

DENSITY FUNCTIONAL THEORY AND TIME-DEPENDENT  
DFT MODELING OF ORGANIC PHOTOVOLTAIC MATERIALS

By

SHUO DAI

Bachelor of Science in Physics  
Nanjing University  
Nanjing, China  
2004

Master of Science in Physics  
Oklahoma State University  
Stillwater, Oklahoma  
2013

Submitted to the Faculty of the  
Graduate College of  
Oklahoma State University  
in partial fulfillment of  
the requirements for  
the Degree of  
DOCTOR OF PHILOSOPHY  
December, 2017

DENSITY FUNCTIONAL THEORY AND TIME-DEPENDENT  
DFT MODELING OF ORGANIC PHOTOVOLTAIC MATERIALS

Dissertation Approved:

Dr. Mario F. Borunda (Dissertation Advisor)

---

Dr. John W. Mintmire

---

Dr. Albert T. Rosenberger

---

Dr. Donghua Zhou

---

Dr. Toby L. Nelson

---

## ACKNOWLEDGMENTS

This dissertation reflects the status of several of my current studies. Although my dissertation writing ends here, my exploration in science is still ongoing. But I would like to take the chance to thank my advisor, my parents, my wife and child, my friends and everyone who accompanied me in the journey of my life so far.

My advisor, Dr. Borunda, is the best person I have ever met. In him, I have seen the combination of every virtue I knew. He is an expert in scientific research in his field while also being talented in each new field he explores. Having considerable insight in science, he nevertheless is a humble person in front of the nature. Although great scientists have been common in history, not many of them were successful speakers who were able to convey excitements about science to the public, not to say being scrupulous in oneself's personal behavior. However, Dr. Borunda is excellent in public speech and can help audience visualize the most complicated idea in a easy way. I still remember his presentation on calculating Raman spectroscopy of graphene. I knew nothing before his presentation but became confident to talk to others about this topic after his speech. Embarrassingly enough, although I learned a lot of useful tips in his presentations, I am still a bad speaker with the hope that one day I can "tunnel" through the barrier to see the same sights as he did. In addition to the communication skills, I learned enormous scientific knowledge and skills from him. However, compared to his personal virtue, everything else dims. I would list his moral excellence as the very first thing I learned from him and am proud of, because, above all, I want to learn how to be as decent a man as he is.

My dissertation committee consists of amazing scientists who guided me through my Ph.D pursuit. I want to express my gratitudes to Dr. Albert Rosenberger, Dr. John Mintmire, Dr. Donghua Zhou and Dr. Toby Nelson. Some of the professors were also in my master committee. So I also would like to thank you again for bearing with me these many years. I also want to thank Susan Cantrell, Tamra Ringer, Alisha Leach and Elizabeth Bridenstine for your help. I can't forget Melissa's emails that are already a unique experience I had during this many years in physics department. She

is so “unfortunate” to have me as her employee. I also want to express my gratefulness to my friends and group members: Paul Smith, Kyle Stoltz, Brian Leininger, Charith Desilva, Mary Catlett, Kazsa Fahrenthold, Logan Thompson, Limu Ke, Tao Tang, Sreekul Rajagopal, Hem Moktan and many others. The pleasant time I spent with all of you will always remind me of these exciting years of education.

Beyond thanks, I felt lucky to have in my life my parents and my wife, who supported me during these many years. English language couldn’t express my gratitude so I would like to save those in the code of GB18030. One of the meanings of my name could refer to “the master’s degree”. Many people told me that it would be difficult to gain an even higher degree because of the destiny reflected in my name. However, everyday is one step closer to the future and I didn’t retreat.

Finally, I want to thank Cowboy Cluster HPCC for providing computation power to make my research possible and I also appreciate the Physics Department at OSU for the support.

---

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: Shuo Dai

Date of Degree: December, 2017

Title of Study: DENSITY FUNCTIONAL THEORY AND TIME-DEPENDENT  
DFT MODELING OF ORGANIC PHOTOVOLTAIC MATERIALS

Major Field: Physics

This thesis has aimed to apply Density Functional Theory (DFT) and Time-dependent DFT (TD-DFT) method to design new organic material for small molecule organic photovoltaic (OPV) and molecules for tandem structure solar cells. The sources of virtual molecules are both from a focus library, consisting of bio-inspired eumelanin based molecules, and from a diverse library, which is the Harvard Clean Energy Project Database (CEPDB).

Chapter 2 validated our method. We first tried to find the density functionals that have best performance among all categories of functionals for HOMO, LUMO and gap calculations. For HOMO only, range separated functional, M06-2x, PBE0 provided the top results; for Gap prediction, PBE, TPSS and HSE06 provided the top results. Then in chapter 3, armed with the selected functionals, we carried out DFT/TD-DFT calculations to study molecules built from a bio-inspired melanin family. After studies on the structure and property relationship of the designed push-pull donors, we further suggested many virtual molecules for small molecule photovoltaic. Finally, in chapter 4, we designed a descriptor, correlation ratio, to describe the alignment between absorption spectra and solar spectrum. With DFT/TD-DFT calculations and correlation ratio, we then able to search for ideal molecules for tandem solar cells, whose spectra would be complimentary so that molecules in the tandem solar cell can achieve a full absorption of the sunlight.

The topic of discovering new material for renewable energy is essential to solve the major challenge in energy. The question is how to discover new materials from the huge chemical space efficiently. One of the approaches is associated with the progress in computational modeling, e.g. machine learning and the progress in hard ware development. In chapter 2 of this dissertation, I used k-fold cross validation to justify the validity of the model. This technique is one of the fundamental algorithms in machine learning. Additionally, there is a great progress in hardware development in recent years, which made the quantum calculation of large molecules done in chapter 3 and a large scale high-throughput screening done in chapter 4 possible.

ADVISOR'S APPROVAL: \_\_\_\_\_

## TABLE OF CONTENTS

Chapter	Page
<b>1 Background and Introductions</b>	<b>1</b>
1.1 Motivations . . . . .	1
1.2 Designing OPV cell's structures according to its operation principle .	2
1.2.1 Simple structure . . . . .	2
1.2.2 Heterojunction structure . . . . .	2
1.3 Improving PCE by investigating the donor materials . . . . .	4
1.3.1 Short-circuit current . . . . .	5
1.3.2 Open-circuit Voltage and the tandem structure . . . . .	6
1.3.3 The Scharber model . . . . .	7
1.4 Outline of the dissertation . . . . .	10
<b>2 Benchmarking DFT/TD-DFT Functionals for Frontier Orbital En- ergies Predictions on Organic Photovoltaic Molecules</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Computational method . . . . .	14
2.3 Results and Analysis . . . . .	23
2.3.1 Calculations from DFT/TD-DFT functionals . . . . .	24
2.3.2 Results on Gap benchmark (Pure, Global Hybrid and RSH functionals) . . . . .	29
2.3.3 Results on LUMO benchmark (Pure, Global Hybrid and RSH functionals) . . . . .	31
2.3.4 Predictions and corrections from linear regression models . . .	34
2.4 Conclusion . . . . .	44
<b>3 DFT/TD-DFT Investigations on Bio-inspired Melanin Molecules for Organic Photovoltaic Applications</b>	<b>45</b>
3.1 Introduction . . . . .	45
3.1.1 Computational methods . . . . .	47

3.1.2	Results and discussions . . . . .	49
3.2	Conclusion . . . . .	55
<b>4</b>	<b>High Throughput Screening Using Correlation Ratio for Tandem Solar Cells Design</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Computational Method . . . . .	59
4.2.1	The ground state geometry calculations using Gaussian09. . . . .	60
4.2.2	The excited state calculation using octopus. . . . .	60
4.2.3	The spectrum calculation using compressed sensing technique. . . . .	62
4.2.4	The spectra correlation ratio. . . . .	62
4.3	Results and Discussions . . . . .	64
4.3.1	The performance of the top ones in one region . . . . .	64
4.3.2	Aiming at a better alignment with the solar spectrum (across the regions) . . . . .	65
4.3.3	Aiming at screening a molecule with special requirement on its spectrum range. . . . .	69
4.4	Conclusion . . . . .	69
<b>5</b>	<b>Conclusions and Outlook</b>	<b>71</b>
5.1	Summary of current research . . . . .	71
5.2	Prospective for future research . . . . .	72
<b>A</b>	<b>A Short Introduction to DFT/TD-DFT and Material Informatics</b>	<b>88</b>
A.0.1	Molecular modeling using DFT . . . . .	88
A.0.2	Molecular modeling using TD-DFT . . . . .	101
A.0.3	Virtual molecule screening and material informatics . . . . .	106
<b>B</b>	<b>Parameters from Linear Regressions in the Chapter 2</b>	<b>108</b>

## LIST OF TABLES

Table		Page
2.1	The chemical formula, SMILES and references for molecules set. . . .	18
2.2	Functionals examined in this study. If two HF percentiles are presented, the first one is for the short range and the second one is for the long range. . . . .	23
4.1	The division scheme of the 26 regions of solar spectrum. The values of correlation ratio were first calculated in each of these regions. Depending on the different ranking tasks as discussed in the “Results” section, the correlation ratio could be summed up to represent certain performances in the different spectrum ranges as labeled in Figure 4.3.	63
4.2	Top performance molecules ID (from the CEPDB database) in each regions. The red character means the molecule shown up in more than one region. And the blue bars in the cell show their values of the spectrum correlation ratio. . . . .	65
A.1	The time complexity of general DFT, MP2 and CCSD codes. . . . .	91
B.1	Linear regression formula for DFT functionals on HOMO and on Gap.	109



## LIST OF FIGURES

Figure	Page	
1.1	Illustration of the bilayer structure (top) and the bulk heterojunction structure (bottom). . . . .	4
1.2	Illustration of energy levels of a donor-acceptor system. . . . .	8
2.1	Structures of the molecules used in this study. "X", "Y", "Ar" and "R" were explained in Table 2.1. The molecules were randomly divided into the training set and the validation set was used in k-fold cross validation process. . . . .	17
2.2	LUMO-Gap distributions of molecules in the training and validation set. Molecules are identified in the graph by their label. Red: hexagon that contains one molecule; Deep red: hexagon that contains two molecules; Blue: hexagon that contains more than two molecules. . .	18
2.3	Calculations from pure functionals compared with experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	25
2.4	Calculations from global hybrid functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	26
2.5	Calculations from RSH functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	28
2.6	Calculations from pure functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	29
2.7	Calculations from global hybrid functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	30

2.8	Calculations from RSH functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	31
2.9	Calculations from pure functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	32
2.10	Calculations from global hybrid functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	33
2.11	Calculations from RSH functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values. . . . .	34
2.12	Average MSEs (Mean Square Errors) of predictions for HOMO values on validation set using 8-fold cross validation. . . . .	36
2.13	Percentage of MAEs (Mean Absolute Errors) on HOMO before (light blue) and after (dark blue) a linear correction. The 0.1 in x-axis represents 10%. . . . .	37
2.14	Average MSEs (Mean Square Errors) of predictions for bandgap values on validation set using 8-fold cross validation. . . . .	39
2.15	Percentage of MAEs (Mean Absolute Errors) on bandgap before (light blue) and after (dark blue) a linear correction. The 0.1 in x-axis represents 10%. . . . .	40
2.16	Average MSEs (Mean Square Errors) of predictions for LUMO values on validation set using 8-fold cross validation. . . . .	42
2.17	Percentage of MAEs (Mean Absolute Errors) on LUMO before (light blue) and after (dark blue) a linear correction. The 0.1 in x-axis represents 10%. . . . .	43
3.1	Structures of tautomer forms of eumelanin building blocks. DHI is for Dihydroxyl indole and DHICA is for Dihydroxyl indole carboxylic acid. DHI and DHICA are both monomers for eumelanin. . . . .	47
3.2	Left: The structure of the molecules designed; Right: The 3D model of the molecule with "H" as the side group and showing the label of each atom. . . . .	49

3.3	Donor and acceptor moieties used in this study. Values of the Hammett constant $\sigma_p$ are from (Hansch, Leo, and Taft 1991) ; and values of Pytela $\sigma^i$ constant are from (Pytela 1996). . . . .	50
3.4	Illustration of the 3 rings in the designed molecule. The two dihedral angles are the angle between the yellow ring and the red ring, and the angle between the red ring and the blue ring. . . . .	51
3.5	Dihedral angles of the molecules. Top: The dihedral angle between the yellow ring and the red ring, measured following the atom label: 4-3-46-45 ; Bottom: The dihedral angle between the red ring and the blue ring, measured following the atom label: 6-5-34-35. . . . .	51
3.6	The heatmap of PCE values, which are varied with different side group combinations. Refer to Figure 3.2 for the position of $R_X$ and $R_Y$ . . .	52
3.7	HOMO and LUMO isosurface graphs (isovalue=0.02) for several typical molecules. HOMO energy values are labeled. . . . .	53
3.8	Scaled HOMO energies and gap from DFT predictions of two virtual molecules. The HOMO and LUMO isosurface graphs are of isovalue=0.02. . . . .	55
4.1	Illustration of a tandem solar cell with a 2-layer structure. . . . .	60
4.2	The basic process of calculating the spectrum correlation ratio. The largest value (scale molecule) of spectrum correlation is used to normalize the data in each region so that spectrum correlation ratios in that region are acquired for every other molecule. For future extension or any changes, we only need to recalculate the scale molecules and the recasts would become comparable with the previous molecules in the database. . . . .	61
4.3	AM 1.5 solar spectrum. The 4 colors represents the 4 ranges that will be used to benchmark the spectrum performance of the molecules. . . .	63
4.4	The distributions of the first peaks of all of the molecules. The order of the molecules was represented by their IDs in CEPDB. . . . .	64
4.5	The molecules that have a correlation ratio higher than the threshold (which is 0.7). Ranked with the number of regions where it satisfies the criteria. . . . .	66
4.6	26 small Divisions then combined to 4 ranges of solar spectrum(mainly IR, red, purple, UV) for later screening to work in. . . . .	66

4.7	Screening result for range I. . . . .	67
4.8	Screening result for range II. . . . .	67
4.9	Screening result for range III. . . . .	68
4.10	Screening result for range IV. . . . .	68
4.11	Screening result across regions. . . . .	69
4.12	The screening result of the molecule that is transparent in visible range (from 1.65 eV to 3.26 eV) but absorbs energy in UV and IR range. . . . .	70
A.1	Illustration of global minimum of energy functional $E[n]$ . . . . .	92
A.2	Levels of accuracy of different DFT functionals (From reference (Mardirossian and Head-Gordon 2014)). . . . .	94
A.3	Functionals used in Chapter 2 for benchmarking purpose. . . . .	96
A.4	Basis sets effect in DFT calculations (from Plumley and Dannenberg 2011). Under a given functional, the result of energy calculation con- verged to a certain value as basis sets grow larger. . . . .	97
A.5	HOMO energies from calculations and experimental ionization poten- tial (IPs) (Zhang and Musgrave 2007). The 100% correlation line is indicated by the dashed line. . . . .	99
A.6	Experimental values compared to theoretical values shows a linear de- pendence: a. Calculations with a larger basis set; b and c. Calculations with 6-31+G* basis set (Zhan, Nichols, and Dixon 2003). . . . .	100
A.7	Curve of RMSE values variation with different value of range-separated parameters (Rohrdanz and Herbert 2008). . . . .	101
A.8	Spectra calculation procedures of TD-DFT implemented in the octopus code. . . . .	102
A.9	Mean absolute errors for $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ predictions (Leang, Zahariev, and Gordon 2012). . . . .	103
A.10	Statistical analysis of the results using VT set as the reference. MSE is Mean Signed Error. MAE is Mean Average Error. RMS is Root Mean Square (Jacquemin et al. 2009). . . . .	104
A.11	Statistical analysis of the results using VE set as the reference (Jacquemin et al. 2009). . . . .	105
A.12	Statistical analysis of the difference between the values from different functionals and the reference values (Dev, Agrawal, and English 2012b). . . . .	106

# CHAPTER 1

## Background and Introductions

“It always seems impossible until it’s done. ”

— Robert H. Goddard

### 1.1 Motivations

There are two major challenges that are associated with the fossil fuel energy. On the one hand, energy is essential to human civilization and the world energy consumption is increasing every year. According to a survey conducted by BP Inc. (BritishPetrol 2017), in the year 2015, the total energy consumption was 13147.3 (million ton oil equivalent), whereas the number in 2005 was around 11000 (million ton oil equivalent). The increase is about 19.5% over the past 10 years. Fossil fuel is our major source of energy (In the U.S., it made up to 81.5% of total energy consumption in 2015 (EnergyIntelligenceAgency 2015)). Its storage is fixed in the earth, and the shortage of fossil fuel energy poses a severe challenge. On the other hand, burning fossil fuel leads to serious environmental problems. The mass production of  $CO_2$  is the major cause of the greenhouse effect. Other harmful gases generated in this process, such as  $SO_2$ , can lead to acid-rain.

The photovoltaic effect, first discovered by the French scientist Becquerel in 1839, is designed to be used as the mechanism to harvest solar energy in solar cells, which has had a huge development in the energy market in recent years. According to Solar Energy Industries Association (SEIA) (SolarEnergyIndustriesAssociation 2016), there was a 95% growth in the solar energy market in 2016, which now makes solar energy first place in new capacity additions among all kinds of sustainable energies. However, most of the growth is from silicon based solar cells. Researchers have already developed the next generation solar cells which are based on organic materials.

Organic photovoltaic (OPV) has been proven to have a faster development in that its power conversion efficiency (PCE) doubled in the last 10 years, whereas it

took silicon cells 40 years to achieve the same progress. Because of that, it will be more promising once it comes on in the market. In the following sections, we will first discuss structures of an OPV cell following a historical order, and then we will concentrate on screening better donor materials based on the general PCE formula and the Scharber model.

## **1.2 Designing OPV cell's structures according to its operation principle**

Since the discovery of the photovoltaic effect, there were many investigations done to various materials, including both inorganic and organic materials to maximize the PCE. However, because they didn't find the right structure for organic material, the development of OPV cells is delayed.

### **1.2.1 Simple structure**

At the early years, most of the material developed are inorganic material and this is due to the fact that the device made from inorganic material doesn't require complicated structures to have a high PCE. The simple structure of the device is a sandwich-like structure with certain testing material in between the two electrodes.

However, organic materials still attracted more attention because they are cheaper and more flexible than traditional materials. The first organic molecule investigated was anthracene by Pochettino (1906) and by Volmer (1913). The device made in 1959 (Kallmann and Pope 1959) has a very small PCE (0.001%). In addition to the anthracene, other organic materials were examined as well, including most of the dye molecules, like porphyrin, methylene blue and biomolecules, like chlorophyll-a, carotene etc. The more molecules people tested, the more results of tiny PCE were reported. Overall, by the 1970s, 0.1% was the PCE limit that organic photovoltaic structure could not go beyond (Spanggaard and Krebs 2004).

### **1.2.2 Heterojunction structure**

A new structure, rather than the same structure inherited from the inorganic solar cell was needed to reflect the uniqueness of organic material in order to have a high PCE. In 1979, a breakthrough was achieved by Tang (Tang 1979) (Tang 1986) who showed the PCE could go beyond 1% by designing a bilayer structure of two different kinds of molecules together between two electrodes. From the physics picture, after

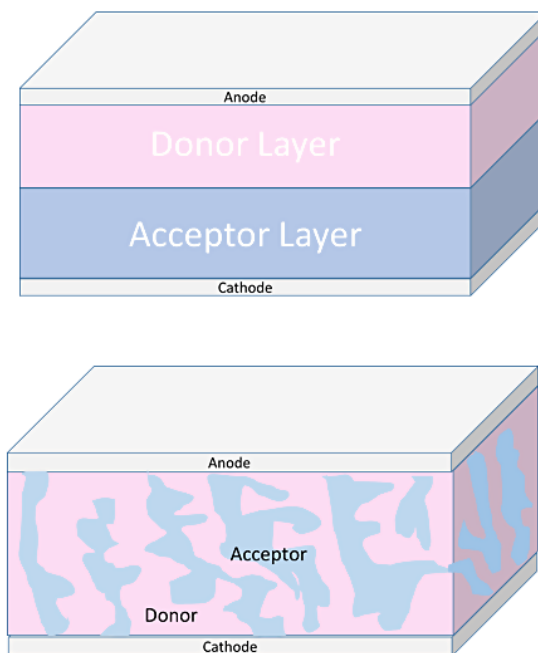
the absorption of a photon, an electron will be promoted from the Highest Occupied Molecular Orbital (HOMO) to the Lowest Unoccupied Molecular Orbital (LUMO) and leave behind it a hole in the HOMO. Then the electron and the hole form a quasi particle called exciton. Lacking of an efficient way to split excitons into free charge carriers is the main reason OPV devices in early days usually had a low PCE. This is because, compared to the inorganic system, the excitons in organic system have a stronger interaction, or binding energy, due to their low dielectric constant (Scholes and Rumbles 2006). Therefore, it is more difficult for excitons in organic material to dissociate by themselves. The mechanism Tang proposed was that the local field near the interface of two molecules would aid the dissociation of local excitons, and therefore, improve the performance.

According to current results, although some details are still under debate, some of the major physics processes involved in the photoinduced charge generation are (Hedley, Ruseckas, and Samuel 2016) (Lu et al. 2015) (Dou et al. 2013) (Derouiche and Djara 2007):

- Exciton formation in the bulk;
- Exciton diffusion;
- Exciton dissociation and Charge-transfer complex formation;
- Formation and collection of free charge carriers;

The first bottleneck of improving PCE lies in the dissociation of excitons. Because of the low dielectric constant, the excitons formed in organic system have a binding energy ( $\sim 0.3$  eV) much larger than those formed in inorganic system ( $\sim 10$  meV); therefore, it is difficult to dissociate by thermal excitations (Dou et al. 2013). However, once an exciton is formed, it is energetically favorable for the electron to further relax into a lower LUMO. So if another kind of molecule (called acceptor) with a lower LUMO was introduced, then an electron from the donor molecule could delocalize into a spatially different location, weakening the coulomb interaction in the exciton, therefore achieving a dissociation. The same would be true for a hole, but this time it would require a suitable energy difference between HOMOs from the donor molecule and the acceptor molecule. Although the details of the above process are still under debate, the exciton dissociation is more efficient when a donor and an acceptor have a suitable HOMOs and LUMOs.

Based on this operation principle, structures could be further optimized into a so-called "bulk heterojunction" structure (shown in Figure 1.1), first proposed by Heeger's group (Yu et al. 1995), to increase the area of the interface between two molecules for excitons to split, leading to a higher PCE. In this new structure, donor molecules and acceptor molecules are randomly mixed together, or interpenetrated into each other's layer. Compared to the bilayer structure, where only excitons near the interface of the two layers would have chance to dissociate, bulk heterojunctions have different phases everywhere. So excitons in the bulk heterojunction, once generated and randomly diffused to any directions, have a better chance to be split into free charges and therefore, the OPV cell would have a high PCE.



**Figure 1.1: Illustration of the bilayer structure (top) and the bulk heterojunction structure (bottom).**

### 1.3 Improving PCE by investigating the donor materials

Once structure is decided (bulk heterojunction), it is important to select a good pair of donor and acceptor material that collaborate well to lead to a high PCE. In recent years, different acceptor and donor materials have been investigated to further improve the PCE. Among different acceptor materials, fullerene and its derivative ( $PC_{61}BM$  and  $PC_{71}BM$ ) have a very high electron affinity. According to Yu *et*



*al.*(1994) and Lee *et al.*(1993), if fullerene and its derivative ( $PC_{61}BM$  and  $PC_{71}BM$ ) are used as the acceptor material, the photoconductivity increases by an order of magnitude than pure MEH-PPV (poly((2-methoxy-5-(2'-ethylhexoxy)-p-phenylene) vinylene)). Also, with fullerene the charge separation time is measured to be less than 100 fs (Zerza et al. 2001). So, fullerene and its derivative ( $PC_{61}BM$  and  $PC_{71}BM$ ) are well recognized to be used as the acceptor material. The structure of bulk heterojunction and acceptor material of fullerene are currently used to ensure an efficient exciton dissociation.

Assuming the exciton dissociation is not the bottleneck, from the general definition of the efficiency and the electricity power (Nelson 2003) :

$$\eta = \frac{J_{SC}V_{OC}F}{P_{in}} \quad (1.1)$$

where  $J_{SC}$  is the short-circuit current,  $V_{OC}$  is the open-circuit voltage,  $F$  is the filling factor and its typical value to be used is 0.65,  $P_{in}$  is the input power to a solar cell, which under the standard AM 1.5 solar radiation spectra would take the value of  $1000 \text{ W/m}^2$ .

From the formula, there are some ways to improve the PCE:

- Control of morphology through varying the side chain or solvent to achieve an ideal deposition (Rumer and McCulloch 2015) (Kim et al. 2016);
- Narrow the bandgap to increase  $J_{SC}$ ;
- Increase the ionization potential of donor materials to deepen  $V_{OC}$ ;

Here we concentrate on the later two points, namely  $J_{SC}$  and  $V_{OC}$ .

### 1.3.1 Short-circuit current

The  $J_{SC}$  is directly connected to the optical bandgap through:

$$J_{SC} = \int EQE(\omega) \times F_{AM1.5}(\omega) d\omega \quad (1.2)$$

where the  $EQE(\Omega)$  is external quantum efficiency,  $F_{AM1.5}$  is the photon flux under the standard AM 1.5 solar radiation spectra.

From the definition, the optical bandgap is  $E_g = E_{LUMO} - E_{HOMO}$ , which means the optical bandgap is decided by both HOMO and LUMO. Since HOMO is related

to  $V_{OC}$ , it is more desirable if we could adjust the molecule’s LUMO only or adjust the molecule’s HOMO only. One way of doing this is using the push-pull design.

In a push-pull design, electron donating groups (EDGs) and electron withdrawing groups (EWGs) are used to modify the molecule, and there is a good chance that only HOMO or LUMO would be adjusted (Selvaraju et al. 2016a). The set of EDG and EWG groups are growing. Aside from their effects on HOMOs and LUMOs, some EDG or EWG also have certain steric effects that could be used to control the morphology. Also, substituent groups like fluorine could effect solubility of the molecule, although it is a strong EWG. This means, choosing a suitable EDG or EWG for a target molecule often needs extensive experiments or theoretical investigations to find out its consequences.

Since the photovoltaic is essentially an energy conversion process, using thermodynamic principles, Shockley and Queisser (Shockley and Queisser 1961a) derived the famous Shockley-Queisser limit for a single junction solar cell under standard AM 1.5 solar radiation spectra to be 33.7%. And to reach this limit, optical bandgap should be around 1.4 eV. This is a value that some of the donors used already reached (Brisset et al. 1994) (You et al. 2013) (Liu et al. 2014). Overall, when employing push-pull design, molecules could be constructed with a satisfying optical bandgap.

### 1.3.2 Open-circuit Voltage and the tandem structure

To have a higher PCE, on one hand, donor’s and acceptor’s LUMO should match each other to minimize the energy loss in the hopping process. On the other hand, although a larger  $V_{OC}$  is preferred,  $V_{OC}$  can not be arbitrarily deep. Otherwise it would affect the narrow bandgap. Thus, once the cell’s structure and the acceptor material are decided (bulk heterojunction and  $PC_{61}BM$ , respectively), the  $V_{OC}$  cannot be arbitrarily larger. To explore the interplay between those factors in bulk heterojunction, Scharber *et al.*(2006) investigated many cells with high PCEs and summarized a semi-empirical model, or the Scharber model, to guide choices of donors. This model maps a pair of donor’s bandgap and HOMOs with the corresponding PCE performance of cells.

Yet another way to have a higher PCE is to adopt a new structure, which is called tandem structure. It could be used to yield a higher PCE even beyond the Shockley-Queisser limit. The idea of the tandem structure is to select different donor molecules according to their absorption spectra such that the combination of them has the

potential to achieve a full absorption of the solar spectrum. By arranging one donor into one layer forming a junction, this multi-layer structure (multi-junctions), or the so-called “tandem structure”, could raise the total  $V_{OC}$ . This is essentially equivalent to the case where multiple “batteries” connected in serial; therefore, it should have a higher open-circuit voltage than individual junction, despite the voltage loss at the interface. Some authors estimated the efficiency of such a device and concluded that, if fabricated properly, it could yield a high PCE (Denkler et al. 2008) and even with currently available material, the PCE could exceed 14% (Li et al. 2013).

### 1.3.3 The Scharber model

The origin of the open-circuit voltage in the organic material was not clear in early years; therefore, it hindered further development of organic solar cells. To understand the open-circuit voltage better, several models have been put forward and the most informative model, which provides researchers with design rules that makes PCE increase from 5% to 11%, was the Scharber model.

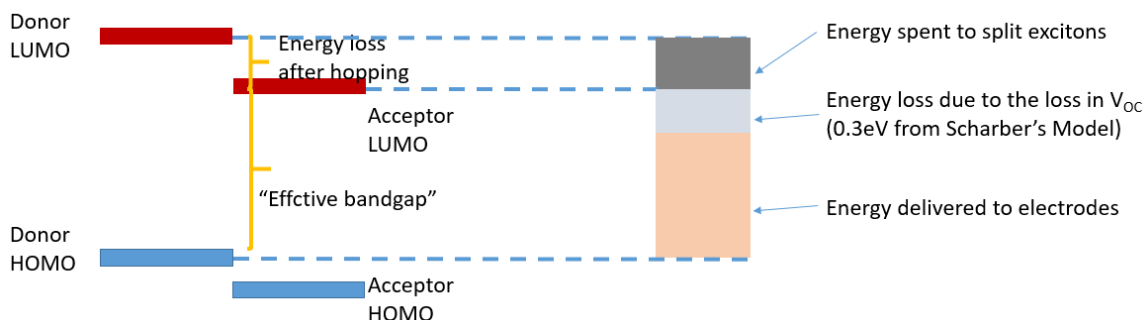
#### Background

Before diving into the Scharber model, an overview of previous models is beneficial. The first model for the organic bulk heterojunction was Metal-Insulator-Metal Model (MIM), which was a successful model for thin film cells. However, some predictions from MIM model were not supported by experimental results. For example, the  $V_{OC}$  was predicted to be independent of the intensity of incident photon flux, but experiments showed the opposite (Blakesley and Neher 2011). So this model is not quite successful for organic bulk heterojunction.

As a first step toward the physics picture inside organic solar cells, we need to be aware of two fundamental driving forces (Nelson 2003): the concentration gradient of charge carriers and the electric fields built up inside. Thus, for the bulk heterojunction, which is thick enough to maintain an effective screening of the electrical field and could re-generate enough excitons to maintain a concentration gradient, the first factor dominates; whereas, for a thin film device ( $< 100$  nm), which is like the case in inorganic solar cells, the second force dominates. Therefore, if we assume the incoming photon flux is constant, the  $V_{OC}$  in organic material should rely on the concentration gradient, which is partly decided by the bandgap of molecules inside the bulk heterojunction, if we consider the external quantum efficiency is 100%.

In 2004, Gadisa *et al.*(2004) undertook a control study, in which they constructed six OPV devices with different donor molecules. But all devices were fabricated under the same conditions and other factors, such as the anode material and cathode material are also the same. In their results, six different OPV devices had six different  $V_{OC}$  and they attributed this variation to the different oxidation potentials of six donor molecules.

Coakley and McGehee (2004) analyzed the performance of the bulk heterojunction, assuming an external quantum efficiency of 100%. Since charge carriers generated are from harvested photons, their energy should be equal to the bandgap energy. However, in order to split organic excitons, there must be an energy offset between donor's LUMO and acceptor's LUMO, which results in an energy loss when charge carriers are hopping from donor to acceptor. And this energy loss should be subtracted when we calculate the final energy converted to the electrical energy. Figure 1.2 illustrated this idea. Assuming this energy loss is 1eV (from hopping of both electrons and holes), they found out the maximum PCE that a bulk heterojunction solar cell can have is 15% and this is achieved when bandgap is 1.75 eV.



**Figure 1.2: Illustration of energy levels of a donor-acceptor system.**

### The Scharber model

From previous studies, the  $V_{OC}$ , donor's LUMO and acceptor's LUMO show correlations but it is not clear how they are correlated in organic material. In order to explore the relations, Scharber *et al.*(2006) conducted a survey study based on a collection of PCE data from 26 different donor molecules (They all used  $PC_{61}BM$  as the acceptor, so the following discussions apply to OPV devices that use  $PC_{61}BM$  as their acceptor). He put together the open-circuit voltage and the onset of oxidation data and he found the linear regression curves passed through -140 mV on the oxidation

axis, which corresponds to an energy level of 4.6 eV. Knowing the LUMO of  $PC_{61}BM$  is -4.3eV, we could deduce the following empirical formula for the  $V_{OC}$ :

$$V_{OC} = \frac{1}{e} (|E_{DONOR}^{HOMO}| - |E_{PC_{61}BM}^{LUMO}|) - 0.3V \quad (1.3)$$

where  $e$  is the elementary charge,  $E$  is the HOMO or LUMO energy level and is in the unit of eV and 0.3 is a parameter.

If we neglected the absorption from the acceptor molecule,  $PC_{61}BM$ , then the  $J_{SC}$  of a given OPV device would be decided by the donor's bandgap only. Then according to Equation 1.1, the PCE of a given OPV device could be calculated if we knew the  $V_{OC}$  and the  $J_{SC}$  (assuming  $PC_{61}BM$  is the acceptor). Based on this principle, Scharber *et al.*(2006) further plot a PCE contour map with different combinations of donor's bandgap and LUMO, which shows the maximum PCE we could have under model assumptions is 11%. Later, in 2013, Scharber *et al.*(2013) used the same model trying to predict the maximum PCE under the current fabrication limit, whose filling factor is 75% and external quantum efficiency is 80%. Their new result was 15%. Nevertheless, the two maximum PCE values correspond to the same bandgap of 1.45 eV.

One of the design rules from the the Scharber Model is that the PCE value depends on the donor's LUMO level more than the donor's bandgap. This conclusion can be drawn by noticing the fact that PCE of a given donor's LUMO varied more than the PCE of a given donor's bandgap. Another design rule is from "0.3" parameter in the formula, which shows the difference between the donor's HOMO and the acceptor's LUMO should be at least 0.3 eV apart to have a non-zero value of  $V_{OC}$ .

A clue from the Scharber Model for how to construct a high PCE donor is the following. Once the acceptor,  $PC_{61}BM$ , is decided, we should engineer the donor to let it fulfill requirements from the Scharber model to have the maximum PCE. The maximum PCE is reached when bandgap is 1.45 eV, although there are two way to achieve it: one is to raise the HOMO and the other is to lower the LUMO; So a better strategy is to first lower the donor's LUMO to around 0.3 eV above the acceptor's LUMO, which is just enough to split an exciton (since exciton binding energy is 0.1 to 0.2 eV) and then adjust the donor's HOMO to make the bandgap closer to 1.45 eV. Then the energy loss from the two sources will be minimum, hence the PCE would be maximum (shown in Figure 1.2).

## 1.4 Outline of the dissertation

This dissertation aims at designing donors used for the small molecule organic photovoltaics. One way we used was bio-inspired design based on melanin by performing Density Functional Theory (DFT) and Time-Dependent DFT (TD-DFT) calculations (chapter 3) and the other way we used was high throughput screening (HTS) on donors for the tandem structure from a database of virtual molecules (chapter 4). However, to validate the functional to be used in the DFT calculations, a benchmark study on functionals was conducted and presented in chapter 2. Please refer to Appendix A for a short introduction on DFT/TD-DFT.

Chapter 2 benchmarks the DFT/TD-DFT functionals to select the best performed functional that could predict OPV properties on molecules from melanin family better;

Chapter 3 investigates OPV donor molecules consisting of a series of organic molecules inspired by melanin structure using push-pull design principle;

Chapter 4 combines DFT/TD-DFT and high throughput virtual screening method to select better donor molecules for the tandem solar cell structure;

Chapter 5 presents the conclusions and outlook;

Appendix A gives a short introduction on DFT/TD-DFT and material informatics;

Appendix B lists the complete fitting parameters from benchmarking study in chapter 2.

## CHAPTER 2

### Benchmarking DFT/TD-DFT Functionals for Frontier Orbital Energies Predictions on Organic Photovoltaic Molecules

“There are two possible outcomes: if the result confirms the hypothesis, then you have made a measurement. If the result is contrary to the hypothesis, then you have made a discovery”

— Enrico Fermi

#### 2.1 Introduction

Efforts made toward finding high-performance organic materials with exceptional electronic and optical properties have shown promising results. In just the last few years the performance of organic light-emitting devices (Tsai et al. 2015) (Pan et al. 2016) (Zhao et al. 2017a), organic field effect transistors (Yuan et al. 2014) (Lim et al. 2017) (Luo et al. 2014) (Xiang et al. 2016), and organic photovoltaic cells (Zhao et al. 2017b) (Zhang et al. 2017) (Li et al. 2017) have advanced tremendously. As reviewed by Liu *et al.* (Liu et al. 2016), some design rules, including alternating Donor-Acceptor units and fused heterocycles (Elsherbini, Hamama, and Zoorob 2017), were summarized. Given the vast ocean of possible organic compounds (Reymond 2015), the computational search for materials with optimal properties capable of filling the appropriate semiconductor niches is ongoing (Curtarolo et al. 2012) (Kanal et al. 2013a) (Hachmann et al. 2014) (Bachman, Curtiss, and Assary 2014) (Phillips et al. 2014) (Er et al. 2015) (Qu et al. 2015) and important. These computational efforts towards the search and design of organic electronics rely on the use of electronic structure calculations that can efficiently compute molecular structure, molecular arrangements, charge transport, photoexcitations, and electron-phonon interactions.

## A. Why do we need frontier orbital energies in designing new organic semiconductors

One important metric of the organic photovoltaic (OPV) performance is the power conversion efficiency (PCE) that describes the ability of converting solar energy into electrical energy. This factor is calculated using the short-circuit current ( $J_{SC}$ ) and the open-circuit voltage ( $V_{OC}$ ) (Nelson 2003). Theoretically, the ( $V_{OC}$ ) can be calculated using the Highest Occupied Molecular Orbital (HOMO) (Scharber et al. 2006) and  $J_{SC}$  can be calculated from the optical bandgap and solar spectrum. Then the photovoltaic performance of a molecule can be predicted after acquiring the information of its frontier orbital energies.

## B. HOMO/LUMO values calculated from DFT/TD-DFT calculations

However, the size of the studied system is limited by the method chosen for quantum chemistry calculation, where a balanced consideration of both calculation speed and accuracy is usually important. Density functional theory (DFT) (Kohn and Sham 1965b) has been extensively used by quantum chemists for its controllable computational accuracy and low computational cost in obtaining electronic structures and electronic properties (Parr 1982). Yet, when calculating the energy levels of the frontier orbitals, there is a need not only for economical results but also precision in predicting the properties of the organic systems in question. However, designed molecules in the optoelectronic field often have an extended conjugation structure and intramolecular charge-transfer excitations, both of which, as we know, cannot be adequately modeled by conventional exchange-correlation functionals and are challenging problems in the DFT (Dev, Agrawal, and English 2012a) (Cohen, Mori-Sánchez, and Yang 2011) (Körzdörfer and Bredas 2014) (Jain, Shin, and Persson 2016) as well as in Time dependent DFT (TD-DFT) (Dreuw, Weisman, and Head-Gordon 2003) (Cai, Sendt, and Reimers 2002) (Grimme and Parac 2003). So it is better to first perform a benchmark study of functionals before diving into the study of molecule systems.

## C. Previous work on benchmarking HOMO predictions from DFT

Zhang *et al.* (Zhang and Musgrave 2007) investigated the accuracy of predicted orbital eigenvalues by using 13 functionals with a set of 27 molecules. Since HOMO and LUMO have correspondence with the ionization potential and the electron affinity, respectively (Perdew et al. 1982) (Stowasser and Hoffmann 1999), Zhang and co-authors



compared the experimental values with calculated results from DFT/ TD-DFT. Also, after applying a simple linear correction to the calculated results, predicting errors are systematically reduced. They also concluded that a certain portion of the HF energy was critical to the accuracy of the HOMO prediction (KMLYP with 55.7% HF and BH&HLYP with 50% HF had the best accuracy) using DFT. McCormick and co-authors (McCormick et al. 2013) examined the accuracy of orbital energy predictions of polymers with different lengths. They found the error of the HOMO energy predictions from B3LYP grew larger when the length of the molecule exceeded certain length.

In a recent benchmark paper by Szczepanik *et al.*(2017), who used a set of small molecules having different conjugation lengths, Szczepanik *et al.*(2017) found the HOMO predictions by conventional hybrid functionals were less accurate than long-range corrected functionals. They attributed this discrepancy to the wrong long range behavior of the conventional functionals. Of all the functionals they tested, one of the range separated functionals, wB97xD, gave the best performance.

#### **D. previous work on benchmarking Gaps from TD-DFT**

One of the schemes to predict the optical bandgap is from the first excitation energy given by TD-DFT. In the benchmark work led by Zhang *et al.*(Zhang and Musgrave 2007), they were able to calculate ten gap results from both DFT and TD-DFT and compared them with their experimental counterparts. They found TD-DFT gave better accuracy in general. For optical bandgap modeling, excitation energies obtained from TD-DFT have been proven to be more accurate than the difference of frontier orbital energies from DFT, as concluded by Zhang *et al.*as well as other authors (Leang, Zahariev, and Gordon 2012) (McCormick et al. 2013).

For valence state predictions, TD-DFT with global hybrid functionals is capable of producing a small error ( $< 0.26\text{eV} \sim 0.28\text{eV}$ ) as examined by Peach *et al.*(Peach et al. 2008) (Mean Absolute Error (MAE): B3LYP=0.22eV), Adamo *et al.*(Adamo, Scuseria, and Barone 1999) (MAE for PBE0 is 0.26eV). Leang *et al.*(Leang, Zahariev, and Gordon 2012) found MAE: B3LYP=0.26eV, X3LYP=0.26eV, PBE0=0.30eV, M06=0.25eV. However, for charge-transfer excitations that can be depicted qualitatively by the overlap quantity Gamma (Peach et al. 2008), Peach *et al.*found that errors arising from B3LYP and PBE predictions increase when Gamma decrease (Peach et al. 2008).

Range separated hybrid (RSH) functionals treat short-range and long-range interactions differently and in general should be able to improve the prediction accuracy for charge transfer excitations. So as a general solution to model molecules with long conjugation lengths and charge transfers, RSH is favored by several benchmarking works (Jacquemin et al. 2007) (Lange, Rohrdanz, and Herbert 2008) (Jacquemin et al. 2008). However, studies also found that the accuracy of RSH varied greatly between different molecules. The factor responsible may be the charge transfer distance (Nguyen, Day, and Pachter 2011a) or the number of cyclic structures (Szczepanik et al. 2017) or molecule families (Jacquemin et al. 2011). And there are also studies pointing out that benchmarking results from model compounds may not be accurate enough. Therefore, some authors begin to tune the  $\omega$  parameter themselves to give better results for their molecule system.

## Research Goal

In this article, we applied DFT and TD-DFT to a set of twenty-nine molecules in the organic semiconductor field and report the accuracy and performance of several exchange-correlation functionals (LDA, hybrid-LDA, GGA, hybrid-GGA, mGGA, hybrid-mGGA and RSH, as listed in Table 2.2) for calculations of frontier orbital energies. After validation studies from k-fold cross validation, we also applied corrections from the linear regression to the DFT/TD-DFT results. Prediction errors from with and without linear correction were compared.

## 2.2 Computational method

DFT can be used to obtain, within a reasonably correct range, the energy levels for occupied states (Kang and Musgrave 2001a), and is what we used to obtain the value of the highest occupied molecular orbital (HOMO). We obtained the values of the first excitation energy from TD-DFT and we considered it to be the bandgap (Runge and Gross 1984a) (Marques and Gross 2004). The lowest unoccupied molecular orbital (LUMO) is then determined from the addition of HOMO and the first excitation energy. To benchmark our results, we carried out a linear correlation study between the calculated values (Kang and Musgrave 2001a) of the  $E_{HOMO}/E_{gap}/E_{LUMO}$  energies and their counterparts found experimentally from cyclic voltammetry measurements.

We carried out DFT calculations with the Gaussian 09 software package (*Gaussian09 Revision C.01*). Our exhaustive study examines the performance of the func-

tionals listed in Table 2.2. We found the optimal geometry of each molecule using each of the exchange correlation functionals mentioned above, and with the the 6-311+G(d,p) level basis set. We calculated the energy of the molecular orbitals using again the appropriate exchange correlation functional and also at the 6-311+G(d,p) level basis set which is sufficient for our calculations. Based on the optimized ground state geometry, Time-dependent DFT (TD-DFT) calculations for singlet electronic configuration were then applied to find the first excitation energies. The first twenty states were used to acquire an accurate result for energy.

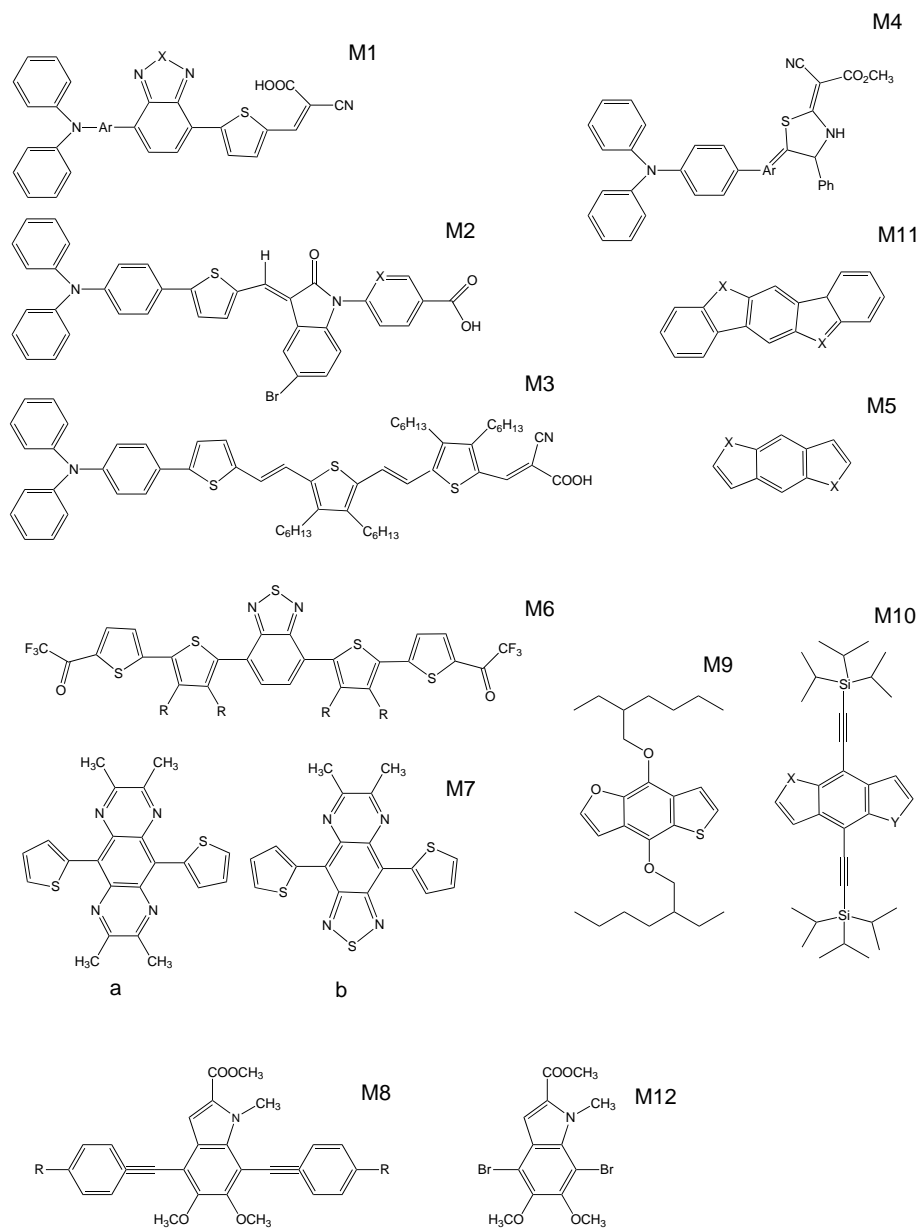
Figure 2.2 shows the distribution of the molecule set that we used in this study. Some of them have a relatively narrow bandgap and a deep HOMO, which favors their photovoltaic application. Also, some of these molecules contain Silicon and Selenium atoms, which could improve their performance. In addition, some of them have extended cyclic structures that are also favored by organic electronic applications. To conclude, molecules in this set represent the novel synthesized molecules in the organic semiconductors field.

Additionally, since one of our purposes is to make better predictions on HOMO, bandgap and LUMO values, we further constructed linear regression models based on the results from the functionals. Generally, linear regression is one of the simple models, which is used by many authors: *e.g.* Zhang and Musgrave in molecular orbital energy prediction (Zhang and Musgrave 2007), Sharma *et al.* in NMR chemical shifts (Sharma, Zhang, and Ohlin 2016), and XM Duan *et al.* in the heat of formation (Duan et al. 2004). Along with a linear regression model, the k-fold cross validation method (k=8) was used, which is often used to evaluate and compare the performance of predictive models in machine learning field (Cawley and Talbot 2010), and is also used to validate the parameters we summarized for the linear model.

In a k-fold cross validation (k=8 in our study), the molecule set was randomly divided into eight groups and each contained a similar number of molecules. Of the eight groups, one group served as a validation set while the remaining seven were used as the training set. The regression formula from the seven groups was validated on the validation set by calculating mean square error (MSE, as shown in Equation (2.1)). Then, this process would repeat again except this time, another group would be picked up as the validation set, while the remaining seven groups were used as the training set. Because we had eight groups, eight different MSEs would be averaged to have one final MSE. Although molecules were randomly put into 8 groups, to avoid accidental bias in the group division scheme and therefore to

have a more accurate estimation on the MSEs and regression parameters, we surveyed 75,000 different random divisions. An average MSE of 75,000 MSEs raised from those divisions was calculated to represent the linear model performance on the results from that functional, and it was also a measure of validity of applying a linear model to the results.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i^{experiment} - y_i^{prediction})^2 \quad (2.1)$$



**Figure 2.1: Structures of the molecules used in this study. "X", "Y", "Ar" and "R" were explained in Table 2.1. The molecules were randomly divided into the training set and the validation set was used in k-fold cross validation process.**

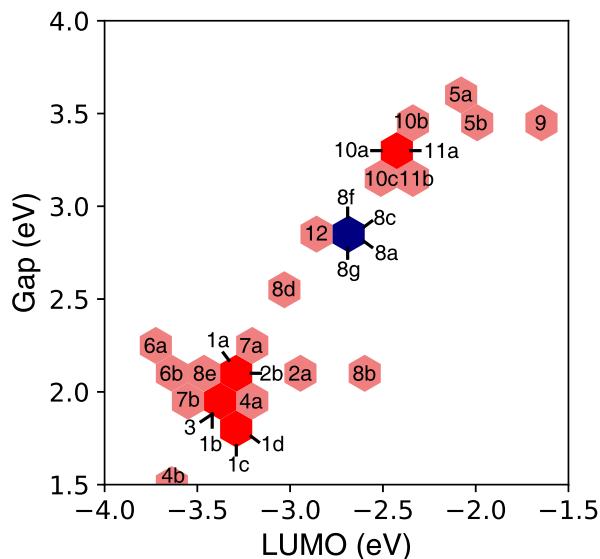


Figure 2.2: LUMO-Gap distributions of molecules in the training and validation set. Molecules are identified in the graph by their label. Red: hexagon that contains one molecule; Deep red: hexagon that contains two molecules; Blue: hexagon that contains more than two molecules.

Table 2.1: The chemical formula, SMILES and references for molecules set.

ID	Variations	Chemical Formula	SMILES	Reference
M1a	$X = S$ $Ar = C_6H_4$	$C_{32}H_{20}O_2N_4S_2$	<chem>OC(=O)\C(=C\C1=CC=C(S1)C2=CC=C(C3=CC=C(C=C3)N(C4=CC=CC=C4)C5=CC=CC=C5)C6=NSN=C26)C#N</chem>	(Velusamy et al. 2005)
M1b	$X = Se$ $Ar = C_6H_4$	$C_{32}H_{20}O_2N_4SSe$	<chem>OC(=O)\C(=C\C1=CC=C(S1)C2=CC=C(C3=CC=C(C=C3)N(C4=CC=CC=C4)C5=CC=CC=C5)C6=N[Se]N=C26)C#N</chem>	

*Continued on the next page*

Table 2.1 – Continued from the previous page

ID	Variations	Chemical Formula	SMILES	Reference
M1c	$X = S$ $Ar = C_4H_2S$	$C_{30}H_{18}O_2N_4S_3$	<chem>OC(=O)\C(=C\C1=CC=C(S1)C2=CC=C(C3=CC=C(S3)N(C4=CC=CC=C4)C5=C(C=CC=C5)C6=NSN=C26)C#N</chem>	
M1d	$X = Se$ $Ar = C_4H_2S$	$C_{30}H_{18}O_2N_4S_2Se$	<chem>OC(=O)\C(=C\C1=CC=C(S1)C2=CC=C(C3=CC=C(S3)N(C4=CC=CC=C4)C5=C(C=CC=C5)C6=N[Se]N=C26)C#N</chem>	
M2a	$X = CN$	$C_{37}H_{23}O_3N_2SBr$	<chem>OC(=O)C1=CC=C(C=C1)N2C(=O)\C(=C\C3=CC=C(S3)C4=CC=C(C=C4)N(C5=CC=CC=C5)C6=CC=CC=C6)C7=CC(=CC=C27)Br</chem>	(Tingare et al. 2013)
M2b	$X = N$	$C_{36}H_{22}O_3N_3SBr$	<chem>OC(=O)C1=NC=C(C=C1)N2C(=O)\C(=C\C3=CC=C(S3)C4=CC=C(C=C4)N(C5=CC=CC=C5)C6=CC=CC=C6)C7=C2C=CC(=C7)Br</chem>	
M3		$C_{62}H_{74}O_2N_2S_3$	<chem>OC(=O)\C(=C\C1=CC=C(S1)\C=C\C2=CC=C(S2)\C=eC\C3=CC=C(S3)C4=C(C=C(C=C4)N(C5=CC=CC=C5)C6=CC=CC=C6)C#N</chem>	(Aljarilla et al. 2012)
M4a	$Ar = CH$	$C_{32}H_{23}ON_3S$	<chem>COC(=O)C(C#N)C1=NC(=C(CC2=CC=C(C=C2)N(C3=CC=CC=C3)C4=CC=CC=C4)S1)C5=CC=CC=C5</chem>	(Esteban et al. 2011)

Continued on the next page

Table 2.1 – Continued from the previous page

ID	Variations	Chemical Formula	SMILES	Reference
M4b	$Ar = C_4H_2S$	$C_{36}H_{25}ON_3S_2$	<chem>COC(=O)C(C#N)C1=NC(=C(CC2=CC=C(S2)C3=CC=C(C=C3)N(C4=CC=CC=C4)C5=eCC=CC=C5)S1)C6=CC=CC=C6</chem>	
M5a	$X = S$	$C_{10}H_6S_2$	<chem>S1C=CC2=CC3=C(C=CS3)C=C12</chem>	(Hideaki Ebata et al. 2007)
M5b	$X = Se$	$C_{10}H_6Se_2$	<chem>[Se]1C=CC2=CC3=C(C=C[Se]3)C=C12</chem>	
M6a	$R = H$	$C_{26}H_{10}O_2N_2S_5F_6$	<chem>CCC1=C(SC(=C1CC)C2=C(C=C(C3=C(CC)C(=C(S3)C4=CC=C(S4)C(=O)C(F)(F)F)CC)C5=NSN=C25)C6=CC=C(S6)C(=O)C(F)(F)F</chem>	(Steinberger et al. 2011)
M6b	$R = C_2H_5$	$C_{34}H_{26}O_2N_2S_5F_6$	<chem>FC(F)(F)C(=O)C1=CC=C(S1)C2=CC=C(S2)C3=CC=C(C4=CC=C(S4)C5=CC=C(S5)C(=O)C(F)(F)F)C6=NSN=C36</chem>	
M7a		$C_{22}H_{18}N_4S_2$	<chem>CC1=NC2=C(N=C1C)C(=C3N=C(C)C(=NC3=C2C4=C(C=CS4)C)C5=CC=CS5</chem>	(Li et al. 2011)
M7b		$C_{18}H_{12}N_4S_3$	<chem>CC1=NC2=C(C3=CC=CS3)C4=NSN=C4C(=C2N=C1C)C5=CC=CS5</chem>	
M8a	$R = t\text{-butyl}$	$C_{37}H_{39}O_4N$	<chem>COC(=O)C1=CC2=C(C#CC3=CC=C(C=C3)C(C)(C)C)C(=C(OC)C(=C2[N]1C)C#CC4=CC=C(C=C4)C(C)(C)C)OC</chem>	(Selvaraju et al. 2016b)

Continued on the next page



Table 2.1 – Continued from the previous page

ID	Variations	Chemical Formula	SMILES	Reference
M8b	$R = N(CH_3)_2$	$C_{37}H_{33}O_4N_3$	<chem>COC(=O)C1=CC2=C([N]1C)C(=C(OC)C(=C2C#CC3=CC=C(C=C3)N(C)C)OC)C#CC4=CC=C(C=C4)N(C)C</chem>	
M8c	$R = F$	$C_{29}H_{21}O_4F_2$	<chem>COC(=O)C1=CC2=C(C#CC3=CC=C(F)C=C3)C(=C(OC)C(=C2[N]1C)C#CC4=C(C=C(F)C=C4)OC</chem>	
M8d	$R = CN$	$C_{31}H_{21}O_4N_3$	<chem>COC(=O)C1=CC2=C(C#CC3=CC=C(C=C3)C#N)C(=C(OC)C(=C2[N]1C)C#CC4=CC=C(C=C4)C#N)OC</chem>	
M8e	$R = NO_2$	$C_{29}H_{23}O_8N_3$	<chem>COC(=O)C1=CC2=C(N)1C)C(=C(OC)C(=C2C#CC3=CC=C(C=C3)[N](=O)=O)OC)C#CC4=CC=C(C=C4)[N](=O)=O</chem>	
M8f	$R = H$	$C_{29}H_{23}O_4N$	<chem>COC(=O)C1=CC2=C(C#CC3=CC=CC=C3)C(=C(OC)C(=C2[N]1C)C#CC4=CC=C(C=C4)OC</chem>	
M8g	$R = OCH_3$	$C_{31}H_{27}O_6N$	<chem>COC1=CC=C(C=C1)C#CC2=C3C=C([N](C)C3=C(C#CC4=CC=C(OC)C=C4)C(=C2OC)OC)C(=O)OC</chem>	
M9		$C_{26}H_{38}O_3S$	<chem>CCCCC(CC)COC1=C2C=CS2=C(C(=O)CC(CC)CCCC)C3=C10C=C3</chem>	(Aeschi et al. 2013)

Continued on the next page

Table 2.1 – Continued from the previous page

ID	Variations	Chemical Formula	SMILES	Reference
M10a	$X = O$ $Y = S$	$C_{32}H_{82}OSSi_2$	<chem>CC(C)[Si](C#CC1=C2C=CSC2=C(C#C[Si](C(C)C)(C(C)C)C(C)C)C3=C10C=C3)(C(C)C)C(C)C</chem>	(Aeschi et al. 2013)
M10b	$X = O$ $Y = O$	$C_{32}H_{82}O_2Si_2$	<chem>CC(C)[Si](C#CC1=C2C=COC2=C(C#C[Si](C(C)C)(C(C)C)C(C)C)C3=C10C=C3)(C(C)C)C(C)C</chem>	
M10c	$X = S$ $Y = S$	$C_{32}H_{82}S_2Si_2$	<chem>CC(C)[Si](C#CC1=C2C=CSC2=C(C#C[Si](C(C)C)(C(C)C)C(C)C)C3=C1SC=C3)(C(C)C)C(C)C</chem>	
M11a	$X = S$	$C_{18}H_{10}S_2$	<chem>S1C2=CC=CC=C2C3=CC4=C(C=C13)C5=C(S4)C=CC=C5</chem>	(Hideaki Ebata et al. 2007)
M11b	$X = Se$	$C_{18}H_{10}Se_2$	<chem>[Se]1C2=CC=CC=C2C3=CC4=C(C=C13)C5=C([Se]4)C=CC=C5</chem>	
M12		$C_{13}H_{13}O_4NBr_2$	<chem>COC(=O)C1=CC2=C(Br)C(=C(OC)C(=C2[N]1C)Br)OC</chem>	(Selvaraju et al. 2016b)

**Table 2.2: Functionals examined in this study. If two HF percentiles are presented, the first one is for the short range and the second one is for the long range.**

Categories	Names	HF percentage	References
LDA	SVWN	0.0%	(Slater 1963) (Vosko, Wilk, and Nusair 1980)
Hybrid-LDA	KMLYP	55.7%	(Kang and Musgrave 2001b)
Pure GGA	PW91	0.0%	(Perdew et al. 1992) (Perdew and Wang 1992) (Perdew, Burke, and Wang 1996)
Pure GGA	BLYP	0.0%	(Becke 1988) (Lee, Yang, and Parr 1988) (Miehlich et al. 1989)
hybrid GGA	B3LYP	20.0%	(Becke 1993) (Lee, Yang, and Parr 1988) (Miehlich et al. 1989) (Stephens et al. 1994)
hybrid GGA	PBE0	25.0%	(Adamo and Barone 1999)
pure mGGA	TPSS	25.0%	(Tao et al. 2003)
hybrid mGGA	M06-2x	54.0%	(Zhao and Truhlar 2008b)
Range-separated	$\omega$ B97XD	22.2%; 100%	(Chai and Head-Gordon 2008)
Range-separated	cam-B3LYP	19.0%; 65.0%	(Yanai, Tew, and Handy 2004)
Range-separated	HSE06	25%; 0.0%	(Heyd, Scuseria, and Ernzerhof 2003a) (Heyd, Scuseria, and Ernzerhof 2006)
Range-separated	LC- $\omega$ PBE	0.0%; 100.0%	(Vydrov and Scuseria 2006) (Vydrov, Scuseria, and Perdew 2007) (Vydrov et al. 2006)

### 2.3 Results and Analysis

In the first part of this section, we will look into the HOMO, Gap and LUMO calculations directly from DFT/TD-DFT functionals that are coming from three categories, namely pure functionals, global hybrid functionals, and range separated functionals (RSHs), respectively. Then, in the second part, we will investigate improving the accuracy by applying additional linear corrections to the results. This was done by

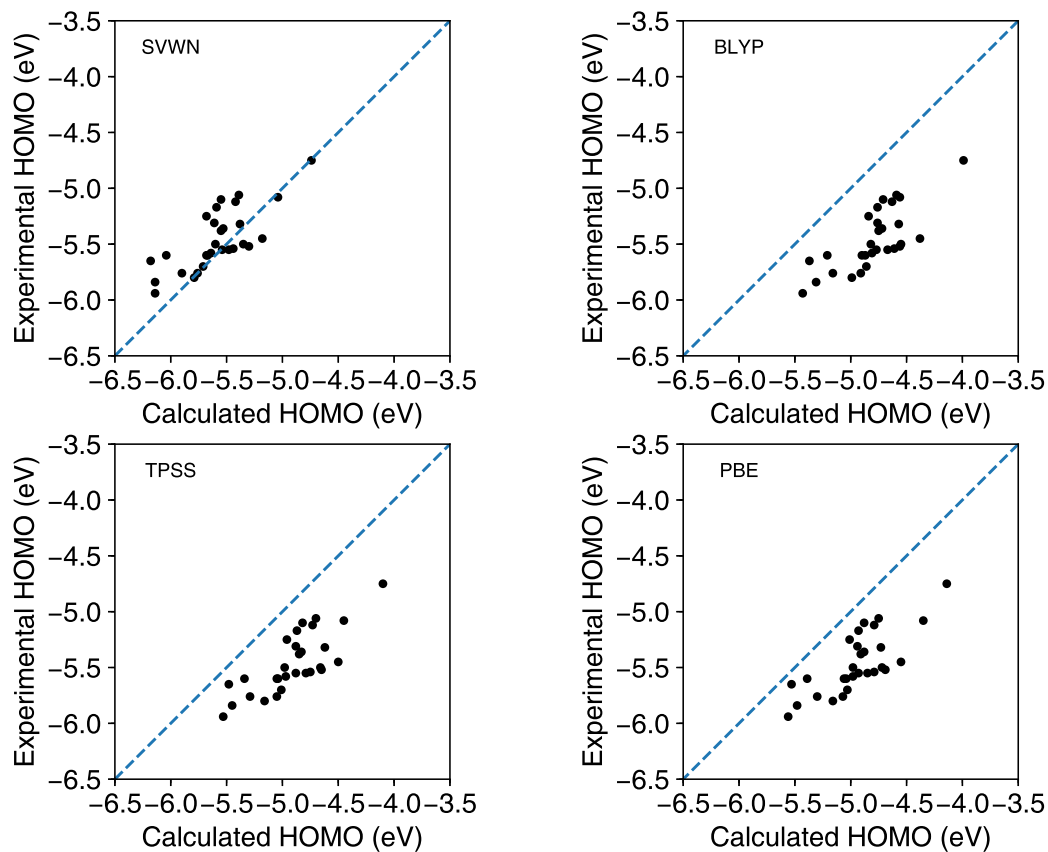
first justifying the linear correction by performing k-fold cross validations and second comparing the percentage errors of both before and after applying linear corrections.

### 2.3.1 Calculations from DFT/TD-DFT functionals

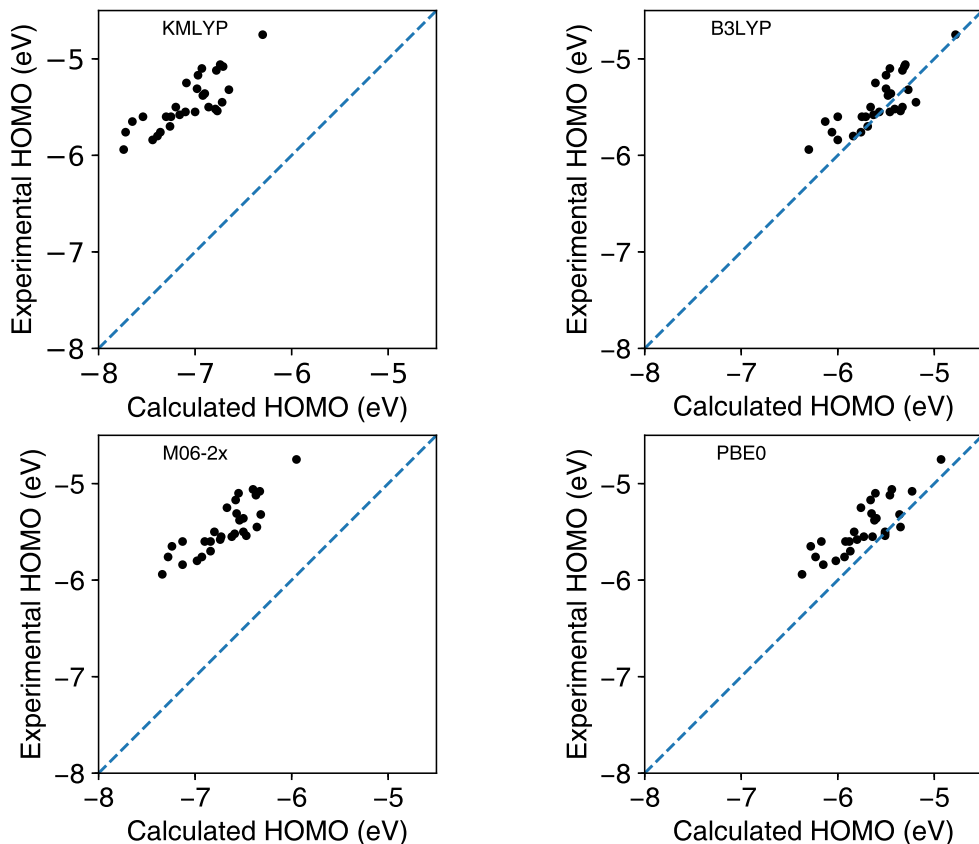
#### Calculations on HOMOs from DFT

***The Pure Functionals*** Figure 2.3 shows the predictions on the set of molecules from four pure functionals, including SVWN, BLYP, PBE and TPSS. Data points from SVWN, an LDA level functional, rest mostly on the diagonal line. Calculated values in the graph have nearly no shift when compared to experimental values in both of the horizontal and vertical directions. The results from GGA functionals (BLYP, PBE), however, are located mostly below the diagonal line, showing a horizontal shift compared to the experiments. This shift indicates the calculated HOMOs from GGA level are larger than experimental values. TPSS is in the mGGA (meta-GGA) level and most of its results are still below the diagonal line. The horizontal offset from TPSS still shows that calculated results are larger than experiments. Overall, pure LDA shows a better correlation between calculated results and experimental values, while predictions from pure GGA and pure mGGA show deeper HOMO values.

Among the calculated results from four functionals, pure LDA has the best correlations with the experiments. This accuracy outperformed other levels of functionals and is mostly due to the error cancellation from the underestimation of correlation energy and overestimation of exchange energy. GGA predictions are all deeper than experimental results. Comparing PBE and BLYP, results from PBE are closer to the diagonal line, meaning a better correlation with the experimental results. In the cases of GGA functionals, parameters in PBE were designed to satisfy exact conditions (Perdew, Burke, and Ernzerhof 1996), while the single parameter in exchange part of BLYP was designed to minimize the error of the exchange energies from rare-gas atom (Becke 1988). Although they have different assumption, the errors of their predictions are close (please refer to Figure 2.13). For the results of mGGA, the data span further along the diagonal line than GGA's results. Mostly, improvements on GGA/mGGA over LDA are the activation energy and transition state structure through correction on the correlation energy, so their prediction is less accurate than LDA in HOMO, which is also observed by other authors (Zhang and Musgrave 2007) (Szczepanik et al. 2017) (McCormick et al. 2013).



**Figure 2.3:** Calculations from pure functionals compared with experimental values. The dashed line indicates 100% correlation between calculated and experimental values.



**Figure 2.4:** Calculations from global hybrid functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.

*The Global Hybrid Functionals* To study the global hybrid model, we chose the functionals of KMLYP, B3LYP, M06-2x and PBE0. Among the four functionals, results from B3LYP and PBE0 are located almost on the diagonal line before any corrections are applied. According to Figure 2.4, for the molecule set we used, it seems there is a small horizontal offset in PBE0, which means the HOMO energies from PBE0 are shallower than experimental values. However, the other two global hybrids, KMLYP and M06-2x, have their calculations above the diagonal line before a linear correction is applied, which means their predictions are deeper than the experimental values. It was observed that, compared to other functionals, data points from the M06-2x are more concentrated.

In the previous study, (Zhang and Musgrave 2007) KMLYP showed smaller error than B3LYP before a linear correction was applied and still performed better after a linear correction was applied. However, in our molecule set, before correction, only B3LYP and PBE0 showed up on the diagonal line. This difference may be attributed

to the different molecule set we used. In their work, the molecule set contained smaller molecules than our molecule set: their molecule set included two-atom or four-atom molecules. The construction of hybrid functionals includes a certain percentage of HF component to compensate for the correction in the correlation part that GGA/mGGA functionals usually have. So results from hybrid GGA/mGGA, compared to pure GGA/mGGA, would shift horizontally to the left, making them sit on the diagonal line. For LDA functionals, adding more HF would shift to the same direction but will move them off-diagonal, since pure LDA's results were already on the diagonal line. We see from Figure 2.4, in the case of GGA/mGGA, 20% of additional HF component would be enough to improve the correlation between BLYP's off-diagonal to B3LYP's on-diagonal.

***RSH*** In addition to the pure functionals and global hybrid functionals, a series of RSH functionals, including cam-B3LYP, HSE06, LC- $\omega$  and  $\omega$ B97X have been examined and the results are shown in Figure 2.5. Except HSE06, the predictions from other three functionals were all shifted off-diagonal to the left direction, showing underestimated predictions when compared to the experimental results without a correction. Compared to the LC- $\omega$ PBE functional,  $\omega$ B97X shifted to the left more. This is due to the difference in short-range HF percentage they have, considering they have the same amount of long range HF component. In addition to the effect of short-range HF, the comparisons from B3LYP in Figure 2.4 and from cam-B3LYP in Figure 2.5 show an offset toward left region to the diagonal due to different long-range HF contributions.

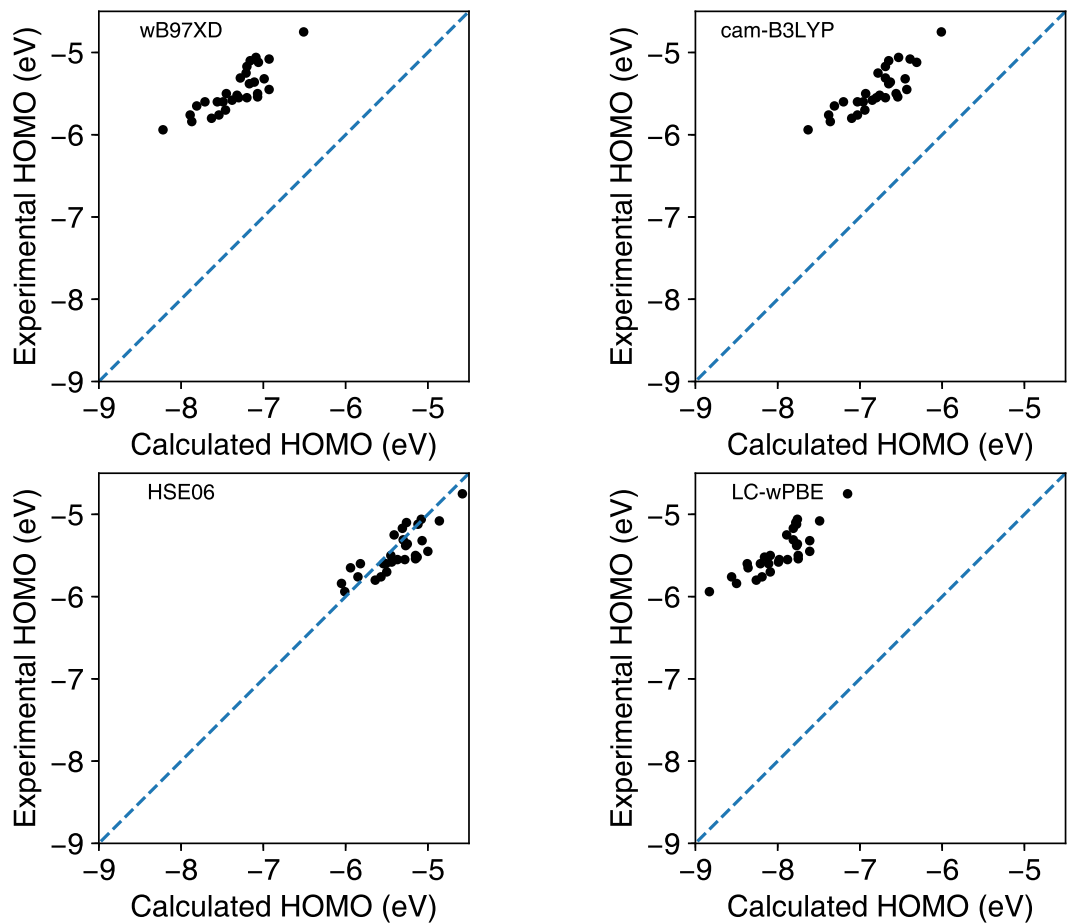
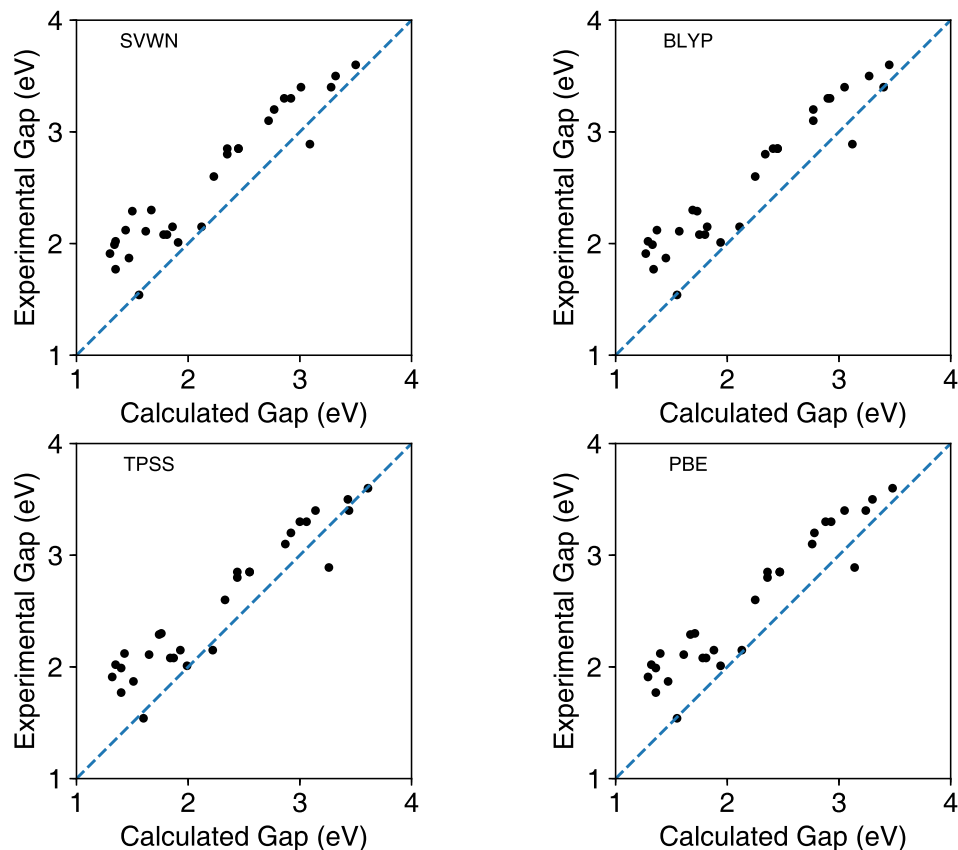


Figure 2.5: Calculations from RSH functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.



### 2.3.2 Results on Gap benchmark (Pure, Global Hybrid and RSH functionals)



**Figure 2.6:** Calculations from pure functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.

*Pure* Predictions from pure functionals of band gaps are generally higher than experimental values (as shown in Figure 2.6). Some of the molecules in the set are consistently shifted to the left. Data points from SVWN, PW91, BLYP and TPSS are all further along the diagonal line than data points from the same functional in HOMO predictions, with TPSS(mGGA) the most distributed one. For all pure functionals here, except for a few points, the gap predictions are all smaller than experimental values. Compared to the results from global hybrids (Figure 2.7), those results are larger than experimental values, owing to the HF components contained in hybrid functionals.

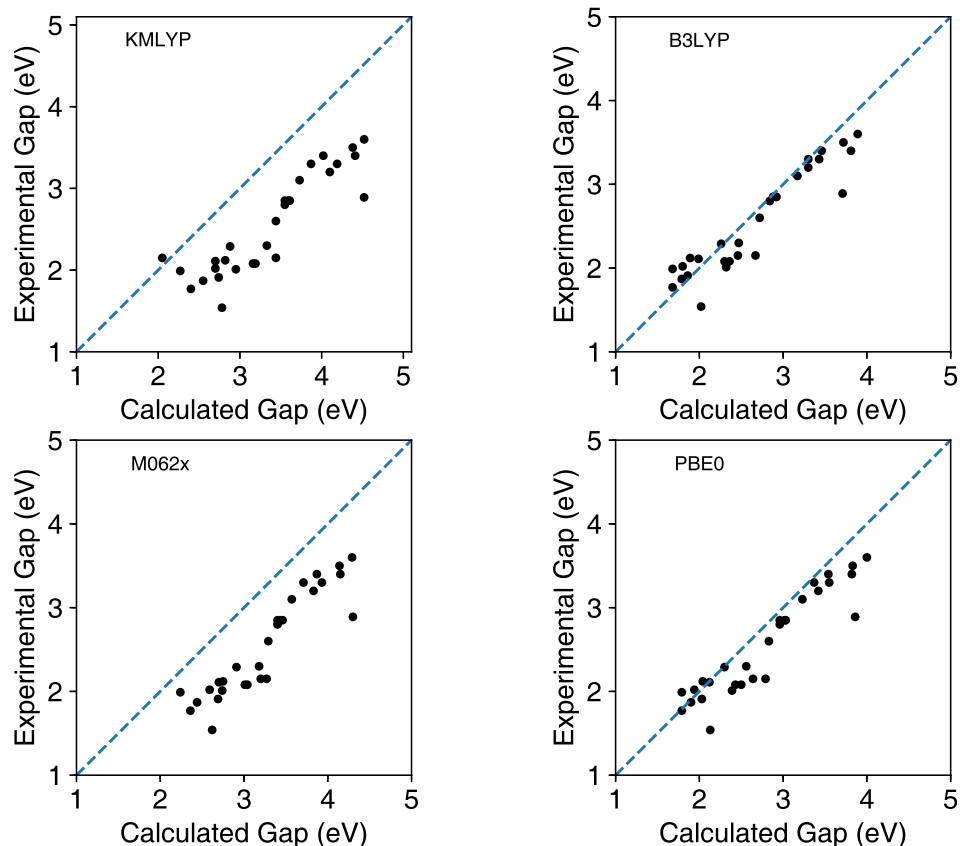
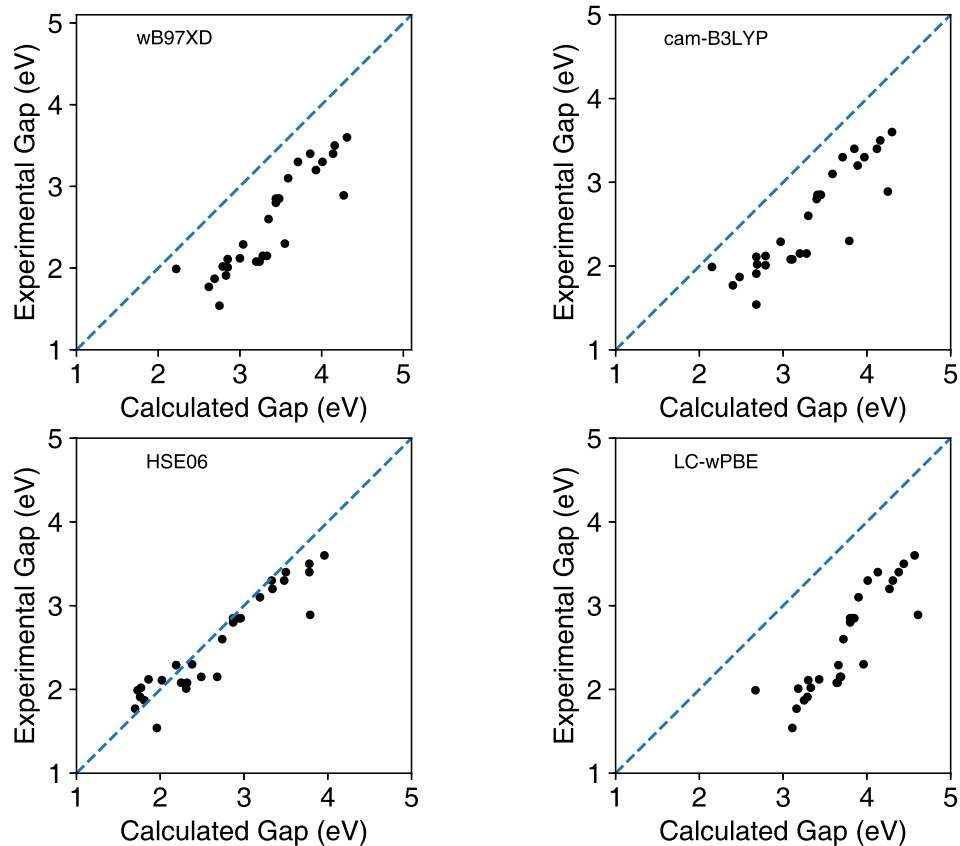


Figure 2.7: Calculations from global hybrid functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.

*Global Hybrid and RSH functionals* Contrary to the predictions from pure functionals, predictions from hybrid functionals are generally higher than experimental values (as shown in Figure 2.7). Although the deviation of predictions from the diagonal line varies, it mostly depends on the HF percentage contained in the hybrid functionals. For example, B3LYP and PBE0(PBE1PBE) have 20% of HF exchange, and their separation from the diagonal line is smaller compared to M06-2x, which has 54% of HF exchange and is therefore shifted more from the diagonal line. In contrast to the case of GGA and mGGA, although KMLYP(hybrid LDA) has 20% of the HF exchange component, it is shifted more than GGA/mGGA functionals with the same amount of HF exchange.

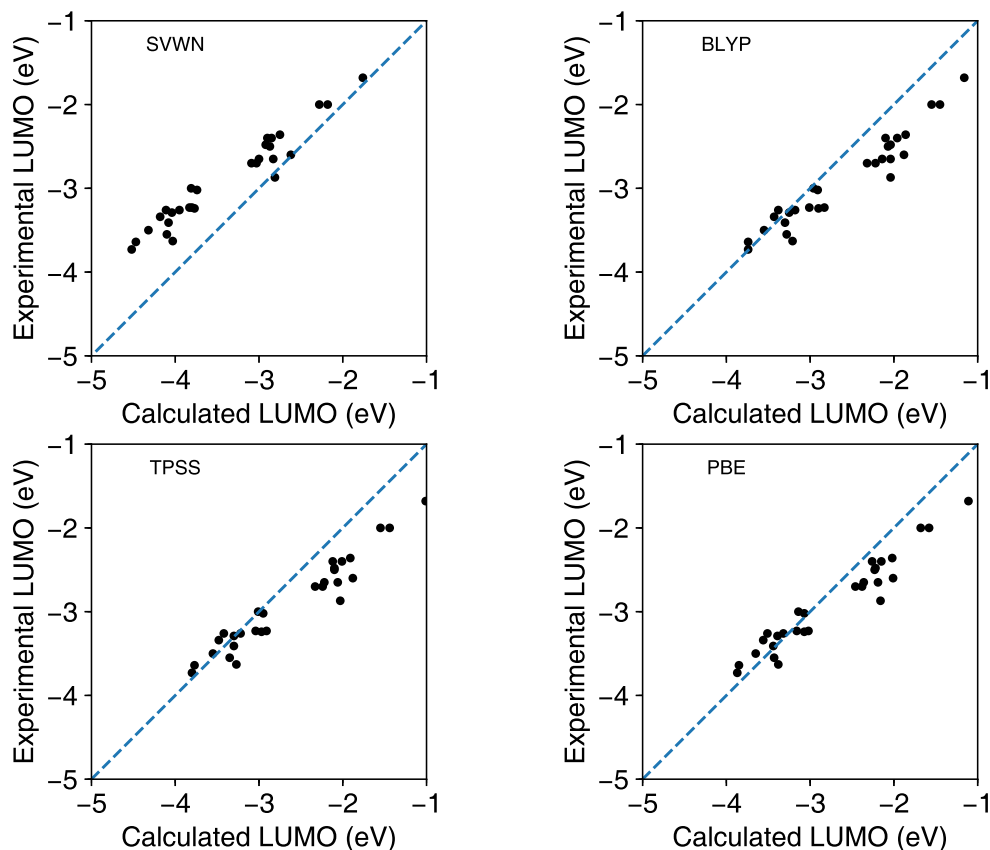


**Figure 2.8:** Calculations from RSH functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.

The predictions from LC-wPBE and wB97XD were distributed mostly along a line that lies in parallel to the diagonal line. The additional long range HF component helps to adjust the correlation by shifting the results the same distance away from the diagonal line. However, HSE06 (HSEH1PBE) is the only RSH where its results mostly rest on the diagonal line, which indicates a good correlation between theoretical and experimental values.

### 2.3.3 Results on LUMO benchmark (Pure, Global Hybrid and RSH functionals)

Rational predictions of LUMO rely on accurate calculations of ground state and on accurate calculations of bandgap. In the following study, LUMO value is acquired from HOMO and gap calculated from the same functionals. Generally speaking, the data points in the LUMO prediction graph are less diverse than gap predictions.



**Figure 2.9:** Calculations from pure functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.

SVWN (LDA) predicts shallower LUMO values nearly exactly but deeper LUMO values were off the diagonal line, as Figure 2.9 shows. On the contrary, pure GGA and mGGA (BLYP, PBE, TPSS) all predict the deeper LUMO nearly exactly but the shallower LUMOs were off the diagonal line. In the case of hybrid functionals, shown in Figure 2.10, B3LYP and PBE0 (having HF component of 20% and 25% respectively) have predictions very close to the experimental values. Among LUMO predictions from RSH, as shown in Figure 2.11, most of the HSE06 results were accurate while other functionals,  $\omega$ B97X, LC- $\omega$ PBE and cam-B3LYP, were off the diagonal line. However, results from these three functionals in gap predictions are nearly distributed on a line that is parallel to the diagonal direction.

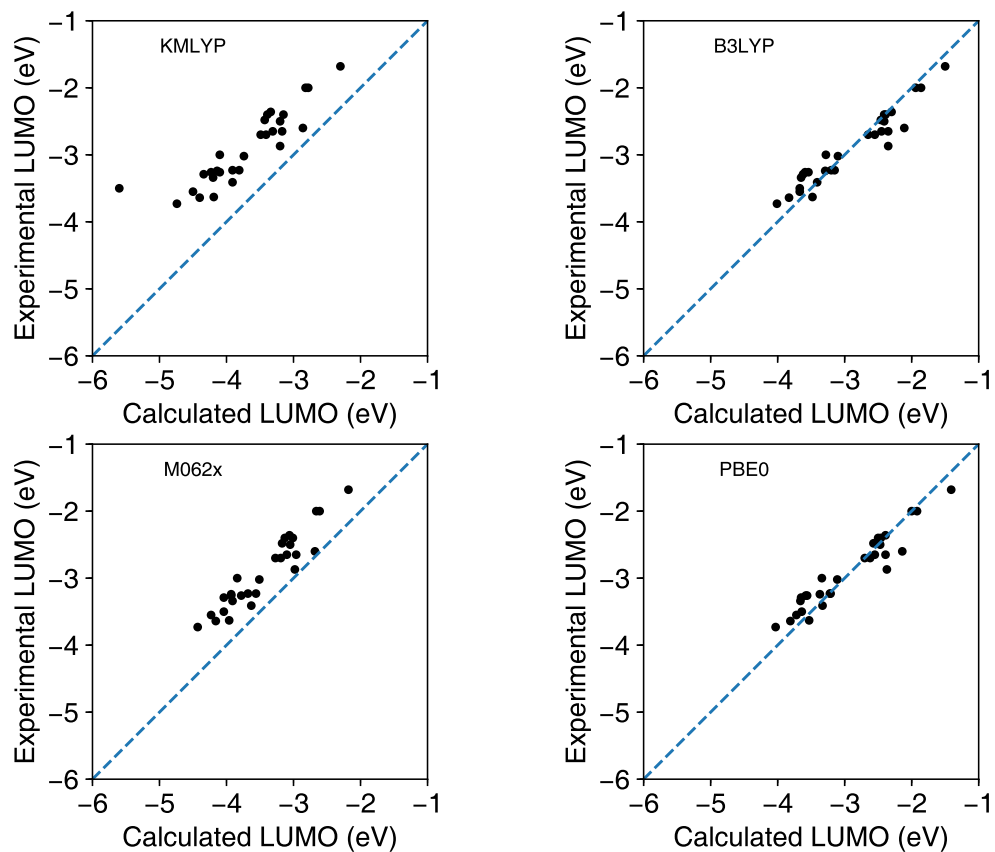
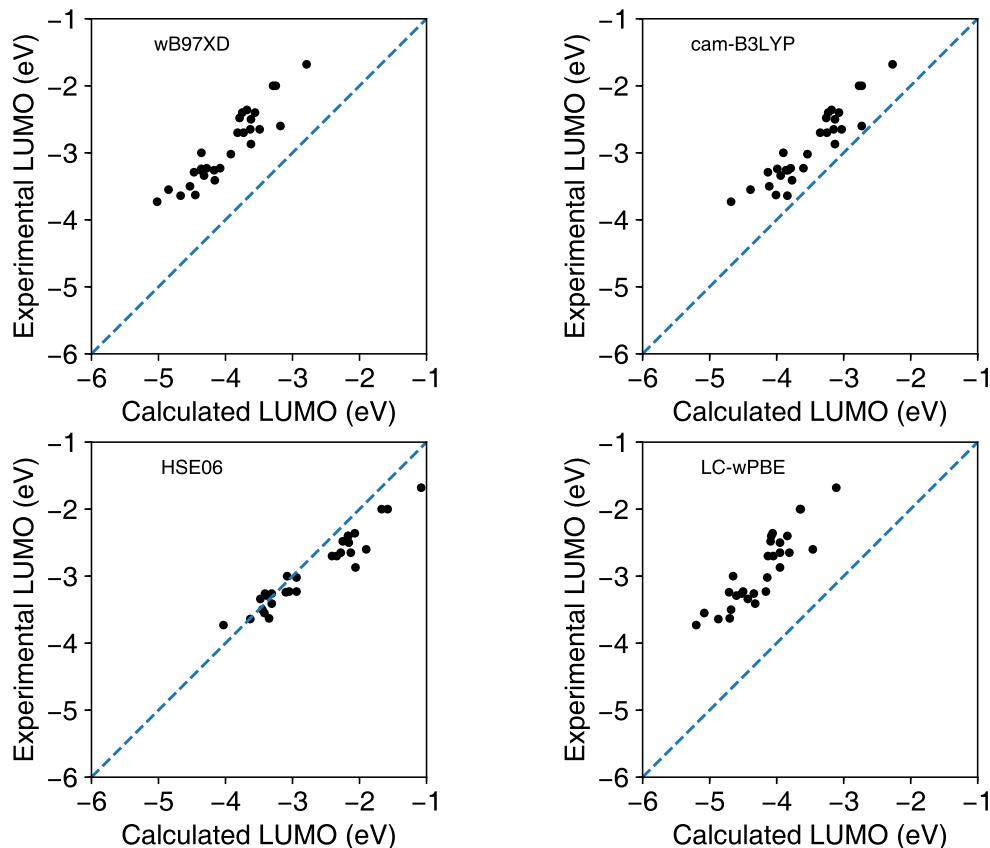


Figure 2.10: Calculations from global hybrid functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.



**Figure 2.11:** Calculations from RSH functionals compared to experimental values. The dashed line indicates 100% correlation between calculated and experimental values.

### 2.3.4 Predictions and corrections from linear regression models

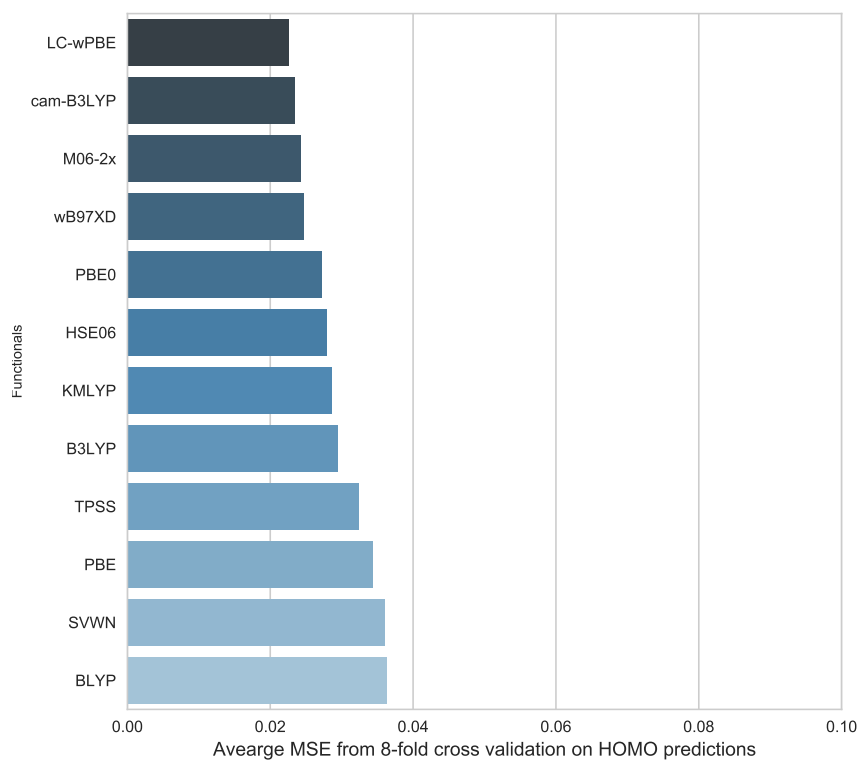
In the last section, we analyzed the performance of functionals from different categories. In this section, we turn our discussion to the possibilities of applying additional linear corrections to the calculated results. First, k-fold cross validation was applied to evaluate the validity of applying a linear model. From the principle of k-fold cross validation, the MSE obtained from this method reflects the linearity of the outputs from DFT/TD-DFT functionals. If the linearity of a certain functional was among the top compared to other functionals, then the errors from this functional could be better corrected by a linear correction. In addition, we will discuss the mean absolute error (MAE) both from before and after the linear correction. We are not aiming at benchmarking how accurate those functionals are based on their raw results; rather we are interested in acquiring better predictions after a linear correction. So the ideal functional must have linearly consistent predictions on all of the data. It is

therefore not important how close each individual prediction is.

## HOMOs

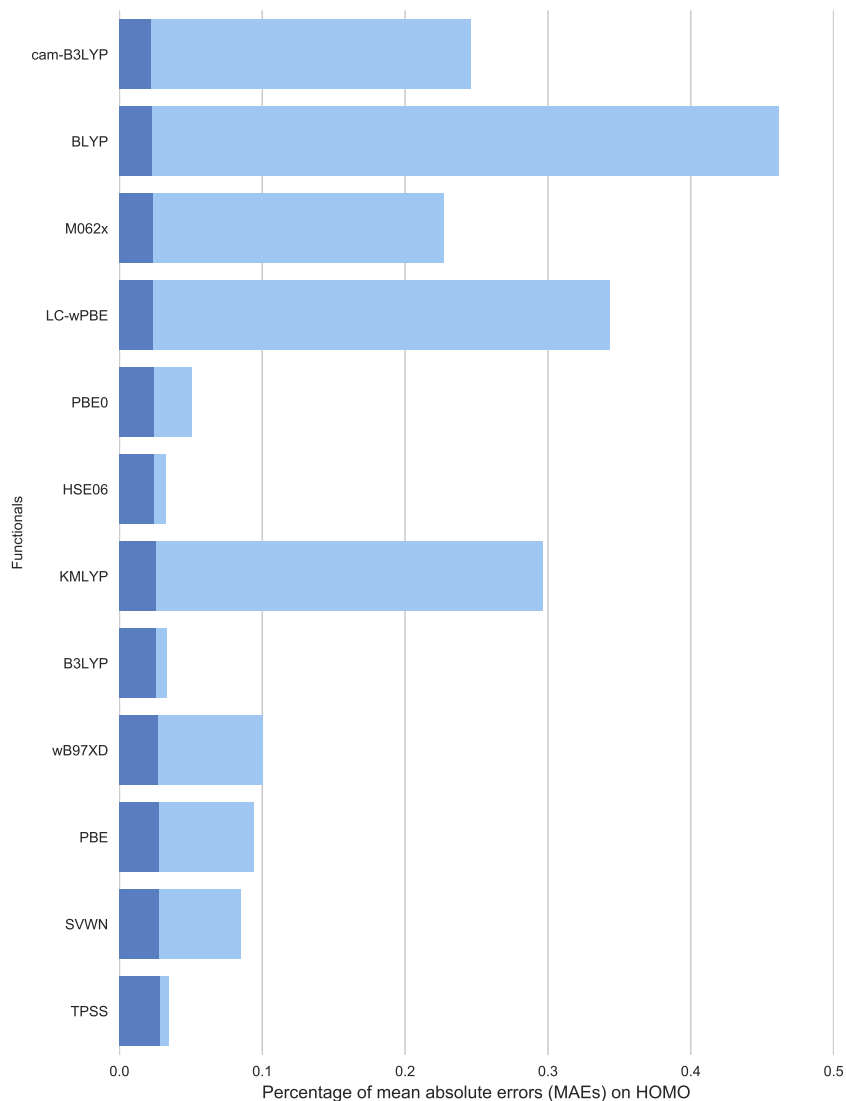
Figure 2.12 show the MSEs from k-fold cross validation on HOMO values (refer to the method section for the procedures of k-fold cross validation). As shown in Figure 2.12, the MSE from LC- $\omega$ PBE was the smallest. From the principle of k-fold cross validation, this means a linear regression applied to LC- $\omega$ PBE functional shows the most predicting power. From the figure, there are two jumps in the MSE values, which essentially divided the functionals into 3 groups. Functionals that fall into those three groups have different features.

The first group is led by LC- $\omega$ PBE and contains cam-B3LYP, M06-2x and  $\omega$ B97XD. This group mostly consists of RSHs, and M06-2x that is from the so-called Minnesota functional series. From Table 2.2, those functionals all have a large percentage ( $> 50\%$ ) of HF. The second group consists of PBE0, HSE06, KMLYP and B3LYP. Those are mostly global hybrid functionals with a certain percentage of Hartree-Fock (HF) exchange component. However, the percentage of HF contained in those functionals is less than 50% for GGA based functionals (PBE0, B3LYP and HSE06). KMLYP, the only exception, has 55.7% of HF component. The third group, having even larger MSE, are from pure functionals: TPSS, PBE, SVWN and BLYP. TPSS is a meta-GGA functional that has the least MSE of the group, while PBE, a GGA functional, follows it. The performance of SVWN and BLYP are nearly the same.



**Figure 2.12:** Average MSEs (Mean Square Errors) of predictions for HOMO values on validation set using 8-fold cross validation.





**Figure 2.13:** Percentage of MAEs (Mean Absolute Errors) on HOMO before (light blue) and after (dark blue) a linear correction. The 0.1 in x-axis represents 10%.

Figure 2.13 shows the mean absolute errors (MAEs) between experimental values and predictions before (light blue bars) and after (dark blue bars) applying a linear correction. Some of the functionals show a big difference in the two cases, while others

stay nearly the same. The functionals in the figure are ordered by the MAEs after applying a linear correction (dark blue). From the figure, linear corrections applied to those results greatly improve the predictions.

The MAEs before applying the linear correction still fall into three groups. Functionals having a heavy percentage of HF exchange ( $> 50\%$ ): cam-B3LYP, LC- $\omega$ PBE, M06-2x,  $\omega$ B97XD and KMLYP show a bigger MAEs among the values in the graph. Other functionals, like TPSS, PBE and BLYP, show medium errors. They are all pure functionals. The rest of the functionals (PBE0, HSE06, B3LYP, SVWN) behave the best, having a deviation less than 0.27eV.

The variation in performance that functionals have tends to be alleviated after a linear correction. The smallest value shown in the figure is 0.12eV from cam-B3LYP and LC- $\omega$ PBE. Then, M06-2x,  $\omega$ B97XD, PBE0 and HSE06 all have a 0.13eV of MAE. The corrected values not only have a high correlation, but they are also more precise. Some of them, like cam-B3LYP, LC- $\omega$ PBE have greater MAEs before the correction, but they are all scaled to the top accuracy with a linear correction. There is no surprise in the results in Figure 2.12 which shows the results from RSH fit to the linear model are the best among the functionals. This means we can always use a linear correction to systematically improve the performance of these functionals.

For HOMO calculation, we have two choices. The first option is using HSE06 without linear correction, since this is the one that has the least error. The second option is choosing RSH or functionals with a heavy HF exchange and after the calculation a linear scaling is needed. Those functionals can be cam-B3LYP and LC- $\omega$ PBE which, according to Figure 2.13, show the least error after scaling.

## Bandgaps

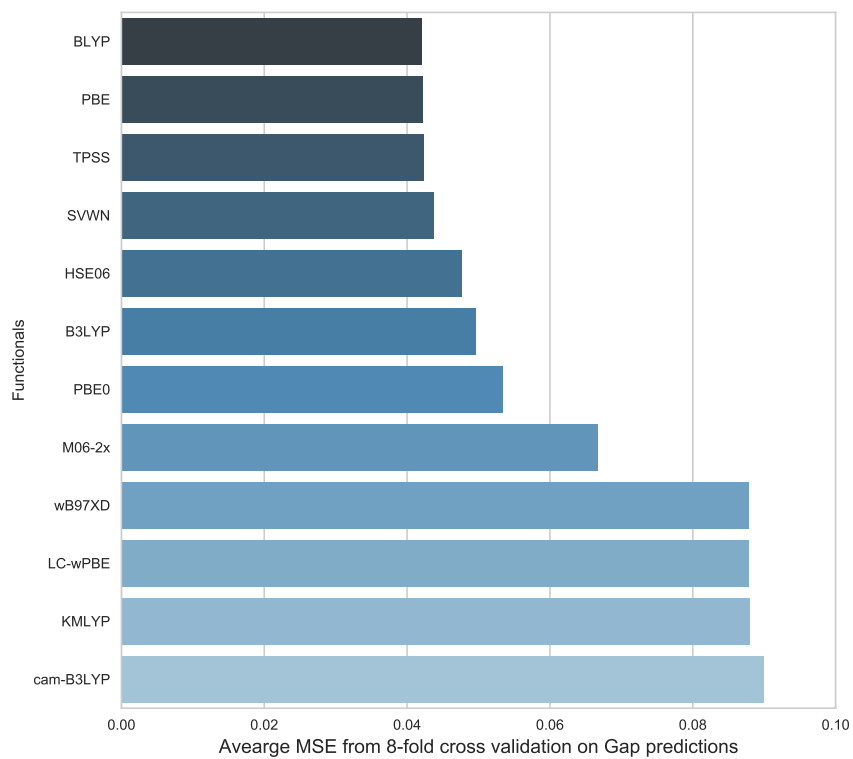
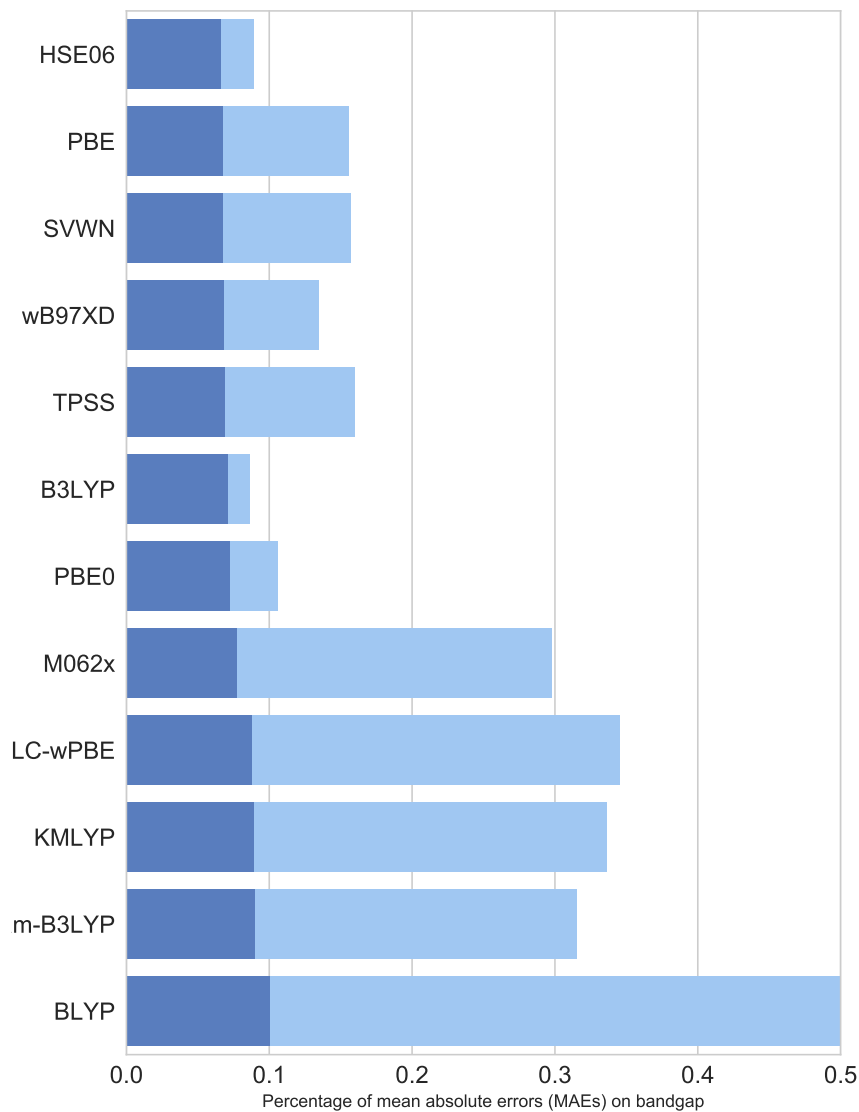


Figure 2.14: Average MSEs (Mean Square Errors) of predictions for bandgap values on validation set using 8-fold cross validation.



**Figure 2.15: Percentage of MAEs (Mean Absolute Errors) on bandgap before (light blue) and after (dark blue) a linear correction. The 0.1 in x-axis represents 10%.**

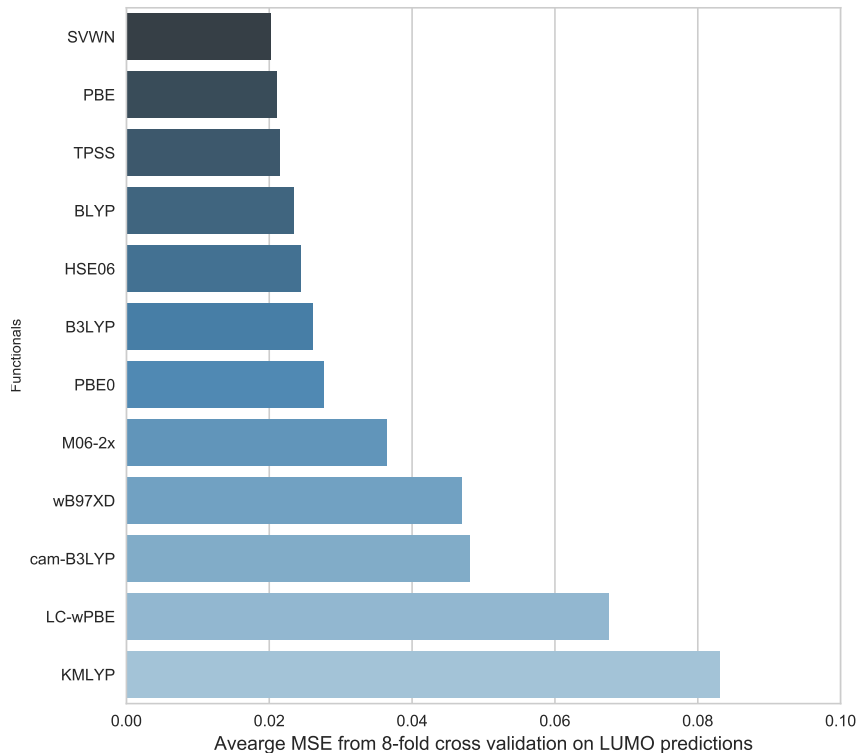
Generally, MSEs for the functionals in Figure 2.14 are all larger than those in Figure 2.12. This fact represents that HOMO results from DFT follow a linear trend better than the bandgap results from TD-DFT. In the graph, all of the pure functionals

(BLYP, PBE, TPSS and SVWN) have similar MSEs and are all smaller than other functionals. This shows that the pure functionals, although having different degrees of approximation, still follow linear trends to a similar degree. In the group of HSE06, B3LYP and PBE0, their MSEs begin to increase but there is hardly any difference. Then the group of RSH and KMLYP form a separate group having the largest MSEs. Similar to Figure 2.12, M06-2x gives a mediocre behavior and its MSE lies between the two groups. This result shows that the validity of a linear model for the data points decreases from the group of pure functionals, to the group of functionals with small HF exchanges, to the group of functionals with heavy HF exchanges and to the group of RSHs. So, as the percentage of HF exchange increases, the linearity of the resulting data points decreases.

The predicting accuracy on bandgap described by the MAEs is shown in Figure 2.15, where the dark blue bars show the errors after scaling with a linear correction and the light blue bars represent the errors before. The functionals are ordered by the MAEs after applying linear corrections. From the figure we see linear corrections applied to the data greatly improve the predictions no matter how much the original deviations from the predictions were. And the light blue bars show, the resulting deviation from PBE, TPSS, HSE06, BLYP, SVWN, B3LYP and PBE0 are either smaller than or close to 0.5eV, with B3LYP, HSE06 and PBE0 being the top. Another group contains M062x, KMLYP,  $\omega$ B97XD, cam-B3LYP and LC- $\omega$ PBE, which show relative larger deviations with LC- $\omega$  being the highest.

For the bandgap calculation, the top three functional choices suggested by the benchmarking results are PBE, TPSS and HSE06 with linear corrections. However, if we planned to perform TD-DFT calculations without linear correction, then HSE06, B3LYP and PBE0 could be the available choices.

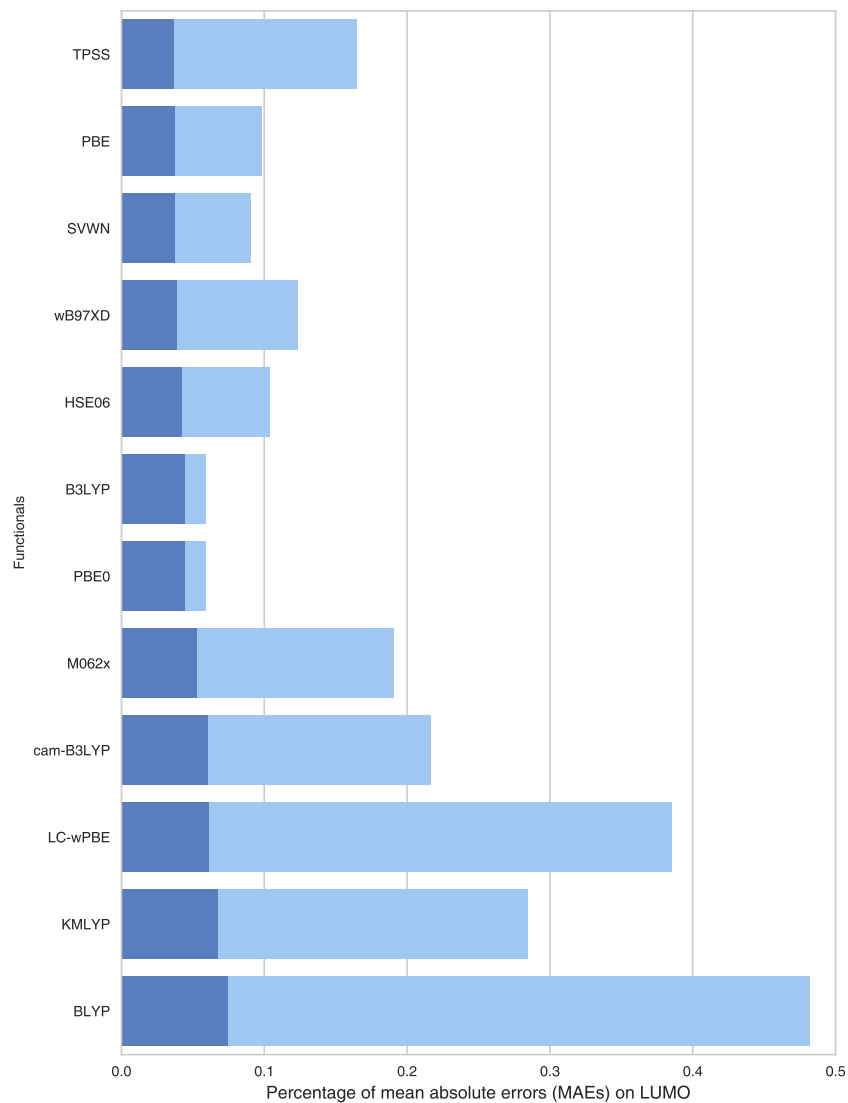
## LUMOs



**Figure 2.16: Average MSEs (Mean Square Errors) of predictions for LUMO values on validation set using 8-fold cross validation.**

As stated, LUMO energies were obtained from the sum of HOMOs (DFT) and the bandgap (TD-DFT). Since the variance in HOMO predictions is relatively small, the variance in LUMO as Figure 2.16 shows, is mostly decided by the variance in bandgaps. So, pure functionals still fit the best into a linear model while RSH or global hybrid with a heavy HF exchange percentage show higher MSEs in the figure.

From the average MSE results in Figure 2.16, the linear correction again shows great potential for scale predictions to a better accuracy. Before the model correction, predictions are diverse in the figure. However, after the correction, the MAEs are all smaller than 0.25eV. The accuracy of scaled results is both decided by the linearity of the prediction results and the performance of the functionals. Because of this, B3LYP performed the best while SVWN, PBE, TPSS, BLYP, HSE06, PBE0 performed almost identically.



**Figure 2.17:** Percentage of MAEs (Mean Absolute Errors) on LUMO before (light blue) and after (dark blue) a linear correction. The 0.1 in x-axis represents 10%.

## 2.4 Conclusion

To select the best functional in predicting HOMO, gap and LUMO energies, we performed a benchmark study on a molecule set containing 29 molecules from the organic electronic field. Various DFT/TD-DFT functionals from nearly all levels were examined. We analyzed the data points both with and without linear corrections and reported corresponding performance of predictions. The k-fold cross validation was used to decide the validity of applying a linear model to the results from different functionals. It was found that in HOMO results from DFT, RSH and M06-2x are more linear than other functionals. And results from GGA and mGGA (BLYP, PBE, TPSS) were shown to be more linear in TD-DFT calculation.

Excited state TD-DFT calculations are based on ground state DFT calculations. According to the results and analysis above, the performance of HSE06 is the top choice for HOMO and is also decent for bandgap predictions. Therefore it is recommended for both HOMO and bandgap calculations if a linear scaling will not be performed. However, if linear corrections will be added, then PBE and TPSS are recommended for the bandgap calculation while cam-B3LYP and LC- $\omega$ PBE are recommended for HOMO calculation.



## CHAPTER 3

### DFT/TD-DFT Investigations on Bio-inspired Melanin Molecules for Organic Photovoltaic Applications

“I do not think that there is any other quality so essential to success of any kind as the quality of perseverance. It overcomes almost everything, even nature.”

— John D. Rockefeller

#### 3.1 Introduction

As we are stretching natural resources to their limits by depleting fossil fuel energies to support our civilization, solar energy harvested by organic photovoltaic (OPV) materials is one of the promising candidates in the concept of renewable energies. Although only in recent years is power conversion efficiency (PCE) for OPV several times higher than it used to be, we have learned a lot in material design rules and fabrication techniques, as reviewed by Nelson (2003), Li *et al.*(2012), Facchetti *et al.*(2013) and Liu *et al.*(2016). These efforts helped achieve new records in PCE exceeding 11% (Baran *et al.* 2017) (Gasparini *et al.* 2017) (Zhao *et al.* 2016a) (Deng *et al.* 2016). Needless to say, better materials still could be discovered if we finished the desperate job of searching over the chemical space, part of which has been estimated to contain ”drug-like” molecules in the order of  $10^{60}$  (Reymond 2015). However, because harvesting solar energy is the first topic for various life forms on earth, evolution over billions of years may have already fostered their reinforced materials to tackle this difficult problem. Adopting a bio-inspired analog into OPV material could potentially boost PCE to even higher values.

Although new types of solar cells have been suggested in the last few decades, in this article we concentrated on the solution-processed OPV with bulk heterojunction (BHJ). BHJ cells utilize two different kinds of molecules, a donor and an acceptor, to form a heterojunction. Once excited by photons, excitons will be generated and

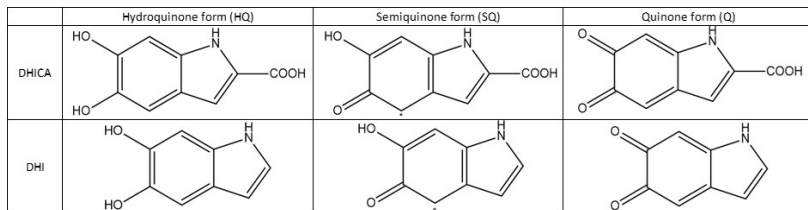
diffuse randomly, and split into electrons and holes at the interface of donors and acceptors (Nelson 2003). Currently, acceptors built from fullerenes and their derivatives are prevailing but non-fullerene acceptors are also under active development (Zhao et al. 2016b). However, in this article, we still assume the acceptor is a fullerene derivative ( $PC_{61}BM$ ) and will focus on the design of the donor. In addition to having the same advantages as the polymer OPV, like being light-weight, flexible, and cheaper, the small molecule OPV has better defined structures and better controlled molecular weight, and less batch to batch variation compared to their polymer counterpart (Roncali, Leriche, and Blanchard 2014) (Collins et al. 2017). In the past few years of development, PCE of solution-processed OPV utilizing small molecule donors achieved 11% (Zhao et al. 2016a) (Deng et al. 2016).

To build bio-inspired small molecule donors, we still followed the design technique in organic electronic material called push-pull design (also called donor-acceptor design, or D-A), which refers to introducing moieties of electron donating and electron withdrawing groups into one molecule core (Jemison and McCullough 2014) (Bureš 2014) (Duan, Huang, and Cao 2012). It has already been proved effective in material design (Bureš 2014) (Jemison and McCullough 2014) (Guo, Baumgarten, and Müllen 2013). Molecules having such a structure usually pose intramolecular charge transfer (ICT) because of the interaction between the electron donating (donor) and the electron withdrawing (acceptor) moieties. Utilizing this interaction, some important properties of the core, such as the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), HOMO-LUMO gap and optical properties can be tuned rationally.

People have explored and discovered many excellent molecules used as the core for small molecule OPV applications. For example, using Benzodithiophene (BDT) as the core moiety to build donors (Yao et al. 2016) (Collins et al. 2017) for small molecule OPV have achieved PCEs around 9% (Kan et al. 2014) (Cui et al. 2015). Another example is diketopyrrolopyrrole (DPP). Unlike BDT, DPP is a widely used electron withdrawing unit that is mostly used to build non-fullerene acceptor. However, it can also be used as the acceptor moiety in the donor (Le Borgne et al. 2016) (Cortizo-Lacalle et al. 2014). From those core moieties, we can conclude that the common design elements for choosing a good core are fused heteroatom ring, quinoid structure, etc (Liu et al. 2016).

Interestingly, melanin, as used by many creatures to protect themselves from sunlight, satisfies several design elements, like quinoid structure as shown in Fig-

ure 3.1 (d’Ischia, Napolitano, and Pezzella 2011). In addition to preferred structure, melanin also has several unique properties, like a broad absorption in UV and visible range (Kollias and Baqer 1987) (Meng and Kaxiras 2008), viable conductivity with different hydration level (Wünsche et al. 2013). All those properties are favored by some opto-electronic and biomedical applications and attract intensive studies from both experimental and theoretical points of view (Manini et al. 2015) (d’Ischia et al. 2009). However, one of the drawbacks of natural melanin lies in its very low solubility. To better serve our purpose of using melanin as the solution-processed small molecule donor, we therefore have to modify its structure to tune its property toward our requirement. More specifically, donor and acceptor moieties in a push-pull molecules can be arranged differently to form many structures fundamental property tuning (Bureš 2014).



**Figure 3.1: Structures of tautomer forms of eumelanin building blocks. DHI is for Dihydroxyl indole and DHICA is for Dihydroxyl indole carboxylic acid. DHI and DHICA are both monomers for eumelanin.**

In this study, we investigated several eumelanin based novel donor molecules with different combinations of electron donating/ electron withdrawing end-capping moieties. Using DFT/TD-DFT calculations, we characterized the related properties for their potential OPV applications. After finding the related rules, we also provide virtual molecules for further experimental work.

### 3.1.1 Computational methods

Generally, one of parameters to characterize photovoltaic property is power conversion efficiency (PCE), which is defined as (Nelson 2003):

$$\eta = \frac{J_{SC} V_{OC} F}{P_{in}} \quad (3.1)$$

where  $J_{SC}$  is the short-circuit current,  $V_{OC}$  is the open-circuit voltage,  $F$  is the filling factor and its typical value to be used is 0.65,  $P_{in}$  is the input power to a solar

cell, which under the standard AM 1.5 solar radiation spectra would take the value of  $1000 \text{ W/m}^2$ .

To theoretically predict PCE for new molecules, we need to have  $J_{SC}$ ,  $V_{OC}$  and  $FF$  for corresponding molecules.  $J_{SC}$  can be calculated from integration of covered solar spectra that determined by optical bandgap and follows the formula(Nelson 2003) (Bérubé et al. 2013):

$$J_{SC} = Q_{external} \int F(\omega_{abs}) d\omega_{abs} \quad (3.2)$$

where  $F(\omega_{abs})$  is the photon flux in the solar spectrum absorbed by certain molecule,  $Q_{external}$  is the external quantum efficiency and the typical value to be used is 0.65.

Second,  $V_{OC}$  can be calculated according to the Scharber Model (Scharber et al. 2006):

$$V_{OC} = \frac{1}{e} (|E_{DONOR}^{HOMO}| - |E_{PC_{61}BM}^{LUMO}|) - 0.3V \quad (3.3)$$

where  $e$  is the elementary charge,  $E_{DONOR}^{HOMO}$  is the HOMO energy level of donors,  $E_{PC_{61}BM}^{LUMO}$  is the LUMO energy level of  $PC_{61}BM$  that was assumed to be the acceptor,

From the theoretical point of view, not is only a narrow bandgap required for a good  $J_{SC}$ , but also energies of frontier orbitals of the donor molecule have to be aligned well, as what the Scharber Model concluded, with  $PC_{61}BM$ , in order to have a good  $V_{OC}$ . Aside from the effect of  $FF$  and quantum efficiency, the performance of solar cells was decided by energy levels of frontier orbitals of donor:acceptor blends. So accurate calculations on HOMOs and LUMOs combined with Equations 3.2, 3.3 and 3.1 will give good predictions on PCE of OPV materials.

To calculate the HOMOs and LUMOs accurately, we carried out a series of DFT/ TD-DFT calculations using the Gaussian09 software package (*Gaussian09 Revision C.01*) with B3LYP and HSE06 ( according to the benchmarking result from chp2) functional at 6-311+G(2d,p) level. For PCE prediction, HSE 06 were applied to first acquire the primary energies of frontier orbitals and then applying the scaling formula concluded from the previous protocol from chapter 2. To be specific, HOMO energies were calculated from optimized ground state geometry which is also used as a starting input structures for TD-DFT calculations. After performing excited state geometry optimizations, bandgap energies were acquired from TD-DFT. Finally, LUMO energies were calculated from HOMO and bandgap results. For general geometry

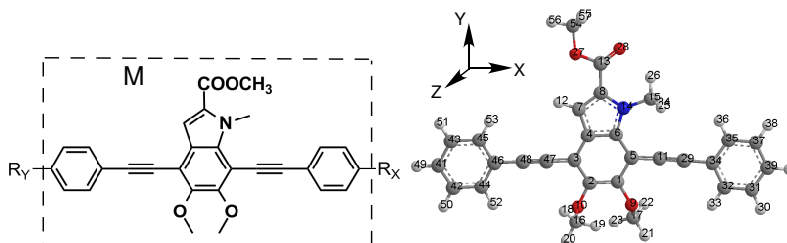
investigation, specifically the dihedral angle comparison, B3LYP was used. This is because it has a better accuracy in describing geometry.

The virtual molecule set was constructed based on a 5,6-dihydroxyindole-2-carboxylic acid molecule (DHICA) that is represented by "M" in this study. Figure 3.2 showed the basic structure of molecules we constructed and side group variations were shown in Figure 3.3. According to the list, we applied groups with different electronic properties (either electron donating or electron withdrawing). This push-pull design or D-A design would give us the ability to better control molecule's energy level and had been widely used in organic solar cell material design (Esteban et al. 2011) (Duan, Huang, and Cao 2012).








### 3.1.2 Results and discussions


We will first look into the geometry structural results of the designed molecules and explore how those results varied with different combinations of side groups. Then we will discuss the metrics of photovoltaic property which is PCE. After locating the best performed molecules in the combination and the corresponding rule, our focus will shift to the electronic structures leading to these results. From the structure-property relationship, we will then try to summarize the structure basis for a high PCE and further suggest several new virtual molecules that may have a better PCE.

#### Ground state geometry and electronic structure




**Figure 3.2:** Left: The structure of the molecules designed; Right: The 3D model of the molecule with "H" as the side group and showing the label of each atom.

Number	Side Groups	Structures	$\sigma_p$	$\sigma^i$
1	-NO <sub>2</sub>		0.78	0.606
2	-CN		0.66	0.525
3	-F		-	-
4	-H		-	-
5	-t-butyl		-	-
6	-OMe		-0.27	0.220
7	-NMe <sub>2</sub>		-0.83	0.089



Stronger  
Electron  
Withdrawing  
Ability



Stronger  
Electron  
Donating  
Ability

**Figure 3.3: Donor and acceptor moieties used in this study. Values of the Hammett constant  $\sigma_p$  are from (Hansch, Leo, and Taft 1991) ; and values of Pytela  $\sigma^i$  constant are from (Pytela 1996).**

The molecules designed in this series have three ring structures: a heteroatom ring (DHICA) and two benzene rings. The overall planar structure is determined by the two dihedral angles formed between each benzene ring and the DHICA ring, or the two dihedral angles that formed between the red ring and the yellow ring, and the red ring and the blue ring (as shown in Figure 3.4), so the values of the dihedral angles are important geometry parameters. If the absolute values of the two dihedral angles are small, then it will lead to a better planar configuration. From Table 3.5, dihedral angles vary with different side groups. In Table 3.5, the values for two dihedral angle were shown for different side group combinations: groups in pink are electron withdrawing groups (EWG or A) and groups in blue are electron donating groups (EDG or D). From the table, all of the molecules in the collection were having nearly planar structures allowing for certain fluctuation. However, the dihedral angle on the right are generally larger than the dihedral angle on the left. The steric hindrance from a methyl group at 1-position of DHICA ring contributes to this value shift.

A planar structure has the advantage of having a good exciton diffusion and hence a better PCE (Liu et al. 2016). If bonds are in similar symmetry, for example, neighboring atom formed a planar structure and are in a similar symmetry, the bond

length alternation (BLA) are expected to decrease since all of the bond are similar. Generally speaking, this decrease would narrow the bandgap, which is another advantage for better PCE (Liu et al. 2016).

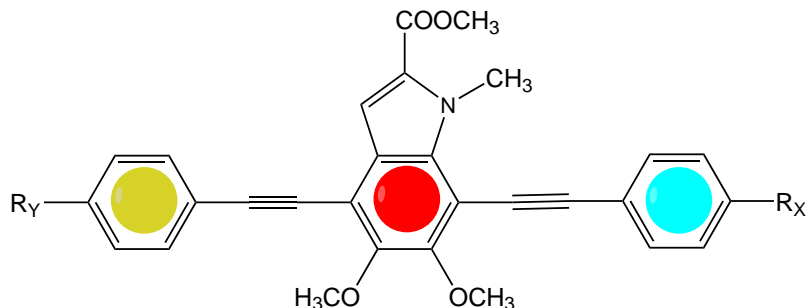


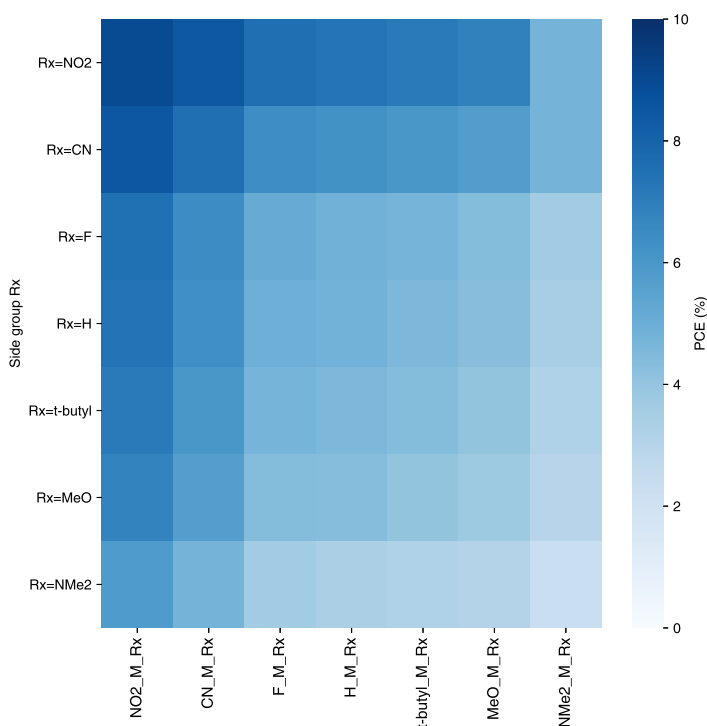
Figure 3.4: Illustration of the 3 rings in the designed molecule. The two dihedral angles are the angle between the yellow ring and the red ring, and the angle between the red ring and the blue ring.

<b>4-3-46-45</b>	NO <sub>2</sub>	CN	F	H	t-butyl	OMe	NMe <sub>2</sub>
NO <sub>2</sub> _M_R <sub>x</sub>	-6.6	-5.7	-4.9	-4.2	-5.6	-5.5	-4.8
CN_M_R <sub>x</sub>	-7.7	-7.4	-6.6	-5.6	-6.7	-6.9	-6.2
F_M_R <sub>x</sub>	-9.1	-9.1	-8.9	-8.3	-9.1	-9.3	-9.0
H_M_R <sub>x</sub>	-8.7	-8.8	-8.7	-8.3	-8.6	-8.5	-8.5
t-butyl_M_R <sub>x</sub>	-6.6	-7.1	-7.4	-7.8	-11.0	-6.6	-6.7
MeO_M_R <sub>x</sub>	-6.2	-5.9	-5.7	-4.7	-6.6	-7.4	-6.7
NMe <sub>2</sub> _M_R <sub>x</sub>	-5.1	-5.5	-6.4	-7.0	-5.9	-6.2	-6.5
<b>6-5-34-35</b>	NO <sub>2</sub>	CN	F	H	t-butyl	OMe	NMe <sub>2</sub>
NO <sub>2</sub> _M_R <sub>x</sub>	-7.6	-9.0	-11.3	-10.8	-11.7	-10.4	-7.5
CN_M_R <sub>x</sub>	-7.0	-8.6	-11.4	-10.8	-8.6	-10.4	-8.2
F_M_R <sub>x</sub>	-5.9	-7.9	-11.2	-10.6	-9.3	-10.6	-9.2
H_M_R <sub>x</sub>	-5.1	-7.1	-10.9	-10.2	-11.7	-10.4	-9.6
t-butyl_M_R <sub>x</sub>	-6.8	-8.1	-11.4	-10.7	-11.9	-11.5	-10.0
MeO_M_R <sub>x</sub>	-5.4	-7.2	-10.9	-10.1	-8.8	-10.9	-9.4
NMe <sub>2</sub> _M_R <sub>x</sub>	-6.3	-7.2	-11.0	-10.3	-7.5	-11.0	-6.5

Figure 3.5: Dihedral angles of the molecules. Top: The dihedral angle between the yellow ring and the red ring, measured following the atom label: 4-3-46-45 ; Bottom: The dihedral angle between the red ring and the blue ring, measured following the atom label: 6-5-34-35.

## PCE prediction

From the Scharber model and scaling formula summarized in chapter 2, PCE values were calculated and shown in the heatmap as Figure 3.6, assuming that  $PC_{61}BM$  is the acceptor molecule used in the bulk heterojunction. In the figure, PCE values varied with different side group combinations, which are represented by different colors. Please refer to the Figure 3.2 for the position of  $R_X$  and  $R_Y$ .



**Figure 3.6:** The heatmap of PCE values, which are varied with different side group combinations. Refer to Figure 3.2 for the position of  $R_X$  and  $R_Y$ .

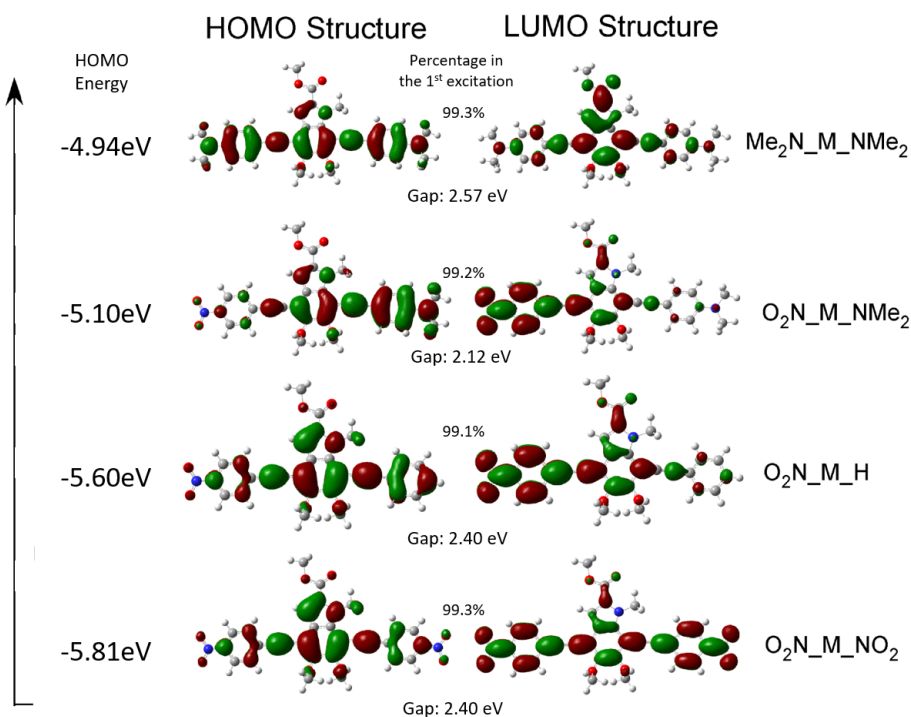
Figure 3.6 can be divided into 4 regions where the configuration of the molecules takes A-Pi-M-Pi-A (top left region), A-Pi-M-Pi-D (top right region), D-Pi-M-Pi-A (bottom left region) and D-Pi-M-Pi-D (bottom right region). The results showed the configuration of A-Pi-M-Pi-A has the best PCE value among all of the combinations and D-Pi-M-Pi-D would be the worst. Since eumelanin core M is rich in electrons, this result essentially means A-Pi-D-Pi-A configuration would be the best one to harvest photon energy. It is interesting to notice that PCE values decrease according to the decrease in electron withdrawing ability of the side group and further decrease ac-



ording to the increase in electron donating ability of the side group. In the following section, we will further look into the electronic structure behind this configuration.

### Isosurface graph of frontier orbitals

To gain more insight into the electronic structures of the molecules, frontier molecular surface of HOMO and LUMO of several typical molecules in the set was calculated and the isosurface graph (isovalue=0.02) is shown in Figure 3.7.



**Figure 3.7:** HOMO and LUMO isosurface graphs (isovalue=0.02) for several typical molecules. HOMO energy values are labeled.

From Figure 3.7, the molecule with  $\text{NO}_2$  (the strongest EWG in the list) at both sides has the deepest HOMO, while the molecule with  $\text{NMe}_2$  at the both sides has the shallowest HOMO. The rest have their HOMOs lying in between. The trend in the graph is, for the molecules in this set, more electron withdrawing group will bring deeper HOMO, and more electron donating groups will raise the HOMO to a shallower level. Comparing HOMO structures among the four molecules, electrons in the molecule with  $\text{NO}_2$  at the both sides span more than in molecules having less  $\text{NO}_2$ . Although electrons in the  $\text{Me}_2\text{N}_M\text{NMe}_2$  also span nearly the same

length horizontally as in the  $O_2N\_M\_NO_2$ , the HOMO isosurface (isovalue=0.02) in  $O_2N\_M\_NO_2$  is more inflated, meaning electrons in the  $O_2N\_M\_NO_2$  occupied a larger space volume. This difference renders  $O_2N\_M\_NO_2$  the deepest HOMO.

The gap values in molecules in Figure 3.7 also followed a certain pattern. The narrowest gap in the graph is from  $O_2N\_M\_NMe_2$ , the molecule that has two different (EDG and EWG) side groups at the two ends. Molecules having the same end-capping groups don't show narrower gaps than others. So the rule here is, to have a relatively narrow gap, two different side groups should be used. Then if we looked into the isosurface structure of those molecules, we would see the HOMO electrons are mostly abundant at the right side of  $O_2N\_M\_NMe_2$ , which is due to the EDG property of  $NMe_2$ , while in the LUMO structure, electrons are abundant at the other side. So this excitation is a typical intra-molecular charge transfer type. Although excitations from other molecules also have different degrees of charge transfer, theirs are much less in the distance and electron density difference. Attaching two different side groups would induce more charge transfer upon excitation, and therefore would result in a narrower gap.

Among the molecules in Figure 3.7, the HOMO to LUMO transition is responsible for more than 99% excitations in the first excitation. For the bottom three molecules, the HOMO electron density in the two oxygen atoms in hydroxyl groups in DHICA has certain electron density from the lone pair electrons, but when those three molecules were promoted to the excited state, the two oxygen atoms in the LUMO surface don't show electron density. The LUMO surface of the bottom three molecules show a typical  $\pi^*$  character. This evidence proves the first transition is  $n \rightarrow \pi^*$  type. For the  $Me_2N\_M\_NMe_2$  molecule, the transition is  $\pi \rightarrow \pi^*$  type, and the transition electrons are from the  $NMe_2$  instead of oxygen atoms in DHICA, which can be seen by comparing the corresponding atoms in HOMO and LUMO surface.

So one feature of D-A design is inducing different degrees of intra-molecular charge transfer that lead to a narrower band gap. Combining with the knowledge on controlling HOMO with EWG group, molecules constructed under this molecule framework (DHICA as the backbone), could fit to different acceptor molecules. The strategies for designing photovoltaic material using this framework (DHICA as the backbone) can be either two suitable EWG groups at both ends, or a suitable EWG group and an EDG group at the ends. Which strategy to use would depend on the acceptor molecule used in the bulk heterojunction, because the energy levels here also have to match the acceptor molecules's energy levels according to the Scherbar

Model.

### New virtual molecule design

From the strategies discovered, we further suggest putting more EWG groups at both sides to induce even deeper HOMO, or EWG and EDG at different sides for an even narrower gap. The isosurface of HOMO and LUMO for new virtual molecules was shown in Figure 3.8. A further idea is to remove the hydroxyl groups at the DHICA to eliminate  $n \rightarrow \pi^*$  transition, therefore enabling absorptions in longer wavelength.

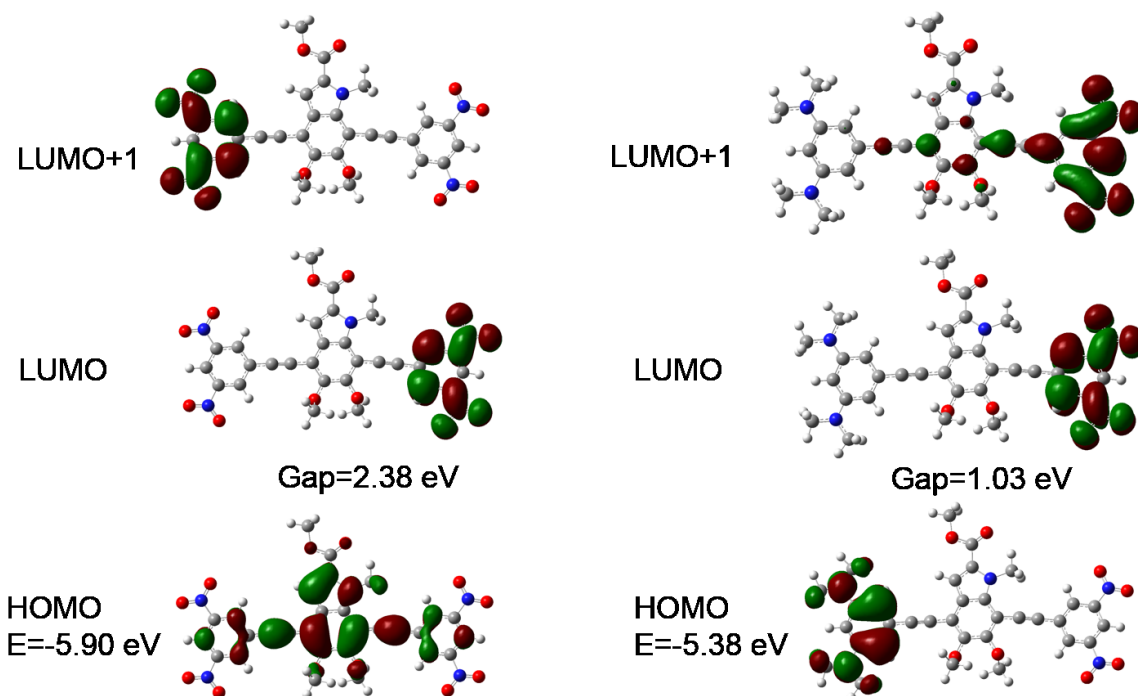


Figure 3.8: Scaled HOMO energies and gap from DFT predictions of two virtual molecules. The HOMO and LUMO isosurface graphs are of isovalue=0.02.

### 3.2 Conclusion

In this study, different side groups with a DHICA core were investigated. Based on the benchmarking study before (chapter2), we scaled predictions from DFT calculations to calculate PCE using the Scharber Model. The best predicted value appeared in A-Pi-M-Pi-A configuration and its value was around 9%. In addition, several design principles for this molecule framework were summarized. It was found that, by changing the side groups into different EWG groups, the designed molecules could show a

systematical change in HOMO energy, which could fit the energy level alignment requirement of different acceptors (Fullerene and beyond) in bulk heterojunction. The gap could be controlled by using different EWG and EDG groups at different ends. Finally, new virtual molecules constructed based on these principles were calculated and properties predicted.

## CHAPTER 4

### High Throughput Screening Using Correlation Ratio for Tandem Solar Cells Design

“We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win, and the others, too.”

— John F. Kennedy

#### 4.1 Introduction

To conquer both the storage and pollution problems from fossil fuel while still providing enough support to the development of human civilization, solar energy is one of the best candidates to replace the fossil fuel energy. Therefore, more researchers devote their efforts to this field (Chu, Cui, and Liu 2017) (Louwen et al. 2016). Bearing the properties of being flexible, lightweight, low in cost and easy to manufacture (Dou et al. 2013) (Denkler, Scharber, and Brabec 2009) (Heeger 2014), organic photovoltaic (OPV) is advantageous over conventional inorganic photovoltaic and is under rapid development in the past few years. Particularly, because of the simple and defined product in the reaction they evolved, the beauty of small organic molecules can gradually be recognized (Chen, Wan, and Long 2013) (Roncali, Leriche, and Blanchard 2014) (Collins et al. 2017).

Despite the recent exciting race in the single junction small molecule OPV that already exceeds 11% (Baran et al. 2017) (Gasparini et al. 2017) (Zhao et al. 2016a) (Deng et al. 2016), detailed analysis pointed out two limiting factors for improving PCE. On one hand, PCE of single junction cells was limited by their thermodynamic

limit or the so-called Shockley-Queisser limit (S-Q limit) (Shockley and Queisser 1961b); on the other hand, PCE of the current OPV cells suffered from thermalization of hot charge carriers that led to a low open-circuit voltage (Scharber 2016). To conquer these two limiting factors, a tandem structure solar cell was first put forward and made by Hiramoto *et al.*(1990). An illustration of a tandem structure solar cell is shown in Figure 4.1. By using more than one junction, S-Q limit is no longer applied to tandem solar cells. In addition, single junction OPV cells usually have one kind of acceptor molecule and, therefore, a donor molecule has to choose between having a narrow bandgap or a deep HOMO, because the two choices can't be both satisfied in certain acceptor molecules (Scharber *et al.* 2006). However, if tandem structure is used, more kinds of acceptor molecules could be used in different layers (Li *et al.* 2013).

Experimentally, among the tandem devices that have shown up so far, a device from You *et al.*(2013) in 2013 is the first one to reach 10%. The double-junction cell they constructed used P3HT:ICBA as the bottom layer and PDTP-DFBT:PCBM as the top layer and the spectra of two layers, split apart around  $630nm$ , have a little overlap. In 2014, another device from Chen *et al.*(Chen *et al.* 2014) gave a PCE of 11.55% using a triple-junction structure and a PCE of 10.70% using a double-junction structure. The absorption spectra in the triple-junction cell split around  $600nm$  and  $730nm$ ; whereas the spectra in the double-junction cell split around  $700nm$ . In the tandem device constructed by Zheng *et al.*(2016), a double-junction structure delivered a PCE of  $11.32 \pm 0.26$ . The spectra of the two molecules split around  $690nm$ . Theoretically, from the model given by Dennler *et al.*(2008), 15% PCE could be expected but there are only a couple of double-junction and triple-junction tandem solar cells that could go beyond 11% (Zhou *et al.* 2015) (Zhang *et al.* 2016) (Zheng *et al.* 2016) (Chen *et al.* 2014). In the work from Li *et al.*(Li *et al.* 2013), they built several different single junction and multi-junction OPV cells and compared the trends from experimental results with the prediction by Dennler *et al.*(2008). Their conclusion was that a PCE beyond 14% was practical. So PCE of tandem device could still march forward if proper molecules could be found.

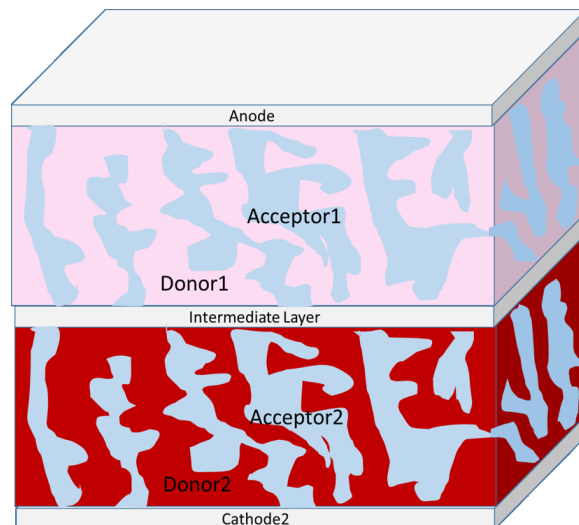
One of the most efficient ways of finding molecules with certain constrains, proved by drug design in the last decade (Macarron *et al.* 2011) (Bajorath 2002), is the high throughput screening (HTS) (Curtarolo *et al.* 2013a). Combining with virtual design and electronic structure calculation, HTS continues to serve in the material information field and is adopted by many modern material projects. Harvard Clean Ener-

gy project (Hachmann et al. 2011) (Hachmann et al. 2014) constructed a database (CEPDB) of 2.3 million virtual molecules that are built from 26 building blocks summarized from experimental work. After performing a large scale of DFT calculations with the help of the IBM World Community Grid, they calculated HOMOs and LUMOs as well as PCEs of those molecules and ranked all of virtual molecules according to their PCE. This result predicts many top performing virtual molecules to guide the experimental synthesis in the future. Another example is a survey study given by Ørnsø *et al.* (Ørnsø, Garcia-Lastra, and Thygesen 2013), which focused on the dye selections and screened out 50 candidates from 1029 porphyrin derivatives. DFT calculations with PBE functional are used to predict the HOMO and LUMO orbital energy. Their results are online and open to public. Kanal *et al.* (Kanal et al. 2013b) adopted a slightly different way for the results analysis. Their aims were not only to find out the candidates but also the trend and the design rules underlying their results. Unlike the combination of Donor-Acceptor pattern, they proposed a new type of Donor-Donor motif pattern which is concluded from their findings.

The goal of this study is to efficiently find out new donor molecules satisfying spectrum requirements for tandem solar cells through a large scale computational screening. Therefore, we applied Time-Dependent Density Functional Theory (TD-DFT) and HTS to calculate the spectra of the top 1000 OPV molecules in CEPDB and used a designed descriptor, correlation ratio (CR), to describe and compare their spectra coverage. Molecules with complementary spectra could be selected out, and those molecules would be potentially used in a tandem device. Further more, this descriptor is not confined to be used in OPV field but any requirements posed to the molecule's spectrum can take advantage of it, for example, sunlight screening products.

## 4.2 Computational Method

The flowchart of screening is shown in the Figure 4.2. Starting from the top, we used the molecules from CEPDB as the diverse library and the first step was calculating their spectrum by performing a ground state geometry optimization (DFT in G09) and excited state absorption spectrum calculation (TDDFT in Octopus). After dividing the whole spectrum into 26 regions, we calculated the correlation ratio in each region. The purpose of 26 region division is to setup a set of "resolution grids" for the spectrum. Then, based on certain score function, screening process were performed.



**Figure 4.1: Illustration of a tandem solar cell with a 2-layer structure.**

#### 4.2.1 The ground state geometry calculations using Gaussian09.

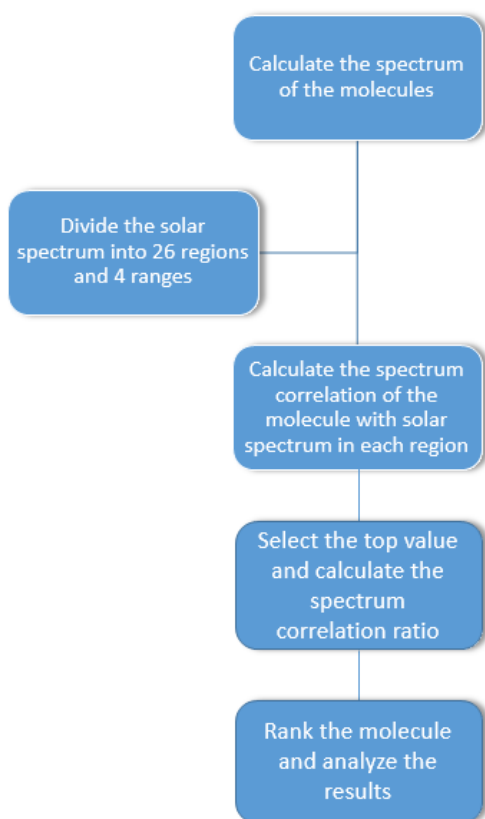
The molecules were retrieved from the SMILES codes given in the website of the CEPDB database and the pybel library provided by the Open Babel software (Boyle et al. 2011) was used to convert the SMILES codes to the corresponding molecules. Ground state geometries of all the compounds were fully optimized under B3LYP functionals with 6-31G++(d,p) level basis set using Gaussian 09 software (*Gaussian09 Revision C.01*). The optimal geometry was also verified by performing vibration analysis.

#### 4.2.2 The excited state calculation using octopus.

The optical absorption was calculated from the Time Dependent Density Functional Theory (TDDFT) as implemented in the octopus software (Marques 2003)(Castro et al. 2006), which uses real time TDDFT (RT-TDDFT). For each of the molecules, octopus calculation applied a disturbance to its electronic ground state structure acquired from the Gaussian calculation and the system described by the LDA functional then evolved for a total time of 25 fs. The molecule was placed in the real space grid with a universal 0.20 Å interval and a simulation box of typical value in Octopus was used around each atom. To allow a stable propagation, a 0.0025 fs time step was used.

Disturbance in three orthogonal directions was applied in the Cartesian coordinates independently and then the response (variation of the multipole strength,  $S(t)$ ) of the molecular system from the three directions was collected. From the Fourier





**Figure 4.2:** The basic process of calculating the spectrum correlation ratio. The largest value (scale molecule) of spectrum correlation is used to normalize the data in each region so that spectrum correlation ratios in that region are acquired for every other molecule. For future extension or any changes, we only need to recalculate the scale molecules and the recasts would become comparable with the previous molecules in the database.

Transform theory, the response in frequency domain, which results in an optical absorption spectrum in one direction of the Cartesian coordinate, would then be given by the following:

$$A(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} S(t)dt \quad (4.1)$$

The total optical absorption is calculated as the average of the three individual absorptions from the different directions in the Cartesian coordinate.

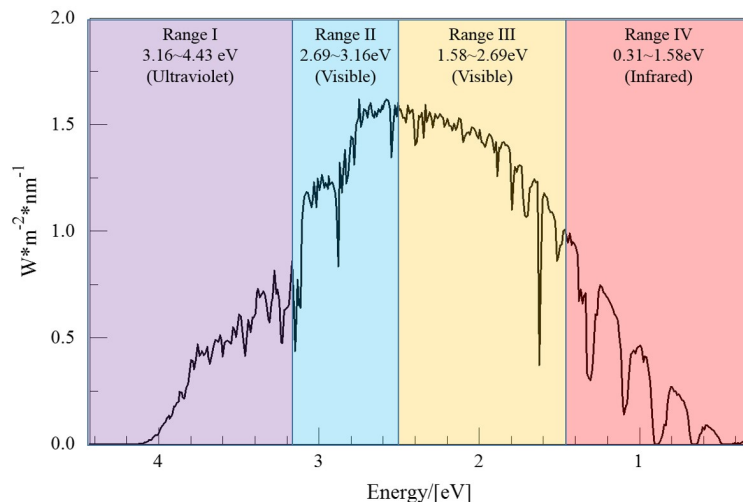
### 4.2.3 The spectrum calculation using compressed sensing technique.

To mathematically find out the corresponding spectrum from the calculation data, we used compressed sensing technique. Compressed Sensing (CS) is a mathematical technique used as a data analysis tool in many fields (Davenport et al. 2012). The major advantage is that it usually needs far less number of samples than what is normally required by traditional Fourier Transform method. Octopus also implemented the Compressed Sensing technique in one of its post calculation processing tools (Andrade, Sanders, and Aspuru-Guzik 2012). It has been shown (Andrade, Sanders, and Aspuru-Guzik 2012) for a spectrum having similar resolution, CS requires far less data points from the time dependent calculations, meaning far less simulation time, which is a big saving of the computer CPU time in high throughput investigations.

### 4.2.4 The spectra correlation ratio.

To quantify the coverage of the solar spectra from a given molecule, we designed the parameter – spectra correlation ratio. Because the higher absorption from the molecule in the given region, the higher spectrum coverage the molecule will have in that region. Therefore the parameter is calculated by taking the correlation of the absorption spectrum from the molecule and the solar spectrum in that region. Then, a higher correlation value would lead to a better coverage of the spectrum in that region. The defined regions and ranges in the paper are shown in Figure 4.3 and Table 4.1.

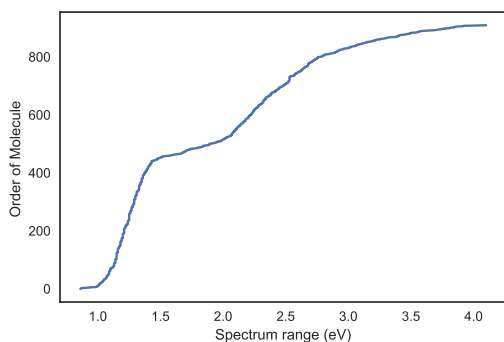
Considering the future extension of the database, we select the top molecules (scaling molecules) in each spectrum regions and use their values to normalize the spectrum correlation, such that in each regions the top molecule would have a spectrum correlation ratio of “1” whereas the value of the rest of the molecules would be given by the ratio with the top molecule. Any systematic changes in the methods or



**Figure 4.3: AM 1.5 solar spectrum.** The 4 colors represents the 4 ranges that will used to benchmark the spectrum performance of the molecules.

Regions	1	2	3	4	5	6
Energy(eV)	4.43-4.26	4.27-4.10	4.11-3.94	3.95-3.78	3.79-3.63	3.64-3.47
Regions	7	8	9	10	11	12
Energy(eV)	3.48-3.31	3.32-3.15	3.16-2.99	3.00-2.84	2.85-2.68	2.69-2.52
Regions	13	14	15	16	17	18
Energy(eV)	2.53-2.36	2.37-2.20	2.21-2.05	2.06-1.89	1.90-1.73	1.74-1.57
Regions	19	20	21	22	23	24
Energy(eV)	1.58-1.41	1.42-1.26	1.27-1.10	1.11-0.94	0.95-0.78	0.79-0.63
Regions	25	26				
Energy(eV)	0.62-0.47	0.48-0.31				

**Table 4.1: The division scheme of the 26 regions of solar spectrum.** The values of correlation ratio were first calculated in each of these regions. Depending on the different ranking tasks as discussed in the “Results” section, the correlation ratio could be summed up to represent certain performances in the different spectrum ranges as labeled in Figure 4.3.



**Figure 4.4: The distributions of the first peaks of all of the molecules. The order of the molecules was represented by their IDs in CEPDB.**

any extensions from the researchers elsewhere can be accounted for by recalculating only the scale molecules in each regions and then the new values would be comparable with the rest of data.

### 4.3 Results and Discussions

The results listed in the Figure 4.4 is the correlation ratio of the top 1000 molecules in the CEPDB and they are expanded along the horizontal axis. The different colors in the graph represent the values of the correlation ratio in different regions. As shown from the figure, one molecule usually has quite different correlation ratios in the different regions. This is mainly due to their different optical activities. The molecules having the correlation ratio of 1.00 in certain region are the ones in the CEPDB that are the best correlated with solar spectrum in that region.

In this section, we are going to discuss some of results from our screening with different criteria.

#### 4.3.1 The performance of the top ones in one region

Because the threshold of CR is set to 0.7 for top candidates, some regions have more molecules than other regions. This can be vividly seen from the length of the deep red color. For example, region 14 has an extraordinarily longer red column than region 1 or 2. This shows there are more molecules in region 14 that have good performance than those in region 1 or 2.4

There is a cross-over point in region 20 and region 21. Then the data points separate symmetrically to each ends forming an "X" shape. This shows the molecules

Table 4.2: Top performance molecules ID (from the CEPDB database) in each regions. The red character means the molecule shown up in more than one region. And the blue bars in the cell show their values of the spectrum correlation ratio.

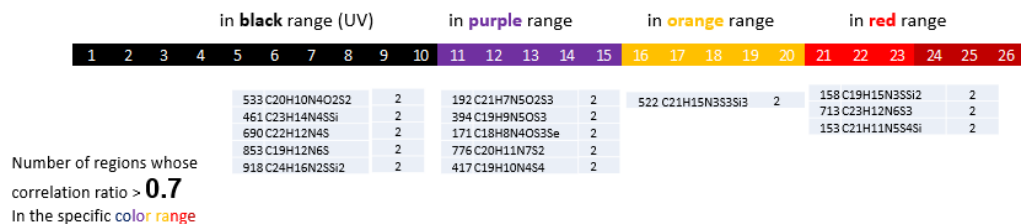
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Region 1	447	285	556	567	765	219	690	839	574	460
Region 2	804	63	303	130	806	580	496	240	807	866
Region 3	842	19	986	123	16	412	283	21	97	162
Region 4	853	871	690	586	52	667	199	501	22	167
Region 5	918	664	223	898	868	111	76	943	530	371
Region 6	461	533	643	219	162	878	556	249	845	834
Region 7	5	575	381	474	260	211	763	342	236	438
Region 8	970	160	146	880	787	341	802	734	121	345
Region 9	776	394	737	627	488	629	214	275	900	208
Region 10	171	657	18	991	827	136	969	144	33	884
Region 11	43	316	305	560	833	132	512	495	192	757
Region 12	794	289	846	958	498	748	623	891	769	922
Region 13	677	243	755	291	140	446	504	427	324	326
Region 14	882	761	404	126	959	687	106	101	952	536
Region 15	522	612	296	460	459	407	270	841	93	941
Region 16	245	143	643	855	100	54	857	898	377	514
Region 17	862	158	817	488	153	600	717	906	798	691
Region 18	104	818	183	657	108	810	339	275	970	963
Region 19	532	279	444	997	571	353	126	651	952	33
Region 20	900	243	329	981	884	321	653	713	143	148
Region 21	791	50	502	219	252	1	901	138	134	640
Region 22	596	502	147	18	85	564	70	881	805	229
Region 23	596	18	881	85	572	147	564	910	536	502
Region 24	85	572	596	18	536	910	564	156	291	504
Region 25	596	85	66	572	175	186	881	841	296	643
Region 26	572	85	536	656	386	675	750	186	949	834

that have the top performance in region 15-20, also tend to have a good performance in region 21-25.

Using the spectrum correlation ratio, we could easily check the alignment with the solar spectrum. From the definition, a larger value in a region would represent a better resemblance with the solar spectrum in that specific region. We therefore could rank the alignment with the solar spectrum in different regions, as listed in Table 4.2.

#### 4.3.2 Aiming at a better alignment with the solar spectrum (across the regions)

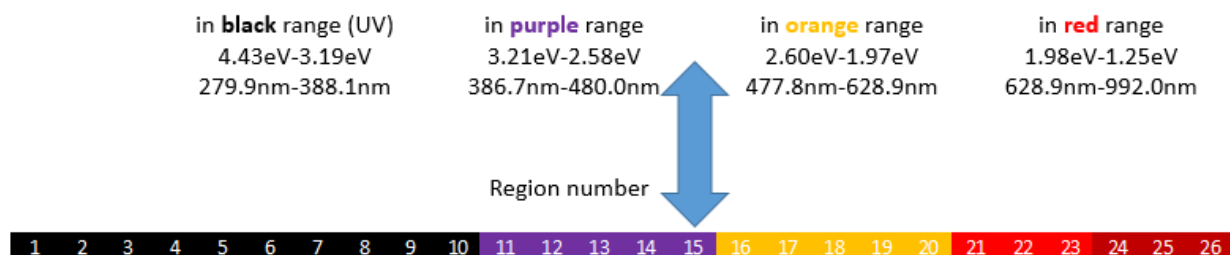
We also mined the data with CRs to select the molecular building blocks that have a good absorption in a specific region. We first divided the 26 spectrum regions into 4 parts, each representing a absorption color range. The scheme is shown in the Figure



**Figure 4.5:** The molecules that have a correlation ratio higher than the threshold (which is 0.7). Ranked with the number of regions where it satisfies the criteria.

4.5. From the definition of spectrum correlation ratio, we know the molecule having a better absorption in a given region should have a higher correlation ratio in that region. Balancing between the variety of the building blocks and the specificity, we set the threshold for the correlation ratio to be 0.7 and the results are shown in Figure 4.5.

In reality, molecules often absorb light in a certain range containing many regions. Using our method, this is a straightforward way to check by comparing the summation of the correlation ratio in different regions, which in turn quantify the performance of the molecule in certain range. In this paper, we presented the results for 4 ranges: Ultraviolet, Visible I, Visible II and Infrared range, which are shown in Figure 4.7, Figure 4.8, Figure 4.9 and Figure 4.10.



**Figure 4.6:** 26 small Divisions then combined to 4 ranges of solar spectrum (mainly IR, red, purple, UV) for later screening to work in.

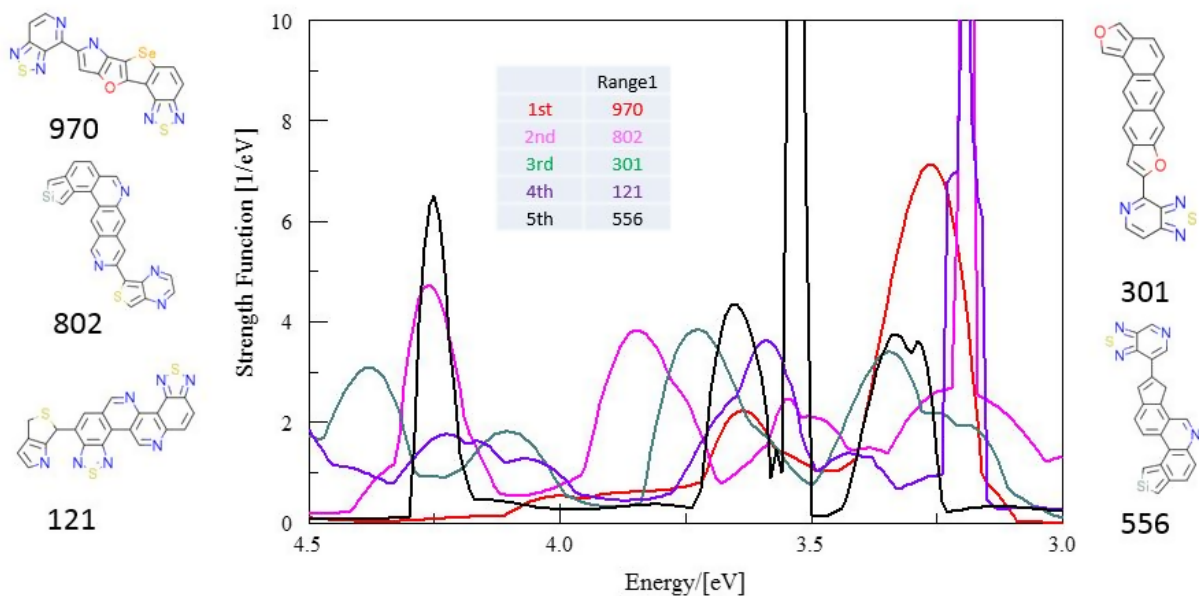


Figure 4.7: Screening result for range I.

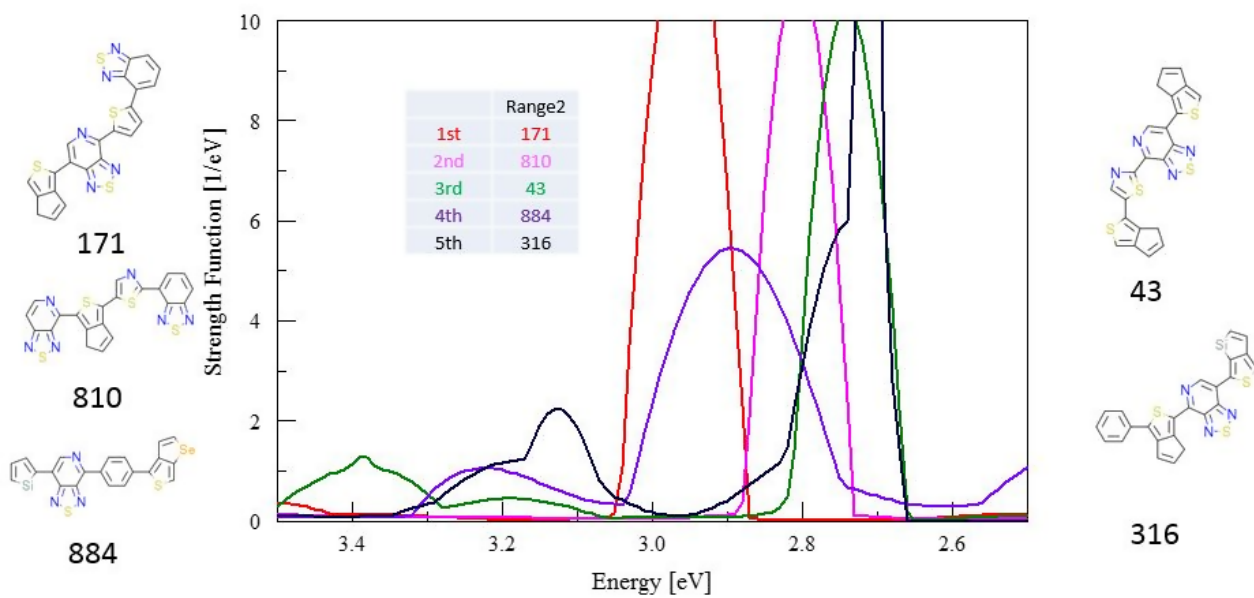


Figure 4.8: Screening result for range II.

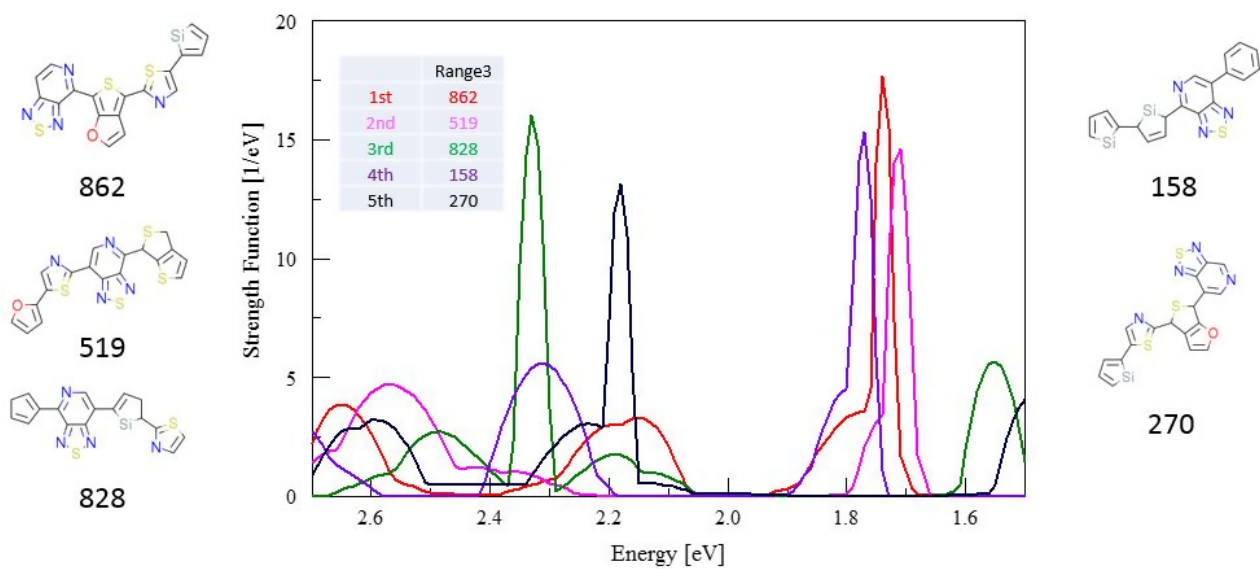


Figure 4.9: Screening result for range III.

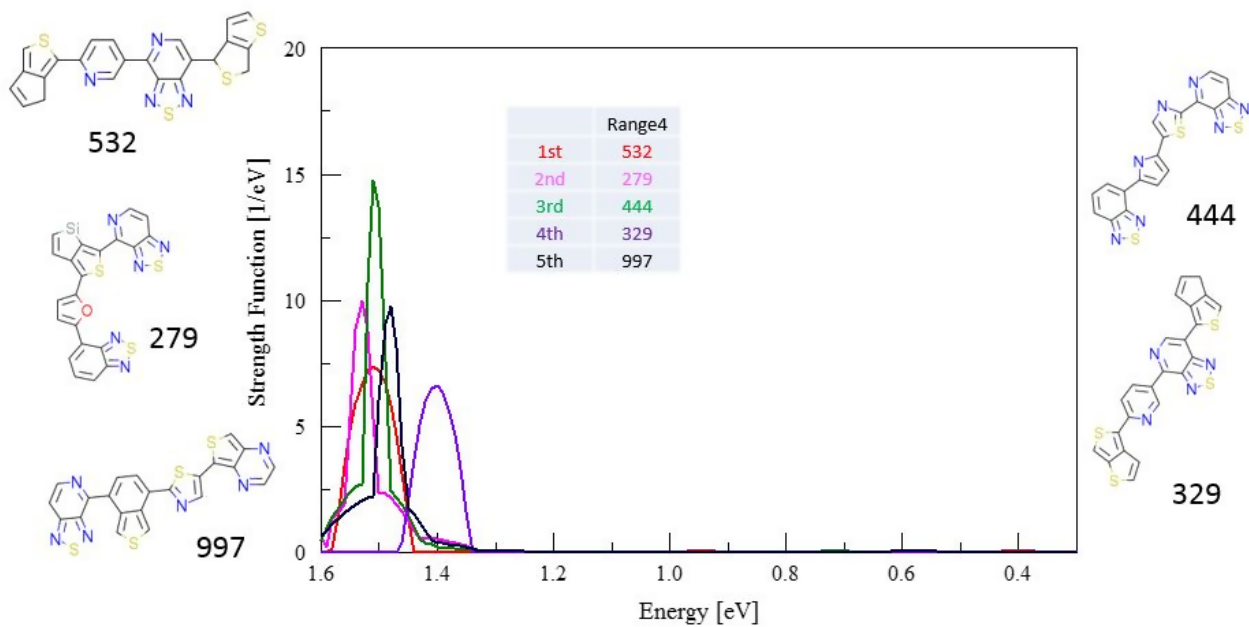
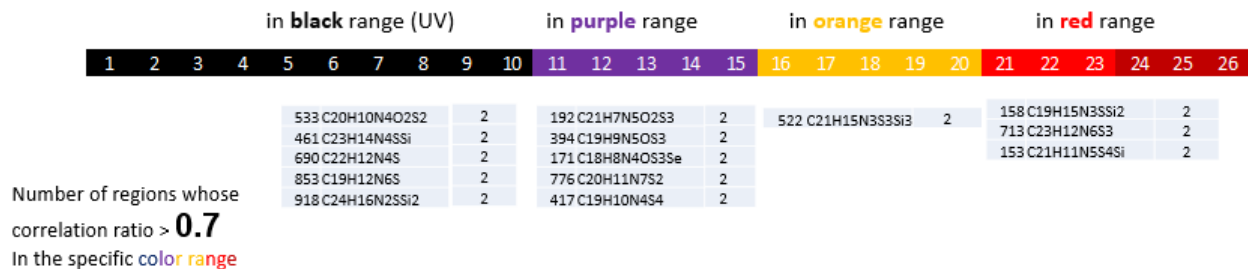


Figure 4.10: Screening result for range IV.





**Figure 4.11: Screening result across regions.**

From the results in Figure 4.11, we could identify the good building block that could help the system to absorb more in the target regions. Also the results show there are more molecules in the UV and purple range that have good performance than that in red and orange regions.

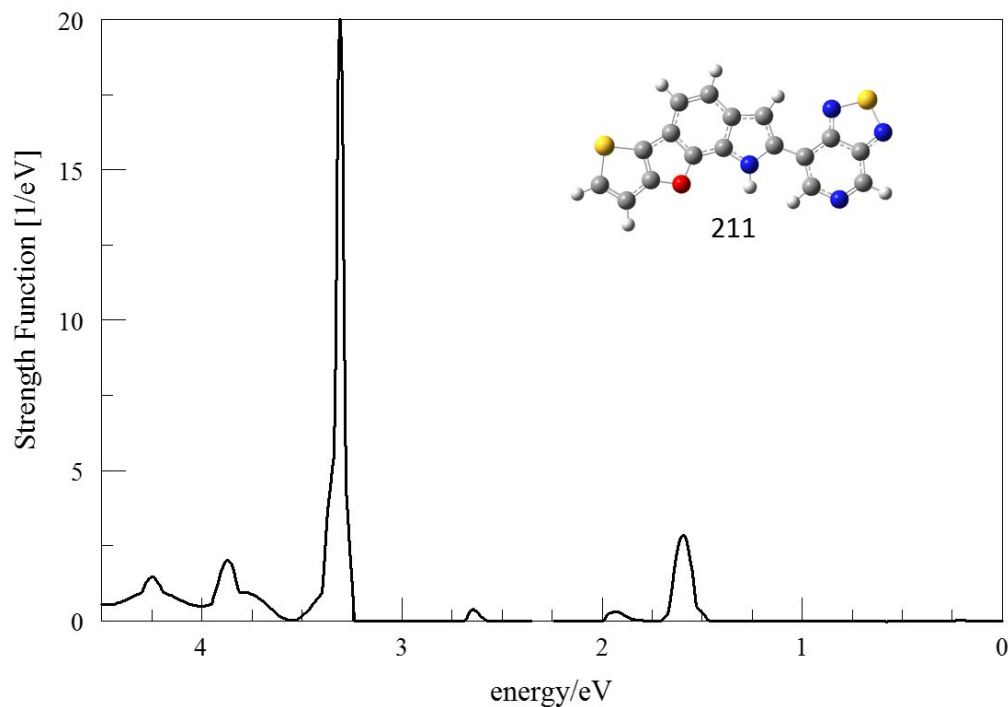
### 4.3.3 Aiming at screening a molecule with special requirement on its spectrum range.

The growing interest in developing power-producing surface (Chen et al. 2012), e.g. special screen cover for mobile devices that could combine energy production and protection together to further slim the device, requires novel materials that are transparent in visible range (Range II and III in our scheme) but absorb photon energy in UV and IR range. This is also useful material for sunglasses, etc. The idea could also applied to a “layered setup” (Betancur et al. 2013).

We can therefore translate the requirement into the screening logic posed on the value of the CR to find out suitable molecules from the spectrum point of view. Here we use a threshold of 0.2 for the correlation ratio of visible range and rank the molecules according to their combined performance in UV and IR range. The selected molecule and its spectrum is shown in Figure 4.12.

## 4.4 Conclusion

In conclusion, we have constructed a descriptor, the spectrum correlation ratio, as a quantitative method to describe the alignment of a molecule’s absorption spectrum with the solar spectrum. By using this quantitative parameter, it is then possible for automated sorting of available material in a large scale to take place. It is designed to handle big data from spectral perspective in available databases, for example, tandem solar cell design. Consequently, we tested the candidates in the CEPDB



**Figure 4.12:** The screening result of the molecule that is transparent in visible range (from 1.65 eV to 3.26 eV) but absorbs energy in UV and IR range.

database and identified the molecules that have a broader spectral coverage or have a better performance in a certain spectral range. The results provided a theoretical screening for the potential OPV materials for tandem solar cells.

In addition, we also gave some examples of translating the demand on the material spectrum property to the screening logic on the spectrum CR. From ranking of the results, we can find out the desired candidate molecule. And a molecule that can be used as a transparent OPV material is screened as an example.

## CHAPTER 5

### Conclusions and Outlook

“...all I know is what the Terminator taught me; never stop fighting. And I never will. The battle has just begun.”

— John Connor, Terminator 3

#### 5.1 Summary of current research

To conquer both the energy problem and pollution problem while still providing enough support to the development of human civilization, solar energy is one of the best candidates to replace conventional energy and therefore more researchers are devoting their efforts to this field (Chu, Cui, and Liu 2017) (Louwen et al. 2016). Bearing the properties of being flexible, lightweight, having low cost and an easy manufacturing process (Dou et al. 2013) (Dennler, Scharber, and Brabec 2009) (Heeger 2014), organic photovoltaic (OPV) is advantageous over conventional inorganic counterparts and is under rapid development in the past few years. Particularly because of the simple and defined product in the reaction they evolved, the advantages of employing small organic molecules has gradually been recognized (Chen, Wan, and Long 2013). The aim of this thesis has been to apply density functional theory (DFT), Time-Dependent DFT (TD-DFT) and High-throughput screening (HTS) method to design new donor material for small molecule organic photovoltaic and molecules for tandem solar cells.

In chapter 2, we first validated the method as doing a functional benchmarking study. We calculated 29 molecules in the organic photovoltaic field using a group of 12 functionals from LDA, GGA, meta-GGA, global hybrid and range separated hybrid functionals. By comparing the predictions from theory and values from experimental measurements, we summarized different linear correction formula for different functionals. And the linear correction model was justified by a k-fold cross validation method. The benchmarking results were used in later chapters.

In chapter 3, we investigated a group of small molecule organic photovoltaic based on a DHICA core with various side groups. Using the functionals summarized from chapter 2, we calculated their HOMO, gap and LUMO. By using the Scharber model, we then predicted their power conversion efficiency (PCE). Since our aim in this chapter was to find the design rule for a high PCE small organic molecule, we further studied their electronic property and suggested new molecules.

In chapter 4, we applied HTS and time dependent density functional theory (TD-DFT) to search for material for tandem solar cells based on a diverse library from the Harvard Clean Energy Project Database (CEPDB). In this study, we designed a new descriptor, correlation ratio (CR), to efficiently screen the molecule based on their spectra property.

## 5.2 Prospective for future research

To continue the study in chapter 2, we would design more molecules based on a melanin core. We would further investigate the electronic structure of melanin and structure control by using side chain groups (end capping groups). Also, the effect of different heteroaromatic rings other than the benzene ring should be explored. Finally, to go to the microscopic scale, we would study the excited state property, e.g. singlet fission, to further boost the quantum efficiency.

Bearing the results from chapter 2 and 3, we would perform an HTS study to discover more useful materials. More design rules are expected to be summarized in the continued chapter 2 study and would be applied to HTS. Also, optical materials not only confined to organic solar cells would be screened out in an HTS study, e.g. material for sunglasses.

## Bibliography

- Adamo, C. and V. Barone (1999). “Toward reliable density functional methods without adjustable parameters: The PBE0 model”. In: *J. Chem. Phys.* 110.13 (cit. on p. 23).
- Adamo, C., G. E. Scuseria, and V. Barone (1999). “Accurate excitation energies from time-dependent density functional theory: Assessing the PBE0 model”. In: *J. Chem. Phys.* 111.7, pp. 2889–2899 (cit. on p. 13).
- Aeschi, Y. et al. (2013). “Directed Metalation Cascade To Access Highly Functionalized Thieno[2,3-f]benzofuran and Exploration as Building Blocks for Organic Electronics”. In: *Org. Lett.* 15.21, pp. 5586–5589 (cit. on pp. 21, 22).
- Aljarilla, A. et al. (2012). “Organic Dyes Incorporating Oligothiophenevinylene for Efficient Dye-Sensitized Solar Cells”. In: *Org. Lett.* 14.22, pp. 5732–5735 (cit. on p. 19).
- Andrade, X., J. N. Sanders, and A. Aspuru-Guzik (2012). “Application of compressed sensing to the simulation of atomic systems”. In: *PNAS* 109.35, pp. 13928–13933 (cit. on p. 62).
- Andrade, X. et al. (2015). “Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems”. In: *Phys. Chem. Chem. Phys.* 17 (47), pp. 31371–31396 (cit. on p. 102).
- Bachman, J. E., L. A. Curtiss, and R. S. Assary (2014). “Investigation of the redox chemistry of anthraquinone derivatives using density functional theory”. In: *J. Phys. Chem. A* 118.38, pp. 8852–8860 (cit. on p. 11).
- Bajorath, J. (2002). “Integration of virtual and high-throughput screening”. In: *Nat. Rev. Drug Discovery* 1.11, p. 882 (cit. on p. 58).
- Baran, D. et al. (2017). “Reducing the efficiency-stability-cost gap of organic photovoltaics with highly efficient and stable small molecule acceptor ternary solar cells”. In: *Nat. Mater.* 16.3, pp. 363–369 (cit. on pp. 45, 57).
- Barone, V. and A. Polimeno (2007). “Integrated computational strategies for UV/vis spectra of large molecules in solution”. In: *Chem. Soc. Rev.* 36.11, pp. 1724–1731 (cit. on p. 102).

- Becke, A. D. (1988). “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Phy. Rev. A* 38.6, pp. 3098–100 (cit. on pp. 23, 24).
- Becke, A. D. (1993). “Density functional thermochemistry. III. The role of exact exchange”. In: *J. Chem. Phys.* 98.7, pp. 5648–5652 (cit. on p. 23).
- Bérubé, N. et al. (2013). “Designing Polymers for Photovoltaic Applications Using ab Initio Calculations”. In: *J. Phys. Chem. C* 117.16, pp. 7964–7972 (cit. on pp. 48, 107).
- Betancur, R. et al. (2013). “Transparent polymer solar cells employing a layered light-trapping architecture”. In: *Nat. Photonics* 7.12, pp. 995–1000 (cit. on p. 69).
- Bickelhaupt, F. M. and E. J. Baerends (2007). “Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry”. In: *Rev. Comput. Chem., Volume 15*, pp. 1–86 (cit. on p. 88).
- Blakesley, J. C. and D. Neher (2011). “Relationship between energetic disorder and open-circuit voltage in bulk heterojunction organic solar cells”. In: *Phys. Rev. B* 84.7, p. 075210 (cit. on p. 7).
- Boyle, N. M. O. et al. (2011). “Open Babel: An open chemical toolbox”. In: *J. Cheminf.* 3.1, p. 33 (cit. on p. 60).
- Brisset, H. et al. (1994). “Novel narrow bandgap polymers from sp<sup>3</sup> carbon-bridged bithienyls: poly (4, 4-ethylenedioxy-4 H-cyclopenta [2, 1-b; 3, 4-b] dithiophene)”. In: *Journal of the Chemical Society, Chemical Communications* 11, pp. 1305–1306 (cit. on p. 6).
- BritishPetrol (2017). “<http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/primary-energy.html>”. In: (cit. on p. 1).
- Bureš, F. (2014). “Fundamental aspects of property tuning in push–pull molecules”. In: *RSC Advances* 4.102, pp. 58826–58851 (cit. on pp. 46, 47).
- Cai, Z.-L., K. Sendt, and J. R. Reimers (2002). “Failure of density-functional theory and time-dependent density-functional theory for large extended  $\pi$  systems”. In: *J. Chem. Phys.* 117.12, pp. 5543–5549 (cit. on p. 12).
- Castro, A. et al. (2006). “octopus: a tool for the application of time-dependent density functional theory”. In: *Phys. Status Solidi B* 243.11, pp. 2465–2488 (cit. on p. 60).
- Cawley, G. C. and N. L. Talbot (2010). “On over-fitting in model selection and subsequent selection bias in performance evaluation”. In: *J. Mach. Learn. Res.* 11.Jul, pp. 2079–2107 (cit. on p. 15).
- Chai, J.-D. and M. Head-Gordon (2008). “Systematic optimization of long-range corrected hybrid density functionals”. In: *J. Chem. Phys.* 128.8 (cit. on p. 23).

- Chen, C.-C. et al. (2012). “Visibly Transparent Polymer Solar Cells Produced by Solution Processing”. In: *ACS Nano* 6.8, pp. 7185–7190 (cit. on p. 69).
- Chen, C.-C. et al. (2014). “An efficient triple-junction polymer solar cell having a power conversion efficiency exceeding 11%”. In: *Adv. Mater.* 26.32, pp. 5670–5677 (cit. on p. 58).
- Chen, Y., X. Wan, and G. Long (2013). “High performance photovoltaic applications using solution-processed small molecules”. In: *Acc. Chem. Res.* 46.11, pp. 2645–2655 (cit. on pp. 57, 71).
- Chu, S., Y. Cui, and N. Liu (2017). “The path towards sustainable energy”. In: *Nat. Mater.* 16.1, pp. 16–22 (cit. on pp. 57, 71).
- Cohen, A. J., P. Mori-Sánchez, and W. Yang (2008). “Insights into current limitations of density functional theory”. In: *Science* 321.5890, pp. 792–794 (cit. on pp. 97, 98).
- (2011). “Challenges for density functional theory”. In: *Chem. Rev.* 112.1, pp. 289–320 (cit. on p. 12).
- Collins, S. D. et al. (2017). “Small is Powerful: Recent Progress in Solution-Processed Small Molecule Solar Cells”. In: *Adv. Energy Mater.* (Cit. on pp. 46, 57).
- Cortizo-Lacalle, D. et al. (2014). “Solution processable diketopyrrolopyrrole (DPP) cored small molecules with BODIPY end groups as novel donors for organic solar cells”. In: *Beilstein J. Org. Chem.* 10, p. 2683 (cit. on p. 46).
- Cui, C. et al. (2015). “High-Performance Organic Solar Cells Based on a Small Molecule with Alkylthio-Thienyl-Conjugated Side Chains without Extra Treatments”. In: *Adv. Mater.* 27.45, pp. 7469–7475 (cit. on p. 46).
- Curtarolo, S. et al. (2012). “AFLOW: an automatic framework for high-throughput materials discovery”. In: *Comput. Mater. Sci* 58, pp. 218–226 (cit. on p. 11).
- Curtarolo, S. et al. (2013a). “The high-throughput highway to computational materials design”. In: *Nat. Mater.* 12.3, pp. 191–201 (cit. on p. 58).
- Curtarolo, S. et al. (2013b). “The high-throughput highway to computational materials design”. In: *Nat Mater* 12.3, pp. 191–201 (cit. on p. 88).
- Davenport, M. et al. (2012). “Introduction to compressed sensing”. In: *Compressed Sensing: Theory and Applications*. Ed. by Y. C. Eldar and G. Kutyniok. Cambridge University Press. Chap. 1, pp. 1–64 (cit. on p. 62).
- Deng, D. et al. (2016). “Fluorination-enabled optimal morphology leads to over 11% efficiency for inverted small-molecule organic solar cells”. In: *Nat. Commun.* 7 (cit. on pp. 45, 46, 57).

- Dennler, G., M. C. Scharber, and C. J. Brabec (2009). “Polymer-fullerene bulk-heterojunction solar cells”. In: *Adv. Mater.* 21.13, pp. 1323–1338 (cit. on pp. 57, 71).
- Dennler, G. et al. (2008). “Design Rules for Donors in Bulk-Heterojunction Tandem Solar Cells Towards 15% Energy-Conversion Efficiency”. In: *Adv. Mater.* 20.3, pp. 579–583 (cit. on p. 7).
- Derouiche, H and V Djara (2007). “Impact of the energy difference in LUMO and HOMO of the bulk heterojunctions components on the efficiency of organic solar cells”. In: *Sol. Energy Mater. Sol. Cells* 91.13, pp. 1163–1167 (cit. on p. 3).
- Dev, P., S. Agrawal, and N. J. English (2012a). “Determining the appropriate exchange-correlation functional for time-dependent density functional theory studies of charge-transfer excitations in organic dyes”. In: *J. Chem. Phys.* 136.22, p. 224301 (cit. on p. 12).
- (2012b). “Determining the appropriate exchange-correlation functional for time-dependent density functional theory studies of charge-transfer excitations in organic dyes”. In: *J. Chem. Phys.* 136.22, p. 224301 (cit. on p. 106).
- Dirac, P. (1930). “On the Stability of Molecules in the Thomas-Fermi Theory”. In: *Proc. Cambridge Philos. Soc.* 26, p. 376 (cit. on p. 93).
- d’Ischia, M., A. Napolitano, and A. Pezzella (2011). “5, 6-Dihydroxyindole Chemistry: Unexplored Opportunities Beyond Eumelanin”. In: *Eur. J. Org. Chem.* 2011.28, pp. 5501–5516 (cit. on p. 47).
- d’Ischia, M. et al. (2009). “Chemical and Structural Diversity in Eumelanins: Unexplored Bio-Optoelectronic Materials”. In: *Angew. Chem. Int. Ed.* 48.22, pp. 3914–3921 (cit. on p. 47).
- Dou, L. et al. (2013). “25th anniversary article: a decade of organic/polymeric photovoltaic research”. In: *Adv. Mater.* 25.46, pp. 6642–6671 (cit. on pp. 3, 57, 71).
- Dreuw, A., J. L. Weisman, and M. Head-Gordon (2003). “Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange”. In: *J. Chem. Phys.* 119.6, pp. 2943–2946 (cit. on pp. 12, 104).
- Duan, C., F. Huang, and Y. Cao (2012). “Recent development of push–pull conjugated polymers for bulk-heterojunction photovoltaics: rational design and fine tailoring of molecular structures”. In: *J. Mater. Chem.* 22.21, pp. 10416–10434 (cit. on pp. 46, 49).



Duan, X.-M. et al. (2004). “Accurate prediction of heat of formation by combining Hartree–Fock/density functional theory calculation with linear regression correction approach”. In: *J. Chem. Phys.* 121.15, pp. 7086–7095 (cit. on p. 15).

Elsherbini, M., W. S. Hamama, and H. H. Zoorob (2017). “Recent advances in the chemistry of selenium-containing heterocycles: Six-membered ring systems”. In: *Coordination Chemistry Reviews* 330, pp. 110–126 (cit. on p. 11).

EnergyIntelligenceAgency (2015). “<https://www.eia.gov/todayinenergy/detail.php?id=26912>”. In: (cit. on p. 1).

Er, S. et al. (2015). “Computational design of molecules for an all-quinone redox flow battery”. In: *Chemical Science* 6.2, pp. 885–893 (cit. on p. 11).

Esteban, S. G. et al. (2011). “Panchromatic PushCPull Chromophores based on T-riphenylamine as Donors for Molecular Solar Cells”. In: *Org. Lett.* 13.19, pp. 5362–5365 (cit. on pp. 19, 49).

Fermi, E. (1927). “Un metodo statistico per la determinazione di alcune priorieta dellatome”. In: *Rend. Accad. Naz. Lincei* 6.602-607, p. 32 (cit. on p. 90).

Frisch, M. J. et al. *Gaussian09 Revision C.01* (cit. on pp. 14, 48, 60, 96).

Gasparini, N. et al. (2017). “High-performance ternary organic solar cells with thick active layer exceeding 11% efficiency”. In: *Energy Environ. Sci.* 10.4, pp. 885–892 (cit. on pp. 45, 57).

Grimme, S. and F. Neese (2007). “Double-hybrid density functional theory for excited electronic states of molecules”. In: *J. Chem. Phys.* 127.15, p. 154116 (cit. on p. 95).

Grimme, S. and M. Parac (2003). “Substantial Errors from Time-Dependent Density Functional Theory for the Calculation of Excited States of Large  $\pi$  Systems”. In: *ChemPhysChem* 4.3, pp. 292–295 (cit. on p. 12).

Guo, X., M. Baumgarten, and K. Müllen (2013). “Designing  $\pi$ -conjugated polymers for organic electronics”. In: *Prog. Polym. Sci.* 38.12, pp. 1832–1908 (cit. on p. 46).

Hachmann, J. et al. (2011). “The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid”. In: *J. Phys. Chem. Letters* 2.17, pp. 2241–2251 (cit. on p. 59).

Hachmann, J. et al. (2014). “Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry - the Harvard Clean Energy Project”. In: *Energy Environ. Sci.* 7 (2), pp. 698–704 (cit. on pp. 11, 59, 106).

Hamasaki, K. and R. R. Rando (1998). “A high-throughput fluorescence screen to monitor the specific binding of antagonists to RNA targets”. In: *Analytical biochemistry* 261.2, pp. 183–190 (cit. on p. 106).

Hansch, C., A. Leo, and R. Taft (1991). "A survey of Hammett substituent constants and resonance and field parameters". In: *Chem. Rev.* 91.2, pp. 165–195 (cit. on p. 50).

Hedley, G. J., A. Ruseckas, and I. D. Samuel (2016). "Light Harvesting for Organic Photovoltaics". In: *Chem. Rev.* (Cit. on p. 3).

Heeger, A. J. (2014). "25th anniversary article: bulk heterojunction solar cells: understanding the mechanism of operation". In: *Adv. Mater.* 26.1, pp. 10–28 (cit. on pp. 57, 71).

Heyd, J., G. E. Scuseria, and M. Ernzerhof (2003a). "Hybrid functionals based on a screened Coulomb potential". In: *J. Chem. Phys.* 118.18 (cit. on p. 23).

Heyd, J., G. E. Scuseria, and M. Ernzerhof (2003b). "Hybrid functionals based on a screened Coulomb potential". In: *J. Chem. Phys.* 118.18, pp. 8207–8215 (cit. on p. 104).

Heyd, J., G. E. Scuseria, and M. Ernzerhof (2006). "Erratum: Hybrid functionals based on a screened Coulomb potential [J. Chem. Phys.118, 8207 (2003)]". In: *J. Chem. Phys.* 124.21 (cit. on p. 23).

Hideaki Ebata, . et al. (2007). "Synthesis, Properties, and Structures of Benzo[1,2-b:4,5-b]bis[b]benzothiophene and Benzo[1,2-b:4,5-b]bis[b]benzoselenophene". In: *Org. Lett.* 9.22, pp. 4499–4502 (cit. on pp. 20, 22).

"[https://en.wikipedia.org/wiki/Time\\_complexity](https://en.wikipedia.org/wiki/Time_complexity)". In: (cit. on p. 90).

Jacquemin, D. et al. (2007). "Assessment of the efficiency of long-range corrected functionals for some properties of large compounds". In: *J. Chem. Phys.* 126.14, p. 144105 (cit. on p. 14).

Jacquemin, D. et al. (2008). "TD-DFT performance for the visible absorption spectra of organic dyes: conventional versus long-range hybrids". In: *J. Chem. Theory Comput.* 4.1, pp. 123–135 (cit. on pp. 14, 97).

Jacquemin, D. et al. (2009). "Extensive TD-DFT benchmark: singlet-excited states of organic molecules". In: *J. Chem. Theory Comput.* 5.9, pp. 2420–2435 (cit. on pp. 102–105).

Jacquemin, D. et al. (2011). "Assessment of the  $\omega$ B97 family for excited-state calculations". In: *Theor. Chem. Acc.* 128.1, pp. 127–136 (cit. on p. 14).

Jain, A., Y. Shin, and K. A. Persson (2016). "Computational predictions of energy materials using density functional theory". In: *Nature Reviews Materials* 1, p. 15004 (cit. on p. 12).

Jemison, R. C. and R. D. McCullough (2014). "Techniques for the Molecular Design of Push-Pull Polymers towards Enhanced Organic Photovoltaic Performance".

In: *Polymer Composites for Energy Harvesting, Conversion, and Storage*. American Chemical Society, pp. 71–109 (cit. on p. 46).

Kallmann, H and M Pope (1959). “Photovoltaic effect in organic crystals”. In: *J. Chem. Phys.* 30.2, pp. 585–586 (cit. on p. 2).

Kan, B. et al. (2014). “Solution-processed organic solar cells based on dialkylthiol-substituted benzodithiophene unit with efficiency near 10%”. In: *Journal of the American Chemical Society* 136.44, pp. 15529–15532 (cit. on p. 46).

Kanal, I. Y. et al. (2013a). “Efficient computational screening of organic polymer photovoltaics”. In: *J. Phys. Chem. letters* 4.10, pp. 1613–1623 (cit. on p. 11).

Kanal, I. Y. et al. (2013b). “Efficient Computational Screening of Organic Polymer Photovoltaics”. In: *J. Phys. Chem. Lett.* 4.10, pp. 1613–1623 (cit. on pp. 59, 106).

Kang, J. K. and C. B. Musgrave (2001a). “Prediction of transition state barriers and enthalpies of reaction by a new hybrid density-functional approximation”. In: *J. Chem. Phys.* 115.24, pp. 11040–11051 (cit. on p. 14).

Kang, J. K. and C. B. Musgrave (2001b). “Prediction of transition state barriers and enthalpies of reaction by a new hybrid density-functional approximation”. In: *J. Chem. Phys.* 115.24, pp. 11040–11051 (cit. on p. 23).

Kim, J.-H. et al. (2016). “Optimization and Analysis of Conjugated Polymer Side Chains for High-Performance Organic Photovoltaic Cells”. In: *Advanced Functional Materials* (cit. on p. 5).

Kohn, W. and L. J. Sham (1965a). “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Phys. Rev.* 140 (4A), A1133–A1138 (cit. on pp. 91–93).

Kohn, W., A. D. Becke, and R. G. Parr (1996). “Density functional theory of electronic structure”. In: *J. Phys. Chem.* 100.31, pp. 12974–12980 (cit. on p. 88).

Kohn, W. and L. J. Sham (1965b). “Self-consistent equations including exchange and correlation effects”. In: *Phy. Rev.* 140.4A, A1133 (cit. on p. 12).

Kollias, N. and A. H. Baqer (1987). “Absorption Mechanisms of Human Melanin in the Visible, 400–720nm”. In: *J. Invest. Dermatol.* 89.4, pp. 384–388 (cit. on p. 47).

Körzdörfer, T. and J.-L. Bredas (2014). “Organic electronic materials: Recent advances in the DFT description of the ground and excited states using tuned range-separated hybrid functionals”. In: *Acc. Chem. Res.* 47.11, pp. 3284–3291 (cit. on pp. 12, 98).

Lange, A. W., M. A. Rohrdanz, and J. M. Herbert (2008). “Charge-transfer excited states in a  $\pi$ -stacked adenine dimer, as predicted using long-range-corrected time-

- dependent density functional theory”. In: *J. Phys. Chem. B* 112.20, pp. 6304–6308 (cit. on p. 14).
- Laurent, A. D. and D. Jacquemin (2013). “TD-DFT benchmarks: A review”. In: *Int. J. Quantum Chem.* 113.17, pp. 2019–2039 (cit. on pp. 102, 104).
- Le Borgne, M. et al. (2016). “Synthesis and properties of a novel narrow band gap oligomeric diketopyrrolopyrrole-based organic semiconductor”. In: *Dyes Pigm.* 131, pp. 160–167 (cit. on p. 46).
- Leang, S. S., F. Zahariev, and M. S. Gordon (2012). “Benchmarking the performance of time-dependent density functional methods”. In: *J. Chem. Phys.* 136.10, p. 104101 (cit. on pp. 13, 102, 103).
- Lee, C., W. Yang, and R. G. Parr (1988). “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Phys. Rev. B* 37 (2), pp. 785–789 (cit. on p. 23).
- Li, H. et al. (2011). “Synthesis of Low Band Gap [1,2,5]-Thiadiazolo[3,4-g]quinoxaline and Pyrazino[2,3-g]quinoxaline Derivatives by Selective Reduction of Benzo[1,2-c;4,5-c]bis[1,2,5]thiadiazole”. In: *Org. Lett.* 13.1, pp. 46–49 (cit. on p. 20).
- Li, M. et al. (2017). “Solution-processed organic tandem solar cells with power conversion efficiencies  $\geq 12\%$ ”. In: *Nature Photonics* 11.2, pp. 85–90 (cit. on p. 11).
- Li, N. et al. (2013). “Towards 15% energy conversion efficiency: a systematic study of the solution-processed organic tandem solar cells based on commercially available materials”. In: *Energy Environ. Sci.* 6.12, pp. 3407–3413 (cit. on pp. 7, 58).
- Lieb, E. H. (2000). *A brief review of Thomas–Fermi theory* (cit. on p. 91).
- Lim, B. et al. (2017). “High Performance Solution Processed Organic Field Effect Transistors with Novel Diketopyrrolopyrrole-Containing Small Molecules”. In: *Sci. Rep.* 7.1, p. 164 (cit. on p. 11).
- Liu, C. et al. (2016). “Low bandgap semiconducting polymers for polymeric photovoltaics”. In: *Chem. Soc. Rev.* 45.17, pp. 4825–4846 (cit. on pp. 11, 46, 50, 51).
- Liu, Y. et al. (2014). “Aggregation and morphology control enables multiple cases of high-efficiency polymer solar cells”. In: *Nat. Commun.* 5 (cit. on p. 6).
- Louwen, A. et al. (2016). “Re-assessment of net energy production and greenhouse gas emissions avoidance after 40 years of photovoltaics development”. In: *Nat. Commun.* 7, p. 13728 (cit. on pp. 57, 71).
- Lu, L. et al. (2015). “Recent advances in bulk heterojunction polymer solar cells”. In: *Chem. Rev.* 115.23, pp. 12666–12731 (cit. on p. 3).

- Luo, C. et al. (2014). “General strategy for self-assembly of highly oriented nanocrystalline semiconducting polymers with high mobility”. In: *Nano letters* 14.5, pp. 2764–2771 (cit. on p. 11).
- Macarron, R. et al. (2011). “Impact of high-throughput screening in biomedical research”. In: *Nat. Rev. Drug Discovery* 10.3, p. 188 (cit. on p. 58).
- Manini, P. et al. (2015). “Melanin-Inspired Organic Electronics: Electroluminescence in Asymmetric Triazatruxenes”. In: *ChemPlusChem* 80.6, pp. 919–927 (cit. on p. 47).
- Mardirossian, N. and M. Head-Gordon (2014). “ $\omega$ B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy”. In: *Phys Chem Chem Phys* 16.21, pp. 9904–9924 (cit. on p. 94).
- Marques, M (2003). “octopus: a first-principles tool for excited electron-ion dynamics”. In: *Comput. Phys. Commun.* 151.1, pp. 60–78 (cit. on p. 60).
- Marques, M. A. and E. K. Gross (2004). “Time-dependent density functional theory”. In: *Annu. Rev. Phys. Chem.* 55, pp. 427–455 (cit. on p. 14).
- Mattsson, A. E. et al. (2005). “Designing meaningful density functional theory calculations in materials science - a primer”. In: *Modell. Simul. Mater. Sci. Eng.* 13.1, R1–R31 (cit. on p. 88).
- McCormick, T. M. et al. (2013). “Conjugated polymers: evaluating DFT methods for more accurate orbital energy modeling”. In: *Macromolecules* 46.10, pp. 3879–3886 (cit. on pp. 13, 24).
- Meng, S. and E. Kaxiras (2008). “Theoretical models of eumelanin protomolecules and their optical properties”. In: *Biophys. J.* 94.6, pp. 2095–2105 (cit. on p. 47).
- Miehlich, B. et al. (1989). “Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr”. In: *Chemical Physics Letters* 157.3, pp. 200–206 (cit. on p. 23).
- Nelson, J. (2003). *The Physics of Solar Cells*. Imperial College Press (cit. on pp. 5, 7, 12, 46–48).
- Neugebauer, J. and T. Hickel (2013). “Density functional theory in materials science”. In: *Wiley Interdiscip. Rev. Comput. Sci. Mol.* 3.5, pp. 438–448 (cit. on p. 88).
- Nguyen, K. A., P. N. Day, and R. Pachter (2011a). “The performance and relationship among range-separated schemes for density functional theory”. In: *J. Chem. Phys.* 135.7, p. 074109 (cit. on p. 14).
- (2011b). “The performance and relationship among range-separated schemes for density functional theory”. In: *J. Chem. Phys.* 135.7, p. 074109 (cit. on p. 104).

- Ørnsø, K. B., J. M. Garcia-Lastra, and K. S. Thygesen (2013). “Computational screening of functionalized zinc porphyrins for dye sensitized solar cells”. In: *Phys. Chem. Chem. Phys.* 15.44, p. 19478 (cit. on p. 59).
- Paier, J. et al. (2006). “Screened hybrid density functionals applied to solids”. In: *J. Chem. Phys.* 124.15, p. 154709 (cit. on p. 104).
- Pan, K.-C. et al. (2016). “Efficient and Tunable Thermally Activated Delayed Fluorescence Emitters Having Orientation-Adjustable CN-Substituted Pyridine and Pyrimidine Acceptor Units”. In: *Adv. Funct. Mater.* 26.42, pp. 7560–7571 (cit. on p. 11).
- Parr, R. and W Yang (1989). “Density Functional Theory of Atoms and Molecules Oxford Univ”. In: *Press, New York* (cit. on p. 88).
- Parr, R. G. (1982). “Density Functional Theory”. In: *Electron distributions and the chemical bond*. Ed. by P. Coppens and M. B. Hall. Springer Science & Business Media. Chap. 3, pp. 95–100 (cit. on p. 12).
- Peach, M. J. et al. (2008). “Excitation energies in density functional theory: an evaluation and a diagnostic test”. In: *J. Chem. Phys.* 128.4, p. 044118 (cit. on pp. 13, 103).
- Perdew, J. P., K. Burke, and M. Ernzerhof (1996). “Generalized Gradient Approximation Made Simple”. In: *Phys. Rev. Lett.* 77 (18), pp. 3865–3868 (cit. on p. 24).
- Perdew, J. P., K. Burke, and Y. Wang (1996). “Generalized gradient approximation for the exchange-correlation hole of a many-electron system”. In: *Phys. Rev. B* 54.23, p. 16533 (cit. on p. 23).
- Perdew, J. P. and Y. Wang (1992). “Accurate and simple analytic representation of the electron-gas correlation energy”. In: *Phys. Rev. B* 45.23, p. 13244 (cit. on p. 23).
- Perdew, J. P. et al. (1982). “Density-functional theory for fractional particle number: derivative discontinuities of the energy”. In: *Phys. Rev. Lett.* 49.23, p. 1691 (cit. on p. 12).
- Perdew, J. P. et al. (1992). “Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation”. In: *Phys. Rev. B* 46.11, p. 6671 (cit. on p. 23).
- Peverati, R. and D. G. Truhlar (2014). “Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics”. In: *Philos. Trans. R. Soc. London, Ser. A* 372.2011, p. 20120476 (cit. on p. 88).
- Phillips, H. et al. (2014). “Orbital gap predictions for rational design of organic photovoltaic materials”. In: *Organic Electronics* 15.7, pp. 1509–1520 (cit. on p. 11).

- Plumley, J. A. and J. Dannenberg (2011). “A comparison of the behavior of functional/basis set combinations for hydrogen-bonding in the water dimer with emphasis on basis set superposition error”. In: *J. Comput. Chem.* 32.8, pp. 1519–1527 (cit. on p. 97).
- Pytela, O. (1996). “Chemometric Analysis of Substituent Effects. IX. Alternative Interpretation of Substituent Effects (AISE)-Orthogonal Model”. In: *Collection of Czechoslovak Chem. Commun.* 61.5, pp. 704–712 (cit. on p. 50).
- Qu, X. et al. (2015). “The Electrolyte Genome project: A big data approach in battery materials discovery”. In: *Comput. Mater. Sci* 103, pp. 56–67 (cit. on p. 11).
- Reymond, J.-L. (2015). “The chemical space project”. In: *Acc. Chem. Res.* 48.3, pp. 722–730 (cit. on pp. 11, 45).
- Rohrdanz, M. A. and J. M. Herbert (2008). “Simultaneous benchmarking of ground- and excited-state properties with long-range-corrected density functional theory”. In: *J. Chem. Phys.* 129.3, p. 034107 (cit. on p. 101).
- Roncali, J., P. Leriche, and P. Blanchard (2014). “Molecular materials for organic photovoltaics: small is beautiful”. In: *Adv. Mater.* 26.23, pp. 3821–3838 (cit. on pp. 46, 57).
- Rumer, J. W. and I. McCulloch (2015). “Organic photovoltaics: Crosslinking for optimal morphology and stability”. In: *Materials Today* 18.8, pp. 425–435 (cit. on p. 5).
- Runge, E. and E. K. Gross (1984a). “Density-functional theory for time-dependent systems”. In: *Phys. Rev. Lett.* 52.12, p. 997 (cit. on p. 14).
- (1984b). “Density-functional theory for time-dependent systems”. In: *Phys. Rev. Lett.* 52.12, p. 997 (cit. on pp. 88, 102).
- Salzner, U et al. (1998). “Accurate Method for Obtaining Band Gaps in Conducting Polymers Using a DFT/Hybrid Approach”. In: *J Phys Chem* 102.15, pp. 2572–2578 (cit. on p. 88).
- Scharber, M. C. (2016). “On the Efficiency Limit of Conjugated Polymer: Fullerene-Based Bulk Heterojunction Solar Cells”. In: *Adv. Mater.* (Cit. on p. 58).
- Scharber, M. et al. (2006). “Design Rules for Donors in Bulk-Heterojunction Solar Cells Towards 10% Energy-Conversion Efficiency”. In: *Adv. Mater.* 18.6, pp. 789–794 (cit. on pp. 12, 48, 58, 106, 107).
- Scholes, G. D. and G. Rumbles (2006). “Excitons in nanoscale systems”. In: *Nature materials* 5.9, pp. 683–696 (cit. on p. 3).

- Selvaraju, S. et al. (2016a). “Effects of structural variations on the optical and electronic properties of eumelanin-inspired small molecules”. In: *Journal of Materials Chemistry C* 4.18, pp. 3995–3999 (cit. on p. 6).
- Selvaraju, S. et al. (2016b). “Effects of structural variations on the optical and electronic properties of eumelanin-inspired small molecules”. In: *J. Mater. Chem. C* 4 (18), pp. 3995–3999 (cit. on pp. 20, 22).
- Sharma, R., J. Zhang, and C. A. Ohlin (2016). “Predicting  $^{17}\text{O}$  NMR chemical shifts of polyoxometalates using density functional theory”. In: *Phys. Chem. Chem. Phys.* 18.11, pp. 8235–8241 (cit. on p. 15).
- Shockley, W. and H. J. Queisser (1961a). “Detailed balance limit of efficiency of p-n junction solar cells”. In: *J. Appl. Phys.* 32.3, pp. 510–519 (cit. on p. 6).
- (1961b). “Detailed balance limit of efficiency of p-n junction solar cells”. In: *J. Appl. Phys.* 32.3, pp. 510–519 (cit. on p. 58).
- Slater, J. (1963). *Quantum Theory of Molecules and Solids: The self-consistent field for molecules and solids*. McGraw-Hill (cit. on p. 23).
- SolarEnergyIndustriesAssociation (2016). “<http://www.seia.org/news/us-solar-market-grows-95-2016-smashes-records>”. In: (cit. on p. 1).
- Sousa, S. F., P. A. Fernandes, and M. J. Ramos (2007). “General performance of density functionals”. In: *J. Phys. Chem. A* 111.42, pp. 10439–10452 (cit. on p. 97).
- Spanggaard, H. and F. C. Krebs (2004). “A brief history of the development of organic and polymeric photovoltaics”. In: *Sol. Energy Mater. Sol. Cells* 83.2, pp. 125–146 (cit. on p. 2).
- Steinberger, S. et al. (2011). “A-D-A-D-A-Type Oligothiophenes for Vacuum-Deposited Organic Solar Cells”. In: *Org. Lett.* 13.1, pp. 90–93 (cit. on p. 20).
- Stephens, P. J. et al. (1994). “Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”. In: *J. Phys. Chem.* 98.45, pp. 11623–11627 (cit. on pp. 23, 95).
- Stowasser, R. and R. Hoffmann (1999). “What Do the Kohn Sham Orbitals and Eigenvalues Mean”. In: *J Am Chem Soc* 121.14, pp. 3414–3420 (cit. on p. 12).
- Szczepanik, D. W. et al. (2017). “The role of the long-range exchange corrections in the description of electron delocalization in aromatic species”. In: *J. Comput. Chem.* 38.18, pp. 1640–1654 (cit. on pp. 14, 24).
- Tang, C. W. (1979). *Multilayer organic photovoltaic elements* (cit. on p. 2).
- (1986). “Two-layer organic photovoltaic cell”. In: *Appl. Phys. Lett.* 48.2, pp. 183–185 (cit. on p. 2).



- Tao, J. et al. (2003). “Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids”. In: *Phys. Rev. Lett.* 91.14, p. 146401 (cit. on p. 23).
- Teller, E. (1962). “On the Stability of Molecules in the Thomas-Fermi Theory”. In: *Rev. Mod. Phys.* 34 (4), pp. 627–631 (cit. on p. 91).
- Thomas, L. H. (1927). “The calculation of atomic fields”. In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 23. 05. Cambridge Univ Press, pp. 542–548 (cit. on p. 90).
- Tingare, Y. S. et al. (2013). “Novel Oxindole Based Sensitizers: Synthesis and Application in Dye-Sensitized Solar Cells”. In: *Org. Lett.* 15.17, pp. 4292–4295 (cit. on p. 19).
- Tsai, W.-L. et al. (2015). “A versatile thermally activated delayed fluorescence emitter for both highly efficient doped and non-doped organic light emitting devices”. In: *Chem. Commun.* 51.71, pp. 13662–13665 (cit. on p. 11).
- Velusamy, M. et al. (2005). “Organic Dyes Incorporating Low-Band-Gap Chromophores for Dye-Sensitized Solar Cells”. In: *Org. Lett.* 7.10, pp. 1899–1902 (cit. on p. 18).
- Vosko, S. H., L. Wilk, and M Nusair (1980). “Accurate spin-dependent electron liquid correlation energies for local spin density calculations a critical analysis”. In: *Can. J. Phys.* 58.8, pp. 1200–1211 (cit. on p. 23).
- Vydrov, O. A. and G. E. Scuseria (2006). “Assessment of a long-range corrected hybrid functional”. In: *J. Chem. Phys.* 125.23, 234109 (cit. on p. 23).
- Vydrov, O. A., G. E. Scuseria, and J. P. Perdew (2007). “Tests of functionals for systems with fractional electron number”. In: *J. Chem. Phys.* 126.15 (cit. on p. 23).
- Vydrov, O. A. et al. (2006). “Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals”. In: *J. Chem. Phys.* 125.7 (cit. on p. 23).
- Wünsche, J. et al. (2013). “Eumelanin thin films: solution-processing, growth, and charge transport properties”. In: *J. Mater. Chem. B* 1.31, pp. 3836–3842 (cit. on p. 47).
- Xiang, L. et al. (2016). “High mobility n-channel organic field-effect transistor based a tetratetracontane interfacial layer on gate dielectrics”. In: *IEEE Electron Device Lett.* 37.12, pp. 1632–1635 (cit. on p. 11).
- Yanai, T., D. P. Tew, and N. C. Handy (2004). “A new hybrid exchange correlation functional using the Coulomb-attenuating method (CAM-B3LYP)”. In: *Chem. Phys. Lett.* 393.1C3, pp. 51–57 (cit. on p. 23).

- Yao, H. et al. (2016). “Molecular design of benzodithiophene-based organic photovoltaic materials”. In: *Chem. Rev* 116.12, pp. 7397–7457 (cit. on p. 46).
- You, J. et al. (2013). “A polymer tandem solar cell with 10.6% power conversion efficiency”. In: *Nat. Commun.* 4, p. 1446 (cit. on p. 6).
- Young, K. et al. (1998). “Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen”. In: *Nature biotechnology* 16.10, pp. 946–950 (cit. on p. 106).
- Yu, G. et al. (1995). “Polymer photovoltaic cells: Enhanced efficiencies via a network of internal donor-acceptor heterojunctions”. In: *Science* 270.5243, p. 1789 (cit. on p. 4).
- Yuan, Y. et al. (2014). “Ultra-high mobility transparent organic thin film transistors grown by an off-centre spin-coating method”. In: *Nat. Commun.* 5, p. 3005 (cit. on p. 11).
- Zerza, G et al. (2001). “Ultrafast charge transfer in conjugated polymer-fullerene composites”. In: *Synthetic metals* 119.1-3, pp. 637–638 (cit. on p. 5).
- Zhan, C.-G., J. A. Nichols, and D. A. Dixon (2003). “Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: molecular properties from density functional theory orbital energies”. In: *J. Phys. Chem. A* 107.20, pp. 4184–4195 (cit. on p. 100).
- Zhang, G. and C. B. Musgrave (2007). “Comparison of DFT methods for molecular orbital eigenvalue calculations”. In: *J. Phys. Chem. A* 111.8, pp. 1554–1561 (cit. on pp. 12, 13, 15, 24, 26, 98, 99).
- Zhang, G. et al. (2017). “High-performance ternary organic solar cell enabled by a thick active layer containing a liquid crystalline small molecule donor”. In: *Journal of the American Chemical Society* 139.6, pp. 2387–2395 (cit. on p. 11).
- Zhang, K. et al. (2016). “High-Performance Polymer Tandem Solar Cells Employing a New n-Type Conjugated Polymer as an Interconnecting Layer”. In: *Adv. Mater.* 28.24, pp. 4817–4823 (cit. on p. 58).
- Zhao, J. et al. (2016a). “Efficient organic solar cells processed from hydrocarbon solvents”. In: *Nature Energy* 1, p. 15027 (cit. on pp. 45, 46, 57).
- Zhao, J. et al. (2017a). “Highly efficient green and red OLEDs based on a new exciplex system with simple structures”. In: *Organic Electronics* 43, pp. 136–141 (cit. on p. 11).
- Zhao, W. et al. (2016b). “Fullerene-Free Polymer Solar Cells with over 11% Efficiency and Excellent Thermal Stability”. In: *Adv. Mater.* 28.23, pp. 4734–4739 (cit. on p. 46).

- Zhao, W. et al. (2017b). “Molecular Optimization Enables over 13% Efficiency in Organic Solar Cells”. In: *Journal of the American Chemical Society* 139.21, pp. 7148–7151 (cit. on p. 11).
- Zhao, Y. and D. G. Truhlar (2008a). “Density functionals with broad applicability in chemistry”. In: *Acc. Chem. Res.* 41.2, pp. 157–167 (cit. on p. 88).
- Zhao, Y. and D. G. Truhlar (2008b). “The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals”. In: *Theor. Chem. Acc.* 120.1, p-p. 215–241 (cit. on p. 23).
- Zheng, Z. et al. (2016). “Over 11% Efficiency in Tandem Polymer Solar Cells Featured by a Low-Band-Gap Polymer with Fine-Tuned Properties”. In: *Adv. Mater.* 28.25, pp. 5133–5138 (cit. on p. 58).
- Zhou, H. et al. (2015). “Polymer Homo-Tandem Solar Cells with Best Efficiency of 11.3%”. In: *Adv. Mater.* 27.10, pp. 1767–1773 (cit. on p. 58).

## APPENDIX A

### A Short Introduction to DFT/TD-DFT and Material Informatics

#### A.0.1 Molecular modeling using DFT

The research method we applied combines computational modeling methods of density functional theory (DFT) and time dependent DFT (TD-DFT) together with a large scale data analysis method (virtual design and screening) (Curtarolo et al. 2013b). To model phenomena at the molecular level, quantum mechanics should be applied. We therefore chose DFT, which is usually employed to solve for ground state electronic structures and properties since it would have the exact solution for the ground state. Due to its good balance between accuracy and efficiency, there is a growing popularity in applying DFT in scientific research which involves studies in electronic structures and properties, including physics (Kohn, Becke, and Parr 1996) (Peverati and Truhlar 2014), chemistry (Bickelhaupt and Baerends 2007) (Zhao and Truhlar 2008a), material science (Neugebauer and Hickel 2013) (Salzner et al. 1998) (Mattsson et al. 2005) etc.

The Time-dependent DFT (Runge and Gross 1984b) (Parr and Yang 1989) is mostly used to study the excited state properties of molecules. In this study, it is used to calculate the vertical singlet/triplet excitation from the ground state geometry acquired from DFT. The first excitation energy should correspond to the energy bandgap in molecules. It is more accurate than DFT when dealing with excited state phenomena.

#### Molecular modeling using DFT

From the time-independent Schrödinger equation (Equation (A.1)) we can safely describe the states of the system using the wave function  $\Psi$ . This is accurate enough for hydrogen and light element atoms (who lay before the transition metal elements in periodic table). Although the physical meaning of  $\Psi$  is not known at first and still many people provide different opinions on it now, the most accepted explanation is from M. Born, which is called the statistical explanation. He related the square of

the wave function to the probability of finding a particle in certain region. And this equation reads like:

$$\left(\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\Psi = -i\hbar\frac{\partial\Psi}{\partial t} \quad (\text{A.1})$$

One particle Schrödinger equation can be solved exactly. However, when we try to apply the same equation to solve the atomic or molecular system, which has more than one electron and nuclues, we met difficulties. The manybody Schrödinger is not solvable, because it would require too many degrees of freedom to describe the system, which is numerically impossible. In general, if a system contains M nuclei (of the same kind) and N electrons, then it has the Hamiltonian:

$$H = H_Z + H_e + H_{Z-e} \quad (\text{A.2})$$

$$H = -\frac{\hbar}{2m_Z} \sum_{i=1}^M \nabla_i^2 + \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^M \frac{ZZ}{|\mathbf{R}_i - \mathbf{R}_j|} \quad (\text{A.3})$$

$$-\frac{\hbar}{2m_e} \sum_{i=1}^N \nabla_i^2 + \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{e}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (\text{A.4})$$

$$-\sum_{i=1}^M \sum_{j=1}^N \frac{Ze}{|\mathbf{R}_i - \mathbf{r}_j|} \quad (\text{A.5})$$

where:

$\mathbf{R}_i$  is the position vector of the  $i$  th nucleus;

$\mathbf{r}_j$  is the position vector of the  $j$  th electron;

$e$  is the fundamental charge for electron;

$Z$  is the atomic number of the nuclei;

$m_Z$  is the mass of a nucleus;

$m_e$  is the mass of an electron.

To solve this equation, we need to make approximations. The first one is called Born-Oppenheimer approximation, which separates the nuclei and electrons into two different parts because they response rates to the disturbance are much different. In this approximation, we could consider the kinetic energy of nuclei to be more or less zero and the potential energy of them to be a constant, so we would ignore them at this moment. After making this approximation, the problem now is how to solve

a system consists of many interacting electrons moving in a background of nuclei potential. So the Hamiltonian now becomes:

$$H = H_e + H_{Z-e} \quad (\text{A.6})$$

$$H = -\frac{\hbar}{2m_e} \sum_{i=1}^N \nabla_i^2 + \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{(Ze)^2}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (\text{A.7})$$

$$- \sum_{i=1}^M \sum_{j=1}^N \frac{Ze^2}{|\mathbf{R}_i - \mathbf{r}_j|} \quad (\text{A.8})$$

### The theory side of DFT

After making Born-Oppenheimer approximation, we have a manybody Schrödinger equation left to solve. People developed some ways to directly attack this equation, *e.g.* diagrammatic perturbation theory based on green's function (from physics community) and configuration interaction (CI) method based on systematic expansion in Slater determinate (from chemistry community). However, there is another formulism that is equivalent to many body Schrodinger equation but is significantly less demanding in calculation time. That is DFT theory. Table A.1 compare the time complexity of calculations from those theories. The time cost is usually measured using the concept of time complexity in computer science field ([https://en.wikipedia.org/wiki/Time\\_complexity](https://en.wikipedia.org/wiki/Time_complexity)). However, it is not a direct link between the size of the current system and its computation cost but rather it is how time cost would increase if the system changed to n times as big as before. From the Table A.1, we choose the theory of DFT since it scales better in time cost. In the following equations, we take  $e = m_e = \hbar = 1$ .

#### *Thomas-Fermi Model and the local exchange*

The first DFT theory is started by E.Fermi who together with L.H.Thomas constructed the Thomas-Fermi model in 1927 (Fermi 1927) (Thomas 1927). The theory gives the energy functional of electron density as in

$$\begin{aligned}
E_{TF}[n(\mathbf{r})] = & C_F \int n^{\frac{5}{3}} d^3\mathbf{r} - Z \int \frac{n(\mathbf{r})}{\mathbf{r}} d^3\mathbf{r} \\
& + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\
& + C_{LE} \int n(\mathbf{r})^{\frac{4}{3}} d^3\mathbf{r}
\end{aligned}$$

Where the first part is for kinetic energy, the second part is for electron-nucleus interaction and the last part is electron-electron interaction. The coefficient  $C_F = \frac{3}{10}(3\Pi^2)^{\frac{2}{3}}$  is a constant and  $C_{LE} = -\frac{3}{4}(\frac{3}{\Pi})^{\frac{1}{3}}$  is a constant for the local exchange introduced later by P.A.M. Dirac (1930). This equation can be solved using lagrange multiplier method (the physical meaning of the parameter is the chemical potential) under the constrain:

$$N = \int n(\mathbf{r})d^3\mathbf{r}$$

The Thomas-Fermi model didn't include the exchange energy, which is later introduced by P.A.M. Dirac and the model becomes Thomas-Fermi-Dirac model. However, because it neglects the electron correlation energy, there is no chemical bonding predicted in this model (Teller 1962) (Lieb 2000). But it initiated the idea to use the density  $n$  as the fundamental variable instead of the wave function and introduced variational method. In modern DFT (e.g. Kohn-Sham DFT or KS-DFT), where we could only solve the equation with certain assumption, we still take the local exchange term in Thomas-Fermi model as the expression for exchange-correlation term and it is therefore called local density approximation (LDA) (Kohn and Sham 1965a). Also, we would still solve the equation through variational method.

**Table A.1: The time complexity of general DFT, MP2 and CCSD codes.**

Theory	Time complexity
DFT	$O(n^3)$
MP2	$O(n^5)$
CCSD	$O(n^6)$

**KS-DFT** The theory of KS-DFT is built on Hohenber-Kohn (HK) theorem and Kohn-Sham equation. The first HK theorem assures us that we could use electron density as the variable because there is a unique correspondent between electron density and external potential. And the second HK theorem asserts the existent

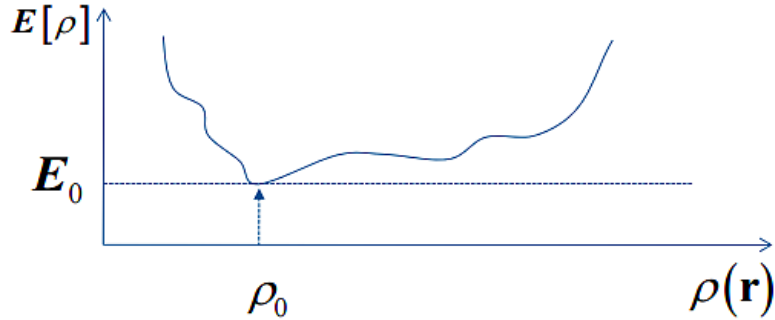
of a functional of density which would lead the energy of the system to minimum. Until this step, all treatments done are accurate. However, in order to solve for the solution, we would use the Kohn-Sham equation (Kohn and Sham 1965a) and SCF (self-consistent field) procedure to find out the approximate solution.

From the 1st HK theorem:

$$E[n] = T[n] + V_{interaction}[n] + \int V_{ext}(\mathbf{r})n(\mathbf{r})d^3r \quad (\text{A.9})$$

where  $T_e$  is the kinetic energy of interacting electrons,  $V_{interaction}$  is the interaction energy of electrons,  $V_{ext}$  is the potential energy of the electrons due to nuclei background. And from the 2nd HK theorem, ground state electron density  $n_0$  is when  $E[n]$  takes the global minimum (As shown in Figure A.1 ). So

$$E_0 = \min_{\forall n \in \{n\}} E[n] \quad (\text{A.10})$$



**Figure A.1:** Illustration of global minimum of energy functional  $E[n]$ .

Although HK theorem proves that there exists a solution satisfying the problem, it provides no clue to how to proceed. To actually find out the solution, Kohn & Sham made a progress (Kohn and Sham 1965a) by constructing a reference system (KS system) of non-interacting electrons with the same density  $n$  as the original system. From the physics point of view, compared to the correlated electrons (original system), the KS system has single particle kinetic energy, coulomb repulsion and potential energy from external nuclei, so we would expect an energy difference which arise from quantum effects of exchange and correlation and denote this part as  $E_{XC}$ . Then under the assumption that the energy could be written separately into different parts, the two systems should have the same energy by adding  $E_{XC}$  to the KS system as in Equation (A.0.1). In the following equations, we use indices "s" to represent



the quantities in single particle system (independent particle system). And  $V_{sext}$  is for single particle system which is different from  $V_{ext}$  for the original system.

$$\begin{aligned} E_{CorrelatedElectrons}[n] &= E_{KS} \\ &= T_s + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \int V_{sext}(\mathbf{r})n(\mathbf{r})d^3r + E_{XC}[n] \end{aligned}$$

The KS system is easier to solve following the 2nd HK theorem in which we set  $\frac{\delta E_{KS}}{\delta n} = 0$  to find out  $n$  that gives minimum  $E$ . And if we use  $\mu$  to denote the lagrange multiplier and the constraint is  $\int n d^3r = 0$ , then we have:

$$\delta[T_s(n) + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \int V_{sext}(\mathbf{r})n(\mathbf{r})d^3r + E_{XC} - \mu(\int n(\mathbf{r})d\mathbf{r} - N)] = 0$$

However, this equation will lead to a schrödinger-like equation, assuming we knew the analytical form of  $E_{XC}$ . To see this, we define  $V_{KS}(\mathbf{r})$  as:

$$V_{KS}(\mathbf{r}) = \delta\left(\frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \int V_{sext}(\mathbf{r})n(\mathbf{r})d^3r + E_{XC}\right)/\delta n(\mathbf{r})$$

And notice  $T_s(n)$  is ( $\Psi_i$  is the single particle wavefunction :  $n(\mathbf{r}) = \sum_{i=1}^N \Psi_i^*(\mathbf{r})\Psi_i(\mathbf{r})$ ):

$$T_s(n) = -\frac{1}{2} \sum_{i=1}^N \int \Psi_i^*(\mathbf{r})\nabla^2\Psi_i(\mathbf{r})d\mathbf{r}$$

Then we would arrive at:

$$\left(-\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})\right)\Psi_i(\mathbf{r}) = \varepsilon_i\Psi_i(\mathbf{r})$$

with  $\varepsilon_i$  being the eigenvalue of the corresponding wave function.

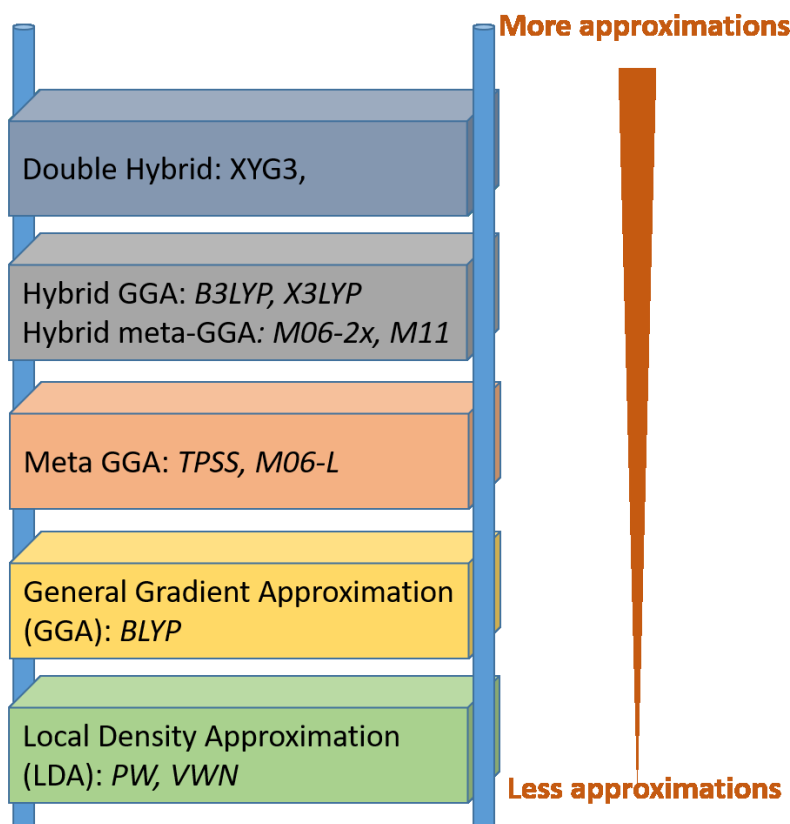
We could then take advantage of the so-called Self Consistent Field (SCF) procedure to solve this single particle schrödinger equation numerically, assuming we know the  $E_{XC}$ . So the  $E_{XC}$  or the exchange-correlation functional is all we need. However, the exact form of the exchange-correlation functional is still unknown so far. In Kohn & Sham 's paper (Kohn and Sham 1965a), they used LDA approximation coming from P.A.M. Dirac Dirac 1930. But we could also make other approximations to make the problem solvable. Other commonly used approximations include: general gradient approximation (GGA), meta-GGA. Here we list the approximations we made in order to solve the KS-DFT equations:

1. Born-Oppenheimer approximation ;
2. Energy functional can be separated into parts ( $E_{XC}$  can be explicitly separated);
3. Exchange-correlation functional takes LDA, GGA, meta-GGA or hybrid approximation

## The calculation side of DFT

### Functionals

The core parts of DFT calculation lies on the exchange-correlation functionals. However, since we haven't found out the exact form of exchange-correlation functional, we have to make approximations if we used KS-DFT. There are different types of approximate exchange-correlation functionals with growing accuracy, which are shown in Figure A.2 (Mardirossian and Head-Gordon 2014)



**Figure A.2:** Levels of accuracy of different DFT functionals (From reference (Mardirossian and Head-Gordon 2014)).

LDA approximation assumes the exchange-correlation is a functional of only local density, which is inherited from uniform gas model. The exchange part of LDA is

shown in Equation (A.11). The correlation part could be acquired through quantum monte carlo simulation.

$$E_{XC}^{LDA}[n] \propto \int n^{\frac{3}{4}}(\mathbf{r}) d\mathbf{r} \quad (\text{A.11})$$

GGA approximation takes the non-uniform density into account so the GGA exchange-correlation is a function of local density and density gradient. In practical, one could use either fitting or physics model to find out the parameters. Examples include PBE, BLYP etc.

The third level, meta-GGA would include the Laplacian and kinetic of the electron density. Example includes TPSS.

On the fourth level lays the hybrid functionals. The Hatree-Fock theory contains the exchange part which arise from pure local exchange (we name it exact exchange), so we would expect the accuracy to improve if we put in the exact exchange accounting for local interaction and leave the difference with real system to be decided by parameters. This idea could be shown in Equation (A.12). In the B3LYP (Stephens et al. 1994) functional, we take the parameters as (A.13) shows.

$$E_{XC}^{Hybrid}[n] = aE_X^{HF} + (1 - a)E_{XC}^{GGA} \quad (\text{A.12})$$

$$E_{XC}^{B3LYP}[n] = E_X^{LDA} + n_0(E_X^{HF} - E_X^{LDA}) + n_x(E_X^{B88}) + E_C^{VWN3} + n_c(E_C^{LYP} - E_C^{VWN3})$$

$$n_0 = 0.2; n_x = 0.72; n_c = 0.81 \quad (\text{A.13})$$

If we consider the coulomb screening effect in the condensed matter system, we would separate the interaction into a short range part and a long range part, which would have different forms (range separation). The length that separate the short and the long is tuned by a parameter ( $\omega$ ). Functionals reconstructed in the way have a  $\omega$  symbol in their names, e.g.  $\omega$ PBE,  $\omega$ B97,  $\omega$ B97XD etc.

The last level is for double hybrid functionals, which have the form as shown in Equation (A.14) (Grimme and Neese 2007)

$$E_{XC}^{DH}[n] = E_X^{LDA} + a(E_X^{HF} - E_X^{LDA}) + bE_X^{GGA} + c(E_C^{PT2} - E_C^{LDA}) + dE_C^{GGA} \quad (\text{A.14})$$

In practice, the calculation time of double hybrid functionals can be very long, sometimes even longer than MP2 level calculation time.

All of the functionals that used in chapter 2 for benchmarking purpose are listed in Figure A.3

Categories	Functional	HF used	Identifier(s) in bibtex file	comment
LDA	SVWN	0.0%	SV4 ; VWN	
hybrid-LDA	KMLYP	55.7%	KMLYP	mixture of slater and hf exchange
pure GGA	BP86	0.0%	Becke88 ; P86	
	PW91	0.0%	Perdew92a ; Perdew92b ; Perdew96	
	BPW91	0.0%	Becke88 ; Perdew92a ; Perdew92b ; Perdew96	
	BLYP	0.0%	Becke88 ; Lee88 ; Miehlich89	
	PBE(PBEPBE)	0.0%	PBE	
hybrid GGA	B3LYP	20.0%	Becke93 ; Lee88 ; Miehlich89 ; B3LYP	
	B3PW91	20.0%	Becke93 ; Perdew92a ; Perdew92b ; Perdew96	
	O3LYP	11.6%	O3LYP	
	BHandHLYP	50.0%	BHandHLYP	
	PBEO (PBE1PBE)	25.0%	PBEO	
	PBEOh1PBE	25.0%	PBEOh1PBE	the same as PBEO except using 1998 replace 1996 version
	B3P86	20.0%	Becke93 ; P86	
pure meta-GGA	TPSS	0.0%	TPSS	
	revTPSS	0.0%	Perdew09	
hybrid meta-GGA	B1B95	28.0%	Becke96	
	M06	27.0%	M06	
	M06L	0.0%	M06L	
	M06-HF	100.0%	M06HF	
	M06-2X	54.0%	M06	
	TPSSH	10.0%	TPSSH	
range-seperated	wB97XD	22.2% 100%	wB97XD	the first number is short range; the second number is long range
	wB97X	15.7% 100%	wB97X	
	Cam-B3LYP	19% 65%	camB3LYP	
	HSE06	25% 0%	HSE061 ; HSE062	

Figure A.3: Functionals used in Chapter 2 for benchmarking purpose.

*Basis set* To actually solve the equation numerically, the  $\Psi$  is expanded in a finite set of basis functions. There are many choices on how to choose the basis functions. Common choices include: Gaussian functions, plane wave functions, augmented wave functions etc. In most quantum chemistry software, especially the Gaussian09 software (*Gaussian09 Revision C.01*), gaussian functions are used. But in periodic system calculations with software like quantum espresso, the plane wave functions are used.

In the Cartesian coordinates, the  $\Psi$  is usually expanded using Gaussian functions in the following format:

$$\Psi_{\nu}[r] = \sum_{i=1} c_{i\nu} x^{l_x} y^{l_y} z^{l_z} e^{\zeta_{i\nu}(\mathbf{r}-\mathbf{R})^2} \quad (\text{A.15})$$

Different expansion series, of course, would give different energy results. However, we would expect the energy to converge to a certain value as the size of basis sets grow larger. This is exactly the case we observed with DFT calculations in the following Table A.4:

CP-OPT calculated (kcal/mol) for the water dimer.

	Basis	B3LYP	M05	M05-2X	M06	M06-2x	B2PLYP	B2PLYPD	X3LYP	MPWB1K	B97D
<b>Functions</b>											
aug-cc-pVQZ	574	-4.55	-4.98	-5.14	-4.54	-5.07	-4.77	-5.19	-4.92	-4.58	-4.42
cc-pV5Z	402	-4.55	-5.01	-5.14	-4.55	-5.06	-4.75	-5.17	-4.92	-4.59	-4.42
aug-cc-pVQZ	344	-4.54	-4.96	-5.16	-4.47	-5.10	-4.76	-5.18	-4.92	-4.59	-4.42
cc-pVQZ	230	-4.55	-4.94	-5.13	-4.52	-5.06	-4.71	-5.12	-4.91	-4.58	-4.41
aug-cc-pVTZ	184	-4.52	-5.05	-5.17	-4.63	-5.11	-4.70	-5.12	-4.90	-4.57	-4.41
6-311++G(3df,2p)	118	-4.54	-5.16	-5.15	-4.80	-5.11	-4.69	-5.10	-4.92	-4.62	-4.44
cc-pVTZ	116	-4.55	-5.03	-5.14	-4.73	-5.09	-4.62	-5.02	-4.90	-4.60	-4.44
aug-cc-pVDZ	82	-4.49	-5.18	-5.05	-4.67	-5.05	-4.60	-5.02	-4.88	-4.64	-4.39
6-311++G(d,p)	72	-5.05	-5.72	-5.74	-5.38	-5.68	-5.00	-5.38	-5.44	-5.18	-4.89
D95++(d,p)	64	-5.22	-5.72	-5.89	-5.33	-5.79	-5.17	-5.57	-5.61	-5.27	-5.07
6-31++G(d,p)	62	-5.20	-5.81	-5.82	-5.27	-5.82	-5.21	-5.61	-5.60	-5.36	-5.09
6-311G(d,p)	60	-5.09	-5.65	-5.74	-5.41	-5.64	-5.03	-5.39	-5.45	-5.23	-4.91
D95(d,p)	52	-5.74	-6.09	-6.25	-5.77	-6.19	-5.62	-6.03	-6.12	-5.72	-5.57
6-31G(d,p)	50	-5.48	-5.95	-6.00	-5.60	-6.03	-5.45	-5.85	-5.84	-5.66	-5.41
cc-pVDZ	48	-4.65	-5.20	-5.12	-4.91	-5.13	-4.61	-4.97	-5.00	-4.84	-4.53
6-31G(d)	38	-5.77	-6.18	-6.28	-5.76	-6.34	-5.74	-6.16	-6.15	-5.90	-5.70

Figure A.4: Basis sets effect in DFT calculations (from Plumley and Dannenberg 2011). Under a given functional, the result of energy calculation converged to a certain value as basis sets grow larger.

### The application of DFT

DFT succeeds in many ways, but it also has some intrinsic defects that should be aware of when we apply them to the real world tasks. The weaknesses include (Cohen, Mori-Sánchez, and Yang 2008) (Jacquemin et al. 2008) (Sousa, Fernandes, and Ramos 2007) :

1. Dispersive interactions (e.g. Van der Waals force) ;
2. Rydberg excitations ;
3. Charge-transfer (CT) state ;
4. Nonlinear optical properties of long conjugated polymers ;
5. Strongly correlated systems .

The origins of those difficulties lie in the approximations that were made in the exchange-correlation functionals, since the exact form is unknown. However, as new concepts emerge, these weaknesses are slowly being overcome: new functionals

are constructed to reflect effort from the functional development community (Cohen, Mori-Sánchez, and Yang 2008).

The molecules being investigated in this research would include long conjugated molecules, where electron delocalization error would be invoked. After previous benchmark studies, people found that tuning of the percentage of the Hartree-Fock exchange and the use of range-separated functionals could improve the results (Cohen, Mori-Sánchez, and Yang 2008) (Körzdörfer and Bredas 2014). In order to choose the best functional for our compounds, we performed a benchmark study with the functionals listed in Figure A.3.

In organic chemistry, the eigenvalues of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are the two most important molecular orbitals, because they determine the bandgap that directly related to the optical property of molecules. Gang Zhao and Charles B. Musgrave (2007) calculated the HOMO and LUMO energies from different DFT functionals and did a comparison study with the corresponding experimental values (Zhang and Musgrave 2007). The functionals they chose include SVWN, BLYP, BP86, BPW91, PBE, B3LYP, KMLYP, BH&HLYP, O3LYP and B1B95. Those functionals ranges from LDA to hybrid meta-GGA level A.3. The calculations include a series of 27 small molecules. As expected, HOMO energies from DFT generally have a better correlation (linear relation) with the experiment than LUMO energies as shown in Figure A.5. Among functionals used, KMLYP (hybrid-LDA level), having an average error of 0.24 eV, behaved the best in HOMO energies predictions. However, in LUMO comparison, they show hardly any correlations. They also used TD-DFT to calculate the HOMO-LUMO gap, which has a better correlation than LUMO energies from DFT. So a better strategy to acquire LUMO energies will be adding HOMO energies and HOMO-LUMO gaps together instead of calculating LUMO energies from DFT.

Zhao *et al.* (2003) performed similar calculations on 52 molecules and their focus is on B3LYP with 6-31 pVTZ+1 basis set for properties calculation and 6-31+G\* for geometry determination. In this study, they were able to perform a linear regression on the calculation results and found a linear relationship between theory and experimental results on the HOMO energy, LUMO energy and other molecular properties' results. This is to show a general and excellent linear correlation between theoretical and experimental results (refer to Figure A.6).

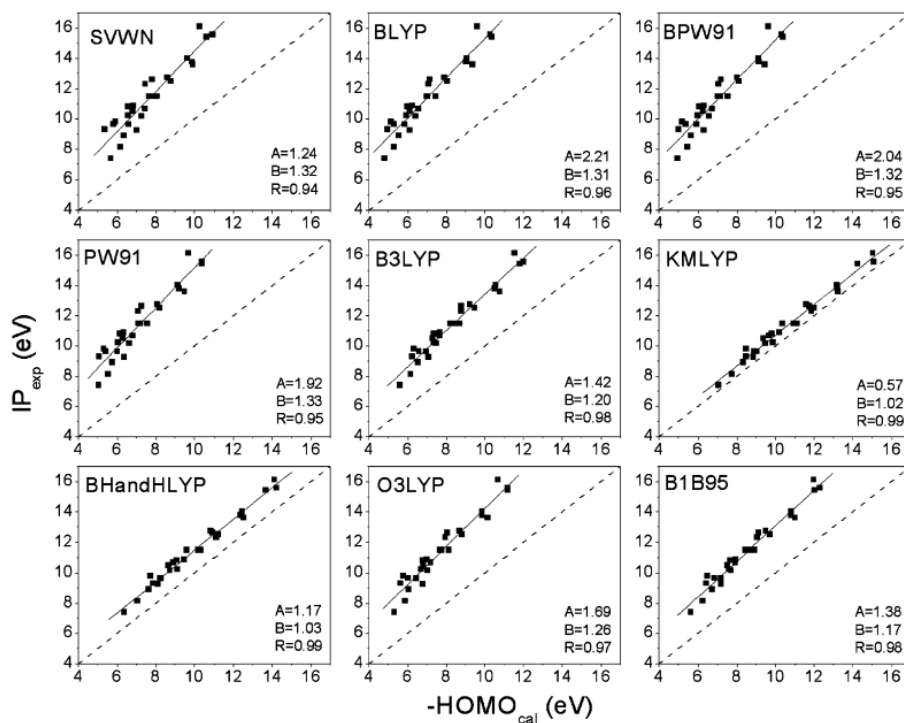


Figure A.5: HOMO energies from calculations and experimental ionization potential (IPs) (Zhang and Musgrave 2007). The 100% correlation line is indicated by the dashed line.

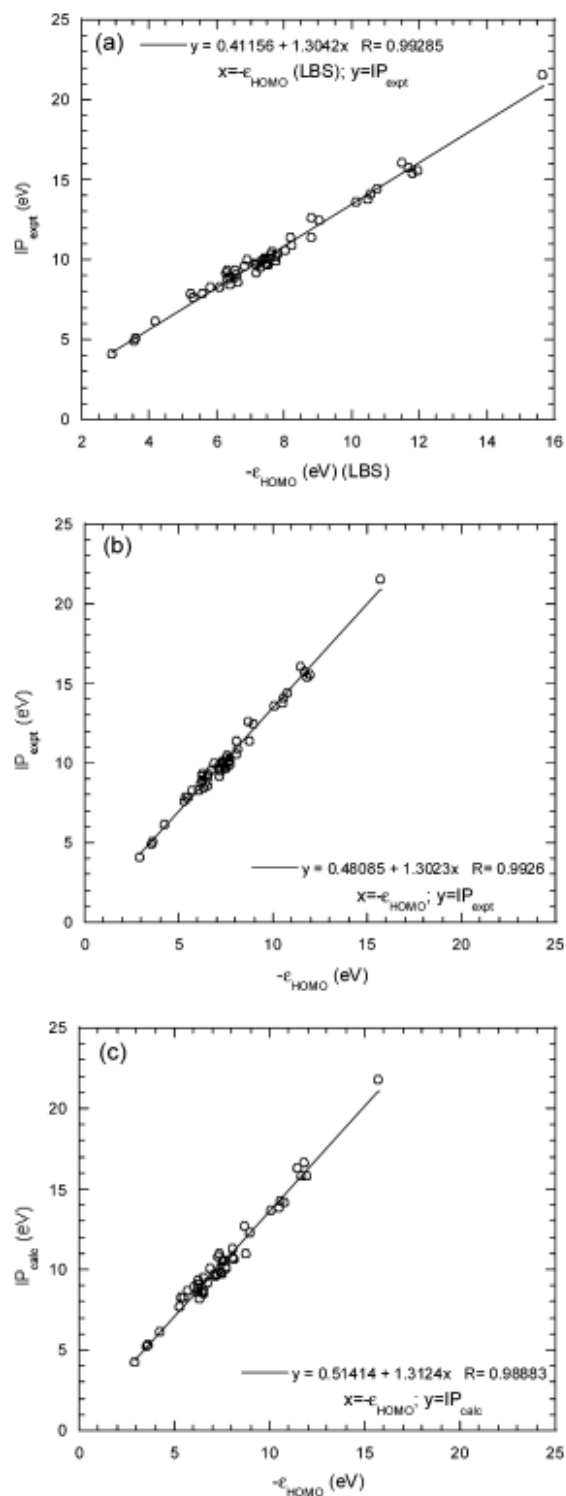


Figure A.6: Experimental values compared to theoretical values shows a linear dependence: a. Calculations with a larger basis set; b and c. Calculations with 6-31+G\* basis set (Zhan, Nichols, and Dixon 2003).



Tortorella *et al.*(2016) and their collaborators Martinelli *et al.*(2014) synthesized a new family of OPV molecules based on benzofulvene. In order to perform the theoretical study on those new molecules, they first performed a benchmark study on DFT and semi-empirical methods and tried to find out which method was better in predicting properties that could match the experimental results. Their candidates in DFT functionals included: B3LYP, cam-B3LYP, HF, PBE0, MPW1K,  $\Omega$ B97XD, M06, M06-2x, M06-HF as well as other ab initio methods like MP2. They found geometry predictions from B3LYP matched x-ray data the best. And concerning predictions on optical bandgaps, MPW1K delivered the most accurate results.

With range-separated functionals (LRC-PBE, LRC-PBE0 and LRC-BLYP), Mary A. Rohrdanz and John M. Herbert (Rohrdanz and Herbert 2008) did a survey study on 109 molecules from the Minnesota Thermochemistry and Thermochemical Kinetics Database. For the ionization energies, three functionals could give a better results than B3LYP's when we properly set the range-separation parameter  $\mu$  to a smaller value (as shown in Figure A.7).

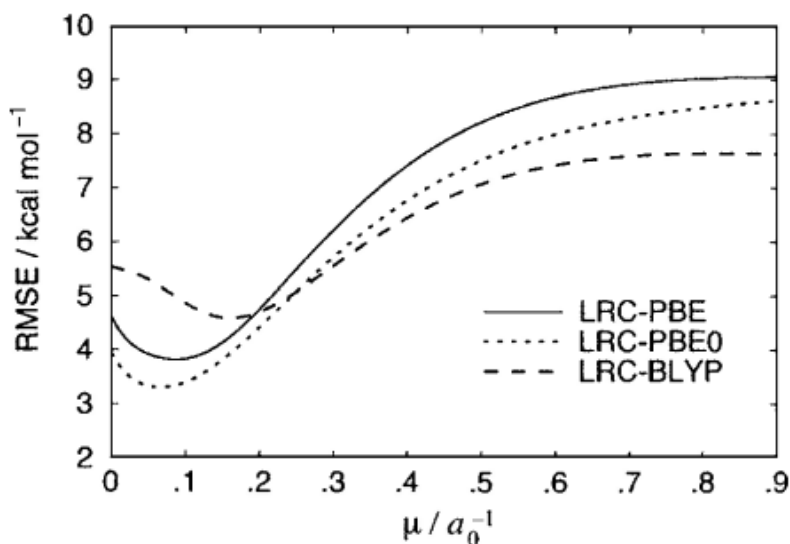


Figure A.7: Curve of RMSE values variation with different value of range-separated parameters (Rohrdanz and Herbert 2008).

### A.0.2 Molecular modeling using TD-DFT

DFT is accurate for ground state energy or the HOMO energy in this research. It could provide excited state information, but it is not accurate. A better method for

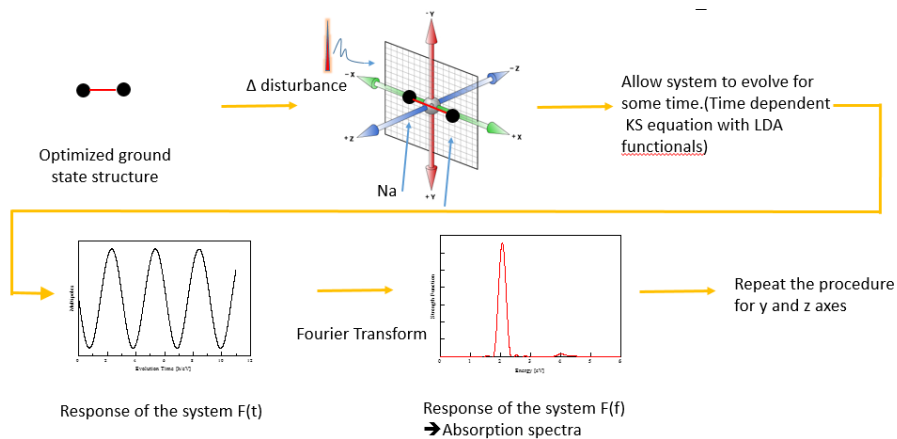
excited state calculation is Time-dependent DFT or TD-DFT.

The fundamental of TD-DFT lies in the Runge-Gross Theorem (RG theorem) (Runge and Gross 1984b). Like HK theorem in DFT, where we set up the relation between single electron density and the energy of the system, the RG theorem sets up the 1-to-1 mapping between the time evolution of the single electron density and the action of the system:

$$A[\rho] = \int_{t_0}^{t_1} \langle \Psi(t) | i\partial_t - H(t) | \Psi(t) \rangle dt \quad (\text{A.16})$$

and the stationary point of the action would give the real density of the system.

TD-DFT method gives better results for optical spectra. In this research work, we use TD-DFT method implemented in the octopus code (Andrade et al. 2015). The procedure of calculating spectra is illustrated in the Figure A.8.



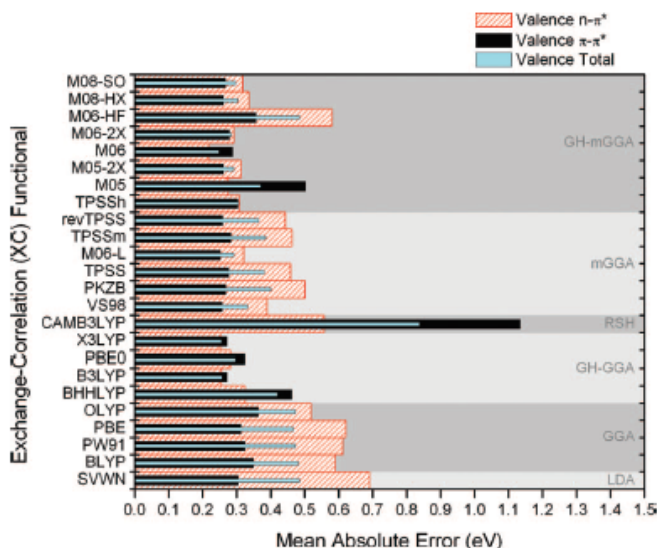
**Figure A.8: Spectra calculation procedures of TD-DFT implemented in the octopus code.**

### The application of Time-dependent DFT (TD-DFT)

Time-dependent DFT or TD-DFT is now the most widely used tool for modeling electronic spectra (Jacquemin et al. 2009) (Laurent and Jacquemin 2013) (Barone and Polimeno 2007). As we know, DFT has limits in treating delocalization but the molecular excitations could be of any type. Specifically,  $n \rightarrow \pi$  is local excitation while  $\pi \rightarrow \pi^*$  is of delocalization nature. So when we deal with excited state problem, it is also important to identify the nature of transitions and then choose the functional correspondingly.

In a study by Sarom and co-workers (Leang, Zahariev, and Gordon 2012), they tested the performance of 24 density functionals on predicting the excited state energy

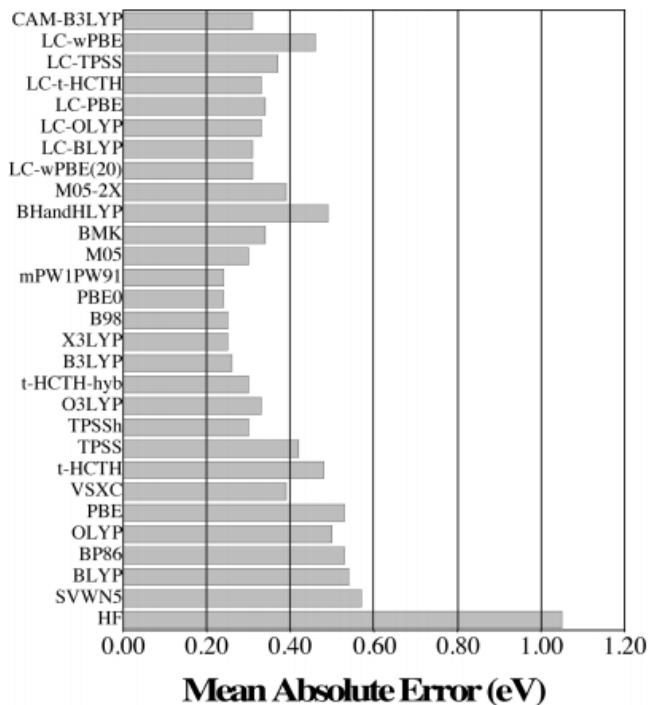
from 110 transitions by using a test set of 14 compounds. Those 24 functionals included LDA, GGA, hybrid GGA, meta GGA (mGGA) and hybrid mGGA. In terms of mean absolute error (MAE), M06-2x ranked first with 0.22 eV but it suffers from convergence problems sometimes. The next best is PBE0 functional with the MAE being 0.28 eV. X3LYP and B3LYP are the third and fourth respectively (For their complete list of functionals see Figure A.9). There is a high error from cam-B3LYP. There is a report saying cam-B3LYP couldn't deal with triplet excitation properly (Peach et al. 2008) and there are quite a number of triplet excitation in this test set.



**Figure A.9: Mean absolute errors for  $\pi \rightarrow \pi^*$  and  $n \rightarrow \pi^*$  predictions (Leang, Zahariev, and Gordon 2012).**

Since singlet excitation is the one seen most, Denis and co-workers (Jacquemin et al. 2009) did an extensive TD-DFT benchmarking using only singlet state excitations. They performed TD-DFT on 500 compounds and more than 700 singlet transitions with 29 functionals including LDA, GGA, hybrid GGA, meta GGA, hybrid meta GGA and range-separated functionals. They used both experimental values (VE set) and theoretical values from CAS-PT2/TZVP (VT set) as the reference. The statistical analysis for both sets are shown in Figure A.10 and Figure A.11. From there, we can see the 4 global hybrid functionals have less than 0.25 eV in MAE in both tests in the VE and VT set. And B3LYP is just a little higher than these 4 functionals, so it is still worth using considering its convergence speed. As one of their conclusions, they think with current TD-DFT functionals, calculation error should be within 0.22 eV and this could be achieved with the 4 global hybrid functionals (X3LYP, B98, PBE0,

mPW91PW91) or the "LC-" hybrid functional with  $\omega = 20$ .



**Figure A.10: Statistical analysis of the results using VT set as the reference. MSE is Mean Signed Error. MAE is Mean Average Error. RMS is Root Mean Square (Jacquemin et al. 2009).**

While those benchmark tests seemed to recognize the importance of global hybrid functionals (*i.e.* PBE0), range-separated functionals (*i.e.* LC-, cam-B3LYP) are not outperformed often. However, because our study objects, organic photovoltaic molecules, usually have a special phenomenon called "charge-transfer (CT)" state, range-separated functionals are the one people think could describe this phenomenon properly (Laurent and Jacquemin 2013) (Dreuw, Weisman, and Head-Gordon 2003). Unlike traditional hybrid functionals, range-separated functionals usually have different exchange percentages for local and long range. For example, the Heyd-Scuseria-Ernzerhof (HSE) functional (Heyd, Scuseria, and Ernzerhof 2003b) (Paier et al. 2006) has 25% of Hartree-Fock exchange for short range and for its long range, it contains 0% of Hartree-Fock exchange.

Even for the range-separated functionals, describing CT properly proved to be a difficult task. In a study on donor-acceptor complexes by Nguyen and coworkers (Nguyen, Day, and Pachter 2011b), they tested several range-separated functionals (LC-PBE, LC- $\Omega$ PBE, cam-B3LYP, ca-B3LYP, LRC- $\Omega$ PBEh, CA-PBE,  $\Omega$ B97,

functional	before fitting			after linear regression		
	MSE	MAE	RMS	$R^2$	MAE	RMS
HF	-0.82	0.85	0.96	0.88	0.37	0.48
SVWN5	0.32	0.41	0.48	0.94	0.28	0.35
BLYP	0.32	0.40	0.47	0.94	0.28	0.35
BP86	0.29	0.38	0.46	0.94	0.28	0.34
OLYP	0.29	0.38	0.45	0.94	0.28	0.35
PBE	0.29	0.39	0.46	0.94	0.28	0.34
VSXC	0.15	0.32	0.39	0.94	0.27	0.34
$\tau$ -HCTH	0.27	0.37	0.44	0.94	0.28	0.34
TPSS	0.20	0.34	0.41	0.94	0.27	0.34
TPSSh	0.05	0.26	0.32	0.95	0.24	0.30
O3LYP	0.11	0.26	0.32	0.95	0.24	0.30
$\tau$ -HCTH-hyb	0.06	0.24	0.31	0.96	0.23	0.29
B3LYP	0.01	0.23	0.29	0.96	0.22	0.28
X3LYP	-0.01	0.22	0.28	0.96	0.22	0.28
B98	-0.04	0.22	0.29	0.96	0.22	0.28
PBE0	-0.08	0.22	0.29	0.96	0.21	0.27
mPW1PW91	-0.08	0.22	0.29	0.96	0.21	0.27
M05	-0.02	0.25	0.31	0.95	0.25	0.30
BMK	-0.26	0.32	0.39	0.96	0.22	0.27
BHLYP	-0.36	0.40	0.47	0.95	0.23	0.29
M05-2X	-0.29	0.38	0.45	0.95	0.25	0.32
LC- $\omega$ PBE(20)	-0.08	0.22	0.27	0.96	0.20	0.26
LC-BLYP	-0.31	0.35	0.41	0.96	0.22	0.27
LC-OLYP	-0.34	0.37	0.43	0.96	0.21	0.27
LC-PBE	-0.34	0.38	0.44	0.96	0.21	0.26
LC- $\tau$ -HCTH	-0.32	0.36	0.42	0.96	0.22	0.28
LC-TPSS	-0.38	0.40	0.46	0.96	0.21	0.26
LC- $\omega$ PBE	-0.46	0.48	0.54	0.96	0.23	0.29
CAM-B3LYP	-0.25	0.30	0.36	0.96	0.21	0.26
MLR, eq 2				0.98	0.16	0.20
MLR-P, eq 3				0.97	0.20	0.25
MLR-B, eq 4				0.97	0.20	0.26

**Figure A.11: Statistical analysis of the results using VE set as the reference (Jacquemin et al. 2009).**

$\Omega$ B97X, HSE, CA0-PBE and CA0-B3LYP) and concluded mean average errors of predictions from  $\Omega$ B97X and LC-PBE are below 0.2 eV. In another study by Dev and others (2012), they chose PBE, M06L, B3LYP, M06, cam-B3LYP and  $\Omega$ B97 to perform calculations on 16 push-pull dyes used in solar cells. This time cam-B3LYP was the best one with a MAE of 0.23 eV as shown in Figure A.12.

There are many different types of excitations and it is difficult to find a universal functional that could predict values matching the experiment well. So an alternative way would be tuning a set of parameters for a certain molecule family. In the future, we would choose to tune the parameters of HSE functionals to predict values for the melanin family.

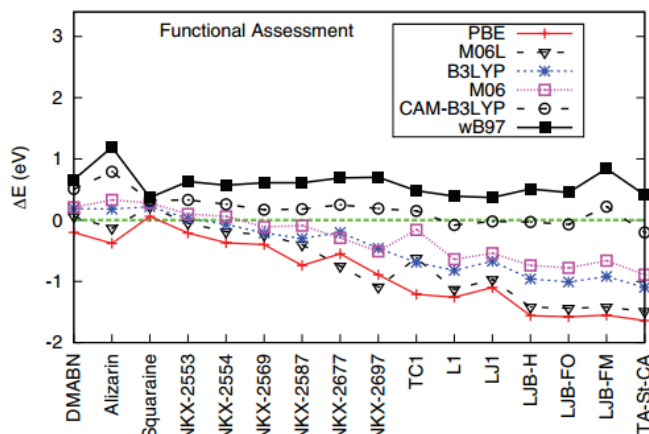


Figure A.12: Statistical analysis of the difference between the values from different functionals and the reference values (Dev, Agrawal, and English 2012b).

### A.0.3 Virtual molecule screening and material informatics

Empowered by the growing technology in computer hardware development, people were able to do high throughput screening (HTS) first in drug design field in the last century and it proved to be an effective method (Young et al. 1998) (Hamasaki and Rando 1998). People in the material design field borrowed the idea and started many new projects utilizing this method.

The Harvard Clean Energy Project (Hachmann et al. 2014), uses chemoinformatics as well as DFT calculations, to investigate a database built from 2.3 million molecular motifs in order to explore their organic electronic properties. After performing large scale DFT calculations using the IBM World Community Grid, they were able to rank the candidates according to the predicted power conversion efficiency (PCE) that was calculated from the Scharber model (Scharber et al. 2006). Their results are online and open to the public. Another survey study by Ørnsø *et al.* (2013) focused on dye selections, screening out 50 candidates out of 1029 molecules that were built from porphyrin derivatives. DFT calculations with PBE functionals were used to predict the HOMO and LUMO energy. Their results are online and open to public. Kanal et al (Kanal et al. 2013b) adopted a slightly different way to analyze their results. Their aims were not only to find out the candidates but also to discover the trend and the design rules underlying their results. Instead of the Donor-Acceptor pattern, they proposed a new type of Donor-Donor pattern which is concluded from their findings. Bérubé et al used DFT as the primary tool to perform

quantum chemistry computations on OPV (Bérubé et al. 2013). By using the Scharber model (Scharber et al. 2006), they set up a prediction protocol and discussed the prediction results of PCE of the material. Overall, the quantum chemistry start to show the possibility of being a versatile yet relatively independent study method.

The last 3 years witnessed an emergence of a new field called material informatics. The current leading projects include: the material genome project and the material project. They usually use combined methods from several disciplines including quantum chemistry, drug design and chemoinformatics to provide guidance in designing and discovering new material for certain purposes.

## APPENDIX B

### Parameters from Linear Regressions in the Chapter 2

This appendix lists the linear regress equation from chapter 2. After performing DFT/TD-DFT calculations, we applied these correction formula to the results to improve the accuracy of the prediction. The correction formula:

$$HOMO_{exp} = HOMO_{DFT} * slope + interception$$

$$Gap_{exp} = Gap_{TD-DFT} * slope + interception$$



**Table B.1: Linear regression formula for DFT functionals on HOMO and on Gap.**

Parameters	HOMO		Gap	
	slope	interception	slope	interception
B3LYP	0.698196693	-1.482445666	0.847553522	0.153889895
HSE06	0.654625822	-1.940596143	0.793098208	0.426161495
PBE	0.660285439	-2.194546975	0.836565766	0.71375629
PBE0	0.680488568044	-1.56432027356	0.829372219306	0.242813994713
BHandHLYP	0.642863289391	-1.31685693837	0.859148916136	-0.310281077773
BLYP	0.68485221044	-2.17601557297	0.820743018835	0.753916335271
KMLYP	0.640248194409	-0.92925053598	0.772131100492	-0.0416018769641
M06	0.668199588454	-1.55863605803	0.876368923513	0.118870582769
M06HF	0.594032137495	-0.528713904122	1.10588106819	-1.60154081271
M06L	0.68547196989	-2.05060424355	0.7979641837	0.632691698669
M062x	0.70641093563	-0.729482631055	0.935951772063	-0.497475930265
O3LYP	0.669431711131	-1.95956466635	0.820492991905	0.476345493934
PW91	0.659997100364	-2.16274552067	0.834275128588	0.719654786427
TPSS	0.672825250192	-2.15642802156	0.792983948119	0.746669318164
RevTPSS	0.675290133666	-2.1450240197	0.791625091404	0.73720741968
SVWN	0.64908365031	-1.8336367923	0.832516308403	0.733279874945
cam-B3LYP	0.647689550556	-1.05392432642	0.90511560809	-0.431626015365
wB97XD	0.648538339186	-0.705319438125	1.00099575609	-0.809206435982
wB97X	0.652626950936	-0.450975629674	1.05687304596	-1.23303351773
LC-wPBE	0.664835817468	-0.157590475097	1.11199228869	-1.61222766889

VITA

Shuo Dai

Candidate for the Degree of

Doctor of Philosophy

Dissertation: Density Functional Theory and Time-Dependent DFT Modeling of Organic Photovoltaic Materials

Major Field: Physics

Biographical:

Education:

Completed the requirements for the degree of Doctor of Philosophy with a major in Physics at Oklahoma State University, Stillwater, Oklahoma in December, 2017.

Received the degree of Master of Science with a major in Biophysics at Oklahoma State University, Stillwater, Oklahoma, in July, 2013.

Received the degree of Bachelor of Science with a major in Physics at Nanjing University, Nanjing, Jiangsu, China, in July 2004.