

Population Analyses Based on Ionic Partition of Overlap Distributions

by

Yiming Wang

B.Sc. in Physics, Dalian University of Technology, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2018

© Yiming Wang 2018

Committee Page

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled: *Population Analyses Based on Ionic Partition of Overlap Distributions* submitted by *Yiming Wang* in partial fulfillment of the requirements for the degree of *Master of Science in Chemistry*.

Examining Committee

Research Supervisor:

Dr. Yan Alexander Wang, Chemistry

Supervisory Committee Members:

Dr. Mark Thachuk, Chemistry

Dr. Gren Patey, Chemistry

Abstract

In this thesis, we bring up several new schemes of partitioning the atomic partial charges for the purpose of reducing the dependency on the basis sets and the inaccuracy from previous methods we did in our group. We analyze all the methods including Mulliken, evaluate them by comparing with Natural Population Analysis (NPA) with several different groups of systems which we divide according to their polarity. We find that when applied to more polarized systems such as compounds containing Fluorine, our Population Analyses Based on Ionic Partition of Overlap Distributions (IPOP) series perform better and produce charges closer to those of NPA method. Within the same system, IPOP series work better for atoms with more polarized bond than for atoms with non-polarized ones. On top of all the analyses for separate groups, we plot the correlation between charges produced by different methods with charges generated by NPA method. From the graph and the slope value we conclude that IPOP2d is the method which gives the most reliable result compared to NPA among all the methods. Also, in order to figure out the best basis set which can represent the result of IPOP2d, we plot the correlation graph between charges produced by IPOP2d and NPA methods for several basis sets. We find that 6-31G basis set is the most representative basis set. Using the 6-31G to calculate charges for certain systems renders us lots of advantages in terms of computational efficiency while still providing a reasonable result.

Lay Summary

A frequent topic in quantum chemistry is the determination of the electronic configuration and the electronic charge distribution of a molecule, especially the net charges associated to each atoms within a molecule. While we know the electronic charge is hard to observe by experiment directly, we need to get the charge distribution among the constituent atoms in a molecule by a given wavefunction. The process to carry out this analysis is considered as population analysis. In order to solve the problems of the method our group previously did, such as the high dependency of basis sets and inaccuracy of the results, we bring up several new schemes to partition the electronic charge. We analyze the results calculated with different basis sets and compare them with the results generated by Natural Population Analysis (NPA) and evaluate how good they perform within polarized and less polarized systems.

Preface

In the theory part, IPOD1 and IPOD2 methods were developed in Dr. Alex Wang's group prior to my arrival. IPOD1 and IPOD2 methods were originally started by Dr. Wang and Mr. Yakun Chen. The following up methods IPOD2b, IPOD2c, and IPOD2d are from Dr. Wang and Mr. Miguel Garcia, a PhD student in our group. I collaborated with Mr. Miguel to put the idea into implementation into NWChem. The choosing of the systems and data analysis was done by myself.

Table of Contents

Abstract	iii
Lay Summary	iv
Preface	v
Table of Contents	vi
List of Tables	viii
List of Figures	x
List of Abbreviations	xii
Acknowledgements	xiii
1 Theory	1
1.1 Electronic Density Analysis	1
1.2 Gaussian	2
1.3 Basis Set	2
1.4 Basics for Quantum Chemistry	9
1.4.1 Hartree-Fock	10
1.4.2 Density Functional Theory	16
1.5 Population Analysis	20
1.5.1 Mulliken and Löwdin Population Analyses	22
1.5.2 Natural Population Analysis (NPA)	23
1.5.3 Ionic Partition of Overlap Distribution (IPOD) methods	24
2 Result and Discussion	27
2.1 Fluorine Systems	27
2.2 Alcohol Systems	31
2.3 Alkene Systems	34

Table of Contents

2.4	Aromatic Systems	43
2.5	Small Inorganic Molecules	43
2.6	Correlation between Different Methods with NPA	49
2.7	The Best Basis Set within IPOD2d Method	49
3	Summary	53
	Bibliography	55

List of Tables

2.1	Average charges for all basis sets and standard deviations for HF and LiF.	28
2.2	Average charges with all basis sets and standard deviations for NaF and KF.	28
2.3	Average charges and standard deviations for atom C in CH ₃ OH and C ₂ H ₅ OH for basis sets from 3-21G to 6-31G*.	31
2.4	Average charges and standard deviations for atom C in C ₃ H ₇ OH and C ₄ H ₉ OH for basis sets from 3-21G to 6-31G*	31
2.5	Average charges and standard deviations for atom Carbon (side) in C ₂ H ₄ and C ₃ H ₄ for basis sets from 3-21G to 6-31G(2pd, 2p).	36
2.6	Average charges and standard deviations for atom Carbon (side) in C ₃ H ₆ and C ₄ H ₆ for basis sets from 3-21G to 6-31G(2pd, 2p).	36
2.7	Average charges and standard deviations for atom Carbon (middle) in C ₂ H ₄ and C ₃ H ₄ for basis sets from 3-21G to 6-31G(2pd, 2p).	36
2.8	Average charges and standard deviations for atom C (middle) in C ₃ H ₆ and C ₄ H ₆ for basis sets from 3-21G to 6-31G(2pd, 2p).	38
2.9	Average charges and standard deviations for atom Carbon (side) in Anthracene and Biphenyl for basis sets from 3-21G to 6-31G**.	39
2.10	Average charges and standard deviations for atom Carbon (side) in Naphtalene and Phenanthrene for basis sets from 3-21G to 6-31G**.	39
2.11	Average charges and standard deviations for atom Carbon (middle) in Anthracene and Biphenyl for basis sets from 3-21G to 6-31G**.	42
2.12	Average charges and standard deviations for atom Carbon (middle) in Naphtalene and Phenanthrene for basis sets from 3-21G to 6-31G**.	42

List of Tables

2.13	Average charges and standard deviations for atom Carbon in CO ₂ and CO for basis sets from 3-21G to 6-31G**.	45
2.14	Average charges and standard deviations for atom Nitrogen in HCN and NH ₃ for basis sets from 3-21G to 6-31G**.	45
2.15	Average charges and standard deviations for atom Nitrogen in N ₂ O and HNO for basis sets from from 3-21G to 6-31G**.	47
2.16	Average charges and standard deviations for atom Nitrogen in H ₂ O and HOOH for basis sets from 3-21G to 6-31G**.	47

List of Figures

2.1	Partial charges for Fluorine diatomic compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	29
2.2	Partial charges for C atom in Alcohol compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	32
2.3	Partial charges for O atom in Alcohol compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	33
2.4	Partial charges for Carbon-side atom in Alkenes compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	35
2.5	Partial charges for C-middle atom in Alkenes compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	37
2.6	Partial charges for C-side atom in Aromatic compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	40
2.7	Partial charges for C-middle atom in Aromatic compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	41

List of Figures

2.8	Partial charges for atom in small inorganic molecules. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	44
2.9	Partial charges for Oxygen atom in small inorganic molecules. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)	48
2.10	Correlation between different methods with NPA	50
2.11	Correlation between different methods with NPA	51
2.12	Correlation for different basis sets in IPOD2d with the same basis sets in NPA	52

List of Abbreviations

IPOD	Partition of Overlap Distribution
MPA	Mulliken Population Analysis
LPA	Löwdin Population Analysis
NPA	Natural Population Analysis
AIM	Atoms In Molecules
LCAO	Linear Combination of Atomic Orbitals
STO	Slater-Type Orbital
GTO	Gaussian-Type Orbital
DFT	Density Functional Theory
SCF	Self-consistent Procedure
CGF	contracted Gaussian functions
PGTO	primitive Gaussian-Type Orbital
DZ	Double Zeta
DZP	Double Zeta plus Polarization
PT	Perturbation Theory
KE	kinetic energy
HF	Hartree-Fock
HK	Hohenberg-Kohn
KS	Kohn-Sham
SD	Slater determinant
MO	Molecular Orbital
AOs	Atomic Orbitals
NOs	Natural Orbitals
NAOs	Natural Atomic Orbitals
CBPA	Christoffersen-Baker population analysis

Acknowledgements

I would like to thank Professor Wang for his patient mentoring and expert advice throughout this project. Also, I would like to thank Miguel for his extraordinary support in this thesis process. And finally, last but by no means least, also to everyone in my lab. It was great sharing laboratory with all of you during last three years. Thanks for all their encouragement!

Chapter 1

Theory

A frequent topic in quantum chemistry is the determination of the electronic configuration and electronic charge distribution of a molecule, especially net charge associated with each atom in a polyatomic molecule [1]. While this charge distribution is hard to observe directly by experiment, it is important for rendering certain chemical interpretation of the wave function which leads to useful understanding of chemical phenomena [2].

To quantify the notions of atomic charge and orbital population in a satisfactory way, there are a number of ways for analysing a calculation when accurate wavefunctions are available [3].

These analysis mainly fall into two categories:

1. Partition of charge between atoms based on the orbital occupancy.
2. Partition of a physical observable derived from the wavefunction, such as electron density.

1.1 Electronic Density Analysis

When we have an optimized structure and the molecular wavefunction Ψ , we can calculate electron density which is the square of the wavefunction. Hence we can examine how electron density distributes over a molecule directly.

For partitioning of a physical observable derived from the wave function, such as electron density, we have Bader Atoms In Molecules (AIM) analysis [4]. It relies on properties of the electron density alone and the information can be obtained from the laplacian of the electron density [5].

For partitioning the molecular wave function by some orbital based scheme, Ψ is written as a Slater determinant of individual molecular orbitals $\psi_i(\mathbf{r})$ [6], and electronic density ρ is defined as

$$\rho(\mathbf{r}) = 2[|\psi_1(\mathbf{r})|^2 + |\psi_2(\mathbf{r})|^2 + \dots + |\psi_{N/2}(\mathbf{r})|^2], \quad (1.1)$$

where N is the electron number, \mathbf{r} is the space vector.

All of the molecular orbitals together must contribute to the electron density. Each molecular orbital is expanded in terms of the atomic orbitals

(AO) basis functions (Slater-Type Orbital or (contracted) Gaussian-Type Orbital) [1]. In the LCAO (Linear Combination of Atomic Orbitals) approximation, if we insert the orbital expansion into ρ , ρ becomes a sum of the contributions. If we go further and see how ρ is distributed in terms of particular atoms and locate the charge on the atomic centers (nucleus), we carry out a population analysis [7].

1.2 Gaussian

Gaussian is a popular and widely used Computational Chemistry Software package, especially for electronic structure calculations [8]. It has been updated continuously since firstly released in 1970 by John Pople. We use 09 version for our project although the latest one is Gaussian 16. Gaussian provides a wide-range modelling capabilities from the prediction of energies, molecular structures and vibrational frequencies to the prediction of reactions in a wide variety of chemical environments [9].

Also, Gaussian offers various methods for modelling compounds and chemical process. Including Hartree-Fock methods (restricted, unrestricted open-shell), Density Functional Theory (DFT) and Molecular Mechanics. In our project, we mainly use DFT methods to do calculations for charge analysis with different basis sets. Among them, DFT and Hartree-Fock method are both Self-consistent Procedure (SCF methods).

1.3 Basis Set

For basis set, many basis sets are stored internally in the Gaussian and the following are what we applied in our projects:

STO-3G, 3-21G, 4-31G, 6-31G, 6-31G*, 6-31G **, 6-31G(2pd, 2p), 6-31G(3df, 3pd), 6-31++G 6-31++G*, 6-31++G**.

In the following section, we will show the definition of different basis sets and illustrate how specific attributes of a basis set influence calculated quantities [10].

The 1s Minimal STO-3G Basis Set

In this section, instead of introducing basis sets for general polyatomic molecule calculations, we describe 1s type basis functions as a start. The extension of these concepts for the general case which includes s,p,d-type basis functions will follow the same rule [11].

1.3. Basis Set

In terms of 1s functions, there are mainly two types of basis functions which are widely used:

1. The normalized 1s Slater-type function of the form

$$\phi_{1s}^{SF}(\xi, r - R_A) = (\xi^3 \pi)^{1/2} e^{-\xi|r-R_A|^2}, \quad (1.2)$$

where R_A is the center, and ξ is the Slater orbital exponent.

2. The normalized 1s Gaussian-type function of the form

$$\phi_{1s}^{GF}(\alpha, r - R_A) = (2\alpha/\pi)^{3/4} e^{-\alpha|r-R_A|^2}, \quad (1.3)$$

where α is the Gaussian orbital exponent [12].

The orbital exponents which are positive numbers control the width of the orbital. Large values of ξ or α give a tight function while small values give a diffuse function. There are two main factors one should consider in terms of choosing of a basis, which are the efficiency and accuracy of describing electronic wave function calculations. For best efficiency, one intends to use fewer possible terms when expanding the molecular orbital ψ_i .

$$\psi_i = \sum_{\mu=1}^k C_{\mu i} \phi_{\mu} \quad (1.4)$$

In this sense, Slater functions have an advantage over Gaussian functions. Fewer Slater basis functions than Gaussian basis functions are needed to provide the same level of accurate results. Another consideration is the time consumption of the two-electron integral evaluation. In an SCF calculation, one of the most expensive steps is the calculation of the two-electron integrals which have the form

$$(\mu_A \nu_B | \lambda_C \sigma_D) = \int dr_1 dr_2 \phi_{\mu}^{*A}(r_1) \phi_{\nu}^B(r_1) r_{12}^{-1} \phi_{\lambda}^{*C}(r_2) \phi_{\sigma}^D(r_2). \quad (1.5)$$

The fact that the evaluation of the four-center integrals is more computationally costly with Slater-basis functions makes Gaussian functions a better choice in this scenario. The reason for that is for a Gaussian functions can be applied with explicit formulas while Slater-basis functions do not have. Also, the name of software Gaussian derives from its use of GTOs.

The reason why these integrals are much easier to calculate with Gaussian basis function, is that the products of Gaussians are Gaussians centered between atom centers. The product of two Gaussian functions is a third Gaussian function centered between two atoms, that is,

$$\phi_{1s}^{GF}(\alpha, r - R_A) \phi_{1s}^{GF}(\beta, r - R_B) = K_{AB} \phi_{1s}^{GF}(p, r - R_p), \quad (1.6)$$

1.3. Basis Set

where K_{AB} is

$$K_{AB} = (2\alpha\beta/[(\alpha + \beta)\pi]^{3/4}) \exp[-\alpha\beta/(\alpha + \beta)|R_A - R_B|^2]. \quad (1.7)$$

The exponent of the new Gaussian centered at R_p is $p = \alpha + \beta$ and the third center P is on a line joining the centers A and B,

$$R_p = (\alpha R_A + \beta R_B)/(\alpha + \beta). \quad (1.8)$$

The same rule applies for four-center integral. For a 1s Gaussian, it can be reduced to two-center integrals.

$$(\mu_A \nu_B | \lambda_C \sigma_D) = K_{AB} K_{CD} \int dr_1 dr_2 \phi_{1s}^{GF}(P, r_1 - R_P) r_{12}^{-1} \phi_{1s}^{GF}(Q, r_2 - R_Q). \quad (1.9)$$

Although the two-electron integrals can be evaluated more efficiently with Gaussian functions, the fact that Gaussian functions have inaccurate functional behaviour of molecular orbitals hinders them from being the optimum basis functions. Thereafter, we can use basis functions which are fixed linear combinations of the primitive Gaussian functions ϕ_p^{GF} as a trade off. These linear combinations called contracted Gaussian functions (CGF) which are given by

$$\phi_\mu^{CGF}(r - R_A) = \sum_{p=1}^L d_{p\mu} \phi_p^{GF}(\alpha_{p\mu}, r - R_A), \quad (1.10)$$

where L is the length of the contraction and $d_{p\mu}$ is a contraction coefficient. There is a functional relationship between the pth normalized primitive Gaussian ϕ_p^{GF} and the basis function ϕ_μ^{CGF} by the Gaussian orbital exponent $\alpha_{p\mu}$ which is also called contraction exponent.

The method consist in choosing a certain contraction length, contraction coefficients and contraction exponents to build a desirable set of basis functions ϕ_μ^{CGF} on the lefthand side. While using these fixed functions to do calculations in molecular wave function, especially an SCF calculation, the contraction coefficients, should remain unchanged.

With the contracted basis set functions $\{\phi_\mu^{CGF}\}$, the two-electron integrals $(\mu\nu|\lambda\alpha)$ can be evaluated as a sum of easily calculated two-electron integrals over the original Gaussian functions. If the contraction parameters are appropriately chosen, one can have atomic Hartree-Fock functions, Slater-type functions, etc and still only use primitive Gaussian functions to get an efficient integral. The idea behind it is to get a linear combination of N primitive Gaussian functions by fitting a Slater-type-orbital (STO),

1.3. Basis Set

which is called STO-NG procedure. STO-3G basis sets are the most widely used in polyatomic calculation to rapidly evaluate integrals since it has been found that using more than three Primitive Gaussian-type-orbital (PGTO) to represent the STO gives little improvement [13]. Let us first consider using ϕ_{1s}^{CGF} to approximate a 1s Slater-type function with $\xi = 1.0$ as an example.

$$\phi_{1s}^{CGF}(\xi = 1.0, STO - 1G) = \phi_{1s}^{GF}(\alpha_{11}), \quad (1.11)$$

$$\phi_{1s}^{CFG}(\xi = 1.0, STO - 2G) = d_{12}\phi_{1s}^{GF}(\alpha_{12}) + d_{22}\phi_{1s}^{GF}(\alpha_{22}), \quad (1.12)$$

$$\phi_{1s}^{CFG}(\xi = 1.0, STO - 3G) = d_{13}\phi_{1s}^{GF}(\alpha_{13}) + d_{23}\phi_{1s}^{GF}(\alpha_{23}) + d_{33}\phi_{1s}^{GF}(\alpha_{33}), \quad (1.13)$$

We only consider up to three contractions where we need to find the best-fit coefficients $d_{p\mu}$ and exponents $\alpha_{p\mu}$ with a fixed $\xi = 1.0$ to get a basis function which is the closest to a Slater-type function. To do this, we can use a least square techniques to minimize the integral [13]

$$I = \int dr [\phi_{1s}^{SF}(\xi = 1.0, r) - \phi_{1s}^{CGF}(\xi = 1.0, STO - LG, r)]^2. \quad (1.14)$$

Equivalently, we get the maximization overlap between the two functions by calculating

$$S = \int dr \phi_{1s}^{SF}(\xi = 1.0, r) \phi_{1s}^{CGF}(\xi = 1.0, STO - LG, r), \quad (1.15)$$

on the condition that both functions in this equation are normalized. For the STO-1G we only need to get the primitive Gaussian exponent α that maximizes the overlap [11],

$$S = (\pi)^{-1/2} (2\alpha/\pi)^{3/4} \int dr e^{-r} e^{-\alpha r^2}. \quad (1.16)$$

The maximized overlap is obtained when $\alpha = 0.270950$. We can observe deficiencies in behaviour of Gaussian functions near the origin and at large distances. The same theory can be applied to the STO-2G and STO-3G to get the optimum fits. The results are shown as follows:

$$\phi_{1s}^{CGF}(\xi = 1.0, STO - 1G) = \phi_{1s}^{CGF}(0.270950), \quad (1.17)$$

$$\begin{aligned} \phi_{1s}^{CGF}(\xi = 1.0, STO - 2G) = & 0.678914\phi_{1s}^{CGF}(0.151623) + \\ & 0.430129\phi_{1s}^{CGF}(0.851819), \end{aligned} \quad (1.18)$$

$$\begin{aligned} \phi_{1s}^{CGF}(\xi = 1.0, STO - 3G) = & 0.444635\phi_{1s}^{CGF}(0.109818) + \\ & 0.535328\phi_{1s}^{CGF}(0.405771) + \\ & 0.1543329\phi_{1s}^{CGF}(2.22766). \end{aligned} \quad (1.19)$$

The general notation for the STO-3G contraction is $(6s3p|3s)|[2s1p|1s]$ [14]. In the parentheses, the number before the slash represents for heavy atoms (first row element). The number after the slash represents for hydrogen [15]. The basis in the square bracket indicates the corresponding number of contracted functions. Note this notation only illustrates the size of the final basis without revealing how the contraction is done.

Minimal basis sets usually provide rough results which are insufficient for research level publications, although they are much cheaper than their larger counterparts. Also, the minimum basis sets have limitation in variational flexibility and are not capable to render accurate representation of orbitals. To solve these issues, we use multiple functions to represent each orbital. For instance, the double-zeta basis set allows us to treat each orbital as

$$\Phi_{2s}(r) = \Phi_{2s}^{STO}(r, \xi_1) + d\Phi_{2s}^{STO}(r, \xi_2), \quad (1.20)$$

so that a 2s atomic orbital can be expressed as the sum of two STOs. The two STOs differ in ξ which determines how large the orbital is. The constant d represents for how much each STO contributes towards the final orbital. The same theory applies for the triple and quadruple-zeta basis sets where each orbital is expanded as sum of three or four STO, respectively. The purpose of the trade-off is to get better accuracy with less time. There are different ways of extending basis sets, such as splitting the valence functions, adding polarized functions and adding diffuse functions [16].

To improve the minimal STO-3G basis set, the first step we do is doubling all basis functions, providing a Double Zeta (DZ) type basis. Doubling the number of basis functions can provide a more accurate description of the electron distribution in molecules while the distribution differs significantly from the one in atoms. Alternatively, we can add polarization functions that benefits the description for chemical bond by introducing directionality which the minimal basis fails to provide. There exists chemical bonding between valence orbitals. For instance, doubling the 1s-functions in carbon results in a better description of 1s-electrons. Nevertheless, when it is getting closer to the atomic case, the 1s orbital is basically independent of the chemical environment. An adjustment made to the DZ type basis is only

1.3. Basis Set

doubling the number of valence orbitals, which is more commonly used as a split valence basis [17].

One of the most popular split-valence basis sets is Pople’s basis sets n-ijG or n-ijkG [18]. n is the number of primitives for the innershells; ij or ijk represents the number of primitives for contractions in the valence shell. The ij/ijk notations describe sets of valence double/triple zeta quality respectively. Here, we use Pople’s 3-21G basis set notation as an example. “3” indicates the number of gaussian functions summed to describe the inner shell orbital. The number “2” is the number of gaussian functions that comprise the first STO of the double zeta. The number “1” is the number of gaussian functions summed in the second STO.

Here, a more concrete example in terms of Hydrogen 1s orbital expanded by Double Zeta Basis sets: 3-21G will be showed. For hydrogen, the contraction are

$$\phi'_{1s}(r) = \sum_{i=1}^2 d'_{i,1s} g_{1s}(\alpha'_{i,1s}, r) \text{ and} \quad (1.21)$$

$$\phi''_{1s}(r) = g_{1s}(\alpha''_{1s}, r). \quad (1.22)$$

In this case, three s-type gaussian primitives are contracted to two basis functions. The inner hydrogen function ϕ'_{1s} is a contraction of two primitive Gaussians. The other one ϕ''_{1s} is uncontracted. It is frequently denoted as (3s)→[2s] contraction. The coefficient in function are then fixed in subsequent molecular calculations [15].

For the atoms Li to F, the contractions are

$$\phi_{1s}(r) = \sum_{i=1}^3 d_{i,1s} g_{1s}(\alpha_{i,1s}, r), \quad (1.23)$$

$$\phi'_{2s}(r) = \sum_{i=1}^2 d'_{i,1s} g_{1s}(\alpha'_{i,2sp}, r), \quad (1.24)$$

$$\phi''_{2s}(r) = g_{1s}(\alpha''_{2sp}, r), \quad (1.25)$$

$$\phi'_{2p}(r) = \sum_{i=1}^2 d'_{i,2p} g_{2p}(\alpha'_{i,2sp}, r), \quad (1.26)$$

$$\phi''_{2p}(r) = g_{2p}(\alpha''_{2sp}, r). \quad (1.27)$$

To determine the 3-21G basis sets, we firstly need to choose a certain form 1.23 to 1.27 for the contractions. Then we further optimize the corresponding parameters. The designation of 3-21G basis is (6s3p|3s) → [3s2p|2s].

Besides, although 3-21G and STO-3G basis contain the same number of primitive GTOs. 3-21G is much more flexible since they include twice as many valence functions which can create free combinations to make MOs.

The other commonly used split-valence basis sets in our project are 4-31G, 6-31G, 6-311G which are similar to 3-21G. Instead of improving a basis set by going to triple zeta, quadruple zeta, etc, one would rather add functions of higher angular quantum number to make the basis set better balanced [12]. These higher angular momentum functions are denoted as polarization functions. Usually we add p-type function to H and d-type functions to the first row atoms Li-F. For example, originally a C-H bond is described by s-orbital(s) for the hydrogen and s- with p_z - orbitals for carbon. If one only involves s-functions for the hydrogen, the difference of the electron distribution between the direction along the bond and perpendicular to the bond cannot be described. To compensate that, a polarization of the s-orbital is introduced by adding p-orbital to the hydrogen. In this way, the p component plays a role in improving the description of the H-C bond. Same theory also applies for using d-orbitals to polarize p-orbitals, using f-orbitals to polarize d-orbitals etc [19].

In arguing whether we need to further add the d-orbital to a hydrogen s-orbital if a p-orbital has already been added, we have a general guideline. The most essential part is the first set of polarization functions (i.e., p-functions for hydrogen, d-functions for heavy atoms). Hereafter, the formation of a Double Zeta plus Polarization (DZP) type basis comes from adding a single set of polarization functions (e.g., p-functions on hydrogens and d-functions on heavy atoms). Polarization functions are denoted in Pople's sets by an asterisk. 6-31G* and 6-31G** basis sets are formed by adding polarization functions to a 6-31G basis. 6-31G* describes a basis set where d-type functions be added to a basis set with valence p orbitals. 6-31G** describes when d-type functions are added to the heavy atoms, and p-type functions are added to hydrogen [17]. The 6-31G** is synonymous to 6-31G(d,p). In terms of degree of contraction, the notation for 6-31G* and 6-31G** are $(11s4p1d/4s)/[4s2p1d/2s]$ and $(11s4p1d/4s1p)[4s2p1d/2s1p]$ respectively. It has been proved by experience that adding polarization functions to the heavy atoms plays a more essential role than adding polarization functions to hydrogen. In addition to polarization functions, the basis sets are also frequently augmented with the diffuse functions. These functions have very small ξ exponents and decay slowly with distance from the nucleus. Diffuse functions are normally s- and p- functions and frequently go before the G. Diffuse gaussians provide an accurate description of anions and weak bond such as hydrogen bonds. Also, they are necessary for cal-

calculation of properties (like dipole moment, Rydberg states, polarizabilities, etc). For Pople's basis sets, diffuse functions are denoted by + or ++. The only one + or the first of two + indicates that adding one set of diffuse s- and p- functions on heavy atoms. The second + indicates that a diffuse s-function is also added to hydrogens [19]. Similar as we discussed about polarization functions, the diffuse functions we add is both on hydrogen and non-hydrogen atoms. Here are some examples about diffuse functions. The 6-31+G(d) represents a double zeta split valence basis which has one set of diffuse sp-functions on heavy atoms only and has a single d-type polarization functions on heavy atoms. The similar theory applies for a more complicated case, like a 6-311++G(2df, 2pd) represents which a triple zeta split valence which has additional diffuse sp-functions, two d- and one f-functions on heavy atoms as well as diffuse s- and two p- and one d- functions on hydrogens [20].

1.4 Basics for Quantum Chemistry

Determining the electronic configuration and charge distribution is a fundamental topic in quantum chemistry. Since charge distribution is important for understanding certain chemical phenomena while cannot be observed directly by experiment, we need the molecular wavefunction Ψ to help interpret certain features of an optimised structure [21]. The wave function $\Psi(x, t)$ is the solution to Schrödinger's equation which contains the information about the system. The wavefunctions themselves do not have physical significance, rather the physical significance lies in the interpretation of the product of the wavefunction and its complex conjugate. As the total electronic density $\rho(x, t)$ is the square of the wavefunction. Here, we interpret $\rho(x, t)$ as the probability of finding the particle at time t at position x.

Although we can solve the equation for part of certain simple cases, it is impossible to get an exact solution for most real molecules. There, we must resort to approximation methods. There are some effective approximate methods such as the Perturbation Theory (PT) and its modifications or Hartree-Fock approximation, which is a non-perturbative method for multi-electron atoms. Hartree-Fock approximation is equivalent to the molecular orbital approximation, which is fundamental to chemistry. The theory provides the idea that a single-particle function (orbital) is enough to describe each electron's motion since the instantaneous motion of electrons is independent from one to another. Besides, the Hartree-Fock theory plays an important role in quantum chemistry not only for its own sake, but also

because it contains effects of electron correlation which provides a good start for more accurate approximation [22].

1.4.1 Hartree-Fock

Born Oppenheimer Approximation

We will give a brief explanation of Hartree-Fock theory from an introductory level and illustrate how to calculate molecular orbitals using Hartree-Fock theory. The aim to develop Hartree-Fock is to approximately solve the Electronic Schrödinger equation

$$\hat{H}_{el}(r; R)\Psi(r; R) = E_{el}(R)\Psi(r; R). \quad (1.28)$$

To achieve that, we need to invoke the Born Oppenheimer approximation first, which guarantees the separation between the motion of the nuclei and the motion of the electrons. As a consequence, the Hamiltonian operator $\hat{H}_{el}(r; R)$ is grouped into four terms

$$\hat{H}_{el}(r; R) = T_e(r) + \hat{V}_{eN} + \hat{V}_{NN} + \hat{V}_{ee}(r). \quad (1.29)$$

We can put it in a more expanded notation as follows:

$$\left[-\frac{1}{2} \sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|R_I - r_i|} + \sum_{I>J} \frac{Z_I Z_J}{|R_I - R_J|} + \sum_{i>j} \frac{1}{|r_i - r_j|} \right] \Psi(r, R) = E_{el} \Psi(r, R), \quad (1.30)$$

where r denotes electronic and R denotes nucleus degrees of freedom [23]. The Born-Oppenheimer Approximation ignores the motion of the atomic nuclear since it is much heavier than an electron. Based on this physical fact, the Born-Oppenheimer Approximation describes the electronic wavefunction under the approximation that the nucleus is stationary. By neglecting the nuclear kinetic energy term $\hat{K}_{NN}(R)$, the problem becomes solving the electronic Schrödinger equation. By solving the Schrödinger equation, we can extract useful information such as electronic dipole moment and polarizability [22].

The Many-electron Wavefunction: the Slater Determinant

To introduce the basic idea of Hartree-Fock theory, we must mention the Slater-determinant since the Hartree-Fock theory assumes ψ is a Slater determinant. An exact wavefunction to solve a multi-electron system should be in the form of $|\psi(r_1, r_2, \dots, r_i)\rangle$ since a single function relies on the coordinate

of all the electrons simultaneously. A possible approximation consists on writing the multi-electron wavefunction as a product of single-electron functions. By writing $|\Psi(r_1, r_2, \dots, r_j)\rangle \approx |\varphi_1(r_1)\rangle|\varphi_2(r_2)\rangle\dots|\varphi_i(r_i)\rangle$, we transform a multi-electron system into a set of independent electrons located in its own orbital. And these single-electron wavefunctions are called atomic orbitals. However, a correct many electron wavefunction must satisfy both the principle of indistinguishability and the principle of antisymmetry. These principles are not satisfied by a simple product. By introducing Slater determinant as follows, we can build an antisymmetric solution [24].

$$\Psi(1, 2, \dots, N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \cdots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \cdots & \phi_2(N) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_N(1) & \phi_N(2) & \cdots & \phi_N(N) \end{vmatrix} \quad (1.31)$$

The exchange of two columns represents the exchange of two particles which results in the change of sign. And two equal rows will cause a zero determinant which corresponds to Pauli's exclusion principle that two (or more) identical fermions cannot occupy the same quantum state. We can also write it in a shorthand form as $|\chi_i\chi_j\dots\chi_k\rangle$ if we know the list of the occupied orbitals are $\{\chi_i(x), \chi_j(x), \dots, \chi_k(x)\}$ or even as simply as $|ij\dots k\rangle$ [25].

Simplified Notation for the Hamiltonian

We now introduce a simplified notation for the Hamiltonian. We define the one-electron operator as:

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}}, \quad (1.32)$$

which represents the kinetic energy (KE) and the attraction to all nuclei of electron i . The two-electron operator $v(i, j) = \frac{1}{r_{ij}}$ stands for the coulomb repulsion between electron i and j . With these definition, the electronic Hamiltonian can be written as:

$$\hat{H}_{el} = \sum_i h(i) + \sum_{i<j} v(i, j) + V_{NN}. \quad (1.33)$$

In Hartree-Fock approximation, the two-electron operator $v(i, j)$ is replaced by Hartree-Fock Potential $v_{HF}(r_i)$ which is the total averaged potential acting on the electron i from the rest of the other $N-1$ electron spin orbitals.

V_{NN} is a constant which can be ignored at present due to the fixed set of nuclear coordinates $\{R\}$. It will only shift the eigenvalues instead of changing the eigenfunctions.

The Hartree-Fock Energy Expression

We already gained the form of a Slater determinant for the Hartree-Fock wavefunction and a simplified notation for the Hamiltonian. To tackle the problem of obtaining the molecular orbitals, we still need to express the Hartree-Fock Energy in an appropriate way. The Hartree-Fock energy will be put as the usual quantum mechanical form as

$$E_{el} = \langle \Psi | \hat{H}_{el} | \Psi \rangle. \quad (1.34)$$

The variational Theorem states that the energy E_{el} is the upper bound energy of the actual ground state. Hence, we will adjust so-called “variational parameters” until the energy of the trial function is minimized. In this way, variational method approximations would result in a good trial wavefunction and its corresponding energy. By employing a linear combination of a set of given basis functions, we can obtain the molecular orbitals successfully. By finding the linear expansion coefficients that minimize the energy. Here, we introduce how to write the Hartree-Fock energy E_{el} by integrals of the one- and two-electron operators [26], that is,

$$E_{HF} = \sum_i \langle i | h | i \rangle + \frac{1}{2} \sum_{ij} [ii|jj] - [ij|ji], \quad (1.35)$$

where one-electron integral is:

$$\sum_i \langle i | h | i \rangle = \int dx_1 \chi_i^*(x_1) \hat{h}(r_1) \chi_i(x_1). \quad (1.36)$$

Each pair of electrons (in orbitals i and j) has a Coulomb integral, given by

$$[ii|jj] = \int dx_1 \int dx_2 \chi_i^*(x_1) \chi_i(x_1) \frac{1}{r_{12}} \chi_j^*(x_2) \chi_j(x_2). \quad (1.37)$$

Since $\chi_i^*(x_1)\chi_i(x_1)$ and $\chi_j^*(x_2)\chi_j(x_2)$ each stands for the probability electron 1 (and 2) in orbital i is located at x_1 (and x_2). $\frac{1}{r_{12}}$ is coulomb repulsion between electron at x_1 and electron at x_2 .

Overall this integral represents the coulomb repulsion between electron 1 in orbital i and electron 2 in orbital j . Another term $[ij|ji] = \int dx_1 \int dx_2 \chi_i^*(x_1) \chi_j(x_1) \frac{1}{r_{12}} \chi_j^*(x_2) \chi_i(x_2)$ is called “Exchange integral”. It looks like as “Coulomb integral” and we can regard it as exchange the two of the orbital indices [27].

The Hartree-Fock Equations

So far we already obtained the wavefunction in the form of a single Slater determinant consisting on one spin orbital per electron. Also, with the variational theorem applied on the energy expression, the best spin orbitals are those that minimize the electron energy.

$$\begin{aligned}
 E_0 = \langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle &= \sum_i \langle i | h | i \rangle + \frac{1}{2} \sum_{ij} \langle ij | ij \rangle \\
 &= \sum_i \langle i | h | i \rangle + \frac{1}{2} \sum_{ij} [ii | jj] - [ij | ji].
 \end{aligned}
 \tag{1.38}$$

Applying the variational theorem on the energy E_0 while constraining the spin orbitals to be orthogonal to each other, we arrive at the Hartree-Fock integro-differential equation:

$$\begin{aligned}
 h(x_1)\chi_i(x_1) + \sum_{j \neq i} \left[\int dx_2 |x_j(x_2)|^2 r_{12}^{-1} \right] \chi_i(x_1) \\
 - \sum_{j \neq i} \left[\int dx_2 \chi_j^*(x_2) \chi_i(x_2) r_{12}^{-1} \right] \chi_j(x_1) = \varepsilon_i \chi_i(x_1)
 \end{aligned},
 \tag{1.39}$$

where ε_i is the orbital energy associated with orbital χ_i . The Hartree-Fock equations can be solved numerically, usually it is solved by expanding χ_i into a linear combination of basis functions (Hartree-Fock-Roothan equations) [28]. By introducing basis sets, we convert the equation into a much simpler linear algebra problem. Regardless on whether we are solving Hartree-Fock equation exactly or using a basis set expansion, Hartree-Fock equation will depend on the orbitals. Therefore, we need to come up with some rough guess as our initial orbitals and use the solution as the second guess to rebuild the Hartree-Fock equation and do this iteratively until we finally refine our solution to a certain standard. In other words, Hartree-Fock is a self-consistent-field (SCF) approach. Before we formally introduce SCF, we will first explain each term in the Hartree-Fock equation [26].

The Coulomb and Exchange Operators

As the Hartree-Fock Equation shows, the first two-electron term

$$\sum_{j \neq i} \left[\int dx_2 |\chi_j(x_2)|^2 r_{12}^{-1} \right]
 \tag{1.40}$$

is the Coulomb term, which gives the Coulomb interaction of an electron in spin orbital χ_i with total average potential from the rest of the N-1 electrons in other spin orbitals. For future convenience, we define a coulomb operator

$$j_i(x_1) = \int dx_2 |\chi_j(x_2)|^2 r_{12}^{-1} \quad (1.41)$$

which represents the average local potential at x_1 contributed from the charge distribution of the electron in orbital χ_j .

The other term in 1.39 arises from the need to satisfy the antisymmetry of the Slater determinant. We call it the exchange term because it looks like Coulomb term with swapping spin orbitals χ_i and χ_j . Correspondingly, we can introduce an exchange operator $\kappa_j(x_1)$, by considering its effect on an arbitrary spin orbital χ_i :

$$\kappa_j(x_1)\chi_i(x_1) = \left[\int dx_2 \chi_j^*(x_2) r_{12}^{-1} \chi_i(x_2) \right] \chi_j(x_1). \quad (1.42)$$

Using this notation, the Hartree-Fock equations can be written more concisely as:

$$\left[h(x_1) + \sum_{j \neq i} j_i(x_1) - \sum_{j \neq i} \kappa_j(x_1) \right] \chi_i(x_1) = \varepsilon_i \chi_i(x_1). \quad (1.43)$$

Since this has the form of an eigenvalue problem and we can show that

$$[j_i(x_1) - \kappa_i(x_1)]\chi_i(x_1) = 0, \quad (1.44)$$

we can remove the restrictions on summation $j \neq i$ hence, and define the Fock operator f as:

$$f(x_1) = h(x_1) + \sum_j j(x_1) - \kappa_j(x_1). \quad (1.45)$$

Therefore the Hartree-Fock equations become

$$f(x_1)\chi_i(x_1) = \varepsilon_i \chi_i(x_1). \quad (1.46)$$

Rootaan solved this equation by introducing a set of κ known basis functions χ and expanded χ in a linear form:

$$\chi_i = \sum_{\mu=1} C_{\mu i} \tilde{\chi}_\mu. \quad (1.47)$$

For every single spin orbital i , this gives:

$$f(x_1) \sum_{\nu} C_{\nu i} \tilde{\chi}_{\nu}(x_1) = \varepsilon_i \sum_{\nu} C_{\nu i} \tilde{\chi}_{\nu}(x_1). \quad (1.48)$$

Multiply $\tilde{\chi}_{\mu}^*(x_1)$ on the left and integrate will lead the integro-differential equation to a matrix equation.

$$\sum_{\nu} C_{\nu i} \int dx_1 \tilde{\chi}_{\mu}^*(x_1) f(x_1) \tilde{\chi}_{\nu}(x_1) = \varepsilon_i \sum_{\nu} C_{\nu i} \int dx_1 \tilde{\chi}_{\mu}^*(x_1) \tilde{\chi}_{\nu}(x_1). \quad (1.49)$$

To simplify it, we introduce the matrix element notation

$$S_{\mu\nu} = \int dx_1 \tilde{\chi}_{\mu}^*(x_1) \tilde{\chi}_{\nu}(x_1), \quad (1.50)$$

$$F_{\mu\nu} = \int dx_1 \tilde{\chi}_{\mu}^*(x_1) f(x_1) \tilde{\chi}_{\nu}(x_1). \quad (1.51)$$

With this definition, the integrated Hartree-Fock equation can be written as

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i}, i = 1, 2, \dots, k. \quad (1.52)$$

These are so-called the Roothaan equations which can be written in an even more compact form as $FC = SC\varepsilon$. If we want to solve it as an eigenvalue equation, we must first make the overlap matrix S to be the identity matrix by finding a transformation matrix to orthogonalize the basis functions [11].

Self-Consistent-Field Procedure

After we reformulate the equation, it becomes a Pseudo-eigenvalue equation $FC = C\varepsilon$. It is not a regular eigenvalue equation because, although C can be solved by diagonalizing F , F also depends on C , which is its own solution. Therefore the equation will be solved in an iterative fashion. That is the reason that Hartree-Fock equation is solved using a SCF procedure. An SCF procedure consists on the following steps [30].

1. Specify the molecule (including the nuclear coordinates $\{R_A\}$, basis functions and number of atoms $\{Z_A\}$).
2. Form overlap matrix $S_{\mu\nu}$.
3. Guess the initial Molecular Orbital coefficients C .
4. Obtain a Fock matrix F using C from last step.
5. Solve $FC=SC\varepsilon$.

6. By getting the new C from the solution of step 5, build a new Fock matrix F.

7. Back to step 5, repeat step 5 and step 6 until C is quite stable and no longer fluctuates from one iteration to another.

8. When the procedure is converged, use the final solution represented by C, F etc. to calculate other quantities of interest.

1.4.2 Density Functional Theory

Hartree-Fock Approximation although plays an important role from many aspects, it still has some drawbacks and weakness. Particularly, the neglect of electron correlation of the Hartree-Fock Algorithm can lead to a large deviation from experimental results. There are a number of approaches to solve this problem. A better alternate to Hartree-Fock calculations is Density Functional Theory (DFT), which is among the most popular and versatile methods in computational chemistry. In this section, we will give a brief explanation of DFT of which name comes from the use of functionals of the electron density [13].

The Kohn-Sham Equations

The framework for Density Functional Theory (DFT) originates from Hohenberg-Kohn theorems (H-K) which demonstrates that the energy of the ground state electron is determined completely by the electron density ρ . The advantage of an electron density approach when compared with the wavefunction approach is that the electron density only depends on three spatial coordinates while the complexity of a wavefunction increases quickly with the number of electrons [31]. The wavefunction is a function of $3N$ variables, where N is the number of electrons and the density is just a function of three variables. Since the electron density is the square of the wavefunction integrated over $N-1$ electron coordinates, the many-body problem of N electrons is reduced to three spatial coordinates. Although H-K theorem proves that there exists a one-to-one connection between the electron density and the ground state energy, the functional connecting these two quantities remains unknown. The goal of DFT is to formulate functionals bridging the electron density and the energy [32].

Many attempts have been made to design DFT models that involve all the energy components expressed as a functional of the electron density. While these methods fail to perform well, Kohn and Sham came up with a modern DFT method in which we work with a fictitious system of non-

interacting electrons, built from an auxiliary set of orbitals such that the real electron kinetic energy part could be calculated to good accuracy. Then the only unknown functional is the exchange-correlation energy, which is the difference in energy between the real system and the non-interacting system. By this method, since we already get much information calculated exactly, there only left a small remainder to be computed by an approximate functional. We will introduce the Kohn-Sham scheme and also discuss some of its major features. We start by introducing orbitals and the Non-Interacting Reference System [33].

The foundation for Kohn-Sham (KS) DFT method is to construct a fictitious non-interacting system by the introduction of orbitals, in such a way, the density of the interacting electrons is the same as that of the system. Our goal in KS formalism is to split the kinetic energy into two parts, one could be calculated exactly, the remainder is just a small correction. With the KS method, one has to pay the price of the increased variable complexity since the orbitals are re-introduced, so the variables increase from three to $3N$. Also, a separate electron correlation term is needed [21].

The KS model is quite similar to HF method, they share identical formulas for the kinetic, coulomb electron-electron and nuclear-electron energies. In Hartree-Fock, the Slater determinant Φ_{SD} represents an approximation of the true N-electron wave function, while in DFT, Φ_{SD} represents the exact wave function of a fictitious system of N non-interacting electrons. The next crucial step is to introduce an effective, local potential $V_s(\mathbf{r})$ within the Hamiltonian, that we need to describe the non-interacting reference system, that is

$$\hat{H}_s = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_i^N V_s(\mathbf{r}_i). \quad (1.53)$$

The Kohn-Sham orbitals are determined the solution of

$$\hat{f}^{ks} \varphi_i = \varepsilon_i \varphi_i, \quad (1.54)$$

where the one-electron Kohn-Sham operator \hat{f}^{ks} introduced as

$$\hat{f}^{ks} = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_i^N V_s(\mathbf{r}). \quad (1.55)$$

For these Kohn-Sham orbitals, choosing an effective V_s plays an important role in bridging the density of the newly-constructed fictitious non-interacting system and that of the real interacting electron system. We need to find a V_s that makes ρ_s which is the summation of the moduli of squared

orbitals $\{\varphi_i\}$ equals to the real density of the ground state for interacting electrons,

$$\rho_s(\mathbf{r}) = \sum_i^N \sum_s |\varphi(\mathbf{r}, s)|^2 = \rho_o(\mathbf{r}). \quad (1.56)$$

The next step is to find an expression for the potential V_s . We start by introducing the expression of the kinetic energy of the non-interacting system [34]

$$T_s = -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle \quad (1.57)$$

The difference between the kinetic energy between the interacting and non-interacting system is included in the exchange correlation term $E_{xc}[\rho(r)]$. This term also includes the difference of the colomb energy between interacting and non-interacting systems. We can write the energy functional as

$$E[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] + E_{ne}[\rho], \quad (1.58)$$

That is, E_{xc} , defined as:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]) = T_c[\rho] + E_{ncl}[\rho], \quad (1.59)$$

where $T_c[\rho]$ could be considered as the kinetic correlation energy and $E_{ncl}[\rho]$ contains both potential correlation and exchange energy. In other words, the exchange-correlation energy E_{xc} is the functional that includes everything that remains unknown. In terms of T_s , we expect it to be a functional of ρ [35].

We now discuss the problem on how to find V_s that will satisfy the requirement that the density of the non-interacting reference system is the same as the real system. We can also regard the energy of the non-interacting system as two components: the kinetic energy and the energy coming from the interaction with the external potential. With this separation, we have introduced from 1.58, we describe the expression for the energy of real,

interacting system as below [36]:

$$\begin{aligned}
 E[\rho(\mathbf{r})] &= T_s[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] + E_{ne}[\rho] \\
 &= T_s[\rho(\mathbf{r})] + \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho] + \int V_{Ne}\rho(\mathbf{r}) d\mathbf{r} \\
 &= -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \iint |\varphi_i(\mathbf{r}_1)|^2 \frac{1}{r_{12}} |\varphi_j(\mathbf{r}_2)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \\
 &\quad + E_{xc}[\rho(\mathbf{r})] - \sum_i^N \sum_A^M \frac{Z_A}{r_{1A}} |\varphi_i(\mathbf{r}_1)|^2 d\mathbf{r}_1
 \end{aligned} \tag{1.60}$$

Except for the unknown part E_{xc} , This is very similar to the Hartree-Fock energy, so we can proceed in the same way here. We apply the variational principle and minimize the energy expression with the constrain of orthogonal orbitals. As a result, we get the equations

$$\left(-\frac{1}{2}\nabla^2 + \left[\int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{xc}(\mathbf{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}}\right]\right)\varphi_i = \left(-\frac{1}{2}\nabla^2 + V_{eff}(\mathbf{r}_1)\right)\varphi_i = \varepsilon_i \varphi_i. \tag{1.61}$$

By comparing it with the one-particle equations from the non-interacting system, we will find that V_{eff} is same as V_s

$$V_s(\mathbf{r}) \equiv V_{eff}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{xc}(\mathbf{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}}, \tag{1.62}$$

where V_{eff} , depends on the density through the coulomb term. Using the same notation from the Hartree Fock integro-differential equation, we write the Kohn-Sham Equation as

$$[h(x_1) + j(x_1) + V_{xc}(x_1)]\chi_i(x_1) = \varepsilon_i \chi_i(x_1). \tag{1.63}$$

Same as what we have done with the Hartree-Fock Approximation, we also resort to basis set approximation and get the similar formalism $FC = SC\varepsilon$. F is Kohn-Sham matrix and hence the Kohn-Sham one-electron equation can be solved iteratively by SCF procedure we have introduced before. The potential V_{xc} is defined as the functional derivative of E_{xc} of ρ .

$$V_{xc} \equiv \frac{\delta E_{xc}}{\delta \rho} \tag{1.64}$$

Another important point we need to mention about Kohn-Sham approach is that unlike the Hartree-Fock model, where the approximation enters from the start, Kohn-Sham theory is in fact exact. It is when we have to figure out an explicit form of the unknown functional for the exchange-correlation energy E_{xc} and the corresponding potential V_{xc} that we need to resort to an approximation. The quality of a DFT calculation will depend on the quality of these two terms.

1.5 Population Analysis

We now proceed to describe the main topic: Population Analysis.

One of the many molecular properties that can be extracted from the wavefunction is the electronic charge distribution. Unfortunately, it is hard to observe this distribution directly from experiment. This raises a question: given a wavefunction, how can we get an idea of the way charge is distributed among the constituent atoms of a molecule? Answering this question is the goal of population analysis. The most commonly used population analysis methods fall into two categories [3]:

- (1) Partitioning the molecular wavefunction using orbital based scheme.
- (2) Partitioning of a physical observable derived from the wavefunction such as electron density.

Population Analysis Based on Basis Function

Two of the most frequently used orbital-based partitioning schemes are Mulliken and Löwdin [22]. They are also highly related to our project. In order to illustrate them clearly, we start by introducing some basic notations. We already know the electron density ρ represents the probability of finding an electron at a certain position \mathbf{r} . We can also define an electron density ρ_i for a single molecular orbital. It can be written as the square of a single molecular orbital containing one electron. $\rho_i^{MO}(\mathbf{r}) = f_i |\psi_i^{MO}(\mathbf{r})|^2$ where $\psi_i^{MO}(\mathbf{r}) = \sum_{\mu}^{AO} \sum_A^{atom} C_{i\mu}^A \phi_{\mu}^A(\mathbf{r})$ represents for a molecular orbital after a SCF calculation has converged. f_i is the occupation number of orbital ψ_i^{MO} . If we sum the electron densities over all orbitals, we get the total electron density [37]:

$$\rho_{tot}(\mathbf{r}) = \sum_i^{MO} \rho_i^{MO}(\mathbf{r}). \quad (1.65)$$

1.5. Population Analysis

The total density is normalized to the total number of electrons in the system n_{tot} :

$$\int \rho_{tot}(\mathbf{r}) d\mathbf{r} = n_{tot}. \quad (1.66)$$

If each orbital is expanded in terms of a set of normalized, but non-orthogonal basis functions ϕ . The electron density can be written as an expansion of the basis functions ϕ and the expansion coefficients $d_{\mu\nu}^{AB}$ define the elements of the density matrix:

$$\rho_{tot}(\mathbf{r}) = \sum_{\mu,\nu} \sum_{A,B}^{AO \text{ atom}} d_{\mu\nu}^{AB} \phi_{\mu}^{A*} \phi_{\nu}^B. \quad (1.67)$$

We can separate $\rho_{tot}(\mathbf{r})$ into two terms, one belonging to the same atomic center and another belonging to two separate centers, as:

$$\rho_{tot}(\mathbf{r}) = \sum_{\mu,\nu} \sum_A^{AO \text{ atom}} d_{\mu\nu}^{AA} \phi_{\mu}^{A*} \phi_{\nu}^A + \sum_{\mu,\nu} \sum_{A \neq B}^{AO \text{ atom}} d_{\mu\nu}^{AB} \phi_{\mu}^{A*} \phi_{\nu}^B. \quad (1.68)$$

The first part stands for the ionic component of the total electron density since it includes the electron density from basis sets centered around a single atom:

$$\rho_{tot}^{ionic}(\mathbf{r}) = \sum_{\mu,\nu} \sum_A^{AO \text{ atom}} d_{\mu\nu}^{AA} \phi_{\mu}^{A*} \phi_{\nu}^A. \quad (1.69)$$

Similarly, the ionic electron density of atom A, ρ_A^{ionic} is defined as

$$\rho_A^{ionic}(\mathbf{r}) = \sum_{\mu,\nu}^{AO} d_{\mu\nu}^{AA} \phi_{\mu}^{A*} \phi_{\nu}^A. \quad (1.70)$$

The second term in Eq.(1.68) represents the covalent part of electronic density since it describes the overlap between two different atoms. If we only sum over unique pairs of atom centers $A \neq B$, since the density matrix is symmetrical we get exactly half of the covalent part.

$$\rho_{tot}^{covalent}(\mathbf{r}) = \sum_{\mu,\nu} \sum_{A \neq B}^{AO \text{ atom}} d_{\mu\nu}^{AB} \phi_{\mu}^{A*} \phi_{\nu}^B. \quad (1.71)$$

Similarly, the covalent contribution between atom A and B are defined as:

$$\rho_A^{covalent}(\mathbf{r}) = \frac{1}{2} \sum_{\mu,\nu} \sum_{A \neq B}^{AO \text{ atom}} d_{\mu\nu}^{AB} \phi_{\mu}^{A*} \phi_{\nu}^B. \quad (1.72)$$

1.5. Population Analysis

Furthermore, by integrating respective electron densities, we can get the ionic and covalent charges as:

$$n_A^{ionic} = \int \rho_A^{ionic}(\mathbf{r}) d\mathbf{r} = \sum_{\mu,\nu}^{AO} d_{\mu\nu}^{AA} S_{\mu\nu}^{AA}, \quad (1.73)$$

$$n_{AB}^{covalent} = \int \rho_{AB}^{covalent}(\mathbf{r}) d\mathbf{r} = 2 \sum_{\mu,\nu}^{AO} d_{\mu\nu}^{AB} S_{\mu\nu}^{AB}, \quad (1.74)$$

$$n_A^{covalent} = \frac{1}{2} n_{AB}^{covalent} = \frac{1}{2} \int \rho_{AB}^{covalent}(\mathbf{r}) d\mathbf{r} = \sum_{\mu,\nu}^{AO} d_{\mu\nu}^{AB} S_{\mu\nu}^{AB}, \quad (1.75)$$

$$n_A = n_A^{ionic} + n_A^{covalent}, \quad (1.76)$$

where S is the overlap matrix:

$$S_{\mu\nu} = \int \phi_{\mu}^*(\mathbf{r}) \phi_{\nu}(\mathbf{r}) d\mathbf{r}. \quad (1.77)$$

1.5.1 Mulliken and Löwdin Population Analyses

Among the methods, the simplest one is the Mulliken scheme which partitions the covalent contribution equally between the two atoms. In Mulliken population analysis, the charge on atom A, is the ionic charge on A, Eq.(1.70) plus half of all the covalent contribution containing atom A, Eq.(1.72). That is, the total number of electrons associated with certain atom A comes from the sum of the contributions from all AOs located on that atom A. Based on different schemes, chemists develop various methods on how the contribution involving basis functions is divided on different atoms. The gross atomic population comes from the sum over all the atomic orbitals on a given atom. We then use this gross atomic population to subtract the nuclear charge and get the partial charge on each atom:

$$Q_A = Z_A - n_A. \quad (1.78)$$

While Mulliken method partitions the matrix $d_{\mu\nu} S_{\mu\nu}$, another frequently used method, Löwdin, uses the symmetric orthogonalization of matrix d,

$$d' = S^{\frac{1}{2}} \cdot d \cdot S^{\frac{1}{2}}. \quad (1.79)$$

The rest of Löwdin proceeds as Mulliken but replacing d by d' . Both Mulliken and Löwdin are just two typical examples among the population analysis using $S^n \cdot d \cdot S^{1-n}$ matrices [38]. Neither of them is superior to the other

one in terms of getting the best result. Mulliken does not require an orthogonal basis, thus it may yield negative gross population or value greater than two. On the contrary, Löwdin analysis performs a transformation of all the atomic orbitals to make it an orthogonal basis. With the orthogonalized basis, the number of electrons are constrained to the correct range [0,2] [39].

1.5.2 Natural Population Analysis (NPA)

Besides Mulliken and Löwdin, there is another popular method called “natural population analysis” which has been developed to make population analysis in general atomic orbital basis sets [40]. This method shows significant improvement when we involve ionic compound compared to conventional Mulliken Population Analysis. In this section, we will give a brief explanation on Natural Population Analysis by first comparing it with conventional natural orbitals (NOs) and then outline the construction of the method.

Relationship to Natural Orbitals

The NAOs resemble the conventional NOs introduced by Löwdin. For isolated atoms, they are both orthogonal atomic orbitals of maximal occupancy. But for polyatomic molecules, a distinction must be mentioned. For conventional NOs, they are defined as the orthonormal molecular orbitals of maximal occupancy and are thus completely delocalized. On the contrary, the NAOs are defined as orthonormal atomic orbitals of maximal occupancy and as a consequence, they are localized on individual atoms within each block of a molecule.

Outline of the Construction of the Method

There are essentially two steps for the construction of NAOs: First, diagonalizing one-center blocks of the density matrix to gain a set of “pre-NAOs”. Secondly, removing interatomic overlap. As a result from the first step, the diagonalization will lead to two categories of the basis of occupancy: (1) the “minimal” set which corresponds to occupied atomic orbitals for the isolated atom. (2) The “Rydberg” set which is the remaining part based on the magnitude of the occupation numbers significantly greater than zero. Since there exists overlap between one center and another in the pre-NAOs, the orbitals are not eligible to assess the atomic charge. Therefore, we need to remove the interatomic overlap as the second step. Our main idea is to orthogonalize the whole set of orbitals while still trying to maintain the atom-centered qualities of pre-NAOs as much as we can. In order to achieve that goal, we

need to perform the orgonalization transformation on our pre-NAOs. Here, we summarize it briefly as four essential steps:

(1) Divide the pre-NAOs within each block into the “minimal” set and Rydberg sets.

(2) Perform the occupancy-weighted symmetric orthogonalization on all the minimal functions.

(3) By applying a standard Gram-Schmidt orthogonalization, make the “minimal” set and Rydberg set on the same center orthogonal to each other.

(4) Perform an occupancy-weighted procedure to make the Rydberg sets on one center orthogonal to Rydberg sets on another center.

After all four procedures, we finally get a set of orthogonal orbitals which are denoted as NAOs. The diagonal elements of this density matrix represent the orbital populations. The atomic charge comes from the sum of contributions of the orbitals centered in that atom. Empirically it is found that the “minimal” set plays an essential role in the contribution of the electron density, as it counts up to 99% of the electron density. Furthermore, like in the case of Löwdin, the electron occupation numbers are constrained between 0 and 2. As the basis set grows, the charges converge to a well-defined number. However, NAOs also have some disadvantages. For example, they can extend too far away from the atom where they are centered. This would lead to the scenario where the contribution of electron density near one nucleus actually belongs to another nucleus [40].

In our calculation, NPA plays an important role and it serves as the standard which we use to evaluate other methods since for most systems NPA generates stable and reasonable results with all basis sets. However, NPA requires orthogonalization during the process, which is computationally expensive. For big molecules and in large basis sets, we should try to avoid using NPA and find other substitute methods. Another motivation we want to develop other method for calculation is that NPA is not widely used in computational software. We need to come up with methods that are easily implemented with most softwares.

1.5.3 Ionic Partition of Overlap Distribution (IPOD) methods

Mulliken and Löwdin Population analyses do not usually generate stable results when calculated with different basis sets. NPA performs well in terms of stability and accuracy. However, since it is computationally expensive, we are looking for some methods that do not require any orthogonalization while still generate the same quality results as NPA. That serves the motivation

1.5. Population Analysis

of developing our IPOD series.

In this section, we introduce the previously published population analysis by Ionic Partition of Charge Distribution (IPOD) methods [37]. For our own project, we made some correction on partitioning the covalent part of the electron charge. We start by introducing the simplest method, IPOD1. The approximation for IPOD1 is consist of distributing the gross atomic population, n_A^{IPOD1} according to the proportion of its ionic charge:

$$n_A^{IPOD1} \propto n_A^{ionic}, \quad (1.80)$$

where n_A^{ionic} is the ionic charge for atom A from Eq.(1.73) while we satisfy this proportionality, we need to make sure that the gross atomic population add up to the total number of electrons n_{tot}

$$n_A^{IPOD1} = \frac{n_A^{ionic}}{n_{tot}^{ionic}} n_{tot}. \quad (1.81)$$

Obviously, this method simply ignores the contribution from the covalent part. IPOD2 differs from IPOD1 in which it takes into account the covalent part which is partitioned by the same ratio as the ionic charge. In mathematical form, the covalent density for atom A becomes:

$$n_A^{covalent} = \frac{n_A^{ionic}}{n_A^{ionic} + n_B^{ionic}} \cdot n_{AB}^{covalent} + \frac{n_A^{ionic}}{n_A^{ionic} + n_C^{ionic}} \cdot n_{AC}^{covalent}. \quad (1.82)$$

The idea of IPOD2 resembles that of Christoffersen-Baker population analysis (CBPA). However, CBPA requires the orthogonalization of the basis function between one atomic center with another. Besides, instead of applying the global density matrix, it performs the partition for every molecular orbital. IPOD2 still has the problem that partitioning by the ionic density ratio results too much weight for heavier atoms which contradicts our rule that the partition should somehow reflect the electro-negativity. To fix that, we propose several other schemes to alleviate the influence of the dependence on atomic number. Usually for heavier atoms, their atomic number is big, so for IPOD2b we want to balance the ratio by dividing the atomic number Z_A . For heavier atoms, sometimes they will likely have larger core electrons, so IPOD2c is to alleviate the influence by removing the core electron number. IPOD2d can be regarded as a combination of IPOD2b and IPOD2c, and we hope it would fix the partition in the right direction and become a trade off if either of 2b or 2c goes a bit over-correction.

a) IPOD2b: First divide n_A^{ionic} by the atomic number Z_A then use the result

1.5. Population Analysis

as the ratio to partition the covalent part.

$$R_A = \frac{n_A^{ionic}}{Z_A}, \quad (1.83)$$

$$n_{A,IPOD2b}^{covalent} = \frac{R_A}{R_A + R_B} \cdot n_{AB}^{covalent} + \frac{R_A}{R_A + R_C} \cdot n_{AC}^{covalent}. \quad (1.84)$$

b) IPOD2c: First subtract the number of core electrons n_{core} from n_A^{ionic} , then renormalize by the whole group.

$$G_A = n_A^{ionic} - n_{A,core}, \quad (1.85)$$

$$n_{A,IPOD2c}^{covalent} = \frac{G_A}{G_A + G_B} \cdot n_{AB}^{covalent} + \frac{G_A}{G_A + G_C} \cdot n_{AC}^{covalent}. \quad (1.86)$$

c) IPOD2d: First subtract the number of core electrons and afterwards divide the subtracted part by the number of valence electron number $n_{valence}$.

$$F_A = \frac{n_A^{ionic} - n_{A,core}}{n_{A,valence}}, \quad (1.87)$$

$$n_{A,IPOD2d}^{covalent} = \frac{F_A}{F_A + F_B} \cdot n_{AB}^{covalent} + \frac{F_A}{F_A + F_C} \cdot n_{AC}^{covalent}. \quad (1.88)$$

Here, we have to especially mention that for oxygen and fluorine. To avoid the over-correction, we divide by five instead the actual number of valence electrons. This number 5 is from our empirical calculation, and it fix the over-correction to a good extent.

Chapter 2

Result and Discussion

Our Ionic Partition of Charge Distribution (IPOD) series algorithms with Mulliken Population Analysis (MPA) and Löwdin Population Analysis (LPA) methods are implemented by Fortran 77 in nwchem program with MPA, LPA methods. NPA method is implemented with Gaussian09 platform. We carried out the studied systems with a series of basis sets: STO-3G, 3-21G, 4-31G, 6-31G, 6-31G*, 6-31G**, 6-31G(2pd, 2p), 6-31G(3df, 3pd), 6-31++G, 6-31++G*, 6-31++G**, using the hybrid density functional B3LYP method in Gaussian09. To visualize the results, we used python to plot the atomic charge of different methods vs. the increased basis sets. We can see from the results (e.g., in Figure 2.1) that the charges of NPA do not generally fluctuate a lot with different basis sets while in other cases, the results generally depend on the basis sets. Therefore, the basis set independency of NPA makes it a good standard to measure the quality of other methods. The choice of systems for our analysis was made based on their functional groups and their polarity. We then analysed the results by comparing them with different basis sets using python graphs.

2.1 Fluorine Systems

For the Fluorine atom in Hydrogen Fluoride (HF) in Table 2.1, compared to the most stable method NPA, our methods of IPOD2b, IPOD2c, IPOD2d all work pretty well in terms of accuracy and stability. Compared to the average of NPA charge which is -0.529, IPOD2b and IPOD2d have the similar average of -0.509 and -0.541 respectively. IPOD2c with an average of -0.620 is slightly more polarized than NPA but is still the most stable among the IPOD methods with a standard deviation of 0.034, just slightly bigger than that of NPA which is 0.024. In diatomic systems, IPOD1 and IPOD2 generate the same results since IPOD2 partitions the covalent part the same ratio as the ionic part. Both of IPOD1 and IPOD2 polarize the molecule too much, therefore our IPOD2b is to balance this partition by dividing the atomic number Z_A . IPOD2c is to penalize the result by removing the core electrons. In terms of HF system from Figure 2.1(a), we can conclude

2.1. Fluorine Systems

Method	HF		LiF	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.385	0.113	-0.545	0.099
LPA	-0.248	0.224	-0.429	0.188
NPA	-0.529	0.024	-0.821	0.083
IPOD1	-0.623	0.033	-0.697	0.064
IPOD2	-0.623	0.033	-0.697	0.064
IPOD2b	-0.509	0.066	-0.585	0.095
IPOD2c	-0.620	0.034	-0.778	0.048
IPOD2d	-0.541	0.059	-0.730	0.066

Table 2.1: Average charges for all basis sets and standard deviations for HF and LiF.

Method	NaF		KF	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.631	0.164	-0.765	0.125
LPA	-0.530	0.193	-0.760	0.140
NPA	-0.833	0.102	-0.881	0.091
IPOD1	-0.622	0.170	-0.735	0.135
IPOD2	-0.622	0.170	-0.735	0.135
IPOD2b	-0.643	0.162	-0.770	0.120
IPOD2c	-0.809	0.102	-0.855	0.085
IPOD2d	-0.776	0.125	-0.840	0.100

Table 2.2: Average charges with all basis sets and standard deviations for NaF and KF.

2.1. Fluorine Systems

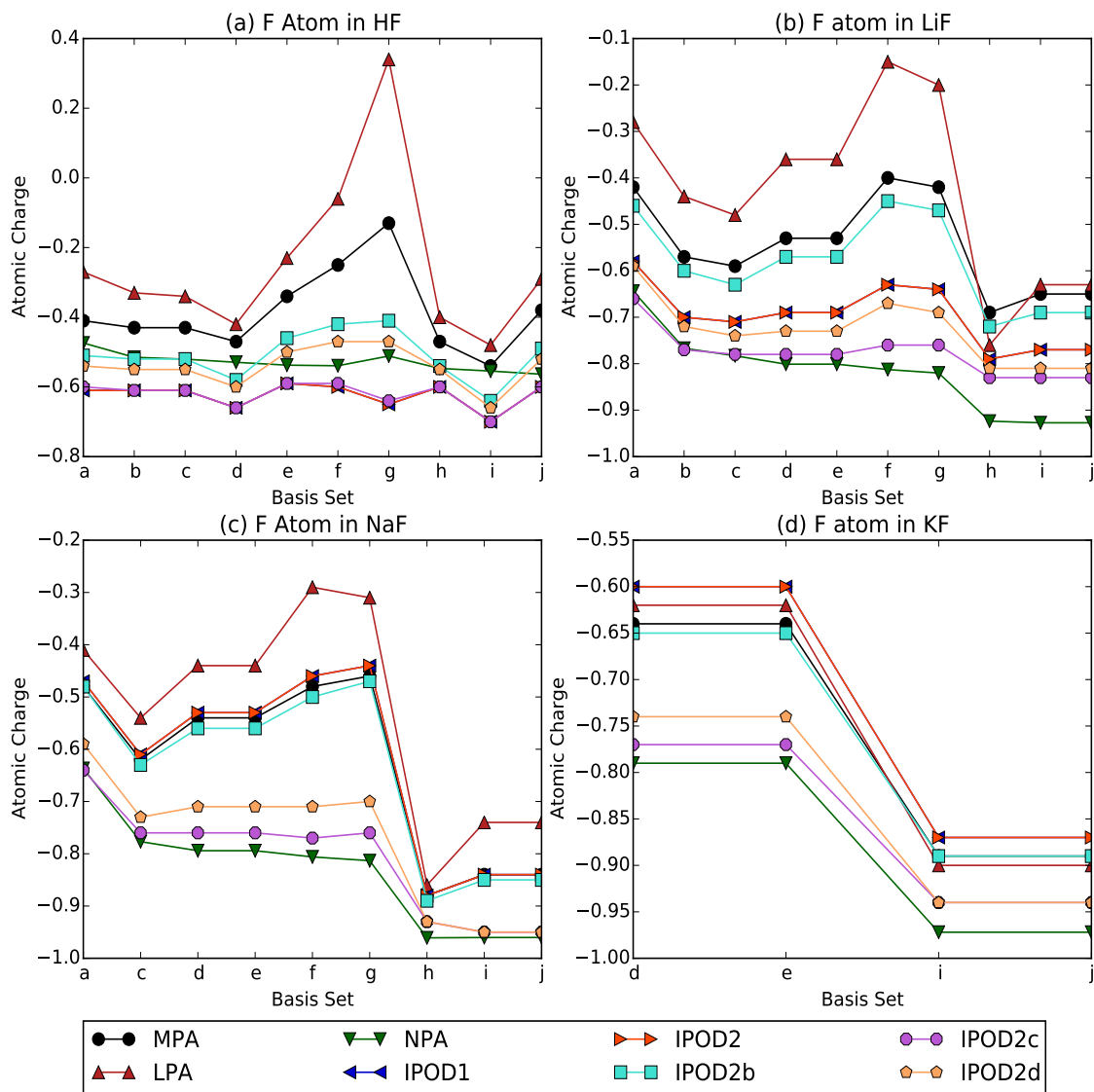


Figure 2.1: Partial charges for Fluorine diatomic compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.1. Fluorine Systems

that our method of IPOD2b, IPOD2c, and IPOD2d work much better than MPA and LPA although our methods do not require any orthogonalization. Especially for LPA, it fluctuates too much with different basis sets. For the basis set 6-31(3df, 3pd), the charge already becomes positive while in reality the charge of Fluorine should always remain negative.

For Fluorine atom in Lithium Fluoride (LiF) from Figure 2.1(b), IPOD2b has the same trend as LPA and MPA and all of the IPOD methods fluctuate severely compared to the rest of the methods. IPOD2b penalize the heavier atom F and makes it less negative. In terms of deviation in Table 2.1, IPOD1, IPOD2, IPOD2c and IPOD2d all work better than NPA since they have smaller standard deviation than NPA. Especially for IPOD2c, the standard deviation is only 0.048 which indicates it is quite stable. And comparing the average of IPOD2c with that of NPA, they are rather close. Since the core electron for Li is only 3, by removing the core electron 2, Li has much less weight in partitioning the covalent charges. As a result, F gets more negative charges. In Figure 2.1(b) the purple line is a relatively close to the green line. Therefore, we can conclude that for LiF, IPOD2c is a better method than NPA considering the advantage of relatively cheaper calculation. Another trend worth mentioning is that with the adding of polarization basis functions, all the methods gradually become less polarized until it comes to the biggest two basis sets: 6-31(2pd,2p) and 6-31(3df,3pd). Afterwards, if we add diffuse functions to our basis set, it becomes polarized again (as the Figure 2.1(b) shows, the lines drop right from g to h), and become stable with diffuse functions.

For Fluorine atom in Sodium Fluoride (NaF) in Figure 2.1(c), similar as the previous two systems HF and LiF, LPA is still the most unstable one. IPOD2, IPOD2b and MPA follow the identical trend of decreasing the polarization gradually with the increasing size of basis sets. Similarly, the decrease bounces back and becomes stable again when we add diffuse functions to the basis sets. When it comes to IPOD2c method, the number of core electrons for Na is 10 which is much bigger than the core of F which is 2. By removing the core electrons, F gets better penalized by sharing more weight in partitioning the covalent contribution which generates a better results compare to NPA method. On the other hand, IPOD2c and IPOD2d stand out for their stability and a relatively reasonable average charge close to that of NPA.

For Fluorine atom in Potassium Fluoride (KF) in Figure 2.1(d), the basis sets 6-31G* and 6-31G** produce the same charge while 6-31++G* and 6-31++G** are equal to each other since double asterisk only adds diffuse functions on hydrogen. Thereafter, for this system, we do not have enough

2.2. Alcohol Systems

Method	CH ₃ OH		C ₂ H ₅ OH	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.160	0.248	-0.023	0.166
LPA	-0.154	0.154	-0.060	0.095
NPA	-0.321	0.033	-0.113	0.026
IPOD1	0.294	0.212	0.146	0.240
IPOD2	-0.933	0.339	-0.396	0.273
IPOD2b	-0.277	0.351	-0.043	0.211
IPOD2c	-0.766	0.368	-0.290	0.272
IPOD2d	-0.177	0.379	0.026	0.225

Table 2.3: Average charges and standard deviations for atom C in CH₃OH and C₂H₅OH for basis sets from 3-21G to 6-31G*.

Method	C ₃ H ₇ OH		C ₄ H ₉ OH	
	Average	Std. Dev.	Average	Std. Dev.
MPA	0.034	0.177	-0.045	0.224
LPA	0.002	0.074	-0.071	0.108
NPA	0.072	0.020	-0.111	0.029
IPOD1	0.024	0.373	0.088	0.412
IPOD2	0.068	0.288	-0.438	0.385
IPOD2b	0.066	0.176	-0.071	0.291
IPOD2c	0.110	0.254	-0.329	0.392
IPOD2d	0.115	0.187	-0.001	0.317

Table 2.4: Average charges and standard deviations for atom C in C₃H₇OH and C₄H₉OH for basis sets from 3-21G to 6-31G*.

basis sets to determine which method works better than the others.

All in all, for fluorine system, our methods of IPOD2c and IPOD2d in general work quite well in terms of accuracy and stability, even better than NPA method in some cases. IPOD2b works better than LPA and MPA but still not polarized enough to yield a good result.

2.2 Alcohol Systems

For Alcohol systems, we mainly concentrate on two atoms: the oxygen and the carbon which attached to that oxygen. For CH₃OH, C₂H₅OH and C₄H₉OH, the -OH are all attached to the carbon from the side. Only for

2.2. Alcohol Systems

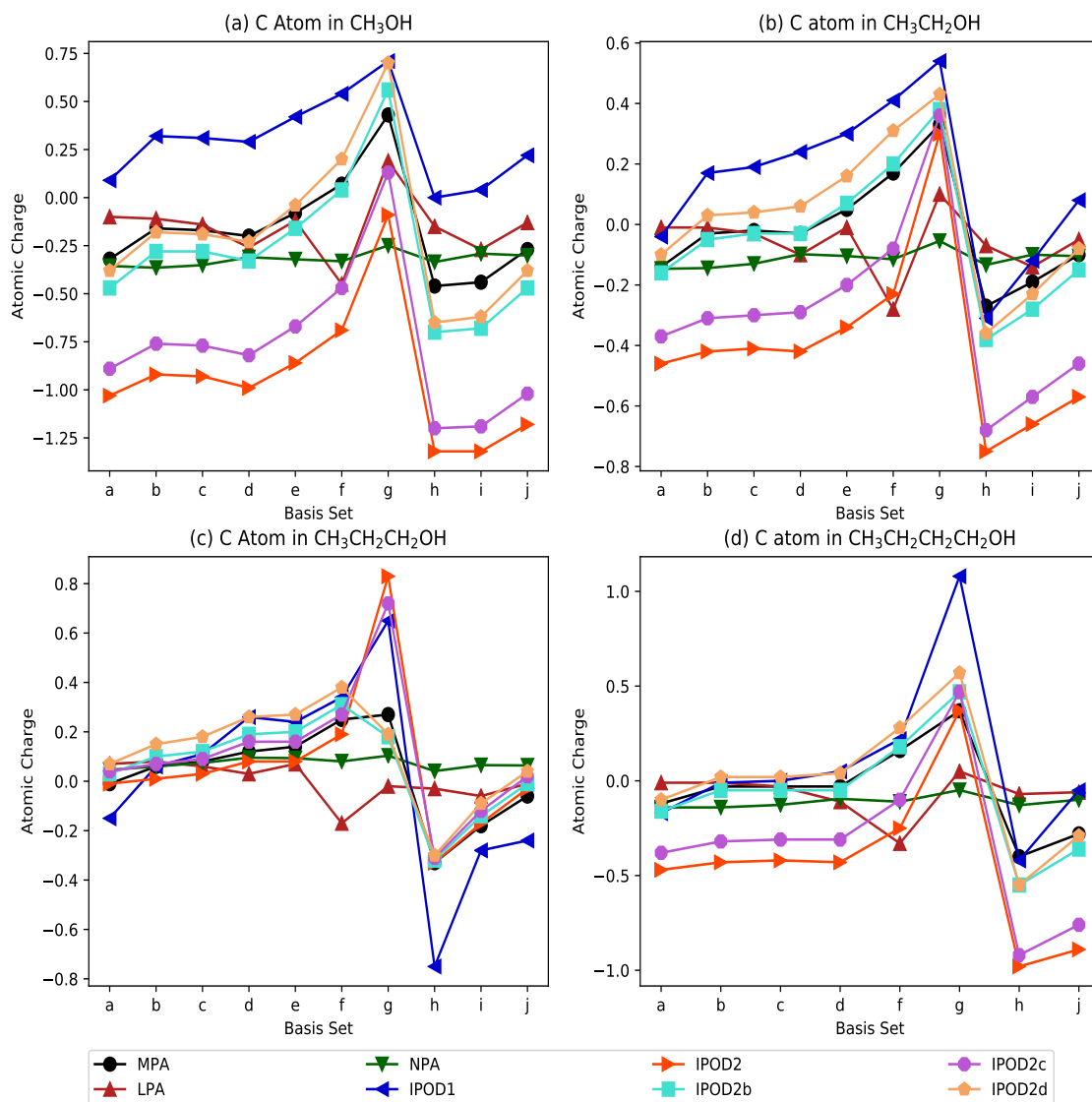


Figure 2.2: Partial charges for C atom in Alcohol compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.2. Alcohol Systems

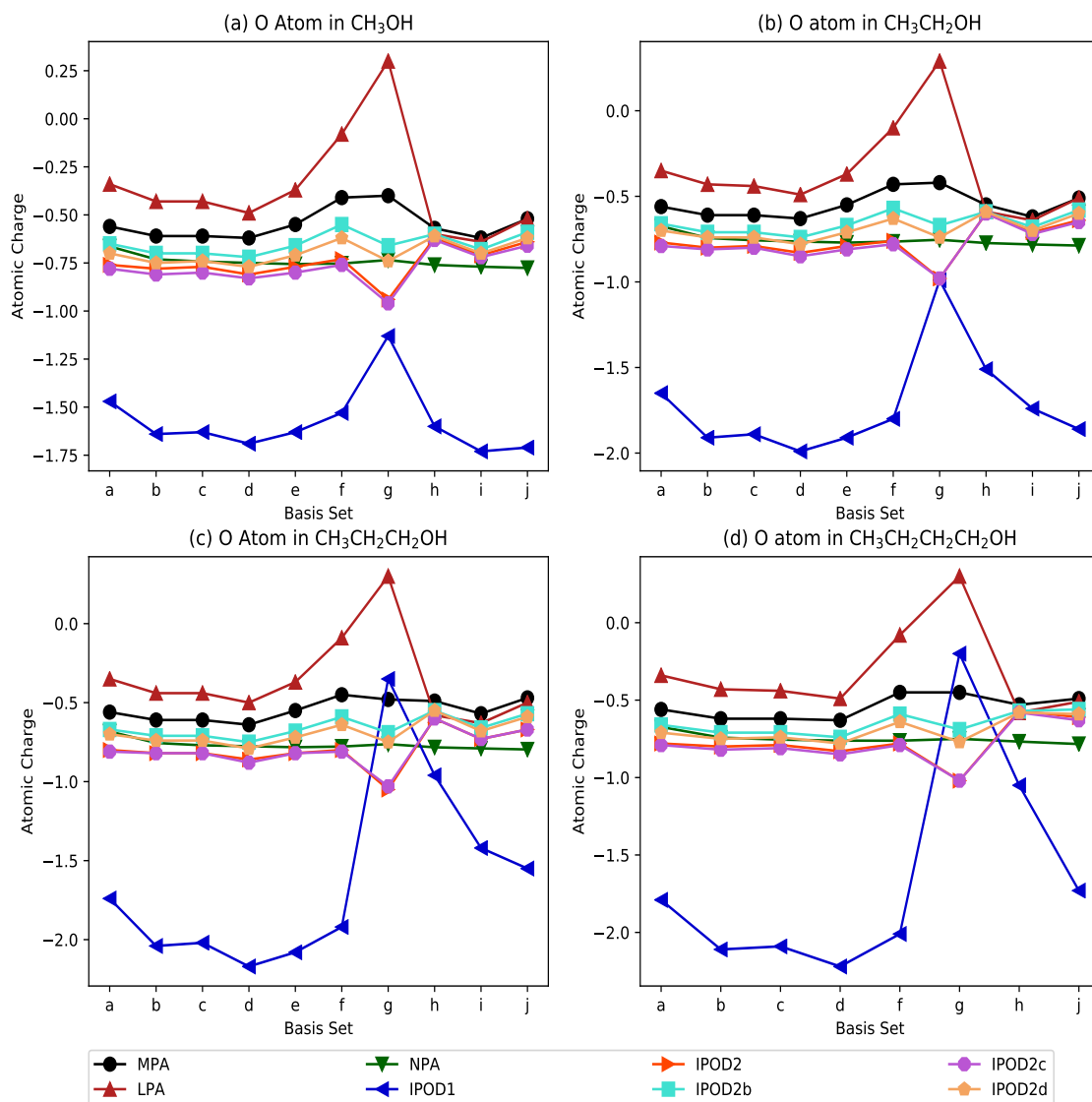


Figure 2.3: Partial charges for O atom in Alcohol compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

C_3H_7OH , the -OH is attached to the carbon in the middle. For the carbon atom in CH_3OH , we can see from Figure 2.3, the only relative stable result is produced by the NPA method, the rest of the methods all have obvious oscillations. Besides, they always share the same trend that with adding more polarization basis sets, the atomic charge becomes more positive until the point we start adding diffuse functions when the atomic charge begin dropping off again and become even more negative than the initial basis set. Although generally all methods are not stable except for NPA (We can see from our deviation Table 2.4), our IPOD series work quite well within a certain range (with basis sets from 3-21G to 6-31G* in Figure 2.3). While IPOD2c is much more negative than NPA which we assume is the most accurate method, IPOD2b and IPOD2d bring it back to the correct level of charge.

For the other two similar systems, C_2H_5OH and C_4H_9OH , they basically have the identical trend as CH_3OH for each method since the -OH are both attached to the carbon on the side. The only system which possess a different trend from the rest of three is C_3H_7OH where the -OH is attached to the carbon in the middle instead of the side. For atom carbon in C_3H_7OH , all the methods are more stable except for certain basis set such as 6-31G(3df, 3pd) and 6-31++G.

Since there are severe oscillation for different basis sets, the basis set averages of all basis sets of each method are not the best way to represent their accuracy. We would rather choose one basis set which is on average the closest to the NPA method and choose that as the final charge we produce from the method. In this way, we can not only have a better result in terms of accuracy, computational wise, we can also save more time and effort. And from the systems we have discussed above, we can crudely pick 6-31G as that basis set we want to use as the standard. Later we will discuss more systems and further prove our statement.

Another atom we care about in the alcohols is the oxygen. Since all the oxygens from the above four systems are attached to hydrogen on the one side and carbon on the other side, their structures are basically all the same. And as a result, All the four systems produce similar trends in terms of the oscillation. Our IPOD series are quite stabilized even compared to NPA.

2.3 Alkene Systems

For Alkene systems, we would choose two kinds of carbon in different environments. The first choice is the carbon at the end which attached to

2.3. Alkene Systems

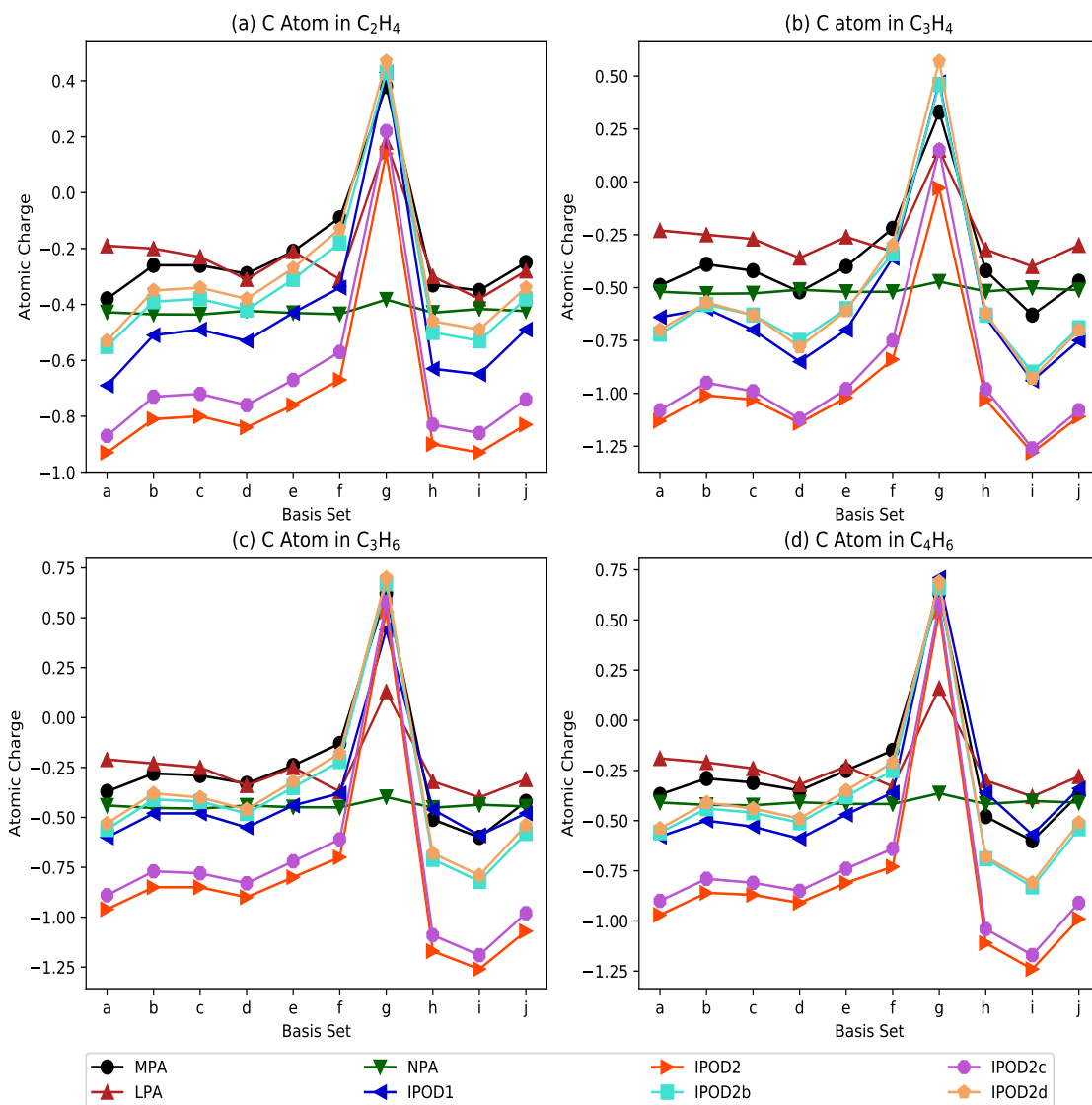


Figure 2.4: Partial charges for Carbon-side atom in Alkenes compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.3. Alkene Systems

Method	C ₂ H ₄		C ₃ H ₄	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.248	0.087	-0.407	0.096
LPA	-0.242	0.050	-0.283	0.046
NPA	-0.431	0.005	-0.521	0.006
IPOD1	-0.498	0.106	0.642	0.148
IPOD2	-0.802	0.079	-1.028	0.099
IPOD2b	-0.372	0.112	-0.603	0.133
IPOD2c	-0.720	0.091	-0.978	0.118
IPOD2d	-0.333	0.120	-0.598	0.150

Table 2.5: Average charges and standard deviations for atom Carbon (side) in C₂H₄ and C₃H₄ for basis sets from 3-21G to 6-31G(2pd, 2p).

Method	C ₃ H ₆		C ₄ H ₆	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.273	0.076	-0.287	0.072
LPA	-0.275	0.059	-0.253	0.053
NPA	-0.448	0.006	-0.416	0.006
IPOD1	-0.448	0.071	0.505	0.077
IPOD2	-0.843	0.081	-0.858	0.075
IPOD2b	-0.407	0.106	-0.433	0.099
IPOD2c	-0.767	0.088	-0.788	0.083
IPOD2d	-0.378	0.110	-0.407	0.106

Table 2.6: Average charges and standard deviations for atom Carbon (side) in C₃H₆ and C₄H₆ for basis sets from 3-21G to 6-31G(2pd, 2p).

Method	C ₂ H ₄		C ₃ H ₄	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.248	0.087	0.223	0.086
LPA	-0.242	0.050	0.015	0.005
NPA	-0.431	0.005	0.068	0.007
IPOD1	-0.498	0.106	0.087	0.211
IPOD2	-0.802	0.079	0.355	0.080
IPOD2b	-0.372	0.112	0.306	0.104
IPOD2c	-0.720	0.091	0.393	0.104
IPOD2d	-0.333	0.120	0.350	0.130

Table 2.7: Average charges and standard deviations for atom Carbon (middle) in C₂H₄ and C₃H₄ for basis sets from 3-21G to 6-31G(2pd, 2p).

2.3. Alkene Systems

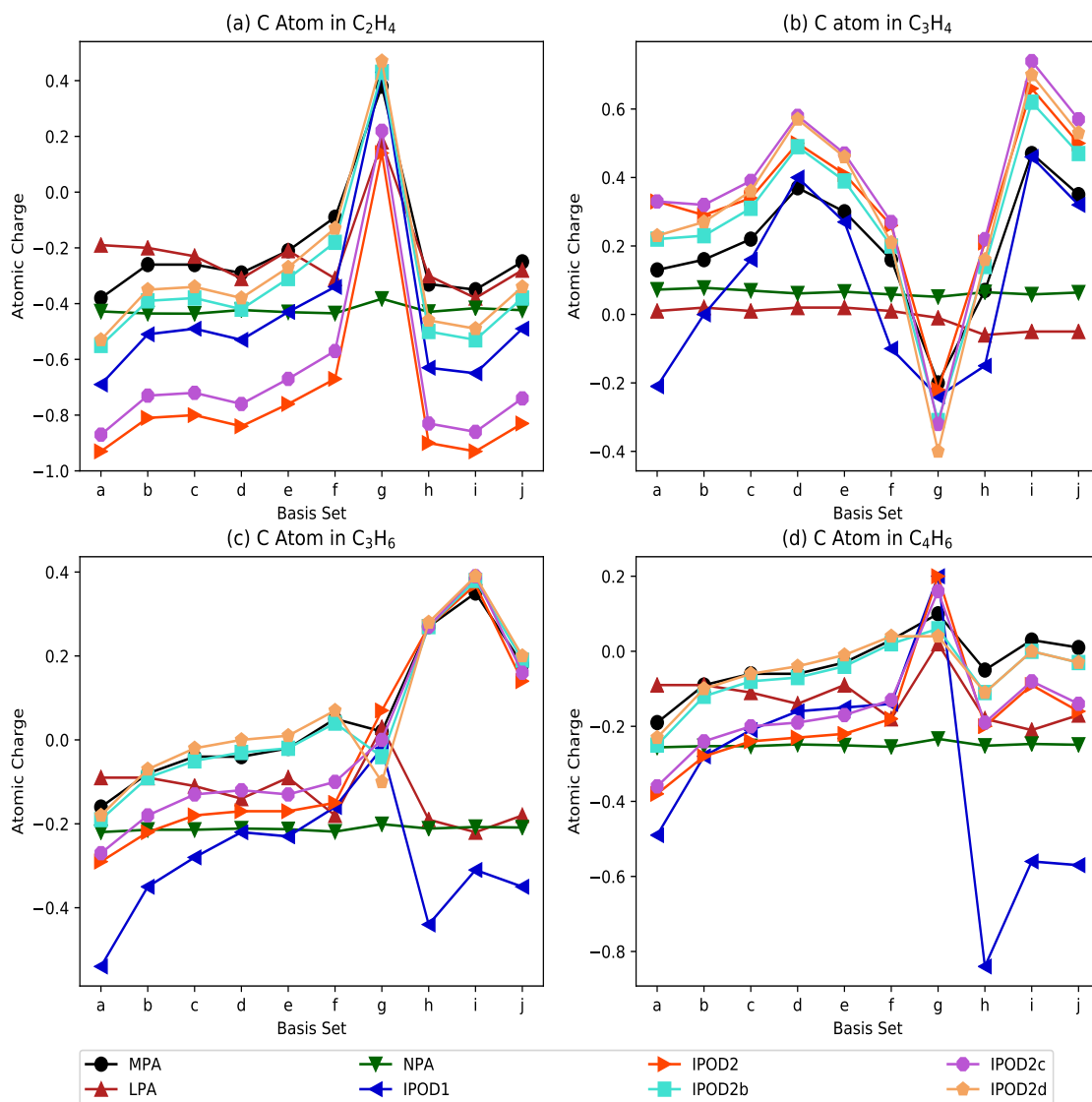


Figure 2.5: Partial charges for C-middle atom in Alkenes compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.3. Alkene Systems

Method	C ₃ H ₆		C ₄ H ₆	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.048	0.063	-0.067	0.066
LPA	-0.117	0.033	-0.117	0.033
NPA	-0.215	0.003	-0.253	0.002
IPOD1	-0.297	0.123	0.248	0.122
IPOD2	-0.197	0.047	-0.255	0.063
IPOD2b	-0.057	0.071	-0.090	0.083
IPOD2c	-0.155	0.057	-0.215	0.073
IPOD2d	-0.032	0.078	-0.067	0.085

Table 2.8: Average charges and standard deviations for atom C (middle) in C₃H₆ and C₄H₆ for basis sets from 3-21G to 6-31G(2pd, 2p).

hydrogen from one side and carbon with double bond on the other side. Another carbon we want to focus on is the carbon in the middle which is attached to two carbons from different sides. We concentrate on these two carbons because we want to evaluate how our method works for more polarized and less polarized atoms even within the same molecule. Hence we plot both the carbons from the end and the carbon from the middle.

Figure 2.4 shows that for different systems, the carbon from the side always share the similar trend. The charges are quite stable for basis sets 3-21G to 6-31G(2pd, 2p). At 6-31G(3df, 3pd), the atomic charge suddenly goes a lot more positive and breaks the flat curve. The numbers drop back when we start adding diffuse functions. Since basis set 6-31G(3df,3pd)), the total average is not very representative, we just calculate the average based on the basis sets from basis set 3-21G to 6-31G(2pd, 2p) and evaluate the standard deviation accordingly. In the range from basis set 3-21G to 6-31G(2pd, 2p), our IPOD2b and IPOD2d methods perform the best. Take C₂H₄ as an example, from the Table 2.5 we can see, the average are -0.372 and -0.333 respectively while the average of NPA is -0.431. Besides, we can observe that for IPOD2b and IPOD2d, the charge produced by 6-31G is quite effective to represent result when compared to NPA.

On the contrary, for the carbon in the middle, the charges produced by our methods are not as good as expected. As the Figure 2.5 shows, for C₃H₆ and C₄H₆, within the range of basis set 3-21G to 6-31G(2pd, 2p), IPOD2c is the best among the three. IPOD2b and IPOD2d become too non-polarized and the charges are almost zero. But at least they are quite stable for these two systems. However, when it comes to C₃H₄, charges fluctuate severely

2.4. Aromatic Systems

Method	Anthracene		Biphenyl	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.138	0.029	-0.128	0.035
LPA	-0.116	0.019	-0.118	0.022
NPA	-0.240	0.003	-0.240	0.003
IPOD1	-0.144	0.074	-0.184	0.069
IPOD2	-0.364	0.043	-0.362	0.047
IPOD2b	-0.186	0.046	-0.178	0.050
IPOD2c	-0.328	0.046	-0.326	0.050
IPOD2d	-0.164	0.048	-0.162	0.053

Table 2.9: Average charges and standard deviations for atom Carbon (side) in Anthracene and Biphenyl for basis sets from 3-21G to 6-31G**.

Method	Naphtalene		Phenanthrene	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.134	0.029	-0.126	0.032
LPA	-0.116	0.019	-0.118	0.022
NPA	-0.239	0.003	-0.237	0.003
IPOD1	-0.190	0.062	-0.134	0.079
IPOD2	-0.358	0.046	-0.354	0.047
IPOD2b	-0.182	0.049	-0.172	0.049
IPOD2c	-0.322	0.050	-0.316	0.051
IPOD2d	-0.160	0.049	-0.154	0.052

Table 2.10: Average charges and standard deviations for atom Carbon (side) in Naphtalene and Phenanthrene for basis sets from 3-21G to 6-31G**.

even it is within the range from basis set 3-21G to 6-31G**. Since the carbon we concern about in C_3H_4 is only attached to two carbons by a double bond and with no hydrogens attached, it is the most non-polarized atom among the four. C_3H_6 and C_4H_6 are both attached to two carbons plus a hydrogen, so they are more non-polarized than C_2H_4 but are still more polarized than C_3H_4 . Accordingly, the quality of result falls in the middle as well. Hence, we can crudely conclude that our methods work better for more polarized atoms even within the same system.

2.4. Aromatic Systems

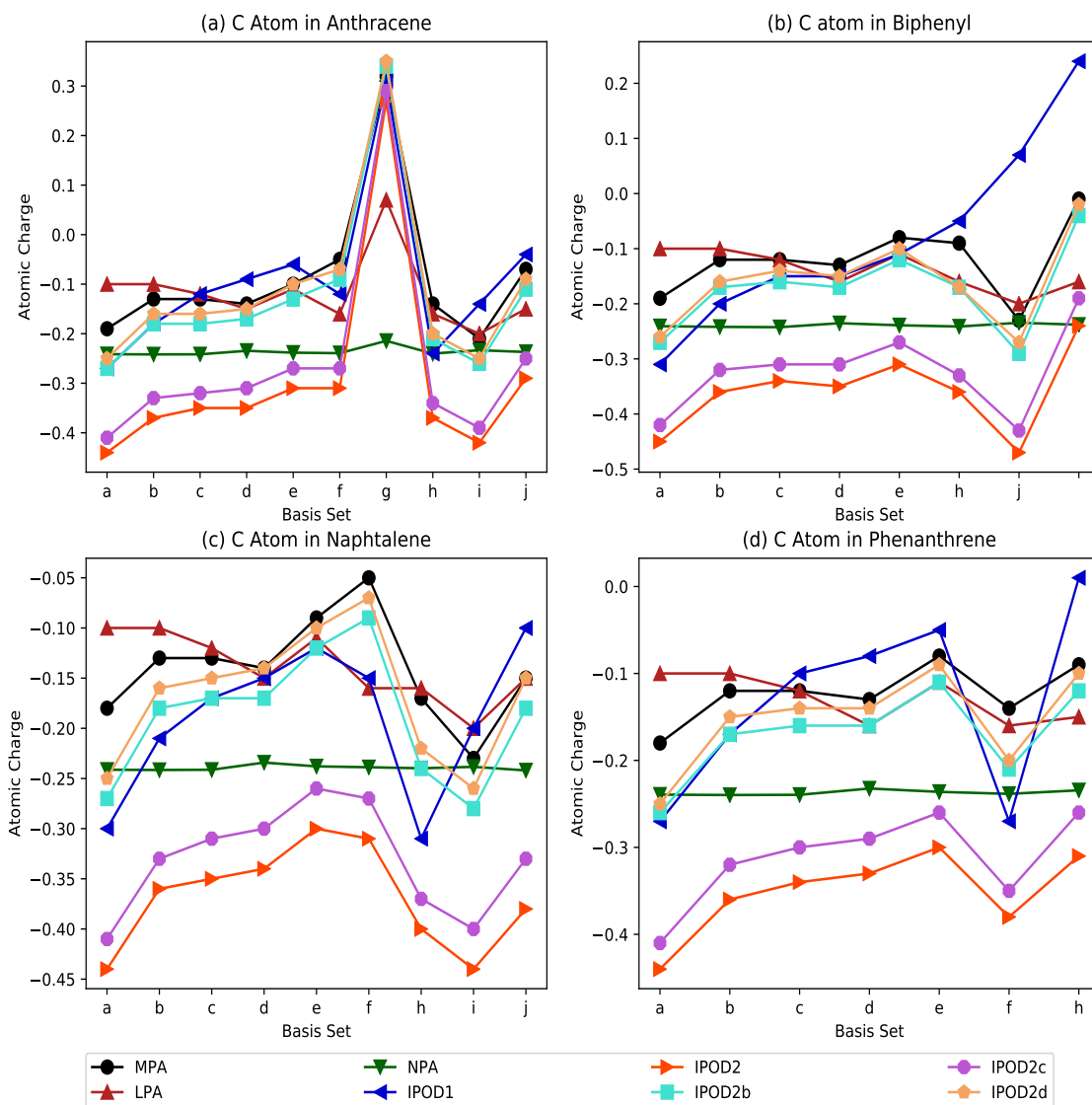


Figure 2.6: Partial charges for C-side atom in Aromatic compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.4. Aromatic Systems

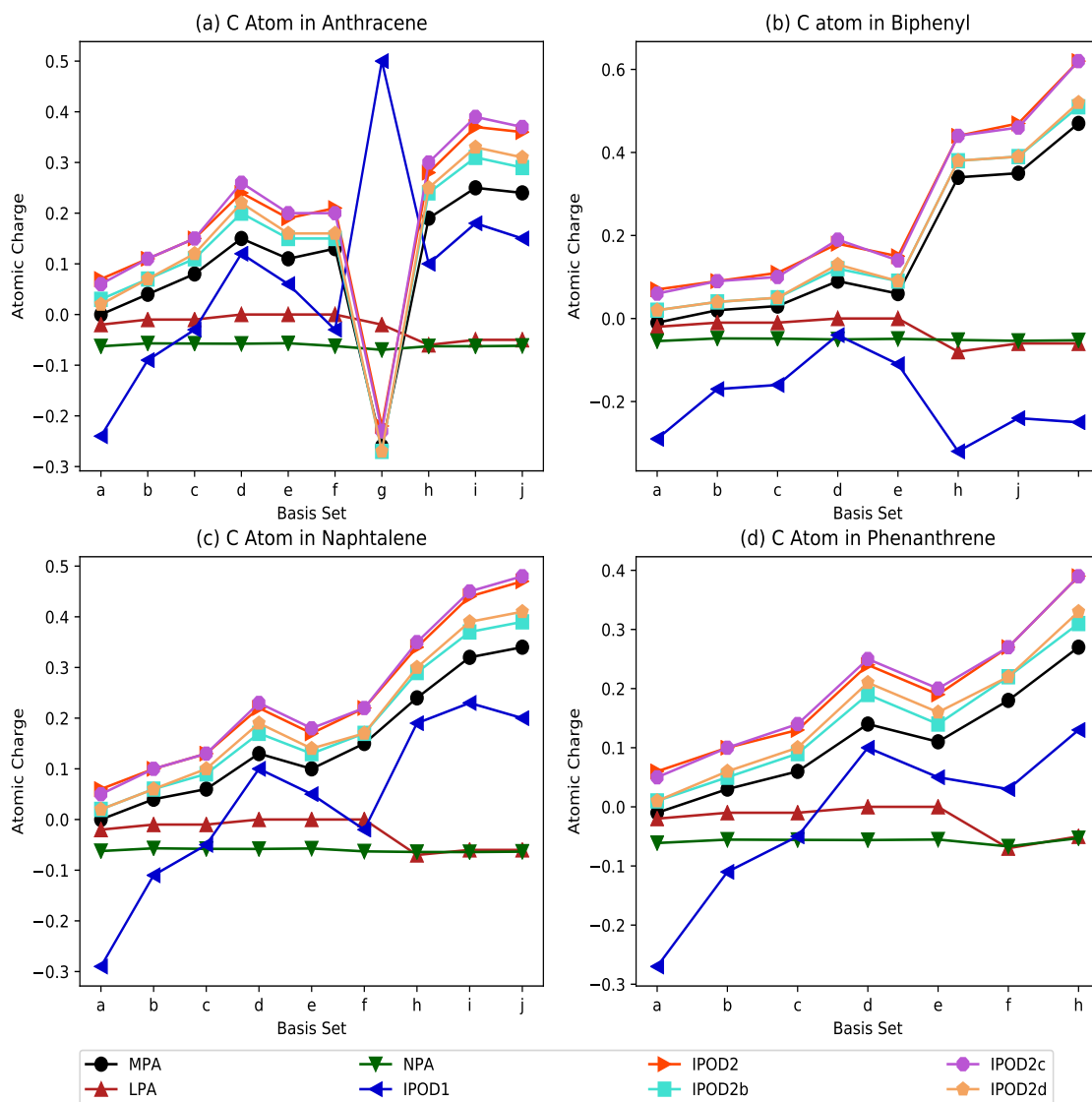


Figure 2.7: Partial charges for C-middle atom in Aromatic compounds. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.4. Aromatic Systems

Method	Anthracene		Biphenyl	
	Average	Std. Dev.	Average	Std. Dev.
MPA	0.076	0.052	0.038	0.034
LPA	-0.008	0.007	-0.008	0.007
NPA	-0.058	0.002	-0.050	0.002
IPOD1	-0.036	0.125	-0.154	0.082
IPOD2	0.152	0.059	0.120	0.040
IPOD2b	0.112	0.059	0.064	0.036
IPOD2c	0.156	0.069	0.116	0.045
IPOD2d	0.118	0.069	0.066	0.039

Table 2.11: Average charges and standard deviations for atom Carbon (middle) in Anthracene and Biphenyl for basis sets from 3-21G to 6-31G**.

Method	Naphthalene		Phenanthrene	
	Average	Std. Dev.	Average	Std. Dev.
MPA	0.066	0.045	0.066	0.054
LPA	-0.008	0.007	-0.008	0.007
NPA	-0.058	0.002	-0.057	0.002
IPOD1	-0.060	0.137	-0.056	0.130
IPOD2	0.136	0.055	0.144	0.064
IPOD2b	0.094	0.052	0.096	0.064
IPOD2c	0.138	0.062	0.148	0.070
IPOD2d	0.102	0.059	0.108	0.071

Table 2.12: Average charges and standard deviations for atom Carbon (middle) in Naphthalene and Phenanthrene for basis sets from 3-21G to 6-31G**.

2.4 Aromatic Systems

So far we already know that our methods have some flaws when applied to non-polarized atoms in a molecule. Here, we will see how our methods work when applied to one of the typical non-polarized system—Aromatic group. For this group, we focus on two carbon atoms, one from the side of the benzene, and another from the joint between the two benzenes. Both of them are quite non-polarized, the joint one is even more non-polarized. The average and standard deviation are shown in Table 2.12.

For the Carbon atom from the side part of the benzenes, as Figure 2.6 shows, in Anthracene and Biphenyl, our IPOD series produce more non-polarized numbers than NPA method. And in terms of stability, the quality of the result is mediocre compared to the deviation of NPA which is only 0.003 for both from Table 2.10. When it comes to the other two systems, Naphtalene and Phenanthrene, the results are pretty much the same. The deviations are around 0.049 and 0.052 respectively which is much bigger than 0.003 from NPA.

Similarly, the results generated by our methods are even worse for the joint carbon between benzenes. The average of the charges become positive from Table 2.12, and the charge of the joint carbon becomes even worse than the performance of the side carbon as we can see from Figure 2.7. Since the joint carbon between benzenes is more non-polarized, our methods cannot handle these cases as well as the more polarized situations. Again, NPA has a very small standard deviation of 0.002 for all of the four systems. It seems the more non-polarized, the more accurate NPA performs.

Based on the above analysis, we find out that our methods are not that reliable for aromatic. It is still meaningful that we know in advance scenarios in which our methods could be applied. All in all, we can generally conclude that our IPOD series do not perform well either for non-polarized atoms or non-polarized systems such as aromatic groups. For these systems we would prefer the NPA method instead. Also for non-polarized atoms, they are usually not very chemically reactive and we do not do much research particular on these atoms.

2.5 Small Inorganic Molecules

The last series of systems we want to analyze by our methods are small inorganic molecules. We choose atom Carbon for CO_2 and CO , and the atom Nitrogen for both HCN and NH_3 to compare how our methods perform. For

2.5. Small Inorganic Molecules

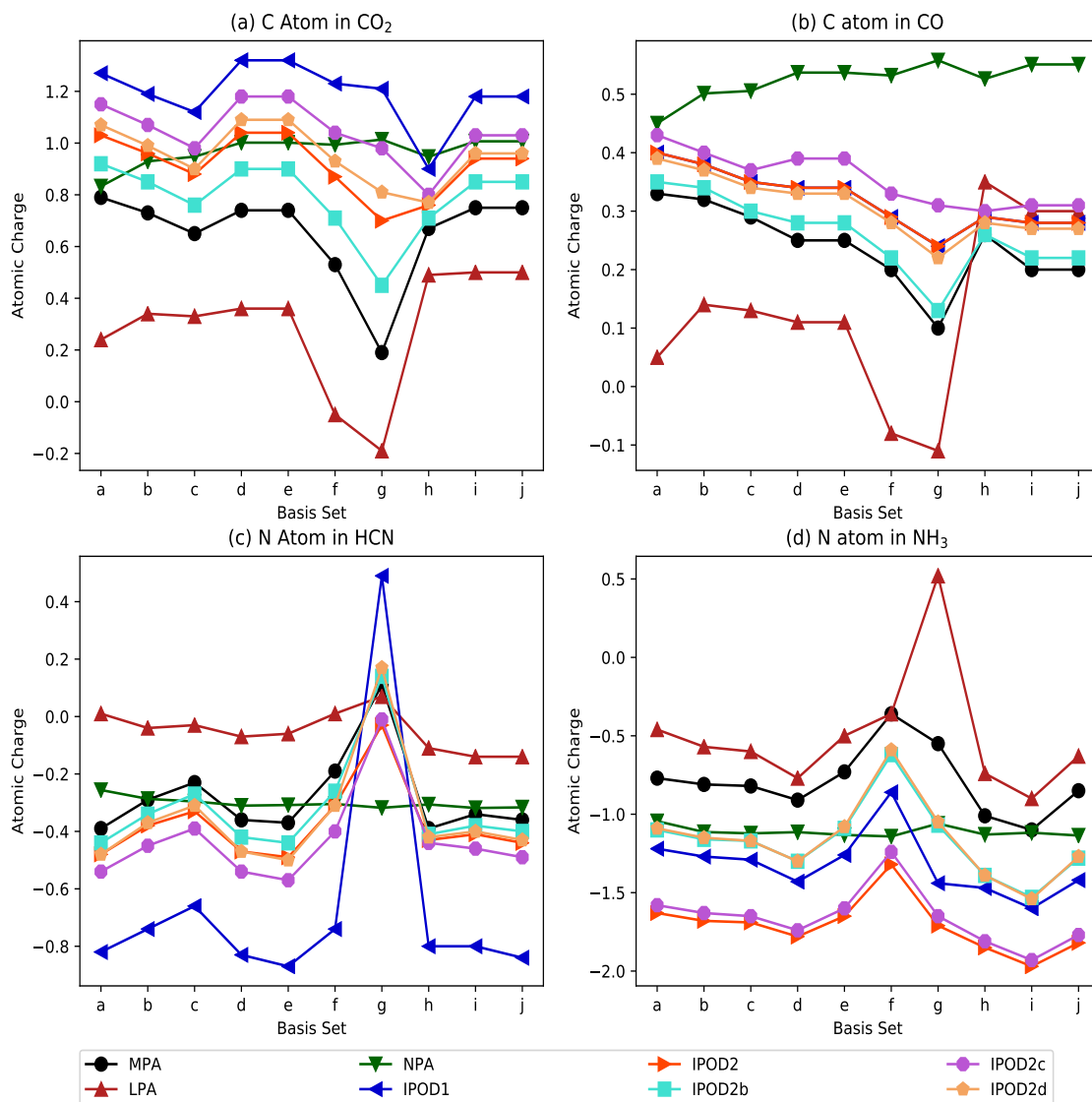


Figure 2.8: Partial charges for atom in small inorganic molecules. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.5. Small Inorganic Molecules

Method	CO ₂		CO	
	Average	Std. Dev.	Average	Std. Dev.
MPA	0.730	0.045	0.288	0.034
LPA	0.326	0.045	0.108	0.031
NPA	0.942	0.062	0.506	0.032
IPOD1	1.244	0.078	0.362	0.024
IPOD2	0.990	0.063	0.362	0.024
IPOD2b	0.866	0.058	0.310	0.030
IPOD2c	1.112	0.077	0.396	0.020
IPOD2d	1.028	0.074	0.352	0.024

Table 2.13: Average charges and standard deviations for atom Carbon in CO₂ and CO for basis sets from 3-21G to 6-31G**.

Method	HCN		NH ₃	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.328	0.059	-0.808	0.060
LPA	-0.038	0.028	-0.580	0.107
NPA	-0.291	0.020	-1.105	0.031
IPOD1	-0.784	0.075	-1.294	0.072
IPOD2	-0.430	0.064	-1.686	0.052
IPOD2b	-0.382	0.067	-1.164	0.075
IPOD2c	-0.498	0.067	-1.640	0.055
IPOD2d	-0.426	0.073	-1.158	0.079

Table 2.14: Average charges and standard deviations for atom Nitrogen in HCN and NH₃ for basis sets from 3-21G to 6-31G**.

2.5. Small Inorganic Molecules

CO₂, we can see from Table 2.13 that within the range from 3-21G to 6-31G*, the average for NPA is 0.942. IPOD2b and IPOD2d have a rather stable result and rather accurate average which are 0.866 and 1.028 respectively. Plus, for IPOD2b, the standard deviation is only 0.058 which is smaller than that of NPA. This indicates IPOD2b is even more stable. Also if we look at the Figure 2.8(a), we can see for basis set 6-31G, IPOD2c and IPOD2d produce a rather close result to NPA which further confirm we can use 6-31G to represent the average in CO₂. All in all, for CO₂, IPOD2 series shows a better performance than the rest of the methods in terms of a more reliable average and less oscillations.

A special case we are particularly interested in here is CO, since in reality the C atom should have negative charge and O have positive charge. However, in Figure 2.8(b), we can see none of these methods produce the correct results. All of the charges produced by our methods are between 0.2 to 0.4 while NPA produces a positive 0.5 charge which is even worse. Therefore, we cannot evaluate our methods for CO, and neither can NPA generate a reasonable result in this system.

Comparing HCN with NH₃, HCN is a more polarized system than NH₃. In HCN, however, we focus on the charge of atom N from C-N bond which is a less polarized bond compared with N-H from NH₃. For C-N bond in HCN system, IPOD2b gives us the closest result to NPA which is -0.382 vs. -0.291. Besides, from Figure 2.8(c), we can see that for basis sets 6-31G, IPOD2b and IPOD2d give charges quite similar to those of NPA. Thereafter, we can conclude that for non-polar bond in a polarized system, using a specific basis set 6-31G for our method to approximate the charges still works. On the other hand, when it comes to N-H bond from NH₃ system, we find from Table 2.16 that IPOD2b and IPOD2d both provide a rather accurate average compared with NPA. And Figure 2.8 also indicates that using 6-31G basis set to approximate the charge for a polarized bond in a non-polar systems works.

For Oxygen atom in N₂O, our IPOD series work relatively well in the range from 3-21G to 6-31G** as shown in Figure 2.9(a), although the number fluctuates a bit when it comes to the basis sets 6-31G* and 6-31G**. All the IPOD methods share the same trend for N₂O system in the range from 3-21G to 6-31G**. And 6-31G can be availed to approximate NPA in this system as well from Figure 2.9(a). For Oxygen atom in HNO system, IPOD methods almost overlap one and another within the range from from 3-21G to 6-31G** as Figure 2.9(b) shows. Also their results are very stable with a standard deviation around 0.13 which is even smaller than that of NPA which is 0.023. There still exists some differences between IPOD methods

2.5. Small Inorganic Molecules

Method	N ₂ O		HNO	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.302	0.058	-0.262	0.010
LPA	-0.204	0.026	-0.088	0.015
NPA	-0.273	0.038	-0.195	0.023
IPOD1	-0.404	0.128	-0.596	0.017
IPOD2	-0.324	0.063	-0.258	0.013
IPOD2b	-0.314	0.063	-0.256	0.014
IPOD2c	-0.332	0.065	-0.260	0.013
IPOD2d	-0.332	0.065	-0.268	0.013

Table 2.15: Average charges and standard deviations for atom Nitrogen in N₂O and HNO for basis sets from from 3-21G to 6-31G**.

Method	H ₂ O		HOOH	
	Average	Std. Dev.	Average	Std. Dev.
MPA	-0.702	0.056	-0.808	0.060
LPA	-0.528	0.093	-0.580	0.170
NPA	-0.918	0.040	-1.105	0.031
IPOD1	-1.062	0.053	-1.294	0.072
IPOD2	-1.166	0.041	-1.686	0.052
IPOD2b	-0.914	0.061	-1.164	0.075
IPOD2c	-1.148	0.041	-1.640	0.055
IPOD2d	-0.952	0.057	-1.158	0.079

Table 2.16: Average charges and standard deviations for atom Nitrogen in H₂O and HOOH for basis sets from 3-21G to 6-31G**.

and NPA since we can see our methods are slightly below NPA which indicates that IPOD gives us a more polarized result than we expect and so does the 6-31G basis set.

For Oxygen atoms in H₂O and HOOH, Figure 2.9(c) and (d) show the same trend for both. The reason for that is both Oxygen come from the O-H bond. IPOD2b and IPOD2d both guarantee good average charges compared with NPA. The fluctuation is reasonable within the range from from 3-21G to 6-31G** as well. As with the previous systems, the basis 6-31G can produce a reasonable result to represent our methods in these two systems as we initially expected.

2.5. Small Inorganic Molecules

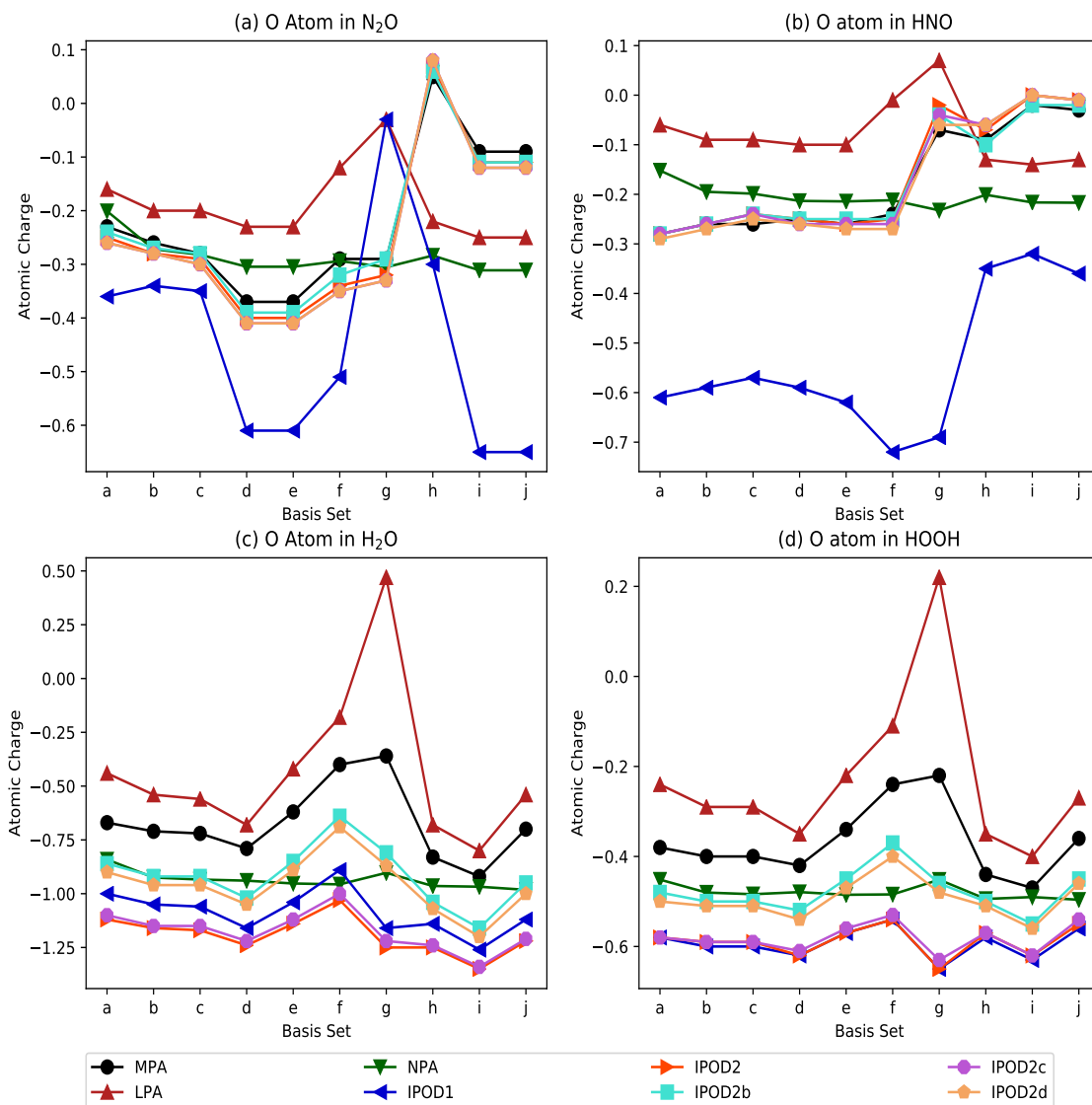


Figure 2.9: Partial charges for Oxygen atom in small inorganic molecules. The basis sets are as follows: (a, 3-21G), (b, 4-31G), (c, 6-31G), (d, 6-31G*), (e, 6-31G**), (f, 6-31G(2pd, 2p)), (g, 6-31G(3df, 3pd)), (h, 6-31++G), (i, 6-31++G*), (j, 6-31++G**)

2.6 Correlation between Different Methods with NPA

We now want to compare how each method performs based on the correlation plot with NPA using all the data we produced from all the systems and basis sets. In the Figure 2.11 and Figure 2.12, we plot the charges produced by the NPA method along the x-axis, and y-axis represents the charges we get from other methods. We can compare how close each method is related to NPA method by comparing the slope value with 1. Also, the r^2 values indicates how reliable is the data shown in the plot. The closer the slope and r^2 are to 1, the better the method performs.

From Figure 2.11, we can read the slope in the NPA and MPA plots are 1.36 and 1.89 respectively which are both much larger than 1. It means both MPA and LPA methods polarize the charges more than they should. On the contrary, the slope for IPOD1 is only 0.67 and it indicates that IPOD1 method is too non-polarized. When it comes to our IPOD2 series from Figure 2.11, the slope shows clearly that our IPOD2b and IPOD2d stand out compared to the rest of two. IPOD2b has a value of 1.056 while IPOD2d has a number even closer to one at 1.007. In terms of reliability, the plots are quite reasonable since r^2 is around 0.95 for both. It is then safe to conclude that IPOD2d is the best method in terms of the correlation to NPA.

2.7 The Best Basis Set within IPOD2d Method

Since we already figured out that IPOD2d is the best method among the all except for NPA, we now want to find the best basis set which can represent IPOD2d method. Figure 2.12 shows how close IPOD2d is to NPA. in terms of four different basis sets, how close IPOD2d is related to NPA method. We choose these four basis sets particularly based on what we previously discussed for different groups of systems. Comparing these four slopes, we can tell that 4-31G and 6-31G are better than the rest of two with slope of 0.987 and 0.990 respectively. 6-31G is the best basis set that we select to represent IPOD2d. This corresponds to what we have analysed before by roughly looking at each graph in different systems which further proves our conclusion is correct. Choosing the best basis set to represent a method means a lot for our practical calculations since our methods are computationally much cheaper than NPA method while still providing a similar result as NPA does.

2.7. The Best Basis Set within IPOD2d Method

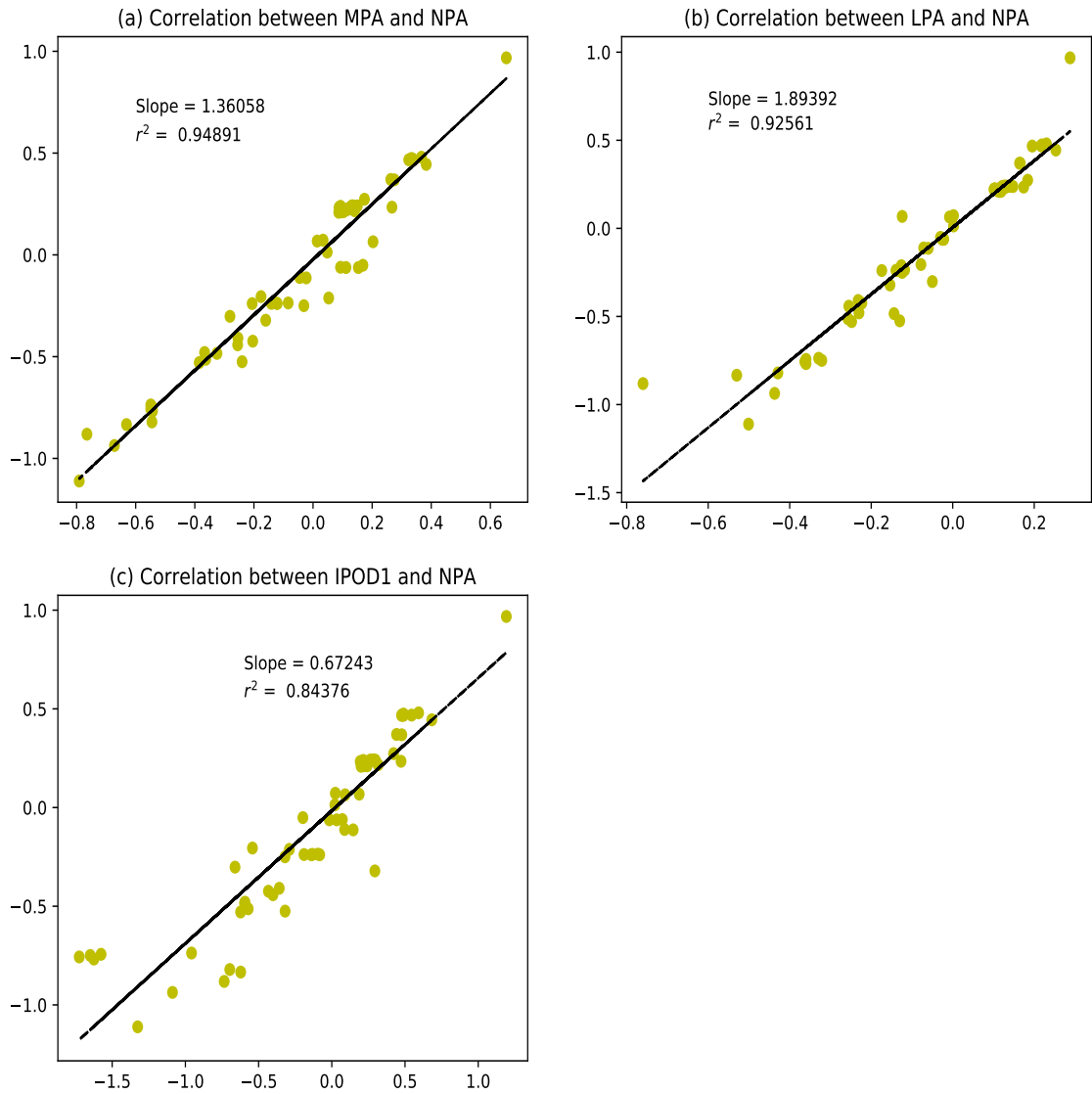


Figure 2.10: Correlation between different methods with NPA

2.7. The Best Basis Set within IPOD2d Method

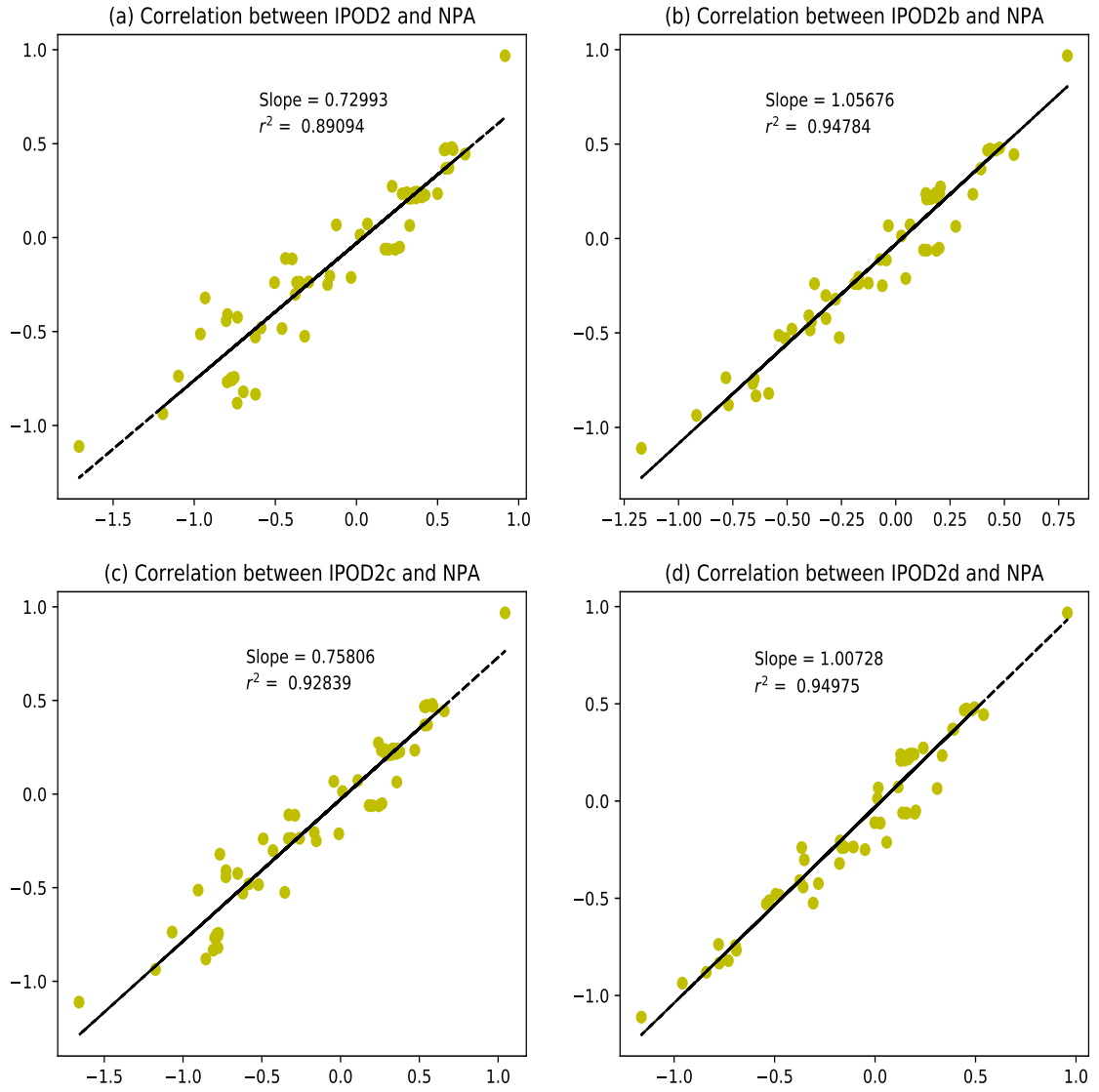


Figure 2.11: Correlation between different methods with NPA

2.7. The Best Basis Set within IPOD2d Method

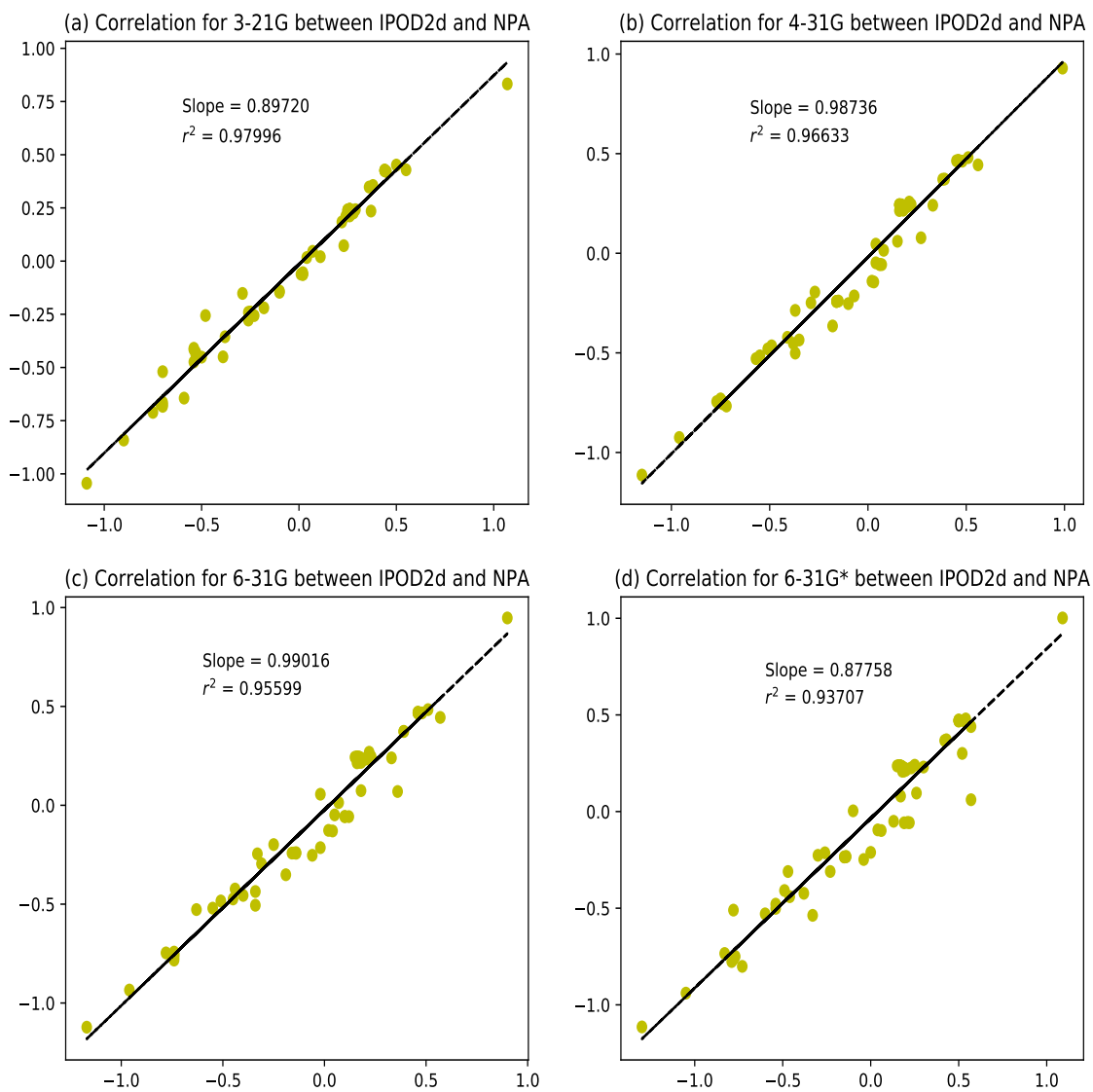


Figure 2.12: Correlation for different basis sets in IPOD2d with the same basis sets in NPA

Chapter 3

Summary

In this thesis, we evaluated the IPOD series of methods to determine partial atomic charges. In order to evaluate our methods in a more systematic way, we put all the molecules into different categories according to their functional groups. As a result, we have more polarized groups such as Fluorine and Alcohol, less polarized ones such as Aromatic, and systems that fall in the middle such as Alkenes and small inorganic groups.

For the more polarized systems, we focus on the atoms from the most polarized bonds within each molecule. We can conclude that IPOD2d performs the best in terms of stability and accuracy when using the NPA method. And since the O-H bond is more polarized than the C-H bond, the result for charges of Oxygen atoms performs even better than that of Carbon atoms.

For non-polarized systems, we mainly concentrate on two kinds of carbon atoms, one from the side of benzenes and another from the joint of two benzenes. The major difference between the two is that the middle one is even less polarized. Our results indicate that in general our method does not work well for non-polarized systems, and the less polarized the bond is, the worse our method performs. Therefore, in the future, if we want to apply our method to a new system, we should rather use NPA method instead of IPOD series if the system is non-polarized.

Besides the above two general cases, there are still many systems that belong neither to extremely polarized category nor to the non-polarized one. We use Alkenes and small inorganic molecules as examples to analyse how our method performs on different types of bonds. The plots show IPOD2d generally gives a rather reasonable result for most cases in these two categories of systems. Plus, when the atom is from a more polarized bond, the result is more stable and more accurate. Hence, we can conclude that even in the same system, IPOD series produce a better result for atoms from more polarized bonds.

At the last part of our discussion, we analysed the correlation between different methods and NPA in order to figure out the best IPOD method. By comparing the slope value from each plot, we can see that IPOD2d has a slope of 1.007 which is the closest to 1. And this result further proves our

previous idea that IPOD2d is the best out of the IPOD series of methods. On top of finding the best method, we still want to find which basis set can provide the most accurate charge which is closer to NPA. By plotting the correlation between the best four basis sets of IPOD2d vs. NPA, we can finally conclude that 6-31G is the best basis set. In conclusion, by using the IPOD2d method with the 6-31G basis set, we can obtain charges very close to those of NPA but at a much lower computational cost.

Bibliography

- [1] RS Mulliken. Electronic population analysis on LCAO–MO molecular wave functions. I. *The Journal of Chemical Physics*, 23:1833, 1955.
- [2] D Cremer and E Kraka. Chemical bonds without bonding electron density does the difference electron-density analysis suffice for a description of the chemical bond? *Angewandte Chemie International Edition*, 23:627, 1984.
- [3] RS Mulliken. Electronic population analysis on LCAO–MO molecular wave functions. II. overlap populations, bond orders, and covalent bond energies. *The Journal of Chemical Physics*, 36:3428, 1962.
- [4] RFW Bader, TT Nguyen-Dang, and Y Tal. A topological theory of molecular structure. *Reports on Progress in Physics*, 44:893, 1981.
- [5] P Coppens and MB Hall. *Electron distributions and the chemical bond*. Springer Science & Business Media, 2012.
- [6] MA Blanco, A Martín Pendás, and E Francisco. Interacting quantum atoms: a correlated energy decomposition scheme based on the quantum theory of atoms in molecules. *Journal of Chemical Theory and Computation*, 1:1096, 2005.
- [7] RS Mulliken. Electronic population analysis on LCAO–MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. *The Journal of Chemical Physics*, 23:1841, 1955.
- [8] JA Pople, M Head-Gordon, DJ Fox, K Raghavachari, and LA Curtiss. Gaussian-1 theory: A general procedure for prediction of molecular energies. *The Journal of Chemical Physics*, 90:5622, 1989.
- [9] JA Pople, PMW Gill, and BG Johnson. Kohn-Sham density-functional theory within a finite basis set. *Chemical Physics Letters*, 199:557, 1992.

Bibliography

- [10] SN Gariseb. *A computational study of the structure, bonding, and thermochemical properties of primary Ozonides derived from substituted Phenol and Thiophenol*. PhD thesis, 2013.
- [11] A Szebo and NS Ostlund. *Modern Quantum Chemistry*. Dover Publications, 1982.
- [12] RBS Krishnan, JS Binkley, R Seeger, and JA Pople. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics*, 72:650, 1980.
- [13] WJ Hehre, RF Stewart, and JA Pople. self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *The Journal of Chemical Physics*, 51:2657, 1969.
- [14] TH Dunning Jr. Gaussian basis functions for use in molecular calculations. I. Contraction of (9s5p) atomic basis sets for the first-row atoms. *The Journal of Chemical Physics*, 53:2823, 1970.
- [15] A Schäfer, H Horn, and R Ahlrichs. Fully optimized contracted gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics*, 97:2571, 1992.
- [16] JS Binkley, JA Pople, and WJ Hehre. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *Journal of the American Chemical Society*, 102:939, 1980.
- [17] MM Francl, WJ Pietro, WJ Hehre, JS Binkley, MS Gordon, DJ DeFrees, and JA Pople. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics*, 77:3654, 1982.
- [18] WJ Pietro, MM Francl, WJ Hehre, DJ DeFrees, JA Pople, and JS Binkley. Self-consistent molecular orbital methods. 24. Supplemented small split-valence basis sets for second-row elements. *Journal of the American Chemical Society*, 104:5039, 1982.
- [19] T Clark, J Chandrasekhar, GW Spitznagel, and PVR Schleyer. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *Journal of Computational Chemistry*, 4:294, 1983.

Bibliography

- [20] GK Chan and M Head-Gordon. Exact solution (within a triple-zeta, double polarization basis set) of the electronic Schrödinger equation for water. *The Journal of Chemical Physics*, 118:8551, 2003.
- [21] F Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, 2017.
- [22] P Löwdin. Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Physical Review*, 97:1474, 1955.
- [23] PR Taylor. Quantum theory of molecular electronic structure. *Chemistry*, 185:285, 1963.
- [24] W Meyer. Theory of selfconsistent electron pairs. An iterative method for correlated manyelectron wavefunctions. *The Journal of Chemical Physics*, 64:2901, 1976.
- [25] HJ Silverstone and O Sinanoğlu. Manyelectron theory of nonclosedshell atoms and molecules. I. Orbital wavefunction and perturbation theory. *The Journal of Chemical Physics*, 44:1899, 1966.
- [26] JC Slater. A simplification of the Hartree-Fock method. *Physical Review*, 81:385, 1951.
- [27] AJC Varandas and J Brandão. A double many-body expansion of molecular potential energy functions. *Molecular Physics*, 57:387, 1986.
- [28] CCJ Roothaan. New developments in molecular orbital theory. *Review of Modern Physics*, 23:69, 1951.
- [29] H Ehrenreich and MH Cohen. Self-consistent field approach to the many-electron problem. *Physical Review*, 115:786, 1959.
- [30] TL Gilbert. Hohenberg-Kohn theorem for nonlocal external potentials. *Physical Review B*, 12:2111, 1975.
- [31] E Runge and EKV Gross. Density-functional theory for time-dependent systems. *Physical Review Letters*, 52:997, 1984.
- [32] M Levy. On the simple constrained-search reformulation of the Hohenberg-Kohn theorem to include degeneracies and more (1964–1979). *International Journal of Quantum Chemistry*, 110:3140, 2010.

Bibliography

- [33] W Koch and MC Holthausen. *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.
- [34] TA Wesolowski and J Weber. Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules. *Chemical Physics Letters*, 248:71, 1996.
- [35] ME Casida and TA Wesolowski. Generalization of the Kohn-Sham equations with constrained electron density formalism and its time-dependent response theory formulation. *International Journal of Quantum Chemistry*, 96:577, 2004.
- [36] YK Chen and YA Wang. Population analyses based on ionic partition of overlap distributions. *Progress in Theoretical Chemistry and Physics*, 30:65, 2017.
- [37] R Carbó-Dorca and P Bultinck. Quantum mechanical basis for Mulliken population analysis. *Journal of Mathematical Chemistry*, 36:231, 2004.
- [38] AE Clark, JL Sonnenberg, PJ Hay, and RL Martin. Density and wave function analysis of actinide complexes: What can fuzzy atom, atoms-in-molecules, Mulliken, Lwdin, and natural population analysis tell us? *The Journal of chemical physics*, 121:2563, 2004.
- [39] AE Reed, RB Weinstock, and F Weinhold. Natural population analysis. *The Journal of Chemical Physics*, 83:735, 1985.