

**Incremental Scheme: A General
Approach For
Electron Correlation Computations of
Large Molecules**

INAUGURAL-DISSERTATION

zur

Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

JUN ZHANG

aus Tianjin, China

Berichtersteller:
(Gutachter)

Prof. Dr. M. Dolg

PD. Dr. M. Hanrath

Tag der mündlichen Prüfung:

15.01.2015

Abstract

The first part of this work introduces incremental scheme as a general approach for electron correlation computations of large molecules, especially its latest implementation: third-order incremental dual-basis set zero buffer (inc3-db-B0) approach. This approach can combine with CCSD, CCSD(T) and their explicit correlation variants to obtain accurate correlation energies in a highly efficient way, and is presented in detail in this work. A program APTS has been developed for a black-box and automatic implementation of these methods. With various strategies, the inc3-db-B0 approach can reduce the wall time of a calculation of a large molecule by up to 10 times, and the error in absolute and especially relative energies can be less than 1 kcal mol⁻¹, making it a reliable method for the treatment of energetically nearly degenerate isomers of large molecules and other kinds of chemical species. A series of applications of the inc3-db-B0 approach in many real chemical problems are then described, including: benchmark set validation; energies of isomers of water clusters; the rotational barrier of biphenyl; hydration of lanthanide trivalent ions; the relative stability of isomers of double fullerene adducts; singlet-triplet gap of biphenylcarbene, and vertical detachment energy of green fluorescent protein chromophore. These problems involve both inorganic and organic chemistry, closed-shell and open-shell molecules. The inc3-db-B0 approach exhibits excellent performance in various kinds of chemical problems, confirming it a promising method for general chemical problems. Finally, the potential direction of further extension of incremental scheme is discussed.

The second part of this work introduces the idea of labile capping bond phenomenon. For a wide range of trivalent lanthanide ion coordination complexes of tricapped trigonal prism or monocapped square antiprism configurations, the bonds between the central lanthanide ions and the capping ligands are found to violate Badger's rule: they can get weaker as they get shorter. We demonstrate that this observation originates from the screening and repulsion effect of the prism ligands. Both effects enhance as the electric field of the central ion or the softness of the prism ligands increases. Thus for heavier lanthanides despite that the capping bond could be shorter, it is more efficient to be weakened by the prism ligands, being inherently labile. This concept of "labile capping bonds" has been successfully used to interpret many experiments, especially we have built an elegant model to solve a problem in the water exchange kinetics of lanthanide ions that has puzzled investigators for a long time: why the exchange rate reaches a maximum for the middle region, but is low at the beginning and end of the lanthanide series. We also use it to interpret why the twisted square antiprism isomer of some lanthanide complexes exhibits much higher water exchange rate than the square antiprism isomer does. We believe that the labile capping bond phenomenon can offer new insights in understanding chemical problems.

Kurzzusammenfassung

Der erste Teil dieser Arbeit stellt die Inkrementenmethode als eine Methode für Elektronenkorrelationsrechnungen an großen Molekülen vor, besonders im Hinblick auf den ‘*third-order incremental dual-basis set zero buffer*’, im Folgenden inc3-db-B0-Ansatz genannt. Dieser ermöglicht die effiziente Berechnung von Korrelationsenergien mit Hilfe von CCSD, CCSD(T) und deren Verbindung mit expliziten Korrelationsmethoden. Das Programm APTS wurde für die automatische Implementierung dieser Methoden entwickelt. Durch verschiedene Strategien kann mit Hilfe des inc3-db-B0-Ansatzes die Rechenzeit bei großen Molekülen stark reduziert werden. Der Fehler in absoluten - und besonders relativen Energien - beträgt dabei weniger als 1 kcal mol^{-1} , so dass dies auch für die Behandlung von energetisch, oft annähernd großer Moleküle, und die Behandlung von anderen Arten von chemischen Problemen eine zuverlässige Methode darstellt. Die Anwendung der inc3-db-B0-Methode bietet verschiedene Anwendungsmöglichkeiten: Benchmark-Berechnungen, Energie von Isomeren verschiedener Wassercluster, Rotationsbarriere von Biphenyl, der Hydratation dreiwertiger Lanthanoidionen, der relativen Stabilität von der Fullerene-Addukt-Isomeren, dem Singlet-Triplett Abstand von Biphenyl-Carbon und vertikale Ablöseenergie von Chromophor des grün fluoreszierenden Proteins. Diese Probleme betreffen anorganische und organische Chemie, geschlossenschalige und offenschalige Moleküle. Der inc3-db-B0-Ansatz ermöglicht also die Beschreibung verschiedener Arten von chemischen Problemen. Abschließend werden weitere Entwicklungsmöglichkeiten der Inkrementenmethode vorgestellt.

Der zweite Teil der Arbeit stellt die das ‘*labile capping bond*’-Phänomen vor. Für ein breites Spektrum von trivalenten Lanthanoidionenkoordinationskomplexe von dreifach-überkappter prismatischer oder einfach überkappter antiprismatischer Struktur gehorchen die Bindungen zwischen Zentralion und Kappenligand nicht der Regel von Badger: Obwohl sie kürzer werden, werden sie schwächer. Diese Beobachtung wird im Hinblick auf Abschirmungs und Abstoßungseffekte diskutiert. Des weiteren wird dies auf die Härte der Prismaaliganden, sowie das elektrische Feld des Zentralions zurückgeführt. Im Falle der späteren Lanthanoide kann die Bindung zum Kappenligand also geschwächt sein, obwohl dies aufgrund der Bindungslänge nicht zu erwarten wäre. Dieses ‘*labile capping bond*’-Konzept wird erfolgreich zur Erklärung vieler Experimente verwendet, wie etwa der Austauschrate von Wasser im Bereich der Kinetik von Lanthanoidionen: Die Austauschrate erreicht ein Maximum im Bereich der mittelschweren Lanthanoide und nimmt zu der schwereren Lanthanoiden deutlich ab. Dieses Phänomen wird auf Hinblick des ‘*labile capping bond*’-Konzepts untersucht und könnte zu neuen Erkenntnissen im Bereich bestimmter chemischer Problem beitragen.

Contents

Contents	iii
List of Figures	viii
List of Tables	ix
List of Codes	xi
1 Basic Principles of Quantum Chemistry	1
1.1 Introduction	1
1.1.1 Molecular Hamiltonian	1
1.1.2 Wave Function	2
1.1.3 Solving the Schrödinger Equation	3
1.2 Single-particle Approximation	4
1.2.1 Hartree–Fock Method	4
1.2.2 Molecular Orbitals	6
1.3 Electron Correlation	8
1.3.1 Fermi and Coulomb Correlation	8
1.3.2 Cusp Conditions	9
1.3.3 Correlation Energy	10
1.4 Many-body Approaches	10
1.4.1 Full Configuration Interaction Expansion	10
1.4.2 Perturbation Theory	11
1.4.3 Coupled-cluster Ansatz	14
1.4.4 Explicit Correlation Methods	15
1.5 Multi-reference Approaches	16
1.6 Density Functional Theory	17
1.6.1 Hohenberg–Kohn Theorem	17
1.6.2 Kohn–Sham Scheme	18
1.6.3 Explicit Functionals	18
1.7 Molecular Geometry Optimization	19
1.8 Wave Function Analysis	19
1.8.1 Electron Localization Function	19
1.8.2 Topology Analysis of Electron Density	21
1.8.3 Noncovalent Interaction Plot	21
1.9 Linear-scaling Methods	22
1.9.1 Locality of Electron Correlation	22
1.9.2 Linear-scaling Correlation Methods	22

2	Incremental Scheme	25
2.1	Incremental Expansion	25
2.1.1	Introduction	25
2.1.2	Incremental Expansion in Molecules	25
2.1.3	Incremental Expansion for High-spin Open-shell Molecules	27
2.2	K-Means Clustering Algorithm	28
2.3	Dual-basis Set Zero-buffer Approximation	29
2.3.1	Dual-basis Set Technique	29
2.3.2	Zero-buffer Approximation	31
2.4	Distance Screening	31
2.5	Inc3-db-B0 Approach	32
3	Implementation of Incremental Scheme	35
3.1	Introduction of APTS	35
3.1.1	Basis Features	35
3.1.2	Code Structures	36
3.1.3	Compilation	36
3.2	Files of APTS	36
3.2.1	Input Files	36
3.2.2	Output Files	39
3.3	Practical Guides of APTS	39
3.3.1	Perform a Standard Calculation	39
3.3.2	Visualization of Domains	40
3.4	Special Issues for F12 and High-spin Open-shell Calculations	41
3.4.1	Perform a F12 Calculation	41
3.4.2	Perform a High-spin Open-shell Calculation	42
4	Applications of the Incremental Scheme	45
4.1	Systems from Benchmark Sets	45
4.2	Water Clusters	47
4.2.1	Minimum Structure of Water Hexamers	47
4.2.2	Larger Clusters	48
4.3	The Rotation Barrier of Biphenyl	50
4.4	Hydration of Trivalent Lanthanide Ions	51
4.5	The Relative Stability of Isomers of Double Fullerene Adducts	54
4.6	Singlet-triplet Gap of Biphenylcarbene	58
4.7	VDE of the GFP Chromophore	60
5	Theory of Labile Capping Bonds and its Applications	63
5.1	Theory of Labile Capping Bonds	63
5.2	Hydration Kinetics of Trivalent Lanthanide Ions	68
5.3	Hydration Kinetics of Lanthanide(III)-DOTAM Complexes	72
6	Summary and Outlook	75
6.1	Summary	75
6.2	Outlook	75
	Bibliography	77
	Abbreviations and Acronyms	95
	Acknowledgments	97

<i>CONTENTS</i>	v
Erklärung	99
Curriculum Vitae	100

List of Figures

1.2.1	A CMO and Boys-, ER-localized orbital of acetaldehyde.	7
1.8.1	The filled-color map of ELF in the plane of water molecule.	20
1.8.2	The NCI plot of benzene dimer. The red and green surface implies the repulsion and weak vdW interaction, respectively.	22
2.1.1	The errors of incremental expansions for four systems versus the truncation order. The energy is calculated at the CCSD(T)/6-31G(d) level. . .	27
2.2.1	Domain decomposition by KM algorithm. The balls stand for the center of the LMOs and the balls with identical color belong to the same domain.	28
2.3.1	The procedure of projecting LMOs in \mathcal{B}' onto those in \mathcal{B}_X	30
2.3.2	Impact of n_d and buffer on the accuracy and efficiency of the incremental scheme. In the right figure t_X stands for the wall time of a calculation by method X at the same machine.	32
2.3.3	Illustration of dual-basis set zero-buffer approximation. Here a HF computation of the molecule is performed with VTZ(p/s) basis set, obtaining the CMOs. After localization the center of the LMOs are computed and represented by balls, and are clustered into domains by KM algorithm. The balls of identical color belong to the same domain. Then the atoms are clustered by FCKM algorithm to determine the buffer region. The last graph illustrates the B0 approximation: to compute the correlation energy of the orange orbital domain, we apply VTZ basis set on the atoms in the orange atom domain, and VTZ(p/s) for the rest of the molecule.	32
2.4.1	The values of the increments for four systems versus the distance. The energy is calculated at CCSD(T)/6-31G(d) level.	33
3.3.1	Visualization of domains by VMD.	41
4.2.1	Some isomers of the water clusters.	49
4.3.1	Biphenyls. The balls stand for the centers of the LMOs and those with the same color constitute a domain.	51
4.4.1	Geometries of water clusters $(\text{H}_2\text{O})_n$ and lanthanide(III) aqua complexes $\text{Ln}(\text{H}_2\text{O})_n^{3+}$ ($n = 8, 9$).	52
4.4.2	The binding energy errors (unit: kJ mol^{-1}) with respect to the CCSD(T) reference for all the considered methods. The errors of method X here are defined as $D_{\text{CCSD(T)}} - D_X$, where D_X is the binding energy computed by method X	54
4.5.1	Molecules investigated for effects of intramolecular dispersion interaction.	56
4.5.2	Geometries and noncovalent interaction visualization. For the surfaces, red regions mean strong repulsion and green ones mean weak attraction.	58

4.6.1 Left top: Diphenylcarbene. Right top: Diphenylcarbene–methanol complex. Bottom: The effect of the methanol molecule suggested by B3LYP-D3 [228] and inc3-db-B0-CCSD(T).	59
4.7.1 Left: The chromophore part of the GFP. Right: The HBDI ⁻ ·Gmd ⁺ complex.	60
5.1.1 TZDO/TZDS-aqua-Ln ³⁺	63
5.1.2 The lengths, n_{BCP} 's, and ELF at BCP of the Ln–O bonds in TZDO/TZDS-aqua-Ln ³⁺	64
5.1.3 Colour filled maps of the ELF on two planes of TZDX-aqua-Eu ³⁺	66
5.1.4 The lengths, n_{BCP} 's, and ELF at BCP of the Ln–Cl bonds in LnCl(H ₂ O) ₈ ²⁺	67
5.1.5 The geometries of p- and c-LuCl(H ₂ O) ₈ ²⁺	68
5.2.1 Water exchange rate of lanthanides(III). The figure is adapted from Helm's work[247].	69
5.2.2 Molecular graphs of lanthanide(III) aqua complexes. Red-white licorice: water molecule; cyan ball: lanthanide(III) ion; yellow ball: bond critical point; green line: bond path; blue line: no physical meaning, just to guide the eye to recognize the coordination polyhedron. There are no bond paths (green lines) connecting lanthanide(III) ions with oxygen atoms since the implementation of PPs for lanthanides depletes the electron density around the nuclei.	70
5.2.3 The lengths, n_{BCP} 's, and ELF at BCP of the Ln–O bonds in aqua lanthanide(III) complexes.	71
5.2.4 The water exchange mechanism. Other possible steps of association of a water molecule or rearrangement of a BTP intermediate for nona-aqua lanthanides(III) are omitted for clarity. Circle: water molecule; A: association; D: dissociation; R: rearrangement.	72
5.3.1 DOTAM-aqua-Ln ³⁺	73
5.3.2 The lengths, n_{BCP} 's, and ELF at BCP of the Ln–O bonds in DOTAM-aqua-Ln ³⁺ complexes.	73
5.3.3 The energy difference between the MSAP and MTSAP isomer of DOTAM-aqua-Ln ³⁺ (A) and bond lengths of Ln–O(prism) in DOTAM-aqua-Ln ³⁺ (B).	74

List of Tables

2.3.1 Benchmark of dual-basis set and standard CCSD, CCSD(T), and their F12 variants.	30
4.1.1 Efficiency of inc3-db-B0-F12 methods. All the calculations were performed with the Intel(R) Xeon(R) E5-4620 CPU cores and 16 GB RAM per core.	46
4.1.2 Interaction energies by inc3-db-B0-F12 methods. Energy unit: kcal mol ⁻¹	47
4.2.1 Inc3-db-B0-F12 results of water hexamers. The interaction energy is defined as: $E_{cage} - E_{prism}$ with units: kcal mol ⁻¹ . All the calculations were performed with the Intel(R) Xeon(R) CPU E7-8837 CPU cores and 8 GB RAM per core.	48
4.2.2 Errors of DLPNO-, OSV- and inc(2/3)-db-B0-CCSD(T)/VDZ methods.	50
4.3.1 Inc3-db-B0-F12 results for the rotation barrier of biphenyl. The interaction energy is defined as: $E_{44^\circ} - E_{0^\circ}$ with unit: kcal mol ⁻¹ . All the calculations were performed with the Intel(R) Xeon(R) E5-4620 CPU cores and 16 GB RAM per core.	51
4.4.1 The RMSDs (unit: kJ mol ⁻¹) of the binding energy errors with respect to the CCSD(T) reference for all the considered methods.	53
4.4.2 The hydration Gibbs free energies (unit: kJ mol ⁻¹) of lanthanide(III) aqua complexes.	55
4.5.1 Energy differences between <i>syn</i> and <i>anti</i> isomers of 2' to 4' by <i>ab initio</i> correlation methods for TPSS-D3/SVP optimized geometries.	56
4.5.2 Energy differences between <i>syn</i> and <i>anti</i> isomers of all molecules considered by HF and DFT methods for TPSS-D3/SVP optimized geometries.	57
4.6.1 Singlet-triplet gap (unit: kcal mol ⁻¹) of diphenylcarbene and its methanol complex.	59
4.7.1 VDEs of dHBDI (unit: eV) in gas phase and protein environment.	61
5.1.1 Electronic structure parameters of An ³⁺ complexes.	66

List of Codes

3.1	Coordinates of CH ₃ CHO.	36
3.2	An example of node file.	37
3.3	Input file for inc3-db-B0-CCSD(T)/VDZ of CH ₃ CHO.	37
3.4	Output file of the inc3-db-B0-CCSD(T)/VDZ calculation of CH ₃ CHO (modified for typesetting).	40
3.5	TCL script for visualization of domains.	40
3.6	Input file for inc3-db-B0-CCSD(T)-F12/VDZ of CH ₃ CHO.	41
3.7	Output file for inc3-db-B0-CCSD(T)-F12/VDZ of CH ₃ CHO.	42
3.8	Output file for CCSD(T)-F12/VDZ of CH ₃ CHO.	42
3.9	Input file for inc3-db-B0-CCSD(T)/VDZ of a triplet state of CH ₃ CHO.	42
3.10	Output file for inc3-db-B0-CCSD(T)/VDZ of triplet state of CH ₃ CHO.	43
3.11	Input file for standard dual basis set CCSD(T)/VDZ of triplet state of CH ₃ CHO, with only the singly occupied orbitals are correlated.	43
3.12	Output file for the MOLPRO calculation.	44

Chapter 1

Basic Principles of Quantum Chemistry

1.1 Introduction

1.1.1 Molecular Hamiltonian

Quantum chemistry has become a routine tool for scientists to understand a chemical problem in a microscopic way. It applies quantum mechanics as well as other modern physics to chemistry to study the movement of electrons in atoms, molecules, clusters, surface, proteins, periodic systems and bulk matters. The starting point to explore the movement of electrons is the molecular *Hamiltonian*. In this work the *Born–Oppenheimer approximation*[1] (BO) (fixed nuclei) is assumed, and the nuclei and electrons will be treated as point charges. Thus, in the nonrelativistic case, the Hamiltonian of a molecule can be written as (in atomic unit):

$$\begin{aligned}\hat{H} &= \hat{T}_E + \hat{V}_{NE} + \hat{V}_{EE} + \hat{V}_{NN} \\ &\equiv \sum_{i=1}^N \left(-\frac{1}{2} \nabla_i^2 \right) + \sum_{i=1}^N \sum_{A=1}^K \left(-\frac{Z_A}{r_{iA}} \right) + \sum_{\substack{i,j=1 \\ i < j}}^N \frac{1}{r_{ij}} + \sum_{\substack{A,B=1 \\ A < B}}^K \frac{Z_A Z_B}{r_{AB}}\end{aligned}\quad (1.1.1)$$

In (1.1.1), N and K stand for the number of electrons and nuclei, respectively; Z_A is the nuclear charge of the nucleus A ; the four terms represent the electron kinetic energy, electron-nucleus attraction energy, electron-electron repulsion energy and nucleus-nucleus repulsion energy, respectively. For \hat{V}_{NE} , we can extract its form for a single electron $v_{\text{ext}}(\mathbf{r})$, which is often called *external potential*. This term is determined by the molecular framework.

Since \hat{V}_{EE} is a two-electron operator, the molecular quantum mechanics problem becomes a many-body problem, solution of whose equation-of-motion (EOM) is quite nontrivial and becomes the central topic of quantum chemistry.

When a molecule contains heavy atoms (heavier than Fe or $Z_A > 26$), the relativistic effects cannot be neglected. In this case one can use *Dirac–Coulomb Hamiltonian*:

$$\hat{H} = \sum_{i=1}^N c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}}_i + \beta mc^2 + \hat{V}_{NE} + \hat{V}_{EE} + \hat{V}_{NN}\quad (1.1.2)$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \mathbf{0} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & \mathbf{0} \end{pmatrix}\quad (1.1.3)$$

$$\boldsymbol{\beta} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_2 \end{pmatrix}\quad (1.1.4)$$

In (1.1.2), c and m are the speed of light and the rest mass of electron, respectively. The three components of $\boldsymbol{\sigma}$ are Pauli matrices. Note that we do not consider higher-order effects like Breit interaction in V_{EE} yet in (1.1.2).

However this Hamiltonian is quite difficult to work with. Since relativistic effects are most significant in the atomic core region, which is usually chemically inert and thus transferable, one can replace the core part of an atom by an effective one-electron potential and only treat the valence electrons explicitly. This is the *effective core potential* (ECP) approach[2]. It cannot only save the computational cost, but can also take the relativistic and quantum electrodynamic effects, etc. into account in the framework of nonrelativistic theory. In this case the Hamiltonian becomes:

$$\hat{H} = \hat{T}_E + \hat{V}_{EE} + \hat{V}_{CV} + \hat{V}_{CC} \quad (1.1.5)$$

where

$$\hat{V}_{CV} = \sum_{i=1}^{N_V} \sum_{A=1}^K \left(-\frac{Q_A}{r_{iA}} + \Delta V_{CV,A}(\mathbf{r}_i) \right) \quad (1.1.6)$$

$$\hat{V}_{CC} = \sum_{\substack{A,B=1 \\ A < B}}^K \frac{Q_A Q_B}{r_{AB}} \quad (1.1.7)$$

From (1.1.5) to (1.1.7), N_V is the number of valence electrons and Q_A is the core charge of the nucleus A . $\Delta V_{CV,A}(\mathbf{r}_i)$ is the *pseudopotential* (PP) that constitutes the ECP of nucleus A , with the form[3]:

$$\Delta \hat{V}_{CV}(\mathbf{r}) = \sum_{l,m} V_l(r) |lm\rangle \langle lm| \quad (1.1.8)$$

In (1.1.8) $|lm\rangle$ is the spherical harmonics function $Y_{lm}(\Omega)$; $V_l(r)$ is the radial part of the PP, which is often written as a linear combination of Gaussian functions multiplied with powers of r [4]:

$$V_l(r) = \sum_k A_{lk} r^{n_{lk}} \exp(-\alpha_{lk} r^2) \quad (1.1.9)$$

In this work we will only consider the Hamiltonian (1.1.1) and (1.1.5).

1.1.2 Wave Function

In quantum mechanics, the motion of an electron is described by its *wave function* (WFN) $\psi(\mathbf{x}, t)$, where \mathbf{x} is a collective variable of spatial and spin coordinates $\mathbf{x} \equiv (\mathbf{r}, s)$ and t is time. For *stationary state* considered in this work ψ is independent of time. The Copenhagen school interprets the WFN as *probability amplitude*, that is, the probability density of finding the particle at \mathbf{x} at time t is $|\psi(\mathbf{x}, t)|^2$.

An exact molecular electron WFN must satisfy the following conditions:

- ψ is a function of N electrons, or *N -representable*.
- For bound states, ψ has to be *square-integrable* (belongs to Hilbert space $L^2(\mathbb{R}^{3N})$), thus normalizable:

$$\langle \psi | \psi \rangle = 1 \quad (1.1.10)$$

- Since electrons are Fermions, ψ has to be *antisymmetric* with respect to the permutation of two electronic indices:

$$\mathcal{P}_{ij} \psi = -\psi \quad (1.1.11)$$

- ψ has to satisfy the spin symmetry, being an eigenfunction of the total and projected spin operators:

$$\hat{S}^2\psi = S(S+1)\psi \quad (1.1.12)$$

$$\hat{S}_z\psi = M\psi \quad (1.1.13)$$

- ψ has to also satisfy the spatial symmetry, being a basis of one irreducible representation of the point group to which the molecule belongs.
- Due to the existence of $1/r$ -like operators in the molecular Hamiltonian, it has to satisfy specific *cusp conditions*, which will be discussed in Subsection 1.3.2.
- At long range the electron density related to ψ

$$n(\mathbf{r}) \equiv N \int d\mathbf{s} d\mathbf{x}_2 \cdots d\mathbf{x}_N |\psi(\mathbf{r}, \mathbf{s}, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 \quad (1.1.14)$$

decays as[5]:

$$n(\mathbf{r}) \rightarrow \exp(-\sqrt{8I}r) \quad (r \rightarrow \infty) \quad (1.1.15)$$

where I is the first ionization potential of the molecule.

Of course this list is far from complete, but a good approximation of the exact WFN should try to satisfy the above conditions as close as possible.

In practice, an approximate WFN is constructed from a set of simple analytical functions, like plane waves or Gaussian functions, which is usually termed as *basis set*.

1.1.3 Solving the Schrödinger Equation

To obtain the eigenfunction and eigenvalue of the Hamiltonian, we should solve the *time-independent Schrödinger equation* (SE)[6]:

$$\hat{H}\psi = E\psi \quad (1.1.16)$$

Except for a few cases, the SE is impossible to be solved analytically. At this stage the *variation principle*[7] can be established: the solution of SE is equivalent to the stationary of the energy functional

$$E[\phi] = \frac{\langle \phi | \hat{H} | \phi \rangle}{\langle \phi | \phi \rangle} \quad (1.1.17)$$

Proof. Let ϕ be an exact WFN that satisfies (1.1.16). Assume a variation $\tilde{\phi} = \phi + \delta\phi$ then

$$\begin{aligned} E[\phi + \delta\phi] &= \frac{\langle \phi | \hat{H} | \phi \rangle + \langle \delta\phi | \hat{H} | \phi \rangle + \langle \phi | \hat{H} | \delta\phi \rangle + \langle \delta\phi | \hat{H} | \delta\phi \rangle}{\langle \phi | \phi \rangle + \langle \delta\phi | \phi \rangle + \langle \phi | \delta\phi \rangle + \langle \delta\phi | \delta\phi \rangle} \\ &= E + \langle \delta\phi | \hat{H} - E | \phi \rangle + \langle \phi | \hat{H} - E | \delta\phi \rangle + O(\delta\phi^2) \\ &= E + O(\delta\phi^2) \end{aligned} \quad (1.1.18)$$

The energy functional proves to be stationary since it does not change at the first-order.

Conversely, if the energy functional is stationary at ϕ , then according to (1.1.18), for variations $\tilde{\phi} = \phi + \delta\phi$ and $\tilde{\phi} = \phi + i\delta\phi$ we have

$$\langle \delta\phi | \hat{H} - E[\phi] | \phi \rangle + \langle \phi | \hat{H} - E[\phi] | \delta\phi \rangle = 0 \quad (1.1.19)$$

$$\langle \delta\phi | \hat{H} - E[\phi] | \phi \rangle - \langle \phi | \hat{H} - E[\phi] | \delta\phi \rangle = 0 \quad (1.1.20)$$

Adding (1.1.19) to (1.1.20) we obtain:

$$\langle \delta\phi | \hat{H} - E[\phi] | \phi \rangle = 0 \quad (1.1.21)$$

Since $\delta\phi$ is arbitrary we have (1.1.16). Therefore, the stationary points of the energy functional (1.1.17) and the solutions of SE (1.1.16) have a one-to-one mapping. \square

In fact, for the exact ground state WFN ψ and a trial WFN ϕ , we always have

$$E[\psi] \leq E[\phi] \quad (1.1.22)$$

In practice the variation principle is always used by constructing a trial WFN with some parameterization ansätze and minimizing the energy functional (1.1.17). One property of parameterization ansatz is the *size-extensivity*[8]: for a system containing two noninteracting subsystems A and B , the Hamiltonian being

$$\hat{H}_{AB} = \hat{H}_A + \hat{H}_B \quad (1.1.23)$$

then the energy of the total system is additive and the WFN is multiplicative:

$$E_{AB} = E_A + E_B \quad (1.1.24)$$

$$\psi_{AB} = \psi_A \otimes \psi_B \quad (1.1.25)$$

A related concept is *size-consistency*[9]: for a system containing two subsystems A and B , when A and B are very far apart, the energy is additive:

$$E_{AB} = E_A + E_B \quad (1.1.26)$$

The size-extensivity and consistency ensure that the accuracy of energies of large and small systems can be consistent, otherwise the error for the larger system tends to be larger, leading to a relative energy of bad quality.

1.2 Single-particle Approximation

1.2.1 Hartree–Fock Method

The first step of approximating an exact N -electron molecular WFN is assuming it be the product of N one-electron WFNs:

$$\psi_{\text{Hartree}}(\mathbf{x}^N) = \prod_{i=1}^N \phi_i(\mathbf{x}_i) \quad (1.2.1)$$

This is called *Hartree product*[10, 11]. The ϕ_i describes the motion of electron i thus it is called *molecular orbital* (MO). To make it satisfy the antisymmetry requirement (1.1.11), an antisymmetrizer \mathcal{A} can be applied, leading to a Slater determinant (SD)[12]:

$$\psi_{\text{SD}}(\mathbf{x}^N) = |\phi_1 \cdots \phi_N\rangle \equiv \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \cdots & \phi_1(\mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{x}_1) & \cdots & \phi_N(\mathbf{x}_N) \end{vmatrix} \quad (1.2.2)$$

A single SD can be used as the trial WFN, which is the *Hartree–Fock* (HF) WFN[13]:

$$|\text{HF}\rangle = \prod_{i=1}^N a_i^\dagger |\text{vac}\rangle \quad (1.2.3)$$

Now we can optimize the energy functional (1.1.17) with respect to the MOs. To ensure $|\text{HF}\rangle$ is normalized, we can restrict that the MOs are orthonormal:

$$\langle \phi_i | \phi_j \rangle = \delta_{ij} \quad (1.2.4)$$

In this case we have to optimize the Lagrangian

$$L[\{\phi\}] = \langle \text{HF} | \hat{H} | \text{HF} \rangle - \sum_{i,j=1}^N \epsilon_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \quad (1.2.5)$$

rather than the energy functional directly. Using the Hamiltonian (1.1.1)

$$E_{\text{HF}} \equiv \langle \text{HF} | \hat{H} | \text{HF} \rangle = \sum_{i=1}^N \langle \phi_i | \hat{h} | \phi_i \rangle + \frac{1}{2} \sum_{i,j=1}^N (\langle \phi_i \phi_j | \phi_i \phi_j \rangle - \langle \phi_i \phi_j | \phi_j \phi_i \rangle) + E_{\text{NN}} \quad (1.2.6)$$

where E_{NN} is the nucleus-nucleus repulsion energy. Now we can compute the variation of (1.2.5):

$$\begin{aligned} \delta L[\{\phi\}] &= \sum_{i=1}^N \langle \delta \phi_i | \hat{h} | \phi_i \rangle + \sum_{i,j=1}^N (\langle \delta \phi_i \phi_j | \phi_i \phi_j \rangle - \langle \delta \phi_i \phi_j | \phi_j \phi_i \rangle) \\ &\quad - \sum_{i,j=1}^N \epsilon_{ij} \langle \delta \phi_i | \phi_j \rangle + \text{c.c.} \\ &= \sum_{i=1}^N \int d\mathbf{x}_1 \delta \phi_i^*(\mathbf{x}_1) \left(\hat{h} \phi_i + \sum_{j=1}^N \left(\int d\mathbf{x}_2 \phi_j^*(\mathbf{x}_2) \frac{1}{r_{12}} \phi_j(\mathbf{x}_2) \right) \phi_i(\mathbf{x}_1) \right. \\ &\quad \left. - \sum_{j=1}^N \left(\int d\mathbf{x}_2 \phi_j^*(\mathbf{x}_2) \frac{1}{r_{12}} \phi_i(\mathbf{x}_2) \right) \phi_j(\mathbf{x}_1) - \sum_{j=1}^N \epsilon_{ij} \phi_j(\mathbf{x}_1) \right) + \text{c.c.} \\ &\equiv \sum_{i=1}^N \int d\mathbf{x}_1 \delta \phi_i^*(\mathbf{x}_1) \left(\left(\hat{h}(1) + \hat{J}(1) - \hat{K}(1) \right) \phi_i(\mathbf{x}_1) - \sum_{j=1}^N \epsilon_{ij} \phi_j(\mathbf{x}_1) \right) \\ &\quad + \text{c.c.} \\ &\equiv \sum_{i=1}^N \int d\mathbf{x}_1 \delta \phi_i^*(\mathbf{x}_1) \left(\hat{f}(1) \phi_i(\mathbf{x}_1) - \sum_{j=1}^N \epsilon_{ij} \phi_j(\mathbf{x}_1) \right) + \text{c.c.} \end{aligned} \quad (1.2.7)$$

In (1.2.7), \hat{J} and \hat{K} are the *Coulomb* and *exchange* operator, respectively. \hat{f} is called *Fock* operator, which is the effective one-electron operator that determines the optimized MOs. Note that Lagrangian (1.2.5) has to be real thus by $L = L^*$ we can prove that $\epsilon_{ij}^* = \epsilon_{ji}$, i.e. ϵ is an Hermite matrix, being diagonalizable. Therefore we can find a unitary transformation \mathbf{U} of the MOs $\{\phi\}$

$$\tilde{\phi}_i = \sum_{j=1}^N \phi_j U_{ji} \quad (1.2.8)$$

$$\phi_i = \sum_{j=1}^N \tilde{\phi}_j U_{ij}^* \quad (1.2.9)$$

that diagonalizes ϵ

$$\sum_{k,l=1}^N U_{ki}^* \epsilon_{kl} U_{lj} = \delta_{ij} \epsilon_i \quad (1.2.10)$$

Substitute (1.2.9) into (1.2.7), and use (1.2.10), let $\delta L = 0$. Note that $\{\delta\phi\}$ is arbitrary, we obtain the *canonical Hartree–Fock equation* (tildes removed):

$$\hat{f}\phi_i = \epsilon_i \phi_i \quad (i = 1, \dots, N) \quad (1.2.11)$$

Since \hat{f} depends on $\{\phi\}$, (1.2.11) has to be solved iteratively. Thus it is also called self-consistent field (SCF) method. Except for atoms and diatomic molecules, (1.2.11) usually cannot be solved numerically. To transform it into an algebraic problem, given a basis set $\{\chi\}$ is used (also known as atomic orbitals (AOs)) that expands MOs:

$$\phi_p = \sum_{\mu} \chi_{\mu} C_{\mu p} \quad (1.2.12)$$

Note that we use subindex p instead of i , implying that ϕ_p is not necessarily an occupied orbital. We will always use i, j, k, l to stand for occupied orbitals, a, b, c, d for unoccupied (virtual) orbitals and p, q, r, s for arbitrary orbitals.

By substituting (1.2.12) into (1.2.11) and taking the inner product with χ^* , we obtain the *Roothaan–Hall equation*[14, 15]:

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (1.2.13)$$

$$F_{\mu\nu} = \langle \mu | \hat{h} | \nu \rangle + \sum_{\rho\tau} D_{\rho\tau} (\langle \mu\rho | \nu\tau \rangle - \langle \mu\rho | \tau\nu \rangle) \quad (1.2.14)$$

$$D_{\mu\nu} = \sum_i C_{\mu i}^* C_{\nu i} \quad (1.2.15)$$

$$S_{\mu\nu} = \langle \mu | \nu \rangle \quad (1.2.16)$$

Here \mathbf{D} is the *density matrix* in AO basis. This equation is much easier to solve. The obtained orbitals from (1.2.14) are called *canonical MOs* (CMOs).

1.2.2 Molecular Orbitals

The Hartree–Fock method is not only the cornerstone of advanced *ab initio* methods, but also has its own value. For the Hilbert space spanned by a basis set $\{\chi\}$, the MO basis can be transformed into another one by a unitary transformation (preserving their inner products), and the WFN constituted by these two bases can be related as:

$$|\tilde{\psi}\rangle = \exp(-\hat{\kappa}) |\psi\rangle \equiv \exp\left(-\sum_{pq} \kappa_{pq} a_p^\dagger a_q\right) |\psi\rangle \quad (1.2.17)$$

Now we look at the HF energy with transformed orbitals:

$$\begin{aligned} E(\boldsymbol{\kappa}) &= \langle \text{HF} | \exp(\hat{\kappa}) \hat{H} \exp(-\hat{\kappa}) | \text{HF} \rangle \\ &= E_{\text{HF}} + \langle \text{HF} | [\hat{\kappa}, \hat{H}] | \text{HF} \rangle + \dots \\ &= E_{\text{HF}} + \sum_{pq} \langle \text{HF} | [a_p^\dagger a_q, \hat{H}] | \text{HF} \rangle \kappa_{pq} + \dots \end{aligned} \quad (1.2.18)$$

If the energy E is already stationary, we will have

$$0 = \left. \frac{\partial E(\boldsymbol{\kappa})}{\partial \kappa_{pq}} \right|_{\boldsymbol{\kappa}=\mathbf{0}} = \langle \text{HF} | [a_p^\dagger a_q, \hat{H}] | \text{HF} \rangle \quad (1.2.19)$$

For nonredundant rotations, e.g. $pq = ai$, we have

$$\langle \text{HF} | [a_a^\dagger a_i, \hat{H}] | \text{HF} \rangle = \left\langle \begin{array}{c} a \\ i \end{array} \middle| \hat{H} | \text{HF} \rangle = f_{ai} = 0 \quad (1.2.20)$$

which is known as *Brillouin theorem*[13].

For redundant rotations, i.e. p, q are both occupied or unoccupied orbitals, $\partial E(\boldsymbol{\kappa}) / \partial \kappa_{pq}$ and hence higher-order terms involving these κ_{pq} will always be zero due to the Pauli's principle, indicating that *orbital rotations within occupied or virtual space do not change the WFN and its energy*. Thus it is possible to transform the occupied MOs obtained from (1.2.14) into a chemically more meaningful set without changing the WFN and energy.

Since CMOs are orthogonal and symmetry-adapted to the molecular point group, they are often very delocalized, thus one can rotate them to obtain localized MOs (LMO) with specific criteria. Two typical choices are:

- *Boys criterion*[16], which aims at minimizing the spatial extent:

$$L_{\text{Boys}}[\{\phi\}] = \max : \sum_i^{\text{occ}} \langle \phi_i \phi_i | (\mathbf{r}_1 - \mathbf{r}_2) | \phi_i \phi_i \rangle \quad (1.2.21)$$

- *Edmiston–Ruedenberg criterion* (ER)[17], which aims at maximizing the self-repulsion energy:

$$L_{\text{ER}}[\{\phi\}] = \max : \sum_i^{\text{occ}} \langle \phi_i \phi_i | \phi_i \phi_i \rangle \quad (1.2.22)$$

One advantage of ER over Boys criterion is that the former can preserve the σ - π symmetry of the MOs[18], but both have a very good localization effect. An example of CMO and LMO can be seen in Figure 1.2.1.

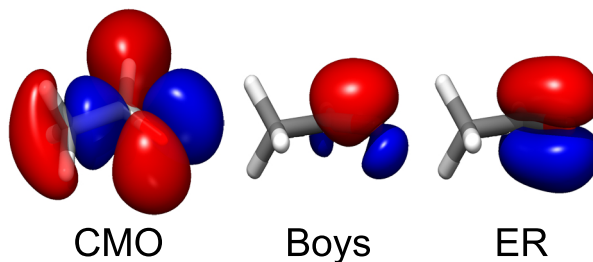


Figure 1.2.1: A CMO and Boys-, ER-localized orbital of acetaldehyde.

Localization is usually performed in an iterative way[17]. In each iteration, every two orbitals are rotated with an optimized angle that satisfies (1.2.21) or (1.2.22):

$$\begin{pmatrix} \tilde{\phi}_i \\ \tilde{\phi}_j \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \phi_i \\ \phi_j \end{pmatrix} \quad (1.2.23)$$

$$\tan 4\theta = -\frac{A_{ij}}{B_{ij}} \quad (1.2.24)$$

$$A_{ij} = \langle \phi_i \phi_i | \hat{\Gamma} | \phi_i \phi_j \rangle - \langle \phi_j \phi_j | \hat{\Gamma} | \phi_i \phi_j \rangle \quad (1.2.25)$$

$$B_{ij} = \langle \phi_i \phi_j | \hat{\Gamma} | \phi_i \phi_j \rangle - \frac{1}{4} \left(\langle \phi_i \phi_i | \hat{\Gamma} | \phi_i \phi_i \rangle + \langle \phi_j \phi_j | \hat{\Gamma} | \phi_j \phi_j \rangle - 2 \langle \phi_i \phi_i | \hat{\Gamma} | \phi_j \phi_j \rangle \right) \quad (1.2.26)$$

where $\hat{\Gamma}$ is the operator related to (1.2.21) or (1.2.22). The iteration continues until convergence is reached.

This method is effective for occupied orbitals of most molecules. For virtual orbitals or very delocalized systems (like graphene), this method can be very difficult to converge[19]. In this case one can explicitly optimize the orbital rotation parameters in (1.2.17) to force convergence[20].

1.3 Electron Correlation

1.3.1 Fermi and Coulomb Correlation

With a molecular WFN, we can define the *one-electron* and *pair probability density*:

$$P(\mathbf{r}) \equiv \int ds_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N |\psi(\mathbf{r}s_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)|^2 \quad (1.3.1)$$

$$P(\mathbf{r}, \mathbf{r}') \equiv \int ds_1 ds_2 d\mathbf{x}_3 \cdots d\mathbf{x}_N |\psi(\mathbf{r}s_1, \mathbf{r}'s_2, \mathbf{x}_3, \cdots, \mathbf{x}_N)|^2 \quad (1.3.2)$$

as well as a *conditional probability density*:

$$P(\mathbf{r}|\mathbf{r}') \equiv \frac{P(\mathbf{r}, \mathbf{r}')}{P(\mathbf{r}')} \quad (1.3.3)$$

For two electrons 1 and 2, their motion is said to be *uncorrelated* if

$$P(\mathbf{r}_1|\mathbf{r}_2) = P(\mathbf{r}_1) \quad (1.3.4)$$

Take the ${}^3\Sigma_u^+$ state of a two-electron molecule H_2 as an example. Assuming it contains only two orbitals: σ_g and σ_u , we can see that for a Hartree product WFN (1.2.1):

$$P(\mathbf{r}_1|\mathbf{r}_2) = P(\mathbf{r}_1) = |\sigma_g(\mathbf{r}_1)|^2 \quad (1.3.5)$$

the two electrons are uncorrelated.

For a SD WFN $|\sigma_g\sigma_u\rangle$, we have

$$P(\mathbf{r}_1) = \frac{1}{2} \left(|\sigma_g(\mathbf{r}_1)|^2 + |\sigma_u(\mathbf{r}_1)|^2 \right) \quad (1.3.6)$$

$$P(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2} \left(|\sigma_g(\mathbf{r}_1)|^2 |\sigma_u(\mathbf{r}_2)|^2 + |\sigma_u(\mathbf{r}_1)|^2 |\sigma_g(\mathbf{r}_2)|^2 - \sigma_u^*(\mathbf{r}_1) \sigma_g(\mathbf{r}_1) \sigma_g^*(\mathbf{r}_2) \sigma_u(\mathbf{r}_2) \right) \quad (1.3.7)$$

Therefore there is electron correlation generated from the antisymmetry requirement of SD, known as *Fermi correlation*. In this case, the Fermi correlation lowers the probability density of finding an electron at \mathbf{r}_1 and an electron at \mathbf{r}_2 , thus it is called *Fermi hole*. For the ${}^1\Sigma_u^+$ state, the Fermi correlation increases this probability, thus it is a *Fermi heap*.

Due to the two-electron Coulomb interactions, a single SD WFN (1.2.2) is far from being accurate although it is a good zero-order approximation. An exact WFN should contain an infinite number of SDs, written as

$$|\psi\rangle = |\text{HF}\rangle + |\text{excited SDs}\rangle \quad (1.3.8)$$

The extra terms also introduce electron correlation, known as *Coulomb correlation*. Recovering Coulomb correlation is critical to get a chemically useful WFN and energy, however it is also quite difficult.

1.3.2 Cusp Conditions

The $1/r$ -like operators in Hamiltonians (1.1.1), (1.1.2) and (1.1.5) are singular at $r = 0$. To keep the eigenvalue of \hat{H} or the energy finite, the kinetic term, i.e. $\nabla^2\psi$ -like term, should also be singular. This implies at $r = 0$ the $\nabla\psi$ might be discontinuous. In fact, Kato stated the famous *cusp condition*[21] in 1957 that for a nonrelativistic Hamiltonian (1.1.1), the first-order derivative of an exact eigenfunction must be discontinuous at the singular points of the Coulomb operators.

This can be understood by analyzing the behavior of SE at the neighborhood of a point where two particles of charge q_1 and q_2 meet, their reduced mass being μ . In this region we can expand the WFN:

$$\psi(\mathbf{r}) = \sum_{lm} \sum_{k=0} r^{l+k} f_{lm}^k Y_{lm}(\Omega) \quad (1.3.9)$$

and only the Hamiltonian involving the two particles becomes significant. Then:

$$\begin{aligned} E\psi(\mathbf{r}) &= \left(-\frac{1}{2\mu} \nabla^2 + \frac{q_1 q_2}{r} \right) \psi(\mathbf{r}) \\ &= \left(-\frac{1}{2\mu} \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{2\mu} \frac{\hat{L}^2}{r^2} + \frac{q_1 q_2}{r} \right) \sum_{lm} \sum_{k=0} r^{l+k} f_{lm}^k Y_{lm}(\Omega) \end{aligned} \quad (1.3.10)$$

After rearrangement and taking an inner product with $Y_{lm}^*(\Omega)$:

$$\begin{aligned} &\left(\frac{l+1}{\mu} f_{lm}^1 - q_1 q_2 f_{lm}^0 \right) r^{l-1} \\ &+ \sum_{k=0} \left(E f_{lm}^k - q_1 q_2 f_{lm}^{k+1} + \frac{1}{2\mu} (k+2)(k+2l+3) f_{lm}^{k+2} \right) r^{l+k} = 0 \end{aligned} \quad (1.3.11)$$

Since r is arbitrary, we have

$$f_{lm}^1 = \frac{\mu q_1 q_2}{l+1} f_{lm}^0 \quad (1.3.12)$$

Also note that

$$f_{lm}^k = \frac{1}{(l+k)!} \left(\widetilde{\frac{\partial^{l+k}\psi}{\partial r^{l+k}}} \right)^{lm} \Big|_{r=0} \quad (1.3.13)$$

where $\widetilde{(\)}^{lm}$ means projection of angular coordinates onto $|lm\rangle$.

For two electrons, $\mu = 1/2$ and $q_1 = q_2 = -1$. If the WFN is totally symmetric, then (1.3.12) holds for $l = 0$:

$$\frac{\partial\psi}{\partial r} \Big|_{r=0} = \frac{1}{2} \psi(r=0) \quad (1.3.14)$$

or

$$\psi(\mathbf{r}) = \left(1 + \frac{1}{2} r \right) \psi(r=0) + O(r^2) \quad (1.3.15)$$

(1.3.14) or (1.3.15) are often called *s-wave coalescence condition* or simply *electron cusp condition*. For triplet symmetry, $l = 1$, we have *p-wave coalescence condition*:

$$\psi(\mathbf{r}) = \left(1 + \frac{1}{4} r \right) \mathbf{r} \cdot \nabla \psi(r=0) + O(r^3) \quad (1.3.16)$$

For WFN of other symmetries or relativistic Hamiltonians, different coalescence conditions are required[22, 23].

The cusp condition poses a tough requirement on the basis set by which a WFN is constructed. Since an exact WFN contains odd powers of the inter-electron distance, basis functions of very high angular momentum and a large number of many-particle basis functions are required to model this behavior.

1.3.3 Correlation Energy

According to Löwdin[24], the *correlation energy* is defined as the difference between the exact nonrelativistic energy and the HF energy (both computed in a complete basis set (CBS)):

$$E_{\text{corr}} = E - E_{\text{HF}} \quad (1.3.17)$$

In practice, quantum chemistry is often used to compute relative energies, and the magnitude of correlation energy is often the largest one. For a typical relative electronic energy of a molecule containing only light elements, the correlation part usually amounts to 1%–2%. Other parts, like relativistic effect and diagonal BO correction, are less than 0.2%. Thus, a large amount of effort has been invested to efficiently obtain accurate correlation energies.

1.4 Many-body Approaches

1.4.1 Full Configuration Interaction Expansion

Linear Parameterization Ansatz

The HF WFN (1.2.6), i.e. a single SD (1.2.2), is not an exact solution of the SE (1.1.16). To approach the exact WFN more many-particle basis functions are needed. In the solution of Roothaan–Hall equation (1.2.14), we obtain not only occupied orbitals \mathbb{O} but also virtual (unoccupied) orbitals \mathbb{V} . The many-particle basis $\{|\mu\rangle\}$ can then be constructed systematically by excitation of 1, 2, \dots , N electrons from the HF WFN:

$$\left| \begin{array}{c} a \\ i \end{array} \right\rangle \equiv a_a^\dagger a_i |\text{HF}\rangle = \{a_a^\dagger a_i\} |\text{HF}\rangle \quad (1.4.1)$$

$$\left| \begin{array}{c} ab \\ ij \end{array} \right\rangle \equiv a_a^\dagger a_b^\dagger a_j a_i |\text{HF}\rangle = \{a_a^\dagger a_i a_b^\dagger a_j\} |\text{HF}\rangle \quad (1.4.2)$$

where $\{\}$ denotes the normal order relative to the *Fermi vacuum*, i.e. HF WFN. The exact WFN can be approximated by the linear combination of all the SDs:

$$\begin{aligned} |\text{FCI}\rangle &= |\text{HF}\rangle + \sum_{ai} c_i^a \{a_a^\dagger a_i\} |\text{HF}\rangle + \frac{1}{4} \sum_{abij} c_{ij}^{ab} \{a_a^\dagger a_i a_b^\dagger a_j\} |\text{HF}\rangle + \dots \\ &\equiv (1 + \hat{C}_1 + \hat{C}_2 + \dots) |\text{HF}\rangle \end{aligned} \quad (1.4.3)$$

This is the *full configuration interaction* (FCI) WFN, where a linear parameterization ansatz is implemented. In practice a truncated version of (1.4.3) has to be used. However a truncated CI WFN suffers from lacking size-extensivity. This can be seen from two noninteracting systems A and B described by CI doubles (only doublet excitations are considered, CID), the direct product of $|\text{CID}\rangle_A$ and $|\text{CID}\rangle_B$ contains quadruple excitations, which is not taken in account in $|\text{CID}\rangle_{AB}$.

Determination of parameters in the CI WFN is a seemingly trivial task. Assuming the WFN is real:

$$|\text{CI}\rangle = \sum_{\mu} c_{\mu} |\mu\rangle \quad (1.4.4)$$

$$E(\mathbf{c}) = \frac{\langle \text{CI} | \hat{H} | \text{CI} \rangle}{\langle \text{CI} | \text{CI} \rangle} = \frac{\sum_{\mu\nu} c_{\mu} c_{\nu} H_{\mu\nu}}{\sum_{\mu} c_{\mu}^2} \quad (1.4.5)$$

$$g_{\mu}(\mathbf{c}) \equiv \frac{\partial E}{\partial c_{\mu}} \quad \mathbf{g} = \frac{2(\mathbf{H} - E(\mathbf{c})\mathbf{I})\mathbf{c}}{\mathbf{c}^T \mathbf{c}} \quad (1.4.6)$$

$$G_{\mu\nu}(\mathbf{c}) \equiv \frac{\partial^2 E}{\partial c_\mu \partial c_\nu} \quad \mathbf{G} = \frac{2 \left(\mathbf{H} - E(\mathbf{c}) \mathbf{I} - \mathbf{g}(\mathbf{c}) \mathbf{c}^T - \mathbf{c} \mathbf{g}(\mathbf{c})^T \right)}{\mathbf{c}^T \mathbf{c}} \quad (1.4.7)$$

Using $\mathbf{g} = \mathbf{0}$, we can obtain the *scalar equation*:

$$\mathbf{H}\mathbf{c} = E\mathbf{c} \quad (1.4.8)$$

Davidson Iteration

Equation (1.4.8) can be solved by, say Jacobi–Givens iteration. However, for large systems, it is sometimes even not possible to store the whole \mathbf{H} in memory, thus one is always trying to avoid the explicit construction of \mathbf{H} . If the solving procedure only involves $\mathbf{H}\mathbf{c}$ then the avoidance is possible. An additional advantage is that we need not to calculate all the eigenvalues but only some interested ones, like ground and lowest excited states.

A popular method is the *Davidson iteration*[25]. We need a reference \mathbf{H}_0 whose eigenvector \mathbf{c}_0 and eigenvalue E_0 are easy to compute (usually the diagonal part of \mathbf{H}). Consider the required solution $\mathbf{c} = \mathbf{c}_0 + \mathbf{d}$, and a Taylor expansion of energy (1.4.5) at \mathbf{c}_0 :

$$E(\mathbf{c}_0 + \mathbf{d}) = E_0 + \mathbf{d}^T \mathbf{g}(\mathbf{c}_0) + \frac{1}{2} \mathbf{d}^T \mathbf{G}(\mathbf{c}_0) \mathbf{d} + \dots \quad (1.4.9)$$

Truncate (1.4.9) at second-order, and optimized E with \mathbf{d} , we obtain

$$\begin{aligned} \mathbf{d} &= -\mathbf{G}(\mathbf{c}_0)^{-1} \mathbf{g}(\mathbf{c}_0) \\ &= -\left(\mathbf{H} - E_0 \mathbf{I} - \mathbf{g}(\mathbf{c}_0) \mathbf{c}_0^T - \mathbf{c}_0 \mathbf{g}(\mathbf{c}_0)^T \right)^{-1} (\mathbf{H} - E_0 \mathbf{I}) \mathbf{c}_0 \end{aligned} \quad (1.4.10)$$

If \mathbf{c}_0 is a good approximation, then $\mathbf{g}(\mathbf{c}_0) \approx \mathbf{0}$, and we replace \mathbf{H} in the inverse by \mathbf{H}_0 , we obtain the final iteration formula:

$$\mathbf{d} = -(\mathbf{H}_0 - E_0 \mathbf{I})^{-1} (\mathbf{H} - E_0 \mathbf{I}) \mathbf{c}_0 \quad (1.4.11)$$

This is much more efficient than the diagonalization of \mathbf{H} (1.4.8). Note that in (1.4.11), the convergence slows down if \mathbf{H}_0 and \mathbf{H} are close to each other. In fact:

$$\mathbf{d} \rightarrow -\mathbf{c}_0 \quad \text{as} \quad \mathbf{H}_0 \rightarrow \mathbf{H} \quad (1.4.12)$$

Convergence to the Exact Solution of the Schrödinger Equation

Finally we want to give some results from functional analysis. If the (one-particle) basis set is complete in the Hilbert space $L^2(\mathbb{R}^3)$, then the set of all possible SDs forms a complete many-particle basis in the Hilbert space $L^2(\mathbb{R}^{3N})$. Therefore a complete basis set can in principle be used to obtain convergence towards the exact WFN to any accuracy in the norm $\sqrt{\langle f|f \rangle}$ (i.e., not pointwisely convergent). However, to ensure the solutions of FCI (1.4.8) converge to the exact WFNs, it has been proved[26, 27] that the basis set has to be complete in the first Sobolev space (the space of functions where the function and its first-order derivative are square-integrable). This is obviously a consequence of the kinetic operator \hat{T}_e .

1.4.2 Perturbation Theory

Formal Theory

Since HF WFN is an eigenfunction of the effective Hamiltonian

$$\hat{H}_0 = \sum_{i=1}^N \hat{f}_i \quad (1.4.13)$$

with eigenvalue $E_0^{(0)} = \sum_{i=1}^N \epsilon_i$, then it is also possible to improve HF ($|\text{HF}\rangle \equiv |0\rangle$) using $\{|\mu\rangle\}$ by perturbation expansion of an *fluctuation potential*:

$$\hat{W} \equiv \hat{H} - \hat{H}_0 \quad (1.4.14)$$

Here we give a general derivation of perturbation theory[28]. For exact WFN ψ and its SE, we project it onto the HF WFN:

$$\langle 0 | \hat{H}_0 | \psi \rangle + \langle 0 | \hat{W} | \psi \rangle = E \langle 0 | \psi \rangle \quad (1.4.15)$$

at this step it is convenient to assume *intermediate normalization*: $\langle 0 | \psi \rangle = 1$ so that

$$\Delta E = \langle 0 | \hat{W} | \psi \rangle \quad (1.4.16)$$

where $\Delta E = E - E_0^{(0)}$. This is the energy expression.

To derive the WFN expression, we introduce a projection operator

$$\hat{P} = |0\rangle \langle 0| \quad (1.4.17)$$

$$\hat{Q} = \sum_{\mu}' |\mu\rangle \langle \mu| \quad (1.4.18)$$

where “'” means that the summation does not include $|0\rangle$. Then we rearrange SE, and add $E_0^{(0)}$ to both sides, and apply \hat{Q} :

$$\hat{Q} \left(E_0^{(0)} - \hat{H}_0 \right) \psi = \hat{Q} \left(E_0^{(0)} + \hat{W} - E \right) \psi \quad (1.4.19)$$

Note that \hat{Q} commutes with \hat{H}_0 and is idempotent, the left operator can completely work in Q -space:

$$\hat{Q} \left(E_0^{(0)} - \hat{H}_0 \right) \hat{Q} \psi = \hat{Q} \left(E_0^{(0)} + \hat{W} - E \right) \psi \quad (1.4.20)$$

Due to the nature of \hat{Q} , the inverse operator in the Q -space (means $\hat{A}\hat{B} = \hat{B}\hat{A} = \hat{Q}$) of $\hat{Q} \left(E_0^{(0)} - \hat{H}_0 \right) \hat{Q}$ exists, which is written as

$$\hat{R}_0 \equiv \frac{\hat{Q}}{E_0^{(0)} - \hat{H}_0} \quad (1.4.21)$$

this is the *resolvent* of \hat{H}_0 . Apply it to (1.4.20):

$$\hat{Q} \psi = \hat{R}_0 \left(E_0^{(0)} + \hat{W} - E \right) \psi \quad (1.4.22)$$

and using $\psi = \hat{Q} \psi + |0\rangle$

$$\psi = |0\rangle + \hat{R}_0 \left(E_0^{(0)} + \hat{W} - E \right) \psi \quad (1.4.23)$$

This expression can be used in a Picard iteration way:

$$|\psi\rangle = \sum_{m=0}^{+\infty} \left(\hat{R}_0 \left(E_0^{(0)} + \hat{W} - E \right) \right)^m |0\rangle \quad (1.4.24)$$

Using (1.4.16), we have *Rayleigh-Schrödinger perturbation theory* (RSPT) formulae:

$$\Delta E = \sum_{m=0}^{\infty} \langle 0 | \hat{W} \left(\hat{R}_0 \left(\hat{W} - \Delta E \right) \right)^m |0\rangle \quad (1.4.25)$$

$$|\psi\rangle = \sum_{m=0}^{\infty} \left(\hat{R}_0 \left(\hat{W} - \Delta E \right) \right)^m |0\rangle \quad (1.4.26)$$

For (1.4.25), we again use Picard iteration to get an explicit expression of ΔE .

$$\begin{aligned} \Delta E &= \langle 0 | \hat{W} | 0 \rangle + \langle 0 | \hat{W} \hat{R}_0 \hat{W} | 0 \rangle + \langle 0 | \hat{W} \hat{R}_0 \left(\hat{W} - \langle 0 | \hat{W} | 0 \rangle \right) \hat{R}_0 \hat{W} | 0 \rangle + \cdots \\ &\equiv E_0^{(1)} + E_0^{(2)} + E_0^{(3)} + \cdots \end{aligned} \quad (1.4.27)$$

In fact,

$$E_0^{(1)} = -\frac{1}{2} \sum_{ij} \langle ij | |ij\rangle \quad (1.4.28)$$

Adding $E_0^{(1)}$ to $E_0^{(0)}$ one obtains E_{HF} . Define:

$$\hat{V} \equiv \hat{W} - \langle 0 | \hat{W} | 0 \rangle \quad (1.4.29)$$

and note that for any constants c , $\hat{R}_0 c | 0 \rangle = 0$, so

$$E_{\text{corr}} = \langle 0 | \hat{V} \hat{R}_0 \hat{V} | 0 \rangle + \langle 0 | \hat{V} \hat{R}_0 \hat{V} \hat{R}_0 \hat{V} | 0 \rangle + \cdots \quad (1.4.30)$$

Linked-diagram Theorem

Linked-diagram theorem [28, 29] states that, only terms of linked-diagram can contribute to the energy and WFN, i.e.

$$|\psi\rangle = \sum_{m=0}^{+\infty} \left(\hat{R}_0 \hat{V} \right)^m |0\rangle_{\text{L}} \quad (1.4.31)$$

$$E_{\text{corr}} = \sum_{m=1}^{+\infty} \langle 0 | \hat{V} \left(\hat{R}_0 \hat{V} \right)^m |0\rangle_{\text{L}} \quad (1.4.32)$$

As a trivial example, the second-order energy $E_0^{(2)}$ is

$$\begin{aligned} E_0^{(2)} &= \langle 0 | \hat{V} \hat{R}_0 \hat{V} | 0 \rangle \\ &= \text{Diagram: Two vertices on a horizontal line. The left vertex has an incoming arrow from the left and an outgoing arrow to the right. The right vertex has an incoming arrow from the left and an outgoing arrow to the right. A dashed line connects the two vertices. Two curved arrows (one above, one below) connect the two vertices, representing a pair of particles being created and then annihilated.} \\ &= \frac{1}{4} \sum_{abij} \frac{\langle ij | |ab\rangle \langle ab | |ij\rangle}{\epsilon_{ij}^{ab}} \end{aligned} \quad (1.4.33)$$

where $\epsilon_{ij}^{ab} \equiv \epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j$. When adding it to the HF energy, the obtained one is called *second-order Møller–Plesset (MP2) energy*.

A nontrivial example is the fourth-order energy $E_0^{(4)}$:

$$E_0^{(4)} = \langle 0 | \hat{V} \hat{R}_0 \hat{V} \hat{R}_0 \hat{V} \hat{R}_0 \hat{V} | 0 \rangle - \langle 0 | \hat{V} \hat{R}_0 \hat{R}_0 \hat{V} | 0 \rangle \langle 0 | \hat{V} \hat{R}_0 \hat{V} | 0 \rangle \quad (1.4.34)$$

The two terms in (1.4.34) are called *principal* and *renormalization* terms, respectively. In fact, the unlinked diagrams in the first term cancel with the renormalization term.

There are only two unlinked diagrams which are

$$\begin{aligned}
& \text{Diagram 1} + \text{Diagram 2} \\
&= \frac{1}{16} \sum_{\substack{abcd \\ ijkl}} \frac{\langle ij|ab\rangle \langle ab|ij\rangle \langle cd|kl\rangle \langle kl|cd\rangle}{\epsilon_{ij}^{ab} \epsilon_{ijkl}^{abcd} \epsilon_{kl}^{cd}} \\
&+ \frac{1}{16} \sum_{\substack{abcd \\ ijkl}} \frac{\langle ij|ab\rangle \langle ab|ij\rangle \langle cd|kl\rangle \langle kl|cd\rangle}{\epsilon_{ij}^{ab} \epsilon_{ijkl}^{abcd} \epsilon_{ij}^{ab}} \tag{1.4.35} \\
&= \left(\frac{1}{4} \sum_{abij} \frac{\langle ij|ab\rangle \langle ab|ij\rangle}{(\epsilon_{ij}^{ab})^2} \right) \left(\frac{1}{4} \sum_{abij} \frac{\langle ij|ab\rangle \langle ab|ij\rangle}{\epsilon_{ij}^{ab}} \right) \\
&= \text{Diagram 1} \times \text{Diagram 2} = \langle 0|\hat{V}\hat{R}_0\hat{R}_0\hat{V}|0\rangle \langle 0|\hat{V}\hat{R}_0\hat{V}|0\rangle
\end{aligned}$$

which confirms the linked-diagram theorem.

1.4.3 Coupled-cluster Ansatz

Coupled-cluster Equation

A more advanced parameterization ansatz can be constructed from some physical consideration. Define m -body excitation operator:

$$\hat{T}_m \equiv \frac{1}{(m!)^2} \sum_{\substack{a_1 a_2 \dots a_m \\ i_1 i_2 \dots i_m}} t_{i_1 i_2 \dots i_m}^{a_1 a_2 \dots a_m} \left\{ a_{a_1}^\dagger a_{i_1} a_{a_2}^\dagger a_{i_2} \dots a_{a_m}^\dagger a_{i_m} \right\} \tag{1.4.36}$$

Here we use t as parameters, different from c in (1.4.3). In a molecule, \hat{T}_m describes that m electrons being excited *simultaneously* (*connected cluster*) from Fermi vacuum. In a FCI expansion, \hat{T}_m also contribute to the case of two *independent* (*unconnected cluster*) m -excitation as $\frac{1}{2}\hat{T}_m^2$, three *independent* m -excitation as $\frac{1}{3!}\hat{T}_m^3$, etc. Thus its whole contribution is:

$$1 + \hat{T}_m + \frac{1}{2}\hat{T}_m^2 + \frac{1}{3!}\hat{T}_m^3 + \dots = e^{\hat{T}_m} \tag{1.4.37}$$

Note that as \hat{T}_m and \hat{T}_n are commutative, the FCI expansion can be written as a product over all levels of excitation, also known as *coupled-cluster* (CC) expansion[30]:

$$|\text{CC}\rangle = \prod_m e^{\hat{T}_m} |\text{HF}\rangle = \exp\left(\sum_m \hat{T}_m\right) |\text{HF}\rangle \equiv e^{\hat{T}} |\text{HF}\rangle \tag{1.4.38}$$

where \hat{T} is called *cluster operator*.

CC is one of the most successful *ab initio* quantum chemistry methods. Due to the exponential parameterization ansatz, even a truncated version can implicitly describe higher-order excitations via unconnected terms, thus its energy can be very accurate. In fact, it is these unconnected terms that make CC an size-extensive method. A variation determination of CC parameters is rather difficult due to the complexity of the equations. Thus, one can project the SE onto HF and excited SD states to obtain the *linked CC equation*:

$$E_{\text{CC}} = \langle \text{HF} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \text{HF} \rangle \tag{1.4.39}$$

$$0 = \langle \mu | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \text{HF} \rangle \quad (1.4.40)$$

When the cluster operator is truncated at the second-order, we obtain CC singles and doubles (CCSD) equations[31]. An explicit expression is rather awkward to derive algebraically. However, diagrammatic notations can facilitate the procedure significantly and *elegantly*. The CC energy:

$$\begin{aligned} E_{\text{CC}} &= E_{\text{HF}} + \langle \text{HF} | e^{-\hat{T}} \hat{H}_{\text{N}} e^{\hat{T}} | \text{HF} \rangle \\ &= E_{\text{HF}} + \langle \text{HF} | \hat{H}_{\text{N}} \left(\hat{T}_1 + \hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) | \text{HF} \rangle_{\text{C}} \\ &= E_{\text{HF}} + \text{diagram 1} + \text{diagram 2} + \text{diagram 3} \\ &= E_{\text{HF}} + \sum_{ai} f_{ai} t_i^a + \frac{1}{4} \sum_{\substack{ab \\ ij}} \langle ij | ab \rangle t_{ij}^{ab} + \frac{1}{2} \sum_{\substack{ab \\ ij}} \langle ij | ab \rangle t_i^a t_j^b \end{aligned} \quad (1.4.41)$$

where \hat{H}_{N} is the normal-ordered Hamiltonian, “C” means that only terms involving contractions of \hat{H}_{N} and all \hat{T} on the right are considered. This leads to the fact that only linked diagrams are involved. Of course for a HF reference all f_{ai} will be zero. For the amplitude equation (1.4.40) more diagrams are required but the procedure is similar. These equations are solved iteratively.

Perturbative Treatment of Coupled-cluster Ansatz

For chemical accuracy, three-body cluster operator \hat{T}_3 must be taken into account, but the extremely high computational cost forbids this for even some small molecules. However, it is found that these three-body cluster amplitudes can be treated using the converged single and double amplitudes in a perturbative way without too much loss of accuracy, leading to the CCSD(T) method[32].

It is argued that the performance of CCSD(T) relies on an error cancellation: the perturbative treatment overestimates the energy associated with triples, and canceled by the neglect of quadruples. In fact, the accuracy of CCSD(T) is comparable with that of all-electron, relativistic CCSDT(Q)[33]. Thus, CCSD(T)/CBS is often used as benchmark reference and recognized as the “gold standard” of quantum chemistry.

1.4.4 Explicit Correlation Methods

The basis set used in practice is often incomplete, leading to the *basis set incompleteness error* (BSIE). In fact, in order to reproduce the cusp condition (1.3.15), basis functions of very high angular momenta up to f, g, \dots are needed. To reduce the basis set size and accelerate the basis set convergence, one can explicitly introduce the inter-electron distance r_{12} to the WFN parameterization ansatz in a (1.3.15)-like way[23, 34].

A popular way is to introduce r_{12} into the CCSD formula, obtaining the simplified CCSD-F12 theory[35, 36]. Its WFN is:

$$|\psi\rangle = \exp\left(\hat{T}_1 + \hat{T}_2 + \hat{T}'_2\right) | \text{HF} \rangle \quad (1.4.42)$$

$$\hat{T}'_2 = \frac{1}{4} \sum_{\alpha\beta ij} t_{ij}^{\alpha\beta} \left\{ a_{\alpha}^{\dagger} a_i a_{\beta}^{\dagger} a_j \right\} \quad (1.4.43)$$

where α, β are indices of a formally “complete virtual space”.

The r_{12} will be introduced in the amplitude $t_{ij}^{\alpha\beta}$:

$$t_{ij}^{\alpha\beta} = \sum_{kl} t_{ij}^{kl} F_{kl}^{\alpha\beta} \quad (1.4.44)$$

$$F_{kl}^{\alpha\beta} \equiv \langle kl | F_{12}(r_{12}) \hat{Q}_{12} | \alpha\beta \rangle \quad (1.4.45)$$

$$\hat{Q}_{12} \equiv (1 - \hat{O}_1) (1 - \hat{O}_2) (1 - \hat{V}_1 \hat{V}_2) \quad (1.4.46)$$

$$F_{12}(r_{12}) = -\frac{1}{\eta} e^{-\eta r_{12}} \quad (1.4.47)$$

where \hat{O} and \hat{V} are the projectors of occupied and virtual space, respectively. In practice (1.4.47) is fitted by linear combinations of Gaussian geminals $\exp(-\alpha r_{12}^2)$ [37].

This ansatz improves the WFN since \hat{T}_2' incorporates terms like

$$|u_{ij}(\mathbf{r}_1, \mathbf{r}_2)\rangle = \sum_{kl} t_{ij}^{kl} \hat{Q}_{12} F_{12}(r_{12}) |\phi_k(\mathbf{r}_1) \phi_l(\mathbf{r}_2)\rangle \quad (1.4.48)$$

In the summation, it seems that the terms for $kl = ij$ or $kl = ji$ are most important since they *directly* affect the configuration $|\phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2)\rangle$. Thus, one can use a *fixed amplitude ansatz* [38]:

$$t_{ii}^{ii} = \frac{1}{2}; \quad t_{ij}^{ij} = \frac{3}{8}; \quad t_{ij}^{ji} = \frac{1}{8}; \quad t_{ij}^{kl} = 0 \quad (kl \neq ij \text{ or } kl \neq ji) \quad (1.4.49)$$

These values are estimated from (1.3.15) and (1.3.16) [38]. In (1.4.42), expanding the cluster operator, we note

$$\begin{aligned} |\psi\rangle = & \dots + t_{ii}^{ii} r_{12} |\phi_i(\mathbf{r}_1) \bar{\phi}_i(\mathbf{r}_2)\rangle \\ & + \left(t_{ij}^{ij} - t_{ij}^{ji} \right) r_{12} \frac{1}{\sqrt{2}} (|\phi_i(\mathbf{r}_1) \bar{\phi}_j(\mathbf{r}_2)\rangle + |\bar{\phi}_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2)\rangle) \\ & + \left(t_{ij}^{ij} + t_{ij}^{ji} \right) r_{12} \frac{1}{\sqrt{2}} (|\phi_i(\mathbf{r}_1) \bar{\phi}_j(\mathbf{r}_2)\rangle - |\bar{\phi}_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2)\rangle) + \dots \end{aligned} \quad (1.4.50)$$

thus we have $t_{ij}^{ij} - t_{ij}^{ji} = \frac{1}{4}$ and $t_{ij}^{ij} + t_{ij}^{ji} = \frac{1}{2}$, resulting (1.4.49).

During the period of solving the amplitude equations, many approximations have to be introduced, leading to CCSD-F12a and CCSD-F12b formulae [35, 36]. Benchmarks reveal that CCSD-F12a is closer to the CBS limit [36]. Also it is observed that diffuse basis functions are necessary for obtaining accurate F12 energies [36].

The F12 methods can reduce the BSIE of the correlation energy very efficiently. However, in this case the BSIE of HF may dominate the error. Therefore, one can add an operator for single excitations (orbital rotation) into the formally complete virtual space and compute an energy correction with perturbation theory [35, 36]. This is called *complementary auxiliary basis sets (CABS) [39] singles correction*. Its computational cost is very low, usually negligible.

1.5 Multi-reference Approaches

For all the methods mentioned above, the reference state is a single SD. However, for molecules containing nearly degenerate MOs, a single SD is usually a bad initial guess. In this case, a linear combination of several SDs can be used as the reference state. These are called *multi-reference* (MR) methods. This class of methods is extremely important for the accurate construction of a potential energy surface (PES) or excited state computations. However they are also quite complex. In this work we do not discuss these methods.

1.6 Density Functional Theory

1.6.1 Hohenberg–Kohn Theorem

The *ab initio* methods in quantum chemistry are based on the WFN, which is a function of $3N$ variables ($4N$ if the spin variables are also to be determined). However, *the first Hohenberg–Kohn theorem*[40] states that *for non-degenerate ground states, the external potential v_{ext} is determined within a constant by the electron density $n(\mathbf{r})$* .

Proof. Assume for two different external potential v_{ext} and v'_{ext} , we have identical $n(\mathbf{r})$. In this case, the number of electrons in the two systems are the same due to (1.1.14), thus the corresponding Hamiltonians \hat{H} and \hat{H}' differ only in the external potential. They have different ground state WFNs ψ , ψ' and energies E , E' . Now, we can apply the variation principle (1.1.22):

$$E < \langle \psi' | \hat{H} | \psi' \rangle = \langle \psi' | \hat{H}' | \psi' \rangle + \langle \psi' | \hat{H} - \hat{H}' | \psi' \rangle = E' + \int d\mathbf{r} n(\mathbf{r}) (v_{\text{ext}}(\mathbf{r}) - v'_{\text{ext}}(\mathbf{r})) \quad (1.6.1)$$

similarly we have

$$E' < E + \int d\mathbf{r} n(\mathbf{r}) (v'_{\text{ext}}(\mathbf{r}) - v_{\text{ext}}(\mathbf{r})) \quad (1.6.2)$$

Combining (1.6.1) and (1.6.2) we have $E + E' < E + E'$ which is impossible. This concludes that the two electron densities $n(\mathbf{r})$ and $n'(\mathbf{r})$ must be different. \square

Therefore the electron density can determine the external potential and thus the Hamiltonian and energy. Then it is possible to write the energy as a functional of the electron density (E_{NN} omitted):

$$\begin{aligned} E[n] &= \int d\mathbf{r}^N \psi^* \hat{T}_e \psi + \int d\mathbf{r}^N \psi^* \hat{V}_{ee} \psi + \int d\mathbf{r} n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) \\ &\equiv T[n] + V_{\text{EE}}[n] + \int d\mathbf{r} n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) \\ &\equiv F_{\text{HK}}[n] + \int d\mathbf{r} n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) \end{aligned} \quad (1.6.3)$$

where F_{HK} is called *universal functional* for an N -electron system.

The second Hohenberg–Kohn theorem[40] states that *for the exact N -electron density $n(\mathbf{r})$ and a trial one $\tilde{n}(\mathbf{r})$, we always have $E[n(\mathbf{r})] \leq E[\tilde{n}(\mathbf{r})]$* .

Due to this theorem, we can optimize the energy with respect to $n(\mathbf{r})$, however with the restriction $\int d\mathbf{r} n(\mathbf{r}) = N$, thus we have a Lagrangian :

$$L[n] = E[n] - \mu \left(\int d\mathbf{r} n(\mathbf{r}) - N \right) \quad (1.6.4)$$

The obtained *Euler equation* is:

$$\mu = \frac{\delta F_{\text{HK}}[n]}{\delta n} + v_{\text{ext}}(\mathbf{r}) \quad (1.6.5)$$

The trial density $n(\mathbf{r})$ has to be N - and v -representable.

1.6.2 Kohn–Sham Scheme

It is not trivial to use (1.6.5) directly. However, if we assume a system of N *noninteracting* electrons, with Hamiltonian $\hat{H}_s = \hat{T}_e + \hat{V}_s$, then its WFN is exactly a SD (1.2.2) constructed from single-electron WFNs ϕ_1, \dots, ϕ_N that satisfy

$$\left(-\frac{1}{2}\nabla^2 + v_s(\mathbf{r})\right)\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}) \quad (i = 1, 2, \dots, N) \quad (1.6.6)$$

thus

$$n_s(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \quad (1.6.7)$$

with the energy functional

$$E_s[n_s] = T_s[n_s] + \int d\mathbf{r} n_s(\mathbf{r}) v_s(\mathbf{r}) \quad (1.6.8)$$

The purpose of the introduction of the fictitious system is to solve the real molecule problems. If we require $n_s(\mathbf{r})$ to be equal to the ground state density $n(\mathbf{r})$ of a real molecule, then the two functional (1.6.8) and (1.6.3) should be the same, then:

$$\begin{aligned} \int d\mathbf{r} n(\mathbf{r}) v_s(\mathbf{r}) &= (T[n] - T_s[n]) + (V_{EE}[n] - J[n]) + J[n] \\ &+ \int d\mathbf{r} n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) \\ &\equiv J[n] + \int d\mathbf{r} n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) + E_{\text{XC}}[n] \end{aligned} \quad (1.6.9)$$

where $J[n]$ is the classical electrostatic potential energy:

$$J[n] \equiv \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (1.6.10)$$

The $E_{\text{XC}}[n]$ is *exchange-correlation functional* (XC) which takes all nonclassical effects into account. Using these quantities, we can solve (1.6.6) in an analog way to the HF method. This is the *Kohn–Sham scheme* (KS)[41].

1.6.3 Explicit Functionals

The essential part of KS equations is $E_{\text{XC}}[n]$. Numerous approximate E_{XC} have been proposed:

- Local density approximation (LDA), depending only on the density. Example: X_α [42].

$$E_{\text{XC}}^{\text{LDA}}[n] = \int d\mathbf{r} f(n(\mathbf{r})) \quad (1.6.11)$$

- Generalized gradient approximation (GGA), depending on the density and its gradient. Example: PBE[43].

$$E_{\text{XC}}^{\text{GGA}}[n] = \int d\mathbf{r} f(n(\mathbf{r}), \nabla n(\mathbf{r})) \quad (1.6.12)$$

- meta-GGA, generalizing GGA by including higher order derivatives or the kinetic energy density $\tau(\mathbf{r}) = \frac{1}{2} \sum_{i=1}^N |\nabla\phi_i(\mathbf{r})|^2$. Example: TPSS[44].

$$E_{\text{XC}}^{\text{mGGA}}[n] = \int d\mathbf{r} f(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r})) \quad (1.6.13)$$

- Hybrid functionals. This class of functionals incorporates the Hartree exchange energy by a parameter. Example: B3LYP[45].
- Others like double-hybrid functionals B2PLYP[46].

All the functionals above fail to treat van der Waals interactions (vdW) due to their incorrect asymptotic behavior[47]. A simple but successful solution is using a r^{-6} -like empirical correction, like DFT-D3[48]:

$$E_{\text{disp}}^{\text{D3BJ}} = \sum_{\substack{A,B=1 \\ A < B}}^K \left(s_6 \frac{C_6^{AB}}{r_{AB}^6 + f(R_{AB}^0)^6} + s_8 \frac{C_8^{AB}}{r_{AB}^8 + f(R_{AB}^0)^8} \right) \quad (1.6.14)$$

with Becke–Johnson damping function[49]

$$f(x) = a_1 x + a_2 \quad (1.6.15)$$

where s_6 , C_6^{AB} , s_8 , C_8^{AB} , R_{AB}^0 and a_1, a_2 are fit by *ab initio* calculations, specifically for each functional.

1.7 Molecular Geometry Optimization

Within the BO approximation, the energy parametrically depends on the nuclear coordinates, forming a so-called potential energy surface (PES). The procedure of locating stationary points on the PES is called *optimization*. For a molecule not involved in a chemical reaction, it is often a (local or global) *minimum*. Searching the minimum requires iterations. In each iteration, one can expand the PES at the initial point \mathbf{R}_0 , truncating at the second-order:

$$E(\mathbf{R}_0 + \mathbf{d}) = E(\mathbf{R}_0) + \mathbf{d}^T \mathbf{g}(\mathbf{R}_0) + \frac{1}{2} \mathbf{d}^T \mathbf{G}(\mathbf{R}_0) \mathbf{d} \quad (1.7.1)$$

Thus one can update the step \mathbf{d} by a *quasi-Newton* strategy[50], i.e. using an approximate Hessian G_0 rather than the accurate one:

$$\mathbf{d} = -\mathbf{G}_0^{-1}(\mathbf{R}_0) \mathbf{g}(\mathbf{R}_0) \quad (1.7.2)$$

where \mathbf{G}_0 can be updated by, e.g. Broyden–Fletcher–Goldfarb–Shanno method[50].

An efficient search also requires suitable coordinate systems (CSs). The Cartesian CS is simple but since it involves very high order couplings between the coordinate components, it is only used for molecules with highly complex topologies. A more natural choice is the internal CS, i.e. bond lengths, bond angles and dihedral angles (and perhaps others). In modern quantum chemistry one often constructs an internal CS of more than $3N - 6$ components, i.e. a *redundant internal CS*[51]. However they must be used carefully since some coordinates like dihedral angles could become singular during optimization.

After optimization, one can compute the eigenvalues of the Hessian at this point. If all the eigenvalues are positive, this point is indeed a minimum, otherwise the stationary point is a maximal or saddle one.

1.8 Wave Function Analysis

1.8.1 Electron Localization Function

After solving a molecular SE, one obtains both energy and WFN. The WFN also contains much valued information, which must be revealed by analysis. The first one is the so-called electron localization function (ELF)[52]. Consider a conditional probability (1.3.4)

of finding a σ -spin electron at \mathbf{r}_1 given that a σ -spin electron is at \mathbf{r}_2 , for which a SD WFN is used:

$$P_{\sigma\sigma}(\mathbf{r}_1|\mathbf{r}_2) \equiv f(\mathbf{r}_1, \mathbf{r}_2) = n_\sigma(\mathbf{r}_1) - \frac{(\sum_i^\sigma \phi_i^*(\mathbf{r}_1) \phi_i(\mathbf{r}_2))^2}{n_\sigma(\mathbf{r}_2)} \quad (1.8.1)$$

with σ -spin electron density

$$n_\sigma(\mathbf{r}) = \sum_i^\sigma \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) \quad (1.8.2)$$

Now we perform a coordinate transformation

$$\mathbf{r}_1 = \mathbf{r} + \frac{1}{2}\mathbf{s} \quad (1.8.3)$$

$$\mathbf{r}_2 = \mathbf{r} - \frac{1}{2}\mathbf{s} \quad (1.8.4)$$

for (1.8.1), and perform a spherical average for \mathbf{s} followed by a Taylor expansion of s [53]:

$$f(\mathbf{r}, s) = \frac{1}{3} \left(2\tau_\sigma - \frac{1}{4} \frac{|\nabla n(\mathbf{r})|^2}{n(\mathbf{r})} \right) s^2 + \dots \equiv \frac{1}{3} D_\sigma(\mathbf{r}) s^2 + \dots \quad (1.8.5)$$

One can see from (1.8.5) that a lower $D_\sigma(\mathbf{r})$ indicates that an electron is more highly localized at \mathbf{r} . To make the index bounded from above and below, Becke introduced an ELF of the following form[52]:

$$\text{ELF}(\mathbf{r}) = \frac{1}{1 + (D_\sigma(\mathbf{r})/D_\sigma^0(\mathbf{r}))^2} \quad (1.8.6)$$

$$D_\sigma^0(\mathbf{r}) \equiv \frac{3}{5} (6\pi^2)^{2/3} n_\sigma(\mathbf{r})^{5/3} \quad (1.8.7)$$

where D_σ^0 is the corresponding D_σ of the uniform gas with density identical to $n_\sigma(\mathbf{r})$. In this case ELF is restricted in the range $[0, 1]$, where $\text{ELF} = 1$ means perfect localization. Generally, a large ELF suggests the existence of core regions, lone pairs or covalent bonds, which is illustrated in Figure 1.8.1.

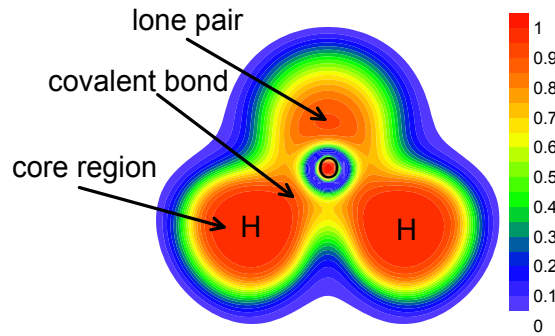


Figure 1.8.1: The filled-color map of ELF in the plane of water molecule.

1.8.2 Topology Analysis of Electron Density

The topology of the electron density $n(\mathbf{r})$ can also reveal much information, forming the so-called *atoms in molecule* (AIM) theory[54]. In real space, $n(\mathbf{r})$ contains many stationary points. These stationary points can be characterized by the signs of the eigenvalues of its Hessian. The number of nonzero eigenvalues is called the *rank* (ω), and the sum of their signs is called the *signature* (σ). Both can be symbolized as (ω, σ) . Then one can obtain[54]:

- *Nuclear critical point* (NCP): (3, -3);
- *Bond critical point* (BCP): (3, -1);
- *Ring critical point* (RCP): (3, +1);
- *Cage critical point* (CCP): (3, +3).

For an isolated molecule, the numbers of these critical points satisfy the *Poincaré-Hopf relationship*[54]:

$$n_{\text{NCP}} - n_{\text{BCP}} + n_{\text{RCP}} - n_{\text{CCP}} = 1 \quad (1.8.8)$$

The electron density at the BCP (n_{BCP}) of a chemical bond can be used as an index of its strength[54, 55, 56, 57]. A larger n_{BCP} indicates a stronger bond.

The Laplacian of the electron density at the BCP, $\nabla^2 n_{\text{BCP}}$ can reveal the nature of a bond. It is well known that in real space $\nabla^2 n(\mathbf{r}) > 0$ and $\nabla^2 n(\mathbf{r}) < 0$ implies that the electron density is locally depleted and concentrated at \mathbf{r} , respectively. Consequently, if $\nabla^2 n_{\text{BCP}} > 0$, the interaction between the two bonding atoms is a closed-shell one, e.g. highly polar covalent bond, ionic or van der Waals (vdW) interactions; if $\nabla^2 n_{\text{BCP}} < 0$, the interaction is usually a typical covalent one.

To calculate the covalent component of a bond, one can integrate the negative Laplacian of the electron density in the “region of bonding”, obtaining the *Laplacian bond order* (LBO)[58]:

$$\text{LBO}_{A,B} = -10 \int_{\nabla^2 n < 0} d\mathbf{r} w_A(\mathbf{r}) w_B(\mathbf{r}) \nabla^2 n(\mathbf{r}) \quad (1.8.9)$$

where w_A is a weighting function that determines the “region” of atom A . An expression proposed by Becke[59] has been widely used.

1.8.3 Noncovalent Interaction Plot

An interesting physical quantity, the *reduced density gradient* (RDG)

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla n|}{n^{4/3}} \quad (1.8.10)$$

has been used in the construction of XC functionals[60]. However, it also has been proved to be useful in detection of noncovalent interactions[61]. For covalent interactions, the regions of low n are those far from the molecule, and it decays exponentially: see (1.1.15). Thus, the RDG can be very large in those regions due to its large denominator.

However, for noncovalent interactions, which formed by the tail regions (low n) of two molecular fragments, the decay rates of n and $|\nabla n|$ are very similar, and $|\nabla n|$ is very close to zero within the region of BCP, thus a *low density and low RDG* becomes the character of covalent interaction. Furthermore, the sign of the second large eigenvalue of its Hessian (λ_2) can be used to identify the interaction nature: $\lambda_2 < 0$ and $\lambda_2 > 0$ implies attractive and repulsive interactions, respectively. Thus, one can plot the isovalue surface of RDG rendered by the sign of λ_2 to identify the noncovalent interactions, as illustrated in Figure 1.8.2. This is often called *NCI plot*.

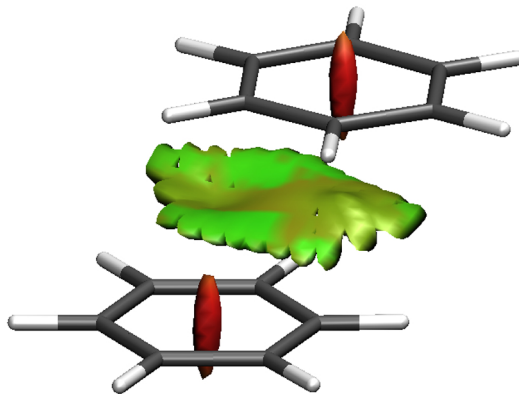


Figure 1.8.2: The NCI plot of benzene dimer. The red and green surface implies the repulsion and weak vdW interaction, respectively.

1.9 Linear-scaling Methods

1.9.1 Locality of Electron Correlation

The high computational barriers of *ab initio* quantum chemical methods prohibit their application to large molecules. For instance, while SCF is a $O(n^4)$ procedure, MP2, CCSD and CCSD(T) are $O(n^5)$, $O(n^6)$ and $O(n^7)$, respectively! However, the high computational cost seems to be unphysical.

First, one can see the physical meaning of E_{HF} and E_{corr} . The former describes most of the physical processes during the formation or change of a molecule, like electrostatic energy, Fermi exchange energy and orbital relaxation. In the computation of E_{corr} , one has to introduce the excited SDs in the WFN, thus it describes the effects of “induced dipole moments”, i.e. the dispersion interaction. Thus E_{HF} and E_{corr} is long- and short-range, thus the latter *should* be easier to calculate.

Second, the molecular electronic equilibrium properties are *nearsighted*[62], i.e., a perturbation at \mathbf{r}' has significant implications on the properties at \mathbf{r} only when $|\mathbf{r} - \mathbf{r}'|$ is sufficiently small. Indeed, Kohn proposed that for molecules with large highest occupied MO-lowest unoccupied MO (HOMO-LUMO) gap G , the implication of effective potential v at \mathbf{r}' on a property Q at \mathbf{r} decays exponentially[62]:

$$\frac{\delta Q(\mathbf{r})}{\delta v(\mathbf{r}')} \approx \exp\left(-\sqrt{G} |\mathbf{r} - \mathbf{r}'|\right) \quad (1.9.1)$$

Therefore, for each point \mathbf{r} , the molecule should only “see” finite surroundings. When the molecule is sufficiently large, the computational scaling should be *linear*, i.e. $O(n)$!

However, none of the *ab initio* or DFT quantum chemical methods satisfy linear-scaling. One reason is that the CMOs used to construct the total one are usually very delocalized, or *farsighted*. To reduce the computational cost and treat only the essential part of electron correlation, the CMOs must be localized, and then for each obtained LMO, the redundant and insignificant numerical quantities (like molecular integrals) can be discarded. This is a possible way of achieving linear-scaling.

1.9.2 Linear-scaling Correlation Methods

In order to make accurate but costly correlation methods feasible for large molecules, in the last three decades quantum chemists have invested large effort in developing new algorithms to lower this computational barrier. Nearly all the methods will first generate

a set of LMOs according to Boys or other criteria, then for each single, pair, triplet, \dots of the LMOs, one can compute its correlation energy by discarding the *small coupling parts*, saving computational cost in this way. These methods mainly fall into two groups: class I and II.

Class I methods divide the system into small fragments and combine the results of the correlation computations of these fragments in specific ways. Some examples are the divide-and-conquer (DC) method by Bartlett[63, 64], the cluster-in-molecule (CIM) strategy by Li[65, 66, 67, 68], the divide-expand-consolidate (DEC) scheme by Jørgensen[69, 70, 71], local CCSD by Head-Gordon[19, 72, 73, 74, 75] and some others[76, 77, 78, 79]. This class of methods is very easy to be realized and parallelized, and is relatively more accurate than class II. They treat correlation energy in a way like

$$E_{\text{corr}} = \sum_D \Delta E_D \quad (1.9.2)$$

where D denotes a “fragment” of the molecular geometrical or electronic structure. For each fragment the virtual space required for correlation to achieve a specific accuracy can be reduced relative to a standard correlation method. However, comparing with class II, their computational costs are larger, the bottleneck being the calculations of the couplings between the fragments.

Class II methods explore the locality of electronic correlation by reconstruction of the occupied and virtual orbital spaces. This group begins with Pulay and Sæbø’s pioneering works [80, 81, 82, 83, 84] which introduced the important concept of projected atomic orbitals (PAOs):

$$|\tilde{\chi}_\mu\rangle = \left(1 - \sum_i |\phi_i\rangle \langle\phi_i|\right) |\chi_\mu\rangle \quad (1.9.3)$$

PAOs were then extensively used by Werner and co-workers in their local correlation methods [85, 86, 87, 88, 89, 90, 91]. Recently Neese proposed that by pair natural orbitals (PNOs) CCSD and CCSD(T) can be highly efficiently implemented [92, 93, 94, 95, 96, 97, 98]. For each pair of LMOs ij its PNOs are constructed by diagonalizing the matrix \mathbf{D}^{ij} :

$$\mathbf{D}^{ij} = (\tilde{\mathbf{t}}^{ij})^\dagger \mathbf{t}^{ij} + \tilde{\mathbf{t}}^{ij} (\mathbf{t}^{ij})^\dagger \quad (1.9.4)$$

where \mathbf{t}^{ij} and $\tilde{\mathbf{t}}^{ij}$ are some kind of MP2 amplitudes. Yang proposed orbital specific virtuals (OSVs) [99, 100] which combine some advantages of PAOs and PNOs. All the methods mentioned above will construct a virtual space that is assumed to contribute to the correlation energy most significant by using a standard of some “indices” being greater than a threshold a_{thres} , like Mulliken-Löwdin charge[85] or density-of-states (DOS)[101]. Class II is usually much faster than class I, however, their accuracy could be very low for a large a_{thres} . For smaller a_{thres} the accuracy can be improved but at the same time the computational cost will increase significantly, comparable with that of a standard implementation.

The *incremental scheme* formally belongs to class I, but it also has virtual space truncation techniques. This method will be presented in detail in the following chapters.

Before concluding this chapter, it is worth pointing out that for most methods mentioned above, the turning point of “linear-scaling” could be rather unpredictable, depending on the molecular shapes, basis sets, correlation methods, and even computational hardware. Molecules of practical interest often lie before the turning point, but significant reduction of computational wall time is indeed possible and useful, which marks great improvements of quantum chemistry in the last three decades.

Chapter 2

Incremental Scheme

2.1 Incremental Expansion

2.1.1 Introduction

In Subsection 1.9.2, we pointed out that the *incremental scheme* is one of the linear-scaling quantum chemical methods. In the 1960s, during the study of the Bethe–Goldstone equations of atomic calculations[102, 103, 104], Bethe wrote down some formulae which can be viewed as the first form of an “increment”:

$$e_{23} = \Delta E_{23} - e_2 - e_3 \quad (2.1.1)$$

After several decades, the incremental scheme was formally first proposed in 1992 during the studies of crystalline solids by Stoll[105, 106, 107]. Thereafter it has been proven highly successful in obtaining the correlation energies of periodic systems[108, 109, 110, 111, 112]. During the 1990s Dolg applied the incremental scheme to calculations of polymer and also got good results[113, 114].

From about 2007, Dolg began to generalize the incremental scheme to finite molecules. The incremental scheme was first applied to CCSD[115], then to MP2, CCSD(T)[116] and F12 methods[117, 118]. After incorporating many strategies to further reduce the computational cost[119, 120, 121, 122], the incremental scheme is recognized as an important method of obtaining *accurate* correlation energies[123, 124]. In fact the results from the incremental scheme can be used as benchmark reference[125, 126]. At the same time the incremental scheme gained many applications, like computations of optical[127, 128] and condensed phase[129, 130] properties.

From 2013, the author of the present work started to improve and realize the incremental scheme, and proposed the latest form of incremental scheme: the third-order incremental dual-basis set zero-buffer approach (inc3-db-B0)[131, 132, 133]. Compared with the previous implementations, this approach has the following advantages: first, it uses a simpler and more efficient domain decomposition algorithm; second, the db-B0 approximation can reduce the computational cost significantly without too much loss of accuracy; third, it is easy to be parallelized, and for users, it is a black-box approach. This approach is accurate and efficient and has been successfully applied into many real chemical problems, including the hydration of trivalent lanthanide (Ln^{3+})[126] and actinide (An^{3+})[134] ions, and the isomer stability of large organic molecules[135]. The inc3-db-B0 approach will be the central topic of this work.

2.1.2 Incremental Expansion in Molecules

In Chapter 1 it is pointed out that for correlation computations the initial HF step will generate the occupied and virtual orbitals (\mathbb{O} and \mathbb{V}). Then all or a (usually valence)

subset of \mathbb{O} will be correlated with the entire \mathbb{V} via the coupled-cluster or other ansätze to get the correlation energy. In the incremental scheme, the \mathbb{O} will be decomposed into several subsets called *1-site domains*, denoted by i, j, k etc. The union of n 1-site domains i_1, \dots, i_n is called *n-site domain*, denoted by $i_1 \cdots i_n$. For an arbitrary n -site domain X , let ϵ_X be the correlation energy obtained by correlating all occupied orbitals in domain X . Then, we define the first-, second-, and third-order increment as:

$$\Delta\epsilon_i \equiv \epsilon_i \quad (2.1.2)$$

$$\Delta\epsilon_{ij} \equiv \epsilon_{ij} - \Delta\epsilon_i - \Delta\epsilon_j \quad (2.1.3)$$

$$\Delta\epsilon_{ijk} \equiv \epsilon_{ijk} - \Delta\epsilon_{ij} - \Delta\epsilon_{ik} - \Delta\epsilon_{jk} - \Delta\epsilon_i - \Delta\epsilon_j - \Delta\epsilon_k \quad (2.1.4)$$

Generally, an n th-order increment is defined as:

$$\Delta\epsilon_{i_1 \cdots i_n} \equiv \epsilon_{i_1 \cdots i_n} - \sum_{I \in \mathcal{P}(\{i_1, \dots, i_n\}) / \{\{i_1, \dots, i_n\}\}} \Delta\epsilon_I \quad (2.1.5)$$

where \mathcal{P} means the power set.

Then, the total correlation energy E_{corr} can be written as[131, 132]:

$$E_{\text{corr}} = \sum_i \Delta\epsilon_i + \frac{1}{2!} \sum_{ij} \Delta\epsilon_{ij} + \frac{1}{3!} \sum_{ijk} \Delta\epsilon_{ijk} + \cdots \quad (2.1.6)$$

Since for finite molecules the number of domains n_d is finite, this expansion is always ending at order n_d and exact. Expressions like (2.1.3) to (2.1.6) can be found in many branches of theoretical physics and chemistry, e.g., N -body expansion of a total intermolecular potential[136], the Hubbard model [137], the discussions in statistical mechanics[138, 139] and the energy expression in fragment molecular orbital theory[140]. Indeed, the concept ‘‘increment’’ has deep relationship[141] with the important physical quantity *cumulant*[142]: the incremental expansion can be derived from the Faddeev-equation for the many-body problem[143, 144] by a cumulant expansion.

In (2.1.6), the locality of electron correlation is explored by domain decomposition, while the many-body effects are taken into account by the higher order increments. The higher the order of increments, the less the increments will contribute to the total correlation energies of systems with reasonably well localized occupied orbitals. Therefore it is possible to truncate the incremental expansion at a certain order n in practice (*incn*).

In Figure 2.1.1, we plot the errors of incremental expansions for different systems versus the truncation order. These systems include dispersion interaction bound cluster $(\text{He})_8$, hydrogen bonding bound cluster $(\text{H}_2\text{O})_8$, saturate and conjugate long chain hydrocarbon C_8H_{18} and C_8H_{10} . Obviously, the smaller the interactions between the atoms, the faster the incremental expansion converges. We can also observe that generally, no matter what n_d is, the errors of *inc1* and *inc2* are too large to accept ($> 10^{-3}$ kJ mol $^{-1}$). The error of *inc3* is smaller for a smaller n_d , since a large n_d makes the decomposition too fragile and the couplings between the domains are still large beyond the third-order. Thus, when the decomposition is not too fragile, the incremental expansion can be truncated at the third-order with best accuracy/cost ratio.

The choice of n_d is critical, since both too small or too large n_d can deteriorate the efficiency and accuracy of the incremental scheme. For the chemical systems with well-defined small ‘‘fragments’’ or ‘‘units’’, n_d should be the number of those chemical groups. For example, for a cluster of 10 solvent molecules and a small organic molecule, n_d should be set to 11. For systems composed of large ‘‘units’’ or ones with no obvious ‘‘fragments’’, n_d should be chosen in such a way that for the highest order increments in the incremental expansion, the number of orbitals to be correlated is still tractable.

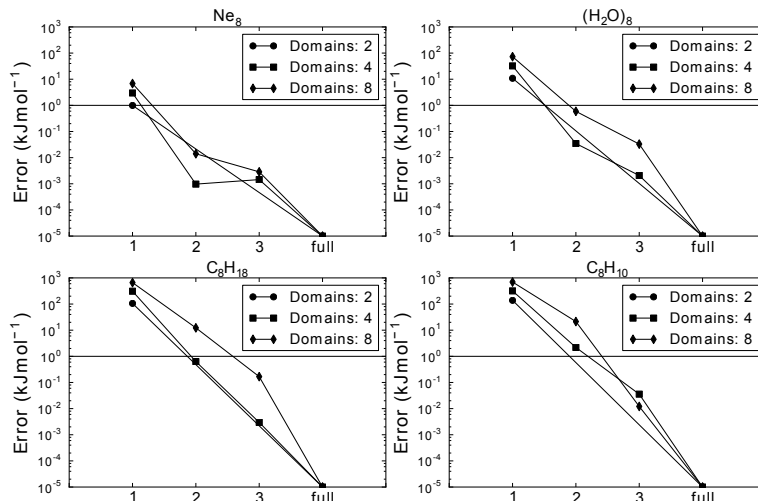


Figure 2.1.1: The errors of incremental expansions for four systems versus the truncation order. The energy is calculated at the CCSD(T)/6-31G(d) level.

Finally, in terms of ϵ_X , the n th-order truncation of (2.1.6) can be written as:

$$E_{\text{corr}}^{(n)} = \sum_{k=0}^{n-1} (-1)^k \binom{n_d - n + k - 1}{k} \sum_{i_1 < \dots < i_{n-k}} \epsilon_{i_1 \dots i_{n-k}} \quad (2.1.7)$$

which can be proved by mathematical induction.

2.1.3 Incremental Expansion for High-spin Open-shell Molecules

For a class of *high-spin* systems, in which the n unpaired electrons occupy n energetically well-separated orbitals and couple ferromagnetically to the highest possible spin $S = n/2$, the static correlation is often not significant, thus these systems can be treated well in the single-reference (SR) framework.

The incremental expansion (2.1.6) in principle still works. However, the singly occupied molecular orbitals, or active molecular orbitals (AMOs), tend to be much more diffuse than the doubly occupied ones. This means that AMOs may couple with all the other occupied MOs to a considerable extent. Thus, for the incremental expansion of high-spin open-shell systems, active orbitals will be correlated in the calculation of each ϵ_X . An additional advantage is that preserving AMOs in each domain X can keep its spin state identical to that of the entire system. In this way (2.1.6) will repeatedly take the correlation of AMOs into account, therefore it can be modified as[133]:

$$E_{\text{corr}} = \epsilon^{\text{A}} + \sum_i \Delta\epsilon_i + \frac{1}{2!} \sum_{ij} \Delta\epsilon_{ij} + \frac{1}{3!} \sum_{ijk} \Delta\epsilon_{ijk} + \dots \quad (2.1.8)$$

$$\Delta\epsilon_i \equiv \epsilon_i^{\text{A}} - \epsilon^{\text{A}} \quad (2.1.9)$$

$$\Delta\epsilon_{ij} \equiv \epsilon_{ij}^{\text{A}} - \epsilon^{\text{A}} - \Delta\epsilon_i - \Delta\epsilon_j \quad (2.1.10)$$

$$\Delta\epsilon_{ijk} \equiv \epsilon_{ijk}^{\text{A}} - \epsilon^{\text{A}} - \Delta\epsilon_{ij} - \Delta\epsilon_{ik} - \Delta\epsilon_{jk} - \Delta\epsilon_i - \Delta\epsilon_j - \Delta\epsilon_k \quad (2.1.11)$$

where “A” stands for AMO and ϵ^{A} is the correlation energy of the AMOs. Note that in each equation ϵ^{A} and the ϵ_X must be calculated with the same X -specific basis set (see Section 2.3); also there is no “A” superscript for $\Delta\epsilon_X$ ’s since the contribution from the

AMOs has been deducted. For the singlet (in high-spin case, closed-shell) states, $\epsilon^A = 0$ and (2.1.8) becomes (2.1.6). In this way, we successfully generalize the incremental scheme to high-spin open-shell systems.

2.2 K-Means Clustering Algorithm

In order to make the truncated incremental expansion sufficiently accurate, the interactions between domains should be as small as possible. First the CMOs to be correlated are localized to obtain the LMOs according to the Boys criterion[16]. Then each LMO ϕ_i is associated to its center of charge $\mathbf{r}_i = \langle \phi_i | \mathbf{r} | \phi_i \rangle$ to get a set of points $\mathcal{P} = \{\mathbf{r}_i | \phi_i \in \mathbb{O}\}$. At this stage we need an algorithm to decompose \mathcal{P} into n_d domains D_1, \dots, D_{n_d} .

There are many algorithms to do this. One way is to create a graph for \mathcal{P} and use some graph theorems for decomposition (e.g. METIS[145]). However, a simpler algorithm is *K-means clustering* (KM)[146, 147], in which we minimize the following function with respect to all possible distributions of the orbitals in domains:

$$J(\mathcal{P}, n_d) = \sum_{k=1}^{n_d} \frac{\sum_{\mathbf{r}_{k_i} \in D_k} |\mathbf{r}_{k_i} - \bar{\mathbf{r}}_k|^2}{|D_k|} \quad (2.2.1)$$

Here $\bar{\mathbf{r}}_k$ is the geometrical center of the $|D_k|$ point \mathbf{r}_{k_i} 's in domain D_k . Obviously, this procedure aims to minimize the distances between the points within a domain. KM algorithm has wide applications in artificial intelligence, statistics and chemoinformatics [148].

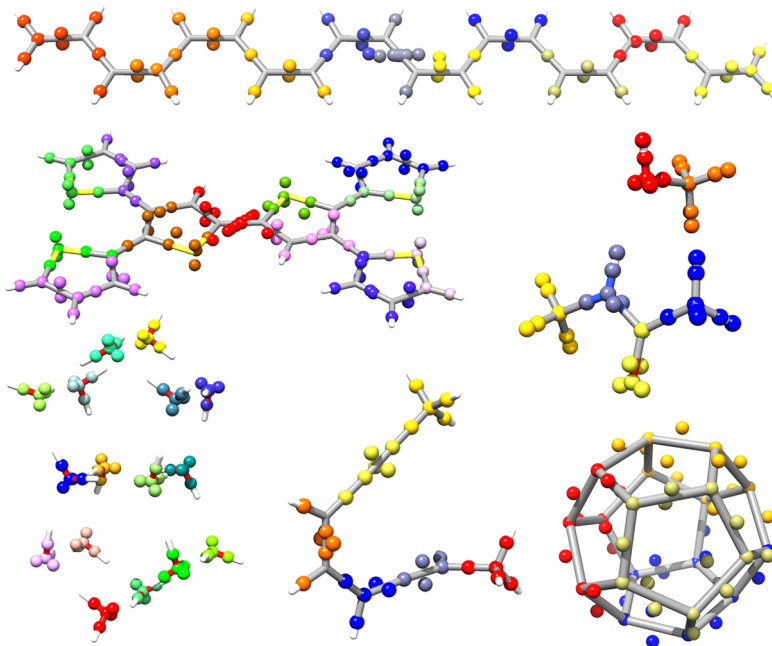


Figure 2.2.1: Domain decomposition by KM algorithm. The balls stand for the center of the LMOs and the balls with identical color belong to the same domain.

To optimize $J(\mathcal{P}, n_d)$, an initial guess (random decomposition) is generated. Then in each iteration, each point in \mathcal{P} is put into all the domains to determine the domain in which it minimizes $J(\mathcal{P}, n_d)$. The iteration stops when the decomposition does not change. This procedure is very close to the Jacobi sweep algorithm [17] in orbital

localization. This algorithm is very sensitive to the initial guess and thus could only be optimized to a local minimum. A simple solution is to use many initial guesses to perform the optimization, and then choose a decomposition with minimal $J(\mathcal{P}, n_d)$. In Figure 2.2.1, we give some decomposition examples, and can see that the KM algorithm can decompose the systems of various topologies in a chemically very meaningful way.

2.3 Dual-basis Set Zero-buffer Approximation

After domain decomposition by KM algorithms, the LMOs in a domain will distribute in a spatially compact way, thus for each domain X it is possible to design a X -dependent virtual space \mathbb{V}_X to compute the increment $\Delta\epsilon_X$ efficiently and accurately. To achieve this we will introduce two strategies: dual-basis set (db) and zero-buffer (B0).

2.3.1 Dual-basis Set Technique

The *dual-basis set* technique has been widely used in quantum chemistry. This terminology was proposed by Jurgens-Lutovsky and Almlöf [149] in 1991, but in fact King and co-workers had implemented similar strategies a few years earlier [150, 151]. This method uses different basis sets at different stages of a quantum chemical computation. Two applications using dual-basis sets are acceleration of MP2 [152, 153] and ab initio molecular dynamics [154].

For the incremental scheme, we use such a strategy: given the basis set \mathcal{B} , at the HF stage we will use a subset \mathcal{B}' of \mathcal{B} , where all the basis functions of angular momenta higher than p (i.e. angular momentum $l = 1$) are removed from \mathcal{B} ; then for the computation of a domain X , a part of the basis functions removed is added again to get the correlation energy.

This strategy enables us, in correlation computations of domain X , to use larger basis sets for the region of the system that has large contribution most to the correlation energy, while for the rest (or “environment”) \mathcal{B}' is still used. The new basis set applied for the domain X is denoted as \mathcal{B}_X . In this way, the virtual space can be significantly reduced compared to the original basis set \mathcal{B} .

Note that in the previous implementation, the reduction of the virtual space is achieved by domain-specific basis sets, using the whole basis set \mathcal{B} at the HF stage, localization and decomposition, whereas for the computations of increments, the environment was treated with a smaller basis set such as STO-3G [120] or SV [124], which leads to considerable difference between the whole basis set at the HF stage and the one used for the increment calculation. Therefore for each increment computation a HF and localization procedure has to be performed, and the obtained LMOs have to be mapped to the LMOs in basis set \mathcal{B} to identify a domain. This mapping is not unique and thus in some cases problematic, e.g., for aromatic systems. In the new implementation, since \mathcal{B}' is a subset of \mathcal{B}_X , the orbital relaxation from basis functions of high angular momenta will be small for SR cases. Thus a repetition of the HF and localization are not necessary, and the LMOs in \mathcal{B}' can be exactly projected onto \mathcal{B}_X (see Figure 2.3.1): the occupied space is preserved and the virtual space is augmented and orthogonalized by the Löwdin procedure [155]. This generates virtual orbitals which are closest in a least-squares sense to the original virtual orbitals and the added higher angular momentum functions [156].

Therefore, besides saving extra HF procedures, there is no ambiguous mapping. Also, the environment can be described more accurately in \mathcal{B}' than with basis sets such as STO-3G or SV which are possibly too small for a good HF description and also unsuitable for correlation computations. In this way, the dual-basis set incremental

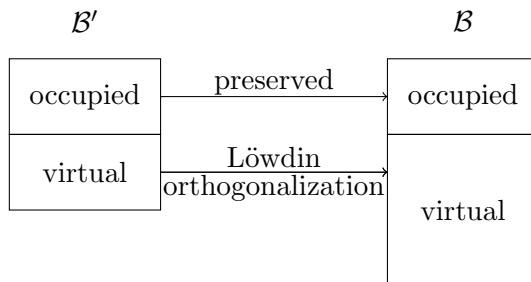
Figure 2.3.1: The procedure of projecting LMOs in \mathcal{B}' onto those in \mathcal{B}_X .

Table 2.3.1: Benchmark of dual-basis set and standard CCSD, CCSD(T), and their F12 variants.

Basis Set	VDZ	VTZ	VQZ	AVDZ	AVTZ	AVQZ
HF	0.07853	0.08196	0.08161	0.08195	0.08275	0.07445
CCSD	0.00066	0.00059	0.00058	0.00077	0.00061	0.00053
CCSD-F12a	0.00235	0.00100	0.00052	0.00240	0.00074	0.00062
CCSD-F12b	0.00303	0.00096	0.00046	0.00327	0.00062	0.00044
CCSD(T)	0.00031	0.00023	0.00022	0.00036	0.00029	0.00020
CCSD(T)-F12a	0.00257	0.00087	0.00045	0.00257	0.00048	0.00030
CCSD(T)-F12b	0.00327	0.00100	0.00053	0.00351	0.00068	0.00037

In the table are the root mean square deviations (RMSDs. Unit: Hartree):

$$\text{RMSD} = \sqrt{\left(\sum_i^N (E_{\text{ex-db}}(i) - E_{\text{st}}(i))^2 / N \right)} \quad (2.3.1)$$

where $E_{\text{ex-db}}(i)$ and $E_{\text{st}}(i)$ are the energy of molecule i computed by the dual-basis set approach and the standard correlation method, respectively. Here, in the dual-basis set computations, the HF was performed with VXZ(p/s) or AVXZ(p/s) where (p/s) means only basis functions of angular momenta s and p are preserved for hydrogen and other elements, respectively; the following correlation computation was performed with original VXZ or AVXZ basis sets.

scheme can enjoy both a high accuracy and efficiency.

Note that as the incremental order increases, the result converges not towards that of a standard correlation calculation, but to the corresponding exact dual-basis set result. To see how the results can differ, we performed a benchmark with 15 small molecules. The results are listed in Table 2.3.1.

We observe that although energies at the HF stage differ significantly, the *total* energies have rather small errors. Meanwhile, the error decreases as the basis set becomes more complete. This can be understood since the single excitation operator \hat{T}_1 in the CC ansatz compensates for the omitted orbital relaxation when going from the smaller to the larger basis set. Thus correlation contributions discussed here also contain some orbital relaxation effect. This implies that for the MP2 method which does not contain \hat{T}_1 the dual-basis set technique could introduce large errors. It is noteworthy that, if relative energies are considered, the error becomes even more negligible.

2.3.2 Zero-buffer Approximation

How to determine the atoms that contribute significantly to the correlation energy of a domain X ? Intuitively they are the atoms distributing closely to the domain. In some previous implementations, these atoms were identified as the atoms whose distance from the orbital centers of a domain to be computed is less than a parameter t_{main} [120]. In this work, we will introduce a new, parameter-free strategy.

As the KM algorithm is used to decompose the MOs, it will then also be used to decompose the atoms of the system \mathcal{A} into n_d atomic domains M_1, \dots, M_{n_d} , but the difference is that the \mathbf{r}_{k_i} 's in (2.2.1) are now the coordinates of the atoms \mathbf{x}_{k_i} and $\bar{\mathbf{r}}_k$ is not dynamically determined but fixed to the geometric center of orbital domain D_k . This is the fixed-center KM (FCKM) algorithm, i.e. we try to minimize:

$$J(\mathcal{A}, n_d) = \sum_{k=1}^{n_d} \frac{\sum_{\mathbf{x}_{k_i} \in M_k} |\mathbf{x}_{k_i} - \bar{\mathbf{r}}_k|^2}{|M_k|} \quad (2.3.2)$$

In this way each orbital domain D_k is associated with M_k that shares the same $\bar{\mathbf{r}}_k$. Thus, in the computation of domain X , only the atoms from M_k 's associated domain D_k 's in X are treated with basis set \mathcal{B} . One may argue that a buffer region "Br" around the M_k 's with a radius r should also be treated with \mathcal{B} to ensure the accuracy. The buffer region resembles the same terminology proposed by Yang in the development of divide-and-conquer density functional theory [157, 158].

To determine a suitable r , we examine its effect using the molecule bicyclooctane. Since bicyclooctane has eight carbon atoms, by chemical intuition the largest n_d is 8. We thus performed the CCSD(T)/VDZ computations of bicyclooctane by the incremental scheme for $5 \leq n_d \leq 10$. In this case, "B10" is the entire molecule, thus all increments are evaluated in basis set \mathcal{B} . The results are shown in Figure 2.3.2. For each $n_d \leq 8$, the accuracies of B0 and B10 correlation energies are similar, with an unsigned error of less than 0.0011 Hartree. However, when n_d goes beyond 8, the accuracy of B0 decreases rapidly. This can be understood by the fact that for $n_d = 6$ each domain is roughly "one-and-a-half" C-C bond, and the basis functions of other domains have a small contribution. For $n_d = 10$ the molecule is broken into too small pieces, where each small domain contains large energy contributions from the basis functions of other domains. Thus the B0 approximation does not work well and must be extended, e.g. to B10. However, when considering the efficiency, B0 is faster than B10 by about 6 times, revealing that $\text{Br}(r > 0)$ is not worth implementing since *as long as a reasonable n_d is selected, B0, or zero-buffer can enjoy both a high accuracy and efficiency*. It is also observed that while smaller n_d yield a more accurate energy, both a too small or a too large n_d will decrease the efficiency.

Thus, as long as the orbital decomposition is *chemically meaningful*, the buffer region is not required at all, i.e., a B0 approximation provides very accurate results. Combining with the dual-basis set technique, this *db-B0 approximation* is shown in Figure 2.3.3 and exhibits excellent performance.

2.4 Distance Screening

When two domains i and j are far away from each other, their coupling is small, thus the increment $\Delta\epsilon_{ij}$ can be approximated as zero and its calculation can be avoided, saving the computational cost. Thus a truncation distance r_t can be defined. If the distance between i and j is greater than r_t , the increments containing the two domains such as $\Delta\epsilon_{ij}$ or $\Delta\epsilon_{ijk}$ can be neglected. This approximation works very efficiently for systems with a large spatial extent like long chain hydrocarbons. As shown in Figure

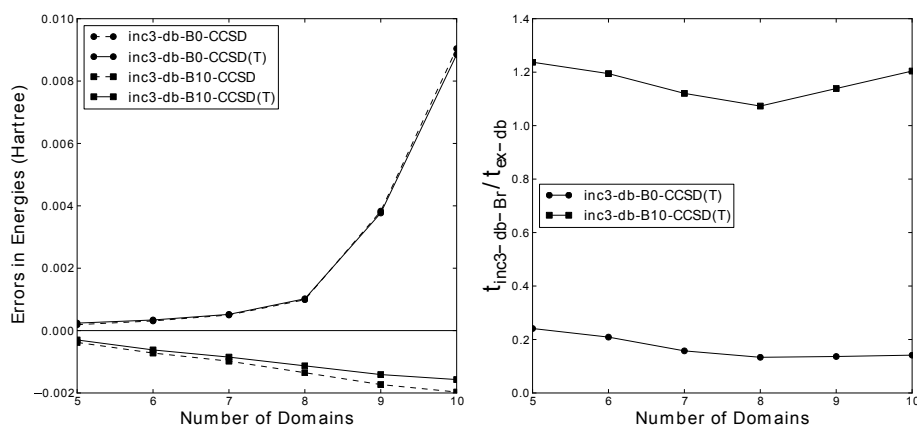


Figure 2.3.2: Impact of n_d and buffer on the accuracy and efficiency of the incremental scheme. In the right figure t_X stands for the wall time of a calculation by method X at the same machine.

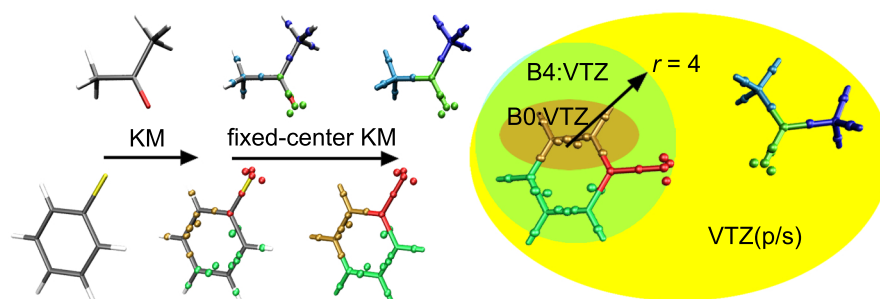


Figure 2.3.3: Illustration of dual-basis set zero-buffer approximation. Here a HF computation of the molecule is performed with VTZ(p/s) basis set, obtaining the CMOs. After localization the center of the LMOs are computed and represented by balls, and are clustered into domains by KM algorithm. The balls of identical color belong to the same domain. Then the atoms are clustered by FCKM algorithm to determine the buffer region. The last graph illustrates the B0 approximation: to compute the correlation energy of the orange orbital domain, we apply VTZ basis set on the atoms in the orange atom domain, and VTZ(p/s) for the rest of the molecule.

2.4.1, where the systems are identical with those in Figure 2.1.1, the increment decays with the distance much faster for the clusters in which the fragments interact weakly like He_8 than the delocalized conjugated hydrocarbon like C_8H_{10} . Usually, it is safe to set r_t greater than 6.0 \AA .

2.5 Inc3-db-B0 Approach

Combining the strategies proposed in the previous sections in this chapter, we obtain the *third-order incremental dual-basis set zero-buffer approach* (inc3-db-B0), which can work with CCSD, CCSD(T) and their F12 variants for all closed-shell and high-spin open-shell molecules with large HOMO-LUMO gap. The basic procedure of inc3-db-B0 approach is:

1. HF computation to obtain the CMOs;

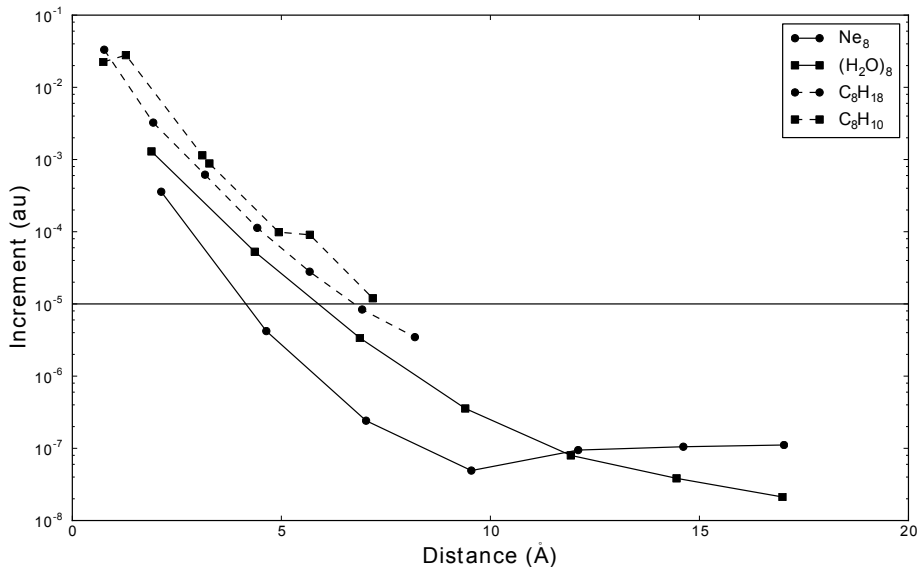


Figure 2.4.1: The values of the increments for four systems versus the distance. The energy is calculated at CCSD(T)/6-31G(d) level.

2. Localization of the CMOs to obtain the LMOs;
3. Compute the charge center of the LMOs, then use KM algorithm to perform the domain decomposition;
4. Correlation computations of all increments, using db-B0 approximation;
5. Collect results and use incremental expansion to obtain the final correlation energy.

Now, for inc3-db-B0-CCSD/CCSD(T), the total energy is written as:

$$E(\mathcal{B}) = E_{\text{HF}}(\mathcal{B}') + \sum_X \Delta\epsilon_X(\mathcal{B}_X) \quad (2.5.1)$$

and for inc3-db-B0-CCSD/CCSD(T)-F12a/b, one should add the CABS singles correction:

$$E(\mathcal{B}) = E_{\text{HF}}(\mathcal{B}') + E_{\text{CABS singles correction}}(\mathcal{B}) + \sum_X \Delta\epsilon_X(\mathcal{B}_X) \quad (2.5.2)$$

In conclusion, inc3-db-B0 approximation explores the locality of electronic correlation in an elegant way: *the orbital localization and decomposition spatially divide the electronic structure into several fragments in such a way that the B0 approximation works well.* Here we emphasize that the incremental scheme decomposes the *electronic* rather than *geometrical* structure, thus there is no need of considering the spin of each fragment or saturating valency by adding chemical units, which may be quite awkward. On one hand, the second- and third-order increments can take the delocalization and conjugation effects into account that might have been lost in the first-order increments; on the other hand, since in an increment the orbitals to be correlated are clustered spatially, the extra basis functions of atoms outside the B0 region have little contribution, thus the truncated virtual space provided by the db-B0 scheme is very accurate. In fact, according to our benchmarks on some medium-sized molecules[131, 132], the error of inc3-db-B0 approach in relative energy is often less than 1 kJ mol⁻¹, thus it can be applied in the cases where the energy difference is very small.

The implementation and application of inc3-db-B0 approach will be presented in detail in the following chapters.

Chapter 3

Implementation of Incremental Scheme

3.1 Introduction of APTS

3.1.1 Basis Features

The program APTS—A Parallel incremenTal Scheme—has been developed by the author since 2013 to realize a user-friendly, black-box, accurate and efficient implementation of the inc3-db-B0 approach introduced in this work. Its main functions and features include:

- Decomposition of a molecule with a given n_d by a KM algorithm.
- Arbitrary order n of the incremental expansion (inc n).
- Arbitrary buffer size (Br) for the dual basis set strategy.
- It can treat both closed-shell and high-spin open-shell molecules.
- It can combine with CCSD, CCSD(T) and their F12 versions.
- It can be run serially or parallelized with an arbitrary number of cores.
- When a work was interrupted, the program can recover and continue the work without doing duplicated calculations.
- It is written in a way that is easy to maintain and to be extended.
- It can work with any mature quantum chemical program packages or in-house codes.
- It is a “black-box”, and user-friendly.

In this work, the geometry optimizations and some DFT calculations were performed with GAUSSIAN03[159] and ORCA 3.0.1[160], and all the correlation calculations were done with MOLPRO2012[161] and our APTS. The AIM and ELF were calculated with MULTIWFN3.2[162]. The visualization of molecules, MOs, etc. are realized with CYLVIEW[163], VMD[164] and CHIMERA[165].

3.1.2 Code Structures

All parts of APTS are written in standard C++, with 24 files and more than 3500 lines of code. The `pthread` library, `STL`, `boosts` library, and the C++0x features are extensively used to simplify the coding. The program consists of 11 modules, with minimal coupling between them:

- `ElapsedTime` Perform the time statistics during the calculation.
- `KMeanClustering` A template class for the KM algorithm.
- `Molecule` A class for reading and treating molecules.
- `Point3D` A class for geometrical controlling.
- `IncrementalDomain` A class for treating domains of the incremental scheme.
- `IncrementalFramework` A class for global control of the incremental scheme.
- `IncrementalInput` A class for treating the input file of the incremental scheme.
- `IncrementalKeeper` A class for treating the incremental expansion.
- `IncrementalParallel` A class for parallelization of the incremental scheme.
- `IncrementalQCInterfaceMolpro` The interface between APTS and MOLPRO.
- `Termination` A “graceful” termination subroutine.

An exhaustive explanation of the codes is a long story. However, the author wants to emphasize that most of the essential modules of APTS have been strictly tested in numerous calculations and are thought to be reliable in practice and those modules should *not* be changed in most cases! Usually in order to add a new function, modification of the `IncrementalInput` and `IncrementalQCInterfaceMolpro` is sufficient!

3.1.3 Compilation

The only requirement of compilation is a C++ compiler like `GCC` and the `boosts` library. Since in the program directory there is a `Makefile`, one simply uses the command `make`, and then the program APTS will appear in the directory.

3.2 Files of APTS

3.2.1 Input Files

For an incremental calculation by APTS, exactly three files are always needed: the *XYZ file*, *node file* and *increment input file*.

XYZ file. This file contains the coordinates of the molecule to be calculated in XYZ file format specified by, say the `OPENBABEL` program. An example for the CH3CHO molecule is given below. Note that the unit must be Å!

Code 3.1: Coordinates of CH3CHO.

1	7			
2	CH3CHO			
3	C	-0.234109	0.399573	0.000009
4	H	-0.301818	1.511786	0.000075

5	O	-1.237942	-0.277006	-0.000003
6	C	1.171086	-0.148488	-0.000008
7	H	1.713840	0.219448	-0.880789
8	H	1.713933	0.219025	0.880891
9	H	1.155718	-1.240727	-0.000163

Node file. This file contains the host names of the nodes used for parallelization. The first line is the number of nodes, followed by the node names, each one occupying one line. Note that for a node, if more than one core is used, the host name must be duplicated! The following example suggests that one will use two cores of the node cc01 and one core of node cc02 and cc03 for parallelization.

Code 3.2: An example of node file.

```

1 4
2 cc01
3 cc01
4 cc02
5 cc03

```

Increment input file. This file contains exactly 28 lines. A typical input file is shown below:

Code 3.3: Input file for inc3-db-B0-CCSD(T)/VDZ of CH₃CHO.

```

1 0          # 0 new mission; pid: restart
2 ch3cho.xyz # Molecule coordinates
3 0          # Charge
4 1          # Multiplicity, i.e. 2S+1
5 boys      # Localization strategy: boys
6 3          # Number of domains
7 3          # Order of incremental expansion
8 6.0       # Truncation distance
9 0          # Distance truncation flag
10 0         # Core correlation
11 No        # (Not used yet)
12 0         # (Not used yet)
13 vdz       # Basis set
14 (p/s)     # Dual basis set flag
15 0.1       # Buffer size
16 0         # ECP information
17 0         # Frozen core in ECP
18 ccsd(t)   # Correlation method
19 2000      # Memory size, in MW
20 molpro    # Quantum Chemistry Program
21 molpro.exe -W /scratch/$USER -d /scratch/$USER # cmd
22 0         # Orbital cubes
23 ./nodes   # File containing nodes to parallelization
24 0         # Canonicalization
25 /scratch/ # Scratch path
26 3.2d-5    # OSV parameter
27 x         # Dummy atoms
28 0         # CC shift

```

For each line, the contents after # is a comment and is ignored.

1. 0 or a positive integer. For a new calculation it must be 0. To restart a failed calculation, this is the number of its process ID, which is given in the output file.
2. String. The file name of the XYZ file.
3. Integer. The charge of the molecule.
4. Positive integer. The spin multiplicity ($2S + 1$) of the molecule.
5. `boys`. The localization method. Although some other settings are valid, we strongly suggest that the user should always `boys`.
6. Positive integer. The number of domains in which a molecule is to be decomposed (n_d).
7. Positive integer. The truncated order of the incremental expansion (2.1.6) or (2.1.8). Usually 3 is a good choice. For molecular clusters 2 could be sometimes suitable.
8. Positive real number. This is the distance screening parameter r_t in Å. For two 1-site increments with their distance greater than r_t , their coupling will be neglected.
9. 0 or 1. For 0 the distance screening is switched off.
10. 0 or 1. For 0 only the valence electrons are correlated; for 1 all the electrons will be correlated.
11. Not used yet.
12. Not used yet.
13. String. The basis set name. We strongly recommend the following items: `VDZ,VTZ,VQZ,AVDZ,AVTZAVQZ`. For explorative calculations `6-31g(d)` or `sto-3g` can be used.
14. 0, (p/s). For 0 no dual-basis set strategy is used; for (p/s) the db-Br approximation is used.
15. Positive real number. The buffer size r . We strongly recommend 0, that is, B0 approximation. For r greater than 0 the computation time can be very long with little increase of accuracy.
16. String. The ECP information. For 0, no ECP will be used. To use ECP, one can write `Eu=ECP52MWB-II`.
17. Positive integer. The number of electrons that is implicitly treated by the ECP. For example, if one use `ECP52MWB-II` for europium, one should write `Eu 52`.
18. `ccsd`, `ccsd(t)`, `ccsd-f12`, `ccsd(t)-f12`, `rccsd` and `rccsd(t)`. For high-spin open-shell molecules, only `rccsd` and `rccsd(t)` are allowed. Some other items are possible but they are still under testing.
19. Positive number. The memory size for each calculation in MW.
20. `molpro`. This is the interface for the quantum chemical program. Currently only `molpro` is possible.
21. String. The command to run the quantum chemical program (currently only `MOLPRO` is available). Note that the scratch path *must* be given explicitly in the form of `-d some path/$USER`.

22. 0 or 1. For 1 the orbital cube files will be output for visualization.
23. String. The file name of the node file.
24. 0 or 1. For 1 an orbital canonicalization will be performed before doing CC iteration. Usually this is desirable. However, in rare cases, this can cause failure of calculation. In this case please change this item to 0.
25. String. This is the scratch path *without* \$USER! For example, if you set /scratch/\$USER in line 21, this item should be set as /scratch/. Note that if you parallelize the program on different nodes, this path must be available for all the nodes!
26. For internal use.
27. String. This is used for counterpoise calculations. For **x**, no counterpoise is applied. To do a counterpoise calculation, please write down the indices of the dummy atoms, say: 0 1 2. Note that the indices are separated by spaces and *start from 0!*
28. 0 or positive integer. For closed-shell molecules, this item can always be 0. For high-spin open-shell molecules, when then CC iteration does not converge, one can increase it to 1, 2, 3, \dots . However, a larger number will increase the computational cost!

With these three files one can perform an incremental calculation.

3.2.2 Output Files

When a calculation is accomplished, some files are appearing:

- `scratch path/*.int*` These are the “file 1” for MOLPRO.
- `scratch path/*.wfu*` These are the “file 2” for MOLPRO.
- `*.out` Main output file.
- A directory `x_pid x` is the input file name, and `pid` is the process ID of the task performed. In the directory the files we have:
 - `*.inp` and `*.out` The input and output files for the calculation of each increment.
 - `*.pdb` This file contains the domain information in standard protein data bank (PDB) format.
 - `*restart*` Information for restart. Never change them!

3.3 Practical Guides of APTS

3.3.1 Perform a Standard Calculation

To perform an inc3-db-B0-CCSD(T)/VDZ calculation for CH₃CHO with four CPUs, we first prepare three files: an XYZ file `ch3cho.xyz` (see Code 3.1), a node file `nodes` (see Code 3.2) and an input file `ch3cho.inp` (see Code 3.3). With these files one can simply use `apts ch3cho.inp > ch3cho.out` to perform the calculation. Note: for F12 or high-spin open-shell calculations, some extra calculations are needed which are introduced in the following section.

When the calculation is accomplished, one can find `ch3cho.out` and a directory (say) `ch3cho_2341`. One can find the following output in `ch3cho.out`:

Code 3.4: Output file of the inc3-db-B0-CCSD(T)/VDZ calculation of CH₃CHO (modified for typesetting).

```

1  -- CCSD(T) Energy --
2  -----
3  domain      Correlation Energy      Incre Corr-Energy
4  -----
5      D_0      -0.2463108771            -0.2463108771
6      D_1      -0.0726881952            -0.0726881952
7      D_2      -0.1409898215            -0.1409898215
8      D_0_1    -0.3856199773            -0.0666209050
9      D_0_2    -0.3946343881            -0.0073336894
10     D_1_2    -0.2598940406            -0.0462160239
11     D_0_1_2  -0.5808648709            -0.0007053587
12  -----
13     order      Incre Corr-Energy      Tot In-Corr-Energy
14  -----
15         1      -0.4599888938            -0.4599888938
16         2      -0.1201706184            -0.5801595122
17         3      -0.0007053587            -0.5808648709
18  -----
19 Scf Energy:          -152.8479006345
20 Total Incremental Correlation Energy: -0.5808648709
21 -----
22 Total MP2 Energy: -153.300852374473010 au
23 Total CCSD Energy: -153.413421130912013 au
24 Total CCSD(T) Energy: -153.428765505367011 au
25 ++++++
26 Hartree-Fock Time:          1s
27 Decomposition Time:        1s
28 Correlation Wall Time:     15s

```

Most of the items are self-explained. For instance, the line starting with D_0_1 is the calculation of increment ϵ_{01} ; the Total CCSD(T) Energy etc. is the corresponding *final* incremental CCSD(T) energy. Note that the MP2 energy is usually not useful.

3.3.2 Visualization of Domains

To visualize of domains, an easy way is to use the following TCL script (with file name domainshow.tcl) with VMD.

Code 3.5: TCL script for visualization of domains.

```

1  # domainshow.tcl
2  proc domainshow { fn } {
3      set id [mol load pdb $fn]
4      mol modselect 0 $id "chain_A"
5      mol modstyle 0 $id "Licorice" 0.100000 10.000000 10
6      .000000
7      mol modcolor 0 $id "ResId"
8      mol addrep $id
9      mol modselect 1 $id "chain_C"
10     mol modstyle 1 $id "VDW" 0.100000 20.000000

```

11 }

1. Get the file `ch3cho_domains.pdb` from the directory `ch3cho_2341`;
2. Load VMD and open the TkConsole;
3. Load the TCL script by: `source domainshow.tcl`;
4. Run the command: `domainshow ch3cho_domains.pdb`, done!

In the OpenGL window one can see the domains (see Figure 3.3.1). One can further render it with more configurations.

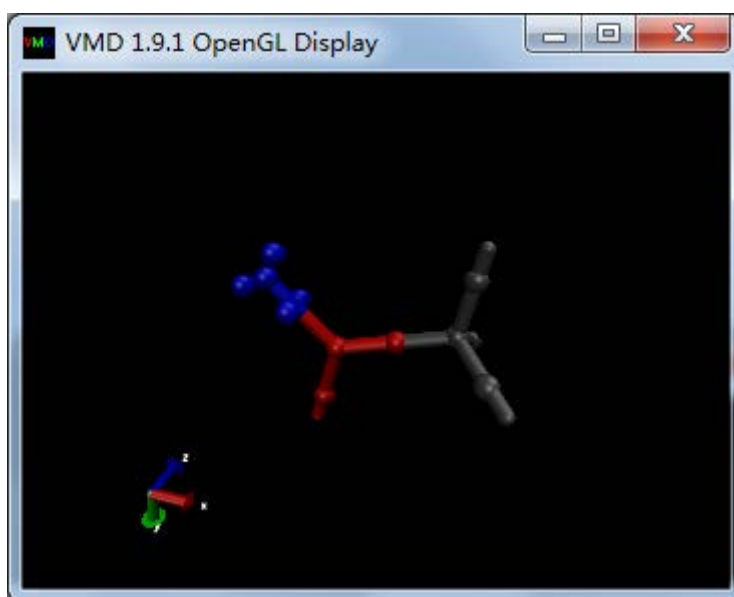


Figure 3.3.1: Visualization of domains by VMD.

3.4 Special Issues for F12 and High-spin Open-shell Calculations

3.4.1 Perform a F12 Calculation

For a F12 calculation, one must add a “F12 correction term”. Taking `inc3-db-B0-CCSD(T)-F12/VDZ` of CH3CHO as an example, one should first run the following calculation:

Code 3.6: Input file for `inc3-db-B0-CCSD(T)-F12/VDZ` of CH3CHO.

```

1 0          # 0 new mission; pid: restart
2 ch3cho.xyz # Molecule coordinates
3 0          # Charge
4 1          # Multiplicity, i.e. 2S+1
5 boys      # Localization strategy: boys
6 3          # Number of domains
7 3          # Order of incremental expansion
8 6.0       # Truncation distance

```

```

9 0          # Distance truncation flag
10 0         # Core correlation
11 No        # (Not used yet)
12 0         # (Not used yet)
13 vdz       # Basis set
14 (p/s)     # Dual basis set flag
15 0.1       # Buffer size
16 0         # ECP information
17 0         # Frozen core in ECP
18 ccsd(t)-f12 # Correlation method
19 2000      # Memory size, in MW
20 molpro    # Quantum Chemistry Program
21 molpro.exe -W /scratch/$USER -d /scratch/$USER # cmd
22 0         # Orbital cubes
23 ./nodes   # File containing nodes to parallelization
24 0         # Canonicalization
25 /scratch/ # Scratch path
26 3.2d-5    # OSV parameter
27 x         # Dummy atoms
28 0         # CC shift

```

The output file contains:

Code 3.7: Output file for inc3-db-B0-CCSD(T)-F12/VDZ of CH₃CHO.

```

1 WARNING: For F12 energies, one must add the "F12 singles
   correction" to the total energies!!
2 Total CCSD-F12a Energy: -153.554846584366004 au
3 Total CCSD-F12b Energy: -153.539346083519007 au
4 Total CCSD(T)-F12a Energy: -153.569110386199014 au
5 Total CCSD(T)-F12b Energy: -153.553609885352017 au

```

Now, perform a *standard* CCSD(T)-F12/VDZ for CH₃CHO. Of course this could be long, but one does not need to wait for the entire calculation to be accomplished, but just wait until the following output appears:

Code 3.8: Output file for CCSD(T)-F12/VDZ of CH₃CHO.

```

1 CABS-singles contribution of -0.04551365 patched into
   reference energy

```

Now, add -0.04551365 to -153.569110386199014 and -153.553609885352017 , one can obtain the *final* results of inc3-db-B0-CCSD(T)-F12/VDZ calculation. This is exactly (2.5.2).

3.4.2 Perform a High-spin Open-shell Calculation

For instance, to calculate inc3-db-B0-CCSD(T)/VDZ of a triplet state of CH₃CHO, one should first run the following calculation:

Code 3.9: Input file for inc3-db-B0-CCSD(T)/VDZ of a triplet state of CH₃CHO.

```

1 0          # 0 new mission; pid: restart
2 ch3cho.xyz # Molecule coordinates
3 0          # Charge
4 3          # Multiplicity, i.e. 2S+1

```



```

5 | boys      # Localization strategy: boys
6 | 3         # Number of domains
7 | 3         # Order of incremental expansion
8 | 6.0      # Truncation distance
9 | 0         # Distance truncation flag
10 | 0         # Core correlation
11 | No        # (Not used yet)
12 | 0         # (Not used yet)
13 | vdz      # Basis set
14 | (p/s)    # Dual basis set flag
15 | 0.1      # Buffer size
16 | 0         # ECP information
17 | 0         # Frozen core in ECP
18 | rccsd(t) # Correlation method
19 | 2000     # Memory size, in MW
20 | molpro   # Quantum Chemistry Program
21 | molpro.exe -W /scratch/$USER -d /scratch/$USER # cmd
22 | 0         # Orbital cubes
23 | ./nodes  # File containing nodes to parallelization
24 | 0         # Canonicalization
25 | /scratch/ # Scratch path
26 | 3.2d-5   # OSV parameter
27 | x         # Dummy atoms
28 | 0         # CC shift

```

The output file contains:

Code 3.10: Output file for inc3-db-B0-CCSD(T)/VDZ of triplet state of CH₃CHO.

```

1 | WARNING: For RCCSD(T) energies, one must add the ACTIVE
   | correction to the total energies!!
2 | Total MP2 Energy: -152.257146651653017 au
3 | Total CCSD Energy: -153.256639898830002 au
4 | Total CCSD(T) Energy: -153.268821998830020 au

```

Now, perform a dual basis set CCSD(T)/VDZ of a triplet state of CH₃CHO, with *only the singly occupied orbitals correlated*. One can copy the file `ch3cho_corr_0.inp` into `ch3cho_open.inp` and change the input below the molecular geometry as:

Code 3.11: Input file for standard dual basis set CCSD(T)/VDZ of triplet state of CH₃CHO, with only the singly occupied orbitals are correlated.

```

1 | ! 1-site domain correlation
2 | basis={
3 | default, vdz;
4 | }
5 | {rhf; maxit, 1; shift, 1.e+10, 1.e+10; wf, 24, 1, 2; start,
   | 2102.2; save, 2500.2}
6 | {locali, boys; core, 3; print, orbital, charge; save,
   | 2500.2; order, fock}
7 | {rccsd(t); maxit, 100; THRESH,ENERGY=1.0D-05,COEFF=1.0D
   | -03;occ, 13; closed, 11; core, 11; wf, 24, 1, 2; orbital,
   | 2500.2, ignore_error=1;}

```

Run this calculation by MOLPRO. The output contains:

Code 3.12: Output file for the MOLPRO calculation.

```
1 Total correlation energy                -0.004083874101
```

Now, add -0.004083874101 to -153.268821998830020 one can obtain the *final* results of the inc3-db-B0-CCSD(T)/VDZ calculation of the triplet state of CH_3CHO .

Chapter 4

Applications of the Incremental Scheme

4.1 Systems from Benchmark Sets

In recent years, numerous benchmark studies appeared establishing the performance of quantum chemical methods in specific classes of systems. The intrinsic accuracy of an *ab initio* method can only be determined when the BSIE is kept small. For benchmark sets of small systems, such as the halogen bonds set XB18 [166], the CBS energies are obtained from an AVQZ-AV5Z extrapolation. If the systems are larger, such extrapolation is too expensive, and several semi-empirical procedures were proposed. For instance, in the biomolecular interaction energies set S66 [167], CCSD(T)/CBS energies were estimated as:

$$E(\text{CCSD(T)/CBS}) = E(\text{HF/AVQZ}) + E_{\text{corr}}(\text{MP2/CBS}) + \Delta E(\text{CCSD(T)}) \quad (4.1.1)$$

$$\Delta E(\text{CCSD(T)}) = E(\text{CCSD(T)/AVDZ}) - E(\text{MP2/AVDZ}) \quad (4.1.2)$$

In the large noncovalent complexes set L7 [168] a similar procedure was used. Here we use inc3-db-B0-F12 to estimate the CBS interaction energies for some selected cases to examine the performance of our new method as well as to test the validity of (4.1.1) and (4.1.2)[132].

The following complexes were selected: CH₃OH-peptide, pyridine-pyridine(π - π), CH₃NH₂-pyridine and N(CH₃)₃-chlorobenzene. The first two are electrostatic and dispersion bound complexes, respectively, and the last two are of mixed intermolecular force nature. The geometries of the first three and the last one are extracted from the S66 [166] and X40 [169] sets, respectively. We have performed inc3-db-B0 and exact (if possible) calculations on the complexes, but only exact ones on monomers.

For all the systems we set $n_d = 6$ and first performed the computations without distance screening. In Table 4.1.1 we list the timings of those calculations. There are two competing factors determining the efficiency: increasing the number of domains raises the computational cost since there are more increments to evaluate; however, since each increment becomes smaller the time saving brought by the db-B0 scheme increases. Therefore, our approach will be more efficient for larger basis sets. For the AVDZ, the inc3-db-B0 calculations take slightly longer than the traditional implementation, since the two factors counteract each other to some degree. But at the AVTZ level, the inc3-db-B0 approach undoubtedly outperforms: our method can save up to 30% of the real time and by parallelization, calculations requiring usually several months become feasible within a few days. For N(CH₃)₃-chlorobenzene, we found from the AVDZ calculation

that a truncation distance $r_t = 5.4 \text{ \AA}$ can be applied with little accuracy loss, leading to a reduction of the real time by 33 days at the AVTZ level. By parallelization with 10 to 20 CPU cores, which are available as standard hardware nowadays, these calculations can be accomplished in less than 5 days. Thus our approach is significantly faster than the traditional implementation and can be efficiently parallelized.

Table 4.1.1: Efficiency of inc3-db-B0-F12 methods. All the calculations were performed with the Intel(R) Xeon(R) E5-4620 CPU cores and 16 GB RAM per core.

system	ne/nbf ^a	inc3-db-B0		ex-db
		wall time ^b	real time	real time
AVDZ				
CH ₃ OH-peptide	44/260	2.9 hours (10)	19.8 hours	12.9 hours
pyridine-pyridine(π - π)	60/366	8.7 hours (20)	5.0 days	4.9 days
CH ₃ NH ₂ -pyridine	44/274	3.8 hours (10)	21.1 hours	18.5 hours
N(CH ₃) ₃ -chlorobenzene	62/383	18.6 hours (20)	7.8 days	6.1 days
N(CH ₃) ₃ -chlorobenzene(t) ^c	62/383	16.6 hours (20)	6.0 days	6.1 days
AVTZ				
CH ₃ OH-peptide	44/575	1.0 days (20)	13.0 days	16.4 days
pyridine-pyridine(π - π)	60/782	4.4 days (20)	72.1 days	N/A ^d
CH ₃ NH ₂ -pyridine	44/598	2.5 days (10)	13.3 days	20.2 days
N(CH ₃) ₃ -chlorobenzene	62/832	14.4 days (20)	87.9 days	N/A ^d
N(CH ₃) ₃ -chlorobenzene(t) ^c	62/832	4.2 days (20)	54.3 days	N/A ^d

^a ne: number of correlated electrons; nbf: number of basis functions.

^b The number in parentheses is the number of CPU cores for parallelization.

^c Performed with truncation distance $r_t = 5.4 \text{ \AA}$.

^d Beyond our computational ability.

Another important aspect is accuracy. We observe that the inc3-B0 error is very small, i.e., less than 0.15 and 0.04 kcal mol⁻¹ for CCSD(T)-F12 x at the AVDZ and AVTZ level, respectively. The error trend is very similar for F12a and F12b. This suggests that the inc3-B0 scheme grasps the essence of the locality of electronic correlation, approaching the exact energy in high precision with minimal computational cost. Maximal errors occur for the dispersion bound pyridine-pyridine(π - π) complex, implying the delocalized nature of the stacked aromatic system and the sensitivity to the basis set. Nevertheless, at the inc3-db-B0-CCSD(T)-F12a/AVTZ level, the energies agree with the standard CCSD(T)-F12a/AVTZ within 0.03 kcal mol⁻¹, and the CCSD(T)/CBS estimations within 0.24 kcal mol⁻¹. The better performance of CCSD(T)-F12a is compatible with the observation that for AVDZ and AVTZ F12a is closer to the CBS limit than F12b[36]. Thus our inc3-db-B0 approach can work with CCSD(T)-F12 theory perfectly to obtain the near CBS energies. At the same time, we also confirm the validity of (4.1.1) and (4.1.2).

Summing up, with the inc3-db-B0-CCSD(T)-F12a/AVTZ method one can obtain energies of CCSD(T)/AV5Z quality which at present corresponds to the limit of accuracy that can be reached for large systems. The results obtained agree with the CBS limit within a few tenths of a kcal mol⁻¹. Note that our method is trying to approximate the absolute energy, thus it is very flexible because one can apply the inc3-db-B0 and traditional implementations simultaneously in practice for molecules of different sizes, e.g. to evaluate the interaction energies. The calculations for large systems of about 60 correlated electrons and 800 basis functions can be accomplished in only 5 days with

Table 4.1.2: Interaction energies by inc3-db-B0-F12 methods. Energy unit: kcal mol⁻¹.

method	AVDZ		AVTZ		CBS in literature
	inc3-db-B0	standard	inc3-db-B0	standard	
CH ₃ OH-peptide					
CCSD-F12a	6.00	6.18	5.81	5.90	
CCSD-F12b	5.93	6.14	5.84	5.94	
CCSD(T)-F12a	6.69	6.75	6.43	6.45	6.19 ^a
CCSD(T)-F12b	6.62	6.70	6.46	6.48	
pyridine-pyridine(π - π)					
CCSD-F12a	3.12	3.82	2.56	N/A ^c	
CCSD-F12b	2.94	3.66	2.64	N/A ^c	
CCSD(T)-F12a	4.81	5.33	4.09	N/A ^c	3.90 ^a
CCSD(T)-F12b	4.62	5.16	4.18	N/A ^c	
CH ₃ NH ₂ -pyridine					
CCSD-F12a	3.54	3.87	3.27	3.37	
CCSD-F12b	3.45	3.79	3.33	3.40	
CCSD(T)-F12a	4.48	4.75	4.11	4.14	3.97 ^a
CCSD(T)-F12b	4.34	4.60	4.16	4.17	
N(CH ₃) ₃ -chlorobenzene					
CCSD-F12a	1.99	2.03	1.65	N/A ^c	
CCSD-F12b	1.94	1.98	1.69	N/A ^c	
CCSD(T)-F12a	2.60	2.63	2.30	N/A ^c	2.11 ^b
CCSD(T)-F12b	2.56	2.58	2.34	N/A ^c	
N(CH ₃) ₃ -chlorobenzene(t) ^d					
CCSD-F12a	1.85	2.03	1.59	N/A ^c	
CCSD-F12b	1.80	1.98	1.63	N/A ^c	
CCSD(T)-F12a	2.42	2.63	2.21	N/A ^c	2.11 ^b
CCSD(T)-F12b	2.39	2.58	2.25	N/A ^c	

^a From S66 set[166].

^b From X40 set[169].

^c Beyond our computational ability.

^d Performed with truncation distance $r_t = 5.4 \text{ \AA}$.

10 to 20 CPU cores. We also point out that of course the inc3-db-B0-F12 approach can also obtain near CBS limit of the CCSD energy. Thus, the inc3-db-B0-CCSD(T)-F12a/AVTZ is able to serve as benchmark reference!

4.2 Water Clusters

4.2.1 Minimum Structure of Water Hexamers

The water hexamer has been attracting the attention of the scientific community for a long time [170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180], since it has the size at which the stable isomer changes from a two-dimensional (2D) to three-dimensional (3D) structure. For the hexamer, the cage and prism isomer (see Figure 4.2.1) are known to be more stable than the other isomers. However, which one is most stable has puzzled investigators for a long time, since they are nearly degenerate in energy. In 2012 a

rotational spectra study suggested that the cage isomer is the minimum structure [175], whereas in the same year a quantum simulation with dynamic effects taken into account argued that the prism is more stable [177]. Since the energy difference between them is at subchemical accuracy (about $0.2 \text{ kcal mol}^{-1}$), a quantum chemical method applied to the problem must be extremely accurate to be successful. The inc3-db-B0 approach can be very efficient for water clusters since the LMOs can be well localized on each water molecule. However, it is also a challenge that due to the small interaction energy, the method must be highly accurate or it may even predict the wrong sign. We performed inc3-db-B0-CCSD(T)-F12 calculations on the two isomers with $n_d = 6$, since in this case the KM algorithm decomposes the cluster perfectly into 6 water domains. The results are listed in Table 4.2.1.

Table 4.2.1: Inc3-db-B0-F12 results of water hexamers. The interaction energy is defined as: $E_{\text{cage}} - E_{\text{prism}}$ with units: kcal mol^{-1} . All the calculations were performed with the Intel(R) Xeon(R) CPU E7-8837 CPU cores and 8 GB RAM per core.

Method	AVDZ		AVTZ	
	inc3-db-B0	standard	inc3-db-B0	standard
CCSD-F12a	0.13	0.23	0.16	0.20
CCSD-F12b	0.14	0.22	0.17	0.20
CCSD(T)-F12a	0.16	0.25	0.22	0.24
CCSD(T)-F12b	0.17	0.24	0.23	0.24
CCSD(T)/CBS ^a	0.25			
ne/nbf ^b	48/246		48/552	
prism isomer				
real time	18.8 hours	14.1 hours	8.6 days	16.1 days
wall time ^c	2.7 hours		1.0 days	
cage isomer				
real time	16.0 hours	13.7 hours	10.6 days	16.1 days
wall time ^c	2.3 hours		1.3 days	

^a Geometries and CCSD(T)/CBS results are taken from Bates' work[174].

^b ne = number of correlated electrons; nbf = number of basis functions.

^c Parallelized by 10 cores.

We can see from Table 4.2.1 that the inc3-db-B0 approach produces nearly the same CCSD(T)-F12/AVTZ energy, which differs from the CCSD(T)/CBS for only $0.02 \text{ kcal mol}^{-1}$. Since this CBS energy is also estimated by an empirical procedure, our result might be even more accurate. Comparing with a very recent study which gives this energy as $0.44 \text{ kcal mol}^{-1}$ at the CCSD(T)/AVDZ level[178], the importance of the elimination of the BSIE is here emphasized again. Even CCSD(T)-F12/AVDZ improves the result significantly. We also see that for AVTZ our method can save almost 50% of the real time necessary for the exact dual-basis set approach and by parallelization, the calculations required only 1 day! We also conclude that the prism isomer is more stable in the nonrelativistic electronic energy. To further clarify the stability order, other physical effects like nuclear and thermal motions must be accurately treated.

4.2.2 Larger Clusters

To further examine the performance of the inc3-db-B0 approach, we performed a series of CCSD(T)/VDZ calculations on some isomers of water hexamer[178], octamer[181],

11-mer[182] and 17-mer[183], the structures of which are shown in Figure 4.2.1. The inc3-db-B0, as well as two other approaches, i.e. OSV[100] and DLPNO[98] schemes, are used for comparison. The results are listed in Table 4.2.2.

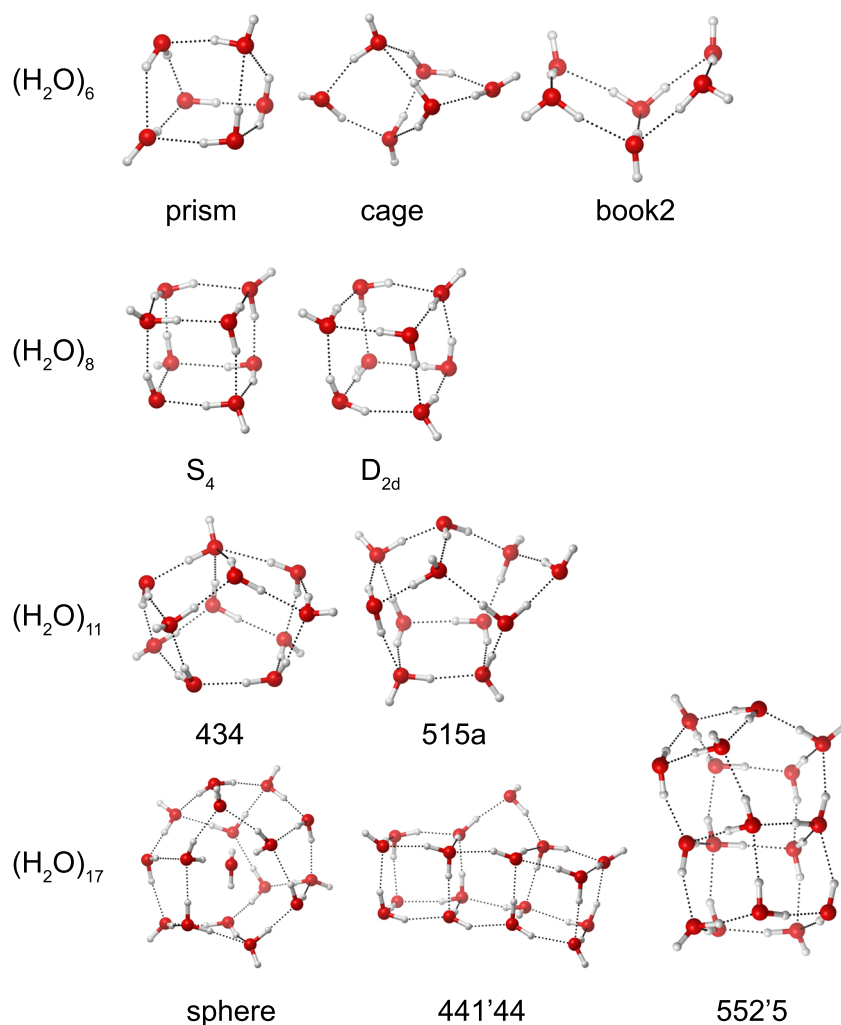


Figure 4.2.1: Some isomers of the water clusters.

From the aspect of accuracy, the incremental scheme always outperforms DLPNO and OSV: the error of inc3-db-B0 is always less than 0.02 kJ mol^{-1} , at the level of subchemical accuracy. Even an inc2-db-B0 implementation exhibited excellent performance, of course this is due to the nature of molecular cluster and for other systems, we cannot expect that “inc2” always shows such high accuracy. The error of the other two approaches is often larger than 1 kJ mol^{-1} , especially that of the binding energy is dozens of kilojoule per mole! Our inc3-db-B0 approach is obviously advanced in accuracy and can be used to treat systems with very small energy difference safely. OSV and DLPNO are perhaps faster than ours, however the cost is the loss of accuracy. In fact, For 17-mers, DLPNO has identified an incorrect minimum! Therefore, from the case of water clusters, the excellent performance of the inc3-db-B0 approach in relative energy is highlighted.

Table 4.2.2: Errors of DLPNO-, OSV- and inc(2/3)-db-B0-CCSD(T)/VDZ methods.

Errors of binding energy ($E[(\text{H}_2\text{O})_n] - nE[\text{H}_2\text{O}]$ in kJ mol^{-1}) ^a					
Isomer	DLPNO	OSV	inc2-db-B0	inc3-db-B0	CCSD(T)
book2	5.02	16.38	0.66	0.08	Ref
cage	5.75	17.69	0.67	0.12	Ref
prism	6.93	17.99	0.73	0.15	Ref
D _{2d}	9.05	28.51	1.51	0.16	Ref
S ₄	7.61	28.56	1.43	0.16	Ref
515a	14.13	39.95	2.29	0.17	Ref
434	16.73	40.66	2.23	0.15	Ref
552'5	30.30	71.89	1.67	Ref	N/A ^b
sphere	29.99	75.83	1.65	Ref	N/A ^b
441'44	26.42	74.34	3.03	Ref	N/A ^b
Errors of conformation energy ($E[\text{isomer}] - E[\text{minimum isomer}]$ in kJ mol^{-1}) ^a					
Isomer	DLPNO	OSV	inc2-db-B0	inc3-db-B0	CCSD(T)
book2	-1.91	-1.61	-0.07	-0.07	Ref
cage	-1.17	-0.30	-0.07	-0.04	Ref
prism	minimum isomer				
D _{2d}	1.44	-0.06	0.08	0.01	Ref
S ₄	minimum isomer				
515a	-2.60	-0.72	0.06	0.02	Ref
434	minimum isomer				
552'5	3.87	-2.45	-1.35	Ref	N/A ^b
sphere	3.57	1.49	-1.38	Ref	N/A ^b
441'44	minimum isomer				

^a In the table, the listed numbers of the errors of the corresponding item relative to the ‘‘Ref’’ one. For instance, the binding energy of the book2 isomer of the water hexamer calculated by DLPNO- and standard CCSD(T)/VDZ is -249.76 and -254.78 kJ mol^{-1} , respectively, thus the error is $(-249.76) - (-254.78) = 5.02$ kJ mol^{-1} .

^b Beyond our computational ability.

4.3 The Rotation Barrier of Biphenyl

The rotation barrier of the biphenyl 1-1' bond is unexpectedly difficult to compute[184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194]. An experiment reported that the barrier of rotation occurs at $44.4 \pm 1.2^\circ$ [195] and the height from the 0° conformation is 1.4 ± 0.5 kcal mol^{-1} [196]. However most calculations gave a too high barrier, and Johansson's high-level theoretical treatment[194], including core correlation energy, extrapolation to CBS and FCI limit, relativistic effects, intramolecular basis set superposition error correction and thermal correction, predicted a barrier of 1.9 kcal mol^{-1} . We tried to compute the valence correlation part[132] with the inc3-db-B0-F12 approach. This might be a difficult case for the inc3-db-B0 approach due to the different conformations of the molecule involved and the aromaticity. The computed barrier height by inc3-db-B0-CCSD(T)-F12 is shown in Table 4.3.1.

From Figure 4.3.1, we see that for the two conformations the domain decompositions remain nearly the same, indicating the stability of our KM decomposition algorithm, which is important for a PES exploration. The results in Table 4.3.1 confirm the accuracy of our approach. The CCSD(T)/AVQZ barrier 2.01 kcal mol^{-1} is only 0.02 kcal

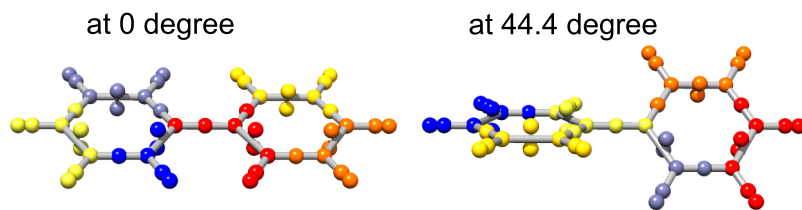


Figure 4.3.1: Biphennyls. The balls stand for the centers of the LMOs and those with the same color constitute a domain.

Table 4.3.1: Inc3-db-B0-F12 results for the rotation barrier of biphenyl. The interaction energy is defined as: $E_{44^\circ} - E_{0^\circ}$ with unit: kcal mol⁻¹. All the calculations were performed with the Intel(R) Xeon(R) E5-4620 CPU cores and 16 GB RAM per core.

Method	AVDZ		AVTZ	
	inc3-db-B0	standard	inc3-db-B0	standard
CCSD-F12a	2.30	2.34	2.22	N/A ^b
CCSD-F12b	2.31	2.35	2.25	N/A ^b
CCSD(T)-F12a	2.19	2.23	1.99	N/A ^b
CCSD(T)-F12b	2.20	2.23	2.02	N/A ^b
CCSD/AVQZ ^a	2.25			
CCSD(T)/AVQZ ^a	2.01			
ne/nbf ^c	58/366		58/782	
0° conformation				
real time	4.4 days	4.7 days	57.4 days	N/A ^b
wall time ^d	15.8 hours		6.9 days	
44° conformation				
real time	4.7 days	4.7 days	58.9 days	N/A ^b
wall time ^d	16.8 hours		8.3 days	

^a Geometries and CCSD(T)/AVQZ results are taken from Johansson’s work[194].

^b Beyond our computational ability.

^c ne = number of correlated electrons; nbf = number of basis functions.

^d Parallelized by 10 cores.

mol⁻¹ higher than the inc3-db-B0-CCSD(T)-F12a/AVTZ result, suggesting that the valence correlation component of the barrier is of high accuracy. Moreover, we note that our approach can accurately predict the barrier accurately with a relatively small basis set (AVTZ) and ordinary hardware (10 CPU cores) in 8 days. An traditional implementation of CCSD(T)/AVQZ, requires abundant computational resources.

4.4 Hydration of Trivalent Lanthanide Ions

Ion hydration is a fundamental phenomenon in nature, determining the solvation dynamics, chemical reactivity and several biological as well as industrial processes. Of the numerous possible atomic cations and anions, the hydration of lanthanides(III) has been an active subject of research for a long time since it is involved in many practical applications. For the extraction and separation of lanthanides, hydration energies and kinetics are necessary quantities to calculate the relative selectivities and binding rates

in the development of effective extracting ligands or solvents[197, 198].

The lanthanide(III) hydration is also of academeical interest because the very subtle changes in electronic structures from La^{3+} to Lu^{3+} nevertheless induce quite complex hydration behaviors[199]. It is accepted that the number of water molecules in the first hydration shell, i.e. the coordination number (CN), is 9 and 8 for light (e.g. La^{3+} , Ce^{3+}) and heavy (e.g. Yb^{3+} , Lu^{3+}) lanthanides, respectively[199]. From Figure 4.4.1 we see that the octa- and nona-lanthanide(III) aqua complexes possess a square antiprism (SAP) and tricapped trigonal prism (TTP) structure, respectively[199].

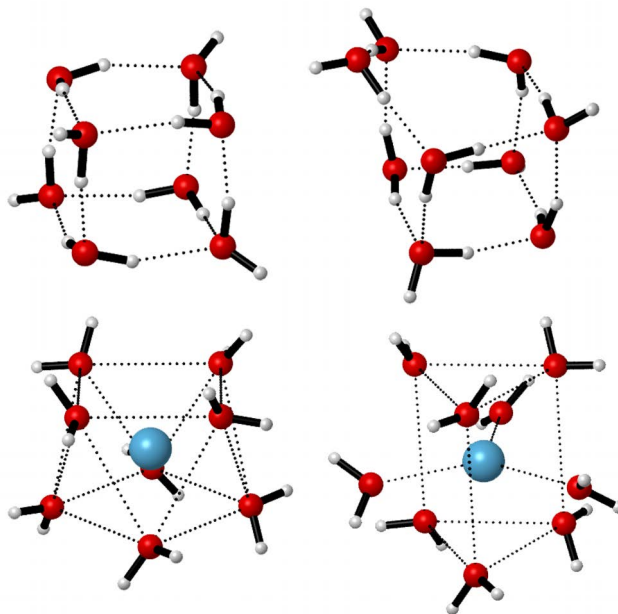


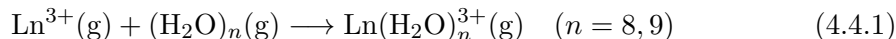
Figure 4.4.1: Geometries of water clusters $(\text{H}_2\text{O})_n$ and lanthanide(III) aqua complexes $\text{Ln}(\text{H}_2\text{O})_n^{3+}$ ($n = 8, 9$).

Modern experimental studies of lanthanide(III) hydration are nearly always coupled with theoretical methods. The X-ray absorption spectroscopy method is the most important one in unraveling lanthanide(III) hydration, and the interpretation of its results must rely on some models and simulations[200, 201, 202]. From a pure theoretical point of view, molecular dynamics and quantum chemistry have both been applied to this field. For quantum chemistry one should realize that the large number of electrons, strong relativistic effects and incomplete 4f shell occupation in lanthanides make their study quite difficult. Nevertheless, the development of ECP improves the computational ability for lanthanides. The large- and small-core energy-consistent PPs for lanthanides[203, 204, 205] have proven to be both accurate and efficient. Both sets of PPs were applied in studies of lanthanide(III) hydration [206, 207, 208, 209, 126].

While DFT and MP2 have been applied to the lanthanide(III) aqua complexes [206, 207, 208, 209], the intrinsic accuracy of these methods remains unknown, because a benchmark with a highly accurate *ab initio* method such as CCSD(T) used to be impossible due to the large computational effort. Thus, the inc3-db-B0 approach was used to perform CCSD(T) calculations on $\text{Ln}(\text{H}_2\text{O})_n^{3+}$ ($n = 8, 9$) in gas phase to produce very accurate results as well to examine the reliability of the previous studies.

The octa- and nona-aqua lanthanide(III) complexes, i.e. $\text{Ln}(\text{H}_2\text{O})_8^{3+}$ and $\text{Ln}(\text{H}_2\text{O})_9^{3+}$, the free Ln^{3+} ions as well as the water clusters $(\text{H}_2\text{O})_8$ and $(\text{H}_2\text{O})_9$ with are considered[126]. For hydrogen and oxygen aug-cc-pVTZ basis sets were used. For the lanthanides, the core shell electrons were substituted by scalar-relativistic 4f-in-core PPs [203]. The va-

lence electrons were represented by (8s7p6d3f2g)/[6s5p5d3f2g] basis sets, which contain a set of (2s1p1d) diffuse and (3f2g) polarization functions[210]. The geometries optimized at the MP2 level were obtained from Ciupka’s work[209] and are shown in Figure 4.4.1. The energies are calculated with the inc3-db-B0-CCSD(T) method. The gas phase binding energies according to the following reaction in gas phase are calculated:



In Table 4.4.1 and Figure 4.4.2 we show the errors of HF, B3LYP, CCSD[126] and MP2, SCS-MP2[209] energies with respect to the CCSD(T) reference. We observe that MP2 and CCSD overestimate the binding energies by about 35 and 5 kJ mol⁻¹, respectively, and the errors for octa- and nona-aqua ions are almost identical. In contrast B3LYP underestimates the binding energies and exhibits size-dependent errors. For octa-aqua ions its error (15.64 kJ mol⁻¹) is much smaller than for MP2, and for nona-aqua ions (37.62 kJ mol⁻¹) slightly larger. A similar trend is observed in the case of SO₄²⁻(H₂O)_n clusters [211]: as *n* increases from 3 to 6, the binding energy error of MP2 changes from 1.33 to 4.64 kJ mol⁻¹ whereas, that of B3LYP increases drastically from 1.17 to 26.48 kJ mol⁻¹. These observations suggest that for these aqua complexes, the error of traditional DFT, at least for B3LYP, increases with the system size, whereas MP2 is more robust.

Table 4.4.1: The RMSDs (unit: kJ mol⁻¹) of the binding energy errors with respect to the CCSD(T) reference for all the considered methods.

Method	Ln(H ₂ O) ₈ ³⁺	Ln(H ₂ O) ₉ ³⁺
HF	3.05	12.71
MP2	35.09	32.06
CCSD	5.72	5.19
B3LYP	15.64	37.62
SCS-MP2	5.13	5.11

Astonishingly, the accuracy of HF is better than the one of MP2 and B3LYP. However, the unsystematic behavior of errors reflects that this is an example of “right answer for wrong reason”, i.e. a lucky error cancellation. SCS-MP2 shows an excellent performance, comparable with the one of CCSD, however the error changes from an overestimation at La gradually to an underestimation at Lu. Splitting the MP2 correlation energy contribution to the energy change of reaction (4.4.1) into triplet and singlet components suggests that the triplet component always favors reaction (4.4.1), whereas the singlet component always disfavors it, implying the former being overestimated or/and the latter underestimated. This is exactly what the spin-component-scaling approach tries to correct and the physical reason why SCS-MP2 improves MP2 here significantly[126].

By CCSD(T) calculations, we proved that B3LYP and MP2 have similar accuracy and SCS-MP2 is nearly comparable with CCSD. However B3LYP as well as MP2 cannot treat the subtle energy change well. This reminds us again that results of DFT or even low-level *ab initio* correlation methods like MP2 should be viewed with caution. SCS-MP2 has been proved very accurate thus its prediction of CNs by Ciupka[209] is reliable. Moreover, combining Ciupka’s[209] and the author’s[126] work, the most accurate hydration Gibbs free energies from first principles so far are obtained[126], see Table 4.4.2: the RMSD of the errors relative to the experimental results[212] is only 25 kJ mol⁻¹. Thus the inc3-db-B0 approach as well as the computational strategies proposed in these works[209, 126] can be applied to, at least, any hydrated charged

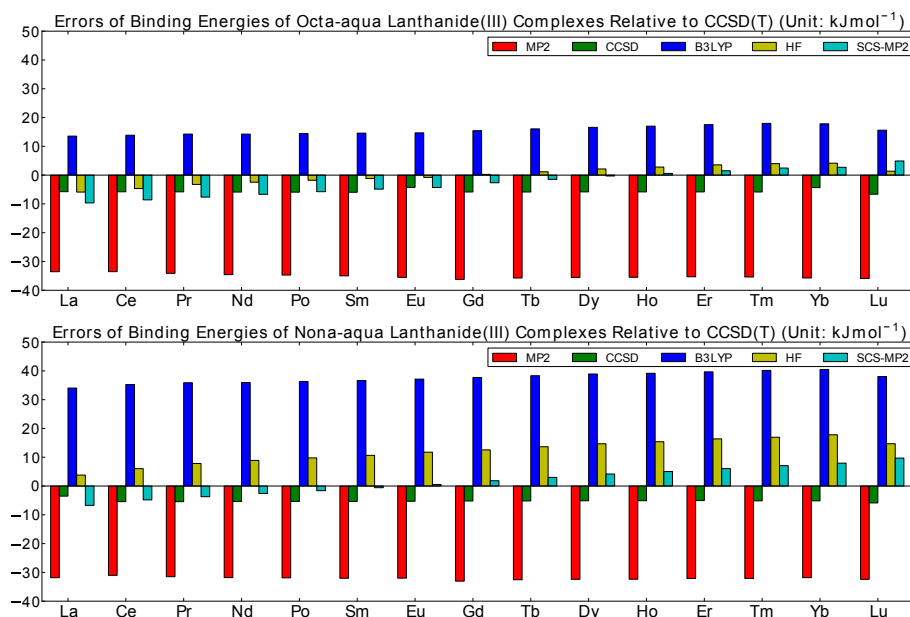


Figure 4.4.2: The binding energy errors (unit: kJ mol⁻¹) with respect to the CCSD(T) reference for all the considered methods. The errors of method X here are defined as $D_{\text{CCSD(T)}} - D_X$, where D_X is the binding energy computed by method X .

compound systems to get accurate energies. This is especially important for chemical species that lack experimental data. In fact, a similar study on actinide(III) hydration has been carried out.

4.5 The Relative Stability of Isomers of Double Fullerene Adducts

Steric effects are generally accepted as a key factor that determines the stability of a chemical system. It originates from Pauli's principle: as two molecular fragments approach, their electron clouds will repel each other in order to decrease the overlapping. Thus, a molecule with compact distribution of non-polar functional groups (i.e. excluding hydroxyl groups or halogens, etc., where electrostatic or hyperconjugation interactions could play a significant role) is usually thought to be less stable than one with a more extensive distribution. A textbook example is that *cis*-2-butene and *cis*-dimethylcyclopropane are less stable than their *trans* isomers. A seemingly exception is that *cis*-1,3-dimethylcyclobutane is more stable than the *trans* isomer. The reason is that in the *cis* isomer both methyl groups can occupy the equatorial position to stay far away from the neighbor C–H bonds, while in the *trans* isomer one methyl group has to be in the axial position, experiencing a larger steric congestion with the neighbor C–H bonds. Thus, the stability order is essentially still determined by the steric effect. Knowing these examples, it is not surprising that for the double C₆₀ adduct of pentacene **1** (see Figure 4.5.1) the sterically less crowded *anti*-1 form was assumed to be the most stable isomer[213].

However, one must keep in mind that the steric effect is only the repulsive part of the vdW interaction between molecules; there is another attractive part: dispersion interaction. According to London's formula, molecular fragments with a larger number of electrons and a concomitant higher polarizability will exhibit stronger dispersion

Table 4.4.2: The hydration Gibbs free energies (unit: kJ mol⁻¹) of lanthanide(III) aqua complexes.

Ln	Ln(H ₂ O) ₈ ³⁺ ^a	Ln(H ₂ O) ₉ ³⁺ ^a	Computed ^b	Experiment ^c
La	-3171.63	-3187.01	3187	3145
Ce	-3201.16	-3216.71	3217	3200
Pr	-3233.51	-3248.23	3248	3245
Nd	-3264.83	-3277.78	3278	3280
Pm	-3294.75	-3307.10	3307	3250
Sm	-3324.28	-3335.82	3336	3325
Eu	-3354.10	-3364.55	3364	3360
Gd	-3378.82	-3387.95	3388	3375
Tb	-3405.57	-3413.09	3413	3400
Dy	-3432.29	-3437.84	3437	3425
Ho	-3458.33	-3462.90	3462	3470
Er	-3484.11	-3486.18	3486	3495
Tm	-3509.48	-3508.43	3509	3515
Yb	-3534.16	-3533.05	3534	3570
Lu	-3556.26	-3553.09	3556	3515
RMSD			25	

^a The computed hydration free energies ΔG_{H} from the author's work[126].

^b Obtained by averaging ΔG_{H} of octa- and nona-aqua lanthanides(III) with a Boltzmann factor $\exp(-\Delta G_{\text{H}}/RT)$, where $T = 298.15$ K.

^c The experimental results are taken from Marcus' work[212].

interaction[214]. Therefore, for bulky functional groups dispersion interactions could have a great effect and result in many unusual chemical phenomena, like the existence of all-*meta-tert*-butylhexaphenylethane[215] and some very long alkane C-C bonds[216]. For the very large C₆₀ cage, dispersion interactions could be of considerable strength. However, whereas HF omits dispersion interactions, standard DFT is well known to fail to treat it accurately. Thus, a reliable prediction of the stabilities of *syn*-**1** and *anti*-**1** requires more advanced methods, such as *ab initio* correlation methods or DFT-D3[48].

Thus, the impact of dispersion interactions on the stability order of **1** were investigated[135]. To gain more insights, we also built three model molecules: the double cyclohexane **2**, the [20]fullerene (C₂₀) **3** and the dodecahedrane (C₂₀H₂₀) **4** adduct of pentacene. To test the accuracy we also considered the corresponding simpler adducts of 1,2,4,5-tetramethylbenzene (durene) **1'**, **2'**, **3'** and **4'**. All the molecules are shown in Figure 4.5.1.

First we consider **1'** to **4'**. We optimized their geometries at the TPSS-D3/SVP level. Since the D3 dispersion correction favors the compact structures, we also performed *ab initio* correlation computations on **2'** to **4'** to examine the possible bias. To make the calculations feasible we have used the inc3-db-B0 and the DLPNO[98] approach to perform the CCSD(T) calculations. From Table 4.5.1 we see that for **2'** at the CCSD(T)/cc-pVDZ level the two methods have an excellent performance and support the DFT-D3 results in Table 4.5.2. However, for **3'** and **4'**, where a standard CCSD(T) is unaffordable, a larger discrepancy appears and DLPNO even gives different signs for cc-pVDZ and cc-pVTZ basis sets! Note that for such large systems all local correlation methods have to discard a great number of numerical quantities to make the calculations feasible, leading to significant potential errors. It is nontrivial to judge the accuracy for

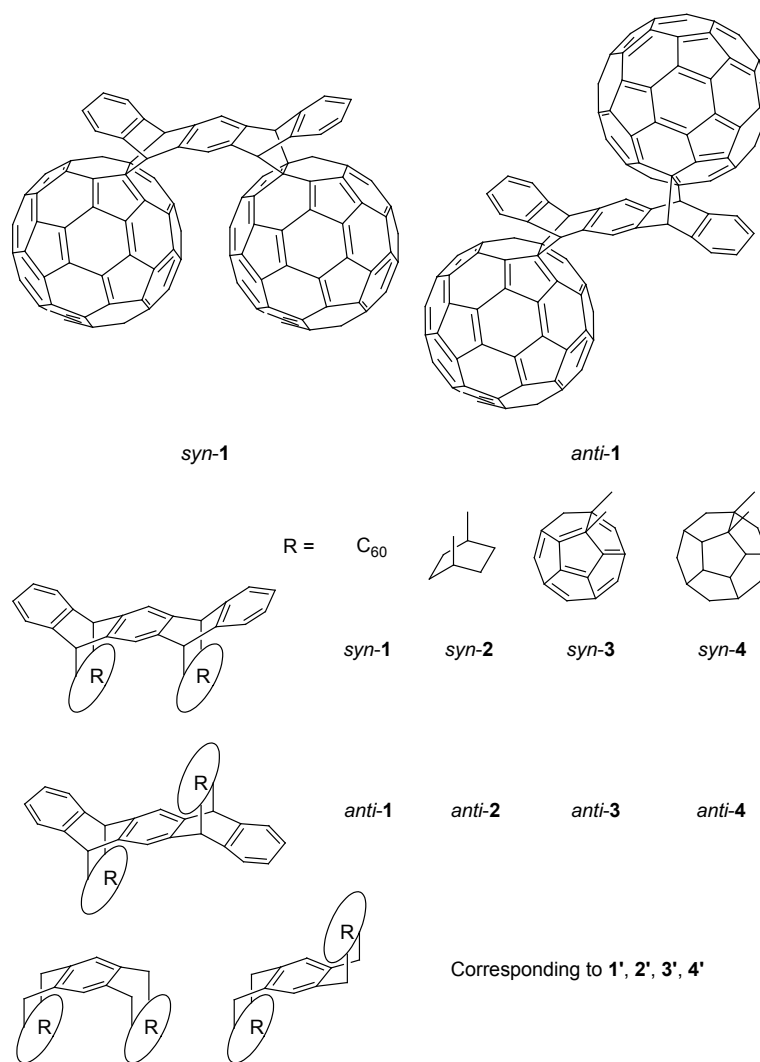


Figure 4.5.1: Molecules investigated for effects of intramolecular dispersion interaction.

3' and **4'**, but based on the performance for **2'**, we believe that DFT-D3 is fair for the relative stability of the *syn* and *anti* isomers.

Table 4.5.1: Energy differences between *syn* and *anti* isomers of **2'** to **4'** by *ab initio* correlation methods for TPSS-D3/SVP optimized geometries.

Energy ^a	2'	3'	4'
MP2/cc-pVDZ	-0.26	N/A ^b	N/A ^b
MP2/cc-pVTZ	-0.26	N/A ^b	N/A ^b
CCSD(T)/cc-pVDZ	-0.23	N/A ^b	N/A ^b
inc3-db-B0 CCSD(T)/cc-pVDZ	-0.22	-0.09	N/A ^b
DLPNO-CCSD(T)/cc-pVDZ	-0.20	-0.66	+0.88
DLPNO-CCSD(T)/cc-pVTZ	-0.06	+0.11	-0.90

^a Electronic energy difference $E(\textit{syn}) - E(\textit{anti})$ in kcal mol⁻¹.

^b Beyond our computational ability.

Confirming the accuracy of DFT-D3, we could determine the stability order of these isomers. The results in Table 4.5.2 suggest that HF and three representative functionals: B97, TPSS and B3LYP all suggest that the *anti* isomer is more stable, and for the bulky adducted fragments, the energy difference is larger, indicating a larger steric effect. However, once the D3 dispersion correction is introduced for DFT, all functionals imply that the *syn* isomer is the more stable one! Even for *syn-2'* the small cyclohexanes can invert the stability order by about 0.5 kcal mol⁻¹. In all cases, due to the dispersion stabilization the more compact isomer, i.e. the *syn* form, is more stable than the *trans* form.

Table 4.5.2: Energy differences between *syn* and *anti* isomers of all molecules considered by HF and DFT methods for TPSS-D3/SVP optimized geometries.

Energy ^a	1'	2'	3'	4'
HF/cc-pVTZ	+10.05	+0.23	+1.09	+2.09
B97/QZVPP	+7.02	+0.20	+0.73	+2.12
TPSS/QZVPP	+4.29	+0.13	+0.49	+1.25
B3LYP/TZVPP	+5.81	+0.21	+0.73	+1.65
B97-D3/QZVPP	-6.26	-0.37	-0.14	-1.25
TPSS-D3/QZVPP	-6.15	-0.34	-0.19	-1.26
B3LYP-D3/TZVPP	-5.71	-0.29	-0.02	-1.11
Energy ^a	1	2	3	4
HF/cc-pVTZ	+9.39	+1.60	+1.94	+2.03
B97/QZVPP	+6.00	+1.26	+1.08	+1.37
TPSS/QZVPP	+3.46	+0.83	+0.85	+0.73
B3LYP/TZVPP	+4.87	+0.95	+1.11	+0.86
B97-D3/QZVPP	-5.65	-0.09	-0.79	-0.88
TPSS-D3/QZVPP	-5.61	-0.06	-0.67	-0.97
B3LYP-D3/TZVPP	-5.19	-0.06	-0.53	-0.99

^a Electronic energy difference $E(\textit{syn}) - E(\textit{anti})$ in kcal mol⁻¹.

Some optimized structures as well as their NCI plots[61] are given in Figure 4.5.2. For **2'** and **2**, the dispersion stabilization effect of the latter is much weaker. For **2'** we see from Figure 4.5.2 that the two cyclohexane fragments can be in the energetically favorable “side-on” arrangement. A related fact is that some anaerobic microbe are found to use ladderane derivatives to build rigid cellular compartments[217], and computations[218] revealed that ladderanes in side-on arrangement can form quite strong interactions. In 4.5.2 due to the rigid benzene constraint the C–H bonds of both cyclohexane fragments have to point towards each other, which is quite repulsive. Thus the strong steric congestion almost cancels the dispersion stability. We also observe that the dispersion effects are very similar for the aromatic C₂₀ and saturated C₂₀H₂₀. This is also compatible with other authors’ conclusions that σ/σ , π/π , σ/π dispersion interactions share many similarities[219, 220].

Taking the thermal effects into account for **1**, we can obtain the Gibbs free energy difference: $\Delta G(\textit{syn} - \textit{anti}) = -6.36$ kcal mol⁻¹ at TPSS-D3 level, and +1.15 kcal mol⁻¹ at TPSS level. Thus, the dispersion stabilization between the two giant C₆₀ cages inverts the energy difference by more than 7 kcal mol⁻¹, making *syn-1* the more stable one. In the original study[213], *anti-1* was computed to be 0.36 kcal mol⁻¹ more stable than *syn-1* by PM3[221], reflecting the inability of this semi-empirical method to treat dispersion interactions. The more recent PM7[222] shows great improvement: using the TPSS-

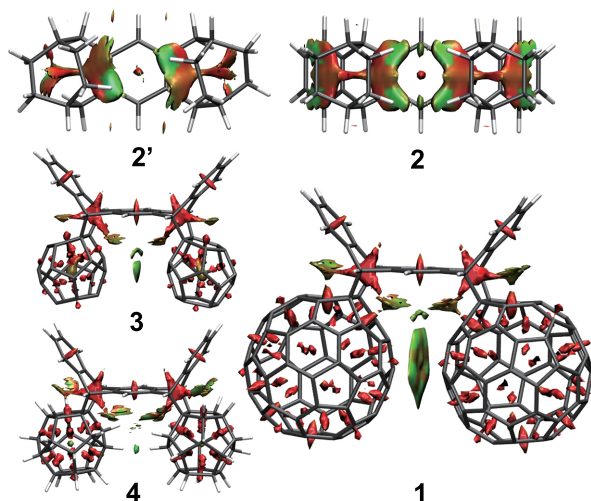


Figure 4.5.2: Geometries and noncovalent interaction visualization. For the surfaces, red regions mean strong repulsion and green ones mean weak attraction.

D3/SVP geometries of this work, the heat of formation difference $\Delta H(\textit{syn} - \textit{anti})$ by PM3 and PM7 is $+3.80$ and -11.72 kcal mol $^{-1}$, respectively. Based on these arguments, we believe that in the high-speed vibration milling solid state synthesis of **1**[213], the *syn* isomer is actually the main product. The tendency of forming *syn* isomers is a positive one, since while they share similar electronic structure, *syn*-facial fullerene cages can facilitate the charge transfer between the moieties or absorption on surfaces, offering a guide of designing molecular devices.

Besides confirming the usefulness of our inc3-db-B0 approach, we demonstrated that for bulky functional groups like C₆₀ the dispersion interaction can outperform the steric effect, making a compact isomer like *syn*-**1** the more stable one. Note that, many similar stereochemical phenomena have been reported in the literature, but they may have quite different origins. For instance, the case of *cis*-1,3-dimethylcyclobutane mentioned above is still due to a steric mechanism. The “*cis* effect”, like *cis*-1,2-difluoroethylene having lower energy than its *trans* isomer[223], is related to the hyperconjugation effect[224]. The “heterodox bond” in bimetallic complexes can also cause such consequences[225, 226]. Therefore, for a chemical problem one must be cautious not to be confined too much by traditional rules and take all significant effects into account, even the usual “negligible” dispersion force. In some aspects it resembles the gravitational force: it is much weaker than the electrostatic force, but it can make huge contributions for massive particles!

4.6 Singlet-triplet Gap of Biphenylcarbene

Carbene is a well-known star in organic chemistry. Its singlet and triplet states can lead to different reactivity[227]. A recent work by Costa and Sander argues that the stability of the spin states of diphenylcarbene can be controlled by the methanol solvent[228]. In that work the authors applied B3LYP-D3 to compute the energies of singlet and triplet states of diphenylcarbene and diphenylcarbene–methanol complex. We will use the more exact CCSD(T) method to examine these energies.

The results are listed in Table 4.6.1. The singlet-triplet gaps (STG) obtained by the inc3-db-B0 approach agree well with those of the standard implementations, the largest error being 0.62 kcal mol $^{-1}$. The efficiency is also very good: at the VDZ level one can

save about 30% computational cost for the diphenylcarbene–methanol complex, and at the VTZ level, the inc3-db-B0-CCSD(T) calculation is even faster than a standard CCSD one! Thus, the incremental scheme shows excellent accuracy and efficiency for these high-spin open-shell molecules.

Table 4.6.1: Singlet-triplet gap (unit: kcal mol⁻¹) of diphenylcarbene and its methanol complex.

Method	diphenylcarbene ^a	diphenylcarbene–methanol ^a
	$E(S - T)^b$	$E(S - T)^b$
standard CCSD/VDZ	5.11	-0.41
inc3-db-B0-CCSD/VDZ	4.63 (0.48)	-0.09 (-0.32)
standard CCSD/VTZ	3.78	N/A ^c
inc3-db-B0-CCSD/VTZ	3.65 (0.13)	-2.56 (N/A)
standard CCSD(T)/VDZ	3.25	-2.20
inc3-db-B0-CCSD(T)/VDZ	2.84 (0.41)	-1.58 (-0.62)
standard CCSD(T)/VTZ	N/A ^c	N/A ^c
inc3-db-B0-CCSD(T)/VTZ	1.57 (N/A)	-4.75 (N/A)
B3LYP-D3/6-311++G(d,p) ^a	5.06	-0.44
Time ^d	diphenylcarbene	diphenylcarbene–methanol
standard CCSD(T)/VDZ	1.7 days	14.0 days
inc3-db-B0-CCSD(T)/VDZ	2.3 days	8.8 days
standard CCSD/VTZ	51.0 days	N/A
inc3-db-B0-CCSD(T)/VTZ	42.7 days	187.8 days

^a The geometries and energies are taken from Costa’s work[228].

^b The number in parentheses is the error of inc3-db-B0 approach relative to the standard implementation.

^c Beyond our computational ability.

^d The listed time is for the triplet molecule. These calculations were performed with Intel(R) Xeon(R) CPU E5-4620 0 @ 2.20GHz.

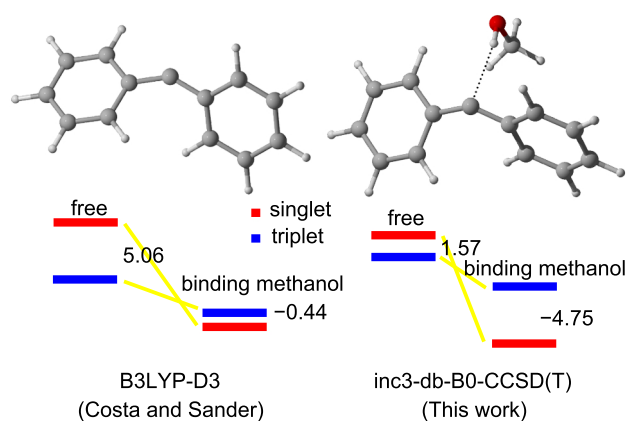


Figure 4.6.1: Left top: Diphenylcarbene. Right top: Diphenylcarbene–methanol complex. Bottom: The effect of the methanol molecule suggested by B3LYP-D3 [228] and inc3-db-B0-CCSD(T).

Based on the reliability of our approach, we can calibrate Costa and Sander’s calculations[228].

CCSD(T) and B3LYP-D3 agree that the ground state of diphenylcarbene changes from a triplet to a singlet state when it complexes with a methanol molecule. However, taking inc3-db-B0-CCSD(T)/VTZ as a reference, the STG of diphenylcarbene and diphenylcarbene–methanol complex at B3LYP-D3 level is over- and underestimated by about 4 kcal mol⁻¹, respectively. The tendency of DFT exhibiting a large error for the STG can also be found for some didehydroazines [229], where B3LYP/VTZ deviates from CCSD(T)/VTZ by 5 to 15 kcal mol⁻¹! Thus, the STG obtained by DFT should be interpreted with cautions. The accurate inc3-db-B0-CCSD(T) results reveal that the methanol molecule can reverse and *enlarge* the STG of diphenylcarbene, rather than reverse and *reduce* the STG as expected by DFT (see Figure 4.6.1). This offers a new guide for rational design of organic reactions involving diphenylcarbene, and possibly other carbenes.

4.7 VDE of the GFP Chromophore

Green fluorescent protein (GFP)[230] is an important tool for biological *in-vivo* studies. The impact of the GFP matrix on its chromophore, i.e. deprotonated 4-hydroxybenzylidene-2,3-dimethylimidazolinone anion (dHBDI⁻), has been discussed theoretically by several authors[231, 232, 233]. It was argued that the protein can increase the vertical detachment energy (VDE) of the dHBDI by more than 2 eV [232], and a possible reason is that the positively charged Arg96 residue of the GFP (PDB ID: 1EMA) [234] stabilizes the anionic form of dHBDI. To examine the effect of this arginine we built a model: the complex of dHBDI and a guanidinium cation (Gmd⁺) (see Figure 4.7.1). The VDEs we try to compute are defined as:

$$\text{VDE}(\text{gas}) = E[\text{dHBDI}] - E[\text{dHBDI}^-] \quad (4.7.1)$$

$$\text{VDE}(\text{protein}) = E[\text{dHBDI} \cdot \text{Gmd}^+] - E[\text{dHBDI}^- \cdot \text{Gmd}^+] \quad (4.7.2)$$

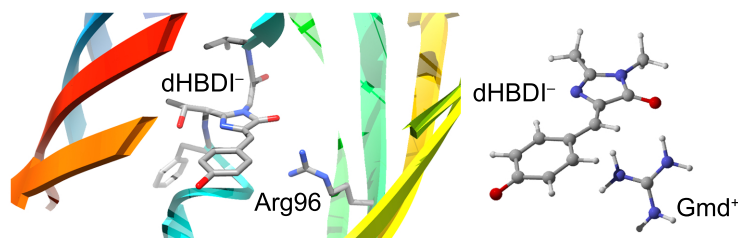


Figure 4.7.1: Left: The chromophore part of the GFP. Right: The HBDI⁻ · Gmd⁺ complex.

The geometries of dHBDI⁻ and dHBDI⁻ · Gmd⁺ were optimized at the B3LYP/6-311++G(d) level and the VDEs calculated by the inc3-db-B0 approach are given in Table 4.7.1.

Obviously, for large basis set like VTZ, the inc3-db-B0 approach saved considerable computational cost with little loss of accuracy. Using inc3-db-B0-CCSD(T)/VTZ as reference, DFT is observed to overestimate the VDE by about 0.2 eV, but the quantitative results are correct. Indeed, the existence of Gmd⁺ stabilizes the dHBDI⁻, yielding the VDE about 2.13 eV larger than that of dHBDI in gas phase, which is very similar to that of dHBDI in GFP. Thus, the positively charged Arg96 residue of GFP could be important in controlling the excitation property of its chromophore.

At this stage it is worth highlighting that inc3-db-B0 exhibits consistent accuracy for systems of different number of electrons.

Table 4.7.1: VDEs of dHBDI (unit: eV) in gas phase and protein environment.

Method	VDE(gas) ^a	VDE(protein) ^a
standard CCSD/VDZ	2.22	5.53
inc3-db-B0-CCSD/VDZ	2.22 (0.00)	5.52 (0.01)
standard CCSD/VTZ	N/A ^b	N/A ^b
inc3-db-B0-CCSD/VTZ	2.67 (N/A)	5.87 (N/A)
standard CCSD(T)/VDZ	2.20	5.51
inc3-db-B0-CCSD(T)/VDZ	2.18 (0.02)	5.48 (0.03)
standard CCSD(T)/VTZ	N/A ^b	N/A ^b
inc3-db-B0-CCSD(T)/VTZ	2.68 (N/A)	5.81 (N/A)
B3LYP/6-311++G(d)	2.88	6.05
Time ^c	VDE(gas)	VDE(protein)
standard CCSD(T)/VDZ	14.3 days	53.5 days
inc3-db-B0-CCSD(T)/VDZ	9.4 days	55.9 days
standard CCSD(T)/VTZ	N/A	N/A
inc3-db-B0-CCSD(T)/VTZ	22.6 days	209.4 days

^a The number in parentheses is the error of inc3-db-B0 approach relative to the standard implementation.

^b Beyond our computational ability.

^c The listed time is for the doublet molecule. These calculations were performed with Intel(R) Xeon(R) CPU E5-4620 0 @ 2.20GHz.

Chapter 5

Theory of Labile Capping Bonds and its Applications

5.1 Theory of Labile Capping Bonds

Low valency, highly charged and large size metal ions, like Ln^{3+} and An^{3+} , are able to form electrostatically bound complexes with rather large CNs. To our best knowledge the complex with the largest CN synthesized up to date is thorium(IV) aminodiborane $\text{Th}(\text{H}_3\text{BNMe}_2\text{BH}_3)_4$ with $\text{CN}=15$ [235], although complexes with $\text{CN}=16$ have been predicted theoretically[236]. The geometries of these complexes are to a large degree determined by the steric and electronic interactions between the ligands. When $\text{CN} \geq 7$, the binding sites cannot always be distributed quasi-equidistantly, instead, some sites will be pushed away from the innermost shell forming a polyhedron, and these sites are the so-called “capping” ones.

The capping bonds in aqua lanthanide(III) complexes are found to exhibit an interesting trend of being *shorter and weaker*[126], this raises two questions: is this a general phenomenon in chemistry, that capping bonds do not obey Badger’s bond length-strength rule[237, 238] and what is the physics behind it?

To explore this problem further[239], we consider two octadentate ligands: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetaldehyde/tetraethanethial (TZDO/TZDS). The complex TZDO/TZDS-aqua- Ln^{3+} has monocapped SAP (MSAP) configuration, the water molecule occupying the capping site (see Figure 5.1.1). We optimized the geometries of 30 TZDX-aqua- Ln^{3+} ($X = \text{O}, \text{S}$) complexes at B3LYP/(cc-pVDZ+ECP(Z-11)MWB[203, 210]) level and computed the electronic structure parameters to examine the capping Ln–O bonds.

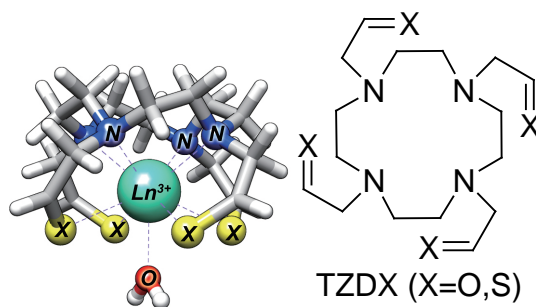


Figure 5.1.1: TZDO/TZDS-aqua- Ln^{3+} .

For the geometrical parameters, we observed that from La^{3+} to Lu^{3+} , as a result

of the enhanced electric field, all the Ln–N and Ln–X bond lengths decrease (see Table S1 in the Supporting Information). The capping Ln–O bond equilibrium lengths are shown in Figure 5.1.2A. It suggests that in TZDO-aqua-Ln³⁺ they also become shorter; however, in TZDS-aqua-Ln³⁺, the capping Ln–O bond first contracts until Eu³⁺, then more or less keeps a constant length, and from Tm³⁺ to Lu³⁺ gets longer. Now we turn to the bond strength. Figure 5.1.2B reveals that the capping Ln–O bond in TZDO is getting stronger until Tm³⁺ and then slightly weakens. Thus for most TZDO-aqua-Ln³⁺ complexes Badger’s rule holds. However, the strength of the capping Ln–O bond in TZDS-aqua-Ln³⁺ keeps decreasing, conflicting with Badger’s rule and usual chemical intuition.

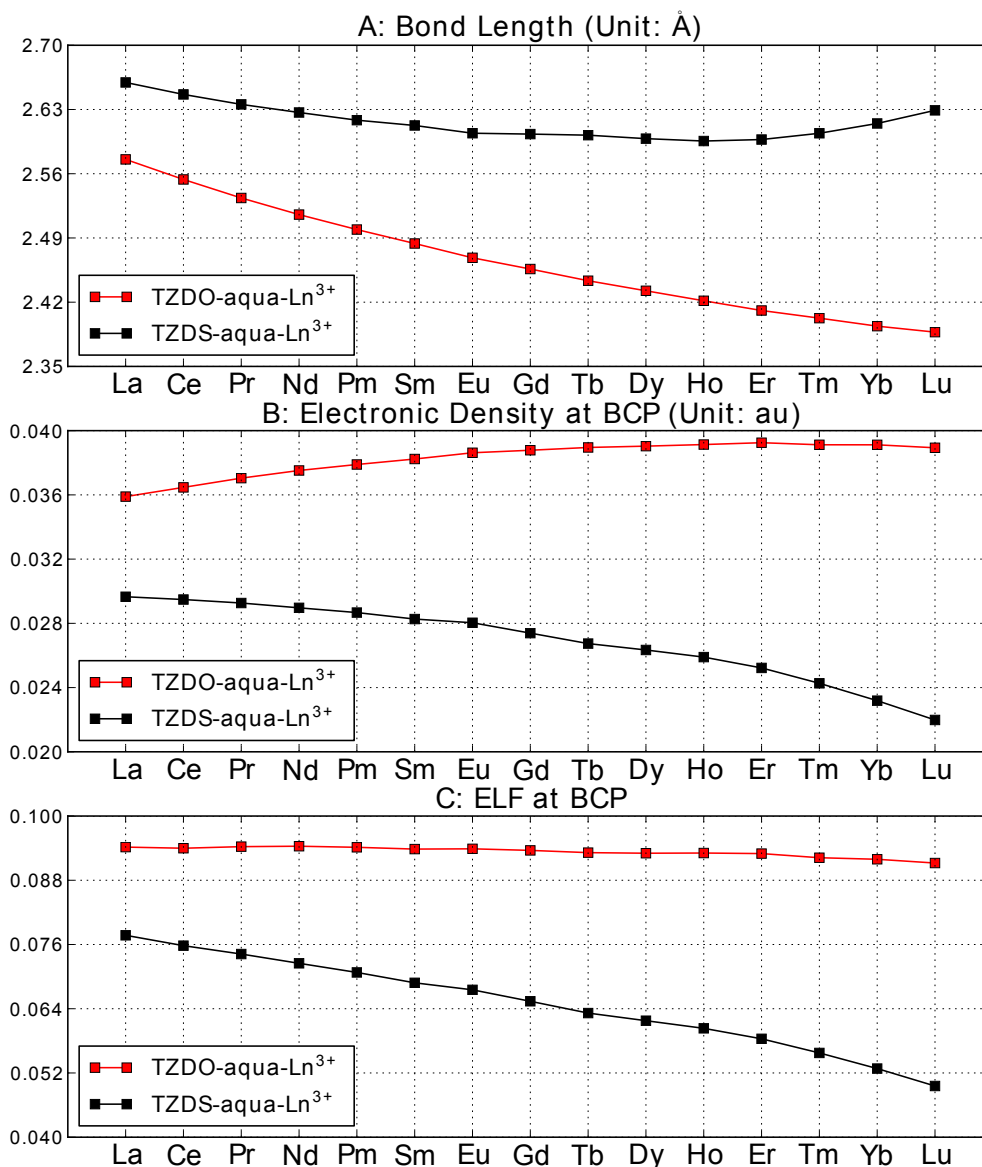


Figure 5.1.2: The lengths, n_{BCP} 's, and ELF at BCP of the Ln–O bonds in TZDO/TZDS-aqua-Ln³⁺.

Thus, the *shorter and weaker capping bond* observed for aqua lanthanide(III) complexes is not an isolate case in lanthanide chemistry. How can a shorter bond be weaker? Such cases have been only reported for covalent bonds in some excited molecules[240,

241]. Since the capping bonds considered are all polar dative ones, this *capping bond phenomenon* could be related to some electrostatic effects.

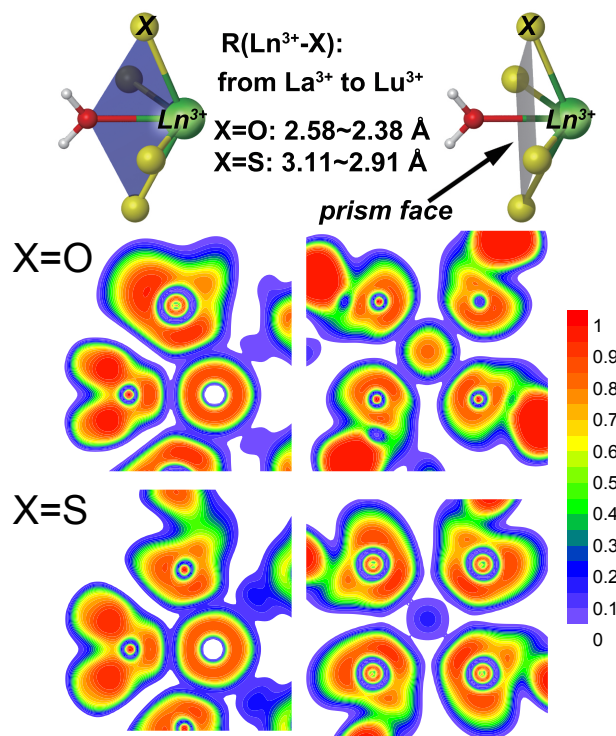
In Figure 5.2.3C and 5.1.2C, we give the values of the ELF at the BCPs of the considered bonds. Electrons are more likely to be localized in the region with large ELF values. The figure reveals that for the shorter and weaker capping bonds, the ELF at BCP has a very *fast* decay rate with nuclear charge while for those “normal” capping bonds, it only slightly changed along the lanthanide series. Thus, in the former case there must be some factors “pushing away” or “depleting” the electrons, and these factors become stronger along the lanthanide series.

To make it clear, we consider TZDX-aqua-Eu³⁺, for which the colour filled maps of the ELF on their two planes are shown in Figure 5.1.3. The lone pairs of the X atoms are all found to point towards the central Eu³⁺ ion as expected. More importantly, the lone pairs of the sulphur atoms are much more diffuse and easier polarizable than those of the oxygen atoms, i.e. sulphur is a softer ligand. The consequence is that Eu³⁺ is “screened” more efficiently by lone pairs on the X ligands in TZDS than in TZDO. This can be seen from the ELF on the “prism face” in Figure 5.1.3 that while in TZDO a large part of the valence region of Eu³⁺ extends outside this face, in TZDS only the tail region of Eu³⁺ exposes. From La³⁺ to Lu³⁺, the prism face formed by 4 X atoms will be closer to the central Ln³⁺, enhancing the screening effect. This is very similar to the case of heavy atoms, where s and p orbitals will contract due to *direct* relativistic effects, thus their screening ability increases, leading as an *indirect* relativistic effect to d and f orbitals being more diffuse[242].

Now the unusual capping bond phenomenon can be interpreted. From La³⁺ to Lu³⁺, the increased electric field pulls the prism as well as capping ligands closer to it. However, this leads to a stronger screening effect of the “prism face”, which makes it more difficult for Ln³⁺ to interact with the capping ligand at the opposite side of the face. Furthermore, the prism ligands can weaken the capping bond by “repelling” (e.g., by the lone pairs of X atoms) the electrons away from the region between the capping ligand and Ln³⁺. Both effects are significant in TZDS-aqua-Ln³⁺. In Ln(H₂O)³⁺₉ the repulsion effect dominates and the screening effect does not become strong enough before Eu³⁺, thus n_{BCP} remains somewhat constant at the beginning of the lanthanide series. For TZDO-aqua-Ln³⁺, since the electrons of the carbonyl oxygen tend to be delocalized over the double bond, it has neither high screening nor repulsion ability. Therefore the strength of the capping bond is still determined by the direct interaction between Ln³⁺ and the capping H₂O, for which Badger’s rule is valid.

We also check whether the capping bond phenomenon occurs in LnCl(H₂O)₈²⁺. To optimize their structures, we put Cl⁻ on the prism and capping sites, respectively, as the starting geometries. Therefore for each LnCl(H₂O)₈²⁺ we got two local minima (denoted by p- and c-, respectively). We give the lengths and n_{BCP} ’s of Ln–Cl bonds as well as ELF at BCPs in Figure 5.1.4. While for all complexes the c-isomer is more stable than the p-isomer by about 0.8 to 1.6 kcal mol⁻¹, interestingly, from La³⁺ to Nd³⁺ the p- and c-LnCl(H₂O)₈²⁺ are optimized to very similar MSAP configurations, and in this range Badger’s rule holds; however, except for Sm³⁺, from Pm³⁺, the geometries of p- and c-LnCl(H₂O)₈²⁺ diverge. In the p-isomers the ligands tend to evenly distribute on a sphere (see Figure 5.1.5), and the Ln–Cl bonds still behave “normally”. The c-isomers are in MSAP configuration (see Figure 5.1.5), and the capping Ln–Cl bonds are shorter and weaker. In fact this can be predicted by their fast decay rates of ELF at the BCPs of Ln–Cl bonds with the nuclear charge. Since Cl⁻ is a large anion and H₂O is a hard ligand, i.e. not efficient in screening, the capping bond phenomenon is not significant before Eu³⁺, similar to the case of Ln(H₂O)₉³⁺.

Therefore, the *shorter and weaker capping bond phenomenon* is essentially an envi-

Figure 5.1.3: Colour filled maps of the ELF on two planes of TZDX-aqua-Eu³⁺.

ronmental effect. For larger ions, like An³⁺, their valence orbitals are well known to be more diffuse than those of their Ln³⁺ analogues. Thus the capping bond phenomenon could be expected only when the prism ligands are sufficiently soft. We performed some calculations and found that (see Table 5.1.1) it occurs only for heavy An³⁺ in TZDS-aqua-An³⁺, implying that ligands with sulphur atoms are very efficient in screening the central ions. Although no capping bond phenomenon was observed for AnCl(H₂O)₈²⁺ and TZDO-aqua-An³⁺, however, the increase rate of n_{BCP} becomes very slow, indicating the existence of the screening and repulsion effect. The optimization of An(H₂O)₉³⁺ leads to an even distribution of water molecules, thus no capping bonds exist at all, or in other words, all bonds are normal ones.

Table 5.1.1: Electronic structure parameters of An³⁺ complexes.

An ^a	AnCl(H ₂ O) ₈ ²⁺	TZDO-aqua-An ₉ ³⁺	TZDS-aqua-An ₉ ³⁺
	2.924 Å	2.671 Å	2.720 Å
Ac	0.0349 au 0.148	0.0323 au 0.087	0.0291 au 0.079
	2.828 Å	2.562 Å	2.629 Å
Np	0.0381 au 0.162	0.0357 au 0.092	0.0307 au 0.080
	2.728 Å	2.425 Å	2.626 Å
Lr	0.0388 au 0.167	0.0383 au 0.093	0.0242 au 0.060

^a The listed data are: bond length, n_{BCP} and ELF at BCP.

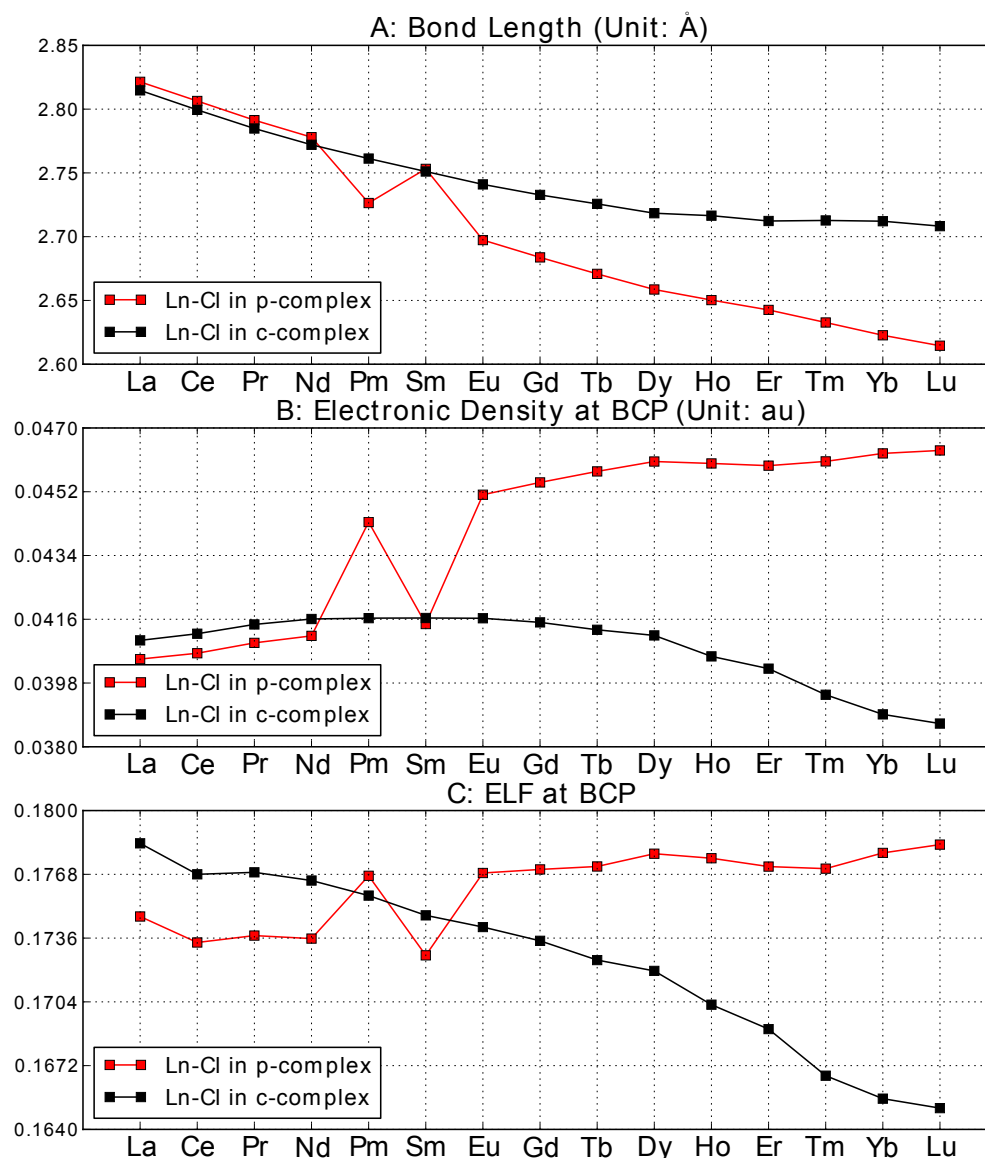


Figure 5.1.4: The lengths, n_{BCP} 's, and ELF at BCP of the Ln–Cl bonds in $\text{LnCl}(\text{H}_2\text{O})_8^{2+}$.

The capping bond phenomenon is expected to occur for large and highly charged metal ions, since otherwise (e.g. K^+) no complexes with high CNs can exist (no capping bonds at all). It also seems unlikely to occur for covalent molecules like ReH_9^{2-} . One should also realize that it is *not* because they are short so they are weak, but because the effects from prism ligands increase. Thus, the fact of the capping Tm–O, Yb–O and Lu–O bonds in TZDS complexes getting longer and weaker (see in Figure 5.1.2) is also due to the capping bond phenomenon rather than the usual intuition of “a longer bond is weaker”.

The capping bond phenomenon is a signature for the bond being hindered by the environment, thus this class of bonds is inherently *labile*, being easily disrupted by external perturbations. This is indeed how we interpreted the water exchange kinetics behaviour and some NMR experiments for hydrated Ln^{3+} [126]. The labile capping bond phenomenon can be applied to more problems. In an experimental study of $\text{Ln}(\text{DODPA})^+$ (DODPA = 6,6'-((1,4,7,10-tetraazacyclododecane-1,7-diyl)bis(methylene))dipicolinic acid)[243], it was found that a capping water molecule can remain bound for Eu^{3+} and Tb^{3+}

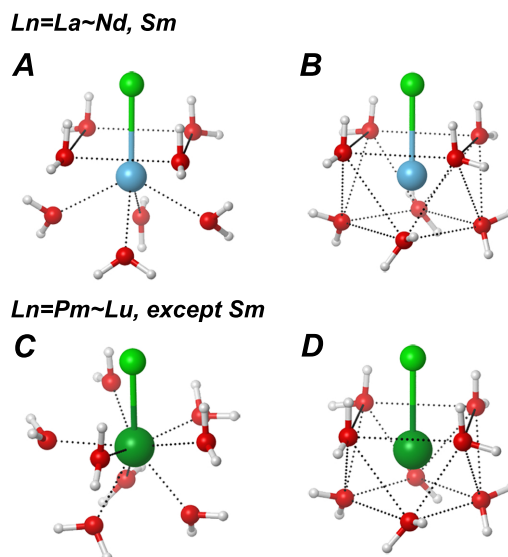


Figure 5.1.5: The geometries of p- and c- $\text{LuCl}(\text{H}_2\text{O})_8^{2+}$.

but not for Yb^{3+} . Another observation is that the water exchange rate of $\text{Ln}(\text{DTPA-BMA})(\text{H}_2\text{O})$ (DTPA-BMA = 1,7-bis[(N-methylcarbamoyl)methyl]-1,4,7-triazaheptane-1,4,7-triacetate) increases along the lanthanide series[244], especially with a steep rise between Gd^{3+} and Tb^{3+} . Both observations are the result of the labile capping Ln-O bond for heavy lanthanides.

That a chemical bond can be affected by its environment has been observed in many cases. Dispersion interaction is able to hold some long and unstable C-C covalent bonds together[215]; the active pocket of human carbonic anhydrase II forces Zn^{2+} to implement an unusual coordination configuration[245]; the electrostatic interactions can sometimes even change a chemical bond from a covalent one to a charge-shift one[246]! Here we demonstrate that in high CN environments, the bond between the capping ligand and the central metal ion can be weakened by the prism ligands through *screening* and *repulsion* effects. If the prism ligands are very hard, or the central metal ions are large, then the screening or repulsion effects on the capping bonds are very small. In this case, it is a normal capping bond and the usual idea of “shorter and stronger bonds”, i.e. Badger’s rule, still holds. If the ligands are sufficiently soft and therefore efficient in screening or repulsion, both effects become stronger as the electric field of the central ions enhances, and the electrons are becoming more difficult to stay in the bonding region. In this case the bond is observed to be “shorter and weaker”, being inherently *labile*. This is the so-called *labile capping bond phenomenon* which we propose in this work. We have seen that in the TTP and MSAP configuration of lanthanides(III) it widely occurs. We believe that it can also exist in higher CN or other suitable chemical environments, providing new insights in understanding chemical problems.

In the following sections we will give two applications to confirm the validity and usefulness of our theory.

5.2 Hydration Kinetics of Trivalent Lanthanide Ions

Besides the hydration free energy problem described in Subsection 4.4, there is another attractive problem involving Ln^{3+} : its kinetics. Lanthanides(III) exhibit very strange hydration kinetics: the exchange rate of the water molecules between the first hydration

sphere and the bulk solvent increases from La^{3+} to Gd^{3+} and decreases thereafter to Lu^{3+} , i.e. it reaches a maximum in the middle region of the lanthanide series[247, 248] (see Figure 5.2.1). The physical reasons for these fascinating phenomena are not well understood. We will see that this is related to the labile capping bond phenomenon.

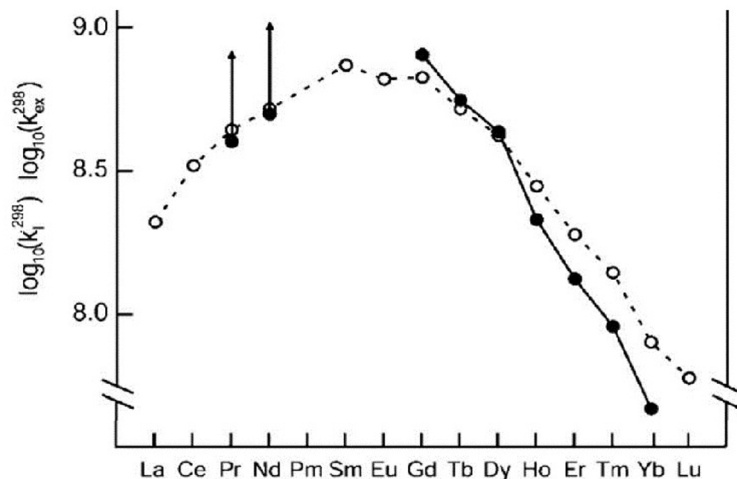


Figure 5.2.1: Water exchange rate of lanthanides(III). The figure is adapted from Helm’s work[247].

To interpret this[126], we first note that the octa- and nona-aqua lanthanide(III) complex have SAP and TTP configuration, respectively (see Figure 4.4.1). For the former there is only one kind of Ln–O bonds, denoted as Ln–O(8); for the latter there are two kinds of Ln–O bonds: six prism (oxygen atoms on the vertices of the trigonal prism, called “Ln–O(9P)” hereafter) and three capping ones (“Ln–O(9C)”).

The AIM theory was applied to analyze the Ln–O bonds. In Figure 5.2.2 we provide the molecular graphs. There exists exactly one BCP for each Ln–O bond. The electron density at the BCP, n_{BCP} , is used to measure the bond strength here.

Since a larger n_{BCP} indicates a stronger bond, a trend that n_{BCP} is larger for a shorter Ln–O bond, is expected. This holds for both the Ln–O(8) and Ln–O(9P) bonds. However, Figure 5.2.3B reveals that the n_{BCP} of the Ln–O(9C) bonds exhibits a non-monotonic behavior: it fluctuates between 0.0365 and 0.0367 a.u. from La^{3+} to Sm^{3+} and then, more astonishingly, decreases from Sm^{3+} to Lu^{3+} , as the corresponding Ln–O(9C) bonds are getting shorter!

These n_{BCP} ’s tell us that for the same Ln^{3+} , the Ln–O(8) bond is stronger than the Ln–O(9P) bond, which is stronger than Ln–O(9C); from La^{3+} to Lu^{3+} , the Ln–O(8) and Ln–O(9P) bonds become stronger in a parallel manner, while the Ln–O(9C) bonds remain somewhat constant in strength before Sm^{3+} and then get weaker. It is just after Sm^{3+} that the octa-aqua lanthanide(III) complexes begin to be more stable. By realizing these facts we can understand the preference of CNs. There is a competition between the formation of eight Ln–O(8) and nine Ln–O(9) bonds in the water exchange process. For light lanthanides, the Ln–O(9C) bond is strong enough that a nona-aqua complex is able to be bound; as going along the lanthanide series, the Ln–O(9C) bond is getting labile, being easier to be disrupted by the environment. Thus the heavy lanthanide(III) complexes will switch to the more stable octa-aqua form.

The particularity of the capping bonds has been noticed before in the literature. In a study of $[\text{Ln}(\text{H}_2\text{O})_9](\text{CF}_3\text{SO}_3)_3$ salts with crystallography and 2D solid state NMR a reduced occupancy of the three capping positions was observed for some heavier lan-

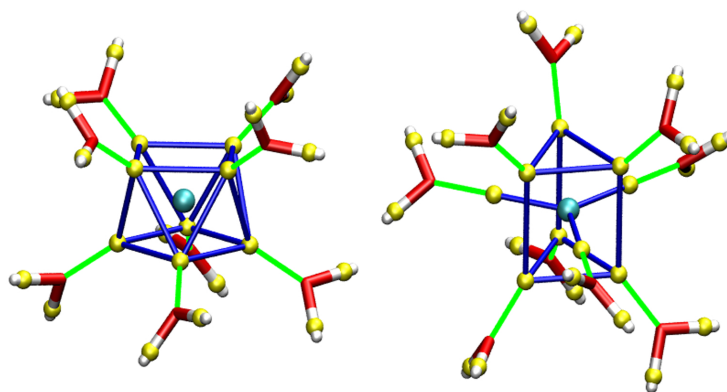


Figure 5.2.2: Molecular graphs of lanthanide(III) aqua complexes. Red-white licorice: water molecule; cyan ball: lanthanide(III) ion; yellow ball: bond critical point; green line: bond path; blue line: no physical meaning, just to guide the eye to recognize the coordination polyhedron. There are no bond paths (green lines) connecting lanthanide(III) ions with oxygen atoms since the implementation of PPs for lanthanides depletes the electron density around the nuclei.

thanide(III) ions[249, 250]: occupancies of 2.91, 2.96, 2.8, 2.7 and 2.4 water molecules were reported for Ho, Er, Tm, Yb and Lu, respectively. Following older ideas[249, 250] it was argued in a recent EXAFS spectroscopy and crystallography study of the same systems that the lanthanide row should be partitioned into four tetrads intersecting at Nd/Pm, Gd, and Ho/Er[200]. In the first tetrad the capping bonds are relatively strong, whereas they get weaker in the second tetrad. In the third and fourth tetrad, due to the smaller ionic radius of the central ion and the increased repulsion between the water ligand, an asymmetry occurs and one capping bond gets stronger again, whereas the other two continue to become weaker. However, a corresponding difference in bond lengths was said to be too small to be detected by the EXAFS technique. The hydrated lanthanide(III) ions in aqueous solution were said to be in this respect very similar to the salts.

Although we agree with the authors of these experimental studies that no sudden structural change occurs at Gd, e.g. a gadolinium break due to reaching a half-filling of the 4f shell, and that the capping bonds play a special role, we do not have any evidence of the postulated asymmetry of the capping bonds for the heavier lanthanide(III) hydrates. Exploratory DFT calculations without imposing any symmetry restrictions on Lu^{3+} hydrates with one as well as two coordination spheres ($\text{Lu}^{3+}(\text{H}_2\text{O})_9$, $\text{Lu}^{3+}(\text{H}_2\text{O})_9(\text{H}_2\text{O})_{18}$) did not give bond length differences larger than 0.01 Å for the capping bonds.

The present work provides a consistent and quantitative picture regarding the strengths of all Ln–O bonds in the equilibrium geometries, revealing a characteristic trend of the *labile capping Ln–O bonds*. This can explain the kinetics of the hydrated water molecules. Duvail et al proposed that the exchange involves a bicapped trigonal prism (BTP) structure of octa-aqua lanthanides[248]. This model can be supported and extended by our work, as illustrated by Figure 5.2.4. A water molecule at capping position is easier to exchange since the Ln–O(9C) bond is much weaker than the Ln–O(9P) or Ln–O(8) bonds. Thus an octa-aqua complex which has a SAP structure will first rearrange to a BTP structure so that the water to exchange is on a capping position; for a nona-aqua ion, a prism water molecule will first rearrange to a capping position

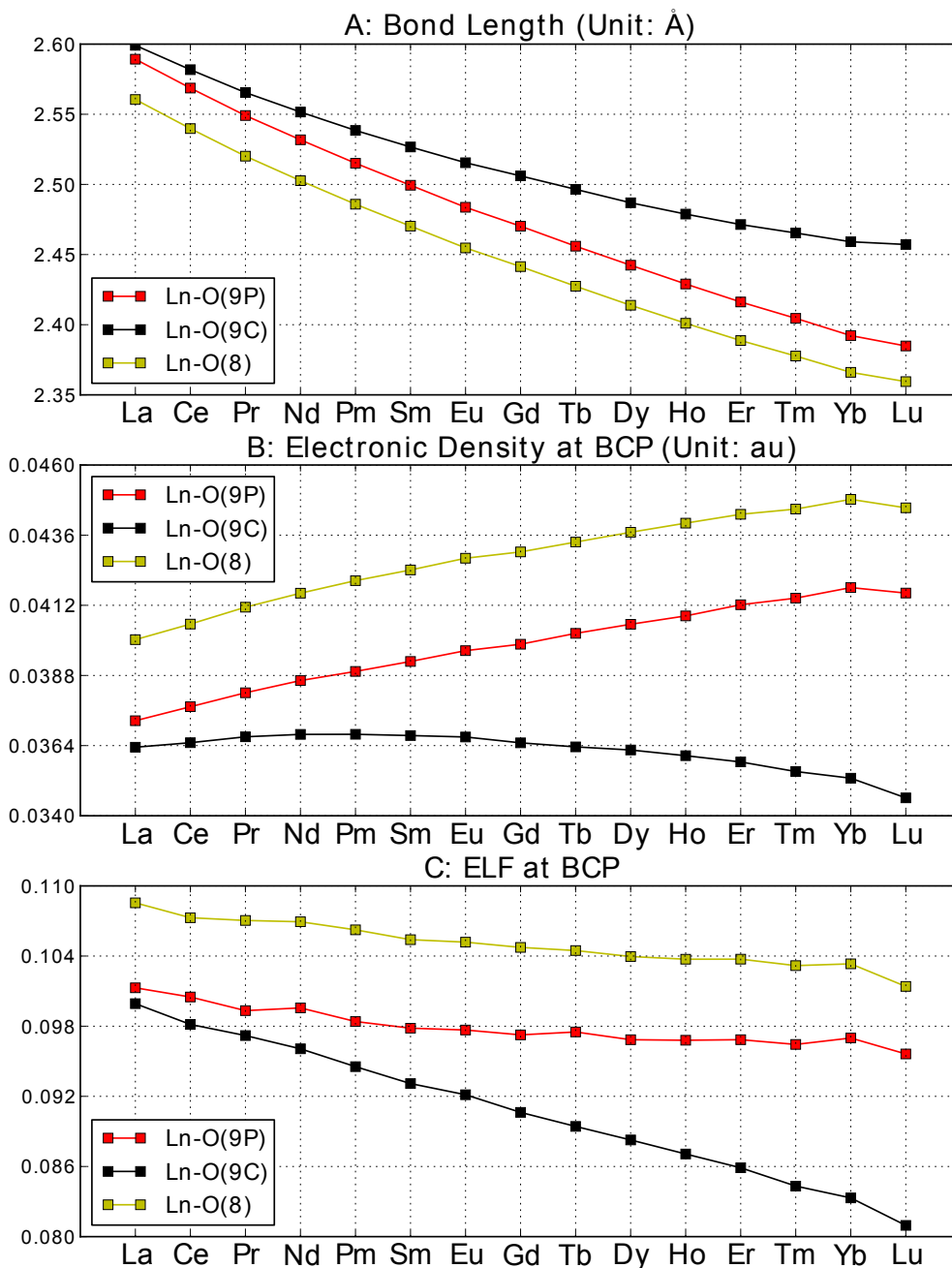


Figure 5.2.3: The lengths, n_{BCP} 's, and ELF at BCP of the Ln–O bonds in aqua lanthanide(III) complexes.

then it can exchange through a BTP intermediate. The possibility of a rearrangement between capping and prism water molecules has been confirmed by NMR experiments on crystals of $[\text{Ln}(\text{H}_2\text{O})_9](\text{CF}_3\text{SO}_3)_3$ [250]. Furthermore, these NMR studies reported that this rearrangement is already fast at 268K for Lu^{3+} , but becomes rapid only at about 300K for La^{3+} , which can be explained by the strength of their Ln–O(9C) bonds.

Now the unusual trend of water exchange rates can be understood. Since the Ln–O(9C) bond is relatively either very strong or very weak for light and heavy lanthanides, respectively, once a nona- or octa-aqua ion forms, it will be very reluctant to dissociate a water molecule or rearrange to form a BTP intermediate. For the intermediate lanthanides (from samarium to holmium), the Ln–O(9C) bond is of moderate strength

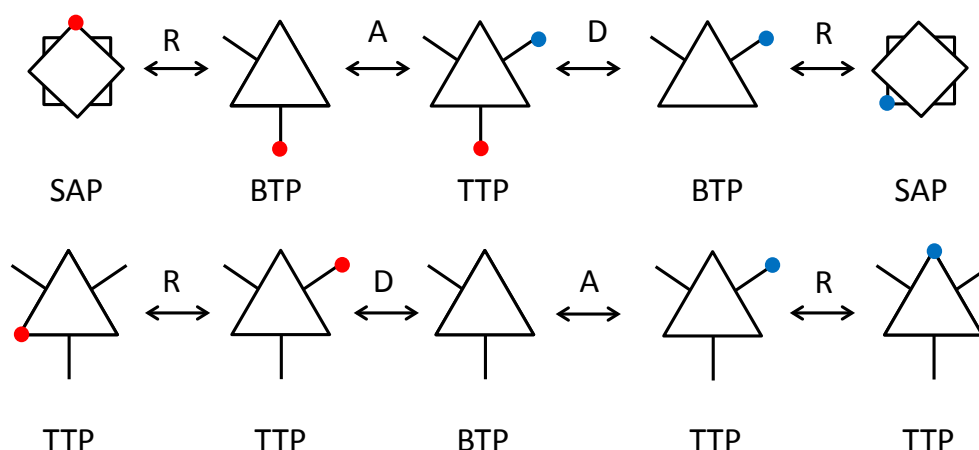


Figure 5.2.4: The water exchange mechanism. Other possible steps of association of a water molecule or rearrangement of a BTP intermediate for nona-aqua lanthanides(III) are omitted for clarity. Circle: water molecule; A: association; D: dissociation; R: rearrangement.

so it can readily both form and break, leading to a faster water exchange. Our model is also compatible with the I_a and I_d mechanism assumed by Helm and Merbach[247]. We also note that the nona- and octa-aqua lanthanide(III) complexes can transform to each other through the BTP intermediate[248], therefore for intermediate lanthanides, both nona- and octa-aqua ions can be long-lived, showing a fractional CN. Therefore, it is the strength of the capping Ln–O bond that determines the preferred CN, exchange rate, and perhaps other hydration behavior.

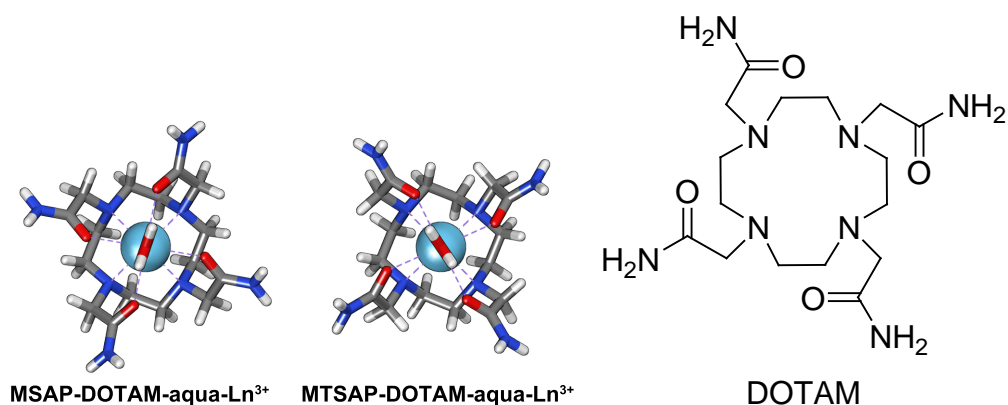
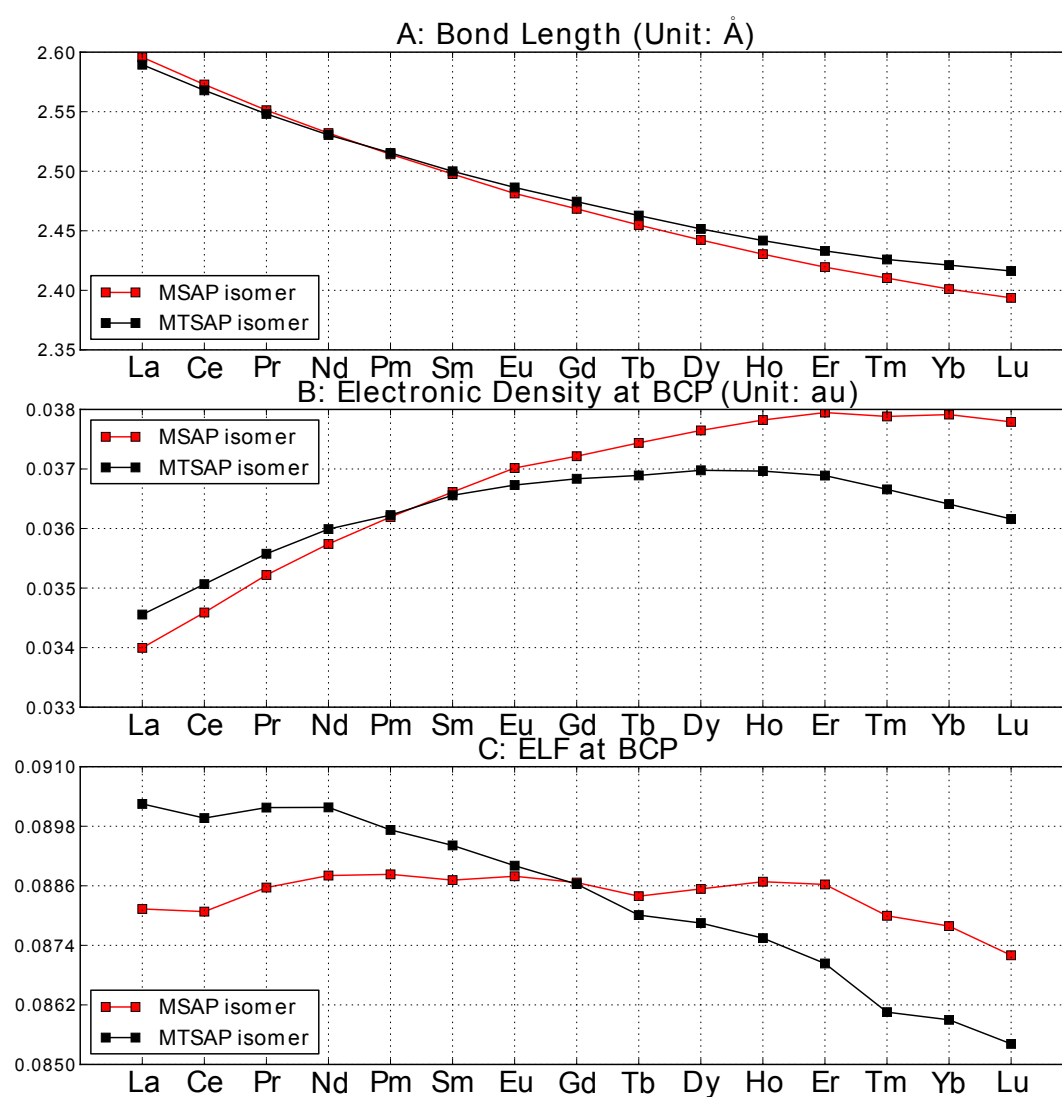
Thus, we see that our theory of labile capping bonds can be supported by various experiments, and successfully interprets this seemingly difficult problem!

5.3 Hydration Kinetics of Lanthanide(III)-DOTAM Complexes

The octadentate ligand 1,4,7,10-tetrakis(carbamoylmethyl)-1,4,7,10-tetraazacyclododecane (DOTAM, see Figure 5.3.1) can form a DOTAM-aqua- Ln^{3+} complex, which exists as M-SAP and monocapped twisted SAP (MTSAP) isomer (also known as M- and m-isomer, respectively) (see Figure 5.3.1). The MTSAP isomer gives a water exchange rate more than 50 times faster than the MSAP isomer does[251]! Is this observation related to the labile capping bond phenomenon? We optimized the geometries of MSAP- and MTSAP-DOTAM-aqua- Ln^{3+} complexes and performed the same analysis for the capping Ln–O bonds as we did for TZDX-aqua- Ln^{3+} . The results are shown in Figure 5.3.2.

The observation from Figure 5.3.2 completely agrees with our theory of labile capping bonds. The MTSAP isomers exhibit a faster decay rate of ELF at BCP of the capping Ln–O bond than the MSAP isomers do, indicating the capping bonds being hindered. Here, the oxygen atom in the prism ligand “ $\text{H}_2\text{N}-\text{C}=\text{O}$ ” of DOTAM is softer than that in “ $\text{H}-\text{C}=\text{O}$ ” of TZDO due to the electron donating nature of the NH_2 group. Therefore its screening and repulsion effect become stronger and the “shorter and weaker” Ln–O bonds appear starting from Dy^{3+} . One can note that the hindrance become also sensible in the MSAP isomers for Er^{3+} to Lu^{3+} , but the effect is much weaker than that of the MTSAP isomers.

Figure 5.3.3A suggests that the more stable isomer is MTSAP and MSAP one before

Figure 5.3.1: DOTAM-aqua-Ln³⁺.Figure 5.3.2: The lengths, n_{BCP} 's, and ELF at BCP of the Ln-O bonds in DOTAM-aqua-Ln³⁺ complexes.

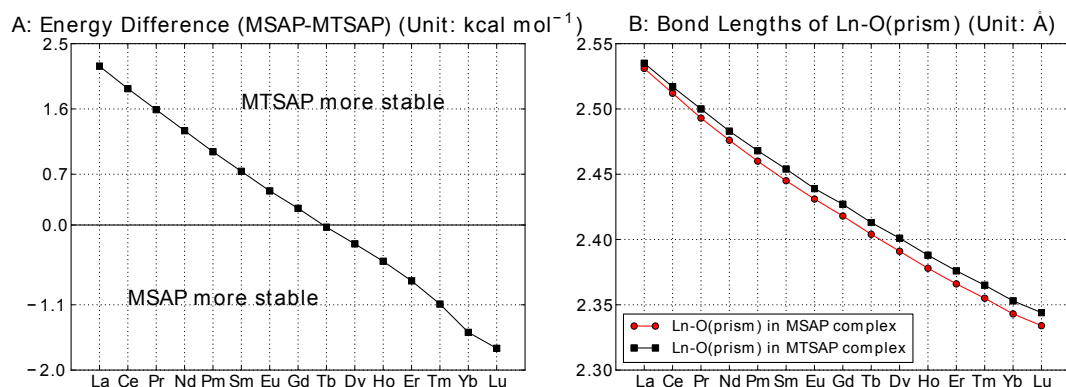


Figure 5.3.3: The energy difference between the MSAP and MTSAP isomer of DOTAM-aqua-Ln³⁺ (A) and bond lengths of Ln-O(prism) in DOTAM-aqua-Ln³⁺ (B).

and after Tb³⁺, respectively, which is compatible with some experiments[252]. This is because the coordination cavity of the MTSAP isomer is larger, which matches better with the lighter and larger Ln³⁺ ions (see Figure 5.3.3B: the distance between Ln³⁺ and the prism oxygen is always larger for the MTSAP isomer). Therefore, DOTAM in TSAP geometry has a better screening ability than in SAP geometry (see Figure 5.3.2C), especially for heavier and smaller Ln³⁺. This is why the labile capping bond phenomenon occurs in the MTSAP but not in the MSAP isomers.

For two isomers of the same Ln³⁺ ion, the bond length-strength relation satisfies the Badger's rule, but the strength difference becomes larger for the heavier lanthanides, resulting from mainly the hindrance of prism ligands in the MTSAP isomer. Both factors (labile and long capping bonds) lead to a much weaker capping Ln-O bond in MTSAP isomers than in MSAP isomers. This not only successfully interprets why the water exchange of the MTSAP isomers is much faster than that of the MSAP ones[251], but also reminds us that the labile capping bond phenomenon can be ligand conformation-dependent[239]. Of course it also confirms the validity of our theory.

Chapter 6

Summary and Outlook

6.1 Summary

In this work, the author has introduced a new quantum chemical approach: the third-order incremental dual-basis set zero buffer approach, abbreviated as inc3-db-B0. This approach can combine with CCSD, CCSD(T) and their F12 variants to compute the accurate energy of large molecules in an efficient and accurate way. The author have also constructed a program APTS to enable anyone implement inc3-db-B0 (and in fact, general incremental scheme) in a black-box mode. It can be easily and efficiently parallelized.

Many examples demonstrate the power of the approach in various kinds of real chemical problems. These includes: benchmark set validation; energies of isomers of water clusters; the rotational barrier of biphenyl; hydration of lanthanide trivalent ions; the relative stability of isomers of double fullerene adducts; singlet-triplet gap of biphenylcarbene, and vertical detachment energy of green fluorescent protein chromophore. These problems range from inorganic to organic chemistry, closed-shell to open-shell and small molecule to biological molecules. The approach exhibits high efficiency, especially for large and open-shell molecules or large basis set. The db-B0 approximation can reduce the computational cost more than 70% and by parallelization, the wall time can be reduced by more than 10 times! For the accuracy, the error of the approach is usually less than 1 kcal mol⁻¹, and for clusters it is even more accurate, making it possible to distinguish the energetically nearly degenerate isomers. The inc3-db-B0 approach may not be so “crazy” fast like some other methods, however, it is the most accurate one! In fact, is was pointed out in Subsection 4.2.1 that DLPNO cannot even give a qualitatively correct result regarding the relative energies of the isomers of water 17-mer, while inc3-db-B0 approach does not have any problem with it. Therefore, the accuracy is the most important advantage of the approach presented here. In fact inc3-db-B0-CCSD(T)-F12a/AVTZ can serve as benchmark reference! In my opinion, if a linear-scaling correlation method has to sacrifice accuracy to achieve efficiency, one can simply use DFT which is much faster and more robust.

6.2 Outlook

Since the application of quantum chemistry has gone beyond atoms, small- and medium-size molecules, to more complex systems, such as clusters, nanoparticles, biological molecules as well as surfaces and solids, the demand for quantum chemistry becomes more and more stringent, requiring higher and higher accuracy and efficiency for different cases. Thus there are still a lot to do with the inc3-db-B0 approach, especially the following three aspects:

- Molecular properties. An efficient realization of this requires application of incremental scheme on analytical derivatives like:

$$\frac{\partial \Delta \epsilon_{ij}}{\partial P} = \frac{\partial \epsilon_{ij}}{\partial P} - \frac{\partial \epsilon_i}{\partial P} - \frac{\partial \epsilon_j}{\partial P} \quad (6.2.1)$$

Some explorative studies[128] showed that the convergence of derivative is much more difficult than that of energy, but if this can be solved, one can greatly generalize the application field of the incremental scheme, say optical, electric or magnetic properties.

- Excited states. The EOM-CC[253] is an accurate method for systems involving electronical excitations, electron detachment or electron attachment. The difficulty in combination of incremental scheme and EOM-CC is that it is extremely difficult to *identity a specific state* in calculations of different increments. One must deeply hack into EOM-CC to solve this problem. Maybe a state-specific EOM-CC is preferred.
- MR methods. Although a robust MRCC is not available, some promising developments have appeared[254, 255, 256]. Also MRCI is an important method for construction of accurate PES. Thus, an “incremental MRCC/MRCI” is desirable for photochemistry of large molecules. The difficulty lies also in the identification of a specific state from different increments.

Finally, we also expect that the incremental scheme can be applied to more kinds of real chemical problems, like the binding energy of large metal-ligand complexes, stability of cluster isomers, etc. This will mark a great success of the incremental scheme.

Bibliography

- [1] M. Born, R. Oppenheimer, Zur Quantentheorie der Molekeln, *Annalen der Physik* **1927**, 389(20):457–484.
- [2] M. Dolg, X. Cao, Relativistic Pseudopotentials: Their Development and Scope of Applications, *Chem. Rev.* **2012**, 112(1):403–480.
- [3] W. Schwarz, Das Kombinierte Näherungsverfahren, *Theor. Chim Acta* **1968**, 11(4):307–324.
- [4] L. R. Kahn, P. Baybutt, D. G. Truhlar, Ab Initio Effective Core Potentials: Reduction of All-electron Molecular Structure Calculations to Calculations Involving Only Valence Electrons, *J. Chem. Phys.* **1976**, 65(10):3826–3853.
- [5] J. Katriel, E. R. Davidson, Asymptotic Behavior of Atomic and Molecular Wave Functions, *Proc. Natl. Acad. Sci. USA* **1980**, 77(8):4403–4406.
- [6] E. Schrödinger, An Undulatory Theory of the Mechanics of Atoms and Molecules, *Phys. Rev.* **1926**, 28:1049–1070.
- [7] S. T. Epstein, *The Variation Method in Quantum Chemistry*, Academic Press **1974**.
- [8] R. J. Bartlett, G. D. Purvis, Many-body Perturbation Theory, Coupled-pair Many-electron Theory, and the Importance of Quadruple Excitations for the Correlation Problem, *Int. J. Quantum Chem.* **1978**, 14(5):561–581.
- [9] J. A. Pople, J. S. Binkley, R. Seeger, Theoretical Models Incorporating Electron Correlation, *Int. J. Quantum Chem.* **1976**, 10(S10):1–19.
- [10] D. R. Hartree, The Wave Mechanics of an Atom with a Non-coulomb Central Field. Part I. Theory and Methods, *Math. Proc. Cambridge Philos. Soc.* **1928**, 24:89–110.
- [11] D. R. Hartree, The Wave Mechanics of an Atom with a Non-coulomb Central Field. Part II. Some Results and Discussion, *Math. Proc. Cambridge Philos. Soc.* **1928**, 24:111–132.
- [12] V. Fock, Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Z. Phys.* **1930**, 61(1-2):126–148.
- [13] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications **1996**.
- [14] C. C. J. Roothaan, New Developments in Molecular Orbital Theory, *Rev. Mod. Phys.* **1951**, 23:69–89.

- [15] G. G. Hall, The Molecular Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials, *Proc. Roy. Soc. (London)* **1951**, 205(1083):541–552.
- [16] J. M. Foster, S. F. Boys, Canonical Configurational Interaction Procedure, *Rev. Mod. Phys.* **1960**, 32:300–302.
- [17] C. Edmiston, K. Ruedenberg, Localized Atomic and Molecular Orbitals, *Rev. Mod. Phys.* **1963**, 35:457–464.
- [18] J. Pipek, P. G. Mezey, A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions, *J. Chem. Phys.* **1989**, 90(9):4916–4926.
- [19] J. E. Subotnik, A. D. Dutoi, M. Head-Gordon, Fast Localized Orthonormal Virtual Orbitals which Depend Smoothly on Nuclear Coordinates, *J. Chem. Phys.* **2005**, 123(11):114108.
- [20] I.-M. Høyvik, B. Jansik, P. Jørgensen, Trust Region Minimization of Orbital Localization Functions, *J. Chem. Theory Comput.* **2012**, 8(9):3137–3146.
- [21] T. Kato, On the Eigenfunctions of Many-particle Systems in Quantum Mechanics, *Commun. Pure Appl. Math.* **1957**, 10(2):151–177.
- [22] Z. Li, S. Shao, W. Liu, Relativistic Explicit Correlation: Coalescence Conditions and Practical Suggestions, *J. Chem. Phys.* **2012**, 136(14):144117.
- [23] C. Hättig, W. Klopper, A. Köhn, D. P. Tew, Explicitly Correlated Electrons in Molecules, *Chem. Rev.* **2012**, 112(1):4–74.
- [24] P.-O. Löwdin, Quantum Theory of Many-Particle Systems. III. Extension of the Hartree-Fock Scheme to Include Degenerate Systems and Correlation Effects, *Phys. Rev.* **1955**, 97:1509–1520.
- [25] E. R. Davidson, The Iterative Calculation of a few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-symmetric Matrices, *J. Comput. Phys.* **1975**, 17(1):87 – 94.
- [26] B. Klahn, W. A. Bingel, Convergence of Rayleigh–Ritz Method in Quantum Chemistry. 1. Criteria of Convergence, *Theo. Chim. Acta* **1977**, 44(1):9–26.
- [27] B. Klahn, W. A. Bingel, Convergence of Rayleigh–Ritz Method in Quantum Chemistry. 2. Investigation of Convergence for Special Systems of Slater, Gauss and 2-electron Functions, *Theo. Chim. Acta* **1977**, 44(1):27–43.
- [28] I. Shavitt, R. J. Bartlett, *Many-body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge University Press **2009**.
- [29] J. Goldstone, Derivation of the Brueckner Many-Body Theory, *Proc. Roy. Soc. (London)* **1957**, 239(1217):267–279.
- [30] J. 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.* **1966**, 45(11):4256–4266.
- [31] G. D. Purvis, R. J. Bartlett, A Full Coupled-cluster Singles and Doubles Model: The Inclusion of Disconnected Triples, *J. Chem. Phys.* **1982**, 76(4):1910–1918.

- [32] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, A Fifth-order Perturbation Comparison of Electron Correlation Theories, *Chem. Phys. Lett.* **1989**, *157*(6):479 – 483.
- [33] J. Řezáč, P. Hobza, Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the “Gold Standard,” CCSD(T) at the Complete Basis Set Limit?, *J. Chem. Theory Comput.* **2013**, *9*(5):2151–2155.
- [34] L. Kong, F. A. Bischoff, E. F. Valeev, Explicitly Correlated R12/F12 Methods for Electronic Structure, *Chem. Rev.* **2012**, *112*(1):75–107.
- [35] T. B. Adler, G. Knizia, H.-J. Werner, A Simple Efficient CCSD(T)-F12 Approximation, *J. Chem. Phys.* **2007**, *127*(22):221106.
- [36] G. Knizia, T. B. Adler, H.-J. Werner, Simplified CCSD(T)-F12 Methods: Theory and Benchmarks, *J. Chem. Phys.* **2009**, *130*(5):054104.
- [37] H.-J. Werner, T. B. Adler, F. R. Manby, General Orbital Invariant MP2-F12 Theory, *J. Chem. Phys.* **2007**, *126*(16):164102.
- [38] S. Ten-no, Initiation of Explicitly Correlated Slater-type Geminal Theory, *Chem. Phys. Lett.* **2004**, *398*(13):56–61.
- [39] E. F. Valeev, Improving on the Resolution of the Identity in Linear R12 Ab Initio Theories, *Chem. Phys. Lett.* **2004**, *395*(46):190 – 195.
- [40] P. Hohenberg, W. Kohn, Inhomogeneous Electron Gas, *Phys. Rev.* **1964**, *136*:B864–B871.
- [41] W. Kohn, L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev.* **1965**, *140*:A1133–A1138.
- [42] J. C. Slater, *The Self-consistent Field for Molecular and Solids, Quantum Theory of Molecular and Solids, Vol. 4*, AMcGraw-Hill **1974**.
- [43] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* **1996**, *77*:3865–3868.
- [44] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, Climbing the Density Functional Ladder: Nonempirical Meta-generalized Gradient Approximation Designed for Molecules and Solids, *Phys. Rev. Lett.* **2003**, *91*:146401.
- [45] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, *J. Phys. Chem.* **1994**, *98*(45):11623–11627.
- [46] S. Grimme, Semiempirical Hybrid Density Functional with Perturbative Second-order Correlation, *J. Chem. Phys.* **2006**, *124*(3):034108.
- [47] A. J. Cohen, P. Mori-Sánchez, W. Yang, Challenges for Density Functional Theory, *Chem. Rev.* **2012**, *112*(1):289–320.
- [48] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, *J. Chem. Phys.* **2010**, *132*(15):154104.
- [49] A. D. Becke, E. R. Johnson, A Density-functional Model of the Dispersion Interaction, *J. Chem. Phys.* **2005**, *123*(15):154101.

- [50] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer-Verlag **1999**.
- [51] C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, Using Redundant Internal Coordinates to Optimize Equilibrium Geometries and Transition States, *J. Comput. Chem.* **1996**, *17*(1):49–56.
- [52] A. D. Becke, K. E. Edgecombe, A Simple Measure of Electron Localization in Atomic and Molecular Systems, *J. Chem. Phys.* **1990**, *92*(9):5397–5403.
- [53] A. D. Becke, Local exchange-correlation approximations and first-row molecular dissociation energies, *Int. J. Quantum Chem.* **1985**, *27*(5):585–594.
- [54] R. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press **1990**.
- [55] M. Domagała, S. J. Grabowski, CH \cdots N and CH \cdots S Hydrogen Bonds—Influence of Hybridization on Their Strength, *J. Phys. Chem. A* **2005**, *109*(25):5683–5688.
- [56] S. J. Grabowski, W. A. Sokalski, J. Leszczynski, How Short Can the H \cdots H Intermolecular Contact Be? New Findings that Reveal the Covalent Nature of Extremely Strong Interactions, *J. Phys. Chem. A* **2005**, *109*(19):4331–4341.
- [57] D. Hugas, S. Simon, M. Duran, Electron Density Topological Properties Are Useful To Assess the Difference between Hydrogen and Dihydrogen Complexes, *J. Phys. Chem. A* **2007**, *111*(20):4506–4512.
- [58] T. Lu, F. Chen, Bond Order Analysis Based on the Laplacian of Electron Density in Fuzzy Overlap Space, *J. Phys. Chem. A* **2013**, *117*(14):3100–3108.
- [59] A. D. Becke, A Multicenter Numerical Integration Scheme for Polyatomic Molecules, *J. Chem. Phys.* **1988**, *88*(4):2547–2553.
- [60] A. Zupan, K. Burke, M. Ernzerhof, J. P. Perdew, Distributions and Averages of Electron Density Parameters: Explaining the Effects of Gradient Corrections, *J. Chem. Phys.* **1997**, *106*(24):10184–10193.
- [61] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang, Revealing Noncovalent Interactions, *J. Am. Chem. Soc.* **2010**, *132*(18):6498–6506.
- [62] W. Kohn, Density-functional Theory for Systems of Very Many Atoms, *Int. J. Quantum Chem.* **1995**, *56*(4):229–232.
- [63] N. Flocke, R. J. Bartlett, A Natural Linear Scaling Coupled-cluster Method, *J. Chem. Phys.* **2004**, *121*(22):10935–10944.
- [64] T. F. Hughes, N. Flocke, R. J. Bartlett, Natural Linear-Scaled Coupled-Cluster Theory with Local Transferable Triple Excitations: Applications to Peptides, *J. Phys. Chem. A* **2008**, *112*(26):5994–6003.
- [65] S. Li, J. Ma, Y. Jiang, Linear Scaling Local Correlation Approach for Solving the Coupled Cluster Equations of Large Systems, *J. Comput. Chem.* **2002**, *23*(2):237–244.
- [66] S. Li, J. Shen, W. Li, Y. Jiang, An Efficient Implementation of the “Cluster-in-molecule” Approach for Local Electron Correlation Calculations, *J. Chem. Phys.* **2006**, *125*(7):074109.

- [67] W. Li, P. Piecuch, J. R. Gour, S. Li, Local Correlation Calculations Using Standard and Renormalized Coupled-cluster Approaches, *J. Chem. Phys.* **2009**, *131*(11):114109.
- [68] W. Li, Y. Guo, S. Li, A Refined Cluster-in-molecule Local Correlation Approach for Predicting the Relative Energies of Large Systems, *Phys. Chem. Chem. Phys.* **2012**, *14*:7854–7862.
- [69] M. Ziolkowski, B. Jansik, P. Jørgensen, J. Olsen, Maximum Locality in Occupied and Virtual Orbital Spaces Using a Least-change Strategy, *J. Chem. Phys.* **2009**, *131*:124112.
- [70] M. Ziolkowski, B. Jansik, T. Kjæbørgard, P. Jørgensen, Linear Scaling Coupled Cluster Method with Correlation Energy Based Error Control, *J. Chem. Phys.* **2010**, *133*:014107.
- [71] K. Kristensen, M. Ziolkowski, B. Jansik, T. Kjæbørgard, P. Jørgensen, A Locality Analysis of the Divide-Expand-Consolidate Coupled Cluster Amplitude Equations, *J. Chem. Theor. Comput.* **2011**, *7*:1677–1694.
- [72] J. E. Subotnik, A. Sodt, M. Head-Gordon, A Near Linear-scaling Smooth Local Coupled Cluster Algorithm for Electronic Structure, *J. Chem. Phys.* **2006**, *125*(7):074116.
- [73] A. Sodt, J. E. Subotnik, M. Head-Gordon, Linear Scaling Density Fitting, *J. Chem. Phys.* **2006**, *125*(19):194109.
- [74] J. E. Subotnik, A. Sodt, M. Head-Gordon, The Limits of Local Correlation Theory: Electronic Delocalization and Chemically Smooth Potential Energy Surfaces, *J. Chem. Phys.* **2008**, *128*(3):034103.
- [75] J. E. Subotnik, M. Head-Gordon, Exploring the Accuracy of Relative Molecular Energies with Local Correlation Theory, *J. Phys.: Condens. Matter* **2008**, *20*(29):294211.
- [76] Y. Mochizuki, K. Yamashita, T. Nakano, Y. Okiyama, K. Fukuzawa, N. Taguchi, S. Tanaka, Higher-order Correlated Calculations Based on Fragment Molecular Orbital Scheme, *Theor. Chem. Acc.* **2011**, *130*:515–530.
- [77] A. Saha, K. Raghavachari, Dimers of Dimers (DOD): A New Fragment-Based Method Applied to Large Water Clusters, *J. Chem. Theory Comput.* **2014**, *10*(1):58–67.
- [78] K. Wang, W. Li, S. Li, Generalized Energy-Based Fragmentation CCSD(T)-F12a Method and Application to the Relative Energies of Water Clusters (H₂O)₂₀, *J. Chem. Theory Comput.* **2014**, *10*(4):1546–1553.
- [79] S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, K. Fukuzawa, Electron-correlated Fragment-molecular-orbital Calculations for Biomolecular and Nano Systems, *Phys. Chem. Chem. Phys.* **2014**, *16*:10310–10344.
- [80] P. Pulay, Localizability of Dynamic Electron Correlation, *Chem. Phys. Lett.* **1983**, *100*(2):151–154.
- [81] S. Sæbø, P. Pulay, Local Configuration Interaction: An Efficient Approach for Larger Molecules, *Chem. Phys. Lett.* **1985**, *113*(1):13–18.

- [82] P. Pulay, S. Sæbø, Orbital-invariant Formulation and Second-order Gradient Evaluation in Møller–Plesset Perturbation Theory, *Theor. Chim. Acta.* **1986**, *69*:357–368.
- [83] S. Sæbø, P. Pulay, Fourth-order Møller–Plessett Perturbation Theory in the Local Correlation Treatment. I. Method, *J. Chem. Phys.* **1987**, *86*(2):914–922.
- [84] S. Sæbø, P. Pulay, The Local Correlation Treatment. II. Implementation and Tests, *J. Chem. Phys.* **1988**, *88*(3):1884–1890.
- [85] C. Hampel, H.-J. Werner, Local Treatment of Electron Correlation in Coupled Cluster Theory, *J. Chem. Phys.* **1996**, *104*(16):6286–6297.
- [86] G. Hetzer, P. Pulay, H.-J. Werner, Multipole Approximation of Distant Pair Energies in Local MP2 Calculations, *Chem. Phys. Lett.* **1998**, *290*(13):143–149.
- [87] M. Schütz, G. Hetzer, H.-J. Werner, Low-order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2, *J. Chem. Phys.* **1999**, *111*(13):5691–5705.
- [88] G. Hetzer, M. Schütz, H. Stoll, H.-J. Werner, Low-order Scaling Local Correlation Methods II: Splitting the Coulomb Operator in Linear Scaling Local Second-order Møller–Plesset Perturbation Theory, *J. Chem. Phys.* **2000**, *113*(21):9443–9455.
- [89] M. Schütz, Low-order Scaling Local Electron Correlation Methods. III. Linear Scaling Local Perturbative Triples Correction (T), *J. Chem. Phys.* **2000**, *113*(22):9986–10001.
- [90] M. Schütz, H.-J. Werner, Low-order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-cluster (LCCSD), *J. Chem. Phys.* **2001**, *114*(2):661–681.
- [91] M. Schütz, A New, Fast, Semi-direct Implementation of Linear Scaling Local Coupled Cluster Theory, *Phys. Chem. Chem. Phys.* **2002**, *4*:3941–3947.
- [92] F. Neese, F. Wennmohs, A. Hansen, Efficient and Accurate Local Approximations to Coupled-electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method, *J. Chem. Phys.* **2009**, *130*(11):114108.
- [93] F. Neese, A. Hansen, D. G. Liakos, Efficient and Accurate Approximations to the Local Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis, *J. Chem. Phys.* **2009**, *131*(6):064103.
- [94] A. Hansen, D. G. Liakos, F. Neese, Efficient and Accurate Local Single Reference Correlation Methods for High-spin Open-shell Molecules Using Pair Natural Orbitals, *J. Chem. Phys.* **2011**, *135*(21):214102.
- [95] D. G. Liakos, A. Hansen, F. Neese, Weak Molecular Interactions Studied with Parallel Implementations of the Local Pair Natural Orbital Coupled Pair and Coupled Cluster Methods, *J. Chem. Theory Comp.* **2011**, *7*(1):76–87.
- [96] L. M. J. Huntington, A. Hansen, F. Neese, M. Nooijen, Accurate Thermochemistry from a Parameterized Coupled-cluster Singles and Doubles Model and a Local Pair Natural Orbital Based Implementation for Applications to Larger Systems, *J. Chem. Phys.* **2012**, *136*(6):064101.
- [97] D. G. Liakos, F. Neese, Improved Correlation Energy Extrapolation Schemes Based on Local Pair Natural Orbital Methods, *J. Phys. Chem. A* **2012**, *116*(19):4801–4816.

- [98] C. Riplinger, F. Neese, An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method, *J. Chem. Phys.* **2013**, *138*(3):034106.
- [99] J. Yang, Y. Kurashige, F. R. Manby, G. K. L. Chan, Tensor Factorizations of Local Second-order Møller–Plesset Theory, *J. Chem. Phys.* **2011**, *134*(4):044123.
- [100] J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz, H.-J. Werner, The Orbital-specific-virtual Local Coupled Cluster Singles and Doubles Method, *J. Chem. Phys.* **2012**, *136*(14):144105.
- [101] P. J. LeStrange, B. Peng, F. Ding, G. W. Trucks, M. J. Frisch, X. Li, Density of States Guided Møller–Plesset Perturbation Theory, *J. Chem. Theory Comput.* **2014**, *10*(5):1910–1914.
- [102] R. K. Nesbet, Atomic Bethe–Goldstone Equations. I. The Be Atom, *Phys. Rev.* **1967**, *155*:51–55.
- [103] R. K. Nesbet, Atomic Bethe–Goldstone Equations. II. The Ne Atom, *Phys. Rev.* **1967**, *155*:56–58.
- [104] R. K. Nesbet, Atomic Bethe–Goldstone Equations. III. Correlation Energies of Ground States of Be, B, C, N, O, F, and Ne, *Phys. Rev.* **1968**, *175*:2–9.
- [105] H. Stoll, The Correlation Energy of Crystalline Silicon, *Chem. Phys. Lett.* **1992**, *191*(6):548–552.
- [106] H. Stoll, On the Correlation Energy of Graphite, *J. Chem. Phys.* **1992**, *97*(11):8449–8454.
- [107] H. Stoll, Correlation Energy of Diamond, *Phys. Rev. B* **1992**, *46*:6700–6704.
- [108] B. Paulus, P. Fulde, H. Stoll, Electron Correlations for Ground-state Properties of Group-IV Semiconductors, *Phys. Rev. B* **1995**, *51*:10572–10578.
- [109] K. Doll, M. Dolg, P. Fulde, H. Stoll, Correlation Effects in Ionic Crystals: The Cohesive Energy of MgO, *Phys. Rev. B* **1995**, *52*:4842–4848.
- [110] B. Paulus, P. Fulde, H. Stoll, Cohesive Energies of Cubic III-V Semiconductors, *Phys. Rev. B* **1996**, *54*:2556–2560.
- [111] H. Stoll, B. Paulus, P. Fulde, On the Accuracy of Correlation-energy Expansions in Terms of Local Increments, *J. Chem. Phys.* **2005**, *123*(14):144108.
- [112] B. Paulus, The Method of Increments—a Wavefunction-based Ab Initio Correlation Method for Solids, *Phys. Rep.* **2006**, *428*(1):1–52.
- [113] M. Yu, S. Kalvoda, M. Dolg, An Incremental Approach for Correlation Contributions to the Structural and Cohesive Properties of Polymers. Coupled-cluster Study of *trans*-polyacetylene, *Chemical Physics* **1997**, *224*(23):121 – 131.
- [114] A. Abdurahman, A. Shukla, M. Dolg, Correlated Ground-state *Ab Initio* Calculations of Polymethineimine, *Chem. Phys.* **2000**, *257*(23):301 – 310.
- [115] J. Friedrich, M. Hanrath, M. Dolg, Fully Automated Implementation of the Incremental Scheme: Application to CCSD Energies for Hydrocarbons and Transition Metal Compounds, *J. Chem. Phys.* **2007**, *126*(15):154110.

- [116] J. Friedrich, M. Dolg, Fully Automated Incremental Evaluation of MP2 and CCSD(T) Energies: Application to Water Clusters, *J. Chem. Theory Comput.* **2009**, *5*(2):287–294.
- [117] J. Friedrich, D. P. Tew, W. Klopper, M. Dolg, Automated Incremental Scheme for Explicitly Correlated Methods, *J. Chem. Phys.* **2010**, *132*(16):164114.
- [118] J. Friedrich, K. Walczak, Incremental CCSD(T)(F12)—MP2-F12—A Method to Obtain Highly Accurate CCSD(T) Energies for Large Molecules, *J. Chem. Theory Comput.* **2013**, *9*(1):408–417.
- [119] J. Friedrich, M. Hanrath, M. Dolg, Energy Screening for the Incremental Scheme: Application to Intermolecular Interactions, *J. Phys. Chem. A* **2007**, *111*(39):9830–9837.
- [120] J. Friedrich, M. Dolg, Implementation and Performance of a Domain-specific Basis Set Incremental Approach for Correlation Energies: Applications to Hydrocarbons and a Glycine Oligomer, *J. Chem. Phys.* **2008**, *129*(24):244105.
- [121] J. Friedrich, M. Hanrath, M. Dolg, Using Symmetry in the Framework of the Incremental Scheme: Molecular Applications, *Chem. Phys.* **2008**, *346*(13):266–274.
- [122] J. Friedrich, K. Walczak, M. Dolg, Evaluation of Core and Core-valence Correlation Contributions Using the Incremental Scheme, *Chem. Phys.* **2009**, *356*(13):47–53.
- [123] J. Friedrich, Incremental Scheme for Intermolecular Interactions: Benchmarking the Accuracy and the Efficiency, *J. Chem. Theory Comput.* **2012**, *8*(5):1597–1607.
- [124] J. Friedrich, J. Hänchen, Incremental CCSD(T)(F12*)—MP2: A Black Box Method To Obtain Highly Accurate Reaction Energies, *J. Chem. Theory Comput.* **2013**, *9*(12):5381–5394.
- [125] T. Anacker, J. Friedrich, New Accurate Benchmark Energies for Large Water Clusters: DFT is Better Than Expected, *J. Comput. Chem.* **2014**, *35*(8):634–643.
- [126] J. Zhang, N. Heinz, M. Dolg, Understanding Lanthanoid(III) Hydration Structure and Kinetics by Insights from Energies and Wave Functions, *Inorg. Chem.* **2014**, *53*(14):7700–7708.
- [127] J. Friedrich, S. Coriani, T. Helgaker, M. Dolg, Implementation of the Incremental Scheme for One-electron First-order Properties in Coupled-cluster Theory, *J. Chem. Phys.* **2009**, *131*(15):154102.
- [128] J. Yang, M. Dolg, Evaluation of Electronic Correlation Contributions for Optical Tensors of Large Systems Using the Incremental Scheme, *J. Chem. Phys.* **2007**, *127*(8):084108.
- [129] J. Friedrich, E. Perlt, M. Roatsch, C. Spickermann, B. Kirchner, Coupled Cluster in Condensed Phase. Part I: Static Quantum Chemical Calculations of Hydrogen Fluoride Clusters, *J. Chem. Theory Comput.* **2011**, *7*(4):843–851.
- [130] C. Spickermann, E. Perlt, M. von Domaros, M. Roatsch, J. Friedrich, B. Kirchner, Coupled Cluster in Condensed Phase. Part II: Liquid Hydrogen Fluoride from Quantum Cluster Equilibrium Theory, *J. Chem. Theory Comput.* **2011**, *7*(4):868–875.

- [131] J. Zhang, M. Dolg, Third-Order Incremental Dual-Basis Set Zero-Buffer Approach: An Accurate and Efficient Way To Obtain CCSD and CCSD(T) Energies, *J. Chem. Theory Comput.* **2013**, *9*(7):2992–3003.
- [132] J. Zhang, M. Dolg, Approaching the Complete Basis Set Limit of CCSD(T) for Large Systems by the Third-order Incremental Dual-basis Set Zero-buffer F12 Method, *J. Chem. Phys.* **2014**, *140*(4):044114.
- [133] J. Zhang, M. Dolg, Incremental Scheme for Large High-spin Open-shell Systems, *J. Chem. Theory Comput.* **2014**, *11*(3):962–968.
- [134] N. Heinz, J. Zhang, M. Dolg, Actinoid(III) Hydration – First Principle Gibbs Energies of Hydration Using High Level Correlation Methods, *J. Chem. Theory Comput.* **2014**, *10*(12):5593–5598.
- [135] J. Zhang, M. Dolg, Dispersion Interaction Stabilizes Sterically Hindered Double Fullerenes, *Chem. Eur. J.* **2014**, *20*:13909–13912.
- [136] M. J. Elrod, R. J. Saykally, Many-Body Effects in Intermolecular Forces, *Chem. Rev.* **1994**, *94*(7):1975–1997.
- [137] J. Malek, S. Flach, K. Kladko, Incremental Expansions for the Ground-state Energy of the Two-dimensional Hubbard Model, *Phys. Rev. B* **1999**, *59*:R5273–R5276.
- [138] H. Matsuda, Physical Nature of Higher-order Mutual Information: Intrinsic Correlations and Frustration, *Phys. Rev. E* **2000**, *62*:3096–3102.
- [139] A. Chervanyov, Incremental Expansion of the Classical Partition Functions, *Phys. Lett. A* **2001**, *286*(1):30–33.
- [140] D. G. Fedorov, K. Kitaura, The Importance of Three-body Terms in the Fragment Molecular Orbital Method, *J. Chem. Phys.* **2004**, *120*(15):6832–6840.
- [141] K. Kladko, P. Fulde, On the Properties of Cumulant Expansions, *Int. J. Quantum Chem.* **1998**, *66*(5):377–389.
- [142] P. Fulde, *Electron Correlations in Molecules and Solids*, Berlin: Springer-Verlag, 3 edition **1995**.
- [143] L. D. Faddeev, Scattering Theory for a Three-particle System, *Sov. Phys.-JETP* **1961**, *12*:1014.
- [144] L. D. Faddeev, S. P. Merkuriev, *Quantum Scattering Theory for Several Particle Systems*, Dordrecht: Kluwer Academic Pub **1993**.
- [145] G. Karypis, V. Kumar, A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs, *SIAM J. Sci. Comput.* **1998**, *20*(1):359–392.
- [146] E. Forgy, Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications, *Biometrics* **1965**, *21*:768–780.
- [147] J. MacQueen, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I, Statistics*, USA: University of California Press **1965**.
- [148] G. M. Downs, J. M. Barnard, Clustering Methods and Their Uses in Computational Chemistry, *Rev. Comput. Chem.* **2002**, *18*:1–40.

- [149] R. Jurgens-Lutovsky, J. Almlöf, Dual Basis Sets in Calculations of Electron Correlation, *Chem. Phys. Lett.* **1991**, *178*(5-6):451–454.
- [150] S. Havriliak, H. F. King, Rydberg Radicals. 1. Frozen-core Model for Rydberg Levels of the Ammonium Radical, *J. Am. Chem. Soc.* **1983**, *105*(1):4–12.
- [151] S. Havriliak, T. R. Furlani, H. F. King, Rydberg Levels of the Sodium Atom and the Ammonium Radical Calculated by Perturbation Theory, *Can. J. Phys.* **1984**, *62*(12):1336–1346.
- [152] K. Wolinski, P. Pulay, Second-order Møller–Plesset calculations with dual basis sets, *J. Chem. Phys.* **2003**, *118*(21):9497–9503.
- [153] J. Deng, P. M. W. Gill, Communication: A New Approach to Dual-basis Second-order Møller–Plesset Calculations, *J. Chem. Phys.* **2011**, *134*(8):081103.
- [154] R. P. Steele, M. Head-Gordon, J. C. Tully, Ab Initio Molecular Dynamics with Dual Basis Set Methods, *J. Phys. Chem. A* **2010**, *114*(43):11853–11860.
- [155] P.-O. Löwdin, On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals, *J. Chem. Phys.* **1950**, *18*(3):365–375.
- [156] B. C. Carlson, J. M. Keller, Orthogonalization Procedures and the Localization of Wannier Functions, *Phys. Rev.* **1957**, *105*:102–103.
- [157] W. Yang, Direct Calculation of Electron Density in Density-functional Theory, *Phys. Rev. Lett.* **1991**, *66*:1438–1441.
- [158] W. Yang, Direct Calculation of Electron Density in Density-functional Theory: Implementation for Benzene and a Tetrapeptide, *Phys. Rev. A* **1991**, *44*:7823–7826.
- [159] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford, CT **2004**.
- [160] F. Neese, The ORCA Program System, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*(1):73–78.
- [161] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar,

- G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, MOLPRO, version 2012.1, a package of ab initio programs **2012**, see "<http://www.molpro.net>".
- [162] T. Lu, F. Chen, Multiwfn: A Multifunctional Wavefunction Analyzer, *J. Comput. Chem.* **2012**, *33*(5):580–592.
- [163] C. Y. Legault, CYLview, 1.0b **2009**, see <http://www.cylview.org>.
- [164] W. Humphrey, A. Dalke, K. Schulten, VMD – Visual Molecular Dynamics, *J. Mol. Graphics.* **1996**, *14*:33–38.
- [165] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera – A visualization System for Exploratory Research and Analysis, *J. Comput. Chem.* **2004**, *25*(13):1605–1612.
- [166] S. Kozuch, J. M. L. Martin, Halogen Bonds: Benchmarks and Theoretical Analysis, *J. Chem. Theory Comput.* **2013**, *9*(4):1918–1931.
- [167] J. Řezáč, K. E. Riley, P. Hobza, S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures, *J. Chem. Theory Comput.* **2011**, *7*(8):2427–2438.
- [168] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, P. Hobza, Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes, *J. Chem. Theory Comput.* **2013**, *9*(8):3364–3374.
- [169] J. Řezáč, K. E. Riley, P. Hobza, Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules, *J. Chem. Theory Comput.* **2012**, *8*(11):4285–4292.
- [170] K. Liu, J. D. Cruzan, R. J. Saykally, Water Clusters, *Science* **1996**, *271*(5251):929–933.
- [171] U. Buck, F. Huisken, Infrared Spectroscopy of Size-Selected Water and Methanol Clusters, *Chem. Rev.* **2000**, *100*(11):3863–3890.
- [172] S. S. Xantheas, C. J. Burnham, R. J. Harrison, Development of Transferable Interaction Models for Water. II. Accurate energetics of the First Few Water Clusters from First Principles, *J. Chem. Phys.* **2002**, *116*(4):1493–1499.
- [173] E. E. Dahlke, R. M. Olson, H. R. Leverentz, D. G. Truhlar, Assessment of the Accuracy of Density Functionals for Prediction of Relative Energies and Geometries of Low-Lying Isomers of Water Hexamers, *J. Phys. Chem. A* **2008**, *112*(17):3976–3984.
- [174] D. M. Bates, G. S. Tschumper, CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures, *J. Phys. Chem. A* **2009**, *113*(15):3555–3559, pMID: 19354314.
- [175] C. Pérez, M. T. Muckle, D. P. Zaleski, N. A. Seifert, B. Temelso, G. C. Shields, Z. Kisiel, B. H. Pate, Structures of Cage, Prism, and Book Isomers of Water Hexamer from Broadband Rotational Spectroscopy, *Science* **2012**, *336*(6083):897–901.

- [176] R. J. Saykally, D. J. Wales, Pinning Down the Water Hexamer, *Science* **2012**, *336*(6083):814–815.
- [177] Y. Wang, V. Babin, J. M. Bowman, F. Paesani, The Water Hexamer: Cage, Prism, or Both. Full Dimensional Quantum Simulations Say Both, *J. Am. Chem. Soc.* **2012**, *134*(27):11116–11119.
- [178] E. Miliordos, E. Aprà, S. S. Xantheas, Optimal Geometries and Harmonic Vibrational Frequencies of the Global Minima of Water Clusters $(\text{H}_2\text{O})_n$, $n = 2-6$, and Several Hexamer Local Minima at the CCSD(T) Level of Theory, *J. Chem. Phys.* **2013**, *139*(11):114302.
- [179] H. W. Qi, H. R. Leverentz, D. G. Truhlar, Water 16-mers and Hexamers: Assessment of the Three-Body and Electrostatically Embedded Many-Body Approximations of the Correlation Energy or the Nonlocal Energy As Ways to Include Cooperative Effects, *J. Phys. Chem. A* **2013**, *117*(21):4486–4499.
- [180] Y. Wang, J. M. Bowman, IR Spectra of the Water Hexamer: Theory, with Inclusion of the Monomer Bend Overtone, and Experiment Are in Agreement, *J. Phys. Chem. Lett.* **2013**, *4*(7):1104–1108.
- [181] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin, W. A. Goddard, Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters, *J. Chem. Theory Comput.* **2009**, *5*(4):1016–1026.
- [182] S. Bulusu, S. Yoo, E. Aprà, S. Xantheas, X. C. Zeng, Lowest-Energy Structures of Water Clusters $(\text{H}_2\text{O})_{11}$ and $(\text{H}_2\text{O})_{13}$, *J. Phys. Chem. A* **2006**, *110*(42):11781–11784.
- [183] S. Yoo, E. Aprà, X. C. Zeng, S. S. Xantheas, High-Level *Ab Initio* Electronic Structure Calculations of Water Clusters $(\text{H}_2\text{O})_{16}$ and $(\text{H}_2\text{O})_{17}$: A New Global Minimum for $(\text{H}_2\text{O})_{16}$, *J. Phys. Chem. Lett.* **2010**, *1*(20):3122–3127.
- [184] G. Häfeli, C. Regelmann, Refined *Ab Initio* 6-31G Splitvalence Basis Set Optimization of the Molecular Structures of Biphenyl in Twisted, Planar, and Perpendicular Conformations, *J. Comput. Chem.* **1987**, *8*(7):1057–1065.
- [185] S. Tsuzuki, K. Tanabe, *Ab Initio* Molecular Orbital Calculations of the Internal Rotational Potential of Biphenyl Using Polarized Basis Sets with Electron Correlation Correction, *J. Phys. Chem.* **1991**, *95*(1):139–144.
- [186] M. Rubio, M. Merchán, E. Ortí, The Internal Rotational Barrier of Biphenyl Studied with Multiconfigurational Second-order Perturbation Theory (CASPT2), *Theor. Chim. Acta* **1995**, *91*(1-2):17–29.
- [187] A. Karpfen, C. H. Choi, M. Kertesz, Single-Bond Torsional Potentials in Conjugated Systems: A Comparison of *Ab Initio* and Density Functional Results, *J. Phys. Chem. A* **1997**, *101*(40):7426–7433.
- [188] S. Tsuzuki, T. Uchimaru, K. Matsumura, M. Mikami, K. Tanabe, Torsional Potential of Biphenyl: *Ab Initio* Calculations with the Dunning Correlation Consistent Basis Sets, *J. Chem. Phys.* **1999**, *110*:2858.
- [189] S. Arulmozhiraja, T. Fujii, Torsional Barrier, Ionization Potential, and Electron Affinity of Biphenyl—A Theoretical Study, *J. Chem. Phys.* **2001**, *115*:10589.

- [190] F. Grein, Twist Angles and Rotational Energy Barriers of Biphenyl and Substituted Biphenyls, *J. Phys. Chem. A* **2002**, *106*:3823.
- [191] F. Grein, New Theoretical Studies on the Dihedral Angle and Energy Barriers of Biphenyl, *J. Mol. Struct. (Theochem)* **2003**, *624*:23.
- [192] F. Grein, Influence of Diffuse and Polarization Functions on the Second-order Møller-Plesset Optimized Dihedral Angle of Biphenyl, *Theor. Chem. Acc.* **2003**, *109*:274.
- [193] J. C. Sancho-García, J. Cornil, Anchoring the Torsional Potential of Biphenyl at the *Ab Initio* Level: The Role of Basis Set versus Correlation Effects, *J. Chem. Theory Comput.* **2005**, *1*:581.
- [194] M. P. Johansson, J. Olsen, Torsional Barriers and Equilibrium Angle of Biphenyl: Reconciling Theory with Experiment, *J. Chem. Theory Comput.* **2008**, *4*(9):1460–1471.
- [195] A. Almenningen, O. Bastiansen, L. Fernholt, B. N. Cyvin, S. J. Cyvin, S. Samdal, Structure and Barrier of Internal Rotation of Biphenyl Derivatives in the Gaseous State: Part 1. The Molecular Structure and Normal coordinate Analysis of Normal Biphenyl and Perdeuterated Biphenyl, *J. Mol. Struct.* **1985**, *128*(13):59–76.
- [196] O. Bastiansen, S. Samdal, Structure and Barrier of Internal Rotation of Biphenyl Derivatives in the Gaseous State: Part 4. Barrier of Internal Rotation in Biphenyl, Perdeuterated Biphenyl and Seven Non-*ortho*-substituted Halogen Derivatives, *J. Mol. Struct.* **1985**, *128*(13):115 – 125.
- [197] J. Chen, Y. Zhu, R. Jiao, The Separation of Am from Lanthanides by Purified Cyanex 301 Extraction, *Sep. Sci. Technol.* **1996**, *31*(19):2723–2731.
- [198] H. H. Dam, D. N. Reinhoudt, W. Verboom, Multicoordinate Ligands for Actinide/Lanthanide Separations, *Chem. Soc. Rev.* **2007**, *36*:367–377.
- [199] P. D’Angelo, R. Spezia, Hydration of Lanthanoids(III) and Actinoids(III): An Experimental/Theoretical Saga, *Chem. Eur. J.* **2012**, *18*(36):11162–11178.
- [200] I. Persson, P. D’Angelo, S. De Panfilis, M. Sandström, L. Eriksson, Hydration of Lanthanoid(III) Ions in Aqueous Solution and Crystalline Hydrates Studied by EXAFS Spectroscopy and Crystallography: The Myth of the “Gadolinium Break”, *Chem. Eur. J.* **2008**, *14*(10):3056–3066.
- [201] P. D’Angelo, V. Barone, G. Chillemi, N. Sanna, W. Meyer-Klaucke, N. V. Pavel, Hydrogen and Higher Shell Contributions in Zn^{2+} , Ni^{2+} , and Co^{2+} Aqueous Solutions: An X-ray Absorption Fine Structure and Molecular Dynamics Study, *J. Am. Chem. Soc.* **2002**, *124*(9):1958–1967.
- [202] P. D’Angelo, A. Zitolo, V. Migliorati, G. Chillemi, M. Duvail, P. Vitorge, S. Abadie, R. Spezia, Revised Ionic Radii of Lanthanoid(III) Ions in Aqueous Solution, *Inorg. Chem.* **2011**, *50*(10):4572–4579.
- [203] M. Dolg, H. Stoll, A. Savin, H. Preuss, Energy-adjusted Pseudopotentials for the Rare Earth Elements, *Theor. Chim. Acta* **1989**, *75*(3):173–194.
- [204] X. Cao, M. Dolg, Valence Basis Sets for Relativistic Energy-consistent Small-core Lanthanide Pseudopotentials, *J. Chem. Phys.* **2001**, *115*:7348–7355.

- [205] X. Cao, M. Dolg, Segmented Contraction Scheme for Small-core Lanthanide Pseudopotential Basis Sets, *J. Mol. Struct.: (THEOCHEM)* **2002**, *581*(13):139–147.
- [206] A. E. Clark, Density Functional and Basis Set Dependence of Hydrated Ln(III) Properties, *J. Chem. Theory Comput.* **2008**, *4*(5):708–718.
- [207] V. Buzko, I. Sukhno, M. Buzko, Ab Initio and DFT Study of Lu³⁺ Hydration, *J. Mol. Struct.: (THEOCHEM)* **2009**, *894*(13):75–79.
- [208] J. Kuta, A. E. Clark, Trends in Aqueous Hydration Across the 4f Period Assessed by Reliable Computational Methods, *Inorg. Chem.* **2010**, *49*(17):7808–7817.
- [209] J. Ciupka, X. Cao-Dolg, J. Wiebke, M. Dolg, Computational Study of Lanthanide(III) Hydration, *Phys. Chem. Chem. Phys.* **2010**, *12*:13215–13223.
- [210] J. Yang, M. Dolg, Valence Basis Sets for Lanthanide 4f-in-core Pseudopotentials Adapted for Crystal Orbital Ab Initio Calculations, *Theor. Chem. Acc.* **2005**, *113*(4):212–224.
- [211] N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas, M. Head-Gordon, The Performance of Density Functionals for Sulfate-Water Clusters, *J. Chem. Theory Comput.* **2013**, *9*(3):1368–1380.
- [212] Y. Marcus, A Simple Empirical Model Describing the Thermodynamics of Hydration of Ions of Widely Varying Charges, Sizes, and Shapes, *Biophys. Chem.* **1994**, *51*(23):111–127.
- [213] Y. Murata, N. Kato, K. Fujiwara, K. Komatsu, Solid-State [4+2] Cycloaddition of Fullerene C₆₀ with Condensed Aromatics Using a High-Speed Vibration Milling Technique, *J. Org. Chem.* **1999**, *64*(10):3483–3488.
- [214] F. London, The General Theory of Molecular Forces, *Trans. Faraday Soc.* **1937**, *33*:8–26.
- [215] S. Grimme, P. R. Schreiner, Steric Crowding Can Stabilize a Labile Molecule: Solving the Hexaphenylethane Riddle, *Angew. Chem. Int. Ed.* **2011**, *50*(52):12639–12642.
- [216] P. R. Schreiner, L. V. Chernish, P. A. Gunchenko, E. Y. Tikhonchuk, H. Hausmann, M. Serafin, S. Schlecht, J. E. P. Dahl, R. M. K. Carlson, A. A. Fokin, Overcoming Lability of Extremely Long Alkane Carbon-carbon Bonds through Dispersion Forces, *Nature* **2011**, *477*(7364):308–311.
- [217] J. S. Sinninghe Damsté, M. Strous, W. I. C. Rijpstra, E. C. Hopmans, J. A. J. Geenevasen, A. C. T. van Duin, L. A. van Niftrik, M. S. M. Jetten, Linearly Concatenated Cyclobutane Lipids Form a Dense Bacterial Membrane **2002**, *419*(6908):708–712.
- [218] J. P. Wagner, P. R. Schreiner, Nature Utilizes Unusual High London Dispersion Interactions for Compact Membranes Composed of Molecular Ladders, *J. Chem. Theory Comput.* **2014**, *10*(3):1353–1358.
- [219] S. Grimme, Do Special Noncovalent π - π Stacking Interactions Really Exist?, *Angew. Chem. Int. Ed.* **2008**, *47*:3430–3434.

- [220] M. Alonso, T. Woller, F. J. Martín-Martínez, J. Contreras-García, P. Geerlings, F. DeProft, Understanding the Fundamental Role of σ/σ , π/π and σ/π Dispersion Interactions in Shaping Carbon-Based Materials, *Chem. Eur. J.* **2014**, *20*(17):4931–4941.
- [221] J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods I. Method, *J. Comput. Chem.* **1989**, *10*(2):209–220.
- [222] J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-optimization of Parameters, *J. Mol. Model.* **2013**, *19*(1):1–32.
- [223] N. C. Craig, L. G. Piper, V. L. Wheeler, Thermodynamics of *cis-trans* Isomerizations. II. 1-chloro-2-fluoroethylenes, 1,2-difluorocyclopropanes, and Related Molecules, *J. Phys. Chem.* **1971**, *75*(10):1453–1460.
- [224] I. V. Alabugin, T. A. Zeidan, Stereoelectronic Effects and General Trends in Hyperconjugative Acceptor Ability of σ Bonds, *J. Am. Chem. Soc.* **2002**, *124*(12):3175–3185.
- [225] C. Bonifaci, A. Ceccon, S. Santi, C. Mealli, R. W. Zoellner, Cofacial and Antarafacial Indenyl Bimetallic Isomers: a Descriptive MO Picture and Implications for the Indenyl Effect on Ligand Substitution Reactions, *Inorg. Chim. Acta* **1995**, *240*(12):541 – 549.
- [226] T. Schwabe, S. Grimme, J.-P. Djukic, Noncovalent Metal-Metal Interactions: The Crucial Role of London Dispersion in a Bimetallic Indenyl System, *J. Am. Chem. Soc.* **2009**, *131*(40):14156–14157.
- [227] W. Kirmse, *Carbene Chemistry*, New York: Academic Press **2001**.
- [228] P. Costa, W. Sander, Hydrogen Bonding Switches the Spin State of Diphenylcarbene from Triplet to Singlet, *Angew. Chem. Int. Ed.* **2014**, *53*(20):5122–5125.
- [229] J. Cioslowski, A. Szarecka, D. Moncrieff, Energetics, Electronic Structures and Geometries of Didehydroazines, *Mol. Phys.* **2003**, *101*(6):839–858.
- [230] Y. Wang, J. Y.-J. Shyy, S. Chien, Fluorescence Proteins, Live-Cell Imaging, and Mechanobiology: Seeing Is Believing, *Ann. Rev. Biomed. Eng.* **2008**, *10*(1):1–38.
- [231] A. Toniolo, S. Olsen, L. Manohar, T. J. Martinez, Conical intersection dynamics in solution: The chromophore of Green Fluorescent Protein, *Faraday Discuss.* **2004**, *127*:149–163.
- [232] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, A. I. Krylov, Effect of Protein Environment on Electronically Excited and Ionized States of the Green Fluorescent Protein Chromophore, *J. Phys. Chem. B* **2011**, *115*(25):8296–8303.
- [233] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov, A. I. Krylov, First-Principles Characterization of the Energy Landscape and Optical Spectra of Green Fluorescent Protein along the A→I→B Proton Transfer Route, *J. Am. Chem. Soc.* **2013**, *135*(31):11541–11549.
- [234] M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, S. J. Remington, Crystal Structure of the *Aequorea Victoria* Green Fluorescent Protein, *Science* **1996**, *273*(5280):1392–1395.

- [235] S. Daly, P. Piccoli, A. Schultz, T. Todorova, L. Gagliardi, G. Girolami, Synthesis and Properties of a Fifteen-Coordinate Complex: The Thorium Aminodiboranate [Th(H₃BNMe₂BH₃)₄], *Angew. Chemie Int. Ed.* **2010**, *49*(19):3379–3381.
- [236] X. Gu, G.-h. Chen, M. Ji, Y.-x. Yao, X.-g. Gong, Superatomic Orbitals in Sixteen-coordinate M@Li₁₆ Bonded by Metallic Bonds, *Nanoscale* **2012**, *4*:2567–2570.
- [237] R. M. Badger, A Relation Between Internuclear Distances and Bond Force Constants, *J. Chem. Phys.* **1934**, *2*(3):128–131.
- [238] R. M. Badger, The Relation Between the Internuclear Distances and Force Constants of Molecules and Its Application to Polyatomic Molecules, *J. Chem. Phys.* **1935**, *3*(11).
- [239] J. Zhang, M. Dolg, Labile Capping Bonds in Lanthanide(III) Complexes: Shorter and Weaker, *J. Phys. Chem. A* **2015**, *119*(4):774–780.
- [240] S. Shaik, A. Shurki, D. Danovich, P. C. Hiberty, Origins of the Exalted *b*_{2u} Frequency in the First Excited State of Benzene, *J. Am. Chem. Soc.* **1996**, *118*(3):666–671.
- [241] C. D. Sherrill, P. Piecuch, The *X*¹Σ_g⁺, *B*¹Δ_g, and *B*¹Σ_g⁺ states of C₂: A Comparison of Renormalized Coupled-cluster and Multireference Methods with Full Configuration Interaction Benchmarks, *J. Chem. Phys.* **2005**, *122*(12):124104.
- [242] K. G. Dyall, J. K. Fægri, *Introduction to Relativistic Quantum Chemistry*, New York: Oxford University Press **2007**.
- [243] A. Rodríguez-Rodríguez, D. Esteban-Gómez, A. de Blas, T. Rodríguez-Blas, M. Fekete, M. Botta, R. Tripier, C. Platas-Iglesias, Lanthanide(III) Complexes with Ligands Derived from a Cyclen Framework Containing Pyridinecarboxylate Pendants. The Effect of Steric Hindrance on the Hydration Number, *Inorg. Chem.* **2012**, *51*(4):2509–2521.
- [244] D. Pubanz, G. Gonzalez, D. H. Powell, A. E. Merbach, Unexpectedly Large Change of Water Exchange Rate and Mechanism on [Ln(DTPA-BMA)(H₂O)] Complexes along the Lanthanide(III) Series, *Inorg. Chem.* **1995**, *34*(17):4447–4453.
- [245] D. P. Martin, P. G. Blachly, A. R. Marts, T. M. Woodruff, C. A. F. de Oliveira, J. A. McCammon, D. L. Tierney, S. M. Cohen, ‘Unconventional’ Coordination Chemistry by Metal Chelating Fragments in a Metalloprotein Active Site, *J. Am. Chem. Soc.* **2014**, *136*(14):5400–5406.
- [246] J. A. Gámez, M. Yañez, [FAAF][−] (A = O, S, Se, Te) or How Electrostatic Interactions Influence the Nature of the Chemical Bond, *J. Chem. Theory Comput.* **2013**, *9*(12):5211–5215.
- [247] L. Helm, A. E. Merbach, Inorganic and Bioinorganic Solvent Exchange Mechanisms, *Chem. Rev.* **2005**, *105*(6):1923–1960.
- [248] M. Duvail, R. Spezia, P. Vitorge, A Dynamic Model to Explain Hydration Behaviour along the Lanthanide Series, *ChemPhysChem* **2008**, *9*(5):693–696.
- [249] E. J. Wheelwright, F. H. Spedding, G. Schwarzenbach, The Stability of the Rare Earth Complexes with Ethylenediaminetetraacetic Acid, *J. Am. Chem. Soc.* **1953**, *75*(17):4196–4201.

- [250] A. Abbasi, P. Lindqvist-Reis, L. Eriksson, D. Sandström, S. Lidin, I. Persson, M. Sandström, Highly Hydrated Cations: Deficiency, Mobility, and Coordination of Water in Crystalline Nonahydrated Scandium(III), Yttrium(III), and Lanthanoid(III) Trifluoromethanesulfonates, *Chem.–Eur. J.* **2005**, *11*(14):4065–4077.
- [251] F. A. Dunand, S. Aime, A. E. Merbach, First ^{17}O NMR Observation of Coordinated Water on Both Isomers of $[\text{Eu}(\text{DOTAM})(\text{H}_2\text{O})]^{3+}$: A Direct Access to Water Exchange and its Role in the Isomerization, *J. Am. Chem. Soc.* **2000**, *122*(7):1506–1512.
- [252] S. Aime, M. Botta, M. Fasano, M. P. M. Marques, C. F. G. C. Geraldes, D. Pubanz, A. E. Merbach, Conformational and Coordination Equilibria on DOTA Complexes of Lanthanide Metal Ions in Aqueous Solution Studied by ^1H -NMR Spectroscopy, *Inorg. Chem.* **1997**, *36*(10):2059–2068.
- [253] A. I. Krylov, Equation-of-motion Coupled-cluster Methods for Open-shell and Electronically Excited Species: The Hitchhiker’s Guide to Fock Space, *Annu. Rev. Phys. Chem.* **2008**, *59*(1):433–462.
- [254] B. Jeziorski, H. J. Monkhorst, Coupled-cluster Method for Multideterminantal Reference States, *Phys. Rev. A* **1981**, *24*:1668–1681.
- [255] L. Meissner, Fock-space Coupled-cluster Method in the Intermediate Hamiltonian Formulation: Model with Singles and Doubles, *J. Chem. Phys.* **1998**, *108*(22):9227–9235.
- [256] M. Hanrath, An Exponential Multireference Wave-function Ansatz, *J. Chem. Phys.* **2005**, *123*(8):084102.

Abbreviations and Acronyms

2D	two-dimensional
3D	three-dimensional
AIM	atoms in molecule
AMO	active molecular orbital
An ³⁺	trivalent actinide ion
AO	atomic orbital
APTS	A Parallel incremenTal Scheme
BCP	bond critical point
BO	Born–Oppenheimer approximation
BSIE	basis set incompleteness error
BTP	bicapped trigonal prism
CABS	complementary auxiliary basis set
CABS	complementary auxiliary basis set
CBS	complete basis set
CBS	complete basis set
CC	coupled-cluster
CCP	cage critical point
CCSD	coupled-cluster singles and doubles
CCSD(T)	CCSD with perturbative treatment of triples
CID	configuration interaction doubles
CIM	cluster-in-molecule
CMO	canonical molecular orbital
CN	coordination number
CS	coordinate system
DC	divide-and-conquer
DEC	divide-expand-consolidate
dHBDI	deprotonated 4-hydroxybenzylidene-2,3-dimethylimidazolinone
DLPNO	domain localized PNO
DODPA	6,6'-((1,4,7,10-tetraazacyclododecane-1,7-diyl)bis(methylene))dipicolinic acid
DOS	density-of-states
DOTAM	1,4,7,10-tetrakis(carbamoylmethyl)-1,4,7,10-tetraazacyclododecane
DTPA-BMA	1,7-bis[(N-methylcarbamoyl)methyl]-1,4,7-triazaheptane-1,4,7-triacetate
ECP	effective core potential
ELF	electron localization function
EOM	equation of motion
ER	Edmiston–Ruedenberg
FCI	full configuration interaction
FCKM	fixed-center K-means clustering
GFP	green fluorescent protein
GGA	generalized gradient approximation
Gmd	guanidinium

HF	Hartree–Fock
HOMO	highest occupied MO
inc3-db-B0	third-order incremental dual-basis set zero-buffer approach
KM	K-means clustering
KS	Kohn–Sham
LBO	Laplacian bond order
LDA	local density approximation
LMO	localized molecular orbital
Ln^{3+}	trivalent lanthanide ion
LUMO	lowest unoccupied MO
MO	molecular orbital
MP2	second-order Møller–Plesset
MR	multi-reference
MSAP	monocapped SAP
MTSAP	monocapped twisted SAP
NCI	noncovalent interaction
NCP	nuclear critical point
OSV	orbital specific virtual
PAO	projected atomic orbital
PDB	protein data bank
PES	potential energy surface
PES	potential energy surface
PNO	pair natural orbitals
PP	pseudopotential
RCP	ring critical point
RDG	reduced density gradient
RMSD	root mean square deviation
RSPT	Rayleigh–Schrödinger perturbation theory
SAP	square antiprism
SCF	self-consistent field
SD	Slater determinant
SE	Schrödinger equation
SR	single-reference
STG	singlet-triplet gap
TTP	tricapped trigonal prism
TZDO/TZDS	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetaldehyde/tetraethanethial
VDE	vertical detachment energy
vdW	van der Waals
WFN	wave function
XC	exchange-correlation functional

Acknowledgments

I would be very pleased to appreciate Prof. Dr. Michael Dolg for giving me the opportunity to work in his group, and his continuous and enthusiastic help for my scientific work and other aspects of my life in Germany! He is an excellent advisor and I am proud of working in his group.

I also want to thank Dr. Xiaoyan Cao-Dolg. She is a kind advisor and friend, and has offered numerous help to my academic activities and daily life. It is a pleasure to work with her.

I am very grateful to Dr. Michael Hanrath. Over these years I have learned a lot from him about theoretical chemistry, software engineering and programming techniques, which are critical to my researches.

Many thanks to my former and current office mates. Martin Böhler has helped me a lot especially in the very early days of my arrival in Germany; Dr. Jan Ciupka, Dr. Tim Hangele, Norah Heinz and Daniel Weißmann have excellent discussions and suggestions on science with me. Especially Norah Heinz and I have many scientific collaborations and she checked my German abstract carefully; Dr. Tim Hangele helped me a lot during the Ph.D. examination period. I would also thank Birgitt Börsch-Pulm, Joseph Held, Oliver Mooßen, Ilyas Türkmen and Dr. Danial Pape for the friendship and their help.

I thank Dr. Haichang Zhang for being a good cooperater and friend. We together accomplished many interesting works and it is a delightful experience. Also I want to express my gratitude to my friends in Department of Chemistry: Dr. Qifang Wang, Dr. Lisong Xiao, Chengzheng Xu, Qingwei Du and Shuangzhou Wang.

I have a nice time working with so many friends from Gemeinschaft Chinesischer Chemiker und Chemieingenieure in Deutschland e.V. (GCCCD). Besides Dr. Xiaoyan Cao-Dolg, Dr. Lisong Xiao, Qingwei Du and Shuangzhou Wang, I also want to thank Dr. Hongchu Du, Dr. Jianwei Tong, and Yu Qiao. We have organized and attended several GCCCD annual conferences together.

I want to thank my old and new friends in Germany. Lang Xu kindly helped me in German language; Yang Zhang and his wife Yiwen Sun, Xiaonan Xu (“Anna”), Qing Sun, and Ao Yuan are my close friends in Cologne; Dr. Li Li encouraged and supported me during a “difficult” period. They have become important parts of my life.

I want to give my deep appreciation to my parents: Yingxin Zhang and Yufang Guan, who gave me my life and a sweet childhood. Without them all the things important to me will be impossible. They are the best and most important people to my life!

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit — einschließlich Tabellen, Karten und Abbildungen — die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie — abgesehen von unten angegebenen Teilpublikationen — noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Michael Dolg betreut worden.

Teilpublikationen:

- J. Zhang, M. Dolg, *J. Chem. Theory Comput.* **2015**, *11*, 962–968.
- J. Zhang, M. Dolg, *J. Phys. Chem. A* **2015**, *119*, 774–780.
- J. Zhang, M. Dolg, *Chem.–Eur. J.* **2014**, *20*, 13909–13912.
- J. Zhang, N. Heinz, M. Dolg, *Inorg. Chem.* **2014**, *53*, 7700–7708.
- J. Zhang, M. Dolg, *J. Chem. Phys.* **2014**, *140*, 044114.
- J. Zhang, M. Dolg, *J. Chem. Theory Comput.* **2013**, *9*, 2992–3003.
- H. Zhang, J. Zhang, B. Tieke, *Polymer* **2015**, *60*, 215–220.
- N. Heinz, J. Zhang, M. Dolg, *J. Chem. Theory Comput.* **2014**, *10*, 5593–5598.
- H. Zhang, J. Zhang, B. Tieke, *Polym. Chem.* **2014**, *5*, 646–652.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und ich verpflichte mich, jedmögliche, die obigen Angaben betreffende, Veränderung dem Dekanat unverzüglich mitzuteilen.

(Datum)

(Unterschrift)

Curriculum Vitae

Lebenslauf

Persönliche Daten

Jun Zhang (in Chinese: 张鋆)

Geboren am 1. Oktober 1986
in Tianjin, V. R. China

Schulbildung

- 09/2012–01/2015 Promotion bei Prof. M. Dolg an der Universität zu Köln zum Thema *Incremental scheme: A general approach for electron correlation computations of large molecules*
- 09/2009–06/2012 Masterarbeit bei Prof. H. Sun an der Nankai Universität zum Thema *Theoretical study on the stereo-selectivity of plant hormone abscisic acid receptor PYL2*
- 09/2005–06/2009 Bachelorarbeit bei Prof. H. Sun an der Nankai Universität zum Thema *The constructing of parallel computational chemistry cluster*
- 09/2002–06/2005 Oberschule in Tianjin
- 09/1999–06/2002 Unterschule in Tianjin
- 09/1993–06/1999 Grundschule in Tianjin

April 12, 2015