations, all our FCI correlation energy estimates have been produced from the set of optimized orbitals.

To this end, we have extrapolated the orbital-optimized variational CIPSI correlation energies to $E_{\text{PT2}} = 0$ via a weighted five-point linear fit using the five largest variational wave functions (see Figure 6.3). The fitting weights have been taken as the inverse square of the perturbative corrections. Our final FCI correlation energy estimates are reported in Tables 6.1 and 6.2 for the five- and six-membered rings, respectively, alongside their corresponding fitting error. The stability of these estimates are illustrated by the results gathered in Table 6.3, where we list the extrapolated correlation energies ΔE_{extrap} and their associated fitting errors obtained via weighted linear fits

varying the number of fitting points from 3 to 7. Although we cannot provide a mathematically rigorous error bar, the data provided by Table 6.3 show that the extrapolation procedure is robust and that our FCI estimates are very likely accurate to a few tenths of a millihartree. Logically, the FCI estimates for the five-membered rings seem slightly more accurate than for the (larger) six-membered rings. It is pleasing to see that, although different geometries are considered, our present estimate of the frozen-core correlation energy of the benzene molecule in the cc-pVDZ basis (-862.9 mE_h) is very close to the one reported in Ref. [249] (-863.4 mE_h).

Table 6.3 does report extrapolated correlation energies and fitting errors for both natural and optimized orbitals. Again, the superiority of the latter set is clear as both the variation in extrapolated values and the fitting error are much larger with the natural set. Taking cyclopentadiene as an example, the extrapolated values vary by almost 1 mE_h with natural orbitals and less than 0.1 mE_h with the optimized set. The fitting errors follow the same trend.

6.4.2 Benchmark of CC and MP methods

Using the CIPSI estimates of the FCI correlation energy produced in section 6.4.1, we now study the performance and convergence properties of three series of methods: i) MP2, MP3, MP4, and MP5, ii) CC2, CC3, and CC4, and iii) CCSD, CCSDT, and CCSDTQ. Additionally, we also report CCSD(T) and CR-CC(2,3) correlation energies. The raw data are reported in Tables 6.1 and 6.2 for the five- and six-membered rings, respectively. In Figure 6.5, we show, for each molecule, the convergence of the correlation energy for each series of methods as a function of the formal computational scaling of the corresponding method. Statistical quantities [mean absolute error (MAE), mean signed error (MSE), minimum (Min) and maximum (Max) absolute errors with respect to the FCI reference values] are also reported in Table 6.4 for each method as well as their formal computational scaling.

First, we investigate the “complete” and well-established series of methods CCSD, CCSDT, and CCSDTQ. Unfortunately, CC with singles, doubles, triples, quadruples, and pentuples (CCSDTQP) calculations are out of reach here [229, 475]. As expected for the present set of weakly correlated systems, going from CCSD to CCSDTQ, one systematically and quickly improves the correlation energies with respective MAEs of 39.4 , 4.5 and 1.8 mE_h for CCSD, CCSDT, and CCSDTQ. As usually observed, CCSD(T) (MAE of 4.5 mE_h) provides similar correlation energies than the more expensive CCSDT method by computing perturbatively (instead of iteratively) the triple excitations, while CCSD(T) and CR-CC(2,3) performs equally well.

Second, we investigate the approximate CC series of methods CC2, CC3, and CC4. As observed in our recent study on excitation energies [397], CC4, which returns a MAE of 1.5 mE_h , is an outstanding approximation to its CCSDTQ parent (MAE of 1.8 mE_h) and is, in the present case, even slightly more accurate in terms of mean errors as well as maximum and minimum absolute errors. Moreover, we observe that CC3 provides very accurate correlation energies with a MAE of 2.7 mE_h , showing that this iterative method is particularly effective for ground-state energetics and outperforms both the perturbative CCSD(T) and iterative CCSDT models. It is important to mention that even if the two families of CC methods studied here are known to be non-variational (see section 6.1), for the present set of weakly-correlated molecular systems, they never produce a lower energy than the FCI estimate as illustrated by the systematic equality between MAEs and MSEs.

Third, let us look into the MP series which is known, as mentioned in section 6.1, to potentially exhibit “surprising” behaviors depending on the type of correlation at play.[436–445] (See Ref. [429] for a detailed discussion). For each system, the MP series decreases monotonically up to MP4 but raises quite significantly when one takes into account the fifth-order correction. We note that the MP4 correlation energy is always quite accurate (MAE of 2.1 mE_h) and is only a few millihartree higher than the FCI value (except in the case of s-tetrazine where the MP4 number is very slightly below the reference value): MP5 (MAE of 9.4 mE_h) is thus systematically worse than MP4 for these weakly-correlated systems. Importantly here, one notices that MP4 [which scales as $\mathcal{O}(N^7)$] is systematically on par with the much more expensive $\mathcal{O}(N^{10})$ CCSDTQ method which exhibits a slightly smaller MAE of 1.8 mE_h .

6.5 Conclusion

Using the SCI algorithm named *Configuration Interaction using a Perturbative Selection made Iteratively* (CIPSI), we have produced FCI-quality frozen-core correlation energies for twelve cyclic molecules (see Figure 6.1) in the correlation-consistent double- ζ Dunning basis set (cc-pVDZ). These estimates, which are likely accurate to a few tenths of a millihartree, have been obtained by extrapolating CIPSI energies to the FCI limit based on a set of orbitals obtained by minimizing the CIPSI variational energy. Using energetically optimized orbitals, one can reduce the size of the variational space by one order of magnitude for the same variational energy as compared to natural orbitals.

Thanks to these reference FCI energies, we have then benchmarked three families of popular electronic structure methods: i) the MP perturbation series up to fifth-order (MP2, MP3, MP4, and MP5), ii) the approximate CC series CC2, CC3, and CC4, and iii) the “complete” CC series CCSD, CCSDT, and CCSDTQ. With a $\mathcal{O}(N^7)$ scaling, MP4 provides an interesting accuracy/cost ratio for this particular set of weakly correlated systems, while MP5 systematically worsen the perturbative estimates of the correlation energy. In addition, CC3 (where the triples are computed iteratively) outperforms the perturbative-triples CCSD(T) method with the same $\mathcal{O}(N^7)$ scaling, its completely renormalized version CR-CC(2,3), as well as its more expensive parent, CCSDT. A similar trend is observed for the methods including quadruple excitations, where the $\mathcal{O}(N^9)$ CC4 model has been shown to be slightly more accurate than CCSDTQ [which scales as $\mathcal{O}(N^{10})$], both methods providing correlation energies within $2 mE_h$ of the FCI limit. Of course, the present trends are only valid for this particular class of (weakly-correlated) molecules and it would be desirable to have a broader variety of systems in the future by including more challenging systems such as, for example, transition metal compounds. Some work along this line is currently being performed.

As perspectives, we are currently investigating the performance of the present approach for excited states in order to expand the QUEST database of vertical excitation energies [149]. We hope to report on this in the near future. The compression of the variational space brought by optimized orbitals could be also beneficial in the context of quantum Monte Carlo methods to generate compact, yet accurate multi-determinant trial wave functions [288–290, 476].

Chapter 7

General conclusion

In this thesis, I have presented the main projects where I was one of the main contributors. The main purpose of the present thesis is to perform a large overview of electronic structure methods that target excited states but also the development of web tools for quantum chemistry.

In order to present the work I have done in this thesis, in the introduction, I have presented, from a general point of view, excited states in molecular systems and computational resources (Chapter 1). In addition, Chapter 2 provides a general overview of the various methods which have been employed in this thesis.

Chapter 3 deals with the study of GW methods, a family of approximations that is highly successful at predicting the electronic properties of solids and molecules in the weakly-correlated regime [57–59]. In this chapter, we have discussed and exemplified a severe limitation of the GW approximation, namely the appearance of discontinuities and/or irregularities in key physical and chemical quantities. These issues are particularly troublesome in the context of the calculation of potential energy surfaces, which are of paramount importance in photochemistry (for example). Although GW is known to be inadequate in the case of strongly correlated systems where the quasiparticle approximation usually breaks down [138–142], these severe shortcomings of two widely-used variants of GW have been observed in the weakly correlated regime. More specifically, we have evidenced that one can hit multiple solution issues within G_0W_0 and $evGW$ due to the location of the quasiparticle solution near poles of the self-energy. In the case of linearized G_0W_0 , this implies irregularities in key experimentally-measurable quantities of simple diatomics, while, and in the case of partially self-consistent $evGW$, discontinuities arise. The correlation energy of the ground states and the excited states depend on the quasiparticle energies so the energy surfaces of these electronic states are affected by the multi-solution issue. My colleagues have also shown that these discontinuities affect the Bethe-Salpeter equation applied on top of GW (BSE@GW) as well because of its dependency on the quasiparticle energies [477]. Recently my coworkers have also shown that the present issues can be cured if one relies on a *fully* self-consistent GW (scGW) scheme since one does not have to make a distinction between quasiparticle and satellites [478]. Thus, one obtains continuous quantities with respect to the geometric parameters. However, one has to mention that scGW is significantly more computationally expensive and cumbersome than G_0W_0 and $evGW$. According to another article recently published by some of my coworkers on this project [479], the Coulomb hole plus screened exchange (COHSEX) scheme fixes the issue of multiple solutions. Indeed, the physics inside the COHSEX self-energy is very similar to that included in the GW self-energy, but, unlike the GW self-energy, it is Hermitian and frequency-independent. As a consequence, COHSEX calculations can be done self-consistently using standard diagonalization of a Fock-like operator.

In the next chapter (Chapter 4), we have discussed molecular excited states in the context of the QUEST database of highly-accurate excitation energies [30, 32, 36, 39, 221] that we started building in 2018 and that is now composed by more than 600 vertical excitations, many of which can be reasonably considered chemically accurate for the considered CC3/*aug-cc-pVTZ* geometry and basis set (*aug-cc-pVTZ*). In particular, we have provided computational details regarding geometries, basis sets, as well as reference and benchmarked computational methods to allow the reproducibility of the results. The content of the five QUEST subsets has been presented in detail, and for each of them, we have provided the number of reference excitation energies, the nature

and size of the molecules, the list of benchmarked methods, as well as other useful specificities. Importantly, we have proposed a new statistical method that produces much safer estimates of the extrapolation error in SCI calculations. This new method based on Gaussian random variables has been tested by computing additional FCI values for five- and six-membered rings. After having discussed the generation of our TBEs, we have reported a comprehensive benchmark for a significant number of methods on the entire QUEST set with, in addition, a specific analysis for each type of excited states. Finally, the main features of the website specifically designed to gather the entire data generated during these past few years have been presented and discussed. A key aspect of this website is the possibility of generating statistics that match the needs of each specific user. In this framework, we have provided in the supplementary material of the paper [149] a file with all our benchmark data. Geometries can be found there, but they can also be downloaded from the website. Recently, my colleagues continued to expand the QUEST database with QUEST#6 that provide reference excitation energies for charge-transfer excited states.

For the QUEST project, we built a web tool to help quantum chemists who work on excited states to easily find reference values and adequate methods for their own interest. In the next chapter (Chapter 5), we continued in this direction by building a web demo for QUANTUM PACKAGE, a quantum chemistry software developed in your research group in order to provide an easy-to-use and easy-to-develop quantum chemistry software to the electronic structure community. The flagship method of QUANTUM PACKAGE is *Configuration Interaction using a Perturbative Selection made Iteratively* (CIPSI) that provides near-FCI energies for a fraction of its computational cost via a sparse exploration of the FCI space. The demo, which has been built using SHELLINABOX and DOCKER, was designed in order to make QUANTUM PACKAGE easy to test via a terminal emulator on a web page. While SHELLINABOX is the tool that provides the terminal emulator web elements and the connection to the remote machine, DOCKER provides the isolated and ephemeral environment for the demo. In practice, all these components are wrapped in a virtual machine (VM) to isolate it from the rest of the server. One can test QUANTUM PACKAGE at <https://quantumpackage.github.io/qp2/page/try> and one can download the ready-to-run QUANTUM PACKAGE DOCKER image on Docker Hub at https://hub.docker.com/r/mveril/qp_demo. Recently, we discovered [408] the ability to create a system-level container in order to use DOCKER in a DOCKER container without security issue. This can be a solution for transferring the host containing DOCKER and SHELLINABOX from a VM to a DOCKER container and thus improve performance without creating a security issue for the host machine. We also heard about a containerization software called SINGULARITY. The main interest of SINGULARITY compared to DOCKER is its features that allow it to be adapted to the HPC (High Performance Computing) world. We shall investigate the compatibility of the QUANTUM PACKAGE demo container with the latter in the foreseeable future.

In the continuity of the search for reference energies, the next chapter (Chapter 6) applied the same type of protocol than the QUEST project but on the ground state of the five- and six-membered cyclic molecules included in the QUEST database. Indeed, using CIPSI, we have produced FCI-quality frozen-core correlation energies for the twelve cyclic molecules in the correlation-consistent double- ζ Dunning basis set (cc-pVDZ). These estimates, which are likely accurate to a few tenths of a millihartree, have been obtained by extrapolating CIPSI energies to the FCI limit. In order to improve the calculation performance we have based our FCI on a set of orbitals obtained by minimizing the CIPSI variational energy. Using energetically optimized orbitals, one can reduce the size of the variational space by one order of magnitude for the same variational energy as compared to natural orbitals. In the same way as for the QUEST project, thanks to these reference FCI energies, we have then benchmarked three families of popular electronic structure methods: i) the Møller-Plesset perturbation series up to fifth-order, ii) the approximate CC series from CC2, to CC4, and iii) the “complete” CC series from CCSD to CCSDTQ. With a $\mathcal{O}(N^7)$ scaling, MP4 provides an interesting accuracy/cost ratio for this particular set of weakly correlated systems, while MP5 systematically worsen the perturbative estimates of the correlation energy. In addition, CC3 (where the triples are computed iteratively) outperforms the perturbative-triples CCSD(T) method with the same $\mathcal{O}(N^7)$ scaling, its completely renormalized version CR-CC(2,3), as well as its more expensive

parent, CCSDT. A similar trend is observed for the methods including quadruple excitations, where the $\mathcal{O}(N^9)$ CC4 model has been shown to be slightly more accurate than CCSDTQ [which scales as $\mathcal{O}(N^{10})$], both methods providing correlation energies within $2 mE_h$ of the FCI limit. Of course, the present trends are only valid for this particular class of (weakly-correlated) molecules and it would be desirable to test this procedure on a larger variety of molecules by following the same principle as for the QUEST project. In particular, we plan on attempting the same approach on more difficult systems such as transition metal compounds. Some work along this line is currently being performed. As perspectives, we are currently investigating the performance of the present approach for excited states in order to expand the QUEST database of vertical excitation energies [149]. We hope to report on this in the near future. The compression of the variational space brought by optimized orbitals could be also beneficial in the context of quantum Monte Carlo methods to generate compact, yet accurate multi-determinant trial wave functions [288–290, 476].

Although GW methods are widely used for the calculation of electronic transitions and band gaps (also in combination with the Bethe-Salpeter equation formalism), we have shown that in some cases one can observe (unphysical) discontinuities in the energy surfaces of several key quantities. Encouraging work makes it possible to see that some workaround are possible like the usage of self-consistent GW (scGW) which unfortunately is computationally heavy [478] or COHSEX that eliminate the discontinuities by using a frequency-independent algorithm [479]. We have also started a long work to provide very accurate reference vertical transition energies and used them to benchmark many low-level quantum chemistry methods. Following the same philosophy as GW100 [185] we have created a website to gather all the data of the QUEST project. We have already started working to complete this database with additional data. An example is the work of my coworkers on intramolecular charge-transfer excitations [397]. Also our work on optimized orbitals [409] could be a key for improving the reference data of the QUEST project. We could also reproduce what we did with the QUEST project, but for ground-state correlation energies, atomization energies, and/or ionization energies and electron affinities. Also we plan on applying the optimized orbitals method again but this time on transition metal compounds. Another interesting work has also been done is the web demo of QUANTUM PACKAGE that allows quantum chemists to test QUANTUM PACKAGE directly on the web. This work led to the creation of a VM containing a web terminal server and a DOCKER image used as a demo environment. An improvement on this architecture could be implemented in the future using a “system level container” to replace the VM in order to improve the performance. Also the DOCKER image used for the demo can be used to enable the usage of QUANTUM PACKAGE on WINDOWS or MACOS desktop computer (even if other solutions exists) or to avoid the long QUANTUM PACKAGE installation process. The DOCKER image of QUANTUM PACKAGE could also be used on HPC machine using SINGULARITY but we need to study the compatibility of this image with this software.

Annexe A

Résumé en français

Après une introduction générale sur les états excités dans les systèmes moléculaires et les ressources de calcul (chapitre 1) ainsi qu'une vue d'ensemble des différentes méthodes utilisées dans cette thèse (chapitre 2), nous allons parler de deux sujets principaux : i) les états électroniques moléculaires excités et ii) les outils numériques pour la chimie quantique théorique.

La thèse actuelle commence par une étude de la famille d'approximations GW (chapitre 3). Celles-ci sont connues pour être des méthodologies précises et relativement bon marché pour le calcul des excitations chargées, telles que les potentiels d'ionisation et les affinités électroniques dans les molécules et les intervalles de bande dans les solides. Ces méthodes peuvent être perturbatives, partiellement ou totalement auto-cohérentes et sont capables de fournir des valeurs de propriétés chimiquement précises pour les systèmes faiblement corrélés. Toutefois, comme indiqué en détails, des discontinuités peuvent être observées pour plusieurs propriétés essentielles, notamment dans les surfaces d'énergie potentielle, même dans les régimes faiblement corrélés. Ces discontinuités sont dues à des pôles dans la self-energy — une quantité clé du formalisme GW — chacune de ses branches est associée à une solution distincte de l'équation quasiparticule. Nous montrons que, dans les molécules diatomiques, le comportement multisolution dans les orbitales frontières est plus probable si l'écart d'énergie entre ces orbitales est faible.

La partie centrale de cette thèse traite du projet QUEST (chapitre 4), une vaste base de données de plus de 600 énergies d'excitation verticales de différentes nature. L'objectif de cette base de données open-source facilement modifiable est de fournir des énergies d'excitation de référence extrêmement précises pour des analyses comparatives et des comparaisons croisées de modèles de calcul. Ces énergies d'excitation de référence sont divisées en six sous-ensembles en fonction de la taille des molécules et des types d'excitation. Afin de rassembler l'énorme quantité de données du projet QUEST, nous avons créé un site web où l'on peut facilement tester et comparer la précision d'une méthode donnée par rapport à diverses variables telles que la taille de la molécule ou sa famille, la nature de l'état excité, le type de base, etc.

La partie suivante concerne les outils numériques, en particulier, la mise en place d'une démo web de QUANTUM PACKAGE (chapitre 5), un logiciel open-source développé principalement par notre groupe de recherche, qui a été conçu pour être facile à utiliser par des chimistes quantiques, qu'ils soient utilisateurs ou développeurs. La possibilité de tester le logiciel sans la nécessité de l'installer est une réelle opportunité de faire que QUANTUM PACKAGE soit plus populaire dans notre communauté. La technologie de conteneurisation fournie par DOCKER et le noyau LINUX nous permet de démarrer rapidement un environnement de démonstration propre pour chaque utilisateur et de l'arrêter facilement avec des performances intéressantes.

Enfin, au chapitre 6, poursuivant les travaux récents de notre équipe sur la molécule de benzène, nous reportons des énergies de corrélation de coeur gelé de référence pour les douze molécules cyclique à cinq et six atomes du projet QUEST, en utilisant l'algorithme d'interaction de configuration à l'aide d'une sélection perturbative effectuée itérativement (CIPSI) dans la base de corrélation standard double ζ (cc-pVDZ). Celles-ci correspondent à des espaces de Hilbert avec des tailles allant de 10^{28} à 10^{36} . Les propriétés de performance et de convergence de plusieurs séries de méthodes sont étudiées. En particulier, nous étudions les propriétés de convergence de la série de perturbations Møller-Plesset jusqu'au cinquième ordre ainsi que divers modèles de clusters couplés à référence unique qui incluent des excitations quadruples.

Nous tirons notre conclusion générale au chapitre 7.

Annexe B

Résumé étendu en français

Introduction

Généralité

Le but de la chimie théorique est de décrire mathématiquement les systèmes chimiques et biochimiques afin de prédire leurs propriétés. Formellement en chimie quantique, nous résolvons l'équation de Schrödinger, sauf pour l'atome d'hydrogène et les ions hydrogénoïdes, nous avons un problème à plusieurs corps qui ne peut pas être résolu sans approximation. L'une des approximations les plus courantes est l'introduction d'une base monoélectronique. Dans celle-ci, la méthode d'interaction de configuration complète (FCI) est la solution exacte de l'équation de Schrödinger, mais malheureusement, cette méthode est d'ordre $\mathcal{O}(N!)$ et ne peut être appliquée que pour des systèmes très petits. Comme point de départ, l'approximation la plus élémentaire de la chimie quantique est la méthode Hartree-Fock (HF) [1]. Avec cette méthode, nous introduisons la notion de champ moyen qui permet d'approximer l'interaction électronique en une interaction entre l'électron et la moyenne des champs de Coulomb des autres électrons. Cette méthode est d'ordre $\mathcal{O}(N^3)$ et peut être appliquée aux gros systèmes. L'erreur entre l'énergie Hartree-Fock et la solution exacte de l'équation de Schrödinger s'appelle l'énergie de corrélation [2, 3].

Même si depuis le 20e siècle et l'avènement de l'ordinateur [4], nous avons accès à de plus en plus de puissance de calcul, nous ne pouvons calculer l'énergie FCI uniquement pour de très petits systèmes. Habituellement, nous utilisons des approximations de la méthode FCI, tant FCI est chère. Ces approximations forment un grand ensemble allant de méthodes de chimie quantique coûteuses aux méthodes de mécanique moléculaire peu coûteuses. Il existe de nombreuses bases et de nombreuses méthodes avec chacune leurs propres approximations et systèmes chimiques sur lesquels nous pouvons les appliquer en fonction de notre domaine d'études.

États excités

Comme représenté sur la Figure B.1, Il existe différents types d'énergies d'excitation. Tout d'abord, nous avons les transitions verticales pour lesquelles la transition électronique entre les deux états se produit sans relaxer la géométrie. Pour l'absorption et la fluorescence, on commence respectivement à partir de la géométrie de l'état d'énergie le plus bas (généralement l'état fondamental) et de la géométrie d'un état excité donné. Dans les transitions adiabatiques, la géométrie du système est relaxée au minimum de la surface d'énergie potentielle. En pratique, il faut donc optimiser non seulement la géométrie de l'état d'énergie le plus bas, mais aussi la géométrie de l'état excité. D'un point de vue computationnel, ce type de calcul peut être un véritable défi technique pour les méthodes ciblant les états excités, car il faut avoir accès aux dérivées premières de l'énergie par rapport aux déplacements nucléaires (voir Tableau B.1). Enfin les énergies 0-0 sont des énergies adiabatiques où l'on prend en compte les corrections d'énergie vibrationnelle de point zéro (ZPVE) pour les deux états. Cependant, la correction de ZPVE représente un important coût de calcul supplémentaire, car on doit calculer les dérivées secondes de l'énergie par rapport aux déplacements nucléaires. Malheureusement, pour les méthodes les plus coûteuses, ce type de calcul est impossible avec les implémentations et les machines actuelles. Notez que, même si

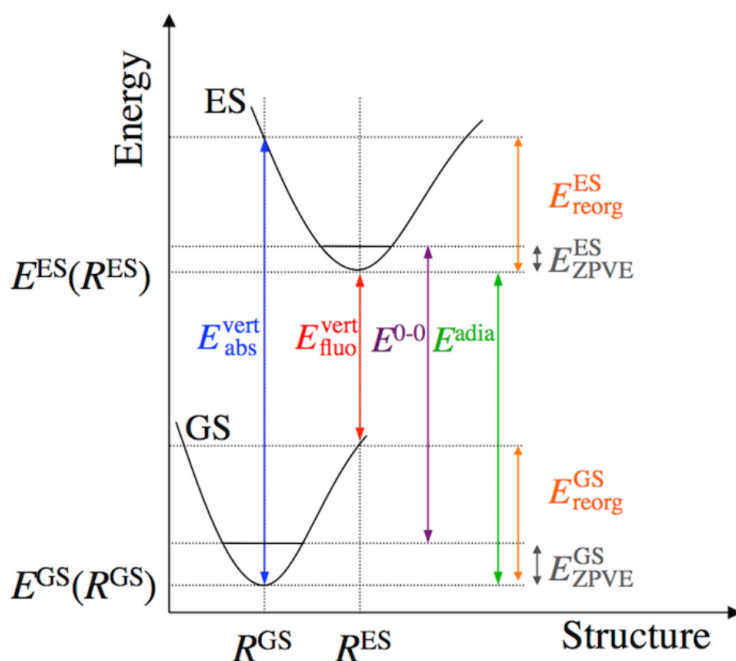


FIGURE B.1 : Représentation des différents types d'énergies de transition où E_{abs}^{vert} en rouge et E_{fluo}^{vert} en bleu sont respectivement les énergies de transition verticale d'absorption et de fluorescence. E^{adia} l'énergie de transition adiabatique en vert et E^{0-0} l'énergie de transition 0-0 en violet. E_{reorg}^{GS} et E_{reorg}^{ES} en orange sont respectivement l'énergie de réorganisation (géométrique) de l'état fondamental et de l'état excité toutes ces énergies sont défini comme des quantités positives [5].

les énergies de transition 0-0 sont beaucoup plus coûteuses d'un point de vue computationnel, elles ont l'énorme avantage d'être directement comparable aux données expérimentales (ce qui n'est pas le cas des énergies d'excitation verticales comme il est fait communément et à tort dans la littérature).

Méthodes

La première méthode conventionnelle *ab initio* pour les états excités était probablement l'interaction de configuration avec les simples (CIS), dans les années 1970 [6]. Globalement, cette méthode surestime de manière significative l'énergie des états excités et il n'est pas inhabituel d'avoir une erreur de ~ 1 eV. Étant donné que CIS fournit souvent un ordre énergétique incorrect des états excités, il est généralement dangereux de l'utiliser comme méthode qualitative. Vingt ans plus tard, une correction perturbative de second ordre a été ajoutée à CIS et donna naissance à CIS(D) [7, 8]. Cette correction de second ordre améliore considérablement la précision par rapport à CIS. En effet, la plage d'erreurs typique est 0,2-0,3 eV. Au début des années 90, la méthode auto-cohérente de l'espace actif complet (CASSCF) [9, 10] et, en particulier, son extension CASPT2 [11] était une véritable percée. Même s'il a fallu plus de 10 ans pour obtenir des gradients analytiques [12], CASPT2 était probablement la première méthode qui pourrait fournir des énergies d'excitation quantitatives pour les systèmes moléculaires chimiquement intéressants [13]. Cependant, dans ses paramètres standard, elle sous-estime clairement les énergies d'excitation verticales des molécules organiques. Au début des années 2000, NEVPT2 (qui corrigent certains problèmes de CASPT2) a été développé par Angeli, Malrieu et leurs coauteurs [14]. CASPT2 et NEVPT2 permettent d'obtenir des résultats plus précis 0,1-0,2 eV et aussi d'avoir accès aux doubles excitations. L'avènement de la théorie de la fonctionnelle de la densité dépendante du temps (TD-DFT) [15, 16] a eu un impact significatif dans la communauté de la chimie théorique, car elle a donné la possibilité d'obtenir des énergies d'excitation relativement précises (0,2-0,4 eV) avec un coût (très) inférieur, car la TD-DFT, dans son implémentation standard, a un coût comme $\mathcal{O}(N^4)$ (où N est proportionnel à la

TABLE B.1 : Scaling formel de diverses méthodes d'état excité (où N est le nombre de fonctions de base N) et l'accessibilité de diverses propriétés principales dans les logiciels de chimie quantique populaires. Pour les dérivés organiques, la plage d'erreurs typique des excitations simples est également fournie comme indicateur qualitatif de la précision de la méthode [39].

Méthode	Scaling formel	Force d'oscillateur	Gradients analytique	Erreur typique (eV)
TD-DFT	$\mathcal{O}(N^4)$	✓	✓	0,2–0,4 ^a
BSE@GW	$\mathcal{O}(N^4)$	✓	✗	0,1–0,3 ^b
CIS	$\mathcal{O}(N^5)$	✓	✓	~1,0
CIS(D)	$\mathcal{O}(N^5)$	✗	✓	0,2–0,3
ADC(2)	$\mathcal{O}(N^5)$	✓	✓	0,1–0,2
CC2	$\mathcal{O}(N^5)$	✓	✓	0,1–0,2
ADC(3)	$\mathcal{O}(N^6)$	✓	✗	0,2
EOM-CCSD	$\mathcal{O}(N^6)$	✓	✓	0,1–0,3
CC3	$\mathcal{O}(N^7)$	✓	✗	~0,04
EOM-CCSDT	$\mathcal{O}(N^8)$	✗	✗	~0,03
EOM-CCSDTQ	$\mathcal{O}(N^{10})$	✗	✗	~0,01
CASPT2/NEVPT2	$\mathcal{O}(N!)$	✓	✓	0,1–0,2
SCI	$\mathcal{O}(N!)$	✗	✗	~0,03
FCI	$\mathcal{O}(N!)$	✓	✓	0,0

^a La marge d'erreurs est fortement dépendante de la fonctionnelle et de l'état. Les valeurs reportées ici correspondent aux cas qui se comportent bien;

^b Barre d'erreurs typique pour les transitions singulet. Des erreurs plus grandes sont souvent observées pour les excitations triplet.

taille du système). C'est actuellement l'une des méthodes de choix de photochimie, bien qu'elle souffre de nombreux problèmes liés, par exemple, des états de transfert de charge [17, 18], de Rydberg [19] et aux doubles excitations [20–22]. Une autre méthode également développée dans les années 2000 est celle du cluster couplé avec équation de mouvement avec des simples et des doubles (EOM-CCSD) [23] où il n'est pas inhabituel d'avoir des erreurs aussi petites que 0,1 eV et 0,2 eV respectivement pour des petits et plus grands composés. Une autre variante (EOM-CCSDT) comprend des excitations triples mais avec un coût plus élevé [24]. Les méthodes du cluster couplé approximé de deuxième et troisième ordre, CC2 [25] et CC3 [26] où on ignore les termes les plus coûteux et évite le stockage des amplitudes d'excitation supérieure, ont été développées quelques années après. La méthode CC3 est particulièrement intéressante car elle fournit des énergies d'excitation quasi systématiquement précises chimiquement pour les transitions simples avec une erreur de ~0,04 eV à un coût réduit ($\mathcal{O}(N^7)$) par rapport à CCSDT ($\mathcal{O}(N^8)$). La réimplémentation récente et plus efficace des méthodes basées sur la fonction de Green ADC(2) [27] et ADC(3) [28, 29] les rend compétitives avec les méthodes CC, en particulier ADC(2) qui donne des résultats satisfaisant (0,1–0,2 eV) avec un coût en $\mathcal{O}(N^5)$ seulement [30–33]. Enfin, le développement récent de méthodes CI sélectionnées (SCI) fortement parallèles (où on sélectionne uniquement les déterminants pertinents) [34, 35] fournit une énergie quasi FCI [30, 36–38] est relativement bon marché par rapport à cette dernière. Toutefois, même si le préfacteur des méthodes SCI est réduit leur scaling globale reste exponentiel avec la taille du système (voir Tableau B.1).

Ce qui a été dit dans le dernier paragraphe est correct dans le cas de monoexcitation, mais

TABLE B.2 : Spécification des processeurs de Curie [41, 42]

	Intel® Xeon®					
	2670@2,60 GHz	2680 v2@2,80 GHz	E5 2680 v3@2,50 GHz	2680 v4@2,40 GHz	6130@2,10 GHz	Gold 5218@2,30 GHz
Cœurs	8	10	12	14	16	16
Threads/cœur (hyper-threading)	2	2	2	2	2	2
Fréquence de base du processeur	2,60 GHz	2,80 GHz	2,50 GHz	2,40 GHz	2,10 GHz	2,30 GHz
Fréquence Turbo maxi	3,30 GHz	3,60 GHz	3,30 GHz	3,30 GHz	3,70 GHz	3,90 GHz
Mémoire cache L3	20 MiB	25 MiB	30 MiB	35 MiB	22 MiB	22 MiB
AVX	AVX	AVX	AVX2	AVX2	AVX-512 (2)	AVX-512 (2)
FMA			FMA3	FMA3	FMA3	FMA3

nous pouvons également avoir des doubles excitations qui sont particulièrement importantes dans les polyènes. D'un point de vue théorique, il n'y a pas vraiment de simples excitations. Chaque excitation est un mélange d'excitation simple, double, triple... Nous définissons donc $\%T_1$ le pourcentage de monoexcitation pour évaluer dans quelle mesure cette excitation peut être considérée comme une simple excitation. Il a été montré [36] que dans ce cas, la méthode CASSCF et ses dérivés sont meilleures que CC3 (qui n'a qu'une description approximée des triples excitations) et toutes les méthodes qui ne décrivent pas correctement les quadruples excitations. Pour décrire efficacement une double excitation, il nous faut au moins des excitations triples, mais pour obtenir de bons résultats, nous avons besoin des excitations quadruples. Ainsi, pour avoir une bonne description des états excités, nous avons besoin d'une bonne description des états doublement excités de ces états excités. Une double excitation sur un état doublement excité équivaut à une quadruple excitation sur l'état fondamental. Cela explique pourquoi CC3 et d'autres méthodes décrivant correctement les états simplement excités décrivent mal les états doublement excités. En résumé au cours des dernières décennies, les chimistes théoriciens ont continué d'innover pour améliorer la précision du calcul des états excités, afin de décrire le mieux possible les transitions électroniques. Nous pouvons voir qu'il n'y a pas de méthodes parfaites et que cela dépend du type d'excitation (simple, double...). Par exemple avec CC3, nous avons une très bonne description des excitations simples, mais nous avons une mauvaise description des doubles excitations. Un autre exemple est la TD-DFT qui donnent un résultat relatif et qui peut être utilisée sur des systèmes plus importants, mais souffre d'un problème avec certains types d'excitation comme les transferts de charge, les états de Rydberg et les doubles excitations. Nous pouvons donc voir qu'il n'y a pas de procédure systématique pour calculer l'énergie d'excitation, mais nous devons rechercher quelle est la bonne méthode selon notre objet d'étude. C'est l'objectif du projet QUEST d'on nous parlons au chapitre 4.

Ressources de calcul

Curie

Nous utilisons différentes stratégies de calcul en fonction de la taille du système étudié et de la méthode utilisée. Nous avons effectivement accès à plusieurs clusters avec des puissances de calculs différentes. Le premier cluster qui nous est accessible est Curie qui appartient à notre laboratoire. Il est utilisé pour les « petits » calculs. Celui-ci a de nombreux types de partitions avec des processeurs (CPUs) différents. Ici, nous ne parlerons que des nœuds xeon (voir Tableau B.3). Avec l'ensemble de processeurs Intel® Xeon® (voir Tableau B.2) disponibles sur Curie, nous pouvons constater que l'évolution des processeurs n'est pas actuellement dû à une amélioration en matière de fréquence, mais plutôt à une augmentation de la parallélisation grâce aux nombres croissants de cœurs par CPU et une vectorisation des calculs via le jeu d'instructions *Advanced Vector Extensions* (AVX) et ses améliorations. Malheureusement, nos programmes de chimie quantique doivent être adaptés à ces nouvelles architectures en proposant des implémentations massivement parallèles de nos méthodes [43].

Nous pouvons calculer théoriquement les Flops par nœud en utilisant l'équation ci-dessous

$$Flops = n_{core} \times freq \times Flops_{cycle} \quad (B.1)$$

TABLE B.3 : Nœuds de curie [41, 42]

Nom	Modèle	Nombre de nœuds	CPUs/nœud	Flops/nœud	Mémoire
xeonv1	Intel® Xeon® E5-2670	14	2	332,8 GFlops	64 GiB
xeonv2	Intel® Xeon® E5-2680 v2	7	2	448 GFlops	64 GiB
xeonv3	Intel® Xeon® E5-2680 v3	9	2	960 GFlops	64 GiB
xeonv4	Intel® Xeon® E5-2680 v4	4	2	1,075 TFlops	128 GiB
xeonv5	Intel® Xeon® Gold 6130	5	2	2,150 TFlops	128 GiB
xeonv6	Intel® Xeon® Gold 5218	1	2	1,075 TFlops	192 GiB

TABLE B.4 : Spécification du processeur d'Olympe (Intel® Xeon® Gold 6140@2,30 GHz) [41, 42, 44]

Spécifications du processeur	valeur
Cœurs	18
Threads/cœur (hyper-threading)	2
Fréquence de base du processeur	2,30 GHz
Fréquence Turbo maxi	3,70 GHz
Cache L3	24,75 MiB
AVX	AVX-512 (2)
FMA	FMA3

où n_{core} est le nombre de cœurs de processeur, $freq$ est la fréquence du processeur, et $Flops_{\text{cycle}}$ est le nombre d'opérations de double précision par cycle. Cette valeur dépend de la capacité qu'a le processeur d'effectuer plusieurs opérations d'addition et de multiplication par cycle, par exemple avec le jeu d'instruction FMA *fused multiply-add* et/ou la possibilité de vectoriser le calcul à l'aide d'AVX. À l'exception de xeonv6, nous avons une augmentation du nombre de Flops/nœud (voir Tableau B.3).

Olympe

Lorsque nous devons effectuer des calculs plus lourds, nous pouvons utiliser les ressources informatiques du CALMIP. En 2018, le CALMIP a abandonné son cluster EOS âgé de 4 ans pour un cluster cinq fois plus puissant nommé Olympe. Chaque nœud CPU d'Olympe a des processeurs Skylake, Intel® Xeon® Gold 6140@2,30 GHz (voir Tableau B.4)

Il existe deux types de nœuds sur Olympe, certains avec uniquement des processeurs, d'autres hybrides avec des processeurs et une carte graphique (GPU) nommée également carte accélératrice, car l'utilisation de celle-ci ne se limite pas au traitement graphique, mais aussi au HPC (calcul haute performance). Comme c'est le cas ici (voir Tableau B.5). Nous pouvons remarquer que sur Olympe, un calcul accéléré par GPU pourrait être 14 fois plus rapide. Cependant, de la même manière que pour les architectures multicœurs, pour bénéficier de la puissance de calcul des nœuds hybrides CPU/GPU, il faut modifier l'implémentation des codes pour l'adapter à l'architecture GPU. Si nous comparons un nœud CPU d'Olympe et un nœud xeonv5, le meilleur type de nœud CPU de Curie, nous pouvons voir qu'il n'y a pas de différence significative en matière de puissance de calcul entre un nœud de Curie xeonv5 et un nœud Sequana X1120 d'Olympe, mais sur Curie, nous avons 180 fois moins de nœuds xeonv5, qu'Olympe a de nœuds Sequana X1120.

TABLE B.5 : Nœuds d'Olympe (voir Tableau B.4) [44]

Nom	Nombre de nœuds	Processeurs/nœud	GPUs/nœud	Flops/nœud	Mémoire
Sequana X1120	360	2 ^a		2,26 TFlops	192 GiB
Sequana X1125	12	2 ^a	4 ^b	33,8 TFlops	384 GiB

^a Intel® Xeon® Gold 6140

^b NVIDIA® Volta 100 SXM2

TABLE B.6 : Spécifications des CPUs de Joliot-Curie [41, 42, 45]

	Intel® Xeon®		AMD Epyc™
	Platinum8168@2,70 GHz	Phi7250	Rome7H12@2,60 GHz
Cœurs	24	68	64
Threads/cœur (hyper-threading)	2	4	2
Fréquence de base du processeur	2,70 GHz	1,40 GHz	2,6 GHz
Fréquence Turbo maxi	3,70 GHz	1,60 GHz	3,30 GHz
Mémoire cache	33 MiB Cache L3	32 MiB Cache L2	250 MiB Cache L3
AVX	AVX-512 (2)	AVX-512 (2)	AVX2
FMA	FMA3	FMA3	FMA3

TABLE B.7 : Nœuds de Joliot-Curie [46] (voir Tableau B.6)

Nom	Nombre de nœuds	CPUs/nœud	GPUs/nœud	Flops/nœud	Mémoire
SKL Irene	1 656	2 ^a		4,4 TFlops	192 GiB
KNL Irene (Knight Landing)	828	1 ^b		2,4 TFlops	112 GiB
AMD Irene ROME	2 292	2 ^c		5,12 TFlops	256 GiB
Irene V100	32	2	4 ^d	35,3 TFlops	3 TB

^a Intel® Xeon® Platinum 8168@2,70 GHz

^b Intel® Xeon Phi™ 7250

^c AMD Epyc™ Rome@2,60 GHz

^d NVIDIA® V100

Joliot-Curie

Le TGCC Joliot-Curie de Genci-CEA est une autre ressource informatique utilisée pour cette thèse. Ce cluster a deux partitions CPU Intel® une appelée SKL Irene avec un processeur Intel® skylake et l'autre appelée KNL Irene avec un processeur Intel® Knight Landing. Il y a aussi la partition AMD Irene Rome avec un CPU AMD Epyc™ Rome (voir Tableau 1.6). Une partition hybride CPU/GPU appelée Irene V100 est également disponible avec un CPU Intel® Cascade Lake et un GPU NVIDIA® v100.

Si nous comparons Olympe et Joliot-Curie, nous pouvons voir que par exemple un nœud SKL Irene est 2 fois plus puissant qu'un nœud Olympe Sequana X1120 et théoriquement, si nous prenons tous les nœuds de chaque partition alors, nous aurions 9 fois plus de puissance de calcul avec les nœuds SKL Irene qu'avec les nœuds Sequana X1120.

Méthodes de chimie quantique

Dans le chapitre 2, nous fournissons des détails sur les concepts principaux et les méthodes utilisées dans cette thèse.

Équation de Schrödinger

Dans l'équation de Schrödinger, on applique l'opérateur Hamiltonien \hat{H} à la fonction d'onde Ψ . Cette équation a deux formes différentes, dépendante du temps et indépendante du temps. Elles sont utilisées respectivement pour les états non stationnaires et stationnaires. La fonction d'onde dépend du vecteur X qui est un vecteur de coordonnées composites rassemblant les coordonnées de spin s et les coordonnées d'espace R avec p le nombre de particules.

L'expression indépendante du temps est une équation aux valeurs propres ; avec l'énergie E comme valeur propre et avec sa fonction d'onde correspondante comme vecteur propre :

$$\hat{H}\Psi(X) = E\Psi(X). \quad (\text{B.2})$$

Approximation de Born-Oppenheimer

Pour tout électron e et noyau n on a $m_e \ll m_n$, par conséquent on peut supposer que les noyaux sont fixes (si on les compare aux électrons). On peut donc découpler les mouvements des électrons et des noyaux, pour séparer l'opérateur Hamiltonien comme la somme d'un Hamiltonien nucléaire \hat{H}_n et d'un Hamiltonien électronique \hat{H}_e , tel que

$$\hat{H} = \hat{H}_e + \hat{H}_n, \quad (\text{B.3})$$

L'Hamiltonien électronique étant

$$\hat{H}_e = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee}, \quad (\text{B.4})$$

Au sein de l'approximation de Born-Oppenheimer, l'Hamiltonien électronique \hat{H}_e est plus explicitement

$$\hat{H}_e = -\sum_i^N \frac{\nabla_i^2}{2} - \sum_i^N \sum_A^M \frac{Z_A}{r_{Ai}} + \sum_{i<j}^N \frac{1}{r_{ij}}, \quad (\text{B.5})$$

où r_{pq} est la distance entre les particules p et q , N est le nombre d'électrons, M est le nombre de noyaux, A est un noyau, i et j sont deux électrons. Ainsi, l'énergie totale de l'Hamiltonien électronique correspond à la somme de l'énergie cinétique des électrons (\hat{T}_e), de l'énergie associée à l'attraction de Coulomb entre les électrons et les noyaux (\hat{V}_{en}) et de la répulsion de Coulomb entre les électrons (\hat{V}_{ee}).

L'approximation Hartree-Fock

Dans l'approximation Hartree-Fock (HF), la fonction d'onde multiélectronique est définie comme un déterminant de Slater,

$$\Psi_{\text{HF}}(\mathbf{X}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \dots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \dots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}, \quad (\text{B.6})$$

composé de N fonctions d'onde monoélectroniques appelées spinorbitales formant une base orthonormée

$$\psi_i(\mathbf{x}) = \phi_i(\mathbf{r})\sigma(s), \quad (\text{B.7})$$

où $\sigma(s)$ est sa partie de spin

$$\sigma(s) = \begin{cases} \alpha(s), & \text{pour les électrons spin-up,} \\ \beta(s), & \text{pour les électrons spin-down,} \end{cases} \quad (\text{B.8})$$

et $\phi_i(\mathbf{r})$ est sa partie spatiale. On définit l'énergie HF, E_{HF} , comme valeurs propres de l'opérateur Hamiltonien appliqué à Ψ_{HF} :

$$E_{\text{HF}} = \frac{\langle \Psi_{\text{HF}} | \hat{H} | \Psi_{\text{HF}} \rangle}{\langle \Psi_{\text{HF}} | \Psi_{\text{HF}} \rangle}. \quad (\text{B.9})$$

Dans la base de spinorbitale, l'énergie HF est donnée par

$$E_{\text{HF}} = \sum_i^N \langle \psi_i(\mathbf{x}_1) | \hat{H}^c | \psi_i(\mathbf{x}_1) \rangle + \sum_{i<j}^N \left[\langle \psi_i(\mathbf{x}_1)\psi_j(\mathbf{x}_2) | \frac{1}{r_{12}} | \psi_i(\mathbf{x}_1)\psi_j(\mathbf{x}_2) \rangle - \langle \psi_i(\mathbf{x}_1)\psi_j(\mathbf{x}_2) | \frac{1}{r_{12}} | \psi_j(\mathbf{x}_1)\psi_i(\mathbf{x}_2) \rangle \right]. \quad (\text{B.10})$$

$$\hat{\mathcal{F}}(\mathbf{x}_1) = \hat{H}^c(\mathbf{x}_1) + \sum_i^N [\hat{\mathcal{J}}_i(\mathbf{x}_1) - \hat{\mathcal{K}}_i(\mathbf{x}_1)], \quad (\text{B.11})$$

où \hat{H}^c est l'Hamiltonien de cœur défini comme la somme de l'énergie cinétique des électrons et de l'énergie potentielle électrons-noyaux, $\hat{\mathcal{J}}$ est l'opérateur de Coulomb et $\hat{\mathcal{K}}$ est l'opérateur d'échange.

Hartree-Fock restreint

L'approximation de Hartree-Fock restreinte (RHF) qui s'applique aux systèmes à couche fermée (c'est-à-dire, un système où chaque orbitale spatiale est doublement occupée par une paire d'électrons), nous supposons que la partie spatiale des spinorbitales est indépendante du spin des électrons :

$$\hat{F}\phi_i(\mathbf{r}_1) = \varepsilon_i\phi_i(\mathbf{r}_1), \quad (\text{B.12})$$

où, dans ce cas, l'opérateur de Fock devient

$$\hat{F} = \hat{H}^c + \sum_i^{N/2} [2\hat{\mathcal{J}}_i - \hat{\mathcal{K}}_i], \quad (\text{B.13})$$

Méthodes post-Hartree-Fock

Nous définissons l'énergie de corrélation comme la différence entre l'énergie exacte (non relativiste) et l'énergie HF

$$E_c = E - E_{\text{HF}} \quad (\text{B.14})$$

De plus, nous définissons Ψ_0 comme déterminant de référence, qui peut ou ne pas être le déterminant de l'état fondamental HF Ψ_{HF} . L'objectif des méthodes post-HF est de récupérer autant d'énergie de corrélation que possible.

Méthode d'interaction de configuration

Dans les méthodes d'interaction de configuration (CI), on définit la fonction d'onde corrélée comme une combinaison linéaire de déterminants de Slater

$$|\Psi_{\text{CI}}\rangle = c_0\Psi_0 + \sum_i^{\text{occ}} \sum_a^{\text{virt}} c_i^a |\Psi_i^a\rangle + \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} c_{ij}^{ab} |\Psi_{ij}^{ab}\rangle + \sum_{ijk}^{\text{occ}} \sum_{abc}^{\text{virt}} c_{ijk}^{abc} |\Psi_{ijk}^{abc}\rangle + \dots \quad (\text{B.15})$$

où occ est le nombre d'orbitales occupées, virt le nombre d'orbitales virtuelles (c'est-à-dire, inoccupée), et Ψ_i^a correspond à un déterminant simplement excité où un électron de l'orbitale occupée i a été excité vers l'orbitale vacante a . De même, Ψ_{ij}^{ab} est un déterminant doublement excité où les électrons des orbitales occupées i et j ont été excité vers les orbitales vacantes a et b . (La même règle s'applique aux excitations plus élevées.) Ici, i, j, \dots sont des spinorbitales occupées, a, b, \dots désignent des spinorbitales virtuelles (inoccupées) et p, q, r, \dots des spinorbitales quelconque (orthonormées). Si on considère que tous les déterminants excités possibles construits en plaçant N électrons dans $2K$ spinorbitales, on parle d'une interaction de configuration complète (FCI) qui correspond à la solution exacte du système pour un ensemble de fonctions de base donné. Inversement, si on considère uniquement les monoexcitations, on obtient CI avec les simples (CIS), tandis que si on considère toutes les excitations simples et doubles, on obtient CI avec des simples et des doubles (CISD).

Au niveau CI, l'énergie est calculée comme

$$E_{\text{CI}} = \frac{\langle \Psi_{\text{CI}} | \hat{H} | \Psi_{\text{CI}} \rangle}{\langle \Psi_{\text{CI}} | \Psi_{\text{CI}} \rangle}. \quad (\text{B.16})$$

Elle est une limite supérieure de l'énergie exacte du système grâce à la propriété variationnelle de l'équation ci-dessus. Même si elle récupère toute l'énergie de corrélation, la méthode FCI est dans la pratique, très coûteuse et limitée à un petit espace de Hilbert.

Méthodes CI sélectionné

Dans les méthodes CI conventionnelles (voir ci-dessus), on choisit les déterminants à inclure dans l'espace dit variationnel en se basant sur leur degré d'excitation maximum. Cependant, dans ce sous-espace prédéfini des déterminants, il est bien connu que seul un petit nombre d'entre eux contribuent de manière significative à l'énergie totale. Par conséquent, l'idée fondamentale derrière certaines méthodes CI sélectionné (SCI) consiste à sélectionner les déterminants parmi l'ensemble des déterminants (c'est-à-dire, l'espace FCI) sur la base de leur contribution estimée à l'énergie et/ou la fonction d'onde FCI. Bien que la famille de méthodes SCI ait de nombreux membres, dans cette section, nous ne discuterons que de l'algorithme CIPSI (*interaction de configuration à l'aide d'une sélection perturbative effectuée itérativement*) [34] implémenté dans QUANTUM PACKAGE, un logiciel de chimie quantique développé dans notre laboratoire.

Dans le processus itératif CIPSI [34], une correction perturbative de second ordre est utilisée pour sélectionner les déterminants. En pratique, nous commençons par un petit ensemble de déterminants ; Ensuite, nous évaluons à chaque itération les contributions des déterminants appartenant à l'espace dit extérieur à l'aide de la méthode perturbative PT2 et incluons le sous ensemble avec la plus grande contribution dans le calcul CI de l'itération suivante, et ce, jusqu'à la convergence. La somme de l'énergie variationnelle et de l'énergie PT2 est une approximation de l'énergie FCI.

Méthodes du cluster couplé

La famille de méthodes du cluster couplé (CC) [3, 48-51] est l'une des approches de la fonction d'onde les plus intéressantes pour la description des systèmes chimiques [52-54]. En particulier, les méthodes CC tronquées à faible ordre, telles que CC avec des simples, des doubles et des triples perturbatifs CCSD(T) [52, 55], décrivent correctement la faible corrélation, tandis que l'inclusion d'excitation d'ordre supérieur est nécessaire pour les systèmes fortement corrélés.

Dans la théorie CC, on représente la fonction d'onde comme suit

$$\Psi_{\text{CC}} = e^{\hat{T}} \Psi_0, \quad (\text{B.17})$$

où l'opérateur de cluster est

$$\hat{T} = \sum_k^N \hat{T}_k, \quad (\text{B.18})$$

avec

$$\begin{aligned} \hat{T}_1 &= \sum_i^{\text{occ}} \sum_a^{\text{virt}} t_i^a \hat{a}_i \hat{a}_a^\dagger \\ \hat{T}_2 &= \frac{1}{4} \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} t_{ij}^{ab} \hat{a}_i \hat{a}_j \hat{a}_a^\dagger \hat{a}_b^\dagger \\ &\vdots \end{aligned} \quad (\text{B.19})$$

Ici, \hat{a}_a^\dagger et \hat{a}_i sont respectivement les opérateurs habituels d'annihilation et de création qui annihilent un électron dans la spinorbitale occupée i et crée un électron dans la spinorbitale vacante a .

En effectuant une expansion de Taylor de l'opérateur exponentialisé

$$e^{\hat{T}} = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!} = \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \frac{\hat{T}^4}{4!} + \dots \quad (\text{B.20})$$

La fonction d'onde CC peut être réécrite comme

$$\Psi_{\text{CC}} = \Psi_0 \left[\hat{1} + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left(\hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{6} \hat{T}_1^3 \right) + \left(\hat{T}_4 + \hat{T}_3 \hat{T}_1 + \frac{1}{2} \hat{T}_2^2 + \frac{1}{2} \hat{T}_2 \hat{T}_1^2 + \frac{1}{24} \hat{T}_1^4 \right) + \dots \right], \quad (\text{B.21})$$

où $\hat{1}$ est l'opérateur d'identité. Le premier terme d'une parenthèse (de forme $\hat{T}_2, \hat{T}_3, \dots$) génèrent les excitations connectées alors que les autres ($\hat{T}_1^2, \hat{T}_2 \hat{T}_1$) génèrent les excitations déconnectées, c'est-à-dire, les excitations qui sont le produit de deux excitations voir plus, d'ordre inférieur. D'un point de vue physique, un terme connecté tel que \hat{T}_4 correspond à quatre électrons interagissant simultanément. Tandis qu'un terme déconnecté comme \hat{T}_2^2 correspond à deux paires non interagissantes d'électron interagissant.

Nous pouvons maintenant comparer CI et CC, tronquées aux simples et doubles, en partant des équations (B.15) et (B.17) nous avons

$$\text{CI: } \hat{T} = 1 + \hat{T}_1 + \hat{T}_2, \quad (\text{B.22})$$

$$\text{CC: } e^{\hat{T}} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left(\hat{T}_2 \hat{T}_1 + \frac{1}{6} \hat{T}_1^3 \right) + \left(\frac{1}{2} \hat{T}_2^2 + \frac{1}{2} \hat{T}_2 \hat{T}_1^2 + \frac{1}{24} \hat{T}_1^4 \right) + \dots \quad (\text{B.23})$$

nous voyons clairement que l'ansatz CC produit en partie des excitations triples, quadruples et supérieures grâce aux termes déconnectés de la forme $\frac{1}{2} \hat{T}_2^2$ ou $\hat{T}_2 \hat{T}_1$. Ces termes ne sont bien sûr pas présent dans son équivalent CI et constituent la cause de la non-extensivité en taille des méthodes CI tronquées.

Méthodes de la perturbation de Møller-Plesset

Dans la théorie de Rayleigh-Schrödinger, on étend à la fois l'énergie et la fonction d'onde en série de Taylor en fonction du paramètre de couplage λ comme suit

$$\hat{H} = \hat{H}_0 + \lambda \hat{V} \quad (\text{B.24})$$

où \hat{H}_0 est l'Hamiltonien d'ordre zéro (référence) et \hat{V} est la perturbation.

Dans le cas de la théorie de perturbation de Møller-Plesset, on a

$$\hat{H}_0 = \sum_i^N \hat{\mathcal{F}}(i) \quad (\text{B.25})$$

qui donne, dans le cas de la théorie de perturbation de Møller-Plesset du second ordre (MP2), les formules bien connues

$$E^{(2)} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{virt}} \frac{\left| \langle \Psi_0 | \sum_{k < l}^n \frac{1}{r_{kl}} | \Psi_{ij}^{ab} \rangle \right|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} = \frac{1}{4} \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \langle ij || ab \rangle t_{ij}^{ab} \quad (\text{B.26})$$

$$t_{ij}^{ab} = \frac{\langle ij || ab \rangle}{\Delta_{ij}^{ab}} \quad (\text{B.27})$$

$$\Delta_{ij}^{ab} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b \quad (\text{B.28})$$

Méthodes de fonction de Green à plusieurs corps

Dans le chapitre 3 nous étudions les discontinuités présentent dans l'approximation GW, qui est une méthode basée sur la fonction de Green. En effet, les méthodes de la théorie de perturbation

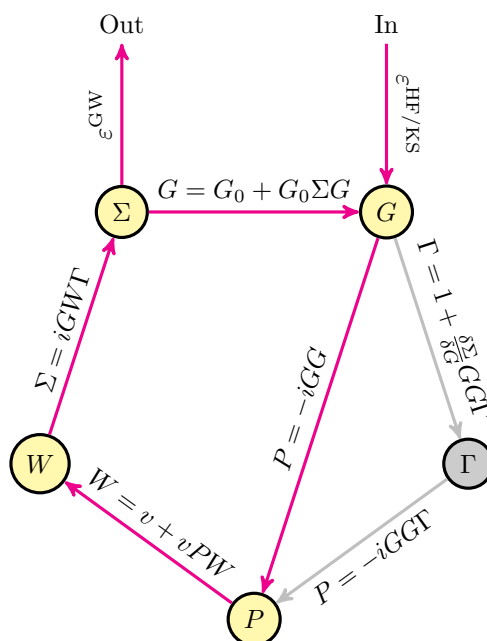


FIGURE B.2 : Le Pentagone d'Hedin [47]. Le chemin rouge montre le processus GW auto-cohérent qui contourne le calcul de la fonction vertex Γ .

à plusieurs corps basées sur la fonction de Green à un corps G sont fascinantes, car elles sont capables de transformer un problème à plusieurs électrons insoluble en un ensemble d'équations non linéaires à un électron, grâce à l'introduction d'un potentiel effectif Σ , la self-energy. La corrélation électronique est explicitement incorporée via une séquence d'étapes auto-cohérente appelée équations d'Hedin [47], qui connecte G , la fonction vertex irréductible Γ , la polarisation irréductible P , l'interaction de Coulomb écrantée dynamiquement W , et Σ à travers un ensemble de cinq équations intégro-différentielles (voir Figure B.2).

En particulier, l'approche d'Hedin utilise une interaction de Coulomb écrantée dynamiquement W au lieu de l'interaction standard de Coulomb v . Des propriétés expérimentales importantes telles que les potentiels d'ionisation, les affinités électroniques ainsi que les fonctions spectrales, qui sont liées aux spectres de photo-émission directe et inverse, peuvent être obtenues directement à partir de la fonction de Green à un corps [57]. Une approximation particulièrement réussie et pratique des équations d'Hedin est ladite approximation GW [57-59] qui contourne le calcul de la partie la plus compliquée des équations d'Hedin, la fonction vertex Γ [47].

Bien que G_0W_0 est probablement la variante de GW la plus simple et la plus largement utilisée [60-65], sa dépendance au point de départ a motivé le développement de plusieurs versions partiellement [66-74] et entièrement [75-83] auto-cohérente afin de réduire ou de supprimer cet effet indésirable. Ici, nous nous sommes concentré sur les méthodes partiellement auto-cohérentes, car elles ont démontré une précision comparable et sont plus légères en calcul que la version entièrement cohérente [84]. De plus, elles sont régulièrement utilisées pour des calculs sur les solides et les molécules et sont disponibles dans divers logiciels de calcul [61, 68, 74, 84-89]. Récemment, un nombre toujours croissant d'applications des méthodes GW partiellement auto-cohérentes sont apparues dans la littérature de physique et de chimie pour les systèmes moléculaires [62-65, 73, 82, 85, 86, 90-96], ainsi que de vastes ensembles de référence élaborés [64, 65, 89, 90, 97-101].

Il existe deux types principaux de méthodes GW partiellement auto-cohérentes : i) *evGW* [66-69], où les énergies de quasiparticules (QP) sont mises à jour à chaque itération, et ii) *qsGW* [70-74], où l'on met à jour les énergies QP et les orbitales correspondantes. Notez qu'une dépendance au point de départ reste dans *evGW* car les orbitales ne sont pas optimisées de manière auto-cohérente dans ce cas.

Dans un article récent [102], tout en étudiant un modèle de système à deux électrons [103-108], nous avons observé que, dans les méthodes GW partiellement auto-cohérentes (telles que *evGW* et

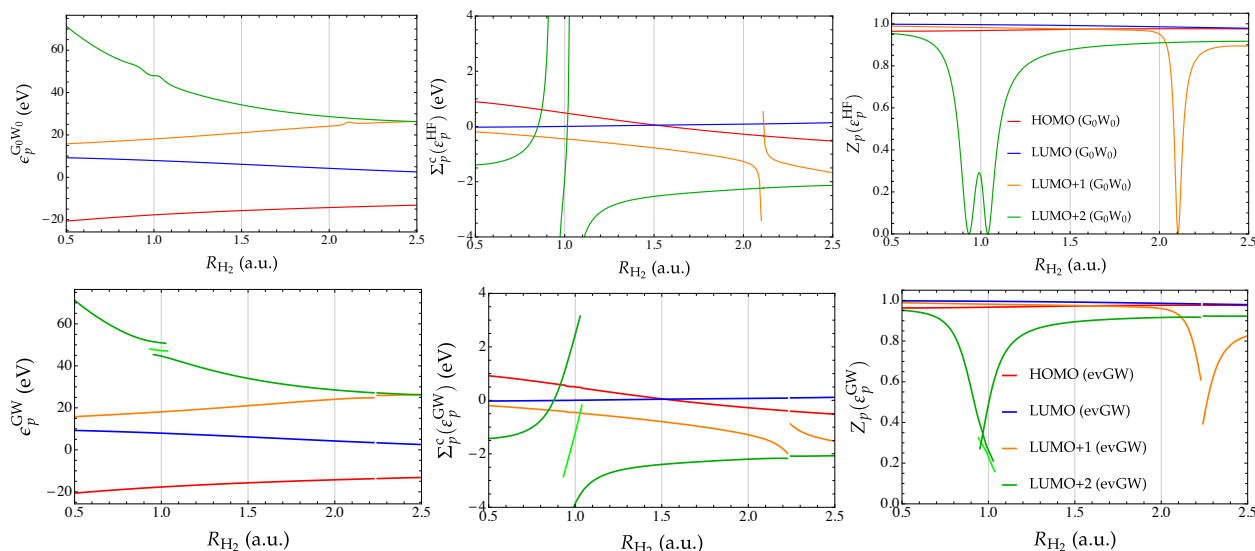


FIGURE B.3 : Énergies QP (à gauche), la partie de corrélation de la self-energy (centre) et le facteur de renormalisation (à droite) en fonction de la distance internucléaire R_{H_2} pour diverses orbitales de H_2 au niveau $G_0W_0@HF/6-31G$ (haut) et $evGW@HF/6-31G$ (bas). Pour plus de commodité, la branche intermédiaire (centre) est présentée en vert clair pour le LUMO+2.

qsGW), on peut observer, dans les régimes faiblement corrélés, des discontinuités (non physique) dans les surfaces d'énergies de plusieurs quantités clés (potentiel d'ionisation, affinité électronique, écart HOMO-LUMO, énergie totale et corrélation, ainsi que les énergies d'excitations verticales). Dans ce chapitre, nous fournissons d'autres preuves et explications de cet effet indésirable dans les systèmes moléculaires. Pour des raisons de simplicité, la présente étude est basée sur de simples molécules diatomiques (H_2 , F_2 et BeO). Cependant, le même phénomène peut être observé dans de nombreux autres systèmes moléculaires, tels que LiF , HeH^+ , LiH , BN , O_3 , etc.

Nous avons trouvé de graves lacunes dans deux variantes largement utilisées de GW dans le régime faiblement corrélé. Nous avons mis en évidence que l'on peut se heurter à des problèmes de solutions multiples dans G_0W_0 et $evGW$ en raison de la localisation de la solution QP près des pôles de la self-energy. Dans G_0W_0 linéarisé, Cela implique des irrégularités dans les principales quantités mesurables expérimentalement, tandis qu'au niveau partiellement auto-cohérent $evGW$, des discontinuités surgissent. Parce que l'énergie de corrélation RPA [86, 120, 143, 144] et les énergies d'excitation Bethe-Salpeter [85, 145, 146] dépendent directement des énergies QP, ces types de discontinuités sont également présents dans ces quantités, et donc dans les surfaces d'énergie des états fondamentaux et excités. De plus, pour les systèmes étendus, ces problèmes peuvent être atténués par les modes de plasmon qui dominent le spectre à haute énergie de l'interaction de Coulomb écrantée. Les résultats de ce travail seront utiles pour les calculs auto-cohérents GW de phénomènes dynamiques, c'est-à-dire, avec le mouvement nucléaire.

La base de données QUEST d'énergies d'excitation verticales

De nos jours, il existe un très grand nombre d'approches de calcul pour la structure électronique, plus ou moins coûteuse en fonction de leur précision globale, capable de prédire quantitativement les énergies absolues et relatives d'états électroniques dans des systèmes moléculaires [1-3, 150]. Un aspect important de certaines de ces méthodes théoriques est leur capacité d'accéder aux énergies des états électroniques excités, [13, 39, 121, 151-158]. La description fidèle des états excités est particulièrement difficile d'un point de vue théorique, mais est essentielle à une compréhension plus profonde des processus photochimiques et photophysiques comme l'absorption, la fluorescence, la phosphorescence et la chimioluminescence [159-165]. Pour un niveau de théorie donné, les

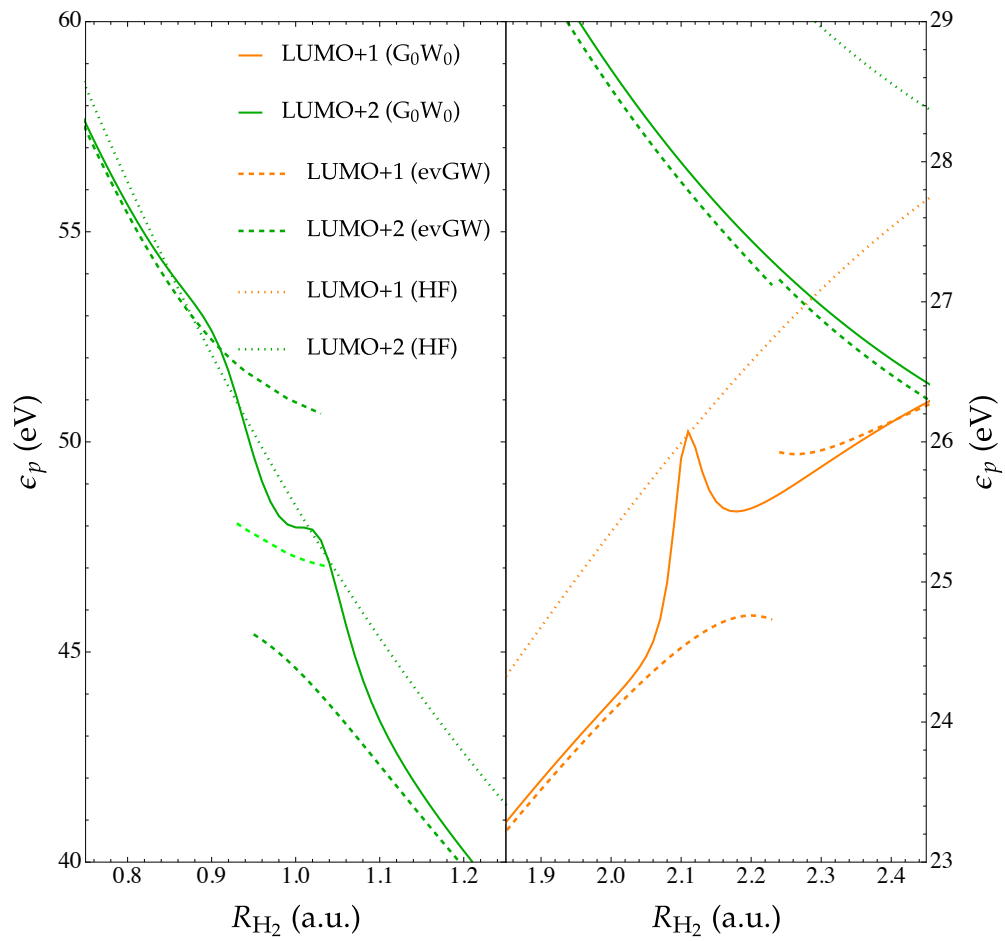


FIGURE B.4 : Energies des orbitales HF (lignes en pointillé) et énergies QP en fonction de la distance internucléaire R_{H_2} pour les orbitales LUMO+1 et LUMO+2 de H_2 aux niveaux $G_0W_0@HF/6-31G$ (lignes solides) et $evGW@HF/6-31G$ (lignes en pointillé). Pour plus de commodité, la branche intermédiaire (centre) est présentée en vert clair pour le LUMO+2.

méthodes d'état fondamental sont généralement plus précises que leurs analogues d'état excité. Les raisons derrière cela sont (au moins) triples : i) modéliser avec précision la structure électronique des états excités nécessite généralement de plus grosses bases monoélectroniques (y compris des fonctions diffuses la plupart des temps) que leur homologue de l'état fondamental, ii) les états excités peuvent être régis par différentes quantités de corrélations dynamiques/statiques, présentant des natures physiques très différentes ($\pi \rightarrow \pi^*$, $n \rightarrow \pi^*$, transfert de charges, double excitation, valence, Rydberg, singulet, doublet, triplet, etc), pourtant très proche en énergie les uns des autres, et iii) il faut généralement compter sur le formalisme de la théorie de la réponse [26, 166-172], qui introduit intrinsèquement un « biais » sur l'état fondamental. Par conséquent, concevoir des méthodes d'état excité capables de cibler simultanément et de la même façon, tous ces types d'états excités à un coût abordable reste un défi ouvert en chimie théorique computationnelle, comme en témoigne le grand nombre d'articles sur ce sujet [13, 39, 121, 151-158, 173].

Lors de la conception d'un nouveau modèle théorique, le premier aspect que l'on pourrait vouloir tester est sa précision globale, c'est-à-dire sa capacité à reproduire des valeurs de référence pour un système donné avec des paramètres bien définis et une même configuration (même géométrie, base, etc). Ces valeurs peuvent être des énergies absolues et/ou relatives, des paramètres géométriques, des propriétés spectroscopiques physiques ou chimiques extraites d'expériences, de calculs théoriques de haut niveau, ou de toute combinaison de ceux-ci. À cette fin, la communauté de structure électronique a conçu au fil des années des données de référence, c'est-à-dire des séries de molécules pour lesquelles on peut calculer avec une grande précision des estimations théoriques et/ou accéder à des données expérimentales solides pour des propriétés données. En ce qui concerne les propriétés de l'état fondamental, deux des ensembles les plus anciens et les plus utilisés sont probablement les ensembles de référence Gaussian-1 et Gaussian-2 [174-176] développés par le groupe de Pople dans les années 1990. Par exemple, l'ensemble Gaussian-2 rassemble des énergies d'atomisation, des énergies d'ionisation, des affinités électroniques, des affinités de proton, des énergies de dissociation et des barrières de réaction. Cet ensemble a ensuite été étendu et raffiné [177, 178]. Un autre ensemble très utile pour la conception de méthodes capables de cibler les effets de dispersion [179] est l'ensemble de référence S22 [180] (et sa version étendue S66 [181]) de Hobza et ses collaborateurs qui fournissent des énergies d'interaction de référence pour des systèmes interagissant faiblement (non covalents). On pourrait aussi mentionner l'ensemble GW100 [64, 182, 183] (et son extension GW5000 [184]) d'énergies d'ionisation qui a contribué énormément à solidifier l'implémentation des méthodes de type GW pour les systèmes moléculaires [61, 84, 86, 185]. L'ensemble *ab initio* de données thermochimique (HEAT) conçu pour atteindre une grande précision sur les enthalpies de formation d'atomes et de petites molécules (sans données expérimentales) est un autre exemple d'un ensemble de référence [186-188]. Plus récemment, mentionnons les données de référence du *Simons Collaboration on the Many-Electron Problem* qui fournit, par exemple, des énergies d'état fondamental très précises pour des chaînes d'hydrogène [189] ainsi que les atomes de métaux de transitions, leurs ions et leurs monoxydes [190]. Mentionnons également l'ensemble de Zhao et Truhlar pour les petits complexes de métaux de transition employés pour comparer la précision des méthodes de la fonctionnelle de la densité (DFT) [134] pour la chimie de métaux de transition *3d* [191], et enfin les bases de données populaires GMTKN24 [192], GMTKN30 [193, 194] et GMTKN55 [195] pour la thermochimie, cinétique, et les interactions non covalentes développées par Goerigk, Grimme et leurs collègues.

Les exemples d'ensembles de référence présentés ci-dessus sont tous conçus pour des propriétés à l'état fondamental, et il existe des protocoles spécifiques adaptés pour modéliser avec précision les énergies d'états excités et leurs propriétés. En effet, les jeux de données de référence des énergies des états excités et/ou de leurs propriétés sont moins nombreux que leurs homologues de l'état fondamental, mais leur nombre a augmenté à un rythme constant au cours des dernières années. Ci-dessous, nous vous proposons une brève description de certains d'entre eux. L'un des exemples les plus caractéristiques est l'ensemble de référence des énergies d'excitation verticales proposé par Thiel et ses collaborateurs [196-199]. Cet ensemble d'énergies d'excitation regroupe un grand nombre d'énergies d'excitation déterminé dans 28 molécules organiques moyennes CNOH avec un

total de 223 états excités de valence (152 singulets et 71 états triplets) pour lesquels les meilleures estimations théoriques (TBE) ont été défini. Dans leur première étude, Thiel et ses collaborateurs ont effectué des calculs CC2 [25, 200], CCSD [23, 168, 201, 202], CC3 [26, 203], et CASPT2 [10, 11, 13, 204] (avec la base TZVP) sur des géométries MP2/6-31G(d) afin de fournir (à partir de données de la littérature de haute qualité) des TBE pour ces transitions. Ces TBE ont été rapidement affiné avec la plus grande base *aug-cc-pVTZ* [198, 199]. Dans le même esprit, il convient également de mentionner l'ensemble de transitions verticales de Gordon (basées sur des valeurs expérimentales) [205] utilisé pour comparer les performances de la TD-DFT [15, 16, 206, 207], ainsi comme sa version étendue par Goerigk et ses collègues qui ont décidé de remplacer les valeurs de référence expérimentales par des énergies d'excitation CC3 [208-210]. Pour les comparaisons avec les valeurs expérimentales, il existe aussi différents ensembles de mesure utilisés dans les énergies 0-0, notamment par le groupe de Furche [211, 212], par celui de Hättig [213], par notre groupe [5, 102, 214] pour des composés en phase gazeuse, par le groupe de Grimme [215, 216] et de Jacquemin [217, 218] pour les colorants solvatés. Mentionnons également la nouvelle série d'états excités de transfert de charge par Szalay et ses collaborateurs [référence basée sur la méthode du cluster couplé avec équation de mouvement (EOM-CC)] [219] ainsi que l'ensemble Gagliardi-Truhlar utilisé pour comparer la précision de la théorie fonctionnelle multiconfiguration pair-density [157] contre la méthode CASPT2 bien établie [220].

Avec une philosophie similaire et une exigence de précision chimique, nous avons récemment rapporté dans plusieurs études des excitations verticales extrêmement précises pour des molécules de petite et moyenne taille [30, 32, 36, 39, 221]. Ces valeurs composent la base de données d'excitation verticales QUEST que nous décrivons en détail dans le chapitre 4 qui est composé de 5 sous-ensembles (voir Figure B.5) : i) un sous-ensemble d'excitation dans de petites molécules contenant de 1 à 3 atomes non-hydrogène appelés QUEST#1, ii) un sous-ensemble de doubles excitations dans des molécules de petites et moyennes tailles connues sous le nom de QUEST#2, iii) un sous-ensemble d'énergies d'excitation pour des molécules de taille moyenne contenant de 4 à 6 atomes non hydrogènes appelés QUEST#3, iv) un sous-ensemble composé de molécules plus « exotiques » et des radicaux appelé QUEST#4 et v) un sous-ensemble appelé QUEST#5, spécialement conçu pour cette thèse, rassemblant des énergies d'excitation sur de plus grandes molécules ainsi que des molécules plus petites supplémentaires. L'un des principaux aspects de la base de données QUEST est qu'elle ne repose pas sur des valeurs expérimentales, en évitant les biais potentiels intrinsèquement liés aux expériences et faciliter les comparaisons entre méthodes théoriques. De plus, notre protocole a été conçu pour être aussi uniforme que possible, ce qui signifie que nous avons conçu une procédure très systématique pour tous les états excités afin de faire des comparaisons croisées aussi facilement que possible.

L'ensemble de données QUEST a la particularité de se fonder dans une large mesure sur les énergies d'excitation de référence SCI ainsi que de méthodes à réponse linéaires (LR) CC d'ordre élevé tels que LR-CCSDT et LR-CCSDTQ [24, 168, 169, 171, 225-230]. Récemment, les méthodes SCI ont été d'une grande importance pour le calcul des énergies très précises dans les petites et moyennes molécules, car SCI donne des énergies d'une qualité proche du FCI pour seulement une très petite fraction du coût de calcul d'un véritable calcul FCI [30, 32, 36-39, 221, 231-250]. En raison de l'idée assez naturelle sous-jacente de ces méthodes, la famille SCI est composée de nombreux membres [34, 36-38, 240, 251-276]. Leur philosophie fondamentale consiste, grosso modo, à ne conserver que les déterminants les plus pertinents de l'espace FCI selon un critère donné pour ralentir l'augmentation exponentielle de la taille de l'expansion du CI. Développé à l'origine à la fin des années 1960 par Bender et Davidson [251] ainsi que Whitten et Hackmeyer [252], de nouveaux algorithmes SCI efficaces ont refait surface récemment. Trois exemples sont iCI [272, 277-279], heat-bath semistochastique CI (SHCI) [37, 240-242, 262, 263] et *Interaction de configuration à l'aide d'une sélection perturbative effectuée itérativement* (CIPSI) [34, 35, 256, 258]. Cette méthode SCI comprend une correction perturbative de second ordre (PT2) qui est la clé pour estimer la « distance » à la solution FCI. Les calculs SCI effectués pour l'ensemble de QUEST des énergies d'excitation repose sur l'algorithme CIPSI, qui est, d'un point de vue historique, l'un des plus anciens algorithmes SCI. Il a été développé en 1973 par Huron, Rancurel et Malrieu [34]. Récemment, l'algorithme CIPSI a été

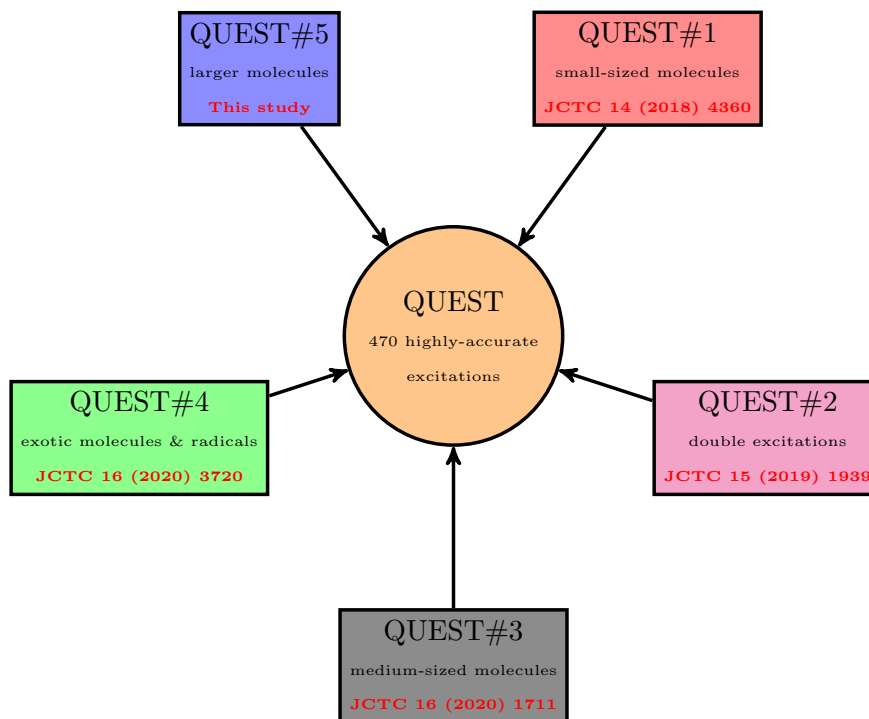


FIGURE B.5 : Composition de chacun des cinq sous-ensembles composant le présent jeu de données d'énergies d'excitation verticales très précises actuelles du projet QUEST.

efficacement implémenté [35] dans l'environnement de programmation open-source QUANTUM PACKAGE par le groupe de Toulouse permettant d'effectuer des calculs massivement parallèles [35, 249, 259, 266]. CIPSI est aussi fréquemment utilisé pour fournir des fonctions d'ondes d'essai précises pour le calcul Monte Carlo quantique dans les molécules [256-258, 260, 261, 264, 265, 285-290] et plus récemment pour les solides périodiques [291].

Nous avons étendu la base de données QUEST d'énergies d'excitation très précises pour les systèmes moléculaires [30, 32, 36, 39, 221] que nous avons commencé à construire en 2018 et qui est maintenant composée de plus de 500 excitations verticales, dont beaucoup peuvent être raisonnablement considérés comme à moins de 1 kcal mol^{-1} de la limite FCI pour la géométrie CC3/*aug-cc-pVTZ* et la base (*aug-cc-pVTZ*) considérées. Nous avons créé un site web spécialement conçu pour rassembler toutes les données générées au cours de ces dernières années. Un aspect clé de ce site web est la possibilité de générer des statistiques correspondant aux besoins de chaque utilisateur spécifique. Il est aussi possible d'importer de nouvelles données pour par exemple comparer de nouvelles méthodes aux méthodes déjà étudiées dans le cadre de ce projet.

En paraphrasant les conclusions de Thiel [196], nous espérons que non seulement la base de données QUEST sera utilisée pour d'autres analyses et tests, mais que d'autres groupes de recherche l'amélioreront également, fournissant non seulement des corrections. (inévitables dans un ensemble de données aussi vaste), mais surtout des extensions avec des estimations améliorées pour certains composés et états, ou de nouvelles molécules.

Démo de Quantum Package dans un navigateur web

QUANTUM PACKAGE [35] est un programme de chimie quantique open source développé principalement dans notre groupe, au Laboratoire de Chimie Théorique (LCT) à l'université de la Sorbonne et par Anouar Benali et ses collègues au Argonne National Lab [291]. Il est également utilisé par plusieurs groupes et chercheurs dans le monde, comme Hugh Burton d'Oxford [398], Joshua Hollett de Winnipeg au Canada [399], Eric Neuscamman de Berkeley [400] ou Claudia Filippi de l'Université de Twente aux Pays-Bas [289].

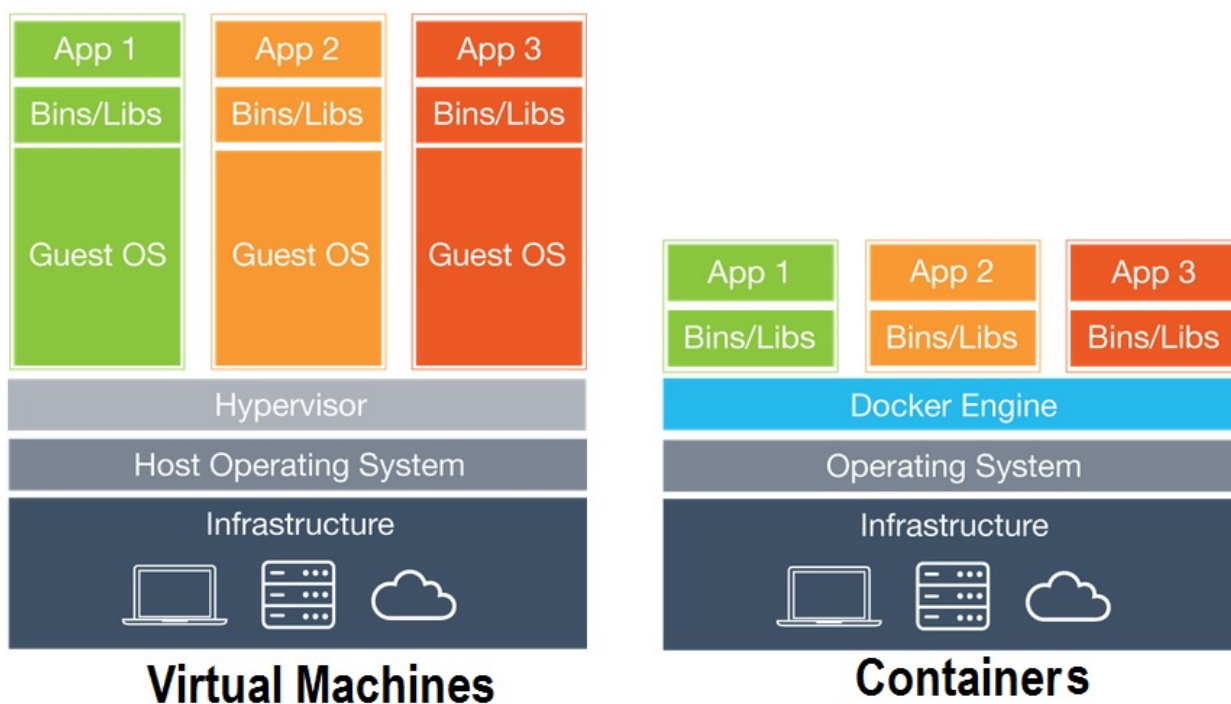


FIGURE B.6 : Conteneur DOCKER comparé à la technologie de virtualisation [403]

L'idée principale conduisant à la création de QUANTUM PACKAGE était de fournir une plateforme pour une implémentation et une expérimentation facile de nouvelles méthodes *ab initio*, tout en faisant du calcul parallèle aussi facilement et efficacement que possible. Suivant cette philosophie, il contient une bibliothèque facile à utiliser pour les développeurs (avec un système de plug-in), son propre shell QUANTUM PACKAGE SHELL (QPSH) et le préprocesseur Fortran IRPF90. Cependant, même si QUANTUM PACKAGE est facile à utiliser, il faut un certain temps pour l'installer et cela peut limiter sa démocratisation dans notre communauté. Pour permettre aux chimistes quantiques de tester QUANTUM PACKAGE rapidement et sans le fardeau de son installation, nous avons décidé de créer une démo qui peut être directement accessible via un navigateur web. Cela permet à quiconque d'exécuter simplement et sans aucun effort de configuration et d'installation QUANTUM PACKAGE à partir d'une page web. Le but du chapitre 5 est d'expliquer comment nous avons intégré la démo de QUANTUM PACKAGE dans son site web.

Quand nous avons imaginé le projet les spécifications étaient les suivantes :

- Un terminal interactif avec toutes les commandes standard d'UNIX et de QUANTUM PACKAGE
- Un nombre limité d'utilisateurs
- Un environnement non persistant, isolé et identique pour chaque utilisateur
- Une limite de temps fixe

Pour respecter ces spécifications, nous avons utilisé les deux outils décrits ci-dessous.

Le premier outil est SHELLINABOX, qui permet de créer un émulateur de terminal intégré à une page web, mais l'outil probablement le plus intéressant que nous avons utilisé est DOCKER [402], un logiciel bien connu dans la communauté des administrateurs système. Il ressemble à un logiciel de virtualisation conventionnel, mais ce n'est pas vraiment le cas. Il s'agit d'un logiciel de conteneurisation. L'hôte partage son noyau avec les conteneurs, ce qui améliore considérablement les performances par rapport aux technologies de virtualisation conventionnelles, car il n'y a pas de système d'exploitation invité ni d'hyperviseur (voir Figure B.6) et pas de matériel virtualisé. De plus, cette technologie de conteneurisation est plus avancée par rapport à l'isolement proposé par un

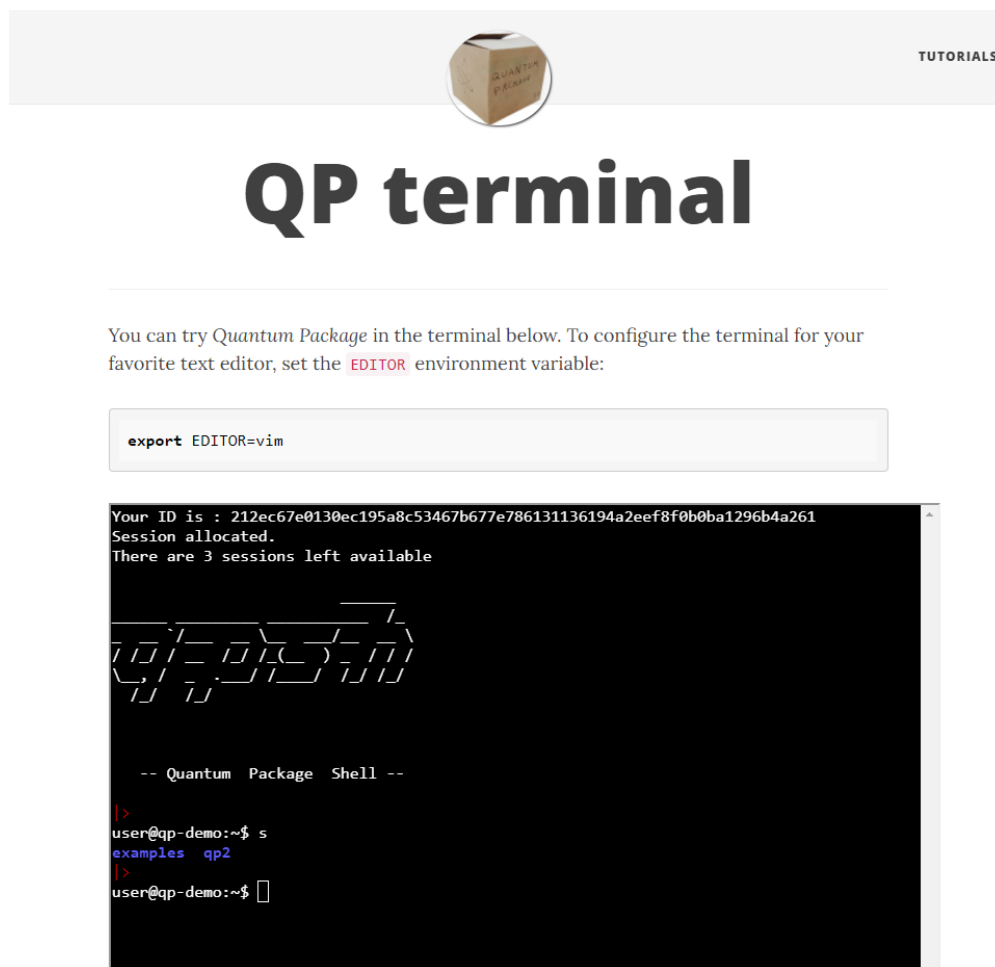


FIGURE B.7 : Capture d'écran du terminal donnant accès à la démo de QUANTUM PACKAGE directement intégré sur le site internet de celui-ci

simple chroot. Pour créer un conteneur, DOCKER utilise une bibliothèque appelée LIBCONTAINER maintenant incluse dans RUNC depuis que cet outil est géré par la *Open Container Initiative* [404]. Cette bibliothèque est responsable du cycle de vie des conteneurs et utilise des composants du noyau LINUX pour les créer.

De plus les conteneurs sont basés sur des images pré-construites et sont des environnements éphémères ce qui est parfait dans le cadre de la démo. Le cycle de vie de celle-ci est géré par un script BASH lancé par SHELLINABOX sur la machine distante. On peut tester QUANTUM PACKAGE sur le site officiel <https://quantumpackage.github.io/qp2/page/try>. On peut également télécharger l'image DOCKER prête à l'emploi utilisée pour la démonstration sur le Docker Hub https://hub.docker.com/r/mveril/qp_demo.

Estimations précises de l'énergie de corrélation d'interaction de configuration complète pour des cycles à cinq et six chaînons

La théorie de la structure électronique repose fortement sur des approximations [1-3]. De façon générale, pour qu'une méthode soit utilisable en pratique, trois approximations principales sont généralement appliquées. La première approximation fondamentale, connue sous le nom d'approximation de Born-Oppenheimer, consiste à supposer que le mouvement des noyaux et des électrons sont découplés [410]. Les coordonnées des noyaux peuvent ensuite être traitées comme paramètres de l'Hamiltonien électronique. La deuxième approximation centrale qui rend les calculs réalisables est l'introduction d'une base dans laquelle l'on introduit un ensemble de fonctions

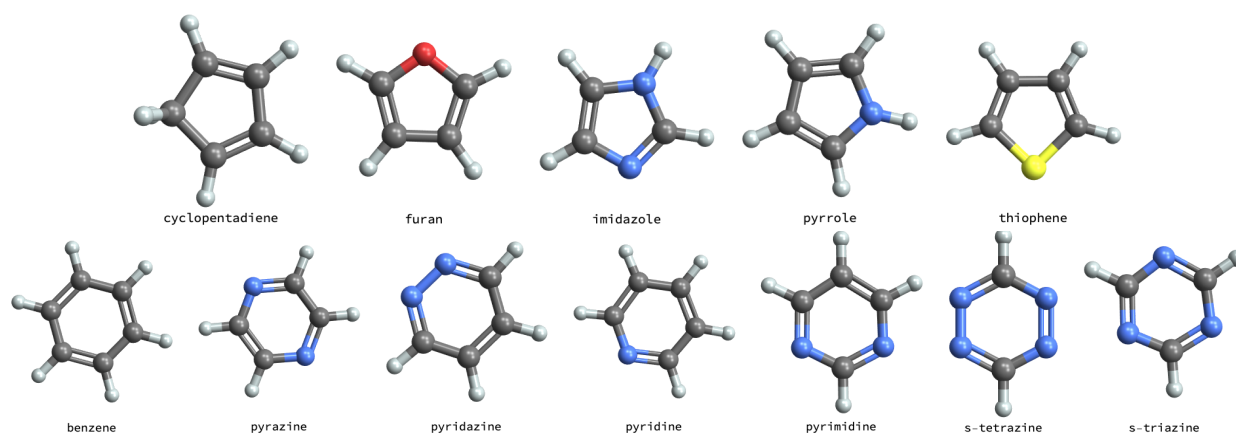


FIGURE B.8 : Les cycles à cinq atomes (en haut) et les cycles à six atomes (en bas) considérés dans cette étude.

de base prédéfinies pour représenter la fonction d'onde à plusieurs électrons du système. Dans la plupart des calculs moléculaires, un ensemble de fonctions de base gaussiennes centrées sur l'atome est introduite pour représenter les orbitales moléculaires monoélectroniques qui sont ensuite utilisées pour construire le(s) déterminant(s) de Slater(s) multiélectroniques. La troisième et la plus pertinente approximation dans le contexte actuel est l'ansatz (ou la forme) de la fonction d'onde électronique Ψ .

La troncation de \hat{T} permet de définir une hiérarchie de méthodes non variationnelles et extensive en taille avec des niveaux de précision croissants : CCSD [48, 52], CCSDT [53, 225], CCSDTQ [412, 413], avec des scaling formels correspondants de $\mathcal{O}(N^6)$, $\mathcal{O}(N^8)$, et $\mathcal{O}(N^{10})$, respectivement. Parallèle à la série CC « complète » la famille de modèles CC itératifs approximatifs CC [169] ignore les termes les plus chers et évite le stockage des amplitudes d'excitation supérieure : CC2 [25], CC3 [26, 203] et CC4 [414, 415]. Ces méthodes itératives sont d'ordre $\mathcal{O}(N^5)$, $\mathcal{O}(N^7)$, et $\mathcal{O}(N^9)$, respectivement, et peuvent être considérées comme des approximations moins chères de CCSD, CCSDT et CCSDTQ. Les méthodes de cluster couplé ont données beaucoup de résultats très satisfaisants pour le calcul des propriétés de l'état fondamental et excité pour les molécules de petite et moyenne taille. [228, 350, 394, 416-419]

Une stratégie de troncation systématique similaire peut être appliquée aux méthodes CI conduisant à la famille de méthodes bien connue appelée CISD, CISDT, CISDTQ, ... où on augmente systématiquement le degré d'excitation maximal des déterminants pris en compte. Sauf pour le CI complet (FCI) où tous les déterminants de l'espace de Hilbert (c'est-à-dire, avec degré d'excitation jusqu'à N) sont considérés, les méthodes CI tronquées sont variationnelles, mais ne sont pas consistantes en taille. La non-variationalité des méthodes CC tronquées étant, dans la pratique, moins un problème que l'inconsistance en taille des méthodes CI tronquées, ces méthodes ont naturellement éclipsé leur équivalent CI dans le domaine de la structure électronique. Cependant, une stratégie différente est récemment revenue sur le devant de la scène dans le cadre des méthodes CI [34, 251-254, 280-284, 420, 421]. En effet, certaines méthodes CI sélectionné (SCI) [231, 256, 258, 261, 262, 266-268, 272, 274-276], où l'on choisit itérativement les déterminants importants de l'espace FCI (généralement) basés sur un critère perturbatif, ont récemment eu beaucoup de succès afin de produire des énergies de référence pour les états fondamentaux et excités de petites et moyennes molécules [30, 32, 35-38, 149, 190, 221, 240-242, 249, 257, 260, 264, 265, 286, 397, 422] grâce à des algorithmes déterministes, stochastiques ou hybrides efficaces bien adaptés à la parallélisation massive. Comme dit précédemment les méthodes SCI sont basées sur un fait bien connu : parmi le très grand nombre de déterminants contenus dans l'espace FCI, seule une infime fraction de ceux-ci contribue de manière significative à l'énergie. Par conséquent, la famille de méthodes SCI+PT2 effectue une exploration de l'espace FCI en sélectionnant les déterminants de manière itérative uniquement les déterminants les plus énergiquement pertinents de l'espace variationnel et en le complétant avec une correction perturbative de second ordre (PT2) [34, 35, 259, 263, 266]. Bien

que le coût de tels algorithmes reste exponentielle, le préfacteur est considérablement réduit, ce qui explique leur attractivité actuelle dans la communauté de la structure électronique grâce à leur applicabilité beaucoup plus large que la méthode FCI standard. Notons que, très récemment, plusieurs groupes [424-426] ont couplé les méthodes CC et SCI via la méthodologie CC corrigée extérieurement [427], montrant des performances prometteuses pour des systèmes faiblement et fortement corrélés.

Une stratégie assez différente pour atteindre la limite FCI est de recourir à la théorie de la perturbation de Møller-Plesset (MP) [428], dont la popularité provient de sa nature « boîte noire », de sa nature extensive en taille et d'une exigence de calcul relativement faible, la rendant facilement appliquée à une large gamme de systèmes moléculaires. La méthode Møller-Plesset de second ordre (MP2) [428] [qui est d'ordre $\mathcal{O}(N^5)$] a été largement adoptée en chimie quantique depuis plusieurs décennies. Ses variantes d'ordre supérieur [MP3 [431], MP4 [432], MP5 [433], et MP6 [434, 435] qui sont d'ordre $\mathcal{O}(N^6)$, $\mathcal{O}(N^7)$, $\mathcal{O}(N^8)$, et $\mathcal{O}(N^9)$ respectivement] ont été étudiées beaucoup plus rarement. Cependant, il est maintenant largement reconnu que la série d'approximations de MP pourrait montrer un comportement erratique, convergent lentement ou divergent ce qui limite son applicabilité et son amélioration systématique [429, 436-445]. Encore une fois, la théorie de la perturbation de MP et les méthodes CC peuvent être couplées. L'exemple le plus emblématique d'un tel couplage, est la méthode CCSD(T) [55], qui comprend de manière itérative les excitations simples et doubles et perturbativement (au niveau MP4 et partiellement MP5) les triples excitations, conduisant à ce qu'on appelle « l'étalon-or » de la chimie quantique pour les systèmes faiblement corrélés grâce à son excellent ratio précision/coût.

Motivé par la récente étude d'Eriksen et ses collaborateurs [422] qui montre les performances d'un grand nombre de méthodes de structure électronique émergentes sur l'énergie de corrélation non relativiste de la molécule de benzène dans la base cc-pVDZ, certains d'entre nous ont récemment étudié la performance de la méthode CIPSI [34, 35, 256, 258, 266] sur le même système [249]. Dans la continuité de ce travail récent, nous rapportons ici une large extension en estimant avec précision l'énergie de corrélation FCI/cc-pVDZ de douze molécules cycliques (cyclopentadiène, furanne, imidazole, pyrrole, thiophène, benzène, pyrazine, pyridazine, pyridine, pyrimidine, s-tétrazine et s-triazine) avec l'aide de CIPSI utilisant des orbitales optimisées énergiquement au même niveau de théorie [241, 452]. Ces systèmes sont représentés dans la Figure B.8. Cet ensemble de systèmes moléculaires correspond à des espaces de Hilbert avec des tailles allant de 10^{29} à 10^{36} . En plus de CIPSI, les propriétés de performance et de convergence de plusieurs séries de méthodes sont étudiées. En particulier, nous étudions i) la série perturbative de MP jusqu'au cinquième ordre (MP2, MP3, MP4 et MP5), ii) les séries approximatives CC2, CC3 et CC4, et iii) la série CC « complète » jusqu'à quadruples (c'est-à-dire CCSD, CCSDT et CCSDTQ). Les performances de CCSD(T) ainsi que du modèle CC entièrement renormalisé, CR-CC(2,3) [151, 453-456], sont également étudiées.

Utilisant l'algorithme CIPSI, nous avons produit des énergies de corrélation de qualité FCI pour douze molécules cycliques (voir Figure B.8) dans la base cc-pVDZ. Ces estimations, probablement précises à quelques dixièmes de millihartree, ont été obtenues en extrapolant les énergies CIPSI à la limite FCI sur la base d'un ensemble d'orbitales obtenue en minimisant l'énergie variationnelle CIPSI. En utilisant des orbitales optimisées énergiquement, on a pu réduire la taille de l'espace variationnel par un ordre de grandeur pour la même énergie variationnelle par rapport aux orbitales naturelles.

Grâce à ces énergies de référence FCI, nous avons ensuite comparé trois familles de méthodes de structure électronique populaires : i) la série perturbative de MP jusqu'au cinquième ordre (MP2, MP3, MP4 et MP5), ii) la série CC approximative CC2, CC3 et CC4, et iii) la série CC « entière » CCSD, CCSDT, et CCSDTQ.

Conclusion générale

Bien que les méthodes GW sont largement utilisées pour le calcul des transitions électroniques et

des intervalles de bande (également en combinaison avec le formalisme d'équation Bethe-Salpeter), nous avons montré que, dans certains cas, on peut observer des discontinuités dans les surfaces d'énergie de plusieurs quantités clés. Des travaux encourageants permettent de constater que certaines solutions de contournement sont possibles comme l'utilisation de la méthode GW auto-cohérente (scGW) qui est malheureusement très lourde numériquement [478] ou COHSEX qui élimine les discontinuités en utilisant un algorithme indépendant de la fréquence [479]. Nous avons également lancé un long travail pour fournir des énergies de transition verticales très précises et les utiliser pour comparer de nombreuses méthodes de chimie quantique de bas niveau. En suivant la même philosophie que GW100 [185] nous avons créé un site web pour recueillir toutes les données du projet QUEST. Nous avons déjà commencé à travailler pour compléter cette base de données avec des données supplémentaires. Un exemple est le travail de mes collègues sur des excitations de type transfert de charges intramoléculaire [397]. De plus, notre travail sur les orbitales optimisées [409] pourrait être une clé pour améliorer les données de référence du projet QUEST. Nous pourrions également reproduire ce que nous avons fait avec le projet QUEST, mais pour les énergies de corrélation de l'état fondamental, les énergies d'atomisation et/ou les énergies d'ionisation et les affinités électroniques. Nous prévoyons également l'application de la méthode d'orbitale optimisée, mais cette fois sur les composés de métaux de transition. Un autre travail intéressant qui a également été fait est la démo de QUANTUM PACKAGE sur le web qui permet aux chimistes quantiques de tester QUANTUM PACKAGE directement sur le Web. Ce travail a conduit à la création d'une machine virtuelle contenant un serveur de terminal web et une image DOCKER utilisée comme environnement de démonstration. Une amélioration de cette architecture pourrait être mise en œuvre à l'avenir à l'aide d'un conteneur « au niveau système » pour remplacer la machine virtuelle afin d'améliorer la performance. De plus, l'image DOCKER utilisée pour la démo peut être employée pour permettre l'utilisation de QUANTUM PACKAGE sur des ordinateurs de bureau WINDOWS ou MACOS ou bien d'éviter le long processus d'installation de QUANTUM PACKAGE. L'image DOCKER de QUANTUM PACKAGE pourrait également être utilisée sur des machines HPC à l'aide de SINGULARITY, mais nous devons étudier la compatibilité de cette image avec ce logiciel.

Bibliography

- ¹A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory* (Dover Publications, Mineola, N.Y, 1996).
- ²T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular electronic-structure theory* (John Wiley & Sons, Inc., 2013).
- ³F. Jensen, *Introduction to computational chemistry*, Third edition (John Wiley & Sons, Chichester, UK ; Hoboken, NJ, 2017).
- ⁴A. M. Turing, "On computable numbers, with an application to the entscheidungsproblem", *Proceedings of the London Mathematical Society* **s2-42**, 230–265 (1937).
- ⁵P.-F. Loos and D. Jacquemin, "Evaluating 0-0 energies with theoretical tools: a short review", *ChemPhotoChem* **3**, 684–696 (2019).
- ⁶J. E. D. Bene, R. Ditchfield, and J. A. Pople, "Self-consistent molecular orbital methods. x. molecular orbital studies of excited states with minimal and extended basis sets", *The Journal of Chemical Physics* **55**, 2236–2241 (1971).
- ⁷M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, "A doubles correction to electronic excited states from configuration interaction in the space of single substitutions", *Chemical Physics Letters* **219**, 21–29 (1994).
- ⁸N. Ishikawa and M. Head-Gordon, "Analytical gradient of the CIS(D) perturbative correction to single-excitation configuration interaction excited states", *International Journal of Quantum Chemistry* **56**, 421–427 (1995).
- ⁹B. O. Roos, P. R. Taylor, and P. E. Sigbahn, "A complete active space scf method (casscf) using a density matrix formulated super-ci approach", *Chemical Physics* **48**, 157–173 (1980).
- ¹⁰K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, "Second-order perturbation theory with a CASSCF reference function", *J. Phys. Chem.* **94**, 5483–5488 (1990).
- ¹¹K. Andersson, P.-A. Malmqvist, and B. O. Roos, "Second-order perturbation theory with a complete active space self-consistent field reference function", *J. Chem. Phys.* **96**, 1218–1226 (1992).
- ¹²P. Celani and H.-J. Werner, "Analytical energy gradients for internally contracted second-order multireference perturbation theory", *The Journal of Chemical Physics* **119**, 5044–5057 (2003).
- ¹³B. O. Roos, K. Andersson, M. P. Fulscher, P.-A. Malmqvist, and L. Serrano-Andrés, "Multiconfigurational perturbation theory: applications in electronic spectroscopy", in , Vol. XCIII, edited by I. Prigogine and S. A. Rice, *Adv. Chem. Phys.* (Wiley, New York, 1996), pp. 219–331.
- ¹⁴C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, "Introduction of n - electron valence states for multireference perturbation theory", *J. Chem. Phys.* **114**, 10252–10264 (2001).
- ¹⁵E. Runge and E. K. U. Gross, "Density-functional theory for time-dependent systems", *Phys. Rev. Lett.* **52**, 997–1000 (1984).
- ¹⁶M. E. Casida, "Time-dependent density functional response theory for molecules", in , edited by D. P. Chong, *Recent Advances in Density Functional Methods* (World Scientific, Singapore, 1995), pp. 155–192.

- ¹⁷D. J. Tozer, R. D. Amos, N. C. Handy, B. O. Roos, and L. Serrano-Andres, "Does density functional theory contribute to the understanding of excited states of unsaturated organic compounds?", *Mol. Phys.* **97**, 859–868 (1999).
- ¹⁸A. Dreuw and M. Head-Gordon, "Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zinbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes", *J. Am. Chem. Soc.* **126**, 4007–4016 (2004).
- ¹⁹D. J. Tozer and N. C. Handy, "Improving virtual Kohn–Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities", *J. Chem. Phys.* **109**, 10180–10189 (1998).
- ²⁰N. T. Maitra, R. J. C. F. Zhang, and K. Burke, "Double excitations within time-dependent density functional theory linear response", *J. Chem. Phys.* **120**, 5932 (2004).
- ²¹B. G. Levine, C. Ko, J. Quenneville, and T. J. Martínez, "Conical intersections and double excitations in time-dependent density functional theory", *Mol. Phys.* **104**, 1039–1051 (2006).
- ²²P. Elliott, S. Goldson, C. Canahui, and N. T. Maitra, "Perspectives on double-excitations in TDDFT", *Chem. Phys.* **391**, 110–119 (2011).
- ²³J. F. Stanton and R. J. Bartlett, "The equation of motion coupled-cluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties", *J. Chem. Phys.* **98**, 7029–7039 (1993).
- ²⁴S. A. Kucharski, M. Włoch, M. Musiał, and R. J. Bartlett, "Coupled-cluster theory for excited electronic states: the full equation-of-motion coupled-cluster single, double, and triple excitation method", *J. Chem. Phys.* **115**, 8263–8266 (2001).
- ²⁵O. Christiansen, H. Koch, and P. Jørgensen, "The second-order approximate coupled cluster singles and doubles model cc2", *Chem. Phys. Lett.* **243**, 409–418 (1995).
- ²⁶O. Christiansen, H. Koch, and P. Jørgensen, "Response functions in the CC3 iterative triple excitation model", *J. Chem. Phys.* **103**, 7429–7441 (1995).
- ²⁷J. Schirmer, "Beyond the random-phase approximation: A new approximation scheme for the polarization propagator", *Phys. Rev. A* **26**, 2395–2416 (1982).
- ²⁸A. B. Trofimov, G. Stelter, and J. Schirmer, "A consistent third-order propagator method for electronic excitation", *J. Chem. Phys.* **111**, 9982–9999 (1999).
- ²⁹P. H. P. Harbach, M. Wormit, and A. Dreuw, "The third-order algebraic diagrammatic construction method (ADC(3)) for the polarization propagator for closed-shell molecules: Efficient implementation and benchmarking", *J. Chem. Phys.* **141**, 064113 (2014).
- ³⁰P. F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, "A mountaineering strategy to excited states: highly-accurate reference energies and benchmarks", *J. Chem. Theory Comput.* **14**, 4360 (2018).
- ³¹A. B. Trofimov, G. Stelter, and J. Schirmer, "Electron excitation energies using a consistent third-order propagator approach: Comparison with full configuration interaction and coupled cluster results", *J. Chem. Phys.* **117**, 6402–6410 (2002).
- ³²P. F. Loos, F. Lipparini, M. Boggio-Pasqua, A. Scemama, and D. Jacquemin, "A mountaineering strategy to excited states: highly-accurate energies and benchmarks for medium size molecules," *J. Chem. Theory Comput.* **16**, 1711 (2020).
- ³³P.-F. Loos and D. Jacquemin, "Is ADC(3) as accurate as CC3 for valence and rydberg transition energies?", *J. Phys. Chem. Lett.* **11**, 974–980 (2020).
- ³⁴B. Huron, J. P. Malrieu, and P. Rancurel, "Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions", *J. Chem. Phys.* **58**, 5745–5759 (1973).

- ³⁵Y. Garniron, K. Gasperich, T. Applencourt, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J. P. Malrieu, M. Vénil, M. Caffarel, P. F. Loos, E. Giner, and A. Scemama, "Quantum package 2.0: a open-source determinant-driven suite of programs", *J. Chem. Theory Comput.* **15**, 3591 (2019).
- ³⁶P.-F. Loos, M. Boggio-Pasqua, A. Scemama, M. Caffarel, and D. Jacquemin, "Reference energies for double excitations", *J. Chem. Theory Comput.* **15**, 1939–1956 (2019).
- ³⁷A. A. Holmes, C. J. Umrigar, and S. Sharma, "Excited states using semistochastic heat-bath configuration interaction", *J. Chem. Phys.* **147**, 164111 (2017).
- ³⁸A. D. Chien, A. A. Holmes, M. Otten, C. J. Umrigar, S. Sharma, and P. M. Zimmerman, "Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction", *J. Phys. Chem. A* **122**, 2714–2722 (2018).
- ³⁹P. F. Loos, A. Scemama, and D. Jacquemin, "The quest for highly-accurate excitation energies: a computational perspective", *J. Phys. Chem. Lett.* **11**, 2374–2383 (2020).
- ⁴⁰K. Burke, J. Werschnik, and E. K. U. Gross, "Time-dependent density functional theory: past, present, and future", *J. Chem. Phys.* **123**, 062206 (2005).
- ⁴¹Intel product specifications, <https://ark.intel.com/content/www/us/en/ark.html>.
- ⁴²Wikichip, <https://en.wikichip.org/wiki/WikiChip>.
- ⁴³H. Sutter and J. Larus, "Software and the concurrency revolution", *Queue* **3**, 54–62 (2005).
- ⁴⁴Caractéristiques techniques du système de calcul olympe, <https://www.calmip.univ-toulouse.fr/spip.php?article586>.
- ⁴⁵AMD EPYC™ 7H12, <https://www.amd.com/en/products/cpu/amd-epyc-7h12>.
- ⁴⁶Tgcc joliot curie, <http://www-hpc.cea.fr/en/complex/tgcc-JoliotCurie.htm>.
- ⁴⁷L. Hedin, "New method for calculating the one-particle Green's function with application to the electron-gas problem", *Phys. Rev.* **139**, A796 (1965).
- ⁴⁸J. Čížek, "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods", *J. Chem. Phys.* **45**, 4256–4266 (1966).
- ⁴⁹J. Paldus, J. Čížek, and I. Shavitt, "Correlation problems in atomic and molecular systems. iv. extended coupled-pair many-electron theory and its application to the bH_3 molecule", *Phys. Rev. A* **5**, 50–67 (1972).
- ⁵⁰T. D. Crawford and H. F. Schaefer, "An Introduction to Coupled Cluster Theory for Computational Chemists", in *Reviews in Computational Chemistry* (John Wiley & Sons, Ltd, 2000), pp. 33–136.
- ⁵¹R. J. Bartlett and M. Musiał, "Coupled-cluster theory in quantum chemistry", *Rev. Mod. Phys.* **79**, 291–352 (2007).
- ⁵²G. P. Purvis III and R. J. Bartlett, "A full coupled-cluster singles and doubles model: the inclusion of disconnected triples", *J. Chem. Phys.* **76**, 1910–1918 (1982).
- ⁵³G. E. Scuseria and H. F. Schaefer, "A new implementation of the full ccsdt model for molecular electronic structure", *Chem. Phys. Lett.* **152**, 382–386 (1988).
- ⁵⁴G. E. Scuseria and H. F. Schaefer, "Is coupled cluster singles and doubles (ccsd) more computationally intensive than quadratic configuration interaction (qcisd)?", *J. Chem. Phys.* **90**, 3700–3703 (1989).
- ⁵⁵K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories", *Chem. Phys. Lett.* **157**, 479–483 (1989).
- ⁵⁶M. Vénil, P. Romaniello, J. A. Berger, and P.-F. Loos, "Unphysical discontinuities in GW methods", *J. Chem. Theory. Comput.* **14**, 5220–5228 (2018).

- ⁵⁷G. Onida, L. Reining, and A. R. and, "Electronic excitations: density-functional versus many-body green's-function approaches", *Rev. Mod. Phys.* **74**, 601–659 (2002).
- ⁵⁸F. Aryasetiawan and O. Gunnarsson, "The GW method", *Rep. Prog. Phys.* **61**, 237–312 (1998).
- ⁵⁹L. Reining, "The GW approximation: content, successes and limitations: The GW approximation", *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, e1344 (2017).
- ⁶⁰M. S. Hybertsen and S. G. Louie, "First-Principles Theory of Quasiparticles: Calculation of Band Gaps in Semiconductors and Insulators", *Phys. Rev. Lett.* **55**, 1418–1421 (1985).
- ⁶¹M. J. van Setten, F. Weigend, and F. Evers, "The GW -Method for Quantum Chemistry Applications: Theory and Implementation", *J. Chem. Theory Comput.* **9**, 232–246 (2013).
- ⁶²F. Bruneval, "Ionization energy of atoms obtained from GW self-energy or from random phase approximation total energies", *J. Chem. Phys.* **136**, 194107 (2012).
- ⁶³F. Bruneval and M. A. L. Marques, "Benchmarking the Starting Points of the GW Approximation for Molecules", *J. Chem. Theory Comput.* **9**, 324–329 (2013).
- ⁶⁴M. J. van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, "GW 100: Benchmarking G_0W_0 for Molecular Systems", *J. Chem. Theory Comput.* **11**, 5665–5687 (2015).
- ⁶⁵M. J. van Setten, R. Costa, F. Viñes, and F. Illas, "Assessing GW Approaches for Predicting Core Level Binding Energies", *J. Chem. Theory Comput.* **14**, 877–883 (2018).
- ⁶⁶M. S. Hybertsen and S. G. Louie, "Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies", *Phys. Rev. B* **34**, 5390–5413 (1986).
- ⁶⁷M. Shishkin and G. Kresse, "Self-consistent G W calculations for semiconductors and insulators", *Phys. Rev. B* **75**, 235102 (2007).
- ⁶⁸X. Blase, C. Attaccalite, and V. Olevano, "First-principles GW calculations for fullerenes, porphyrins, phtalocyanine, and other molecules of interest for organic photovoltaic applications", *Phys. Rev. B* **83**, 115103 (2011).
- ⁶⁹C. Faber, C. Attaccalite, V. Olevano, E. Runge, and X. Blase, "First-principles GW calculations for DNA and RNA nucleobases", *Phys. Rev. B* **83**, 115123 (2011).
- ⁷⁰S. V. Faleev, M. van Schilfhaarde, and T. Kotani, "All-Electron Self-Consistent G W Approximation: Application to Si, MnO, and NiO", *Phys. Rev. Lett.* **93**, 126406 (2004).
- ⁷¹M. van Schilfhaarde, T. Kotani, and S. Faleev, "Quasiparticle Self-Consistent G W Theory", *Phys. Rev. Lett.* **96**, 226402 (2006).
- ⁷²T. Kotani, M. van Schilfhaarde, and S. V. Faleev, "Quasiparticle self-consistent G W method: A basis for the independent-particle approximation", *Phys. Rev. B* **76**, 165106 (2007).
- ⁷³S.-H. Ke, "All-electron G W methods implemented in molecular orbital space: Ionization energy and electron affinity of conjugated molecules", *Phys. Rev. B* **84**, 205415 (2011).
- ⁷⁴F. Kaplan, M. E. Harding, C. Seiler, F. Weigend, F. Evers, and M. J. van Setten, "Quasi-Particle Self-Consistent GW for Molecules", *J. Chem. Theory Comput.* **12**, 2528–2541 (2016).
- ⁷⁵A. Stan, N. E. Dahlen, and R. van Leeuwen, "Fully self-consistent GW calculations for atoms and molecules", *Europhys. Lett. EPL* **76**, 298–304 (2006).
- ⁷⁶A. Stan, N. E. Dahlen, and R. van Leeuwen, "Levels of self-consistency in the GW approximation", *J. Chem. Phys.* **130**, 114105 (2009).
- ⁷⁷C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, "Fully self-consistent GW calculations for molecules", *Phys. Rev. B* **81**, 085103 (2010).
- ⁷⁸F. Caruso, P. Rinke, X. Ren, M. Scheffler, and A. Rubio, "Unified description of ground and excited states of finite systems: The self-consistent G W approach", *Phys. Rev. B* **86**, 081102(R) (2012).

- ⁷⁹F. Caruso, D. R. Rohr, M. Hellgren, X. Ren, P. Rinke, A. Rubio, and M. Scheffler, "Bond Breaking and Bond Formation: How Electron Correlation is Captured in Many-Body Perturbation Theory and Density-Functional Theory", *Phys. Rev. Lett.* **110**, 146403 (2013).
- ⁸⁰F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, "Self-consistent G W : All-electron implementation with localized basis functions", *Phys. Rev. B* **88**, 075105 (2013).
- ⁸¹F. Caruso, "Self-consistent GW approach for the unified description of ground and excited states of finite systems", PhD Thesis (Freie Universität Berlin, 2013).
- ⁸²P. Koval, D. Foerster, and D. Sánchez-Portal, "Fully self-consistent G W and quasiparticle self-consistent G W for molecules", *Phys. Rev. B* **89**, 155417 (2014).
- ⁸³J. Wilhelm, D. Golze, L. Talirz, J. Hutter, and C. A. Pignedoli, "Toward GW Calculations on Thousands of Atoms", *J. Phys. Chem. Lett.* **9**, 306–312 (2018).
- ⁸⁴F. Caruso, M. Dauth, M. J. van Setten, and P. Rinke, "Benchmark of GW approaches for the GW100 test set", *J. Chem. Theory Comput.* **12**, 5076 (2016).
- ⁸⁵X. Blase, I. Duchemin, and D. Jacquemin, "The Bethe–Salpeter equation in chemistry: relations with TD-DFT, applications and challenges", *Chem. Soc. Rev.* **47**, 1022–1043 (2018).
- ⁸⁶F. Bruneval, T. Rangel, S. M. Hamed, M. Shao, C. Yang, and J. B. Neaton, "Molgw 1: Many-body perturbation theory software for atoms, molecules, and clusters", *Comput. Phys. Commun.* **208**, 149–161 (2016).
- ⁸⁷F. Kaplan, F. Weigend, F. Evers, and M. J. van Setten, "Off-Diagonal Self-Energy Terms and Partially Self-Consistency in GW Calculations for Single Molecules: Efficient Implementation and Quantitative Effects on Ionization Potentials", *J. Chem. Theory Comput.* **11**, 5152–5160 (2015).
- ⁸⁸K. Krause and W. Klopper, "Implementation of the Bethe-Salpeter equation in the TURBOMOLE program", *J. Comput. Chem.* **38**, 383–388 (2017).
- ⁸⁹E. Maggio, P. Liu, M. J. van Setten, and G. Kresse, "GW 100: A Plane Wave Perspective for Small Molecules", *J. Chem. Theory Comput.* **13**, 635–648 (2017).
- ⁹⁰F. Bruneval, S. M. Hamed, and J. B. Neaton, "A systematic benchmark of the *ab initio* Bethe-Salpeter equation approach for low-lying optical excitations of small organic molecules", *J. Chem. Phys.* **142**, 244101 (2015).
- ⁹¹F. Bruneval, "Optimized virtual orbital subspace for faster GW calculations in localized basis", *J. Chem. Phys.* **145**, 234110 (2016).
- ⁹²L. Hung, F. H. da Jornada, J. Souto-Casares, J. R. Chelikowsky, S. G. Louie, and S. Ögüt, "Excitation spectra of aromatic molecules within a real-space G W -BSE formalism: Role of self-consistency and vertex corrections", *Phys. Rev. B* **94**, 085125 (2016).
- ⁹³P. Boulanger, D. Jacquemin, I. Duchemin, and X. Blase, "Fast and Accurate Electronic Excitations in Cyanines with the Many-Body Bethe–Salpeter Approach", *J. Chem. Theory Comput.* **10**, 1212–1218 (2014).
- ⁹⁴D. Jacquemin, I. Duchemin, A. Blondel, and X. Blase, "Benchmark of Bethe-Salpeter for Triplet Excited-States", *J. Chem. Theory Comput.* **13**, 767–783 (2017).
- ⁹⁵J. Li, M. Holzmann, I. Duchemin, X. Blase, and V. Olevano, "Helium Atom Excitations by the G W and Bethe-Salpeter Many-Body Formalism", *Phys. Rev. Lett.* **118**, 163001 (2017).
- ⁹⁶L. Hung, F. Bruneval, K. Baishya, and S. Ögüt, "Benchmarking the GW Approximation and Bethe–Salpeter Equation for Groups IB and IIB Atoms and Monoxides", *J. Chem. Theory Comput.* **13**, 2135–2146 (2017).
- ⁹⁷R. M. Richard, M. S. Marshall, O. Dolgounitcheva, J. V. Ortiz, J.-L. Brédas, N. Marom, and C. D. Sherrill, "Accurate Ionization Potentials and Electron Affinities of Acceptor Molecules I. Reference Data at the CCSD(T) Complete Basis Set Limit", *J. Chem. Theory Comput.* **12**, 595–604 (2016).

- ⁹⁸L. Gallandi, N. Marom, P. Rinke, and T. Körzdörfer, "Accurate Ionization Potentials and Electron Affinities of Acceptor Molecules II: Non-Empirically Tuned Long-Range Corrected Hybrid Functionals", *J. Chem. Theory Comput.* **12**, 605–614 (2016).
- ⁹⁹J. W. Knight, X. Wang, L. Gallandi, O. Dolgounitcheva, X. Ren, J. V. Ortiz, P. Rinke, T. Körzdörfer, and N. Marom, "Accurate Ionization Potentials and Electron Affinities of Acceptor Molecules III: A Benchmark of GW Methods", *J. Chem. Theory Comput.* **12**, 615–626 (2016).
- ¹⁰⁰O. Dolgounitcheva, M. Díaz-Tinoco, V. G. Zakrzewski, R. M. Richard, N. Marom, C. D. Sherrill, and J. V. Ortiz, "Accurate Ionization Potentials and Electron Affinities of Acceptor Molecules IV: Electron-Propagator Methods", *J. Chem. Theory Comput.* **12**, 627–637 (2016).
- ¹⁰¹D. Jacquemin, I. Duchemin, and X. Blase, "Benchmarking the Bethe–Salpeter Formalism on a Standard Organic Molecular Set", *J. Chem. Theory Comput.* **11**, 3290–3304 (2015).
- ¹⁰²P. F. Loos, P. Romaniello, and J. A. Berger, "Green functions and self-consistency: insights from the spherium model", *J. Chem. Theory Comput.* **14**, 3071–3082 (2018).
- ¹⁰³M. Seidl, "Adiabatic connection in density-functional theory: Two electrons on the surface of a sphere", *Phys. Rev. A* **75**, 062506 (2007).
- ¹⁰⁴P.-F. Loos and P. M. W. Gill, "Ground state of two electrons on a sphere", *Phys. Rev. A* **79**, 062517 (2009).
- ¹⁰⁵P.-F. Loos and P. M. W. Gill, "Two Electrons on a Hypersphere: A Quasiexactly Solvable Model", *Phys. Rev. Lett.* **103**, 123008 (2009).
- ¹⁰⁶P.-F. Loos and P. M. Gill, "Excited states of spherium", *Mol. Phys.* **108**, 2527–2532 (2010).
- ¹⁰⁷P.-F. Loos and P. M. W. Gill, "Thinking outside the box: The uniform electron gas on a hypersphere", *J. Chem. Phys.* **135**, 214111 (2011).
- ¹⁰⁸P. M. W. Gill and P.-F. Loos, "Uniform electron gases", *Theor. Chem. Acc.* **131**, 1069 (2012).
- ¹⁰⁹M. E. Casida and D. P. Chong, "Physical interpretation and assessment of the Coulomb-hole and screened-exchange approximation for molecules", *Phys. Rev. A* **40**, 4837–4848 (1989).
- ¹¹⁰M. E. Casida and D. P. Chong, "Simplified Green-function approximations: Further assessment of a polarization model for second-order calculation of outer-valence ionization potentials in molecules", *Phys. Rev. A* **44**, 5773–5783 (1991).
- ¹¹¹G. Stefanucci and R. van Leeuwen, *Nonequilibrium many-body theory of quantum systems: a modern introduction* (Cambridge University Press, Cambridge, 2013).
- ¹¹²J. V. Ortiz, "Electron propagator theory: an approach to prediction and interpretation in quantum chemistry: Electron propagator theory", *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 123–142 (2013).
- ¹¹³J. J. Phillips and D. Zgid, "Communication: The description of strong correlation within self-consistent Green's function second-order perturbation theory", *J. Chem. Phys.* **140**, 241101 (2014).
- ¹¹⁴J. J. Phillips, A. A. Kananenka, and D. Zgid, "Fractional charge and spin errors in self-consistent Green's function theory", *J. Chem. Phys.* **142**, 194108 (2015).
- ¹¹⁵A. A. Rusakov, J. J. Phillips, and D. Zgid, "Local Hamiltonians for quantitative Green's function embedding methods", *J. Chem. Phys.* **141**, 194105 (2014).
- ¹¹⁶A. A. Rusakov and D. Zgid, "Self-consistent second-order Green's function perturbation theory for periodic systems", *J. Chem. Phys.* **144**, 054106 (2016).
- ¹¹⁷S. Hirata, M. R. Hermes, J. Simons, and J. V. Ortiz, "General-Order Many-Body Green's Function Method", *J. Chem. Theory Comput.* **11**, 1595–1606 (2015).
- ¹¹⁸S. Hirata, A. E. Doran, P. J. Knowles, and J. V. Ortiz, "One-particle many-body Green's function theory: Algebraic recursive definitions, linked-diagram theorem, irreducible-diagram theorem, and general-order algorithms", *J. Chem. Phys.* **147**, 044108 (2017).

- ¹¹⁹P. M. W. Gill, "Molecular integrals over gaussian basis functions", *Adv. Quantum Chem.* **25**, 141–205 (1994).
- ¹²⁰M. E. Casida, "Generalization of the optimized-effective-potential model to include electron correlation: A variational derivation of the Sham-Schlüter equation for the exact exchange-correlation potential", *Phys. Rev. A* **51**, 2005–2013 (1995).
- ¹²¹A. Dreuw and M. Head-Gordon, "Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules", *Chem. Rev.* **105**, 4009–4037 (2005).
- ¹²²F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, eds., *Nist handbook of mathematical functions* (Cambridge University Press, New York, 2010).
- ¹²³P. C. Martin and J. Schwinger, "Theory of Many-Particle Systems. I", *Phys. Rev.* **115**, 1342–1373 (1959).
- ¹²⁴G. Baym and L. P. Kadanoff, "Conservation Laws and Correlation Functions", *Phys. Rev.* **124**, 287–299 (1961).
- ¹²⁵G. Baym, "Self-Consistent Approximations in Many-Body Systems", *Phys. Rev.* **127**, 1391–1401 (1962).
- ¹²⁶U. von Barth and B. Holm, "Self-consistent GW 0 results for the electron gas: Fixed screened potential W_0 within the random-phase approximation", *Phys. Rev. B* **54**, 8411 (1996).
- ¹²⁷P. Pulay, "Convergence acceleration of iterative sequences. the case of scf iteration", *Chem. Phys. Lett.* **73**, 393–398 (1980).
- ¹²⁸P. Pulay, "Improved SCF convergence acceleration", *J. Comput. Chem.* **3**, 556–560 (1982).
- ¹²⁹K. P. Huber and G. Herzberg, *Molecular spectra and molecular structure: iv. constants of diatomic molecules* (van Nostrand Reinhold Company, 1979).
- ¹³⁰R. J. Cave, F. Zhang, N. T. Maitra, and K. Burke, "A dressed TDDFT treatment of the 21Ag states of butadiene and hexatriene", *Chem. Phys. Lett.* **389**, 39–42 (2004).
- ¹³¹P. Romaniello, D. Sangalli, J. A. Berger, F. Sottile, L. G. Molinari, L. Reining, and G. Onida, "Double excitations in finite systems", *J. Chem. Phys.* **130**, 044108 (2009).
- ¹³²D. Sangalli, P. Romaniello, G. Onida, and A. Marini, "Double excitations in correlated systems: a many-body approach", *J. Chem. Phys.* **134**, 034115 (2011).
- ¹³³J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple", *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- ¹³⁴R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules* (Oxford, Clarendon Press, 1989).
- ¹³⁵M. F. Lange and T. C. Berkelbach, "On the relation between equation-of-motion coupled-cluster theory and the gw approximation", arXiv, 1805.00043 (2018).
- ¹³⁶A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange", *J. Chem. Phys.* **98**, 5648–5652 (1993).
- ¹³⁷X. Gui, C. Holzer, and W. Klopper, "Accuracy Assessment of GW Starting Points for Calculating Molecular Excitation Energies Using the Bethe–Salpeter Formalism", *J. Chem. Theory Comput.* **14**, 2127–2136 (2018).
- ¹³⁸P. Romaniello, S. Guyot, and L. Reining, "The self-energy beyond GW: Local and nonlocal vertex corrections", *J. Chem. Phys.* **131**, 154111 (2009).
- ¹³⁹P. Romaniello, F. Bechstedt, and L. Reining, "Beyond the G W approximation: Combining correlation channels", *Phys. Rev. B* **85**, 155131 (2012).
- ¹⁴⁰S. Di Sabatino, J. A. Berger, L. Reining, and P. Romaniello, "Reduced density-matrix functional theory: correlation and spectroscopy", *J. Chem. Phys.* **143**, 024108 (2015).

- ¹⁴¹S. Di Sabatino, J. A. Berger, L. Reining, and P. Romaniello, "Photoemission spectra from reduced density matrices: The band gap in strongly correlated systems", *Physical Review B* **94**, 155141 (2016).
- ¹⁴²W. Tarantino, P. Romaniello, J. A. Berger, and L. Reining, "Self-consistent Dyson equation and self-energy functionals: An analysis and illustration on the example of the Hubbard atom", *Phys. Rev. B* **96**, 045124 (2017).
- ¹⁴³N. E. Dahlen, R. van Leeuwen, and U. von Barth, "Variational energy functionals of the Green function and of the density tested on molecules", *Phys. Rev. A* **73**, 012511 (2006).
- ¹⁴⁴F. Furche, "Developing the random phase approximation into a practical post-Kohn–Sham correlation model", *J. Chem. Phys.* **129**, 114105 (2008).
- ¹⁴⁵G. Strinati, "Application of the Green's functions method to the study of the optical properties of semiconductors", *Riv. Nuovo Cimento* **11**, 1–86 (1988).
- ¹⁴⁶X. Leng, F. Jin, M. Wei, and Y. Ma, "GW method and Bethe-Salpeter equation for calculating electronic excitations: GW method and Bethe-Salpeter equation", *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **6**, 532–550 (2016).
- ¹⁴⁷Y. Pavlyukh, "Pade resummation of many-body perturbation theory", *Nature* **7**, 504 (2017).
- ¹⁴⁸J. Lee and M. Head-Gordon, "Regularized orbital-optimized second-order Møller–Plesset perturbation theory: a reliable fifth-order-scaling electron correlation model with orbital energy dependent regularizers", *J. Chem. Theory Comput., ASAP article* (2018).
- ¹⁴⁹M. Véril, A. Scemama, M. Caffarel, F. Lipparini, M. Boggio-Pasqua, D. Jacquemin, and P.-F. Loos, "QUESTDB : a database of highly accurate excitation energies for the electronic structure community", *WIREs Comput. Mol. Sci.* **11**, e1517 (2021).
- ¹⁵⁰C. J. Cramer, *Essentials of computational chemistry: theories and models* (Wiley, 2004).
- ¹⁵¹P. Piecuch, K. Kowalski, I. S. O. Pimienta, and M. J. McGuire, "Recent advances in electronic structure theory: method of moments of coupled-cluster equations and renormalized coupled-cluster approaches", *Int. Rev. Phys. Chem.* **21**, 527–655 (2002).
- ¹⁵²A. I. Krylov, "Spin-flip equation-of-motion coupled-cluster electronic structure method for a description of excited states, bond breaking, diradicals, and triradicals", *Accounts of Chemical Research* **39**, PMID: 16489727, 83–91 (2006).
- ¹⁵³K. Snegov and O. Christiansen, "Excited state coupled cluster methods", *WIREs Comput. Mol. Sci.* **2**, 566–584 (2012).
- ¹⁵⁴L. González, D. Escudero, and L. Serrano-Andrés, "Progress and challenges in the calculation of electronic excited states", *ChemPhysChem* **13**, 28–51 (2012).
- ¹⁵⁵A. D. Laurent and D. Jacquemin, "TD-DFT benchmarks: a review", *Int. J. Quantum Chem.* **113**, 2019–2039 (2013).
- ¹⁵⁶C. Adamo and D. Jacquemin, "The calculations of excited-state properties with time-dependent density functional theory", *Chem. Soc. Rev.* **42**, 845–856 (2013).
- ¹⁵⁷S. Ghosh, P. Verma, C. J. Cramer, L. Gagliardi, and D. G. Truhlar, "Combining wave function methods with density functional theory for excited states", *Chem. Rev.* **118**, 7249–7292 (2018).
- ¹⁵⁸X. Blase, I. Duchemin, D. Jacquemin, and P. F. Loos, "The bethe-salpeter formalism: from physics to chemistry", *J. Phys. Chem. Lett.* **11**, 7371 (2020).
- ¹⁵⁹F. Bernardi, M. Olivucci, and M. A. Robb, "Potential energy surface crossings in organic photochemistry", *Chem. Soc. Rev.* **25**, 321 (1996).
- ¹⁶⁰M. Olivucci, *Computational photochemistry*, OCLC: 800555856 (Elsevier Science, Amsterdam; Boston (Mass.); Paris, 2010).

- ¹⁶¹M. A. Robb, M. Garavelli, M. Olivucci, and F. Bernardi, "A Computational Strategy for Organic Photochemistry", in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2007), pp. 87–146.
- ¹⁶²I. Navizet, Y.-J. Liu, N. Ferre, D. Roca-Sanjuan, and R. Lindh, "The chemistry of bioluminescence: an analysis of chemical functionalities", *ChemPhysChem* **12**, 3064–3076 (2011).
- ¹⁶³R. Crespo-Otero and M. Barbatti, "Recent advances and perspectives on nonadiabatic mixed quantum–classical dynamics", *Chem. Rev.* **118**, 7026–7068 (2018).
- ¹⁶⁴M. A. Robb, *Theoretical chemistry for electronic excited states*, Theoretical and Computational Chemistry Series (The Royal Society of Chemistry, 2018), P001–225.
- ¹⁶⁵S. Mai and L. González, "Molecular photochemistry: recent developments in theory", *Angew. Chem. Int. Ed.* **59**, 16832–16846 (2020).
- ¹⁶⁶H. J. Monkhorst, "Calculation of properties with the coupled-cluster method", *International Journal of Quantum Chemistry* **12**, 421–432 (1977).
- ¹⁶⁷T. Helgaker, P. Jørgensen, and N. Handy, "A numerically stable procedure for calculating møller-plesset energy derivatives, derived using the theory of lagrangians", *Theoret. Chim. Acta* **76**, 227–245 (1989).
- ¹⁶⁸H. Koch, H. J. A. Jensen, P. Jørgensen, and T. Helgaker, "Excitation energies from the coupled cluster singles and doubles linear response function (CCSDLR). Applications to Be, CH⁺, CO, and H₂O", *J. Chem. Phys.* **93**, 3345–3350 (1990).
- ¹⁶⁹O. Christiansen, P. Jørgensen, and C. Hättig, "Response functions from fourier component variational perturbation theory applied to a time-averaged quasienergy", *Int. J. Quantum Chem.* **68**, 1–52 (1998).
- ¹⁷⁰C. Hättig, "Geometry optimizations with the coupled-cluster model CC2 using the resolution-of-the-identity approximation", *J. Chem. Phys.* **118**, 7751–7761 (2003).
- ¹⁷¹M. Kállay and J. Gauss, "Calculation of excited-state properties using general coupled-cluster and configuration-interaction models", *J. Chem. Phys.* **121**, 9257–9269 (2004).
- ¹⁷²C. Hättig, "Structure optimizations for excited states with correlated second-order methods: cc2 and adc(2)", in *Response theory and molecular properties (a tribute to jan linderberg and poul jørgensen)*, Vol. 50, edited by H. Å. Jensen, Advances in Quantum Chemistry (Academic Press, 2005), pp. 37–60.
- ¹⁷³A. Dreuw and M. Wormit, "The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states", *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **5**, 82–95 (2015).
- ¹⁷⁴J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, "Gaussian-1 theory: a general procedure for prediction of molecular energies", *J. Chem. Phys.* **90**, 5622–5629 (1989).
- ¹⁷⁵L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, "Gaussian-2 theory for molecular energies of first- and second-row compounds", *J. Chem. Phys.* **94**, 7221–7230 (1991).
- ¹⁷⁶L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, "Assessment of gaussian-2 and density functional theories for the computation of enthalpies of formation", *J. Chem. Phys.* **106**, 1063–1079 (1997).
- ¹⁷⁷L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, "Gaussian-3 (g3) theory for molecules containing first and second-row atoms", *J. Chem. Phys.* **109**, 7764–7776 (1998).
- ¹⁷⁸L. A. Curtiss, P. C. Redfern, and K. Raghavachari, "Gaussian-4 theory", *J. Chem. Phys.* **126**, 084108 (2007).
- ¹⁷⁹J. Ángyán, J. Dobson, G. Jansen, and T. Gould, *London dispersion forces in molecules, solids and nano-structures, An introduction to physical models and computational methods*, Theoretical and Computational Chemistry Series (The Royal Society of Chemistry, 2020), P001–434.

- ¹⁸⁰P. Jurečka, J. Šponer, J. Černý, and P. Hobza, "Benchmark database of accurate (mp2 and ccSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs", *Phys. Chem. Chem. Phys.* **8**, 1985–1993 (2006).
- ¹⁸¹J. Řezáč, K. E. Riley, and P. Hobza, "S66: a well-balanced database of benchmark interaction energies relevant to biomolecular structures", *J. Chem. Theory Comput.* **7**, PMID: 21836824, 2427–2438 (2011).
- ¹⁸²K. Krause, M. E. Harding, and W. Klopper, "Coupled-cluster reference values for the GW27 and GW100 test sets for the assessment of GW methods", *Mol. Phys.* **113**, 1952 (2015).
- ¹⁸³E. Maggio and G. Kresse, "Correlation energy for the homogeneous electron gas: exact Bethe-Salpeter solution and an approximate evaluation", *Phys. Rev. B* **93**, 235113 (2016).
- ¹⁸⁴A. Stuke, C. Kunkel, D. Golze, M. Todorović, J. T. Margraf, K. Reuter, P. Rinke, and H. Oberhofer, "Atomic structures and orbital energies of 61,489 crystal-forming organic molecules", *Sci. Data* **7**, 58 (2020).
- ¹⁸⁵M. Govoni and G. Galli, "GW100: comparison of methods and accuracy of results obtained with the west code", *J. Chem. Theory Comput.* **14**, 1895–1909 (2018).
- ¹⁸⁶A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, "Heat: high accuracy extrapolated ab initio thermochemistry", *J. Chem. Phys.* **121**, 11599–11613 (2004).
- ¹⁸⁷Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, and J. F. Stanton, "High-accuracy extrapolated ab initio thermochemistry. ii. minor improvements to the protocol and a vital simplification", *J. Chem. Phys.* **125**, 064108 (2006).
- ¹⁸⁸M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, "High-accuracy extrapolated ab initio thermochemistry. iii. additional improvements and overview", *J. Chem. Phys.* **128**, 114111 (2008).
- ¹⁸⁹M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, et al., "Towards the solution of the many-electron problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods", *Phys. Rev. X* **7**, 031059 (2017).
- ¹⁹⁰K. T. Williams, Y. Yao, J. Li, L. Chen, H. Shi, M. Motta, C. Niu, U. Ray, S. Guo, R. J. Anderson, et al., "Direct comparison of many-body methods for realistic electronic hamiltonians", *Phys. Rev. X* **10**, 011041 (2020).
- ¹⁹¹Y. Zhao and D. G. Truhlar, "Comparative assessment of density functional methods for 3d transition-metal chemistry", *J. Chem. Phys.* **124**, 224105 (2006).
- ¹⁹²L. Goerigk and S. Grimme, "A general database for main group thermochemistry, kinetics, and noncovalent interactions – assessment of common and reparameterized (meta-)GGA density functionals", *J. Chem. Theory Comput.* **6**, PMID: 26614324, 107–126 (2010).
- ¹⁹³L. Goerigk and S. Grimme, "A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions", *Phys. Chem. Chem. Phys.* **13**, 6670–6688 (2011).
- ¹⁹⁴L. Goerigk and S. Grimme, "Efficient and accurate double-hybrid-meta-GGA density functionals—evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions", *J. Chem. Theory Comput.* **7**, PMID: 26596152, 291–309 (2011).
- ¹⁹⁵L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, "A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions", *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017).

- ¹⁹⁶M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer, and W. Thiel, "Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3", *J. Chem. Phys.* **128**, 134110 (2008).
- ¹⁹⁷M. R. Silva-Junior, M. Schreiber, S. P. A. Sauer, and W. Thiel, "Benchmarks for electronically excited states: time-dependent density functional theory and density functional theory based multireference configuration interaction", *J. Chem. Phys.* **129**, 104103 (2008).
- ¹⁹⁸M. R. Silva-Junior, M. Schreiber, S. P. A. Sauer, and W. Thiel, "Benchmarks of electronically excited states: Basis set effects on CASPT2 results", *J. Chem. Phys.* **133**, 174318 (2010).
- ¹⁹⁹M. R. Silva-Junior, S. P. Sauer, M. Schreiber, and W. Thiel, "Basis set effects on coupled cluster benchmarks of electronically excited states: CC3, CCSDR(3) and CC2", *Mol. Phys.* **108**, 453–465 (2010).
- ²⁰⁰C. Hättig and F. Weigend, "CC2 excitation energy calculations on large molecules using the resolution of the identity approximation", *J. Chem. Phys.* **113**, 5154–5161 (2000).
- ²⁰¹D. J. ROWE, "Equations-of-motion method and the extended shell model", *Rev. Mod. Phys.* **40**, 153–166 (1968).
- ²⁰²H. Koch, R. Kobayashi, A. Sanchez de Merás, and P. Jørgensen, "Calculation of size-intensive transition moments from the coupled cluster singles and doubles linear response function", *J. Chem. Phys.* **100**, 4393–4400 (1994).
- ²⁰³H. Koch, O. Christiansen, P. Jørgensen, A. M. Sanchez de Merás, and T. Helgaker, "The CC3 model: an iterative coupled cluster approach including connected triples", *J. Chem. Phys.* **106**, 1808–1818 (1997).
- ²⁰⁴K. A. B. O. Roos, M. P. Fulscher, P.-A. Malmqvist, and L. Serrano-Andres, "Adv. chem. phys.", in Vol. XCIII, edited by I. Prigogine and S. A. Rice (Wiley, New York, 1996), pp. 219–331.
- ²⁰⁵S. S. Leang, F. Zahariev, and M. S. Gordon, "Benchmarking the performance of time-dependent density functional methods", *J. Chem. Phys.* **136**, 104101 (2012).
- ²⁰⁶M. Casida and M. Huix-Rotllant, "Progress in time-dependent density-functional theory", *Annu. Rev. Phys. Chem.* **63**, 287 (2012).
- ²⁰⁷C. Ullrich, *Time-dependent density-functional theory: concepts and applications*, Oxford Graduate Texts (Oxford University Press, New York, 2012).
- ²⁰⁸T. Schwabe and L. Goerigk, "Time-dependent double-hybrid density functionals with spin-component and spin-opposite scaling", *J. Chem. Theory Comput.* **13**, 4307.
- ²⁰⁹M. Casanova-Paez, M. B. Dardis, and L. Goerigk, " ω B2p1yp and ω b2gp1yp: the first two double-hybrid density functionals with long-range correction optimized for excitation energies", *J. Chem. Theory Comput.* **15**, 4735 (2019).
- ²¹⁰M. Casanova-Paez and L. Goerigk, "Assessing the tamm–dancoff approximation, singlet–singlet, and singlet–triplet excitations with the latest long-range corrected double-hybrid density functionals", *J. Chem. Phys.* **153**, 064106 (2020).
- ²¹¹F. Furche and R. Ahlrichs, "Adiabatic time-dependent density functional methods for excited state properties", *J. Chem. Phys.* **117**, 7433 (2002).
- ²¹²R. Send, M. Kühn, and F. Furche, "Assessing excited state methods by adiabatic excitation energies", *J. Chem. Theory Comput.* **7**, 2376–2386 (2011).
- ²¹³N. O. C. Winter, N. K. Graf, S. Leutwyler, and C. Hättig, "Benchmarks for 0–0 transitions of aromatic organic molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 compared to high-resolution gas-phase data", *Phys. Chem. Chem. Phys.* **15**, 6623–6630 (2013).
- ²¹⁴P. F. Loos and D. Jacquemin, "Chemically accurate 0-0 energies with not-so-accurate excited state geometries", *J. Chem. Theory Comput.* **15**, 2481 (2019).
- ²¹⁵M. Dierksen and S. Grimme, "A density functional calculation of the vibronic structure of electronic absorption spectra", *J. Chem. Phys.* **120**, 3544–3554 (2004).

- ²¹⁶L. Goerigk and S. Grimme, "Assessment of td-dft methods and of various spin scaled CIS_nD and CC2 versions for the treatment of low-lying valence excitations of large organic dyes", *J. Chem. Phys.* **132**, 184103 (2010).
- ²¹⁷D. Jacquemin, A. Planchat, C. Adamo, and B. Mennucci, "A TD-DFT assessment of functionals for optical 0-0 transitions in solvated dyes", *J. Chem. Theory Comput.* **8**, 2359–2372 (2012).
- ²¹⁸D. Jacquemin, I. Duchemin, and X. Blase, "0–0 energies using hybrid schemes: benchmarks of TD-DFT, CIS(D), ADC(2), CC2, and BSE/GW formalisms for 80 real-life compounds", *J. Chem. Theory Comput.* **11**, 5340–5359 (2015).
- ²¹⁹B. Kozma, A. Tajti, B. Demoulin, R. Izsak, M. Nooijen, and P. G. Szalay, "A new benchmark set for excitation energy of charge transfer states: systematic investigation of coupled cluster type methods", *J. Chem. Theory Comput.* **16**, 4213–4225 (2020).
- ²²⁰C. E. Hoyer, S. Ghosh, D. G. Truhlar, and L. Gagliardi, "Multiconfiguration pair-density functional theory is as accurate as caspt2 for electronic excitation", *J. Phys. Chem. Lett.* **7**, 586–591 (2016).
- ²²¹P. F. Loos, A. Scemama, M. Boggio-Pasqua, and D. Jacquemin, "A mountaineering strategy to excited states: highly-accurate energies and benchmarks for exotic molecules and radicals", *J. Chem. Theory Comput.* **16**, 3720–3736 (2020).
- ²²²A. Köhn and C. Hättig, "Analytic gradients for excited states in the coupled-cluster model CC2 employing the resolution-of-the-identity approximation", *J. Chem. Phys.* **119**, 5021–5036 (2003).
- ²²³C. Fang, B. Oruganti, and B. Durbeej, "How method-dependent are calculated differences between vertical, adiabatic and 0-0 excitation energies?", *J. Phys. Chem. A* **118**, 4157–4171 (2014).
- ²²⁴B. Oruganti, C. Fang, and B. Durbeej, "Assessment of a composite CC2/DFT procedure for calculating 0–0 excitation energies of organic molecules", *Mol. Phys.* **114**, 3448–3463 (2016).
- ²²⁵J. Noga and R. J. Bartlett, "The full CCSDT model for molecular electronic structure", *J. Chem. Phys.* **86**, 7041–7050 (1987).
- ²²⁶S. A. Kucharski and R. J. Bartlett, "Recursive intermediate factorization and complete computational linearization of the coupled-cluster single, double, triple, and quadruple excitation equations", *Theor. Chim. Acta* **80**, 387–405 (1991).
- ²²⁷K. Kowalski and P. Piecuch, "The active-space equation-of-motion coupled-cluster methods for excited electronic states: full EOMCCSDt", *J. Chem. Phys.* **115**, 643–651 (2001).
- ²²⁸M. Kállay, J. Gauss, and P. G. Szalay, "Analytic first derivatives for general coupled-cluster and configuration interaction models", *J. Chem. Phys.* **119**, 2991–3004 (2003).
- ²²⁹S. Hirata, M. Nooijen, and R. J. Bartlett, "High-order determinantal equation-of-motion coupled-cluster calculations for electronic excited states", *Chem. Phys. Lett.* **326**, 255–262 (2000).
- ²³⁰S. Hirata, "Higher-order equation-of-motion coupled-cluster methods", *J. Chem. Phys.* **121**, 51–59 (2004).
- ²³¹G. H. Booth, A. J. W. Thom, and A. Alavi, "Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space", *J. Chem. Phys.* **131**, 054106 (2009).
- ²³²G. H. Booth and A. Alavi, "Approaching chemical accuracy using full configuration-interaction quantum monte carlo: a study of ionization potentials", *J. Chem. Phys.* **132**, 174104 (2010).
- ²³³D. Cleland, G. H. Booth, and A. Alavi, "Communications: survival of the fittest: accelerating convergence in full configuration-interaction quantum monte carlo", *J. Chem. Phys.* **132**, 041103 (2010).
- ²³⁴G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, "Breaking the carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods", *J. Chem. Phys.* **135**, 084104 (2011).

- ²³⁵C. Daday, S. Smart, G. H. Booth, A. Alavi, and C. Filippi, "Full Configuration Interaction Excitations of Ethene and Butadiene: Resolution of an Ancient Question", *J. Chem. Theory Comput.* **8**, 4441–4451 (2012).
- ²³⁶N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, "An excited-state approach within full configuration interaction quantum Monte Carlo", *J. Chem. Phys.* **143**, 134117 (2015).
- ²³⁷K. Ghanem, A. Y. Lozovoi, and A. Alavi, "Unbiasing the initiator approximation in full configuration interaction quantum monte carlo", *J. Chem. Phys.* **151**, 224108 (2019).
- ²³⁸J. E. Deustua, J. Shen, and P. Piecuch, "Converging high-level coupled-cluster energetics by monte carlo sampling and moment expansions", *Phys. Rev. Lett.* **119**, 223003 (2017).
- ²³⁹J. E. Deustua, I. Magoulas, J. Shen, and P. Piecuch, "Communication: approaching exact quantum chemistry by cluster analysis of full configuration interaction quantum monte carlo wave functions", *J. Chem. Phys.* **149**, 151101 (2018).
- ²⁴⁰J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, "Fast semistochastic heat-bath configuration interaction", *J. Chem. Phys.* **149**, 214110 (2018).
- ²⁴¹Y. Yao, E. Giner, J. Li, J. Toulouse, and C. J. Umrigar, *Almost exact energies for the gaussian-2 set with the semistochastic heat-bath configuration interaction method*, 2020.
- ²⁴²J. Li, Y. Yao, A. A. Holmes, M. Otten, Q. Sun, S. Sharma, and C. J. Umrigar, "Accurate many-body electronic structure near the basis set limit: application to the chromium dimer", *Phys. Rev. Research* **2**, 012015 (2020).
- ²⁴³J. J. Eriksen, F. Lipparini, and J. Gauss, "Virtual orbital many-body expansions: a possible route towards the full configuration interaction limit", *J. Phys. Chem. Lett.* **8**, 4633–4639 (2017).
- ²⁴⁴J. J. Eriksen and J. Gauss, "Many-body expanded full configuration interaction. i. weakly correlated regime", *J. Chem. Theory Comput.* **14**, 5180 (2018).
- ²⁴⁵J. J. Eriksen and J. Gauss, "Many-body expanded full configuration interaction. ii. strongly correlated regime", *J. Chem. Theory Comput.* **15**, 4873 (2019).
- ²⁴⁶J. J. Eriksen and J. Gauss, "Generalized many-body expanded full configuration interaction theory", *J. Phys. Chem. Lett.* **27**, 7910–7915 (2019).
- ²⁴⁷E. Xu, M. Uejima, and S. L. Ten-no, "Full coupled-cluster reduction for accurate description of strong electron correlation", *Phys. Rev. Lett.* **121**, 113001 (2018).
- ²⁴⁸E. Xu, M. Uejima, and S. L. Ten-no, *Towards near-exact solutions of molecular electronic structure: full coupled-cluster reduction with a second-order perturbative correction*, 2020.
- ²⁴⁹P.-F. Loos, Y. Damour, and A. Scemama, "The performance of CIPSI on the ground state electronic energy of benzene", *J. Chem. Phys.* **153**, 176101 (2020).
- ²⁵⁰J. J. Eriksen, "The shape of full configuration interaction to come", *J. Phys. Chem. Lett.* **12**, 418–432 (2021).
- ²⁵¹C. F. Bender and E. R. Davidson, "Studies in configuration interaction: the first-row diatomic hydrides", *Phys. Rev.* **183**, 23–30 (1969).
- ²⁵²J. L. Whitten and M. Hackmeyer, "Configuration interaction studies of ground and excited states of polyatomic molecules. i. the CI formulation and studies of formaldehyde", *J. Chem. Phys.* **51**, 5584–5596 (1969).
- ²⁵³M. L. Abrams and C. D. Sherrill, "Important configurations in configuration interaction and coupled-cluster wave functions", *Chem. Phys. Lett.* **412**, 121–124 (2005).
- ²⁵⁴C. F. Bunge and R. Carbó-Dorca, "Select-divide-and-conquer method for large-scale configuration interaction", *J. Chem. Phys.* **125**, 014108 (2006).
- ²⁵⁵L. Bytautas and K. Ruedenberg, "A priori identification of configurational deadwood", *Chem. Phys.* **356**, 64–75 (2009).

- ²⁵⁶E. Giner, A. Scemama, and M. Caffarel, "Using perturbatively selected configuration interaction in quantum monte carlo calculations", *Can. J. Chem.* **91**, 879–885 (2013).
- ²⁵⁷M. Caffarel, E. Giner, A. Scemama, and A. Ramírez-Solís, "Spin density distribution in open-shell transition metal systems: a comparative post-hartree-fock, density functional theory, and quantum monte carlo study of the cucl₂molecule", *J. Chem. Theory Comput.* **10**, 5286–5296 (2014).
- ²⁵⁸E. Giner, A. Scemama, and M. Caffarel, "Fixed-node diffusion monte carlo potential energy curve of the fluorine molecule f₂ using selected configuration interaction trial wavefunctions", *J. Chem. Phys.* **142**, 044115 (2015).
- ²⁵⁹Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, "Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory", *J. Chem. Phys.* **147**, 034101 (2017).
- ²⁶⁰M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Communication: Toward an improved control of the fixed-node error in quantum Monte Carlo: The case of the water molecule", *J. Chem. Phys.* **144**, 151103 (2016).
- ²⁶¹M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Using CIPSI nodes in diffusion monte carlo", in *Recent progress in quantum monte carlo* (2016) Chap. 2, pp. 15–46.
- ²⁶²A. A. Holmes, N. M. Tubman, and C. J. Umrigar, "Heat-bath configuration interaction: an efficient selected configuration interaction algorithm inspired by heat-bath sampling", *J. Chem. Theory Comput.* **12**, 3674–3680 (2016).
- ²⁶³S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, "Semistochastic heat-bath configuration interaction method: selected configuration interaction with semistochastic perturbation theory", *J. Chem. Theory Comput.* **13**, 1595–1604 (2017).
- ²⁶⁴A. Scemama, Y. Garniron, M. Caffarel, and P.-F. Loos, "Deterministic Construction of Nodal Surfaces within Quantum Monte Carlo: The Case of FeS", *J. Chem. Theory Comput.* **14**, 1395–1402 (2018).
- ²⁶⁵A. Scemama, A. Benali, D. Jacquemin, M. Caffarel, and P.-F. Loos, "Excitation energies from diffusion Monte Carlo using selected configuration interaction nodes", *J. Chem. Phys.* **149**, 034108 (2018).
- ²⁶⁶Y. Garniron, A. Scemama, E. Giner, M. Caffarel, and P. F. Loos, "Selected configuration interaction dressed by perturbation", *J. Chem. Phys.* **149**, 064103 (2018).
- ²⁶⁷F. A. Evangelista, "Adaptive multiconfigurational wave functions", *J. Chem. Phys.* **140**, 124114 (2014).
- ²⁶⁸N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, "A deterministic alternative to the full configuration interaction quantum Monte Carlo method", *J. Chem. Phys.* **145**, 044112 (2016).
- ²⁶⁹N. M. Tubman, C. D. Freeman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, "Modern approaches to exact diagonalization and selected configuration interaction with the adaptive sampling ci method", *J. Chem. Theory Comput.* **16**, 2139 (2020).
- ²⁷⁰J. B. Schriber and F. A. Evangelista, "Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy", *J. Chem. Phys.* **144**, 161106 (2016).
- ²⁷¹J. B. Schriber and F. A. Evangelista, "Adaptive Configuration Interaction for Computing Challenging Electronic Excited States with Tunable Accuracy", *J. Chem. Theory Comput.*, **10**. 1021/acs.jctc.7b00725 (2017).
- ²⁷²W. Liu and M. R. Hoffmann, "iCI: Iterative CI toward full CI", *J. Chem. Theory Comput.* **12**, 1169–1178 (2016).

- ²⁷³M. C. Per and D. M. Cleland, "Energy-based truncation of multi-determinant wavefunctions in quantum monte carlo", *J. Chem. Phys.* **146**, 164101 (2017).
- ²⁷⁴Y. Ohtsuka and J.-y. Hasegawa, "Selected configuration interaction method using sampled first-order corrections to wave functions", *J. Chem. Phys.* **147**, 034102 (2017).
- ²⁷⁵P. M. Zimmerman, "Incremental full configuration interaction", *J. Chem. Phys.* **146**, 104102 (2017).
- ²⁷⁶J. P. Coe, "Machine learning configuration interaction", *J. Chem. Theory Comput.* **14**, 5739 (2018).
- ²⁷⁷W. Liu and M. Hoffmann, "SDS: the static–dynamic–static framework for strongly correlated electrons", *Theor. Chem. Acc.* **133**, 1481 (2014).
- ²⁷⁸Y. Lei, W. Liu, and M. R. Hoffmann, "Further development of SDSPT2 for strongly correlated electrons", *Mol. Phys.* **115**, 2696–2707 (2017).
- ²⁷⁹N. Zhang, W. Liu, and M. R. Hoffmann, "Iterative configuration interaction with selection", *J. Chem. Theory Comput.* **16**, 2296–2316 (2020).
- ²⁸⁰S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, "Convergence of an improved cipsi algorithm", *Chem. Phys.* **75**, 91–102 (1983).
- ²⁸¹R. Cimiraglia, "Second order perturbation correction to CI energies by use of diagrammatic techniques: An improvement to the CIPSI algorithm", *J. Chem. Phys.* **83**, 1746–1749 (1985).
- ²⁸²R. Cimiraglia and M. Persico, "Recent advances in multireference second order perturbation CI: The CIPSI method revisited", *J. Comput. Chem.* **8**, 39–47 (1987).
- ²⁸³F. Illas, J. Rubio, and J. M. Ricart, "Approximate natural orbitals and the convergence of a second order multireference many-body perturbation theory (CIPSI) algorithm", *J. Chem. Phys.* **89**, 6376–6384 (1988).
- ²⁸⁴A. Povill, J. Rubio, and F. Illas, "Treating large intermediate spaces in the CIPSI method through a direct selected CI algorithm", *Theor. Chem. Acc.* **82**, 229–238 (1992).
- ²⁸⁵A. Scemama, E. Giner, T. Applencourt, and M. Caffarel, *QMC using very large configuration interaction-type expansions*, Pacifichem, Advances in Quantum Monte Carlo, 2015.
- ²⁸⁶A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, "Quantum monte carlo with very large multideterminant wavefunctions", *J. Comput. Chem.* **37**, 1866–1875 (2016).
- ²⁸⁷A. Scemama, M. Caffarel, A. Benali, D. Jacquemin, and P. F. Loos., "Influence of pseudopotentials on excitation energies from selected configuration interaction and diffusion monte carlo", *Res. Chem.* **1**, 100002 (2019).
- ²⁸⁸M. Dash, S. Moroni, A. Scemama, and C. Filippi, "Perturbatively selected configuration-interaction wave functions for efficient geometry optimization in quantum monte carlo", arXiv:1804.09610 (2018).
- ²⁸⁹M. Dash, J. Feldt, S. Moroni, A. Scemama, and C. Filippi, "Excited states with selected configuration interaction-quantum monte carlo: chemically accurate excitation energies and geometries", *J. Chem. Theory. Comput.* **15**, PMID: 31348645, 4896–4906 (2019).
- ²⁹⁰A. Scemama, E. Giner, A. Benali, and P.-F. Loos, "Taming the fixed-node error in diffusion monte carlo via range separation", *J. Chem. Phys.* **153**, 174107 (2020).
- ²⁹¹A. Benali, K. Gasperich, K. D. Jordan, T. Applencourt, Y. Luo, M. C. Bennett, J. T. Krogel, L. Shulenburger, P. R. C. Kent, P.-F. Loos, A. Scemama, and M. Caffarel, "Towards a systematic improvement of the fixed-node approximation in diffusion monte carlo for solids", *J. Chem. Phys.* **153**, 184111 (2020).
- ²⁹²Š. Budzák, G. Scalmani, and D. Jacquemin, "Accurate excited-state geometries: a caspt2 and coupled-cluster reference database for small molecules", *J. Chem. Theory Comput.* **13**, 6237–6252 (2017).

- ²⁹³K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettrema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawłowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Salek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren, "The Dalton quantum chemistry program system", *WIREs Comput. Mol. Sci.* **4**, 269–284 (2014).
- ²⁹⁴CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package by J.F. Stanton, J. Gauss, L. Cheng, M.E. Harding, D.A. Matthews, P.G. Szalay with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see <http://www.cfour.de>.
- ²⁹⁵M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16 Revision A.03*, Gaussian Inc. Wallingford CT, 2016.
- ²⁹⁶J. S. Binkley and J. A. Pople, "Self-consistent molecular orbital methods. xix. split-valence gaussian-type basis sets for beryllium", *J. Chem. Phys.* **66**, 879–880 (1977).
- ²⁹⁷T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer, "Efficient diffuse function-augmented basis sets for anion calculations. iii. the 3-21+g basis set for first-row elements, li-f", *J. Comput. Chem.* **4**, 294–301 (1983).
- ²⁹⁸J. D. Dill and J. A. Pople, "Self-consistent molecular orbital methods. xv. extended gaussian-type basis sets for lithium, beryllium, and boron", *J. Chem. Phys.* **62**, 2921–2923 (1975).
- ²⁹⁹R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules", *J. Chem. Phys.* **54**, 724–728 (1971).
- ³⁰⁰M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, "Self-consistent molecular orbital methods. xxiii. a polarization-type basis set for second-row elements", *J. Chem. Phys.* **77**, 3654–3665 (1982).
- ³⁰¹M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, and W. J. Hehre, "Self-consistent molecular-orbital methods. 22. small split-valence basis sets for second-row elements", *J. Am. Chem. Soc.* **104**, 2797–2803 (1982).

- ³⁰²W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self-consistent molecular orbital methods. xii. further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules", *J. Chem. Phys.* **56**, 2257–2261 (1972).
- ³⁰³T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen", *J. Chem. Phys.* **90**, 1007–1023 (1989).
- ³⁰⁴R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions", *J. Chem. Phys.* **96**, 6796–6806 (1992).
- ³⁰⁵B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. vii. valence, core-valence, and scalar relativistic basis sets for li, be, na, and Mg", *Theor. Chem. Acc.* **128**, 69–82 (2011).
- ³⁰⁶D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon", *J. Chem. Phys.* **98**, 1358–1371 (1993).
- ³⁰⁷D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. iv. calculation of static electrical response properties", *J. Chem. Phys.* **100**, 2975–2988 (1994).
- ³⁰⁸D. Feller, "The role of databases in support of computational chemistry calculations", *J. Comput. Chem.* **17**, 1571–1586 (1996).
- ³⁰⁹B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibsom, and T. L. Windus, "A new basis set exchange: an open, up-to-date resource for the molecular sciences community", *J. Chem. Inf. Model.* **59**, 4814–4820 (2019).
- ³¹⁰K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, "Basis set exchange: a community database for computational sciences", *J. Chem. Inf. Model.* **47**, 1045–1052 (2007).
- ³¹¹Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, "An efficient linear-scaling CCSD(T) method based on local natural orbitals", *J. Chem. Phys.* **139**, 094105 (2013).
- ³¹²M. Kállay, Z. Rolik, J. Csontos, P. Nagy, G. Samu, D. Mester, J. Csóka, B. Szabó, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and B. Hégyely., *MRCC, quantum chemical program*, 2017.
- ³¹³T. Applencourt, K. Gasperich, and A. Scemama, *Spin adaptation with determinant-based selected configuration interaction*, 2018.
- ³¹⁴*Turbomole v7.3 2018, a development of university of karlsruhe and forschungszentrum karlsruhe gmbh, 1989-2007, turbomole gmbh, since 2007; available from <http://www.turbomole.com> (accessed 13 june 2016).*
- ³¹⁵M. Head-Gordon, D. Maurice, and M. Oumi, "A perturbative correction to restricted open-shell configuration-interaction with single substitutions for excited-states of radicals", *Chem. Phys. Lett.* **246**, 114–121 (1995).
- ³¹⁶A. I. Krylov and P. M. Gill, "Q-chem: an engine for innovation", *WIREs Comput. Mol. Sci.* **3**, 317–326 (2013).
- ³¹⁷J. F. Stanton and J. Gauss, "Perturbative treatment of the similarity transformed hamiltonian in equation-of-motion coupled-cluster approximations", *J. Chem. Phys.* **103**, 1064–1076 (1995).
- ³¹⁸A. Trofimov and J. Schirmer, "Polarization propagator study of electronic excitation in key heterocyclic molecules i. pyrrole", *Chem. Phys.* **214**, 153–170 (1997).
- ³¹⁹O. Christiansen, H. Koch, and P. Jørgensen, "Perturbative triple excitation corrections to coupled cluster singles and doubles excitation energies", *J. Chem. Phys.* **105**, 1451–1459 (1996).
- ³²⁰J. D. Watts and R. J. Bartlett, "Iterative and non-iterative triple excitation corrections in coupled-cluster methods for excited electronic states: the EOM-CCSDT-3 and EOM-CCSD(\bar{T}) methods", *Chem. Phys. Lett.* **258**, 581–588 (1996).

- ³²¹E. Prochnow, M. E. Harding, and J. Gauss, "Parallel calculation of CCSDT and Mk-MRCCSDT energies", *J. Chem. Theory Comput.* **6**, 2339–2347 (2010).
- ³²²F. Neese, "The ORCA program system", *WIREs Comput. Mol. Sci.* **2**, 73–78 (2012).
- ³²³M. Nooijen and R. J. Bartlett, "A new method for excited states: similarity transformed equation-of-motion coupled-cluster theory", *J. Chem. Phys.* **106**, 6441–6448 (1997).
- ³²⁴A. K. Dutta, M. Nooijen, F. Neese, and R. Izsák, "Exploring the accuracy of a low scaling similarity transformed equation of motion method for vertical excitation energies", *J. Chem. Theory Comput.* **14**, 72–91 (2018).
- ³²⁵C. M. Krauter, M. Pernpointner, and A. Dreuw, "Application of the scaled-opposite-spin approximation to algebraic diagrammatic construction schemes of second order", *J. Chem. Phys.* **138**, 044107 (2013).
- ³²⁶A. Hellweg, S. A. Grün, and C. Hättig, "Benchmarking the performance of spin-component scaled CC2 in ground and electronically excited states", *Phys. Chem. Chem. Phys.* **10**, 4119–4127 (2008).
- ³²⁷R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "Psi4 1.1: an open-source electronic structure program emphasizing automation, advanced libraries, and interoperability", *J. Chem. Theory Comput.* **13**, PMID: 28489372, 3185–3197 (2017).
- ³²⁸M. Pitoňák, P. Neogrády, J. Černý, S. Grimme, and P. Hobza, "Scaled MP3 non-covalent interaction energies agree closely with accurate CCSD(T) benchmark data", *ChemPhysChem* **10**, 282–289 (2009).
- ³²⁹H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package", *WIREs Comput. Mol. Sci.* **2**, 242–253 (2011).
- ³³⁰C. Angeli, R. Cimiraglia, and J.-P. Malrieu, "N-electron valence state perturbation theory: a fast implementation of the strongly contracted variant", *Chem. Phys. Lett.* **350**, 297–305 (2001).
- ³³¹C. Angeli, R. Cimiraglia, and J.-P. Malrieu, "N -electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants", *J. Chem. Phys.* **117**, 9138–9153 (2002).
- ³³²J. Finley, P.-Å. Malmqvist, B. O. Roos, and L. Serrano-Andrés, "The multi-state caspt2 method", *Chem. Phys. Lett.* **288**, 299–306 (1998).
- ³³³T. Shiozaki, W. Györfly, P. Celani, and H.-J. Werner, "Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients", *J. Chem. Phys.* **135**, 081106 (2011).
- ³³⁴M. Dierksen and S. Grimme, "The vibronic structure of electronic absorption spectra of large molecules: a time-dependent density functional study on the influence of 'Exact' hartree-fock exchange", *J. Phys. Chem. A* **108**, 10225–10237 (2004).
- ³³⁵J. K. Hurst, P. Wormell, E. Krausz, and A. R. Lacey, "The electronic spectrum of 1, 4, 5, 8-tetraazanaphthalene", *Chem. Phys.* **246**, 229–246 (1999).
- ³³⁶R. Pou-Amérigo, M. Merchán, and E. Orti, "Theoretical study of the electronic spectrum of p-benzoquinone", *J. Chem. Phys.* **110**, 9536–9546 (1999).
- ³³⁷J. Weber, K. Malsch, and G. Hohlneicher, "Excited electronic states of p-benzoquinone", *Chem. Phys.* **264**, 275–318 (2001).
- ³³⁸Y. Honda, M. Hada, M. Ehara, and H. Nakatsuji, "Excited and ionized states of p-benzoquinone and its anion radical: sac-ci theoretical study", *J. Phys. Chem. A* **106**, 3838–3849 (2002).

- ³³⁹D. Bousquet, R. Fukuda, P. Maitarad, D. Jacquemin, I. Ciofini, C. Adamo, and M. Ehara, "Excited-state geometries of heteroaromatic compounds: a comparative td-dft and sac-ci study", *J. Chem. Theory Comput.* **9**, 2368–2379 (2013).
- ³⁴⁰D. Bousquet, R. Fukuda, D. Jacquemin, I. Ciofini, C. Adamo, and M. Ehara, "Benchmark study on the triplet excited-state geometries and phosphorescence energies of heterocyclic compounds: comparison between td-pbe0 and sac-ci", *J. Chem. Theory Comput.* **10**, 3969–3979 (2014).
- ³⁴¹D. Jacquemin, V. Wathelet, E. A. Perpète, and C. Adamo, "Extensive TD-DFT benchmark: singlet-excited states of organic molecules", *J. Chem. Theory Comput.* **5**, 2420–2435 (2009).
- ³⁴²D. B. Jones, P. Limão-Vieira, M. Mendes, N. C. Jones, S. V. Hoffmann, R. F. da Costa, M. T. d. N. Varella, M. H. F. Bettega, F. Blanco, G. García, O. Ingólfsson, M. A. P. Lima, and M. J. Brunger, "An experimental and theoretical investigation into the electronically excited states of *para*-benzoquinone", *J. Chem. Phys.* **146**, 184303 (2017).
- ³⁴³R. F. da Costa, J. C. Ruivo, F. Kossoski, M. T. d. N. Varella, M. H. F. Bettega, D. B. Jones, M. J. Brunger, and M. A. P. Lima, "An ab initio investigation for elastic and electronically inelastic electron scattering from *para*-benzoquinone", *J. Chem. Phys.* **149**, 174308 (2018).
- ³⁴⁴H. P. Trommsdorff, "Electronic states and spectra of *p*-benzoquinone", *J. Chem. Phys.* **56**, 5358–5372 (1972).
- ³⁴⁵J. Goodman and L. E. Brus, "Distant intramolecular interaction between identical chromophores: the $n - \pi^*$ excited states of *p*-benzoquinone", *J. Chem. Phys.* **69**, 1604–1612 (1978).
- ³⁴⁶P. Brint, J.-P. Connerade, P. Tsekeris, A. Bolovinos, and A. Baig, "Vacuum ultraviolet absorption spectrum of *p*-benzoquinone", *J. Chem. Soc., Faraday Trans. 2* **82**, 367–375 (1986).
- ³⁴⁷L. Serrano-Andrés, R. Pou-Amérigo, M. P. Fülcher, and A. C. Borin, "Electronic excited states of conjugated cyclic ketones and thioketones: a theoretical study", *J. Chem. Phys.* **117**, 1649–1659 (2002).
- ³⁴⁸E. W. Garbisch and R. F. Sprecher, "2,4-di-*t*-butyl- and 3-*t*-butylcyclopentadienones", *J. Am. Chem. Soc.* **88**, 3433–3434 (1966).
- ³⁴⁹D. Khuseynov, J. F. Stanton, and A. Sanov, "Low-lying electronic states of cyclopentadienone", *J. Phys. Chem. A* **118**, PMID: 25093249, 6965–6970 (2014).
- ³⁵⁰R. Sarkar, M. Boggio-Pasqua, P. F. Loos, and D. Jacquemin, "Benchmark of td-dft and wavefunction methods for oscillator strengths and excited-state dipoles", *J. Chem. Theory Comput.* (submitted).
- ³⁵¹M. S. Han, H.-G. Cho, and B.-S. Cheong, "Theoretical studies of the low-lying electronic states of diazirine and 3,3'-dimethyldiazirine", *Bull. Korean. Chem. Soc.* **20**, 1281–1287 (1999).
- ³⁵²I. Fedorov, L. Koziol, A. K. Mollner, A. I. Krylov, and H. Reisler, "Multiphoton ionization and dissociation of diazirine: a theoretical and experimental study", *J. Phys. Chem. A* **113**, 7412–7421 (2009).
- ³⁵³L. C. Robertson and J. A. Merritt, "The electronic absorption spectra of *n*-diazirine, n_1^15 -diazirine, d_1 -diazirine, and d_2 -diazirine", *J. Mol. Spectrosc.* **19**, 372–388 (1966).
- ³⁵⁴W. M. Flicker, O. A. Mosher, and A. Kuppermann, "Low energy, variable angle electron-impact excitation of 1,3,5-hexatriene", *Chem. Phys. Lett.* **45**, 492–497 (1977).
- ³⁵⁵D. G. Leopold, R. D. Pendley, J. L. Roebber, R. J. Hemley, and V. Vaida, "Direct absorption spectroscopy of jet-cooled polyenes. ii. the $1^1B_u^+ \leftarrow 1^1A_g^-$ transitions of butadienes and hexatrienes", *J. Chem. Phys.* **81**, 4218–4229 (1984).
- ³⁵⁶D. G. Leopold, V. Vaida, and M. F. Granville, "Direct absorption spectroscopy of jet-cooled polyenes. i. the $1^1B_u^+ \leftarrow 1^1A_g^-$ transition of *trans,trans*-1,3,5,7-octatetraene", *J. Chem. Phys.* **81**, 4210–4217 (1984).
- ³⁵⁷T. Fujii, A. Kamata, M. Shimizu, Y. Adachi, and S. Maeda, "Two-photon absorption study of 1,3,5-hexatriene by cars and csrs", *Chem. Phys. Lett.* **115**, 369–372 (1985).

- ³⁵⁸L. Serrano-Andrés, M. Merchán, I. Nebot-Gil, R. Lindh, and B. O. Roos, "Towards an accurate molecular orbital theory for excited states: Ethene, butadiene, and hexatriene", *J. Chem. Phys.* **98**, 3151–3162 (1993).
- ³⁵⁹A. Sabljčić and R. McDiarmid, "The lower rydberg states of *trans*-hexatriene", *J. Chem. Phys.* **82**, 2559–2565 (1985).
- ³⁶⁰K. Nakayama, H. Nakano, and K. Hirao, "Theoretical study of the $\pi \rightarrow \pi^*$ excited states of linear polyenes: The energy gap between 1^1Bu^+ and 2^1Ag^+ states and their character", *Int. J. Quantum Chem.* **66**, 157–175 (1998).
- ³⁶¹M. Allan, "The excited states of 1,3-butadiyne determined by electron energy loss spectroscopy", *J. Chem. Phys.* **80**, 6020–6024 (1984).
- ³⁶²A. V. Copan and A. Y. Sokolov, "Linear-response density cumulant theory for excited electronic states", *J. Chem. Theory Comput.* **14**, 4097–4108 (2018).
- ³⁶³T. Climent, R. González-Luque, and M. Merchán, "Theoretical analysis of the excited states in maleimide", *J. Phys. Chem. A* **107**, 6995–7003 (2003).
- ³⁶⁴A. Lehr, S. Gómez, M. A. Parkes, and G. A. Worth, "The role of vibronic coupling in the electronic spectroscopy of maleimide: a multi-mode and multi-state quantum dynamics study", *Phys. Chem. Chem. Phys.* **22**, 25272–25283 (2020).
- ³⁶⁵D. Tuna, Y. Lu, A. Koslowski, and W. Thiel, "Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks of electronically excited states", *J. Chem. Theory Comput.* **12**, 4400–4422 (2016).
- ³⁶⁶C. J. Seliskar and S. P. McGlynn, "Electronic spectroscopy of maleimide and its isoelectronic molecules. i. maleimide and n-alkylmaleimides", *J. Chem. Phys.* **55**, 4337–4342 (1971).
- ³⁶⁷M. Rubio, M. Merchán, E. Ortí, and B. O. Roos, "A theoretical study of the electronic spectrum of naphthalene", *Chem. Phys.* **179**, 395–409 (1994).
- ³⁶⁸M. J. Packer, E. K. Dalskov, T. Enevoldsen, H. J. A. Jensen, and J. Oddershede, "A new implementation of the second-order polarization propagator approximation (soppa): the excitation spectra of benzene and naphthalene", *J. Chem. Phys.* **105**, 5886–5900 (1996).
- ³⁶⁹V. Sauri, L. Serrano-Andrés, A. R. Moughal Shahi, L. Gagliardi, S. Vancoillie, and K. Pierloot, "Multiconfigurational second-order perturbation theory restricted active space (RASPT2) method for electronic excited states: a benchmark study", *J. Chem. Theory Comput.* **7**, 153–168 (2011).
- ³⁷⁰H. Fliegl and D. Sundholm, "Coupled-cluster calculations of the lowest 0–0 bands of the electronic excitation spectrum of naphthalene", *Phys. Chem. Chem. Phys.* **16**, 9859–9865 (2014).
- ³⁷¹A. Prlj, M. E. Sandoval-Salinas, D. Casanova, D. Jacquemin, and C. Corminboeuf, "Low-lying $\pi\pi^*$ states of heteroaromatic molecules: a challenge for excited state methods", *J. Chem. Theory Comput.* **12**, 2652–2660 (2016).
- ³⁷²F. Bettanin, L. F. A. Ferrão, M. Pinheiro, A. J. A. Aquino, H. Lischka, F. B. C. Machado, and D. Nachtigallova, "Singlet 1a and 1b bands for n-acenes (n = 2–7): a casscf/caspt2 study", *J. Chem. Theory Comput.* **13**, PMID: 28719203, 4297–4306 (2017).
- ³⁷³R. Huebner, S. Meilczarek, and C. Kuyatt, "Electron energy-loss spectroscopy of naphthalene vapor", *Chem. Phys. Lett.* **16**, 464–469 (1972).
- ³⁷⁴G. George and G. Morris, "The intensity of absorption of naphthalene from 30 000 cm^{-1} to 53 000 cm^{-1} ", *J. Mol. Spectrosc.* **26**, 67–71 (1968).
- ³⁷⁵N. Mikami and M. Ito, "Two-photon excitation spectra of naphthalene and naphthalene-d8", *Chem. Phys. Lett.* **31**, 472–478 (1975).
- ³⁷⁶B. Dick and G. Hohlneicher, "Two-photon spectroscopy of the low-lying singlet states of naphthalene and acenaphthene", *Chem. Phys. Lett.* **84**, 471–478 (1981).

- ³⁷⁷A. Bree and T. Thirunamachandran, "The crystal spectrum of naphthalene in the region 3200 Å to 2200 Å", *Mol. Phys.* **5**, 397–405 (1962).
- ³⁷⁸H. B. Klevens and J. R. Platt, "Spectral resemblances of cata-condensed hydrocarbons", *J. Chem. Phys.* **17**, 470–481 (1949).
- ³⁷⁹H. E. Hunziker, "Gas-phase absorption spectrum of triplet naphthalene in the 220-300 nm and 410-620 nm wavelength regions", *J. Chem. Phys.* **56**, 400–403 (1972).
- ³⁸⁰H. Hunziker, "Formation of triplet naphthalene by energy transfer from $hg(6^3 p_0)$ atoms, observed by a modulation technique", *Chem. Phys. Lett.* **3**, 504–507 (1969).
- ³⁸¹A. Luna, M. Merchán, and B. O. Ross, "A theoretical analysis of the lowest excited states in HNO/NOH and HPO/POH", *Chem. Phys.* **196**, 437–445 (1995).
- ³⁸²B. Le Guennic and D. Jacquemin, "Taking up the cyanine challenge with quantum tools", *Acc. Chem. Res.* **48**, 530–537 (2015).
- ³⁸³R. Send, O. Valsson, and C. Filippi, "Electronic excitations of simple cyanine dyes: reconciling density functional and wave function methods", *J. Chem. Theory Comput.* **7**, 444–455 (2011).
- ³⁸⁴J. Fabian, "Electronic excitation of sulfur-organic compounds - performance of time-dependent density functional theory", *Theor. Chem. Acc.* **106**, 199–217 (2001).
- ³⁸⁵R. H. Judge and D. C. Moule, "A vibronic analysis of the lower $\tilde{A}^1 A' \leftarrow \tilde{X}^1 A'$ singlet–singlet and $\tilde{a}^3 A' \leftarrow \tilde{X}^1 A'$ triplet–singlet band systems of thioacrolein (2-propenethial)", *J. Chem. Phys.* **80**, 4646–4650 (1984).
- ³⁸⁶M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and $pt(p(t-bu)_3)_2 + h_2$ oxidative addition", *J. Phys. Chem.* **100**, 19357–19363 (1996).
- ³⁸⁷M. Svensson, S. Humbel, and K. Morokuma, "Energetics using the single point imomo (integrated molecular orbital+molecular orbital) calculations: choices of computational levels and model system", *J. Chem. Phys.* **105**, 3654–3661 (1996).
- ³⁸⁸M. Caricato, G. W. Trucks, M. J. Frisch, and K. B. Wiberg, "Electronic transition energies: a study of the performance of a large range of single reference density functional and wave function methods on valence and rydberg states compared to experiment", *J. Chem. Theory Comput.* **6**, 370–383 (2010).
- ³⁸⁹T. J. Watson, V. F. Lotrich, P. G. Szalay, A. Perera, and R. J. Bartlett, "Benchmarking for perturbative triple-excitations in ee-eom-cc methods", *J. Phys. Chem. A* **117**, 2569–2579 (2013).
- ³⁹⁰D. Kánnár and P. G. Szalay, "Benchmarking coupled cluster methods on valence singlet excited states", *J. Chem. Theory Comput.* **10**, 3757–3765 (2014).
- ³⁹¹D. Kánnár, A. Tajti, and P. G. Szalay, "Accuracy of coupled cluster excitation energies in diffuse basis sets", *J. Chem. Theory Comput.* **13**, 202–209 (2017).
- ³⁹²M. Hodecker, D. R. Rehn, A. Dreuw, and S. Höfener, "Similarities and differences of the lagrange formalism and the intermediate state representation in the treatment of molecular properties", *J. Chem. Phys.* **150**, 164125 (2019).
- ³⁹³J. J. Eriksen and J. Gauss, "Ground and excited state first-order properties in many-body expanded full configuration interaction theory", *J. Chem. Phys.* **153**, 154107 (2020).
- ³⁹⁴A. Chrayteh, A. Blondel, P. F. Loos, and D. Jacquemin, "A mountaineering strategy to excited states: highly-accurate oscillator strengths and dipole moments of small molecules", *J. Chem. Theory Comput.* **17**, 416–438 (2021).
- ³⁹⁵D. Jacquemin, "What is the key for accurate absorption and emission calculations ? energy or geometry ?", *J. Chem. Theory Comput.* **14**, 1534–1543 (2018).

- ³⁹⁶T. Gould, "'Diet GMTKN55' offers accelerated benchmarking through a representative subset approach", *Phys. Chem. Chem. Phys.* **20**, 27735–27739 (2018).
- ³⁹⁷P.-F. Loos, M. Comin, X. Blase, and D. Jacquemin, "Reference energies for intramolecular charge-transfer excitations", *J. Chem. Theory. Comput.* **0**, null (0).
- ³⁹⁸H. G. A. Burton and A. J. W. Thom, "General approach for multireference ground and excited states using nonorthogonal configuration interaction", *J. Chem. Theory. Comput.* **15**, PMID: 31381337, 4851–4861 (2019).
- ³⁹⁹J. W. Hollett and P.-F. Loos, "Capturing static and dynamic correlation with Δ NO-MP2 and Δ NO-CCSD", *The Journal of Chemical Physics* **152**, 014101 (2020).
- ⁴⁰⁰S. D. Pineda Flores and E. Neuscamman, "Excited state specific multi-slater jastrow wave functions", *The Journal of Physical Chemistry A* **123**, PMID: 30702890, 1487–1497 (2019).
- ⁴⁰¹L. Kraijger and T. Spalinger, *Shellinabox*, version 2.20, Nov. 10, 2016.
- ⁴⁰²D. Merkel, "Docker: lightweight linux containers for consistent development and deployment", *Linux J.* **2014** (2014).
- ⁴⁰³<https://www.lebigdata.fr/docker-definition>.
- ⁴⁰⁴Open container initiative, <https://opencontainers.org>.
- ⁴⁰⁵Docker, Inc, *Docker documentation*, <https://docs.docker.com>.
- ⁴⁰⁶Namespaces - overview of linux namespaces, <http://man7.org/linux/man-pages/man7/namespaces.7.html>.
- ⁴⁰⁷<https://www.distrowatch.com/table.php?distribution=ubuntu>.
- ⁴⁰⁸<https://blog.nestybox.com/2019/11/11/docker-sandbox.html>.
- ⁴⁰⁹Y. Damour, M. Vénil, F. Kossoski, M. Caffarel, D. Jacquemin, A. Scemama, and P.-F. Loos, "Accurate full configuration interaction correlation energy estimates for five- and six-membered rings", *The Journal of Chemical Physics* **155**, 134104 (2021).
- ⁴¹⁰M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln", *Ann. Phys.* **389**, 457–484 (1927).
- ⁴¹¹I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge Molecular Science (Cambridge University Press, Cambridge, 2009).
- ⁴¹²N. Oliphant and L. Adamowicz, "Coupled-cluster method truncated at quadruples", *J. Chem. Phys.* **95**, 6645 (1991).
- ⁴¹³S. A. Kucharski and R. J. Bartlett, "The coupled-cluster single, double, triple, and quadruple excitation method", *J. Chem. Phys.* **97**, 4282 (1992).
- ⁴¹⁴M. Kállay and J. Gauss, "Approximate treatment of higher excitations in coupled-cluster theory", *J. Chem. Phys.* **123**, 214105 (2005).
- ⁴¹⁵D. A. Matthews, "Analytic gradients of approximate coupled cluster methods with quadruple excitations", *J. Chem. Theory Comput.* **16**, 6195–6206 (2020).
- ⁴¹⁶M. Kállay and J. Gauss, "Analytic second derivatives for general coupled-cluster and configuration-interaction models", *J. Chem. Phys.* **120**, 6841–6848 (2004).
- ⁴¹⁷J. Gauss, A. Tajti, M. Kállay, J. F. Stanton, and P. G. Szalay, "Analytic calculation of the diagonal born-oppenheimer correction within configuration-interaction and coupled-cluster theory", *J. Chem. Phys.* **125**, 144111 (2006).
- ⁴¹⁸M. Kállay and J. Gauss, "Calculation of frequency-dependent polarizabilities using general coupled-cluster models", *J. Mol. Struct. THEOCHEM* **768**, 71–77 (2006).
- ⁴¹⁹J. Gauss, M. Kállay, and F. Neese, "Calculation of electronic g-tensors using coupled cluster theory", *J. Phys. Chem. A* **113**, 11541–11549 (2009).

- ⁴²⁰S. Shih, R. Butscher, W. ans Buenker, and S. Peyerimhoff, "Calculation of vertical electronic spectrum of nitrogen molecule using MRD-CI method", *Chem. Phys.* **29**, 241–252 (1978).
- ⁴²¹R. Buenker, S. Peyerimhoff, and W. Butscher, "Applicability of multi-reference double-excitation ci (mrd-ci) method to calculation of electronic wavefunctions and comparison with related techniques", *Mol. Phys.* **35**, 771–791 (1978).
- ⁴²²J. J. Eriksen, T. A. Anderson, J. E. Deustua, K. Ghanem, D. Hait, M. R. Hoffmann, S. Lee, D. S. Levine, I. Magoulas, J. Shen, N. M. Tubman, K. B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G. K.-L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S. L. Ten-no, C. J. Umrigar, and J. Gauss, "The ground state electronic energy of benzene", *J. Phys. Chem. Lett.* **11**, 8922–8929 (2020).
- ⁴²³J. Ivanic and K. Ruedenberg, "Identification of deadwood in configuration spaces through general direct configuration interaction", *Theor. Chem. Acc.* **106**, 339–351 (2001).
- ⁴²⁴G. J. R. Aroeira, M. M. Davis, J. M. Turney, and H. F. Schaefer, "Coupled cluster externally corrected by adaptive configuration interaction", *J. Chem. Theory Comput.* **17**, 182–190 (2021).
- ⁴²⁵S. Lee, H. Zhai, S. Sharma, C. J. Umrigar, and G. K.-L. Chan, "Externally corrected ccsd with renormalized perturbative triples (R-ecCCSD(T)) and the density matrix renormalization group and selected configuration interaction external sources", *J. Chem. Theory Comput.* **17**, 3414–3425 (2021).
- ⁴²⁶I. Magoulas, K. Gururangan, P. Piecuch, J. E. Deustua, and J. Shen, "Is externally corrected coupled cluster always better than the underlying truncated configuration interaction?", *J. Chem. Theory Comput.* **17**, 4006–4027 (2021).
- ⁴²⁷J. Paldus, "Externally and internally corrected coupled cluster approaches: an overview", *J. Math. Chem.* **55**, 477–502 (2017).
- ⁴²⁸C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems", *Phys. Rev.* **46**, 618–622 (1934).
- ⁴²⁹A. Marie, H. G. A. Burton, and P.-F. Loos, "Perturbation theory in the complex plane: exceptional points and where to find them", *J. Phys.: Condens. Matter* **33**, 283001 (2021).
- ⁴³⁰S. Grimme, "Semiempirical hybrid density functional with perturbative second-order correlation", *J. Chem. Phys.* **124**, 034108 (2006).
- ⁴³¹J. A. Pople, J. S. Binkley, and R. Seeger, "Theoretical models incorporating electron correlation", *Int. J. Quantum Chem.* **10**, 1–19 (1976).
- ⁴³²R. Krishnan, M. J. Frisch, and J. A. Pople, "Contribution of triple substitutions to the electron correlation energy in fourth order perturbation theory", *J. Chem. Phys.* **72**, 4244–4245 (1980).
- ⁴³³S. A. Kucharski, J. Noga, and R. J. Bartlett, "Fifth-order many-body perturbation theory for molecular correlation energies", *J. Chem. Phys.* **90**, 7282–7290 (1989).
- ⁴³⁴Z. He and D. Cremer, "Sixth-order many-body perturbation theory. i. basic theory and derivation of the energy formula", *Int. J. Quantum Chem.* **59**, 15–29 (1996).
- ⁴³⁵Z. He and D. Cremer, "Sixth-order many-body perturbation theory. ii. implementation and application", *Int. J. Quantum Chem.* **59**, 31–55 (1996).
- ⁴³⁶W. D. Laidig, G. Fitzgerald, and R. J. Bartlett, "Is fifth-order MBPT enough?", *Chem. Phys. Lett.* **113**, 151 (1985).
- ⁴³⁷P. J. Knowles, K. Somasundram, N. C. Handy, and K. Hirao, "The calculation of higher-order energies in the many-body perturbation theory series", *Chem. Phys. Lett.* **113**, 8–12 (1985).
- ⁴³⁸N. C. Handy, P. J. Knowles, and K. Somasundram, "On the convergence of the Møller-Plesset perturbation series", *Theoret. Chim. Acta* **68**, 87–100 (1985).
- ⁴³⁹P. M. W. Gill and L. Radom, "Deceptive convergence in Møller-Plesset perturbation energies", *Chem. Phys. Lett.* **132**, 16–22 (1986).

- ⁴⁴⁰W. D. Laidig, P. Saxe, and R. J. Bartlett, "The description of N₂ and F₂ potential energy surfaces using multireference coupled cluster theory", *J. Chem. Phys.* **86**, 887 (1987).
- ⁴⁴¹R. H. Nobes, J. A. Pople, L. Radom, N. C. Handy, and P. J. Knowles, "Slow convergence of the Møller–Plesset perturbation series: the dissociation energy of hydrogen cyanide and the electron affinity of the cyano radical", *Chem. Phys. Lett.* **138**, 481 (1987).
- ⁴⁴²P. M. W. Gill, J. A. Pople, L. Radom, and R. H. Nobes, "Why does unrestricted Møller–Plesset perturbation theory converge so slowly for spin-contaminated wave functions?", *J. Chem. Phys.* **89**, 7307–7314 (1988).
- ⁴⁴³P. M. W. Gill, M. W. Wong, R. H. Nobes, and L. Radom, "How well can RMP4 theory treat homolytic fragmentations?", *Chem. Phys. Lett.* **148**, 541 (1988).
- ⁴⁴⁴M. B. Lepetit, M. Péliissier, and J. P. Malrieu, "Origins of the poor convergence of many-body perturbation theory expansions from unrestricted Hartree-Fock zeroth-order descriptions", *J. Chem. Phys.* **89**, 998–1008 (1988).
- ⁴⁴⁵J.-P. Malrieu and C. Angeli, "The Møller–Plesset perturbation revisited: origin of high-order divergences", *Mol. Phys.* **111**, 1092–1099 (2013).
- ⁴⁴⁶N. M. Tubman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, *An efficient deterministic perturbation theory for selected configuration interaction methods*, 2018.
- ⁴⁴⁷S. R. White, "Density matrix formulation for quantum renormalization groups", *Phys. Rev. Lett.* **69**, 2863 (1992).
- ⁴⁴⁸S. R. White, "Density-matrix algorithms for quantum renormalization groups", *Phys. Rev. B* **48**, 10345 (1993).
- ⁴⁴⁹G. K.-L. Chan and S. Sharma, "The density matrix renormalization group in quantum chemistry", *Annu. Rev. Phys. Chem.* **62**, 465–481 (2011).
- ⁴⁵⁰J. Lee, F. D. Malone, and D. R. Reichman, "The performance of phaseless auxiliary-field quantum monte carlo on the ground state electronic energy of benzene", *J. Chem. Phys.* **153**, 126101 (2020).
- ⁴⁵¹M. Motta and S. Zhang, "Ab initio computations of molecular systems by the auxiliary-field quantum monte carlo method", *WIREs Comput. Mol. Sci.* **8**, e1364 (2018).
- ⁴⁵²Y. Yao and C. J. Umrigar, "Orbital optimization in selected configuration interaction methods", *J. Chem. Theory Comput.* **17**, 4183–4194 (2021).
- ⁴⁵³K. Kowalski and P. Piecuch, "The method of moments of coupled-cluster equations and the renormalized CCSD[T], CCSD(T), CCSD(TQ), and CCSDT(Q) approaches", *J. Chem. Phys.* **113**, 18–35 (2000).
- ⁴⁵⁴K. Kowalski and P. Piecuch, "Renormalized CCSD(T) and CCSD(TQ) approaches: Dissociation of the N₂ triple bond", *J. Chem. Phys.* **113**, 5644–5652 (2000).
- ⁴⁵⁵P. Piecuch, S. A. Kucharski, K. Kowalski, and M. Musiał, "Efficient computer implementation of the renormalized coupled-cluster methods: The R-CCSD[T], R-CCSD(T), CR-CCSD[T], and CR-CCSD(T) approaches", *Comput. Phys. Commun.* **149**, 71–96 (2002).
- ⁴⁵⁶P. Piecuch and M. Włoch, "Renormalized coupled-cluster methods exploiting left eigenstates of the similarity-transformed Hamiltonian", *J. Chem. Phys.* **123**, 224105 (2005).
- ⁴⁵⁷E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices", *J. Comput. Phys.* **17**, 87–94 (1975).
- ⁴⁵⁸H.-J. Werner and W. Meyer, "A quadratically convergent multiconfiguration–self-consistent field method with simultaneous optimization of orbitals and CI coefficients", *J. Chem. Phys.* **73**, 2342–2356 (1980).
- ⁴⁵⁹H.-J. Werner and P. J. Knowles, "A second order multiconfiguration SCF procedure with optimum convergence", *J. Chem. Phys.* **82**, 5053–5063 (1985).

- ⁴⁶⁰Q. Sun, J. Yang, and G. K.-L. Chan, "A general second order complete active space self-consistent-field solver for large-scale systems", *Chem. Phys. Lett.* **683**, 291–299 (2017).
- ⁴⁶¹D. A. Kreplin, P. J. Knowles, and H.-J. Werner, "Second-order MCSCF optimization revisited. i. improved algorithms for fast and robust second-order casscf convergence", *J. Chem. Phys.* **150**, 194106 (2019).
- ⁴⁶²D. A. Kreplin, P. J. Knowles, and H.-J. Werner, "MCSCF optimization revisited. ii. combined first- and second-order orbital optimization for large molecules", *J. Chem. Phys.* **152**, 074102 (2020).
- ⁴⁶³U. Bozkaya, J. M. Turney, Y. Yamaguchi, H. F. Schaefer, and C. D. Sherrill, "Quadratically convergent algorithm for orbital optimization in the orbital-optimized coupled-cluster doubles method and in orbital-optimized second-order møller-pleeset perturbation theory", *J. Chem. Phys.* **135**, 104103 (2011).
- ⁴⁶⁴T. M. Henderson, I. W. Bulik, T. Stein, and G. E. Scuseria, "Seniority-based coupled cluster theory", *J. Chem. Phys.* **141**, 244104 (2014).
- ⁴⁶⁵J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, NY, New York, NY, USA, 1999).
- ⁴⁶⁶D. A. Matthews, L. Cheng, M. E. Harding, F. Lipparini, S. Stopkowicz, T.-C. Jagau, P. G. Szalay, J. Gauss, and J. F. Stanton, "Coupled-cluster techniques for computational chemistry: the CFOUR program package", *J. Chem. Phys.* **152**, 214108 (2020).
- ⁴⁶⁷G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Wloch, P. Xu, F. Zahariev, and M. S. Gordon, "Recent developments in the general atomic and molecular electronic structure system", *J. Chem. Phys.* **152**, 154102 (2020).
- ⁴⁶⁸M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09 Revision E.01*, Gaussian Inc. Wallingford CT 2009.
- ⁴⁶⁹J. M. Foster and S. F. Boys, "Canonical configurational interaction procedure", *Rev. Mod. Phys.* **32**, 300 (1960).
- ⁴⁷⁰C. Angeli, C. J. Calzado, R. Cimiraglia, S. Evangelisti, N. Guihéry, T. Leininger, J.-P. Malrieu, D. Maynau, J. V. P. Ruiz, and M. Sparta, "The use of local orbitals in multireference calculations", *Mol. Phys.* **101**, 1389–1398 (2003).
- ⁴⁷¹C. Angeli, R. Cimiraglia, and M. Cestari, "A multireference n-electron valence state perturbation theory study of the electronic spectrum of s-tetrazine", *Theor. Chem. Acc.* **123**, 287–298 (2009).
- ⁴⁷²N. Ben Amor, F. Bessac, S. Hoyau, and D. Maynau, "Direct selected multireference configuration interaction calculations for large systems using localized orbitals", *J. Chem. Phys.* **135**, 014101 (2011).
- ⁴⁷³N. Suaud and J.-P. Malrieu, "Natural molecular orbitals: limits of a Lowdin's conjecture", *Mol. Phys.* **115**, 2684–2695 (2017).

- ⁴⁷⁴V. G. Chilkuri, T. Applencourt, K. Gasperich, P.-F. Loos, and A. Scemama, "Spin-adapted selected configuration interaction in a determinant basis", *Adv. Quantum Chem.*, in press (2021).
- ⁴⁷⁵M. Kállay and P. R. Surján, "Higher excitations in coupled-cluster theory", *J. Chem. Phys.* **115**, 2945–2954 (2001).
- ⁴⁷⁶M. Dash, S. Moroni, C. Filippi, and A. Scemama, "Tailoring CIPSI expansions for QMC calculations of electronic excitations: the case study of thiophene", *J. Chem. Theory Comput.* **17**, 3426–3434 (2021).
- ⁴⁷⁷P.-F. Loos, A. Scemama, I. Duchemin, D. Jacquemin, and X. Blase, "Pros and cons of the bethe–salpeter formalism for ground-state energies", *J. Phys. Chem. Lett.* **11**, 3536–3545 (2020).
- ⁴⁷⁸S. D. Sabatino, P.-F. Loos, and P. Romaniello, *Scrutinizing GW-based methods using the hubbard dimer*, 2021.
- ⁴⁷⁹J. A. Berger, P.-F. Loos, and P. Romaniello, "Potential energy surfaces without unphysical discontinuities: the coulomb hole plus screened exchange approach", *J. Chem. Theory. Comput.* **17**, PMID: 33306908, 191–200 (2021).