DEVELOPMENT AND BENCHMARKING OF A SEMILOCAL DENSITY-FUNCTIONAL APPROXIMATION INCLUDING DISPERSION

by

Felix O. Kannemann

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
February 2013

© Copyright by Felix O. Kannemann, 2013
The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “DEVELOPMENT AND BENCHMARKING OF A SEMILOCAL DENSITY-FUNCTIONAL APPROXIMATION INCLUDING DISPERSION” by Felix O. Kannemann in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: February 22, 2013

External Examiner: ____________________________
Ajit J. Thakkar

Research Supervisor: ____________________________
Axel D. Becke

Examinining Committee:
Russell J. Boyd

Donald F. Weaver

Josef W. Zwanziger

Departmental Representative: ____________________________
Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions. I understand that my thesis will be electronically available to the public.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author’s written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in the thesis (other than brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

Signature of Author
“En raison de l’indéterminisme, le futur de maintenant apparaît comme distinct du passé qui sera… Le futur de maintenant est beaucoup plus riche en possibilités que le passé qui sera.”

Contents

List of Tables .................................. ix

List of Figures .................................. xi

Abstract ........................................ xiv

List of Abbreviations Used ..................... xv

Acknowledgements ................................ xix

Chapter 1 Introduction .......................... 1

Chapter 2 Theoretical Background ............. 4

2.1 Overview .................................... 4

2.2 Density-Functional Theory ................... 5

2.3 Exchange-Correlation Functionals .......... 10
  2.3.1 Local Density Approximation ............ 11
  2.3.2 Generalized Gradient Approximation ... 11

2.4 Dispersion In DFT .......................... 14
  2.4.1 The Exchange-Hole Dipole Moment Model 16
  2.4.2 Perturbation Theory Of The Dispersion Interaction 17
  2.4.3 Interatomic Decomposition ............... 19

2.5 Exchange-Correlation Holes In DFT ........ 20
  2.5.1 The Becke-Roussel Exchange Hole Model 21

2.6 Basis Sets ................................. 23
  2.6.1 Overview ................................ 23
  2.6.2 Examples ................................ 24
  2.6.3 Pople-Type Basis Sets ................... 29
  2.6.4 Ahlrichs-Type (def2) Basis Sets ........ 30
  2.6.5 Correlation-Consistent Basis Sets ....... 31
  2.6.6 Polarization Consistent Basis Sets ...... 32
6.5.1 Basis-Set Dependence Of XDM Dispersion Energies . 110
6.5.2 Counterpoise Correction And Basis-Set Superposition Error . 110
6.5.3 Rare-Gas Binding Energies With Pople-Type Basis Sets . 118
6.5.4 Rare-Gas Binding Energies With The Density Fitting Approximation . 122
6.6 Conclusions . 127

Chapter 7 Benchmarking Of The PW86PBE-XDM Method For Intermolecular Interactions 128

7.1 Introduction . 128
7.2 Computational Details . 129
7.3 Intermolecular Interactions At Equilibrium Geometries . 130
7.3.1 Overall Accuracy Of PW86PBE-XDM For The S22 And S66 Benchmark Sets . 130
7.3.2 Basis Set Effects . 135
7.3.3 Basis Set Effects By Interaction Type . 140
7.4 Intermolecular Interactions At Non-Equilibrium Geometries . 147
7.4.1 Overall Accuracy Of Intermolecular Dissociation Curves . 148
7.4.2 Detailed Assessment . 153
7.4.3 Exemplary Potential Energy Curves And Basis Set Effects . 158
7.4.4 Angular-Displaced Geometries . 164
7.5 Conclusions . 164
7.6 Appendix . 167

Chapter 8 Atomic Volumes And Polarizabilities 175

8.1 Introduction . 175
8.2 Computational Methods . 177
8.3 Results And Discussion . 177
8.4 Conclusions . 181
8.5 Appendix: Statistical Theory Of Atomic Polarizability . 182

Chapter 9 Conclusion 185

Appendix A Copyright Permissions 187
List of Tables

2.1 Double zeta basis sets employed in this work. ............ 27
2.2 Triple zeta basis sets employed in this work. ............ 28
2.3 Quadruple zeta and larger basis sets employed in this work. 28
2.4 Summary of basis sets employed in this work ............. 33
3.1 Hartree-Fock repulsion energies of rare-gas diatomics ..... 39
3.2 Repulsion energy errors of exchange GGAs (post-HF) .... 39
3.3 Repulsion energy errors of exchange GGAs (post-xLDA) ... 39
3.4 Exact and PW86 repulsion energies of rare-gas diatomics . 40
3.5 Best-fit dispersion damping parameters .................. 43
3.6 Reference and calculated reduced equilibrium separations . 45
3.7 Reference and calculated reduced equilibrium binding energies 45
3.8 Atomization energy errors ............................. 46
4.1 Binding energies and binding energy errors for the training set of 65 vdW complexes .......................... 55
4.2 Optimized dispersion damping parameters and error statistics for the training set of 65 vdW complexes .............. 56
4.3 Error statistics for the current set of 65 vdW complexes using rare-gas-optimized damping parameters .............. 56
4.4 Mean absolute errors of various DFT methods for the S22 set and its subsets ................................ 57
5.1 Repulsion Energies in Rare-Gas Diatomics .................. 63
5.2 Binding Energies of Intermolecular Complexes ............. 68
6.1 Basis sets and abbreviated names used in this Chapter ...... 73
6.2 Basis sets used in this Chapter ............................. 74
6.3 Equilibrium separations, reference binding energies, and Hartree-Fock binding energies of rare-gas diatomics ............. 77
6.4 HF repulsion energies of rare-gas diatomics (μEh) .......... 78
6.5 CP-corrected HF repulsion energies of rare-gas diatomics ($\mu E_h$) ..................................... 83
6.6 HF and PW86 repulsion energies of ten rare-gas diatomics ($\mu E_h$) 90
6.7 PW86 repulsion energies of rare-gas diatomics ($\mu E_h$) ...................................................... 91
6.8 CP-corrected PW86 repulsion energies of rare-gas diatomics ($\mu E_h$) .............................................. 95
6.9 HF and PW86 repulsion energy error statistics .......................................................... 100
6.10 Rare-gas binding energies at various levels of theory .................................................. 104
6.11 Total energies and total-energy deviations of rare-gas atoms at various levels of theory .......................................................... 106
6.12 Rare-gas dispersion coefficients at various levels of theory ........................................ 106
6.13 PW86PBE-XDM binding energies of rare-gas diatomics ($\mu E_h$) 108
6.14 Damped XDM dispersion energies of rare-gas diatomics ($\mu E_h$) 111
6.15 CP-corrected PW86PBE-XDM binding energies of rare-gas diatomics ($\mu E_h$) ...................................................... 113
6.16 Pople-type basis sets for rare-gas elements .......................................................... 120
6.17 PW86PBE-XDM binding energies of rare-gas diatomics for Pople-type basis sets ($\mu E_h$) ...................................................... 121
6.18 CP-corrected PW86PBE-XDM binding energies of rare-gas diatomics for Pople-type basis sets ($\mu E_h$) ...................................................... 121
6.19 Per-atom deviation of W06 density-fitted PW86PBE-XDM rare-gas total energies ...................................................... 124
6.20 Deviation of W06 density-fitted PW86PBE-XDM rare-gas binding energies ...................................................... 126

7.1 Binding energies for the S22 benchmark set ...................................................... 131
7.2 Mean absolute deviations of various DFAs for the S22 and S66 benchmarks set ...................................................... 133
7.3 Basis functions, Gaussian primitives, CPU time and S66 MAD ...................................................... 139
7.4 Binding energies for the S66 benchmark set ...................................................... 169

8.1 Free atomic volumes of elements H-Ar and Kr with various model chemistries ...................................................... 178
8.2 Free atomic polarizabilities and volumes of elements H-Lr ...................................................... 179
**List of Figures**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Enhancement factors of exchange GGAs</td>
<td>13</td>
</tr>
<tr>
<td>2.2</td>
<td>def2-SVP basis set for hydrogen</td>
<td>25</td>
</tr>
<tr>
<td>2.3</td>
<td>def2-SVP basis set for oxygen</td>
<td>26</td>
</tr>
<tr>
<td>3.1</td>
<td>Hartree-Fock repulsion, xLDA and exchange-only GGA interaction energies of Ne$_2$</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>He$_2$, Ne$_2$, Ar$_2$, and Kr$_2$ binding energy curves</td>
<td>44</td>
</tr>
<tr>
<td>6.1</td>
<td>HF repulsion energy % deviations of rare-gas diatomics</td>
<td>80</td>
</tr>
<tr>
<td>6.2</td>
<td>CP-corrected HF repulsion energy % deviations of rare-gas diatomics</td>
<td>84</td>
</tr>
<tr>
<td>6.3</td>
<td>CP-corrected and uncorrected (noCP) HF repulsion energy MAPD by basis set</td>
<td>86</td>
</tr>
<tr>
<td>6.4</td>
<td>CP-corrected and uncorrected (noCP) HF repulsion energy MAPD ordered by basis set size</td>
<td>88</td>
</tr>
<tr>
<td>6.5</td>
<td>PW86 repulsion energy % deviations of rare-gas diatomics</td>
<td>92</td>
</tr>
<tr>
<td>6.6</td>
<td>HF and PW86 repulsion energy mean absolute percent deviations (MAPD)</td>
<td>94</td>
</tr>
<tr>
<td>6.7</td>
<td>CP-corrected PW86 repulsion energy % deviations of rare-gas diatomics</td>
<td>96</td>
</tr>
<tr>
<td>6.8</td>
<td>CP-corrected and uncorrected (noCP) PW86 repulsion energy MAPD</td>
<td>97</td>
</tr>
<tr>
<td>6.9</td>
<td>CP-corrected HF and PW86 repulsion MAPD</td>
<td>99</td>
</tr>
<tr>
<td>6.10</td>
<td>CP-corrected and uncorrected HF and PW86 repulsion MAPD</td>
<td>101</td>
</tr>
<tr>
<td>6.11</td>
<td>CP-corrected and uncorrected HF and PW86 repulsion MAPD</td>
<td>102</td>
</tr>
<tr>
<td>6.12</td>
<td>PW86PBE-XDM binding energy percent deviations of rare-gas diatomics</td>
<td>109</td>
</tr>
<tr>
<td>6.13</td>
<td>Damped XDM dispersion energies of rare-gas diatomics</td>
<td>112</td>
</tr>
<tr>
<td>6.14</td>
<td>CP-corrected PW86PBE-XDM binding energy percent deviations of rare-gas diatomics</td>
<td>114</td>
</tr>
<tr>
<td>6.15</td>
<td>CP-corrected and uncorrected PW86PBE-XDM binding energy MAPD</td>
<td>116</td>
</tr>
</tbody>
</table>
6.16 CP-corrected and uncorrected PW86PBE-XDM binding energy MAPD .......................................................... 117

6.17 Uncorrected (no counterpoise) and CP-corrected PW86PBE-XDM binding energy percent deviations of rare-gas diatomics with Pople-type basis sets ........................................... 119

7.1 S66 and S22 mean absolute deviations of various DFAs for noncovalent interactions ........................................... 134

7.2 CP-corrected and uncorrected PW86PBE-XDM binding energy MAD (S66) ......................................................... 136

7.3 Correlation between CPU time and number of basis functions 141

7.4 Correlation between S66 MAD and CPU time ................. 142

7.5 Deviation of PW86PBE-XDM binding energies from S22 reference for Pople-type basis sets ............................. 144

7.6 Deviation of PW86PBE-XDM binding energies from S22 reference for cc-, def2, and pc-type basis sets ..................... 146

7.7 S66x8 mean absolute percent deviation (MAPD) by distance multiplier for various methods with the def2-QZVP basis set . 149

7.8 S66x8 mean percent deviation (MPD) by distance multiplier for various methods with the def2-QZVP basis set ............. 151

7.9 S66x8 mean percent deviation by distance multiplier for PW86PBE-XDM/def2-QZVP ........................................... 152

7.10 S66x8 reference binding energies by distance multiplier .... 154

7.11 S66x8 reference binding energies as percent fraction of the S66 reference equilibrium binding energy .................. 155

7.12 Percent deviation of PW86PBE-XDM/haug-cc-pVTZ binding energies from S66x8 reference by distance multiplier .... 157

7.13 Deviation of PW86PBE-XDM/haug-cc-pVTZ binding energies from S66x8 reference by distance multiplier .................. 159

7.14 Selected potential energy curves from the S66x8 set (π-π-stacked complexes) ..................................................... 160

7.15 Selected potential energy curves from the S66x8 set (hydrogen bonds and aliphatic-aliphatic dispersion interactions) . 162

7.16 Selected potential energy curves from the S66x8 set (π-aliphatic, T-shaped, XH-π and other interactions) .................. 163
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.17</td>
<td>Deviation of PW86PBE-XDM/haug-cc-pVTZ binding energies from S66a8 reference</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>values</td>
<td></td>
</tr>
<tr>
<td>7.18</td>
<td>CP-corrected and uncorrected PW86PBE-XDM binding energy MAD (S22)</td>
<td>170</td>
</tr>
<tr>
<td>7.19</td>
<td>Deviation of PW86PBE-XDM binding energies from S66 reference for Pople-type</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>basis sets (no CP)</td>
<td></td>
</tr>
<tr>
<td>7.20</td>
<td>Deviation of PW86PBE-XDM binding energies from S66 reference for Pople-type</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>basis sets (CP)</td>
<td></td>
</tr>
<tr>
<td>7.21</td>
<td>Deviation of PW86PBE-XDM binding energies from S66 reference for cc-, def2,</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>and pc-type basis sets (no CP)</td>
<td></td>
</tr>
<tr>
<td>7.22</td>
<td>Deviation of PW86PBE-XDM binding energies from S66 reference for cc-, def2,</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td>and pc-type basis sets (CP)</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Free atomic volumes and polarizabilities plotted against atomic number Z</td>
<td>180</td>
</tr>
</tbody>
</table>
Abstract

Density-functional theory has become an indispensable tool for studying matter on the atomic level, being routinely applied across diverse disciplines from solid-state physics to chemistry and molecular biology. Its failure to account for dispersion interactions has spurred intensive research over the past decade. In this thesis, a semilocal density-functional approximation including dispersion is developed by combining standard functionals for exchange and correlation with the nonempirical “exchange-hole dipole moment” (XDM) dispersion model of Becke and Johnson.

With a minimum of empiricism, the method accurately describes all types of noncovalent interactions, from the extremely weak dispersion forces in rare-gas systems to the hydrogen bonding and stacking interactions responsible for the structure and function of biological macromolecules such as DNA and proteins. The method is compatible with a wide variety of standard Gaussian basis sets, and is easily applied to any system that can be modeled with density-functional theory.
## List of Abbreviations Used

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACFDT</td>
<td>adiabatic-connection fluctuation-dissipation theorem</td>
</tr>
<tr>
<td>ALL</td>
<td>Andersson-Langreth-Lundqvist functional</td>
</tr>
<tr>
<td>B86</td>
<td>Becke (1986) exchange GGA</td>
</tr>
<tr>
<td>B86b</td>
<td>Becke (1986) exchange GGA</td>
</tr>
<tr>
<td>B88</td>
<td>Becke (1988) exchange GGA</td>
</tr>
<tr>
<td>BE</td>
<td>binding energy</td>
</tr>
<tr>
<td>BJ</td>
<td>Becke-Johnson</td>
</tr>
<tr>
<td>BR</td>
<td>Becke-Roussel exchange meta-GGA</td>
</tr>
<tr>
<td>BSIE</td>
<td>basis-set incompleteness error</td>
</tr>
<tr>
<td>BSSE</td>
<td>basis-set superposition error</td>
</tr>
<tr>
<td>CBS</td>
<td>complete basis set</td>
</tr>
<tr>
<td>cc</td>
<td>correlation-consistent</td>
</tr>
<tr>
<td>CC</td>
<td>coupled cluster</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>coupled cluster singles and doubles with perturbative triples</td>
</tr>
<tr>
<td>cGTO</td>
<td>contracted Gaussian-type orbital</td>
</tr>
<tr>
<td>CI</td>
<td>configuration interaction</td>
</tr>
<tr>
<td>CP</td>
<td>counterpoise</td>
</tr>
<tr>
<td>DCP</td>
<td>dispersion-correcting pseudopotential</td>
</tr>
<tr>
<td>DFA</td>
<td>density functional approximation</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>DFT-D</td>
<td>density functional theory with empirical dispersion correction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>DKH</td>
<td>Douglas-Kroll-Hess scalar relativistic Hamiltonian</td>
</tr>
<tr>
<td>DZ</td>
<td>double zeta</td>
</tr>
<tr>
<td>GGA</td>
<td>generalized gradient approximation</td>
</tr>
<tr>
<td>GTO</td>
<td>Gaussian-type orbital</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree–Fock</td>
</tr>
<tr>
<td>HK</td>
<td>Hohenberg-Kohn</td>
</tr>
<tr>
<td>LCAO</td>
<td>linear combination of atomic orbitals</td>
</tr>
<tr>
<td>LDA</td>
<td>local density approximation</td>
</tr>
<tr>
<td>LRD</td>
<td>local response dispersion</td>
</tr>
<tr>
<td>LSDA</td>
<td>local spin density approximation</td>
</tr>
<tr>
<td>KS</td>
<td>Kohn-Sham</td>
</tr>
<tr>
<td>KS-DFT</td>
<td>Kohn-Sham density functional theory</td>
</tr>
<tr>
<td>MAD</td>
<td>mean absolute deviation</td>
</tr>
<tr>
<td>MAE</td>
<td>mean absolute error</td>
</tr>
<tr>
<td>MAPD</td>
<td>mean absolute percent deviation</td>
</tr>
<tr>
<td>MAPE</td>
<td>mean absolute percent error</td>
</tr>
<tr>
<td>MaxAPD</td>
<td>maximum absolute percent deviation</td>
</tr>
<tr>
<td>MaxAPE</td>
<td>maximum absolute percent error</td>
</tr>
<tr>
<td>MaxPD</td>
<td>maximum percent deviation</td>
</tr>
<tr>
<td>MaxPE</td>
<td>maximum percent error</td>
</tr>
<tr>
<td>MBPT</td>
<td>many-body perturbation theory</td>
</tr>
<tr>
<td>MP2</td>
<td>second-order Møller-Plesset perturbation theory</td>
</tr>
</tbody>
</table>
MD  mean deviation
ME  mean error
MPD mean percent deviation
MPE mean percent error
MaxAD maximum absolute deviation
MaxAE maximum absolute error
MO  molecular orbital
NAO numerical atom-centered orbital
NCI noncovalent interaction
P86 Perdew (1986) correlation GGA
PBE Perdew-Burke-Ernzerhof (1996) exchange-correlation GGA
pc  polarization consistent
PW86 Perdew-Wang (1986) exchange GGA
PW91 Perdew-Wang (1991) exchange-correlation GGA
QZ  quadruple zeta
RE  repulsion energy
revPBE revised PBE exchange GGA
RG  rare-gas
RMSPD root-mean-square percent deviation
RMSPE root-mean-square percent error
RPA random phase approximation
RSH range-separated hybrid
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCF</td>
<td>self-consistent field</td>
</tr>
<tr>
<td>STO</td>
<td>Slater-type orbital</td>
</tr>
<tr>
<td>SV</td>
<td>split-valence</td>
</tr>
<tr>
<td>TF</td>
<td>Thomas-Fermi</td>
</tr>
<tr>
<td>TFD</td>
<td>Thomas-Fermi-Dirac</td>
</tr>
<tr>
<td>TT</td>
<td>Tang-Toennies</td>
</tr>
<tr>
<td>TZ</td>
<td>triple zeta</td>
</tr>
<tr>
<td>UEG</td>
<td>uniform electron gas</td>
</tr>
<tr>
<td>UGBS</td>
<td>universal Gaussian basis set</td>
</tr>
<tr>
<td>vdW</td>
<td>van der Waals</td>
</tr>
<tr>
<td>vdW-DF</td>
<td>van der Waals density functional</td>
</tr>
<tr>
<td>VWN</td>
<td>Vosko, Wilk, and Nusair correlation LSDA</td>
</tr>
<tr>
<td>W06</td>
<td>Weigend (2006) auxiliary (density fitting) basis set</td>
</tr>
<tr>
<td>WFT</td>
<td>wavefunction theory</td>
</tr>
<tr>
<td>XC</td>
<td>exchange-correlation</td>
</tr>
<tr>
<td>XDM</td>
<td>exchange-hole dipole moment</td>
</tr>
<tr>
<td>xLDA</td>
<td>exchange-only local density approximation</td>
</tr>
<tr>
<td>XX</td>
<td>exact exchange</td>
</tr>
</tbody>
</table>
Acknowledgements

First of all I would like to express my gratitude to Axel D. Becke for his guidance and support during the duration of my PhD studies at Dalhousie. I am very grateful to have had the opportunity to work with such an exceptional scientist.

Thank you to Alya A. Arabi for her friendliness and cordiality during our time as coworkers, and to Erin R. Johnson and Hugo J. Bohórquez for many interesting discussions. Also thanks to J. Stuart Grossert for his advice during the later stages of this thesis.

I am grateful to Russell J. Boyd for introducing me to the world of quantum chemistry during my honours project in his group, and for encouraging me to pursue graduate studies in this field. Two of his former students, Jason Pearson and Alexis Taylor, taught me the “nuts and bolts” of computational chemistry calculations, and much is owed to them.

I would like to thank Dalhousie University, the Faculty of Graduate Studies, and the Department of Chemistry for providing facilities, services, and financial support.

All numerical calculations in this thesis were performed on clusters of the Atlantic Computational Excellence network (ACEnet), and I am grateful to the organization and their staff (especially Ross M. Dickson) for providing such a useful resource and excellent technical support.

This thesis would not have been possible without the software developed by the open source community, especially Gnuplot, $\LaTeX$, $\LaTeX$, and the whole Linux operating system.

Finally, I would like to thank my teachers, parents, and friends both near and far for their support and encouragement.
Chapter 1

Introduction

Chemistry yields an understanding of the incredibly complex material world in terms of simple building blocks, the elements of the periodic table. Every element is composed of a single type of microscopic particle known as the atom. Atoms interact with one another in a characteristic way to form molecules, which in turn interact to form macroscopic substances and compounds. Over 70 million substances have been described in the chemical literature and are available commercially [1]. This number is exponentially rising, making chemistry the most prolific of all sciences [2].

This chemical diversity is due to the variety of possible interactions between atoms, which may form strong “chemical bonds” (covalent, dative, ionic or metallic) and weaker “noncovalent” or “van der Waals” interactions. Noncovalent interactions occur both within (intramolecular) and between (intermolecular) molecules. They are responsible for the structure and function of biological macromolecules, molecular self-assembly and the formation of supramolecular structures (clusters, micelles, films, layers, macromolecular assemblies), the condensed phases of matter (solid and liquid substances), and many related phenomena (solvation, absorption, molecular recognition, colloids and mixtures). Understanding noncovalent interactions is essential for the description of macroscopic substances and materials in terms of their microscopic building blocks (atoms and molecules).

Phenomenologically, important noncovalent interaction motifs are hydrogen bonding, halogen bonding, charge transfer, π–π stacking and cation–π interactions. The physical basis for all these interactions is the Coulomb interaction between charge distributions, which can be formulated in terms of multipole moments [3]. This approach yields a rigorous classification of noncovalent intermolecular interactions at long range:

- interactions between permanent multipole moments (electrostatic)
- interactions between multipole moments induced by the electric field of neighbouring molecules (induction)
interactions between instantaneous multipole moments induced by fluctuating charge distributions (dispersion)

Of these, electrostatic interactions are often dominant and can be either attractive or repulsive, depending on the relative orientation of the interacting systems. Induction and dispersion interactions are always attractive. Despite their relative weakness, dispersion interactions are important due to their ubiquity, existing even for otherwise “inert” elements such as noble gases.

The Coulomb interaction in macroscopic systems can be described by classical electrodynamics. However, the behaviour of microscopic charged particles such as electrons and nuclei is governed by quantum mechanics. Indeed, quantum mechanics is the theoretical foundation for essentially all of chemistry, giving rise to the subdisciplines of “quantum chemistry” and “electronic structure theory”.

Density-functional theory (DFT) [4, 5] is an approach that allows an efficient quantum mechanical treatment of atoms, molecules, and periodic systems. It has become the most widely used method in quantum chemistry. Advances in theory, algorithms, and computational resources have made DFT applicable to extended systems such as proteins, DNA, supramolecular aggregates, liquids, crystals, and many more. In principle an exact theory, practical applications of DFT rely on approximations to the unknown “exchange-correlation” functional. The most widely used approximations fall in the category of “semilocal” Kohn-Sham DFT [6]. While accurate for chemical bonds and electrostatic interactions, these approximations are fundamentally unable to describe dispersion interactions [7–9]. This deficiency has sparked intense research over the past two decades, and multiple density-functional methods that treat dispersion are now available [10–14].

This thesis documents the development and benchmarking of one such method by combining standard semilocal approximations for the exchange-correlation functional with the recent “exchange-hole dipole moment” (XDM) model for dispersion. To set the stage, theoretical background on Kohn-Sham DFT, the XDM model, and technical details such as basis sets is outlined in Chapter 2. Chapter 3 concerns the selection of an appropriate exchange-correlation functional to be combined with the XDM model. In Chapter 4, the method is tested on intermolecular complexes using a fully numerical approach, whereas Chapter 5 introduces standard Gaussian basis sets. The basis-set dependence of the method is studied in Chapter 6 for rare-gas systems, leading to an improved parametrization that supersedes those of Chapters 3-5. Systematic testing using standard benchmark sets
of noncovalent interactions is performed in Chapter 7. Chapter 8 details the use of atomic volumes and polarizabilities in the XDM model. Finally, a summary and conclusions are given in Chapter 9.
Chapter 2

Theoretical Background

2.1 Overview

Quantum mechanics enables a description of atoms and molecules from first principles (ab initio), in terms of the fundamental physical laws that govern the behaviour of electrons and nuclei. For all but the smallest systems (one or two electrons), this is a daunting task. The central quantity of quantum mechanics, the wavefunction $\Psi$, is a function of the coordinates\(^1\) of each particle in the system, and contains all information about all possible interactions between these particles. Its information content grows factorially with the number of particles, and so does the effort required to determine the wavefunction.

From the beginning, approximations have been used to make quantum mechanics tractable for many-electron systems. For quantum chemistry, the most important such approximations fall in the category of wavefunction theory (WFT). However, instead of trying to calculate the wavefunction of many-electron systems directly, it is typically constructed from an antisymmetrized product\(^2\) (Slater determinant) of one-electron wavefunctions known as orbitals. The linear combination of atomic orbitals (LCAO) is a common method to construct orbitals for molecular systems, i.e. molecular orbitals (MOs). This approach reflects one of the fundamental aspects of chemistry, that molecules are composed of atoms. Even though the precise definition of an atom in a molecule is ambiguous, any molecule can be transformed into its constituent atoms by the process of atomization.\(^3\) Hence, it is both conceptually and computationally useful to construct a molecular wavefunction from atomic contributions. In practice, the orbitals in the LCAO scheme are further approximated by a linear combination of suitable basis functions, which will be discussed in Section 2.6.

If a single Slater determinant of orbitals is used to represent a many-electron

---

\(^1\)in a quantum mechanical sense, including spatial and spin coordinates
\(^2\)thus obeying the Pauli exclusion principle by construction
\(^3\)typically requiring a substantial amount of energy, the atomization energy
wavefunction, and each one-electron orbital is optimized in the field\textsuperscript{4} of all other electrons, the Hartree-Fock (HF) method is obtained. Since each orbital depends on the other orbitals through the Hartree-Fock potential, an iterative self-consistent field (SCF) scheme is used. Electron-electron interactions are treated in a mean-field approximation, including classical Coulomb repulsion and quantum mechanical electron exchange but neglecting electron correlation. The latter is defined in WFT as the difference between HF and the exact wavefunction. Electron correlation can be treated by expanding the wavefunction in a linear combination of multiple Slater determinants\textsuperscript{5} or via perturbation theory\textsuperscript{6}. While extremely accurate, the high computational cost of these methods severely restricts their applicability to small systems.

### 2.2 Density-Functional Theory

The use of one-electron orbitals and the LCAO approximation to construct a many-electron wavefunction was a major step towards applying quantum mechanics to arbitrary molecular systems. However, systems of increasing size quickly become too complex to be treated by wavefunction theory due to its computational expense and unfavourable scaling.

Since electrons are indistinguishable, an alternative approach might be to use the electron density as the fundamental variable, which depends only on one set of spatial coordinates. At least formally, its complexity does not increase with the number of particles, in contrast to the exact wavefunction (factorial growth) or the one-electron orbitals (linear growth in the number of orbitals). Despite this simplicity, the Hohenberg-Kohn theorem \textsuperscript{15} proves that the electron density uniquely determines its generating potential, and therefore the ground-state wavefunction, of any system with a nondegenerate ground state. In other words, the ground-state wavefunction is a unique functional\textsuperscript{7} of the density. Furthermore, the Hohenberg-Kohn variational principle states that the exact electron density yields the lowest ground-state energy, and can be found by minimizing an energy functional over all possible trial densities. The resulting theory is known as Density-Functional Theory (DFT) \textsuperscript{4,5,16}.

\textsuperscript{4}equivalently, the Hartree-Fock potential
\textsuperscript{5}multiconfiguration SCF (MCSCF), configuration interaction (CI), and coupled cluster (CC) methods
\textsuperscript{6}Møller-Plesset (MP) or many-body perturbation theory (MBPT)
\textsuperscript{7}i.e., those that are positive definite and integrate to the total number of electrons in the system
The earliest DFT methods did indeed model atoms and molecules with the electron density alone, without recourse to wavefunctions or orbitals. Derived from the Fermi-Dirac statistics of the free-electron gas, and including quantum mechanical electron-electron interactions in an approximate way, they are collectively known as the Statistical Theory of Atoms. The simplest model due to Thomas and Fermi is semiclassical in the sense that electrons are “pulverized” and continuously distributed throughout the system, and the electron-electron interaction is approximated by the Coulomb repulsion of a classical charge distribution. The latter results in spurious self-interaction error for one-electron systems. By including quantum mechanical exchange effects, in particular “Dirac” exchange, this self-interaction is partly compensated for. The resulting model is known as the Thomas-Fermi-Dirac (TFD) method. These simple models yield useful information about the electronic structure of atoms, ions, and extended systems. They are not applicable to chemistry as they fail to bind molecular systems. Nowadays, they remain important as conceptual tools due to their simplicity, semiclassical nature, and exactness in the thermodynamic limit \( N_{\text{el}} \to \infty \). Contemporary methods that use only the electron-density as the basic variable are known as orbital-free density-functional theory (OFDFT).

A different approach was followed by Slater, who approximated the orbital-dependent Hartree-Fock potential by a model potential that depends only on the electron density, thus greatly simplifying HF theory. The density-functional part in the resulting Hartree-Fock-Slater or “X\( \alpha \)” method differs from Dirac exchange only by a constant, \( \alpha \).

Shortly after the Hohenberg-Kohn theorem, a second theoretical breakthrough was achieved by Kohn and Sham. Since the Kohn-Sham (KS) method forms the basis of almost all modern applications of DFT including the method developed in this thesis, it will be presented in some detail.

The fundamental equation of quantum mechanics with relevance for quantum chemistry is the time-independent, nonrelativistic Schrödinger equation

\[
\hat{H} \Psi = E \Psi
\]  

(2.1)

In the Born-Oppenheimer approximation and using atomic units, the Hamiltonian \( \hat{H} \) for a system of \( N \) electrons with coordinates \( \mathbf{r}_i \) in an external potential \( \hat{V}_{\text{ext}} \) is

---

8or uniform electron gas (UEG)
9i.e., the local density approximation (LDA) for exchange
given by

\[ \hat{H} = \hat{T} + \hat{V}_{\text{ext}} + \hat{V}_{\text{ee}} = -\frac{1}{2} \sum_i^{N} \nabla_i^2 + \sum_i^{N} \phi_{\text{ext}} + \sum_i^{N} \sum_{j>i}^{N} \frac{1}{|r_i - r_j|} \]  \hspace{1cm} (2.2)

\( \hat{T}, \hat{V}_{\text{ext}}, \text{and} \hat{V}_{\text{ee}} \) are the operators for the kinetic energy, external potential, and electron-electron repulsion, respectively. In a molecular system of \( M \) nuclei at positions \( R_a \) with no applied electric or magnetic fields, the external potential is just the nuclear-electron attraction potential \( \hat{V}_{\text{ne}} \):

\[ \hat{V}_{\text{ne}} = -\sum_i^{N} \sum_a^{M} \frac{Z_a}{|R_a - r_i|} \]  \hspace{1cm} (2.3)

According to the WFT variational principle, the electronic ground-state energy \( E_0 \) can be obtained by minimizing the energy as a functional (expectation value) of \( \Psi \) as follows:

\[ E_0 = \min E [\Psi] = \min \langle \Psi | \hat{T} + \hat{V}_{\text{ext}} + \hat{V}_{\text{ee}} | \Psi \rangle \]  \hspace{1cm} (2.4)

The Hohenberg-Kohn theorem states that \( E_0 \) may also be obtained by minimizing the energy as a functional of the density \( \rho \):

\[ E_0 = \min E [\rho] \]  \hspace{1cm} (2.5)

where

\[ E [\rho] = T [\rho] + V_{\text{ne}} [\rho] + V_{\text{ee}} [\rho] = V_{\text{ne}} [\rho] + F_{\text{HK}} [\rho] \]  \hspace{1cm} (2.6)

\( F_{\text{HK}} [\rho] \) is a system-independent, universal Hohenberg-Kohn functional. The exact form of this functional in terms of the density is unknown and has to be approximated. The Thomas-Fermi model uses simple density functionals (inspired by the uniform electron gas) for the kinetic energy

\[ T_{\text{TF}} [\rho] = C_F \int \rho^{5/3} (r) \, dr \]  \hspace{1cm} (2.7)

and the classical Coulomb repulsion energy of a charge distribution

\[ J [\rho] = \int \int \frac{\rho (r) \rho (r')}{|r - r'|} dr dr' \]  \hspace{1cm} (2.8)
to approximate \( F_{HK}[\rho] \):

\[
F_{HK}[\rho] = T[\rho] + V_{ee}[\rho] \approx T_{TF}[\rho] + J[\rho] \tag{2.9}
\]

These are very severe approximations, especially for the kinetic energy. Another fundamental drawback of Thomas-Fermi theory is its failure to account for wavefunction antisymmetry in its trial densities. The Hohenberg-Kohn theorem holds only for electron densities that arise from an antisymmetric \( N \)-electron wavefunction (\( N \)-representability) and are associated with an external potential \( \hat{V}_{\text{ext}} \) (\( v \)-representability). Neither of these conditions is enforced in Thomas-Fermi theory, and the densities resulting from application of Eq. 2.5 lack atomic shell structure. Thus, the attempt to approximate the full Hohenberg-Kohn functional—a large portion of the total energy functional, Eq. 2.6—in terms of the electron density alone does not yield a viable electronic structure method. A less direct approach is needed, which is the essence of Kohn-Sham density-functional theory (KS-DFT) [6].

Introducing an electron-electron coupling strength parameter \( \lambda \), the Hamiltonian is written as

\[
\hat{H}_\lambda = \hat{T} + \hat{V}_{\text{ext}}(\lambda) + \lambda \hat{V}_{ee} \tag{2.10}
\]

with \( 0 \leq \lambda \leq 1 \). \( \lambda = 1 \) corresponds to the system of interacting electrons, where the external potential is identical to the nuclear-electron attraction potential (Eq. 2.3). Its exact kinetic energy, although in principle a density functional by virtue of the Hohenberg-Kohn theorem, is explicitly known only in terms of the exact wavefunction \( \Psi \)

\[
T[\rho] = \langle \Psi | \hat{T} | \Psi \rangle = \left\langle \Psi \left| -\frac{1}{2} \sum_i N_i^2 \right| \Psi \right\rangle \tag{2.11}
\]

Kohn and Sham introduced a hypothetical, noninteracting system (\( \lambda = 0 \)) whose external potential \( \hat{V}_{\text{ext}}(\lambda) \) is adjusted so that the density is the same as in the interacting system. Its virtue lies in the fact that the noninteracting ground state is exactly a Slater determinant of one-electron orbitals \( \psi_i \), whose kinetic energy is given by

\[
T_0[\rho] = -\frac{1}{2} \sum_i N_i \left\langle \psi_i \left| \hat{\nabla}_i^2 \right| \psi_i \right\rangle \tag{2.12}
\]

This noninteracting kinetic energy is a much better approximation to the exact kinetic energy of Eq. 2.11 than the Thomas-Fermi expression (Eq. 2.7). The \( \psi_i \) are
called “Kohn-Sham orbitals”. In the Kohn-Sham approach, the density is given by
\[
\rho = \sum_i |\psi_i|^2
\] (2.13)

and is by construction both N-representable\textsuperscript{10} and v-representable\textsuperscript{11}.

Using \( T_0 [\rho] \) and \( J [\rho] \) in the energy functional of Eq. 2.5
\[
E [\rho] = T [\rho] + V_{ne} [\rho] + V_{ee} [\rho] = T_0 [\rho] + V_{ne} [\rho] + J [\rho] + E_{xc} [\rho]
\] (2.14)
defines the exchange-correlation functional \( E_{xc} [\rho] \):
\[
E_{xc} [\rho] = (T [\rho] - T_0 [\rho]) + (V_{ee} [\rho] - J [\rho])
\] (2.15)
\( E_{xc} [\rho] \) accounts for the difference between the interacting and noninteracting kinetic energies, \( T [\rho] - T_0 [\rho] \), as well as the quantum mechanical contributions (exchange and correlation) to the electron-electron interaction, \( V_{ee} [\rho] - J [\rho] \).

By minimizing the energy (Eq. 2.14) with respect to variations in the KS orbitals, one obtains the Kohn-Sham equations
\[
h^{KS} \psi_i = \epsilon_i \psi_i
\] (2.16)
with the Kohn-Sham operator
\[
h^{KS} = -\frac{1}{2} \nabla^2 + v_0
\] (2.17)
and the Kohn-Sham potential
\[
v_0 (r) = v_{ne} (r) + v_{Coulomb} (r) + v_{XC} (r)
\] (2.18)
\( v_0 (r) \) is a local, effective (one-body) potential that contains the nuclear-electron potential
\[
v_{ne} (r) = \sum_a \frac{Z_a}{|\mathbf{R}_a - r|}
\] (2.19)
\textsuperscript{10}since it arises from a Slater determinant of Kohn-Sham orbitals
\textsuperscript{11}Eq. 2.10 with \( \lambda = 0 \)
the Coulomb potential
\[ v_{\text{Coulomb}}(r) = \int \frac{\rho'(r')}{|r - r'|} dr' \] (2.20)
and the exchange-correlation potential \( v_{\text{XC}}(r) \). The latter is the functional derivative of \( E_{\text{xc}}[\rho] \):
\[ v_{\text{XC}}(r) = \frac{\delta E_{\text{xc}}}{\delta \rho(r)} \] (2.21)

### 2.3 Exchange-Correlation Functionals

Up to this point, KS-DFT is a formally exact theory. In practice, suitable density-functional approximations (DFAs) to \( E_{\text{xc}}[\rho] \) must be made, whose functional forms may be classified as follows:

1. \( E_{\text{xc}}[\rho] \): local density approximation (LDA)
2. \( E_{\text{xc}}[\rho, \nabla \rho] \): generalized gradient approximation (GGA)
3. \( E_{\text{xc}}[\rho, \nabla \rho, \nabla^2 \rho, \tau] \): meta-GGA\(^{12}\)
4. \( E_{\text{xc}}[\rho(r), \rho(r')] \): nonlocal functionals
5. \( E_{\text{xc}}[\rho, \psi_{\text{occ}}^i] \): hybrid functionals and hyper-GGAs
6. \( E_{\text{xc}}[\rho, \psi_{\text{occ}}^i, \psi_{\text{virt}}^i] \): “5th rung” functionals

Of these, forms 2 and 3 are known as semi-local approximations, since they only depend on the density, its derivatives, and the kinetic energy density at a given point \( r \) (i.e., local) but go beyond the basic LDA by including gradient information. They can be expressed in terms of an exchange-correlation energy density per unit volume \( e_{\text{XC}}[\rho(r)] \) as
\[ E_{\text{XC}}[\rho(r)] = \int e_{\text{XC}}[\rho(r)] dr \] (2.22)

It is customary to separate \( E_{\text{xc}} \) and \( e_{\text{XC}} \) into “exchange” and “correlation” components:
\[ E_{\text{XC}} = E_X + E_C \] (2.23)
\[ e_{\text{XC}} = e_X + e_C \] (2.24)

\(^{12}\)using the kinetic energy density of the Kohn-Sham orbitals, \( \tau(r) = \frac{1}{2} \sum_{\text{occ}} |\nabla \psi_i(r)|^2 \)
Definitions of the exchange and correlation parts will be discussed momentarily. Approximations 1–4 are referred to as "pure" KS-DFT since they involve only local variables, i.e. densities. Items 5 and 6 are orbital-dependent functionals involving occupied ($\psi_{\text{occ}}^i$) and/or virtual ($\psi_{\text{virt}}^i$) Kohn-Sham orbitals. In this thesis, approximations of types 1, 2, and 3 will be used.

2.3.1 Local Density Approximation

In the LDA, $e_{\text{XC}}[\rho(r)]$ is the exchange-correlation energy density of a uniform electron gas of density $\rho(r)$. The exchange part can be analytically derived from theory, with the LDA exchange energy density per unit volume $e_{\text{X}}^{\text{LDA}}$ given by [17]:

$$e_{\text{X}}^{\text{LDA}}[\rho(r)] = -C_X\rho^{4/3}$$  \hspace{1cm} (2.25)

with

$$C_X = \frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3}$$  \hspace{1cm} (2.26)

For spin-polarized (e.g., open-shell) systems the LDA can be generalized to the local spin density approximation (LSDA)

$$e_{\text{X}\sigma}^{\text{LSDA}}[\rho_{\sigma}(r)] = -2^{1/3}C_X \sum_{\sigma} \rho_{\sigma}^{4/3}$$  \hspace{1cm} (2.27)

where the spin label $\sigma = \{\alpha, \beta\}$ denotes either up or down spin.

No analytical derivation is available for the correlation energy of a uniform electron gas, but accurate parametrizations $e_{\text{C}}^{\text{LSDA}}$ of quantum Monte Carlo simulations have been published by Vosko, Wilk, and Nusair (VWN) [19] as well as by Perdew and Wang (PW) [20]. Using the LDA for molecular systems makes the assumption that $e_{\text{XC}}[\rho(r)]$ at each point $r$ is identical to the exchange-correlation energy density of a uniform electron gas of density $\rho(r)$. It does not assume a constant electron density for the molecule, but it does (implicitly) assume a slowly varying density [21,22]. This explains the relative success of the LDA in describing the electronic structure of nonuniform systems such as atoms or molecules.

2.3.2 Generalized Gradient Approximation

An obvious way to improve upon the LDA for nonuniform systems is to make $E_{\text{xc}}$ dependent also on the gradient of the density, leading to the “GGA” class of
functionals. GGA exchange functionals can be expressed in the form

\[ E_{X}^{\text{GGA}} = \int e_{X}^{\text{LDA}} f_{X}(\chi) \]  \hspace{1cm} (2.28)

or

\[ E_{X}^{\text{GGA}} = \sum_{\sigma} \int e_{X\sigma}^{\text{LSDA}} f_{X}(\chi_{\sigma}) \]  \hspace{1cm} (2.29)

for the spin-polarized case. The “reduced” density gradients \( \chi \) and \( \chi_{\sigma} \) are defined as

\[ \chi = \frac{|\nabla \rho|}{\rho^{4/3}} = \frac{|\nabla 2\rho_{\sigma}|}{(2\rho_{\sigma})^{4/3}} = 2^{-1/3} \frac{|\nabla \rho_{\sigma}|}{\rho_{\sigma}^{4/3}} = 2^{-1/3} \chi_{\sigma} \]  \hspace{1cm} (2.30)

The function \( f_{X}(\chi) \) or \( f_{X}(\chi_{\sigma}) \) is called the “exchange enhancement factor” and characterizes the modification of LDA exchange (\( f_{X}^{\text{LDA}} = 1 \)) by gradient dependence. The enhancement factors for the GGAs B86 [23], B86b [24] and B88 [25] are given by

\[ f_{X}^{\text{B86}}(\chi_{\sigma}) = 1 + \left(2^{1/3} C_{X}\right)^{-1} \frac{0.0036 \chi_{\sigma}^{2}}{1 + 0.004 \chi_{\sigma}^{2}} \]  \hspace{1cm} (2.31)

\[ f_{X}^{\text{B86b}}(\chi_{\sigma}) = 1 + \left(2^{1/3} C_{X}\right)^{-1} \frac{0.0036 \chi_{\sigma}^{2}}{(1 + 0.006 \chi_{\sigma}^{2})^{4/5}} \]  \hspace{1cm} (2.32)

\[ f_{X}^{\text{B88}}(\chi_{\sigma}) = 1 + \left(2^{1/3} C_{X}\right)^{-1} \frac{0.0042 \chi_{\sigma}^{2}}{1 + 6(0.0042) \chi_{\sigma} \sinh^{-1} \chi_{\sigma}} \]  \hspace{1cm} (2.33)

B86, B86b and B88 were derived by analyzing exchange energies and exchange holes in atomic systems. By contrast, the exchange GGAs PW86 [26], PW91 [27,28], and PBE [29] are based on an analysis of the exchange hole in the slowly varying electron gas. Additionally, PW91 and PBE satisfy various exact constraints. These functionals use an equivalent definition of the dimensionless reduced density gradient in terms of the Fermi wavevector \( k_{F} = \left(3\pi^{2}\rho\right)^{1/3} \):

\[ s = \frac{|\nabla \rho|}{2\rho k_{F}} = \frac{|\nabla \rho|}{2 \left(3\pi^{2}\right)^{1/3} \rho^{4/3}} = \left(24\pi^{2}\right)^{-1/3} \chi = \left(48\pi^{2}\right)^{-1/3} \chi_{\sigma} \]  \hspace{1cm} (2.34)

\[ f_{X}^{\text{PW86}}(s) = \left(1 + 1.296 s^{2} + 14 s^{4} + 0.2 s^{6}\right)^{1/15} \]  \hspace{1cm} (2.35)
Finally, revPBE [30] is a reparametrization of PBE to give better total atomic energies and atomization energies:

$$f^{\text{revPBE}}_X(s) = 1 + 1.245 - \frac{1.245}{1 + 0.21951/1.245s^2}$$ (2.38)

Plots of the enhancement factor as a function of the dimensionless reduced density gradient are shown in Figure 2.1. The reduced density gradient $\chi$ has small values in the bonding regions of molecules (characterized by small gradient), and large values in the asymptotic tails far from nuclei, due to the exponential decay of the electron density [5]. All standard exchange GGAs have similar behaviour.
for small $\chi < 10$, but behave very differently at large $\chi$. The description of van der Waals interactions is very sensitive to the large-$\chi$ behaviour of the exchange functional (see Chapter 3). The correct asymptotic limit $F_X(\chi \to \infty) = \infty$ is satisfied by the B88 functional [25]. However, this does not guarantee good performance for van der Waals interactions due to lack of systematic error cancellation between complexes and their separated monomers.

As is the case with the LDA, generalized gradient approximations for the correlation functional tend to be much more complicated than their exchange counterparts. In general, correlation GGAs describe “dynamical”\textsuperscript{13} electron correlation by enforcing interelectronic cusp and sum rule conditions on the correlation hole. One of the most popular correlation functionals was derived by Lee, Yang, and Parr [31] (LYP) as a density-functional approximation to the opposite-spin correlation energy in the helium atom. Other correlation functionals due to Perdew and coworkers (P86 [32], PW91 [27,28], and PBE [29]) are based on the gradient expansion of the correlation hole in the slowly varying electron gas. P86 contains empirical parameters fitted to the correlation energy of the neon atom. On the other hand, the parameters in PW91 and PBE are not fitted to reference data but obtained by enforcing particular choices of exact constraints\textsuperscript{14}, leading to the classification of these functionals as “nonempirical”.

2.4 Dispersion In DFT

None of the functionals discussed so far includes a description of the dispersion interaction. In fact, semilocal (LDA or GGA) functionals are fundamentally unable to treat long-range dispersion interactions [7].

When two neutral atoms are separated by a distance $R$, large enough that their charge distributions\textsuperscript{15} don’t overlap and no chemical bond is formed, they experience an attractive force known as the London dispersion force. This force is responsible for the condensation of noble gas atoms into liquid and solid phases. The extremely low boiling points (4.2 K for helium-4) of liquid noble gases indicate the weakness of the dispersion interaction. At long range, the dispersion energy can

\textsuperscript{13}i.e., short-range
\textsuperscript{14}such as uniform and nonuniform scaling conditions, recovering the linear response of a uniform electron gas, the Lieb-Oxford bound, etc.
\textsuperscript{15}or wavefunctions
be written as an inverse power series in $R$:

$$E_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} \cdots$$

(2.39)

with the $C_i$ being system-dependent dispersion coefficients. Dispersion is a universal force that exists for any type of matter. In polar materials, it is often overshadowed by the much stronger electrostatic and induction forces. Nonetheless, any method that aims to predict the properties of matter from first principles must include dispersion.

Dispersion is a purely quantum mechanical phenomenon, arising from instantaneous (virtual) fluctuations of charge distributions. In wavefunction theory, it can be effectively described with perturbation theory or coupled-cluster methods, albeit at high computational expense. Within DFT, a rigorous description of dispersion starts from the adiabatic-connection fluctuation-dissipation theorem (ACFDT), which yields an exact formula for the correlation energy in terms of “response” functions and an interaction operator involving Coulomb, exchange and correlation effects. For dispersion, exchange and correlation effects are usually neglected\(^\text{16}\), and if the ACFDT expansion is further truncated at second order, one arrives at the Casimir-Polder formula for the long-range correlation (i.e., dispersion) energy between two well-separated monomers $A$ and $B$ [33–35]

$$E_{\text{disp}}^{(2)} = -\frac{1}{2\pi} \int_0^\infty \, d\omega \int \int \int \int \, dr_A \, dr_A' \, dr_B \, dr_B' \chi_A (r_A, r_A', \omega) \times \chi_B (r_B, r_B', \omega) \frac{1}{|r_A - r_B||r_A' - r_B'|}$$

(2.40)

The imaginary frequency-dependent linear response functions $\chi (r, r', \omega)$ are related to the underlying charge density fluctuations in the isolated monomers through the fluctuation-dissipation theorem. A local density-functional approximation for the response of a uniform electron gas [36–38] and more recent local polarizability approximations [39,40] form the basis for various tractable methods of treating dispersion interactions. These are known as the Andersson-Langreth-Lundqvist (ALL) functional [37], local response dispersion (LRD) method [41, 42], van der Waals density functional (vdW-DF) [43–45], and its variants vdW-DF2 [46], VV09 [47] and VV10 [48]. For recent reviews see Refs [14, 40, 49].

\(^{16}\)due to their short-ranged nature, yielding the random-phase approximation (RPA)
2.4.1 The Exchange-Hole Dipole Moment Model

A completely different approach to model dispersion was followed by Becke and Johnson, starting from the question *What is the source of the instantaneous charge fluctuations leading to the dispersion interaction?* [50] This led to the exchange-hole dipole moment (XDM) model of dispersion [50–53], an essential part of the density-functional methodology developed in this thesis. In the following, a review of the conceptual basis and important formulas of the XDM model is provided.

Consider an isolated hydrogen atom in its ground state. The atom is spherically symmetric, with no *permanent* multipole moments. However, as the electron moves through the atom, a position-dependent dipole moment is created at every instant. If the nucleus is at the origin and the electron at distance $r$ from the nucleus, this pair of point charges generates a dipole moment\(^{17}\) of magnitude $|M_1| = r$ [52]. The electric field due to this instantaneous dipole moment can induce another dipole moment in a neighbouring, polarizable atom. The resulting attractive interaction is known as dispersion.

In a many-electron atom, the nucleus is shielded by the electronic charge distribution. Now consider an electron at position $\mathbf{r}_1$ in the atom (a “reference” electron). Due to the quantum mechanical effects of exchange and correlation between electrons, the shielding is not isotropic but depends on the spin $\sigma$ and instantaneous position $\mathbf{r}_1$ of the reference electron. In particular, no other electron of spin $\sigma$ can be found at $\mathbf{r}_1$ due to the Pauli exclusion principle. The depletion in probability of finding another electron of spin $\sigma$ near the reference electron is known as the *Fermi hole*, or *exchange hole* $h_{X\sigma}(\mathbf{r}_1, \mathbf{r}_2)$ in the language of density-functional theory. Furthermore, Coulomb repulsion between electrons yields a further depletion in probability of finding a second electron near $\mathbf{r}_1$, known as the *correlation hole* $h_{C}(\mathbf{r}_1, \mathbf{r}_2)$. A general overview of exchange and correlation holes in DFT is given in Section 2.5.

For the current purpose, it suffices to say that the exchange hole represents a positive charge distribution whose center of charge\(^{18}\) is in general displaced from the reference electron. This position-dependent charge separation can be viewed as an instantaneous charge fluctuation depending on $\mathbf{r}_1$. It gives rise to instantaneous multipole moments depending on both the position of the reference electron and

---

\(^{17}\) as well as higher-order multipole moments given by $|M_\ell| = r^\ell$

\(^{18}\) or mean position
the shape and mean position of the exchange hole\textsuperscript{19}. The dipole moment $d_{X\sigma}(r_1)$ of an electron at $r_1$ and its associated exchange hole $h_{X\sigma}(r_1, r_2)$ was termed the exchange-hole dipole moment (XDM) by Becke and Johnson \cite{50}, and is found by integration over $r_2$ as

$$d_{X\sigma}(r_1) = \left[ \int r_2 h_{X\sigma}(r_1, r_2) \, dr_2 \right] - r_1$$ \hfill (2.41)

In spherical atomic systems, the center of charge $h^+$ of the exchange hole lies on the line connecting the nucleus and the reference electron.\textsuperscript{20} Its distance from the nucleus is given by $r - d_{X\sigma}$, where $d_{X\sigma}$ is the magnitude of the exchange-hole dipole moment. Thus, approximate higher-order multipole moments with respect to the nucleus are generated as

$$|M_{\ell\sigma}| = \left[ r^\ell - (r - d_{X\sigma})^\ell \right]$$ \hfill (2.42)

where the electron and its exchange hole are modeled by two point charges $e^-$ and $h^+$, at distances $r$ and $(r - d_{X\sigma})$ from the nucleus.

Over a series of papers \cite{50–57}, Becke and Johnson were able to transform the concept of the XDM model into a practical density-functional approximation for the dispersion energy of arbitrary, polyatomic systems. An overview of the derivations is given in the following sections.

### 2.4.2 Perturbation Theory Of The Dispersion Interaction

Using second-order perturbation theory and an expansion in terms of multipole moments $\ell$, the dispersion energy between two spherical, noninteracting ground-state atoms $A$ and $B$ is given by \cite{51, 58}

$$E_{disp} = - \sum_{\ell_A=1}^\infty \sum_{\ell_B=1}^\infty \frac{\varepsilon_2(\ell_A, \ell_B)}{R^{2(\ell_A+\ell_B+1)}}$$ \hfill (2.43)

$$\varepsilon_2(\ell_A, \ell_B) = \lambda(\ell_A, \ell_B) \sum_a \sum_b \frac{\left| \langle \Psi_A^a | M_{\ell_A} | \Psi_A^0 \rangle \right|^2 \left| \langle \Psi_B^b | M_{\ell_B} | \Psi_B^0 \rangle \right|^2}{E_A^a + E_B^b - E_A^0 - E_B^0}$$ \hfill (2.44)

\textsuperscript{19}the latter of which is determined by the former
\textsuperscript{20}In atoms containing only s-electrons $h^+$ is independent of $r_1$ and centered on the nucleus.
where the summation runs over excited states denoted by $a$ and $b$, and 
$
\lambda (\ell_A, \ell_B) = \frac{(2\ell_A + 2\ell_B)!}{(2\ell_A + 1)(2\ell_B + 1)(2\ell_A)!}\frac{(2\ell_B + 1)!}{(2\ell_B)!}
\]

is a numerical prefactor. $M_\ell$ is the $m = 0$ component of the multipole moment operator

$$
\hat{M}_\ell = \sqrt{4\pi} \sum_i r_i^\ell Y^0_\ell (\Omega_i)
$$

where the summation runs over individual particles $i$, $r$ is the distance to the origin, and $\Omega_i$ is the solid angle. In the “closure” or “Unsöld” approximation [59], the energy differences in the denominator of Eq. 2.44 are replaced by average excitation energies $\Delta E$

$$
E^a_A + E^b_B - E^0_A - E^0_B \approx \Delta E_A + \Delta E_B
\] (2.46)

Since the ground states of both atoms have no permanent multipole moments

$$
\langle \Psi^0_A | \hat{M}_\ell^A | \Psi^0_A \rangle = \langle \Psi^0_B | \hat{M}_\ell^B | \Psi^0_B \rangle = 0
$$

the summation over excited states in Eq. 2.44 can be extended to include the ground states. Using the resolution of the identity $I = \sum_{\text{all states}} |\Psi^a \rangle \langle \Psi^a|$ yields

$$
\sum_a \left| \langle \Psi^a_A | \hat{M}_\ell^A | \Psi^0_A \rangle \right|^2 = \sum_a \langle \Psi^0_A | \hat{M}_\ell^A | \Psi^a_A \rangle \langle \Psi^a_A | \hat{M}_\ell^A | \Psi^0_A \rangle = \langle \Psi^0_A | \hat{M}_\ell^2 A | \Psi^0_A \rangle
$$

(2.48)

Thus, Eq. 2.44 is simplified from a summation of excited-state expectation values of $\hat{M}_\ell$ to expectation values of the squared multipole moment operator $\hat{M}_\ell^2$ over the ground-state wavefunctions of atoms $A$ and $B$.

$$
\epsilon_2 (\ell_A, \ell_B) \approx \lambda (\ell_A, \ell_B) \frac{\langle \Psi^0_A | \hat{M}_\ell^2 A | \Psi^0_A \rangle \langle \Psi^0_B | \hat{M}_\ell^2 B | \Psi^0_B \rangle}{\Delta E_A + \Delta E_B}
$$

(2.49)

If the expansion in Eq. 2.43 is truncated at $\ell_A + \ell_B = 4$ one arrives at the dispersion series up to the $R^{-10}$ term

$$
E_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}}
$$

(2.50)

with

$$
C_6 = \frac{2}{3} \frac{\langle \hat{M}_1^2 \rangle_A \langle \hat{M}_1^2 \rangle_B}{\Delta E_A + \Delta E_B}
$$

(2.51)
The average excitation energies \( \Delta E \) in Eqs. 2.51–2.53 can be written in terms of polarizability \( \alpha \) as

\[
\Delta E = \frac{2}{3} \frac{\langle \hat{M}_1^2 \rangle}{\alpha}
\]  (2.54)

by invoking the same “closure” approximation in the second-order perturbation theory of polarizability [59].

Becke and Johnson now approximate the multi-electron expectation values \( \langle \hat{M}_1^2 \rangle \) by the simple one-electron integral [51]

\[
\langle \hat{M}_1^2 \rangle \approx \sum_{\sigma} \int \rho_\sigma (r) M_{1,\ell,\sigma}^2 dr
\]  (2.55)

where \( M_{1,\ell,\sigma}^2 \) is the square of the \( \ell \)-th multipole moment of the electron and its exchange hole, Eq. 2.42. This semiclassical approximation enables an interatomic decomposition and transformation to a density-functional model, since the dispersion coefficients of Eqs. 2.51–2.53 depend only on the electron density \( \rho_\sigma (r) \) and the magnitude of the exchange-hole dipole moment \( d_{X,\sigma} \).

### 2.4.3 Interatomic Decomposition

The dispersion coefficients of isolated atoms \( A \) and \( B \), Eqs. 2.51–2.53, are now generalized to atoms within a molecule \( i \) and \( j \) [51, 55]:

\[
C_{6,ij} = \frac{\alpha_i \alpha_j \langle \hat{M}_1^2 \rangle_i \langle \hat{M}_1^2 \rangle_j}{\langle \hat{M}_1^2 \rangle_i \alpha_j + \langle \hat{M}_1^2 \rangle_j \alpha_i}
\]  (2.56)

\[
C_{8,ij} = \frac{3 \alpha_i \alpha_j \left( \langle \hat{M}_1^2 \rangle_i \langle \hat{M}_2^2 \rangle_j + \langle \hat{M}_2^2 \rangle_i \langle \hat{M}_1^2 \rangle_j \right)}{2 \langle \hat{M}_1^2 \rangle_i \alpha_j + \langle \hat{M}_1^2 \rangle_j \alpha_i}
\]  (2.57)
Becke and Johnson use the Hirshfeld [60] partitioning scheme to decompose Eq. 2.59 into atomic contributions

$$
\langle \hat{M}_l^2 \rangle_i \approx \sum_{\sigma} \int w_i (r) \rho_{\sigma} (r) \left[ r^\ell - (r - d_{X\sigma})^\ell \right]^2 dr
$$

(2.59)

where \( w_i (r) \) is the Hirshfeld atomic partitioning weight

$$
w_i (r) = \frac{\rho_{i,\text{free}} (r)}{\sum_n \rho_{n,\text{free}} (r)}
$$

(2.60)

with the sphericalized free atomic densities \( \rho_{i,\text{free}} (r) \).

The effective atom-in-molecule polarizability \( \alpha_i \) is approximated by

$$
\alpha_i = \frac{V_{i,\text{eff}}}{V_{i,\text{free}}} \alpha_{i,\text{free}}
$$

(2.61)

in terms of free atomic polarizability \( \alpha_{i,\text{free}} \) [61], effective atom-in-molecule volume \( V_{i,\text{eff}} \) [51, 56]

$$
V_{i,\text{eff}} = \langle r^3 \rangle_{i,\text{eff}} = \int r^3 w_i (r) \rho (r) dr
$$

(2.62)

and free atomic volume \( V_{i,\text{free}} \)

$$
V_{i,\text{free}} = \langle r^3 \rangle_{i,\text{free}} = \int r^3 \rho_{i,\text{free}} (r) dr
$$

(2.63)

The relation between polarizability and volume is explored in Chapter 8. The equations in this section constitute the XDM model as employed in this thesis.

### 2.5 Exchange-Correlation Holes In DFT

With the reference electron at \( r_1 \), the exchange-correlation hole \( h_{XC} (r_1, r_2) \) measures the depletion in probability of finding another electron at position \( r_2 \). It is a nonlocal (two-electron) quantity of great use in DFT, since \( E_{XC} \) can be written
exactly in terms of $h_{XC}(r_1, r_2)$ [4, 5]:

$$E_{XC} = \frac{1}{2} \int \frac{\rho(r_1) h_{XC}(r_1, r_2)}{|r_1 - r_2|} dr_1 dr_2$$  \hspace{1cm} (2.64)

Furthermore, useful approximations to $E_{XC}$ are obtained by modelling $h_{XC}(r_1, r_2)$ and exploiting its well-known properties. One universal property of the hole is its normalization, which is equal to exactly minus one electron no matter where the reference point:

$$\int h_{XC}(r_1, r_2) \, dr_2 = -1$$  \hspace{1cm} (2.65)

Introducing spin-dependence yields a decomposition of the exchange-correlation hole into an exchange hole (Fermi hole, between same-spin electrons only) and a correlation hole (same and opposite spins) [5]. The exchange hole has a value of $-\rho_\sigma(r_1)$ for $r_1 = r_2$, completely extinguishing the probability of finding two same-spin electrons at $r_1$ as a consequence of the Pauli exclusion principle. It has the same normalization as the full XC hole, Eq. 2.65. Correlation holes, on the other hand, integrate to zero and are a consequence of the Coulomb repulsion between electrons.

In the noninteracting limit ($\lambda = 0$) of KS-DFT, the exchange hole is given in terms of the Kohn-Sham orbitals $\psi_{i\sigma}$ as

$$h_{X\sigma}^{(r)}(r_1, r_2) = -\frac{\sum_{ij} \psi_{i\sigma}(r_1) \psi_{j\sigma}(r_1) \psi_{i\sigma}(r_2) \psi_{j\sigma}(r_2) \hat{r}_2 \psi_{i\sigma}(r_2) \psi_{j\sigma}(r_2) \, dr_2}{\rho_\sigma(r_1)}$$  \hspace{1cm} (2.66)

The dipole moment of this exact exchange hole and its associated electron can be obtained by integration over $r_2$ (Eq. 2.67) to yield [50]

$$d_{X\sigma}(r_1) = \left[ \frac{1}{\rho_\sigma(r_1)} \sum_{ij} \psi_{i\sigma}(r_1) \psi_{j\sigma}(r_1) \times \int r_2 \psi_{i\sigma}(r_2) \psi_{j\sigma}(r_2) \, dr_2 \right] - r_1$$  \hspace{1cm} (2.67)

Using the exchange-hole dipole moment of Eq. 2.67 to compute the dispersion coefficients of Eqs. 2.56–2.58 constitutes the “exact exchange” (XX) version of the XDM model. It is used in Chapters 3 and 4.

2.5.1 The Becke-Roussel Exchange Hole Model

An appealing alternative is to use a density-functional exchange hole model instead of the exact expression. In the Becke-Roussel (BR) model [62], the exchange
hole is represented by an exponential function $-Ae^{-ar}$, centered at distance $b$ from the reference electron. Hence, the magnitude of the exchange-hole dipole moment is simply $d_{X\sigma} = b$ [54]. The BR hole perfectly reproduces the exact exchange hole in hydrogenic atoms, and the $r \to \infty$ asymptotic limit of the exchange potential in any finite system. The parameters $A$, $a$, and $b$ are determined by enforcing three conditions:

1. Normalization of the exchange hole to integrate to $-1$ electrons at all $r_1$

2. A value of $-\rho_\sigma (r_1)$ for $r_1 = r_2$

3. Reproduces the correct curvature $Q_\sigma$ of the exchange hole around the reference point $r_1$

Constraints 2 and 3 are enforced by equating the spherical average (around $r_1$) of the BR exchange hole to the exact exchange hole, whose Taylor expansion on a shell of radius $s$ around $r_1$ is

$$h_{X\sigma}^{\text{exact}} (r_1, s) = -\rho_\sigma (r_1) - Q_\sigma s^2 + \ldots$$  \hspace{1cm} (2.68)

$$Q_\sigma = \frac{1}{6} \left( \nabla^2 \rho_\sigma - 2D_\sigma \right)$$  \hspace{1cm} (2.69)

$$D_\sigma = \tau_\sigma - \frac{1}{4} \frac{\left( \nabla \rho_\sigma \right)^2}{\rho_\sigma}$$  \hspace{1cm} (2.70)

This yields a nonlinear equation whose solution determines the values of the parameters $A$, $a$ and $b$ for every point $r_1$ in the atom or molecule. Since the BR exchange hole depends on both the Laplacian $\nabla^2 \rho_\sigma$ and the kinetic energy density $\tau_\sigma = \sum_i^{\text{occ}} |\nabla \psi_i \sigma|^2$, it is a second-order or meta-GGA functional. Although it depends on the Kohn-Sham orbitals $\psi_i \sigma$ via the kinetic energy density, $\tau_\sigma$ is a local quantity, and hence BR is a semi-local DFA. This has both computational and practical advantages over the exact exchange hole of Eq. 2.66, as detailed in Chapter 4. If used to compute the exchange-hole dipole moment ($d_{X\sigma} = b$) and the dispersion coefficients of Eqs. 2.56–2.58, this constitutes the “Becke-Roussel” (BR) version of the XDM model [54].
2.6 Basis Sets

2.6.1 Overview

The solution of the Kohn-Sham equations (Eq. 2.16) yields the Kohn-Sham orbitals \( \psi_i \) and the associated electron density, Eq. 2.13. Various approaches exist on how to represent \( \psi_i \) and \( \rho (\mathbf{r}) \). The fully numerical approach of Becke and Dickson represents orbitals and densities as discrete quantities on atom-centered grids [63], and solves the Kohn-Sham and Poisson’s equation for \( v_{\text{Coulomb}} \) as implemented in the \texttt{Numol} program [64–68]. This method allows the self-consistent determination of Hartree–Fock and LDA orbitals, which can then be used to evaluate exchange-correlation energies with GGA or even more sophisticated functionals. \texttt{Numol} is used in Chapters 3 and 4 of this thesis.

The more common approach, used in most electronic structure software packages, is to expand the Kohn-Sham orbitals in a set of suitable basis functions, the \textit{basis set}. For computational reasons, the basis functions are most commonly chosen to be Gaussian functions\footnote{Gaussian-type orbitals (GTOs)}. Other options are exponential\footnote{Slater-type orbitals (STOs)} or numerical basis functions\footnote{numerical atom-centered orbitals (NAOs)}. In this thesis, basis sets of Gaussian functions are explored in Chapters 5, 6, 7 and 8.

A Gaussian-type orbital is written in Cartesian coordinates as

\[
\phi_{\zeta, i, j, k}(x, y, z) = Nx^i y^j z^k e^{-\zeta r^2} \tag{2.71}
\]

where \( N \) is a normalization constant, \( \zeta \) is the basis function exponent, and \( r = \sqrt{x^2 + y^2 + z^2} \) is the distance to the origin of the coordinate system, typically taken to be the position of a nucleus. The angular dependence is specified through the set of integers \( i, j, k \): \( i = j = k = 0 \) defines a spherical s-type GTO, \( l = i + j + k = 1 \) yields a p-type GTO\footnote{\( p_x \) for \( i = 1 \), \( p_y \) for \( j = 1 \), \( p_z \) for \( k = 1 \)}, and so on for higher angular momenta \( l = i + j + k \). Alternatively, GTOs can be written in spherical polar coordinates as

\[
\phi_{\zeta, n, l, m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{n-1}e^{-\zeta r^2} \tag{2.72}
\]

where \( n \) is the principal quantum number and \( Y_{l,m}(\theta, \varphi) \) is a spherical harmonic function that describes the angular dependence via the integers \( l \) and \( m \), with \( m \)}
taking values of \(-l\) to \(+l\).

The exact orbitals of an atom have a cusp at the nucleus \((r \to 0)\) and feature exponential decay at \(r \to \infty\). GTOs fail to meet both conditions, having a slope of zero at the nucleus and falling off too rapidly. Furthermore, the primitive GTOs of Eqns. 2.71 and 2.72 are nodeless and unable to describe the nodal structure of real atomic orbitals. However, it is possible to better approximate the behaviour of atomic orbitals by taking a linear combination of \(N\) GTOs:

\[
\phi_{\{\zeta,c\},i,j,k}(x, y, z) = x^i y^j z^k \sum_{n=1}^{N} c_n N_n e^{-\zeta_n r^2}
\]  

(2.73)

This is called a \textit{contracted} Gaussian-type orbital (cGTO), where the individual functions \(e^{-\zeta_n r^2}\) are known as Gaussian \textit{primitives}. Typically between 3 and 6 primitives are contracted to form a cGTO, with larger values of \(N\) possible for extended basis sets. The contraction coefficients \(c_n\) and primitive exponents \(\zeta_n\) are chosen by basis set developers to yield the optimal description of an atomic orbital for a given \textit{degree of contraction} \(N\). Usually, this is done by minimizing the Hartree–Fock energy of the isolated atom with respect to \(c_n\) and \(\zeta_n\).

### 2.6.2 Examples

As examples, the def2-SVP basis sets for hydrogen and oxygen are shown in Figures 2.2 and 2.3. The def2-SVP basis set for hydrogen contains three basis functions (cGTOs), two s-type functions and one p-type function. The first s-type function is a linear combination of three primitives, whose exponents and contraction coefficients are specified in the first and second column, respectively. The primitive with exponent 0.4445 has the largest contribution to the cGTO. The other two primitives with larger exponents (“tight” functions) improve the basis set in the region of the nucleus. The def2-SVP basis set is of so called \textit{split-valence} (SV) type, meaning that each atomic valence orbital is described by two basis functions. This allows the orbital to adapt to different molecular environments or external fields, by varying the relative contribution of each cGTO to the orbital. Hence the occurrence of a second (uncontracted) s-type function that consists of a single Gaussian primitive with exponent 0.122. The basis set also contains a set\(^{25}\) of p-type functions with exponent 0.8. Such higher angular momentum functions are unnecessary for the description of the isolated atom, but allow the orbital to deviate from spherical

---

\(^{25}\)consisting of \(p_x\), \(p_y\), and \(p_z\) functions with the same exponents and contraction coefficients
The def2-SVP basis set for hydrogen can be specified as \((4s1p)/[2s1p]\), where the text in parentheses refers to the numbers and types of primitive GTOs, and the bracketed string refers to the number and type of contracted GTOs.

The def2-SVP basis set for oxygen is constructed in a similar manner. It contains three s-type cGTOs, the first of which is composed of five primitives and describes the \(1s\) core orbital of oxygen. The \(2s\) valence orbital is described by the two remaining s-type functions. Similarly, the \(2p\) valence orbitals are described by two p-type functions, the first of which is composed of three primitives. Finally, a set of d-type polarization functions with exponent 1.2 is present. The abbreviated notation of def2-SVP for oxygen\(^\text{26}\) is \((7s4p1d)/[3s2p1d]\). Its contraction pattern is given by \(511/31/1\), referring to the number of primitives contracted to form each s/p/d-type basis function.

Using two cGTOs to represent a valence orbital is also termed double zeta (DZ), from the Gaussian exponent \(\zeta\). Basis sets with three, four, and five cGTOs per valence orbital are known as triple zeta (TZ), quadruple zeta (QZ), and quintuple zeta (5Z). The cardinality \(n\) of the basis set refers to the number of valence basis functions, with \(n = 2, 3, 4\) corresponding to DZ, TZ, and QZ, respectively. The number and maximum angular momentum of polarization functions are typically increased in concert with the valence basis set. Thus, for first-row atoms a DZ basis set typically contains one set of d-type polarization functions (abbreviated as d), whereas a TZ basis set contains two sets of d-type and one set of f-type polarization functions (abbreviated as 2df). This polarization pattern ensures the

\(^{26}\text{and other first-row atoms}\)
Figure 2.3: def2-SVP basis set for oxygen

proportion proper balance between valence and polarization functions [22]. For QZ and 5Z basis sets, it leads to 3d2f1g and 4d3f2g1h polarization functions. Functions of g or h angular momentum are considered irrelevant for Hartree-Fock or DFT calculations of ground-state properties, and may be omitted without affecting basis set balance [22].

Finally, basis sets can be augmented with diffuse functions of low exponent $\xi$\textsuperscript{27} to improve the description of orbitals far from the nucleus. This is especially important for anions, excited states, and the calculation of molecular properties such as polarizabilities. Diffuse functions can also be beneficial for the calculation of van der Waals interactions, which occur primarily at large interatomic separations. The presence of diffuse functions on both hydrogen and nonhydrogen atoms is indicated in the names of standard basis sets by the prefix “aug-”, the suffix “D”, or the string “++”. It is also common to omit diffuse functions on hydrogens and include them on heavy atoms only, leading to the “aug-”, “heavy-aug-”, “min-aug-”, and “+”-type basis sets.

\begin{verbatim}
O  0
S  5  1.00
  2266.1767785 -0.53431809926E-02
  340.87010191  -0.39890039230E-01
  77.363135167  -0.17853911985
  21.479644940  -0.46427684959
  6.6589433124  -0.44309745172
S  1  1.00
  0.80975975668  1.0000000
S  1  1.00
  0.25530772234  1.0000000
P  3  1.00
  17.721504317  0.43394573193E-01
  3.8635505440  0.23094120765
  1.0480920883  0.51375311064
P  1  1.00
  0.27641544411  1.0000000
D  1  1.00
  1.2000000  1.0000000
\end{verbatim}

\textsuperscript{27}usually between 0.01 and 0.1
Table 2.1: Double zeta basis sets employed in this work.

All basis sets used in this thesis are shown in Tables 2.1–2.3 with their standard names, cardinality, diffuse functions, and contracted basis functions listed for hydrogen and first-row elements. Basis functions that were omitted from the calculations are indicated by curly brackets. In these tables, basis sets are grouped according to the presence of diffuse functions:

1. no diffuse functions ($nZ$, $n = D,T,Q,5,7$)
2. diffuse functions on non-hydrogen atoms ("heavy augmented", aug'-nZ)
3. diffuse functions on all atoms ("fully augmented", aug-nZ)

In many cases, the presence of diffuse functions formally increases the cardinality $n$ of the basis set by one. Since diffuse functions serve a rather specific purpose, an abbreviation consistent with the underlying basis set (e.g., "aug-DZ") is used in the “type” column. Still, it may be noted that a fully augmented DZ basis set such as aug-cc-pVDZ contains the same number of basis functions as a TZ basis set.
<table>
<thead>
<tr>
<th>basis set</th>
<th>type</th>
<th>n-zeta</th>
<th>diffuse functions</th>
<th>contracted basis functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-311G(d,p)</td>
<td>TZ</td>
<td>3/3</td>
<td>3s1p</td>
<td>4s3p1d</td>
</tr>
<tr>
<td>6-311G(2df,p)</td>
<td>3/3</td>
<td>3s1p</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>6-311G(2df,2pd)</td>
<td>3/3</td>
<td>3s2p1d</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>6-311G(3df,3pd)</td>
<td>3/3</td>
<td>3s3p1d</td>
<td>4s3p3d1f</td>
<td></td>
</tr>
<tr>
<td>def2-TZVP</td>
<td>3/3</td>
<td>3s1p</td>
<td>5s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>3/3</td>
<td>3s2p1d</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>pc-2</td>
<td>3/3</td>
<td>3s2p1d</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>6-311+G(d,p)</td>
<td>aug'-TZ</td>
<td>3/4</td>
<td>sp</td>
<td>3s1p</td>
</tr>
<tr>
<td>6-311+G(2df,p)</td>
<td>3/4</td>
<td>sp</td>
<td>4s1p</td>
<td>5s4p1d</td>
</tr>
<tr>
<td>6-311+G(2df,2pd)</td>
<td>3/4</td>
<td>sp</td>
<td>4s2p1d</td>
<td></td>
</tr>
<tr>
<td>6-311+G(3df,3pd)</td>
<td>3/4</td>
<td>sp</td>
<td>4s3p1d</td>
<td></td>
</tr>
<tr>
<td>min-aug-cc-pVTZ</td>
<td>3/4</td>
<td>sp</td>
<td>5s4p3d1f</td>
<td></td>
</tr>
<tr>
<td>heavy-aug-cc-pVTZ</td>
<td>3/4</td>
<td>spd</td>
<td>5s4p3d2f</td>
<td></td>
</tr>
<tr>
<td>6-311++G(d,p)</td>
<td>aug-TZ</td>
<td>4/4</td>
<td>s</td>
<td>4s1p</td>
</tr>
<tr>
<td>6-311++G(2df,p)</td>
<td>4/4</td>
<td>s</td>
<td>4s1p</td>
<td>5s4p2d1f</td>
</tr>
<tr>
<td>6-311++G(2df,2pd)</td>
<td>4/4</td>
<td>s</td>
<td>4s2p1d</td>
<td></td>
</tr>
<tr>
<td>6-311++G(3df,3pd)</td>
<td>4/4</td>
<td>s</td>
<td>4s3p1d</td>
<td></td>
</tr>
<tr>
<td>def2-TZVPD</td>
<td>3/4</td>
<td>p</td>
<td>spd(sd)</td>
<td>3s2p</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>4/4</td>
<td>spd</td>
<td>4s3p2d</td>
<td></td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>4/4</td>
<td>spd</td>
<td>5s4p3d2f</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.2:** Triple zeta basis sets employed in this work.

<table>
<thead>
<tr>
<th>basis set</th>
<th>type</th>
<th>n-zeta</th>
<th>diffuse functions</th>
<th>contracted basis functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>def2-QZVP</td>
<td>QZ</td>
<td>4/4</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>4/4</td>
<td>4s3p2d1f</td>
<td>5s4p3d2f[1g]</td>
<td></td>
</tr>
<tr>
<td>min-aug-cc-pVQZ</td>
<td>aug'-QZ</td>
<td>4/5</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>heavy-aug-cc-pVQZ</td>
<td>4/5</td>
<td>spd</td>
<td>4s3p2d1f</td>
<td></td>
</tr>
<tr>
<td>def2-QZVPD</td>
<td>aug-QZ</td>
<td>4/5</td>
<td>spd(sd)</td>
<td>4s4p2d1f</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td>5/5</td>
<td>spd</td>
<td>5s4p3d2f</td>
<td></td>
</tr>
<tr>
<td>cc-pV5Z</td>
<td>5Z</td>
<td>5/5</td>
<td>5s4p3d2f[1g]</td>
<td></td>
</tr>
<tr>
<td>pc-3</td>
<td>5/5</td>
<td>5s4p2d1f</td>
<td>6s5p4d2f[1g]</td>
<td></td>
</tr>
<tr>
<td>aug-cc-pV5Z</td>
<td>aug-5Z</td>
<td>6/6</td>
<td>spdfg</td>
<td></td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>6/6</td>
<td>spdfg</td>
<td>6s5p4d3f[2g]</td>
<td></td>
</tr>
<tr>
<td>pc-4</td>
<td>7Z</td>
<td>7/7</td>
<td>7s6p3d2f[1g]</td>
<td></td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>aug-7Z</td>
<td>8/8</td>
<td>spdfg</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.3:** Quadruple zeta and larger basis sets employed in this work.
such as cc-pVTZ, except for the higher angular-momentum polarization functions in the TZ basis. In other words, full augmentation with diffuse functions increases the basis set size almost as much as an increase in cardinality by one. However, the resulting basis sets will have different exponents and contraction coefficients, and thus perform differently for a given system.

Tables 2.1–2.3 contain many basis sets with an identical or very similar composition in terms of cGTOs. Hence, some comments on the design principles and naming conventions of various basis set families are warranted.

### 2.6.3 Pople-Type Basis Sets

The 3-21G, 6-31G, and 6-311G basis sets [69] of Pople and coworkers are some of the oldest and most widely used basis sets for electronic structure calculations on molecules. Their exponents and contraction coefficients were optimized on Hartree–Fock calculations of atomic and molecular systems.

The G stands for “Gaussian” basis set, and the integers refer to the number of primitives that make up each cGTO. The integer in front of the dash refers to the core orbitals, and the integers after the dash refer to the valence orbitals. Hence, the 3-21G basis sets uses one cGTO composed of three primitives to represent core orbitals, and two cGTOs to represent valence orbitals, which are in turn composed of two and one primitives, respectively. The 6-311G basis set uses one cGTO composed of six primitives to represent core orbitals, and three cGTOs (composed of three, one, and one primitives, respectively) to represent valence orbitals. 3-21G and 6-31G are double zeta basis sets, while 6-311G is a triple zeta basis set.

Pople-type basis sets may be augmented with a set of s- and p-type diffuse functions on nonhydrogen atoms (“+”) and additional s-type diffuse functions on hydrogens (“++”), the exponents of which were optimized for anionic hydrides. Polarization functions can be added to Pople-type basis sets in a very flexible, user-defined manner [70]. For example, the widely used 6-31G(d) basis set contains a single set of d-type polarization functions on nonhydrogen atoms, while 6-311(2df,p) has two sets of d- and one set of f-type polarization functions on nonhydrogen atoms, and a single set of p-type functions on hydrogen. This flexibility has the drawback that no standard polarization pattern beyond a single set of d-type functions is defined (2d, 2df, and 3df-type functions are all widely used), and
unbalanced basis sets may result if too few\textsuperscript{29} or too many\textsuperscript{30} polarization functions are added. Such unbalanced basis sets are avoided in this thesis, and only “reasonable” combinations\textsuperscript{31} of diffuse and polarization functions are employed. Still, many different combinations of diffuse and polarization functions on hydrogen and nonhydrogen atoms are possible with Pople-type basis sets, as shown in Tables 2.1 and 2.2.

\subsection{Ahrluchs-Type (def2) Basis Sets}

Another basis set family that is similar to Pople-type basis sets in terms of size, range of applicability, and design principles are the “Karlsruhe” basis sets of Ahlrichs and coworkers. Slightly improved variants\textsuperscript{71} of the original split valence (SV, equivalent to DZ)\textsuperscript{72}, triple zeta valence (TZV)\textsuperscript{73}, and quadruple zeta valence (QZV)\textsuperscript{74} basis sets have been given the prefix “def2-”, since they are the default basis sets in the Turbomole\textsuperscript{234} electronic structure package.

Ahlrichs-type basis sets use individually optimized exponents\textsuperscript{32} for s- and p-type functions of the same principal quantum number (e.g., 2s and 2p), whereas Pople-type basis sets use the same exponents (but different contraction coefficients), yielding for example a “2sp” shell. The latter scheme saves computational resources, but requires more primitives to achieve the same accuracy as the former “fully optimized” basis sets. Another feature of Ahlrichs-type basis sets is their uniform availability for almost the whole periodic table, up to radon in conjunction with effective core potentials (ECPs). Thirdly, these basis sets have been designed and tested both with wavefunction theory (HF and correlated methods) and density-functional theory. Due to their age, Pople-type basis sets were originally designed and tested for wavefunction theory only (HF for 3-21G and 6-31G, MP2 for 6-311G), but are nonetheless often used in conjunction with DFT.

Finally, whereas Pople-type basis sets allow a user-defined combination of diffuse and polarization functions, polarization functions for Ahlrichs-type basis sets come in only two versions. The larger set of “PP” functions was designed for correlated methods (e.g., MP2) that require high-angular momentum functions to

\footnotesize
\begin{itemize}
\item \textsuperscript{29}e.g., 6-311++G is \textit{underpolarized} since it contains four cGTOs per valence orbital (4s/4p on first-row atoms) but no polarization functions
\item \textsuperscript{30}e.g., 6-31G(3df,3pd) is \textit{overpolarized} since it contains more polarization functions (3d and 1f on first-row atoms) than valence basis functions (2s/2p)
\item \textsuperscript{31}i.e., basis sets that are widely used and/or have a similar composition in terms of cGTOs as other standard basis sets (cc-pVnZ, def2-, pc-n) of the same class (DZ or TZ).
\item \textsuperscript{32}determined in HF calculations of atomic ground states
\end{itemize}
describe electron correlation. The smaller set of “P” functions is sufficient for HF and DFT calculations, where polarization functions are only needed to describe the distortion of orbitals due to the molecular environment (i.e., chemical bonding). This yields the def2-SV(P), def2-SVP, def2-TZVP, and def2-QZVP basis sets for HF and DFT calculations, and the def2-TZVPP and def2-QZVPP basis sets for correlated WFT calculations.

Diffuse functions optimized for molecular response properties (e.g., polarizabilities) have been described recently [75], yielding the def2-SVPD, def2-TZVPD, and def2-QZVPD basis sets (and their PP variants). Since the diffuse functions of Pople- and Ahlrichs-type basis sets were developed for different purposes (anions and response properties, respectively), they are quite different. Pople-type diffuse functions are of s- and p-type for first row atoms and s-type for hydrogen. Diffuse functions for Ahlrichs-type basis sets are of s- and d-type for “early” p-block elements (B, C, N); s-, p-, and d-type for “late” p-block elements (O, F, Ne); and p-type for s- and d-block elements (H; alkali, alkaline earth, and transition metals). See Tables 2.1–2.3.

2.6.5 Correlation-Consistent Basis Sets

A third family of Gaussian basis sets used in this thesis are the correlation-consistent (cc) basis sets of Dunning and coworkers [76–83]. They were designed to systematically approach the complete basis set (CBS) limit in correlated wavefunction calculations, but are nonetheless often used with HF and DFT methods due to their systematic construction. Their standard notation is cc-pV\(n\)Z for “correlation-consistent polarized valence \(n\)-tuple zeta”. Here, \(n\) is the cardinality of the basis set and takes values of D, T, Q, 5, and higher.

The attribute “correlation-consistent” refers to the fact that for each increase in \(n\) (e.g., from DZ to TZ to QZ), the maximum angular momentum and the number of polarization functions of a given angular momentum are both increased by one (e.g., from 1d to 2d1f to 3d2f1g), since each set of functions makes a similar contribution to the total electron correlation energy [76].

Diffuse functions were determined in a similar manner by correlated calculations on atomic anions [77–79], and are indicated by the prefix “aug-” standing for “augmented”. A set of diffuse functions is added for every angular momentum present in the underlying basis set, leading to spd, spdf, and spdfg diffuse sets on

\[33\text{no polarization functions on hydrogen}\]
first-row atoms for aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ, respectively. A popular variant omits diffuse functions on hydrogen atoms to give the heavy-augmented (heavy-aug-cc-pVnZ/haug-cc-pVnZ) basis sets, sometimes denoted as aug’-cc-pVnZ or jul-cc-pVnZ [84].

Truhlar and coworkers examined the effect of cc-type diffuse functions on DFT calculations and found that only sp-type functions on nonhydrogen atoms are important for a wide range of thermochemical quantities\textsuperscript{34} (especially barrier heights, electron affinities and hydrogen bond energies), and thus defined the minimally-augmented (maug-cc-pVnZ) basis sets [84]. The various augmentation schemes can be compared by inspection of Tables 2.1–2.3.

2.6.6 Polarization Consistent Basis Sets

Finally, basis sets that converge monotonically to the CBS limit in HF and DFT calculations were constructed by Jensen in analogy to the correlation-consistent basis sets, and are termed “polarization consistent” (pc-n) [85–90]. An important difference is that n does not refer strictly to the number of basis functions per valence orbital (i.e., cardinality), but rather to the “level of polarization beyond the isolated atom”. Thus, pc-1 contains a single set of d-type polarization functions on first-row atoms, whereas pc-2, pc-3, and pc-4 contain 2d1f, 4d2f1g, and 6d3f2g1h polarization functions, respectively. Note that pc-1 and pc-2 are typical DZ and TZ basis sets (Tables 2.1 and 2.2), but pc-3 is already of quintuple-zeta quality\textsuperscript{35} (comparable to cc-pV5Z), and pc-4 is a septuple-zeta (7Z) basis set (see Table 2.3).

PC-type basis sets have more low-angular momentum polarization functions optimized for describing charge polarization effects in small molecules, as opposed to the high-angular momentum polarization functions optimized for recovering atomic correlation energies in the cc-type basis sets. Diffuse functions have been determined based on calculations of anions and molecular properties such as dipole moment and polarizability. Recognizing the importance of diffuse s- and p-type functions for electron affinities and the importance of diffuse d-type functions for properties, Jensen chose a “full augmentation” scheme akin to aug-cc-pVnZ, leading to the aug-pc-n basis sets [85, 89].

Although the pc basis sets have many attractive features, one drawback is that

\textsuperscript{34}however, note that diffuse d-type functions are needed for accurate dipole/quadrupole moments and polarizabilities, i.e. for properties sensitive to the wavefunction tail [75, 85]

\textsuperscript{35}for hydrogen and first-row atoms only, for second-row atoms pc-3 and pc-4 are equivalent to QZ and 5Z, respectively [89]
<table>
<thead>
<tr>
<th>Polynomial (nZ)</th>
<th>Pople</th>
<th>def2</th>
<th>cc</th>
<th>pc</th>
</tr>
</thead>
<tbody>
<tr>
<td>D, T</td>
<td>D, T</td>
<td>D, T,Q</td>
<td>D, T, Q, 5</td>
<td>D, T, 5, 7</td>
</tr>
<tr>
<td>Polarization functions</td>
<td>user-defined, up to (3df,3pd)</td>
<td>P, PP atoms (CI/MP2), nearly degenerate excited states of atoms (HF)</td>
<td>included in definition</td>
<td>included in definition</td>
</tr>
<tr>
<td>Optimized on</td>
<td>d: molecules (HF) f: hydrides (MP4)</td>
<td>d: molecules (HF) f: hydrides (MP4)</td>
<td>d: molecules (HF) f: hydrides (MP4)</td>
<td>d: molecules (HF) f: hydrides (MP4)</td>
</tr>
<tr>
<td>Diffuse functions</td>
<td>sp on 1st-row (+) s on H, sp on 1st-row (++)</td>
<td>p on H, Li, Be sd on B, C, N spd on O, F, Ne</td>
<td>full (aug-) heavy (haug-) minimal (maug-)</td>
<td>full (aug-)</td>
</tr>
<tr>
<td>Optimized on</td>
<td>anions (HF)</td>
<td>atomic polarizabilities (HF)</td>
<td>atomic anions (HF/MRCI)</td>
<td>molecular electron affinities, multipole moments, and polarizabilities (HF)</td>
</tr>
<tr>
<td>Contraction scheme</td>
<td>segmented</td>
<td>segmented</td>
<td>general</td>
<td>general</td>
</tr>
<tr>
<td>Availability</td>
<td>H-Kr</td>
<td>H-Rn</td>
<td>H-Kr H-Ar (maug-)</td>
<td>H-Ar</td>
</tr>
<tr>
<td>References</td>
<td>[69,70]</td>
<td>[71–75]</td>
<td>[76–84,92,93]</td>
<td>[85–90]</td>
</tr>
</tbody>
</table>

Table 2.4: Summary of basis sets employed in this work

their development and testing was focused on the *uncontracted* versions of these basis sets. Jensen did provide so-called *general* contraction schemes [22, 86, 91], but due to the larger number of primitives these are less computationally efficient with typical electronic structure programs such as Gaussian than the *segmented* [22] contraction schemes of Pople- and Ahlrichs-type basis sets.

Within each group of Tables 2.1–2.3, Pople-type basis sets are followed by Ahlrichs-type (def2), correlation-consistent (cc), and polarization consistent (pc) basis sets. This ordering corresponds roughly to increasing numbers of contracted basis functions for first-row elements as shown in the last column. A concise summary of the basis sets employed in this thesis is given in Table 2.4. All basis sets were used as implemented in Gaussian 03/09 [94,95] or obtained from the basis set exchange website [96,97].
Chapter 3

Van Der Waals Interactions In Density-Functional Theory: Rare-Gas Diatomics

© Copyright 2009 American Chemical Society

Abstract

The application of conventional GGA, and meta-GGA, density functionals to van der Waals interactions is fraught with difficulties. Conventional functionals do not contain the physics of the dispersion interaction. To make matters worse, the exchange part alone can yield anything from severe over-binding to severe over-repulsion depending on the choice of functional. We have assessed a variety of exchange GGAs for their ability to reproduce exact Hartree-Fock repulsion energies in rare-gas systems, and we find that PW86 [Phys. Rev. B 33, 8800 (1986)] performs remarkably well. Addition of a dynamical correlation GGA and the nonempirical dispersion model of Becke and Johnson [J. Chem. Phys. 127, 154108 (2007)] to PW86 gives a simple GGA+dispersion theory yielding excellent rare-gas interaction curves for pairs involving He through Kr, with only two adjustable parameters for damping of the dispersion terms.

3.1 Introduction

Density functional theory (DFT) is a successful and accurate method for electronic structure calculations of atoms, molecules and solids. Its application to soft matter and weakly bound systems, including intermolecular complexes, biomolecules, molecular crystals and polymers, depends on a realistic description of van der Waals (vdW) interactions including the London dispersion force. Dispersion arises from correlated motions of electrons on well-separated systems, and is an inherently nonlocal electron correlation effect. Density functionals based on the local
density approximation (LDA) or the semilocal generalized gradient approximation (GGA) do not account for such long-range correlations and, consequently, fail to reproduce the attractive $R^{-6}$ behaviour of the interatomic potential between closed-shell atoms at large separation $R$ [98].

Problems with LDA and GGA functionals in vdW interactions arise not only in the asymptotic region, however. At the equilibrium separation of typical dispersion-bound van der Waals complexes, such as rare-gas diatomics or stacked aromatic rings, there is a weak wavefunction overlap leading to considerable Hartree-Fock (HF) repulsion between the monomers. Lacks and Gordon [99] found that LDA and GGA exchange functionals applied to the helium and neon dimers give widely varying interaction energies, from too repulsive (B88 [100]) to spurious binding (LDA, PW91 [101, 102]), when compared to exact Hartree-Fock repulsion.

A DFT treatment of vdW interactions should address both issues, the erratic behaviour of approximate exchange functionals near equilibrium separations as well as the missing long-range attraction due to dispersion [103]. Many different approaches have been used in the literature and only a superficial overview is given here. For more complete discussions see Refs. [98, 103–105].

The van der Waals density functional (vdW-DF) approach of Langreth et al. [106, 107] uses an approximate nonlocal correlation functional derived from response theory to account for long-range dispersion interactions. It is combined with revPBE [108] exchange and LDA correlation to give a seamless functional valid at all interatomic distances. The revPBE GGA was chosen for exchange as it does not produce artificial binding in vdW complexes [109,110]. The original vdW-DF method consistently overestimated the separation in vdW complexes, which was ascribed to the overestimation of Hartree-Fock repulsion by revPBE [107]. The substitution of revPBE by HF exchange indeed improves the separations and establishes the correct relative stability of different benzene dimer conformations, but leads to significant overestimation of binding energies [111–113], suggesting that the nonlocal correlation part in vdW-DF is not fully compatible with HF exchange.

Other nonempirical approaches circumvent the problems associated with approximate exchange functionals by using a separation of the electron-electron interaction into a long-range part treated exactly and a short-range part described by an exchange GGA. The method of Hirao et al. [114,115] combines long-range HF with short-range B88 [100] exchange. Short-range correlation is treated by the OP [116] functional, while the long-range correlation comes from the nonlocal Andersson-Langreth-Lundqvist [117] (ALL) vdW functional multiplied by a
damping function depending on empirical atomic radii. The range-separated hybrid (RSH+MP2) method of Ángyán et al. [103, 118] employs the LDA for short-range exchange and correlation, HF for long-range exchange, and second-order perturbation theory (MP2) to account for the dispersion interaction in a seamless manner.

The three methods mentioned so far have in common that long-range dispersion is treated by explicitly nonlocal correlation functionals of the vdW-DF, ALL or MP2 type. These are inherently more computationally demanding than the (semi-)local functionals of conventional DFT. In order to retain the computational efficiency of DFT, empirical dispersion terms of the form $-\frac{C_6}{R^6}$ have been added to conventional density functionals to give methods collectively known as DFT-D [105, 119–121]. The $C_6$ coefficients are empirical parameters derived from fits, atomic calculations, or approximate formulas. Also, the dispersion term must be attenuated at small separations $R$ by damping functions depending on empirical vdW radii. Some DFT-D methods use additional global scaling factors that depend on the underlying density functional to account for differences in the description of Hartree-Fock repulsion [105, 121] and may include reparametrization [121].

Another approach involves modifying standard DFT functionals to give a good description of van der Waals interactions without an explicit dispersion or nonlocal correlation correction. This is done by refitting the exchange-correlation functional while including van der Waals complexes in the training set [122–126]. An exception is the Wilson-Levy correlation functional [127] combined with HF exchange, which gives good results for various vdW complexes near their equilibrium distances without adjustment, although lacking theoretical justification [128].

In summary, previous attempts to make DFT applicable to van der Waals complexes either (a) include explicitly nonlocal correlation which is computationally demanding, (b) introduce highly empirical dispersion corrections, or (c) refit functionals that are fundamentally unable to account for dispersion interactions to van der Waals data nevertheless.

The nonempirical dispersion model of Becke and Johnson [129–134] has previously been used in conjunction with exact Hartree-Fock exchange [135, 136] to give an accurate treatment of vdW interactions. In this work, we replace Hartree-Fock exchange with an exchange GGA that well reproduces Hartree-Fock repulsion in rare-gas diatomics. This functional is then combined with various correlation GGAs and the Becke-Johnson dispersion model. The method is calibrated on all pair interactions between the atoms He, Ne, Ar and Kr, and yields excellent
equilibrium separations, binding energies and interatomic potential energy curves.

3.2 Hartree-Fock Repulsion In Rare-Gas Diatomics

Hartree-Fock theory gives (in the complete basis set limit) repulsive potentials for rare-gas diatomics. Exchange functionals should accurately reproduce this Hartree-Fock repulsion if DFT is to be applied to vdW interactions.

The only systematic benchmark study of the ability of exchange functionals to reproduce Hartree-Fock repulsion in rare-gas diatomics is by Lacks and Gordon [99]. They showed that most exchange functionals give large errors, over 100%, for the exchange-only interaction energy in He$_2$ and Ne$_2$, even though total atomic exchange energies are within 1%. Some functionals such as B88 are too repulsive, others such as PW91 or the LDA give an artifactual attractive interaction.

The broad range of exchange-only interaction energies obtained from approximate functionals is illustrated in Figure 3.1 for the neon dimer. The interaction energy of two Ne atoms is plotted as a function of interatomic separation for exact HF exchange, the exchange-only LDA (xLDA), and the GGA exchange functionals B86 [137], B86b [138], B88 [100], PW86 [139], PW91 [101, 102], PBE [140, 141] and revPBE [108]. The HF and xLDA curves are self-consistent. The GGA curves are evaluated using the xLDA orbitals (i.e., post-xLDA). All calculations are performed with the fully numerical, basis-set-free Numol program of Becke and Dickson [142, 143].

Van der Waals interaction curves are extremely sensitive to the choice of GGA exchange functional. [144–147] We obtain in Figure 3.1 everything from massive “binding” of $\approx 500\mu E_h$ (xLDA and PW91) to repulsion of $\approx 800\mu E_h$ (B88) compared to the exact Hartree-Fock repulsion energy of $\approx 100\mu E_h$ at the experimental equilibrium separation of Ne$_2$ [148]. The curves can be ordered according to the behaviour of the exchange GGA at large reduced density gradient $\chi$:

$$\chi(r) = \frac{\left|\nabla \rho(r)\right|}{\rho(r)^{4/3}}$$

(3.1)

GGA exchange functionals can be expressed in the form

$$E_{\chi}^{GGA} = \int e_{\chi}^{LDA}(r)F[\chi(r)]$$

(3.2)

where $e_{\chi}^{LDA}$ is the LDA exchange energy density, and $F[\chi(r)]$ is the “exchange
enhancement factor”. Standard exchange GGAs, such as those in Fig. 3.1, all have similar behaviour for small $\chi$. However, they behave very differently at large $\chi$ (see Fig. 1 in Ref. [144]), corresponding in real space to the region very far from nuclei (i.e., the asymptotic tail density). The functionals displaying artifactual binding in Fig. 3.1 are those with relatively small enhancement factor at large $\chi$; those being overly repulsive have relatively large enhancement factor at large $\chi$. Inspection of Fig. 3.1 suggests that PW86 best reproduces the exact Hartree-Fock repulsion curve in Ne$_2$, especially near the equilibrium separation.

Lacks and Gordon concluded that PW86 and B86b best reproduced Hartree-Fock repulsion in He$_2$ and Ne$_2$ on the basis of post-HF calculations with the xLDA, PW86, PW91, B86, B86b, B88 and DK87 [149] functionals.

We have extended their assessment to all pair interactions between the rare-gas atoms He, Ne, Ar and Kr, using both Hartree-Fock and xLDA orbitals, and including the exchange functionals PBE [140,141] and revPBE [108]. Table 3.2 shows
Table 3.1: Hartree-Fock repulsion energies ($\mu E_h$) of rare-gas diatomics at Tang-Toennies reference equilibrium separations $r_{ref}^{TT}$ (Å).

<table>
<thead>
<tr>
<th></th>
<th>$r_{ref}^{TT}$ [148]</th>
<th>$\Delta E^{HF}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>He$_2$</td>
<td>2.97</td>
<td>28</td>
</tr>
<tr>
<td>He-Ne</td>
<td>3.05</td>
<td>47</td>
</tr>
<tr>
<td>He-Ar</td>
<td>3.50</td>
<td>85</td>
</tr>
<tr>
<td>He-Kr</td>
<td>3.69</td>
<td>103</td>
</tr>
<tr>
<td>Ne$_2$</td>
<td>3.09</td>
<td>96</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>3.48</td>
<td>190</td>
</tr>
<tr>
<td>Ne-Kr</td>
<td>3.65</td>
<td>227</td>
</tr>
<tr>
<td>Ar$_2$</td>
<td>3.76</td>
<td>435</td>
</tr>
<tr>
<td>Ar-Kr</td>
<td>3.89</td>
<td>538</td>
</tr>
<tr>
<td>Kr$_2$</td>
<td>4.01</td>
<td>674</td>
</tr>
</tbody>
</table>

Table 3.2: Hartree-Fock repulsion energy errors (%) of exchange GGAs at reference equilibrium separations $r_{ref}^{TT}$ (post-HF).

<table>
<thead>
<tr>
<th></th>
<th>B86</th>
<th>B86b</th>
<th>B88</th>
<th>PW86</th>
<th>PW91</th>
<th>PBE</th>
<th>revPBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPE</td>
<td>-76</td>
<td>32</td>
<td>374</td>
<td>34</td>
<td>-337</td>
<td>-106</td>
<td>2</td>
</tr>
<tr>
<td>MAPE</td>
<td>85</td>
<td>34</td>
<td>374</td>
<td>34</td>
<td>337</td>
<td>106</td>
<td>60</td>
</tr>
<tr>
<td>MaxAPE</td>
<td>239</td>
<td>57</td>
<td>679</td>
<td>53</td>
<td>832</td>
<td>259</td>
<td>153</td>
</tr>
</tbody>
</table>

Table 3.3: Hartree-Fock repulsion energy errors (%) of exchange GGAs at reference equilibrium separations $r_{ref}^{TT}$ (post-xLDA).

<table>
<thead>
<tr>
<th></th>
<th>B86</th>
<th>B86b</th>
<th>B88</th>
<th>PW86</th>
<th>PW91</th>
<th>PBE</th>
<th>revPBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPE</td>
<td>-96</td>
<td>53</td>
<td>542</td>
<td>20</td>
<td>-418</td>
<td>-158</td>
<td>51</td>
</tr>
<tr>
<td>MAPE</td>
<td>111</td>
<td>53</td>
<td>542</td>
<td>22</td>
<td>418</td>
<td>158</td>
<td>75</td>
</tr>
<tr>
<td>MaxAPE</td>
<td>309</td>
<td>72</td>
<td>1126</td>
<td>82</td>
<td>1133</td>
<td>381</td>
<td>124</td>
</tr>
</tbody>
</table>

exact Hartree-Fock repulsion energies $\Delta E^{HF}$ computed with Numol at the reference equilibrium separations of Tang and Toennies [148]. Table 3.2 gives the mean percent error (MPE), mean absolute percent error (MAPE) and maximum absolute percent error (MaxAPE) of GGA exchange functionals with respect to $\Delta E^{HF}$, evaluated using Hartree-Fock orbitals (i.e., post-HF). In agreement with Lacks and Gordon, B86b and PW86 best reproduce Hartree-Fock repulsion. With xLDA orbitals however, we find that PW86 is superior to all other functionals tested (Table 3.3). Self-consistent GGA calculations are currently not possible with Numol.
Table 3.4: Exact (Hartree-Fock) and PW86 repulsion energies (μEₜₚ) of rare-gas diatomics at equilibrium separations. Worst cases (MaxAPE) in bold.

<table>
<thead>
<tr>
<th></th>
<th>Δ(E^{HF}) SCF, numerical</th>
<th>Δ(E^{PW86}) post-xLDA, numerical</th>
<th>Δ(E^{PW86}) SCF, aug-cc-pV5Z-CP (aug-cc-pV6Z, aug-pc-4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>He₂</td>
<td>28</td>
<td>51</td>
<td>29 (30, 30)</td>
</tr>
<tr>
<td>He-Ne</td>
<td>47</td>
<td>64</td>
<td>52 (50, 51)</td>
</tr>
<tr>
<td>He-Ar</td>
<td>85</td>
<td>107</td>
<td>87 (89, 88)</td>
</tr>
<tr>
<td>He-Kr</td>
<td>103</td>
<td>116</td>
<td>98</td>
</tr>
<tr>
<td>Ne₂</td>
<td>96</td>
<td>104</td>
<td>117 (113, 115)</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>190</td>
<td>189</td>
<td>198 (197, 198)</td>
</tr>
<tr>
<td>Ne-Kr</td>
<td>227</td>
<td>201</td>
<td>219</td>
</tr>
<tr>
<td>Ar₂</td>
<td>435</td>
<td>505</td>
<td>503 (507, 512)</td>
</tr>
<tr>
<td>Ar-Kr</td>
<td>538</td>
<td>611</td>
<td>626</td>
</tr>
<tr>
<td>Kr₂</td>
<td>674</td>
<td>764</td>
<td>796</td>
</tr>
<tr>
<td>MPE</td>
<td>20</td>
<td></td>
<td>8 (6, 6)</td>
</tr>
<tr>
<td>MAPE</td>
<td>22</td>
<td></td>
<td>10 (6, 6)</td>
</tr>
<tr>
<td>MaxAPE</td>
<td>82</td>
<td></td>
<td>22 (18, 20)</td>
</tr>
</tbody>
</table>

We then used Gaussian03 [94] to obtain fully self-consistent interaction energies with the PW86 functional using the keyword iop(3/74=800). A numerical integration grid of 400 radial shells and 590 angular points per atom was used, the SCF convergence criterion was set to \(10^{-8}\) with full integral accuracy during the SCF cycles. We used the aug-cc-pV5Z [150–153] basis set throughout as well as aug-cc-pV6Z [154,155] and aug-pc-4 [156] for the subset of systems containing only He, Ne and Ar atoms, as these basis sets are not available for Kr. Basis sets were obtained from the EMSL basis-set library [157, 158]. The counterpoise (CP) [159] procedure was used to correct for basis set superposition error, which was found to be up to 5% of Δ\(E^{PW86}\) for He-Ne and Ne₂ with the aug-cc-pV5Z basis set. The resulting PW86/aug-cc-pV5Z-CP interaction energies are given in Table 3.4, with results for aug-cc-pV6Z and aug-pc-4 in parentheses. The self-consistent PW86 interaction energies are even more accurate than the post-HF and post-xLDA results. In particular the overestimation of Hartree-Fock repulsion in He₂ is significantly improved.

PW86 is a remarkably accurate exchange GGA for reproducing Hartree-Fock
repulsion energies in rare-gas diatoms. Compared to more sophisticated approaches to correct for deficiencies in approximate exchange functionals when applied to van der Waals complexes, such as the use of long-range Hartree-Fock exchange [103, 114, 115, 118] or exact-exchange-based functionals themselves [128, 133, 134], it has the obvious benefit of computational efficiency.

PW86 is also appealing due to its simple functional form and the absence of any empirical parameters [139]. As with any typical GGA, it gives atomic exchange energies accurate to within 1% [26, 99], and is therefore of similar utility for general quantum chemistry as the more popular functionals B88, PW91 or PBE (see Section 3.5).

3.3 Dynamical Correlation And Dispersion Corrections

Having found a suitable exchange GGA for use with van der Waals interactions, we now consider the correlation functional. Typical semilocal correlation functionals account for short-range dynamical correlation arising from interelectronic cusp conditions [160]. Semilocal functionals are inherently unable to describe the long-range electron correlations between non-overlapping fragments that give rise to the asymptotic dispersion interaction [7, 98, 103, 109]. The application of GGAs to van der Waals interactions therefore requires an explicit long-range dispersion correction. Empirical corrections have been used successfully but suffer from limited applicability and transferability [104, 105, 119–121, 161–163]. The recent dispersion model of Becke and Johnson [132], however, is nonempirical.

The Becke-Johnson model employs the dipole moment of an electron and its associated exchange hole (as an approximation to the full exchange-correlation hole) as the source of position-dependent multipole moments leading to the dispersion interaction [129, 132, 164]. System-dependent interatomic dispersion coefficients $C_{6,ij}$, $C_{8,ij}$, and $C_{10,ij}$ are thus obtained from the exchange-hole dipole moment, effective atomic polarizabilities, and second-order perturbation theory. The dispersion model has two variants, an exact-exchange version where the dipole moment of the exchange hole and its reference electron is calculated using occupied orbitals (XX), and a density-functional version based on the Becke-Roussel model of the exchange hole (BR) [165].

The asymptotic dispersion terms need to be damped at small separations $R$. 
The Becke-Johnson scheme [134] uses

\[ E_{\text{disp}} = -\frac{1}{2} \sum_{i \neq j} \left( \frac{C_{6,ij}}{R_{\text{vdW},ij}^6} + \frac{C_{8,ij}}{R_{\text{vdW},ij}^8} + \frac{C_{10,ij}}{R_{\text{vdW},ij}^{10}} \right) \]  

(3.3)

where the sum is over all pairs of atoms \( i \) and \( j \), and the effective interatomic van der Waals separation \( R_{\text{vdW},ij} \) is related to a “critical” interatomic separation \( R_{c,ij} \) by

\[ R_{\text{vdW},ij} = a_1 R_{c,ij} + a_2 \]  

(3.4)

The critical separation \( R_{c,ij} \) is given by the average value of the ratios \( \left( \frac{C_{8,ij}}{C_{6,ij}} \right)^{1/2} \), \( \left( \frac{C_{10,ij}}{C_{6,ij}} \right)^{1/4} \) and \( \left( \frac{C_{10,ij}}{C_{8,ij}} \right)^{1/2} \), and corresponds to the separation where the three asymptotic dispersion terms are approximately equal in magnitude:

\[ \frac{C_{6,ij}}{R_{ij}^6} \approx \frac{C_{8,ij}}{R_{ij}^8} \approx \frac{C_{10,ij}}{R_{ij}^{10}} \]  

(3.5)

This is where the asymptotic series expansion is expected to break down. The two coefficients \( a_1 \) and \( a_2 \) in Eq. (3.4) are empirical parameters assumed to be universal and determined as described below.

We combine the PW86 exchange functional with various GGA functionals for dynamical correlation \( E_{\text{GGA}}^C \) and the Becke-Johnson dispersion model as follows:

\[ E_{\text{XC}} = E_{\text{PW86}} + E_{\text{GGA}}^C + E_{\text{BJ}}^\text{disp} \]  

(3.6)

We use the standard correlation GGAs P86 [166,167], PW91 [101,102] and PBE [140, 141]. All following calculations were performed with the Numol program using LDA orbitals and the Perdew-Wang uniform-gas exchange-correlation parametrization [168] (i.e., post-xcLDA). We used numerical grids of 302 angular points per atom, and 80, 120, 160 and 200 radial shells for the He, Ne, Ar and Kr atoms, respectively.

The coefficients \( a_1 \) and \( a_2 \) in Eq. (3.4) were determined by minimizing the root-mean-square percent error (RMSPE) for the interaction energies of the ten rare-gas pairs involving He, Ne, Ar and Kr with respect to reference binding energies at the experimental equilibrium separations of Table 3.1. We chose the rare-gas reference data of Tang and Toennies (TT) [148] following the suggestion by Gerber and Ángyán [103].
Table 3.5: Best-fit dispersion damping parameters and interaction-energy RMS percent errors for various combinations of correlation GGA and Becke-Johnson dispersion.

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>$a_2/\text{Å}$</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>P86-XX</td>
<td>1.80</td>
<td>-1.16</td>
<td>20.5</td>
</tr>
<tr>
<td>P86-BR</td>
<td>1.54</td>
<td>-0.53</td>
<td>28.7</td>
</tr>
<tr>
<td>PW91-XX</td>
<td>0.96</td>
<td>0.87</td>
<td>8.1</td>
</tr>
<tr>
<td>PW91-BR</td>
<td>0.76</td>
<td>1.35</td>
<td>12.5</td>
</tr>
<tr>
<td>PBE-XX</td>
<td>0.95</td>
<td>0.87</td>
<td>7.8</td>
</tr>
<tr>
<td>PBE-BR</td>
<td>0.75</td>
<td>1.25</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Results of our fits are shown in Table 3.5. The exact-exchange version (XX) of the dispersion model gives somewhat better fits than the density-functional version (BR) for all correlation GGAs. The PW91 and PBE correlation functionals give almost identical results, not surprising given that PBE generally behaves very much like PW91 [140]. P86 however, yields errors more than twice as large compared to PW91 and PBE. We therefore omit P86 from further consideration.

3.4 Binding Energy Curves

We have calculated binding energy curves using PW86x and PBEc and the dispersion damping parameters in Table 3.5. These are shown for the homonuclear dimers of He, Ne, Ar and Kr in Figure 3.2 together with the TT reference potentials. The PW91c curves are essentially identical to the PBEc binding energy curves. For clarity we also omit the curves for the BR version of the Becke-Johnson dispersion model. They are qualitatively similar to the XX-based curves and only slightly inferior. PW86xPBEc+disp gives binding energy curves that closely reproduce the reference curves. The largest deviations are observed for the heavier diatomics Ne-Kr, Ar-Ar, Ar-Kr and Kr-Kr.

Equilibrium separations were found by potential-energy scans in steps of 0.01Å, and are shown in Table 3.6 together with the TT reference values (Å). Our results are reported as reduced quantities, defined as equilibrium separation divided by the reference value. The methods using either PW91 or PBE dynamical correlation are very accurate, with mean absolute and maximum errors of about 1% and 2%, respectively.

Binding energies obtained at the equilibrium separations are given in Table 3.7
Figure 3.2: He₂, Ne₂, Ar₂, and Kr₂ binding energy curves as reduced quantities along with the TT reference values ($\mu E_h$). For comparison, we also include CCSD(T) [103], DFT+CCSD(T) [169] and RSH+MP2 [103] results. We obtain excellent binding energies with the dispersion-corrected GGA functionals, surpassing even CCSD(T) in accuracy, though our results are admittedly fits to the reference data.

3.5 Conclusions

Standard semilocal density functionals without an explicitly nonlocal correlation part or an asymptotic dispersion correction are unable to describe van der Waals interactions. Not only do they miss the attractive $R^{-6}$ behaviour at long-range, but
<table>
<thead>
<tr>
<th></th>
<th>reference (TT) [148]</th>
<th>PW91-XX</th>
<th>PW91-BR</th>
<th>PBE-XX</th>
<th>PBE-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>He₂</td>
<td>2.97</td>
<td>1.009</td>
<td>1.019</td>
<td>1.012</td>
<td>1.022</td>
</tr>
<tr>
<td>He-Ne</td>
<td>3.05</td>
<td>0.994</td>
<td>0.994</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>He-Ar</td>
<td>3.50</td>
<td>0.995</td>
<td>0.998</td>
<td>0.998</td>
<td>1.001</td>
</tr>
<tr>
<td>He-Kr</td>
<td>3.69</td>
<td>0.988</td>
<td>0.991</td>
<td>0.994</td>
<td>0.996</td>
</tr>
<tr>
<td>Ne₂</td>
<td>3.09</td>
<td>1.006</td>
<td>1.000</td>
<td>1.010</td>
<td>1.003</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>3.48</td>
<td>1.007</td>
<td>0.998</td>
<td>1.010</td>
<td>1.004</td>
</tr>
<tr>
<td>Ne-Kr</td>
<td>3.65</td>
<td>0.993</td>
<td>0.990</td>
<td>0.996</td>
<td>0.993</td>
</tr>
<tr>
<td>Ar₂</td>
<td>3.76</td>
<td>1.019</td>
<td>1.014</td>
<td>1.022</td>
<td>1.017</td>
</tr>
<tr>
<td>Ar-Kr</td>
<td>3.89</td>
<td>1.018</td>
<td>1.010</td>
<td>1.018</td>
<td>1.013</td>
</tr>
<tr>
<td>Kr₂</td>
<td>4.01</td>
<td>1.012</td>
<td>1.007</td>
<td>1.015</td>
<td>1.010</td>
</tr>
<tr>
<td>MPE</td>
<td>0.4</td>
<td>0.2</td>
<td>0.7</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>1.0</td>
<td>0.8</td>
<td>1.0</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>MaxAPE</td>
<td>1.9</td>
<td>1.9</td>
<td>2.2</td>
<td>2.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.6: Reference equilibrium separations (Å) and calculated reduced equilibrium separations for various combinations of correlation GGA and Becke-Johnson dispersion.

<table>
<thead>
<tr>
<th></th>
<th>reference (TT) [148]</th>
<th>CCSD(T) [103]</th>
<th>DFT+ CCSD(T) [169]</th>
<th>RSH+ MP2 [103]</th>
<th>PW91-XX</th>
<th>PW91-BR</th>
<th>PBE-XX</th>
<th>PBE-BR</th>
</tr>
</thead>
<tbody>
<tr>
<td>He₂</td>
<td>34.8</td>
<td>0.85</td>
<td>0.62</td>
<td>0.58</td>
<td>0.86</td>
<td>0.77</td>
<td>0.87</td>
<td>0.78</td>
</tr>
<tr>
<td>He-Ne</td>
<td>66</td>
<td>0.85</td>
<td>0.75</td>
<td>0.68</td>
<td>1.09</td>
<td>1.11</td>
<td>1.08</td>
<td>1.10</td>
</tr>
<tr>
<td>He-Ar</td>
<td>93.7</td>
<td>0.86</td>
<td>0.84</td>
<td>0.85</td>
<td>1.03</td>
<td>0.97</td>
<td>1.02</td>
<td>0.96</td>
</tr>
<tr>
<td>He-Kr</td>
<td>99.6</td>
<td>0.85</td>
<td>0.82</td>
<td>0.85</td>
<td>1.04</td>
<td>0.94</td>
<td>1.02</td>
<td>0.93</td>
</tr>
<tr>
<td>Ne₂</td>
<td>134</td>
<td>0.79</td>
<td>0.86</td>
<td>0.73</td>
<td>1.05</td>
<td>1.19</td>
<td>1.03</td>
<td>1.17</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>211</td>
<td>0.81</td>
<td>0.97</td>
<td>0.87</td>
<td>1.04</td>
<td>1.09</td>
<td>1.02</td>
<td>1.07</td>
</tr>
<tr>
<td>Ne-Kr</td>
<td>224</td>
<td>0.81</td>
<td>1.02</td>
<td>0.89</td>
<td>1.15</td>
<td>1.16</td>
<td>1.13</td>
<td>1.14</td>
</tr>
<tr>
<td>Ar₂</td>
<td>454</td>
<td>0.83</td>
<td>1.00</td>
<td>1.00</td>
<td>0.92</td>
<td>0.93</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>Ar-Kr</td>
<td>531</td>
<td>0.84</td>
<td>1.02</td>
<td>1.00</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Kr₂</td>
<td>638</td>
<td>0.83</td>
<td>1.04</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>MPE</td>
<td>-16.8</td>
<td>-10.5</td>
<td>-15.5</td>
<td>1.3</td>
<td>0.9</td>
<td>0.1</td>
<td>-0.3</td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>16.8</td>
<td>12.3</td>
<td>15.6</td>
<td>6.5</td>
<td>10.1</td>
<td>5.9</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>MaxAPE</td>
<td>21.1</td>
<td>37.8</td>
<td>41.9</td>
<td>14.6</td>
<td>23.2</td>
<td>13.4</td>
<td>22.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.7: Reference binding energies (μEₜ) and calculated reduced equilibrium binding energies for various combinations of correlation GGA and Becke-Johnson dispersion.
Table 3.8: Atomization energy errors (kcal/mol) for the exchange-correlation functional of Eq. 3.7.

<table>
<thead>
<tr>
<th>$E_{\text{GGA}}^{\text{EX}}$</th>
<th>ME</th>
<th>MAE</th>
<th>MaxAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>B86</td>
<td>6.8</td>
<td>9.2</td>
<td>37.4</td>
</tr>
<tr>
<td>B86b</td>
<td>11.1</td>
<td>12.2</td>
<td>50.5</td>
</tr>
<tr>
<td>B88</td>
<td>6.9</td>
<td>9.5</td>
<td>37.4</td>
</tr>
<tr>
<td>PW86</td>
<td>7.1</td>
<td>10.0</td>
<td>41.8</td>
</tr>
<tr>
<td>PBE</td>
<td>23.3</td>
<td>23.7</td>
<td>87.0</td>
</tr>
</tbody>
</table>

those that give vdW binding at shorter separations do so as a result of artifactual binding in their exchange parts. We have examined a variety of exchange GGAs for their ability to reproduce exact Hartree-Fock repulsion in rare-gas diatomics. While the selected functionals gave widely differing results, the PW86 exchange functional was found to be the most accurate.

PW86x was then combined with dynamical correlation functionals P86, PW91, and PBE, and the nonempirical dispersion model of Becke and Johnson, to give binding energy curves for the diatomics of the rare-gas atoms He, Ne, Ar and Kr. The P86 correlation functional yielded poor results in these systems. The PW91 and PBE correlation functionals gave binding energy curves of excellent quality. Our scheme contains only two empirical parameters in the dispersion damping function. In future work the method will be tested on vdW complexes beyond rare-gas diatomics.

We expect that functionals like Eq. 3.6, containing PW86 for exchange, will have wide applicability beyond van der Waals systems. PW86x has not been extensively benchmarked in the past, but should perform similar to other exchange GGAs. The results of atomization-energy calculations on the 222 molecules of the G3/99 benchmark set of Curtiss et al. [170] are presented in Table 3.8. The computations are done post-xcLDA with the Numo1 program. We compare mean errors (ME), mean absolute errors (MAE), and maximum absolute errors (MaxAE) for the functional

$$E_{\text{XC}} = E_{\text{GGA}}^{\text{EX}} + E_{\text{C}}^{\text{PBE}} + E_{\text{disp}}^{\text{BJ}}$$

using the exchange GGAs B86, B86b, B88, PW86, and PBE. We use the exact-exchange version of the Becke-Johnson dispersion model (XX) with the optimized damping parameters from Table 3.5, $a_1 = 0.95$ and $a_2 = 0.87\text{Å}$.

All exchange GGAs perform similarly with the exception of PBE, which gives
noticeably larger atomization energy errors compared to the other functionals. These results are consistent with previous atomization energy benchmarks for plain GGAs without a dispersion correction [171]. We conclude that PW86 exchange, combined with a dynamical correlation GGA and the Becke-Johnson dispersion model, yields excellent results for vdw interactions in rare-gas diatomics, and its performance for atomization energies is comparable to other standard GGAs.
Chapter 4

Van Der Waals Interactions In Density-Functional Theory: Intermolecular Complexes

© Copyright 2010 American Chemical Society

Abstract

In previous work (J. Chem. Theory Comput. 2009, 5, 719) we assessed the performance of standard semi-local exchange-correlation density functionals plus the nonempirical dispersion model of Becke and Johnson (J. Chem. Phys. 2007, 127, 154108) on binding energy curves of rare-gas diatomics. The results were encouraging. In this work we extend our study to 65 intermolecular complexes representing a wide variety of van der Waals interactions including dispersion, hydrogen bonding, electrostatic and stacking. Comparisons are made with other density-functional methods for van der Waals interactions in the literature.

4.1 Introduction

An accurate description of van der Waals (vdW) interactions is required for electronic structure calculations on biomolecules, intermolecular complexes, molecular crystals and polymers. Although conventional semi-local density-functional theory (DFT) gives accurate predictions for many molecular and solid-state properties, semi-local functionals are inherently unable to describe dispersion, a non-local correlation effect [172]. Thus, generalized gradient approximation (GGA), meta-GGA, and hybrid functionals are unreliable for systems where van der Waals interactions are important, even though they may give accidentally good results in limited cases.

Many methods have been developed to treat van der Waals interactions with DFT. These range from physically rigorous dispersion functionals derived from
first principles to entirely empirical corrections or parametrizations. A comprehensive review of such methods is given by Johnson et al. [12] We provide an overview of the established methods and mention more recent approaches not covered in the review of Johnson et al.

The most rigorous description of dispersion interactions is provided by explicitly nonlocal correlation functionals. However, these methods are more computationally demanding and complicated than standard DFT. An example is the Andersson-Langreth-Lundqvist (ALL) functional [117] for non-overlapping systems, also derived by Dobson and Dinte (DD) [36], which has been applied to intermolecular complexes in conjunction with long-range corrected (LC) exchange-correlation functionals [114, 115]. A seamless van der Waals density functional (vdW-DF), valid at all interatomic distances, has been developed by Langreth and coworkers [106, 107] and applied to many molecular, solid-state and biochemical systems [45]. Other approaches use \textit{ab initio} methods such as MP2 or CCSD(T) to describe long-range electron correlation, which is combined with short-range DFT [103, 118, 169, 173]. The advantages of these range-separated hybrid (RSH) methods compared to standard MP2 or CCSD(T) are reduced basis-set dependence and smaller basis-set superposition error (BSSE).

An entirely empirical approach to dispersion interactions involves the parameterization of highly flexible density functionals by including van der Waals complexes in their fitting sets, in spite of the failure of semi-local DFT to capture dispersion physics. Early functionals of this type such as X3LYP [123], M05 and M05-2X [174] give large errors for stacking interactions [175–177], while the newer M06-2X functional performs better due to additional empirical parameters [178]. Also, these functionals do not improve the description of prototypical vdW systems such as rare-gas diatomics, compared to the standard GGAs B97-1 and PBE [179]. Neither are they able to reproduce the asymptotic $R^{-6}$ behaviour of the dispersion energy at large separation [12,176].

Theoretically more sound, but still highly empirical, is the addition of explicit dispersion corrections to conventional functionals. These can take the form of $C_6 R^{-6}$ corrections (DFT-D) [105,119–121,161,163,180], or atom-centered dispersion-correcting pseudopotentials (DCP) [181–184]. Both approaches are easy to implement in existing electronic structure codes, and have negligible computational cost. However, the necessary $C_6$ coefficients and vdW radii (DFT-D), or pseudopotential parameters (DCP), must be empirically determined for each element and have limited transferability. Real dispersion coefficients depend on the oxidation state
of an atom and its molecular environment, which is disregarded in the DFT-D and DCP approaches [12,185].

Much more satisfying dispersion corrections are now available using system-dependent dispersion coefficients calculated from first principles. Three recently proposed methods provide dispersion coefficients from the ground-state electron density of the system. Tkatchenko and Scheffler [185] use Hirshfeld atomic volumes to calculate atom-in-molecule dispersion coefficients from corresponding free-atom dispersion coefficients. Their method yields accurate $C_6$ coefficients, but relies on free-atom reference data for dispersion coefficients and polarizabilities. Sato and Nakai [41] have formulated the local response dispersion (LRD) method, which evaluates polarizabilities and dispersion coefficients from first principles without the need for free-atom reference data. It uses the same local density approximation to the response function as in the DD/ALL and vdW-DF approaches. A multipole expansion of the dispersion energy is introduced, and numerical evaluation of a Casimir-Polder imaginary frequency integral is used to obtain dispersion coefficients [41], replacing the double numerical integrations required in the ALL and vdW-DF methods.

Preceeding both the Tkatchenko-Scheffler and LRD methods is the nonempirical dispersion model of Becke and Johnson [134]. This model generates dispersion coefficients from the exchange-hole dipole moment (XDM) [132] using occupied orbitals or the electron density [130]. Its theoretical foundations have been investigated by various other authors [34, 35, 164]. In addition to the exchange-hole dipole moment, the method uses atom-in-molecule polarizabilites derived from atomic reference data and Hirshfeld atomic volumes [134]. A modification introduced by Krishtal et al. [186] employs intrinsic atomic polarizabilites, obtained by a Hirshfeld partitioning of molecular polarizibility tensors, which also gives anisotropy corrections. Kong et al. [187] recently implemented the XDM model self-consistently, and assessed the importance of self-consistency on the calculation of dispersion energies and forces.

The XDM dispersion model is part of the DF07 functional [136], a universal DFT for thermochemistry, kinetics, and van der Waals interactions [188]. DF07 is exact-exchange-based, using 100% Hartree-Fock (HF) exchange. HF exchange is computationally expensive, however, and a combination of XDM dispersion with a pure GGA functional is desirable. Unfortunately, standard exchange GGAs give everything from artificial binding (eg LDA, PW91 [27, 102], PBE [140], B86 [137]) to strong over-repulsion (B88 [100]) in rare-gas diatomic tests, compared to HF
repulsion [99].

In a previous paper [189], we benchmarked standard exchange GGAs for their ability to reproduce HF repulsion in ten rare-gas diatomics, and found that the nonempirical exchange GGA of Perdew and Wang (PW86) [139,190] performs best. This conclusion was reached by comparing GGA exchange-only interaction energies, at equilibrium separations, to exact Hartree-Fock interaction energies. More recently, Murray, Lee and Langreth [191] have examined standard exchange GGAs in interacting molecular systems (dimers of H₂, N₂, CO₂, ammonia, methane, ethene, benzene and pyrazine). They also find that PW86 best reproduces Hartree-Fock repulsion energies over a range of intermolecular separations.

We then combined [189] the PW86 exchange functional with the PBE [140] correlation functional and the XDM dispersion model as follows,

\[
E_{XC} = E_{X}^{PW86} + E_{C}^{PBE} + E_{disp}^{XDM}
\]

(4.1)

and obtained excellent binding energy curves for our ten rare-gas diatomics. In the present work, we extend the benchmarking of Eq. (4.1) from rare-gas diatomics to intermolecular complexes. A comprehensive test set of 65 complexes has been assembled (see Sec. 4.2), containing vDW interactions from He₂ through electrostatic, hydrogen bonding and stacking interactions of importance in biochemistry. With only two parameters, an excellent fit is obtained to binding energies spanning three orders of magnitude in strength.

In Sec. 4.3, we compare our results with results from a variety of other methods in the literature. Our method compares quite favourably, especially considering its small number (2) of fitted parameters.

### 4.2 Fitting Of Dispersion Damping Parameters

In the XDM model of Becke and Johnson, the dispersion energy is given by

\[
E_{disp}^{XDM} = -\frac{1}{2} \sum_{i \neq j} \left( \frac{C_{6,ij}}{R_{vdW,ij}^6} + \frac{C_{8,ij}}{R_{vdW,ij}^8} + \frac{C_{10,ij}}{R_{vdW,ij}^{10}} \right)
\]

(4.2)

The nonempirical, system-dependent dispersion coefficients \(C_{6,ij}, C_{8,ij},\) and \(C_{10,ij}\) are obtained from the exchange-hole dipole moment and atom-in-molecule polarizabilities using second-order perturbation theory [132]. In the “exact-exchange”
(XX) version of the XDM model, the dipole moment of the exchange hole is calculated using occupied orbitals [134]. Alternatively, the Becke-Roussel density-functional model of the exchange hole [165] can be used to calculate an approximate dipole moment, giving the “BR” variant of the XDM model [130].

The van der Waals separations $R_{vdW,ij}$ in Eq. (4.2) are assumed to be linearly related to “critical” interatomic separations $R_{c,ij}$ by

$$R_{vdW,ij} = a_1 R_{c,ij} + a_2$$

where $a_1$ and $a_2$ are universal parameters and $R_{c,ij}$ is the average value of the ratios $(C_{8,ij}/C_{6,ij})^{1/2}$, $(C_{10,ij}/C_{6,ij})^{1/4}$ and $(C_{10,ij}/C_{8,ij})^{1/2}$. At this separation the three asymptotic dispersion terms are approximately equal to each other:

$$\frac{C_{6,ij}}{R_{c,ij}^6} \approx \frac{C_{8,ij}}{R_{c,ij}^8} \approx \frac{C_{10,ij}}{R_{c,ij}^{10}}$$

and the asymptotic expansion of the dispersion energy is no longer valid [134].

In our previous work [189], the $a_1$ and $a_2$ parameters in Eq. (4.3) were fit to the binding energies of ten rare-gas diatomics. Good performance in rare-gas systems does not, however, guarantee good performance in intermolecular complexes [175, 192]. In this work we therefore fit the damping parameters to a larger set of 65 complexes. This set includes:

- The 22 complexes of the “S22” biochemical benchmark set [192]. S22 uses CCSD(T) or MP2 geometries, and the binding energies are CCSD(T)/complete basis set estimates. Monomer deformations are not considered.

- 10 rare-gas diatomics involving He through Kr. We use the experimentally-derived data of Tang and Toennies (TT) [148].

- 12 complexes from the NC31/05 “non-covalent” database of Zhao and Truhlar [124, 179], excluding those systems duplicated in the S22 and TT sets. The NC31/05 database uses mainly MC-QCISD/3 geometries and W1 binding energies, including monomer deformation energies. We also exclude charge-transfer complexes from our training set, as GGAs strongly overestimate charge-transfer interactions due to severe self-interaction error [193, 194]. This error is partly removed by hybrid functionals that mix in a fraction of HF exchange [193, 194], and more completely by LC-hybrid methods
that use 100% long-range HF exchange [180, 195]. However, the focus of the present work is on a pure GGA functional without inclusion of HF exchange.

- 21 systems from the 45 vdw complexes of Johnson and Becke (JB) [134], excluding systems contained in the preceding databases. Binding energies for the JB systems are mainly at the estimated CCSD(T)/complete basis-set limit, and do not include monomer deformation energies.

We compiled reference data from various sources to obtain a diverse set of intermolecular complexes with binding energies ranging from 0.022 kcal/mol (He2) to 20.65 kcal/mol (hydrogen bonded uracil dimer), and including dispersion, hydrogen bonding, electrostatic and stacking interactions.

The damping parameters $a_1$ and $a_2$ are determined by minimizing the root-mean-square percent error (RMS%E)

$$\text{RMS}\%\text{E} = 100 \times \sqrt{\frac{1}{N} \sum_{i} \left( \frac{\text{BE}_{i}^{\text{calc}} - \text{BE}_{i}^{\text{ref}}}{\text{BE}_{i}^{\text{ref}}} \right)^2} \quad (4.5)$$

of our calculated binding energies $\text{BE}_{i}^{\text{calc}}$ with respect to the reference binding energies $\text{BE}_{i}^{\text{ref}}$ at the reference geometries. Cartesian coordinates for the complexes and monomers of this training set are provided in the supporting information. Our calculations were performed with the fully numerical, basis-set-free Numol program of Becke and Dickson [65,142,143] using LDA orbitals (i.e., “post-LDA”) and the Perdew-Wang uniform-electron-gas exchange-correlation parametrization [168]. We use numerical grids of 302 angular points per atom, and 80 ($Z \leq 2$), 120 ($2 < Z \leq 10$), 160 ($10 < Z \leq 18$) and 200 ($18 < Z \leq 36$) radial shells per atom.

The binding energies of the 65 complexes are shown in Table 4.1. Binding energies are taken to be positive quantities, i.e., negative values indicate a repulsive interaction. Table 4.1 also lists the dispersion contribution to the binding energy, calculated as

$$\%\text{disp} = 100 \times \frac{\text{BE}(\text{PW86PBE-XDM}) - \text{BE}(\text{PW86PBE})}{\text{BE}(\text{PW86PBE-XDM})} \quad (4.6)$$

Hydrogen bonded complexes have dispersion contributions $< 20\%$, dipolar and “mixed” interactions $< 75\%$ (with the exception of the T-shaped benzene dimer), while in dispersion-bound and “stacked” complexes the contribution of the dispersion energy exceeds 50\%.
Table 4.2 contains the optimized $a_1$ and $a_2$ values and error statistics. As can be seen from Tables 4.1 and 4.2, PW86PBE describes hydrogen bonding and dipolar (electrostatic) interactions well but fails for dispersion. Addition of the XDM dispersion energy gives accurate binding energies for the whole set of 65 complexes. A few systems (C2H4 · HF, HF · HF, NH3 · H2O, H2S · HCl, CH3SH · HCl) are slightly overbound by PW86PBE itself, and addition of the dispersion energy worsens the agreement with the reference binding energies.

Table 4.2 also shows that, for this set of 65 complexes, the BR variant of the XDM dispersion model is significantly more accurate than the XX version. The opposite result was found in our work on rare-gas diatomics [189]. This can be understood by considering how well the exchange hole in the XDM dispersion model actually approximates the full exchange-correlation (XC) hole [34, 164]. In rare-gas systems, which do not have nondynamical correlation, the exact-exchange hole (XX) is apparently a better approximation to the XC hole than the approximate BR hole. In molecular systems, however, nondynamical (left-right) correlation leads to a multi-centre-to-single-centre localization of the XC hole [196, 197]. As the localized XC holes in molecules are more effectively modelled by semi-local (meta-)GGAs such as BR than by the delocalized exact-exchange hole, the BR version of the XDM dispersion model can be expected to work better in intermolecular complexes. Dynamical correlation also contributes to the XC hole, but the dipole moment of the XC hole should be rather insensitive to the effects of dynamical correlation [34], thus justifying the use of the exchange-only hole in the XDM dispersion model.

Table 4.3 contains the $a_1$ and $a_2$ damping parameters for the PW86PBE-XDM functional obtained in our previous work [189] on rare-gas diatomics. Using these rare-gas-optimized parameters to calculate binding energies for the current set of 65 vdW complexes, we obtain similar error statistics (Table 4.3) compared to the fit in Table 4.2. In other words, the damping parameters optimized for rare-gas systems are transferable to more complex intermolecular interactions. Conversely, the damping parameters obtained in this work give good results for the binding energies of rare-gas diatomics, with mean absolute percentage errors (MA%E) of 24.7% for XDM(XX) and 10.3% for XDM(BR). This is very gratifying. The functional of Eq. (4.1), with the damped XDM dispersion model of Eq. (4.2), is apparently universally applicable to vdW interactions spanning three orders of magnitude in strength, with only two fitted parameters.
Table 4.1: Binding energies (kcal/mol), binding energy errors (kcal/mol) and dispersion contribution to binding energy (%) for the training set of 65 vDW complexes.

<table>
<thead>
<tr>
<th>complex</th>
<th>database type</th>
<th>BE/PW86PBE</th>
<th>BE/PW86PBE-XDM(XX)</th>
<th>BE/PW86PBE-XDM(BR)</th>
<th>BE/PW86PBE-XDM(BR)</th>
<th>BE/PW86PBE-XDM(BR)</th>
<th>BE/PW86PBE-XDM(BR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>He · H2O</td>
<td>dispersion</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.36</td>
</tr>
<tr>
<td>Ne · Ne</td>
<td>dispersion</td>
<td>-0.43</td>
<td>-0.43</td>
<td>-0.43</td>
<td>-0.43</td>
<td>-0.43</td>
<td>-0.43</td>
</tr>
<tr>
<td>Ar · Ar</td>
<td>dispersion</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.09</td>
</tr>
<tr>
<td>Kr · Kr</td>
<td>dispersion</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
</tr>
<tr>
<td>N2 · H2O</td>
<td>dispersion</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.4</td>
</tr>
<tr>
<td>O2 · O2</td>
<td>dispersion</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>CO2 · CO2</td>
<td>dispersion</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>CO2 · CO2</td>
<td>dispersion</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>CO2 · CO2</td>
<td>dispersion</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>CO2 · CO2</td>
<td>dispersion</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Note: The table includes binding energies (BE) and binding energy errors (BE error) for various complexes, along with the dispersion contribution to binding energy (%). The complexes are listed with their respective databases and types, such as dispersion, hydrogen bonding, and dipole-induced dipole interactions. The data is presented for both PW86PBE and PW86PBE-XDM(BR) calculations.
Table 4.2: Optimized dispersion damping parameters and error statistics for the training set of 65 vdW complexes

<table>
<thead>
<tr>
<th>dispersion</th>
<th>none</th>
<th>XDM(XX)</th>
<th>XDM(BR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.68</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>$a_2$ (Å)</td>
<td>1.43</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td>RMS%E (%)</td>
<td>96.8</td>
<td>24.0</td>
<td>15.8</td>
</tr>
<tr>
<td>MA%E (%)</td>
<td>79.1</td>
<td>19.9</td>
<td>12.6</td>
</tr>
<tr>
<td>MAE (kcal/mol)</td>
<td>1.72</td>
<td>0.53</td>
<td>0.33</td>
</tr>
<tr>
<td>MaxE(−) (kcal/mol)</td>
<td>-11.28 (A · T stack)</td>
<td>-3.93 (A · T stack)</td>
<td>-2.16 (A · T stack)</td>
</tr>
<tr>
<td>MaxE(+) (kcal/mol)</td>
<td>0.73 (CH$_3$SH · HCl)</td>
<td>2.30 (CH$_3$SH · HCl)</td>
<td>1.98 (CH$_3$SH · HCl)</td>
</tr>
</tbody>
</table>

Table 4.3: Error statistics for the current set of 65 vdW complexes using rare-gas-optimized damping parameters [189]

<table>
<thead>
<tr>
<th>dispersion</th>
<th>XDM(XX)</th>
<th>XDM(BR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.95</td>
<td>0.75</td>
</tr>
<tr>
<td>$a_2$ (Å)</td>
<td>0.87</td>
<td>1.25</td>
</tr>
<tr>
<td>RMS%E (%)</td>
<td>30.5</td>
<td>18.9</td>
</tr>
<tr>
<td>MA%E (%)</td>
<td>23.3</td>
<td>14.2</td>
</tr>
<tr>
<td>MAE (kcal/mol)</td>
<td>0.73</td>
<td>0.35</td>
</tr>
<tr>
<td>MaxE(−) (kcal/mol)</td>
<td>-5.95 (A · T stack)</td>
<td>-0.76 (A · T stack)</td>
</tr>
<tr>
<td>MaxE(+) (kcal/mol)</td>
<td>1.74 (CH$_3$SH · HCl)</td>
<td>2.26 (CH$_3$SH · HCl)</td>
</tr>
</tbody>
</table>

4.3 Performance On The S22 Benchmark Set

The “S22” database of Jurecka et al. [192] contains 22 intermolecular complexes of biochemical interest and covers hydrogen bonding, dispersion and stacking interactions. It provides CCSD(T) binding energies at the estimated complete basis-set limit, and has been widely adopted to assess the performance of electronic structure methods for intermolecular interactions. In Table 4.4, we list mean absolute errors (MAE, kcal/mol) and mean absolute percent errors (MA%E) for a variety of DFT methods for which benchmark data on the S22 set are available in the literature [41, 176, 178, 180, 183, 185, 198–202].

The Becke-Roussel variant of the XDM dispersion model, XDM(BR), gives excellent binding energies for the S22 set as demonstrated by its low MAE and MA%E values. Its accuracy is comparable to the empirical DFT-D methods, the highly
<table>
<thead>
<tr>
<th>method</th>
<th>type</th>
<th>MAE (S22)</th>
<th>MAE (HB)</th>
<th>MAE (disp)</th>
<th>MAE (mix)</th>
<th>MA%E</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW86PBE-XDM(BR)</td>
<td>GGA-XDM</td>
<td>0.46</td>
<td>0.33</td>
<td>0.68</td>
<td>0.34</td>
<td>7.4</td>
</tr>
<tr>
<td>PW86PBE-XDM(BR) TT [189]</td>
<td>GGA-XDM</td>
<td>0.31</td>
<td>0.52</td>
<td>0.27</td>
<td>0.14</td>
<td>6.2</td>
</tr>
<tr>
<td>PW86PBE-XDM(XX)</td>
<td>GGA-XDM</td>
<td>0.81</td>
<td>0.31</td>
<td>1.39</td>
<td>0.64</td>
<td>14.3</td>
</tr>
<tr>
<td>PW86PBE-XDM(XX) TT [189]</td>
<td>GGA-XDM</td>
<td>1.33</td>
<td>0.43</td>
<td>2.39</td>
<td>1.03</td>
<td>26.7</td>
</tr>
<tr>
<td>vDW-DF(revPBE) [198]</td>
<td>GGA+vdW-DF</td>
<td>1.39</td>
<td>2.81</td>
<td>0.79</td>
<td>0.65</td>
<td>18.3</td>
</tr>
<tr>
<td>vDW-DF(B86) [199]</td>
<td>GGA+vdW-DF</td>
<td>0.53</td>
<td>0.76</td>
<td>0.58</td>
<td>0.23</td>
<td>na</td>
</tr>
<tr>
<td>LC-BOP+LRD(6+8+10) [41]</td>
<td>LC hybrid GGA+LRD</td>
<td>0.27</td>
<td>0.35</td>
<td>0.20</td>
<td>0.28</td>
<td>5.7</td>
</tr>
<tr>
<td>PBE+TS [185]</td>
<td>GGA+TS</td>
<td>0.30</td>
<td>0.46</td>
<td>0.30</td>
<td>0.14</td>
<td>na</td>
</tr>
<tr>
<td>ωB97X-D [180]</td>
<td>LC hybrid GGA-D</td>
<td>0.22</td>
<td>0.24</td>
<td>0.26</td>
<td>0.17</td>
<td>5.4</td>
</tr>
<tr>
<td>B97-D [180]</td>
<td>GGA-D</td>
<td>0.50</td>
<td>0.84</td>
<td>0.43</td>
<td>0.24</td>
<td>6.4</td>
</tr>
<tr>
<td>B3LYP-D [180]</td>
<td>hybrid GGA-D</td>
<td>0.48</td>
<td>0.81</td>
<td>0.35</td>
<td>0.28</td>
<td>8.5</td>
</tr>
<tr>
<td>BLYP-D [180]</td>
<td>GGA-D</td>
<td>0.33</td>
<td>0.28</td>
<td>0.52</td>
<td>0.16</td>
<td>8.6</td>
</tr>
<tr>
<td>revPBE+LAP [183]</td>
<td>GGA-DCP(LAP)</td>
<td>0.57</td>
<td>1.11</td>
<td>0.42</td>
<td>0.22</td>
<td>7.0</td>
</tr>
<tr>
<td>B3LYP-DCP [200]</td>
<td>hybrid GGA-DCP</td>
<td>0.93</td>
<td>1.34</td>
<td>0.90</td>
<td>0.56</td>
<td>20.4</td>
</tr>
<tr>
<td>M06-2X [178]</td>
<td>hybrid meta-GGA</td>
<td>0.40</td>
<td>0.70</td>
<td>0.17</td>
<td>0.35</td>
<td>6.4</td>
</tr>
<tr>
<td>M05-2X [176]</td>
<td>hybrid meta-GGA</td>
<td>0.86</td>
<td>0.75</td>
<td>1.26</td>
<td>0.53</td>
<td>14.8</td>
</tr>
<tr>
<td>mPW2PLYP-D [202]</td>
<td>double hybrid GGA-D</td>
<td>0.46</td>
<td>0.50</td>
<td>0.70</td>
<td>0.16</td>
<td>8.4</td>
</tr>
<tr>
<td>B2PLYP-D [202]</td>
<td>double hybrid GGA-D</td>
<td>0.27</td>
<td>0.18</td>
<td>0.48</td>
<td>0.12</td>
<td>6.6</td>
</tr>
<tr>
<td>ωB97X-2(LP) [201]</td>
<td>LC double hybrid GGA</td>
<td>0.24</td>
<td>0.21</td>
<td>0.30</td>
<td>0.22</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Table 4.4: Mean absolute errors (MAE, kcal/mol) of various DFT methods for the S22 set and the subsets of hydrogen-bonded (HB), dispersion-dominated (disp) and mixed (mix) complexes; mean absolute percentage errors (MA%E) for the S22 set. Methods using optimized geometries instead of the S22 reference geometries are shown in italics.
parametrized M06-2X and ωB97X functionals, and the much more expensive “double hybrid” functionals (which include non-local correlation through second order MP2 perturbation theory). By coincidence, our previous damping parameter fit to rare-gas systems [189] (denoted as “TT” in Table 4.4) gives slightly better error statistics than the current fit to 65 intermolecular complexes. We also note that the exact-exchange version of the XDM dispersion model, XDM(XX), is much less accurate for the S22 set for reasons explained in Section 4.2. Given its higher accuracy and lower computational cost, we prefer the XDM(BR) variant over XDM(XX). XDM(BR) is also the method which was recently implemented self-consistently [187].

The nonempirical dispersion approaches of Sato and Nakai (LRD) and Tkatchenko and Scheffler (TS) also give excellent binding energies for the S22 set, as do the empirical (DFT-D) dispersion corrections. The van der Waals density functional (vdW-DF) is less accurate and, as shown by Gulans et al. [198] and Klimes et al. [199], the results depend on the underlying exchange functional. With revPBE [108] exchange the complexes of the S22 set are systematically underbound [198, 199], while the combination of B86 [137] exchange and vdw-DF gives substantially improved results [199].

A similar case is B3LYP-DCP [200], which combines the dispersion-correcting pseudopotentials of DiLabio and coworkers [203] with the B3LYP [204, 205] functional. Due to the use of overly repulsive B88 exchange [100] in B3LYP, B3LYP-DCP underbinds all systems in the S22 database except for the water dimer [200]. The revPBE+LAP [183] method, which combines the revPBE [108] exchange-correlation functional with a dispersion-correcting local atomic potential (LAP) [183], does not suffer from this deficiency, but is still rather inaccurate for hydrogen-bonded complexes.

A variety of DFT methods reproduce the binding energies of the intermolecular complexes in the S22 database very well. Those that use nonempirical dispersion coefficients and only a few parameters in their damping functions are the present XDM(BR), the LRD of Sato and Nakai, and the Tkatchenko-Scheffler methods. The nonempirical vdw-DF is comparably less accurate with the original revPBE exchange, but can be improved by changing the underlying exchange functional [199]. The influence of the exchange functional is well known [99,144,145], but has been somewhat overlooked until the recent studies of Refs. [189,191,199] and the present work.
4.4 Conclusions

We have shown that the XDM dispersion model of Becke and Johnson can be combined with standard GGAs for exchange (PW86) and correlation (PBE) to give an excellent description of van der Waals interactions. The XDM dispersion model contains only two empirical parameters in the damping function. These have been fit to a set of 65 complexes ranging from rare-gas systems to nucleic acid base pairs and spanning three orders of magnitude in binding energy strength. Also, the dispersion damping parameters optimized for rare-gas diatomics in our previous work [189] are found to be highly transferable to the larger set of intermolecular interactions.

The Becke-Roussel variant of XDM, XDM(BR), is more accurate for intermolecular complexes than the exact-exchange-hole variant, XDM(XX), and we have rationalized this result. The performance of the XDM dispersion model on the S22 database has been compared to a variety of alternative DFT methods that account for dispersion, and the XDM(BR) method compares very favourably. In future work, we will explore geometry optimizations of intermolecular complexes using XDM-derived dispersion forces.
Chapter 5

Nonempirical Density-Functional Theory For Van Der Waals Interactions

© Copyright 2010 Canadian Science Publishing or its licensors

Abstract

In previous work, Kannemann and Becke [J. Chem. Theory Comput. 5, 719 (2009) and J. Chem. Theory Comput. 6, 1081 (2010)] have demonstrated that the generalized gradient approximations (GGAs) of Perdew and Wang for exchange [Phys. Rev. B 33, 8800 (1986)] and Perdew, Burke, and Ernzerhof for correlation [Phys. Rev. Lett. 77, 3865 (1996)], plus the dispersion density functional of Becke and Johnson [J. Chem. Phys. 127, 154108 (2007)], comprise a nonempirical density-functional theory of high accuracy for thermochemistry and van der Waals complexes. The theory is nonempirical except for two universal cutoff parameters in the dispersion energy. Our calculations so far have been grid based and have employed the local density approximation (LDA) for the orbitals. In this work we employ orbitals from self-consistent GGA calculations using Gaussian basis sets. The results, on a benchmark set of 65 van der Waals complexes, are similar to our grid-based post-LDA results. This work sets the stage for van der Waals force computations and geometry optimizations.

5.1 Introduction

The treatment of weak chemical interactions in density-functional theory (DFT) is fraught with difficulties. The weakest interaction, dispersion, is the result of long range correlations which are impossible to model with standard local or semilocal density-functional approximations [7] such as the local density approximation (LDA) or generalized gradient approximations (GGAs). To make matters worse,
the exchange part alone of the LDA or standard GGAs gives anything from massive over-binding to massive over-repulsion in dispersion-bound complexes depending on the choice of functional [189]. Compared to the purely repulsive curves given by Hartree-Fock theory, the performance of the various standard exchange GGAs in the DFT literature is unacceptably erratic (see Fig. 1 in Ref. [189]).

The latter of these problems, reproducing Hartree-Fock repulsion in closed-shell interactions, has been addressed by Lacks and Gordon [99] and, more recently, by Kannemann and Becke [189] and Murray, Lee, and Langreth [191]. It has been discovered that, among the many available exchange GGAs in the literature, the nonempirical exchange GGA of Perdew and Wang [26] (PW86) reproduces Hartree-Fock repulsions remarkably well. For the ten pair interactions between the rare-gas atoms He, Ne, Ar, and Kr, self-consistent PW86 repulsion energies have a mean absolute percent error with respect to Hartree-Fock of only 10 percent [189]. Considering that the model underlying PW86 is not explicitly designed to capture weak interatomic interactions, this agreement is perhaps fortuitous. It has been pointed out by Zhang, Pan, and Yang [206] that weak interactions are extremely sensitive to the behaviour of exchange GGAs in the limit of large reduced density gradient

\[ \chi(r) = \frac{\left| \nabla \rho(r) \right|}{\rho(r)^{1/3}} \]  

Fortuitous or not, the large-\(\chi\) behaviour of PW86 is ideal for reproducing closed-shell Hartree-Fock repulsions. The “B86b” exchange GGA of Becke [23] also performs well at this task, and has nearly the same asymptotic behaviour as PW86 (Fig. 1 in Ref. [206]).

The long-range London dispersion interaction responsible for binding in van der Waals (vdW) complexes cannot be obtained from local or semilocal density functionals, as first argued by Kristyan and Pulay [7]. Explicit models incorporating the necessary long-range physics are known [36, 37, 43, 44, 207, 208], but are complicated. Models of the asymptotic dispersion interaction (i.e. dispersion coefficients) of relatively simple form have been proposed by several groups [41, 50, 52, 185]. We use in our work the model of Becke and Johnson [50, 52], which exploits non-sphericity of the exchange hole around its reference point to generate \(C_{6,ij}, C_{8,ij},\) and \(C_{10,ij}\) interatomic dispersion coefficients in a nonempirical density-functional manner. We will refer to this model as the exchange-hole dipole moment
(XDM) dispersion model. Although XDM generates asymptotic dispersion coefficients without empirical fit parameters, two (universal) fit parameters are required in order to damp the dispersion energy at small internuclear separations.

We have reported in Refs. [189] and [209] that the PW86 exchange GGA combined with the nonempirical correlation GGA of Perdew, Burke, and Ernzerhof [29] (PBE) and the XDM dispersion model yields excellent binding-energy curves in rare-gas diatomics [189] and excellent binding energies in a wide variety of intermolecular complexes [209]. Our work thus far has been grid-based fully numerical, using LDA orbitals computed by the basis-set-free \texttt{Numol} program [65, 66] (i.e. “post-LDA”). In the present work, we instead use self-consistent PW86+PBE orbitals (i.e., “post-GGA”) computed with the Gaussian09 program [95]. The objective is to obtain PW86+PBE+XDM forces suitable for geometry optimization of intermolecular complexes. Forces cannot be obtained from our \texttt{Numol} post-LDA approach, and self-consistent \texttt{Numol} GGA calculations are not yet feasible.\footnote{\texttt{Numol} numerical grid noise in the GGA exchange-correlation potential is too large for application to dispersion-bound systems.} We therefore explore a post-Gaussian09-GGA approach in this work.

In Section 2, we discuss the suitability for our purposes of the Dunning [76] aug-cc-pVDZ and aug-cc-pVTZ basis sets by considering rare-gas Hartree-Fock repulsion energies. In Section 3, the two dispersion damping parameters in our PW86+PBE+XDM functional are fit to benchmark binding energies of 65 intermolecular complexes compiled in Ref. [209]. The results are of excellent quality, similar in quality to our \texttt{Numol} post-LDA results [209]. Future directions are discussed in Section 4.

### 5.2 Hartree-Fock Repulsion In Rare-Gas Diatomics: Basis-Set Considerations

In Ref. [189], the ability of the PW86 exchange GGA to reproduce Hartree-Fock repulsion energies in rare-gas systems was carefully studied. Our test set consisted of the ten pair interactions between the atoms He, Ne, Ar, and Kr. Post-Hartree-Fock, post-exchange-only-LDA, and self-consistent PW86 repulsion energies, computed at experimental Tang-Toennies internuclear separations [210], were compared to exact Hartree-Fock repulsion energies from fully numerical \texttt{Numol} calculations. In Table 5.1, the Hartree-Fock and the self-consistent PW86 repulsion energies are listed. The deviation of the PW86 from the Hartree-Fock results is remarkably small. The mean absolute error of PW86 is only 10 percent, and the
The PW86 benchmarks in Table 5.1 were obtained with Gaussian09 and the Dunning aug-cc-pV5Z basis set, including counterpoise (CP) corrections for basis-set superposition error [211]. This basis set is much too large, however, for practical applications. CP corrections are impractical as well, especially for the computation of intra-molecular dispersion energies. In this work we therefore undertake additional PW86 calculations with small Dunning basis sets, and without CP corrections, in search of a practical PW86 scheme able to mimic Hartree-Fock repulsions in rare-gas systems.

We consider the Dunning basis sets aug-cc-pVDZ and aug-cc-pVTZ to be of practical size for general applications and have therefore assessed them, without CP correction, on our repulsion-energy test set. Integration grids of 200 radial $\times$ 590 angular points (unpruned) were used throughout. The results are reported in the last two columns of Table 5.1. Both the DZ and the TZ basis sets give repulsion for all ten systems considered. However, the triple-zeta results are clearly superior to the double-zeta results. Indeed the triple-zeta data is similar in quality to the
CP-corrected quintuple-zeta data. The worst cases for aug-cc-pVDZ noticeably involve the Ne atom, but this doesn’t necessarily rule out the utility of aug-cc-pVDZ in general.

We are therefore encouraged to adopt aug-cc-pVTZ, without CP corrections, as a viable and practical basis set for GGA computations of weak chemical interactions. The viability of non-CP-corrected aug-cc-pVDZ remains to be seen. We apply both to the computation of vdW binding energies in the next section.

### 5.3 Binding Energies Of Intermolecular Complexes

A comprehensive binding-energy (BE) test set of 65 intermolecular complexes, spanning three orders of binding-energy strength from 0.022 kcal/mol (He₂) to 20.65 kcal/mol (uracil dimer), has been assembled in Ref. [209] for the assessment of vdW computational methods. The set includes the ten rare-gas diatomics in Table 5.1 with experimentally derived BEs from Tang and Toennies [210], 21 complexes from the test set of Johnson and Becke [56] with CCSD(T)-computed BEs, the 22 complexes of the “S22” biochemical benchmark set of Jurecka et al [212] also with CCSD(T)-computed BEs, and 12 complexes from Zhao and Truhlar [213, 214] with BEs computed by W1 theory. These 65 complexes are listed in Table 5.2 along with their reference binding energies. Cartesian coordinate files of all the complexes and their monomers are available as supplementary data to Ref. [209].

We assess the exchange-correlation functional

\[
E_{XC} = E_X^{PW86} + E_C^{PBE} + E_{disp}^{XDM}
\]

on our 65 benchmark complexes and report the results in Table 5.2. The PW86 exchange GGA is given in Ref. [26] and the PBE correlation GGA in Ref. [29]. The XDM dispersion energy is given by [56]

\[
E_{disp}^{XDM} = -\frac{1}{2} \sum_{i \neq j} \left( \frac{C_{6,ij}}{R_{vdW,ij}^6 + R_{ij}^6} + \frac{C_{8,ij}}{R_{vdW,ij}^8 + R_{ij}^8} + \frac{C_{10,ij}}{R_{vdW,ij}^{10} + R_{ij}^{10}} \right)
\]

with dispersion coefficients \(C_{6,ij}, C_{8,ij},\) and \(C_{10,ij}\) computed as in Refs. [52] and [54]. The effective vdW interatomic separations are linearly related to “critical” interatomic separations by

\[
R_{vdW,ij} = a_1 R_{c,ij} + a_2
\]
with the critical separation given by the average of \( \left( \frac{C_{8,ij}}{C_{6,ij}} \right)^{1/2}, \left( \frac{C_{10,ij}}{C_{6,ij}} \right)^{1/4} \) and \( \left( \frac{C_{10,ij}}{C_{8,ij}} \right)^{1/2}. \)

We use the “BR” variant of XDM, since the BR variant was found to be superior to the “XX” variant in Ref. [209]. XDM-BR is based on a density-functional approximation of the exchange hole, whereas XDM-XX uses the exact orbital-based exchange hole. That the former outperforms the latter in intermolecular complexes is rationalized in Ref. [209].

Self-consistent PW86+PBE orbitals were computed with Gaussian09 using the aug-cc-pVDZ and aug-cc-pVTZ basis sets, without CP corrections, as motivated in the previous section. The XDM dispersion energy was then computed by an interface program [215] that reads Gaussian09 orbital information from “wfn” files. These calculations are therefore not fully self-consistent, rather “post-GGA”. Since the dispersion energy is a very small fraction of the total energy, however, we expect that ignoring self-consistency in the \( E_{\text{disp}}^{XDM} \) term is of negligible consequence. Investigations of the XDM dispersion model by Kong et al [187] support this expectation.

The binding energies in Table 5.2 of the present methods are quite similar to those of our previous numerical post-LDA method [209]. The present mean absolute percent error (MAPE) with the aug-cc-pVDZ basis is 16.6, and with the aug-cc-pVTZ basis is 11.5, compared to our previous 12.6 percent. Best-fit values (minimum RMS percent error) of the damping parameters are

\[
\begin{align*}
    a_1 &= 0.80, \quad a_2 = 1.49 \text{ Å (SCF aug-cc-pVDZ)} \\
    a_1 &= 0.79, \quad a_2 = 1.36 \text{ Å (SCF aug-cc-pVTZ)} \\
    a_1 &= 0.82, \quad a_2 = 1.16 \text{ Å (numerical post-LDA [209])}
\end{align*}
\]

Here we also see similarity between the basis-set and our previous numerical results, especially for the \( a_1 \) parameter.

The quality of the double-zeta results, though not as good overall as triple-zeta, is pleasantly surprising. Even the rare-gas diatomic binding energies involving Ne are in good agreement with reference data, despite the notably poor repulsion energies in Section 2. The under-repulsive tendency of the DZ basis set compared to TZ (see Table 5.1) is apparently compensated by a slightly larger best-fit \( a_2 \) parameter. The only significant qualitative failure of the double-zeta method is the parallel benzene dimer (the last entry in Table 5.2), for which only one third of the reference binding energy is obtained. This failure is not seen in our numerical or triple-zeta results.
5.4 Conclusions And Outlook

The nonempirical (except for two universal dispersion damping parameters) exchange-correlation density functional of Eq. 5.2 was found in Refs. [189] and [209] to be an excellent model for van der Waals interactions in chemistry. The GGA exchange and nondynamical correlation parts, PW86+PBE, are a very good model for intramolecular chemistry as well, as are most standard GGAs. On the atomization energies of the 222 molecules of the G3/99 thermochemical test set of Curtiss et al [216] the mean absolute error of PW86+PBE+XDM is 10.0 kcal/mol [189]. This is similar to the mean absolute error of 9.3 kcal/mol [216] for the popular BLYP [25,31] GGA. Thus PW86+PBE+XDM is a promising GGA-DFT for simultaneous optimization of intra- and inter-molecular geometries in complex chemical systems.

Our previous post-LDA approach [189,209], while useful for assessing the performance of density functionals on energetics, cannot provide forces for optimization of geometries. Here we have tested a methodology with the promise to provide energies, and forces, at reasonable computational cost. This PW86+PBE/aug-cc-pVnZ scheme, with $n = 2$ or 3 and XDM dispersion added post-GGA, will provide useful forces (and Hessians) if orbital self-consistency can be ignored in the XDM part, and changes in the XDM dispersion coefficients $C_{6,ij}$, $C_{8,ij}$, and $C_{10,ij}$ with molecular geometry can be ignored in the energy derivatives.

Preliminary tests of these suppositions by Kong et al [187] are encouraging. We will undertake our own independent tests in future work. Also in future work, we will refit the dispersion damping parameters $a_1$ and $a_2$ for inclusion of exact Hartree-Fock exchange (i.e. “hybrid” functionals plus XDM dispersion) for even better accuracy in the intramolecular energetics.
<table>
<thead>
<tr>
<th>Complex</th>
<th>Set</th>
<th>Reference</th>
<th>Numerical</th>
<th>aTZ</th>
<th>aDZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>post-LDA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He - He</td>
<td>TT</td>
<td>0.022</td>
<td>0.018</td>
<td>0.021</td>
<td>0.015</td>
</tr>
<tr>
<td>He - Ne</td>
<td>TT</td>
<td>0.041</td>
<td>0.048</td>
<td>0.051</td>
<td>0.050</td>
</tr>
<tr>
<td>He - Ar</td>
<td>TT</td>
<td>0.059</td>
<td>0.059</td>
<td>0.058</td>
<td>0.052</td>
</tr>
<tr>
<td>He - Kr</td>
<td>TT</td>
<td>0.063</td>
<td>0.065</td>
<td>0.064</td>
<td>0.055</td>
</tr>
<tr>
<td>Ne - Ne</td>
<td>TT</td>
<td>0.084</td>
<td>0.100</td>
<td>0.088</td>
<td>0.097</td>
</tr>
<tr>
<td>Ne - Ar</td>
<td>TT</td>
<td>0.132</td>
<td>0.143</td>
<td>0.133</td>
<td>0.132</td>
</tr>
<tr>
<td>Ne - Kr</td>
<td>TT</td>
<td>0.141</td>
<td>0.164</td>
<td>0.154</td>
<td>0.148</td>
</tr>
<tr>
<td>Ar - Ar</td>
<td>TT</td>
<td>0.285</td>
<td>0.255</td>
<td>0.245</td>
<td>0.217</td>
</tr>
<tr>
<td>Ar - Kr</td>
<td>TT</td>
<td>0.333</td>
<td>0.311</td>
<td>0.302</td>
<td>0.256</td>
</tr>
<tr>
<td>Kr - Kr</td>
<td>TT</td>
<td>0.400</td>
<td>0.381</td>
<td>0.378</td>
<td>0.311</td>
</tr>
<tr>
<td>(NH3)2 [C2h]</td>
<td>S22</td>
<td>3.17</td>
<td>3.24</td>
<td>3.23</td>
<td>3.29</td>
</tr>
<tr>
<td>(H2O)2 [Cs]</td>
<td>S22</td>
<td>5.02</td>
<td>5.23</td>
<td>5.22</td>
<td>5.25</td>
</tr>
<tr>
<td>Formamide dimer [C2h]</td>
<td>S22</td>
<td>15.96</td>
<td>16.09</td>
<td>16.08</td>
<td>16.20</td>
</tr>
<tr>
<td>Uracil dimer [C2h]</td>
<td>S22</td>
<td>20.65</td>
<td>20.28</td>
<td>20.45</td>
<td>20.70</td>
</tr>
<tr>
<td>2-pyridoxine-2-aminopyridine</td>
<td>S22</td>
<td>16.71</td>
<td>17.44</td>
<td>17.53</td>
<td>17.89</td>
</tr>
<tr>
<td>Adenine-thymine WC</td>
<td>S22</td>
<td>16.37</td>
<td>16.68</td>
<td>16.89</td>
<td>17.17</td>
</tr>
<tr>
<td>(CH4)2 [D3d]</td>
<td>S22</td>
<td>0.53</td>
<td>0.54</td>
<td>0.53</td>
<td>0.77</td>
</tr>
<tr>
<td>(C2H4)2 [D2d]</td>
<td>S22</td>
<td>1.51</td>
<td>1.36</td>
<td>1.38</td>
<td>1.55</td>
</tr>
<tr>
<td>Benzene - CH4 [C3]</td>
<td>S22</td>
<td>1.50</td>
<td>1.48</td>
<td>1.48</td>
<td>1.50</td>
</tr>
<tr>
<td>Benzene dimer PD [C2h]</td>
<td>S22</td>
<td>2.73</td>
<td>2.63</td>
<td>2.64</td>
<td>2.63</td>
</tr>
<tr>
<td>Pyrazine dimer [Cs]</td>
<td>S22</td>
<td>4.42</td>
<td>3.69</td>
<td>3.62</td>
<td>3.64</td>
</tr>
<tr>
<td>Uracil dimer stack [C2]</td>
<td>S22</td>
<td>10.12</td>
<td>8.78</td>
<td>8.71</td>
<td>8.47</td>
</tr>
<tr>
<td>Indole-benzene stack</td>
<td>S22</td>
<td>5.22</td>
<td>4.27</td>
<td>4.23</td>
<td>4.16</td>
</tr>
<tr>
<td>Adenine-thymine stack</td>
<td>S22</td>
<td>12.23</td>
<td>10.07</td>
<td>9.96</td>
<td>9.82</td>
</tr>
<tr>
<td>Ethene-ethyne [C2v]</td>
<td>S22</td>
<td>1.53</td>
<td>1.71</td>
<td>1.76</td>
<td>1.92</td>
</tr>
<tr>
<td>Benzene-H2O [Cs]</td>
<td>S22</td>
<td>3.28</td>
<td>3.17</td>
<td>3.19</td>
<td>3.12</td>
</tr>
<tr>
<td>Benzene-NH3 [Cs]</td>
<td>S22</td>
<td>2.35</td>
<td>2.22</td>
<td>2.24</td>
<td>2.24</td>
</tr>
<tr>
<td>Benzene- HCN [Cs]</td>
<td>S22</td>
<td>4.46</td>
<td>4.15</td>
<td>4.30</td>
<td>4.26</td>
</tr>
<tr>
<td>Benzene dimer T-shaped [C2v]</td>
<td>S22</td>
<td>2.74</td>
<td>2.40</td>
<td>2.52</td>
<td>2.79</td>
</tr>
<tr>
<td>Indole-benzene T-shaped</td>
<td>S22</td>
<td>5.73</td>
<td>4.95</td>
<td>5.16</td>
<td>5.42</td>
</tr>
<tr>
<td>Phenol dimer</td>
<td>S22</td>
<td>7.05</td>
<td>6.52</td>
<td>6.55</td>
<td>6.65</td>
</tr>
<tr>
<td>He - N2 L-shaped</td>
<td>JB</td>
<td>0.053</td>
<td>0.040</td>
<td>0.046</td>
<td>0.035</td>
</tr>
<tr>
<td>He - N2 T-shaped</td>
<td>JB</td>
<td>0.066</td>
<td>0.062</td>
<td>0.066</td>
<td>0.082</td>
</tr>
<tr>
<td>He - FCI</td>
<td>JB</td>
<td>0.097</td>
<td>0.077</td>
<td>0.080</td>
<td>0.096</td>
</tr>
<tr>
<td>FCI - He</td>
<td>JB</td>
<td>0.182</td>
<td>0.157</td>
<td>0.181</td>
<td>0.191</td>
</tr>
<tr>
<td>CH4 - C2H4</td>
<td>JB</td>
<td>0.50</td>
<td>0.67</td>
<td>0.67</td>
<td>0.75</td>
</tr>
<tr>
<td>CF4 - CF4</td>
<td>JB</td>
<td>0.78</td>
<td>0.70</td>
<td>0.76</td>
<td>0.82</td>
</tr>
<tr>
<td>SiH4 - CH4</td>
<td>JB</td>
<td>0.81</td>
<td>0.88</td>
<td>0.90</td>
<td>1.13</td>
</tr>
<tr>
<td>CO2 - CO2</td>
<td>JB</td>
<td>1.37</td>
<td>1.15</td>
<td>1.19</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Complex</th>
<th>Set&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reference&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Numerical&lt;sup&gt;c&lt;/sup&gt; post-LDA</th>
<th>aTZ&lt;sup&gt;d&lt;/sup&gt; post-GGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCS - OCS</td>
<td>JB</td>
<td>1.40</td>
<td>1.38</td>
<td>1.35 1.19</td>
</tr>
<tr>
<td>C10H8 - C10H8 parallel</td>
<td>JB</td>
<td>3.78</td>
<td>4.50</td>
<td>4.55 3.98</td>
</tr>
<tr>
<td>C10H8 - C10H8 parallel crossed</td>
<td>JB</td>
<td>5.28</td>
<td>5.85</td>
<td>5.92 5.28</td>
</tr>
<tr>
<td>C10H8 - C10H8 T-shaped</td>
<td>JB</td>
<td>4.34</td>
<td>4.46</td>
<td>4.62 4.68</td>
</tr>
<tr>
<td>C10H8 - C10H8 T-shaped crossed</td>
<td>JB</td>
<td>3.09</td>
<td>3.50</td>
<td>3.59 3.63</td>
</tr>
<tr>
<td>CH4 - NH3</td>
<td>JB</td>
<td>0.73</td>
<td>0.97</td>
<td>0.96 1.12</td>
</tr>
<tr>
<td>SiH4 - HF</td>
<td>JB</td>
<td>0.73</td>
<td>0.62</td>
<td>0.62 0.91</td>
</tr>
<tr>
<td>CH4 - HF</td>
<td>JB</td>
<td>1.65</td>
<td>1.76</td>
<td>1.79 1.94</td>
</tr>
<tr>
<td>C2H4 - HF</td>
<td>JB</td>
<td>4.47</td>
<td>5.16</td>
<td>5.21 5.27</td>
</tr>
<tr>
<td>CH3F - CH3F</td>
<td>JB</td>
<td>2.33</td>
<td>2.16</td>
<td>2.16 2.17</td>
</tr>
<tr>
<td>H2CO - H2CO</td>
<td>JB</td>
<td>3.37</td>
<td>2.99</td>
<td>3.00 3.09</td>
</tr>
<tr>
<td>CH3CN - CH3CN</td>
<td>JB</td>
<td>6.16</td>
<td>6.13</td>
<td>6.18 6.20</td>
</tr>
<tr>
<td>HCN - HF</td>
<td>JB</td>
<td>7.30</td>
<td>7.71</td>
<td>7.83 7.91</td>
</tr>
<tr>
<td>HF - HF</td>
<td>ZT</td>
<td>4.57</td>
<td>4.91</td>
<td>4.86 4.90</td>
</tr>
<tr>
<td>NH3 - H2O</td>
<td>ZT</td>
<td>6.41</td>
<td>7.12</td>
<td>7.17 7.45</td>
</tr>
<tr>
<td>H2S - H2S</td>
<td>ZT</td>
<td>1.66</td>
<td>2.02</td>
<td>2.25 2.35</td>
</tr>
<tr>
<td>HCl - HCl</td>
<td>ZT</td>
<td>2.01</td>
<td>2.43</td>
<td>2.45 2.50</td>
</tr>
<tr>
<td>H2S - HCl</td>
<td>ZT</td>
<td>3.35</td>
<td>4.43</td>
<td>4.53 4.76</td>
</tr>
<tr>
<td>CH3Cl - HCl</td>
<td>ZT</td>
<td>3.55</td>
<td>4.06</td>
<td>4.08 4.16</td>
</tr>
<tr>
<td>HCN - CH3SH</td>
<td>ZT</td>
<td>3.59</td>
<td>4.20</td>
<td>4.21 4.44</td>
</tr>
<tr>
<td>CH3SH - HCl</td>
<td>ZT</td>
<td>4.16</td>
<td>6.14</td>
<td>6.22 6.59</td>
</tr>
<tr>
<td>CH4 - Ne</td>
<td>ZT</td>
<td>0.22</td>
<td>0.14</td>
<td>0.17 0.24</td>
</tr>
<tr>
<td>C6H6 - Ne</td>
<td>ZT</td>
<td>0.47</td>
<td>0.35</td>
<td>0.35 0.37</td>
</tr>
<tr>
<td>C2H2 - C2H2</td>
<td>ZT</td>
<td>1.34</td>
<td>1.64</td>
<td>1.49 1.57</td>
</tr>
<tr>
<td>C6H6 - C6H6 parallel</td>
<td>ZT</td>
<td>1.81</td>
<td>1.49</td>
<td>1.50 0.56</td>
</tr>
</tbody>
</table>

| MPD (%)                           |                 |                        |                               | 2.1 4.1                |
| MAD (kcal/mol)                    | 0.33            | 0.34                   | 0.39                          |
| MAPD (%)                          | 12.60           | 11.50                  | 16.60                         |
| MaxAPD (%)                        | 49.5            | 69.3                   |                               |

Table 5.2: Binding Energies of Intermolecular Complexes (kcal/mol)

<sup>a</sup>Tang-Toennies [210] (TT), Johnson-Becke [56] (JB), S22 [212], Zhao-Truhlar [213, 214] (ZT).

<sup>b</sup>Reference binding energy from respective data set.

<sup>c</sup>Numerical post-LDA binding energy from Ref. [209].

<sup>d</sup>Present aug-cc-pVTZ post-GGA binding energy.

<sup>e</sup>Present aug-cc-pVDZ post-GGA binding energy.
Chapter 6

Basis Set Effects In The HF And DFT Descriptions Of Rare-Gas Interactions

6.1 Introduction

The accurate description of noncovalent interactions is an important aspect of computational quantum chemistry with special relevance in biochemistry [217], condensed matter physics, and materials science [45]. In ab initio wavefunction theory, it is achieved with post-HF correlation methods such as perturbation theory or coupled cluster theory [218]. The CCSD(T) method, in particular, has emerged as the benchmark method for highly accurate noncovalent interaction energies [82, 217, 219–222].

Correlated ab initio methods require the use of extended basis sets including multiple valence, polarization and diffuse functions, as the electron-electron cusp [223] must be constructed by the basis functions. In addition to basis-set requirements, the inherent computational complexity and expense of correlated ab initio methods limits their applicability to small molecular systems mainly outside the realms of biomolecular or materials science [217, 219, 222].

Density-functional theory is an attractive and widely used alternative to correlated ab initio methods, as it incorporates electron correlation effects at the cost of a SCF calculation. Furthermore, the electron-electron cusp is modeled by the correlation functional, leading to relatively minor basis-set requirements for the DFT calculation of molecular geometries, properties, and energetics. Accordingly, DFT calculations with small basis sets such as 6-31G(d) are highly prevalent in the computational chemistry literature [5].

However, noncovalent interactions (NCIs) have long been a challenge for DFT as semi-local XC functionals lack a proper account of dispersion [7, 9]. Recent advances in semi-empirical corrections and DFT-based dispersion models have addressed this challenge, producing highly accurate density functional approximations (DFAs) for noncovalent interactions with a much lower computational cost than correlated ab initio methods [12, 13].
Compared to the effort in development and benchmarking of DFAs for NCIs, the related issue of basis-set requirements of such DFT calculations has received much less attention. Compared to molecular systems, NCIs place greater demands on basis-set quality for numerous reasons:

- the relatively small noncovalent interaction energies (compared to covalent bonding) necessitate a higher absolute accuracy than the typical "chemical accuracy" of 1 kcal/mol
- basis-set superposition error (BSSE) can significantly impact NCI energies
- the long-ranged nature of NCIs requires diffuse functions to properly account for the overlap between spatially separated fragments
- where induction and dispersion are relevant, high flexibility in the valence region and extended polarization functions are needed to properly describe polarizability and related molecular properties
- accurate electrostatic potentials are needed to account for the electrostatic part of NCIs

Errors due to basis set incompleteness and BSSE can be eliminated by performing calculations in the complete basis set (CBS) limit, e.g. using Numol. This allows us to assess the accuracy inherent in DFAs without additional complications due to basis set errors. In previous work [189], we benchmarked the ability of DFAs for exchange to reproduce HF interaction (repulsion) energies in rare-gas diatomics using the fully numerical, basis-set free method of Becke & Dickson [64–66, 68]. With this information, we constructed a nonempirical DFA incorporating the DFT-based dispersion model of Becke & Johnson, and benchmarked the accuracy of this method for a wide range of NCI energies [209].

While the basis-set free approach of Becke & Dickson has been useful for developing widely used DFAs [25, 224, 225], it is not a practical quantum chemical method. The calculations in Refs [189, 209] were performed non-self-consistently using HF or LDA orbitals, but a fully self-consistent scheme is needed in order to compute forces, optimize geometries, and perform calculations of vibrational frequencies and other molecular properties.

Accordingly, we abandoned the fully-numerical approach and tested the DFA of Ref [189] with the standard Gaussian basis sets aug-cc-pVDZ and aug-cc-pVTZ
in Ref. [226]. The XDM dispersion model of Becke & Johnson contains two empirical parameters in the damping function, which were optimized individually for each basis set. The deviations between the optimized damping parameters indicate a non-negligible basis set dependence. In this scheme, basis set errors are partially compensated by the damping parameter fit. A side effect is that basis set errors become conflated with errors intrinsic to the DFA, complicating the assessment.

Similar schemes are used by other developers of DFT-based methods for NCIs. The DFT-D [227], DFT-D2 [228] & DFT-D3 [229] schemes of Grimme and coworkers are parametrized for the TZV(2d,2p) [73,227], TZVPP [73] & def2-QZVP [71,74] basis sets, respectively, which are claimed to be sufficiently close to the CBS limit that remaining basis-set errors are negligible [227]. The VV10 method of Vydrov and van Voorhis [48] is parametrized for the aug-cc-pVTZ [76,78,79,81] basis set with counterpoise (CP) [211] correction. Although it is reasonable to assume that these basis sets are close enough to the CBS limit to not introduce significant basis set artifacts, this conjecture has not been studied in much detail [227,228]. Furthermore, the accuracy of these parametrizations with other (specifically smaller) basis sets has not been calibrated extensively. Experience has shown that users are likely to combine any given DFA with their basis set of choice, whether this combination has been calibrated or not.

Kruse and Grimme recently studied the performance of the DFT-D3 and VV10 methods for NCIs with a limited number of small basis sets (MINIS, SV, 6-31G*, def2-SVP, def2-TZVP) in the context of their empirical "geometrical CP" scheme [230]. They found that after correction for BSSE, small basis sets in conjunction with both the DFT-D3 and VV10 methods reproduce NCI energies from the S22 and S66 test sets with MADs of 0.6-1.5 kcal/mol (1.6-3.5 kcal/mol without correction), whereas the def2-TZVP basis set yields MADs around 0.3 kcal/mol [after correction for BSSE]. While this study demonstrates the large effect of BSSE on NCI energies with small basis sets, it is limited to a few chosen basis sets and does not address what basis sets are accurate enough for NCIs without additional correction for BSSE. This question is highly relevant for users, as the standard counterpoise correction for BSSE has numerous drawbacks (computational expense, not applicable to intramolecular NCIs) and alternative schemes such as the newly proposed gCP method are not widely available.

The benchmarking and basis-set studies of Grimme et al and others employed the widely used S22 [212], S66 [231], and GMTKN [232,233] test sets. While these
sets cover a wide variety of NCIs, they neglect fundamental van der Waals systems such as rare-gas dimers. The extremely small binding energies in rare-gas systems arise purely from dispersion, hence providing a fundamental and stringent test for the description of dispersion-dominated NCIs by DFAs.

We have previously benchmarked semilocal DFAs for rare-gas dimers using \textsc{numol} [189], and now turn our attention to basis sets. In contrast to previous DFAs for NCIs [48,226,229], which compensate for basis set errors through the empirical parametrization of the dispersion damping function, we take a more systematic approach by parametrizing our method in the CBS limit, thus separating basis set effects from the intrinsic accuracy of the DFA. Our first task, therefore, will be to determine how to achieve the CBS limit using standard Gaussian basis sets.

We employ a wide variety of standard Gaussian basis sets from the literature:

1. The 6-31G(d), 6-311G(2df,p), and 6-311G(3df,3pd) and the diffuse-augmented 6-31+G(d), 6-311+G(2df,p), and 6-311++G(3df,3pd) basis sets of Pople and coworkers [69]

2. The (augmented) correlation-consistent (aug)-cc-p\(VnZ\) (n=D,T,Q,5) basis sets of Dunning and coworkers [76,78,79,81]

3. The “minimally augmented” min-aug-cc-p\(VnZ\) (n=D,T,Q) basis sets of Truhlar and coworkers [84] (diffuse functions truncated to s and p functions, no diffuse functions on H and He)

4. The def2-SVP, def2-TZVP, and def2-QZVP basis sets of Weigend and Ahlrichs [71]

5. The augmented def2-SVPD, def2-TZVPD, and def2-QZVPD basis sets of Rapporto and Furche [75] optimized for response properties

6. The (augmented) polarization-consistent aug-pc-\(\eta\) (n=1,2,3,4) basis sets of Jensen and coworkers [85–90]

A list of basis sets together with the abbreviations used in this Chapter is given in Table 6.1. Table 6.2 provides details on the size and construction of the basis sets, listing both the set of diffuse functions and the total set of contracted basis functions in standard basis set notation.

This Chapter is organized as follows: Details on the computational procedure are given in Section 6.2. In Section 6.3, we test the ability of a wide range of
<table>
<thead>
<tr>
<th>Basis Set</th>
<th>Abbreviation</th>
<th>Availability</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-31G(d)</td>
<td>31(d)</td>
<td>H-Kr</td>
<td>[69]</td>
</tr>
<tr>
<td>6-311G(2df,p)</td>
<td>311(2dfp)</td>
<td>H-Kr</td>
<td>[69]</td>
</tr>
<tr>
<td>6-31G(3df,3pd)</td>
<td>311(3df3pd )</td>
<td>H-Kr</td>
<td>[69]</td>
</tr>
<tr>
<td>6-31+G(d)</td>
<td>31+(d)</td>
<td>H-Kr</td>
<td>[69]</td>
</tr>
<tr>
<td>6-311+G(2df,p)</td>
<td>311+(2dfp)</td>
<td>H-Kr</td>
<td>[69]</td>
</tr>
<tr>
<td>6-311++G(3df,3pd)</td>
<td>311++(3df3pd)</td>
<td>H-Kr</td>
<td>[69]</td>
</tr>
<tr>
<td>cc-pVDZ</td>
<td>pVDZ</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>pVTZ</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>pVQZ</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>cc-pV5Z</td>
<td>pV5Z</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>min-aug-cc-pVDZ</td>
<td>maug-pVDZ</td>
<td>H-Ar</td>
<td>[84]</td>
</tr>
<tr>
<td>min-aug-cc-pVTZ</td>
<td>maug-pVTZ</td>
<td>H-Ar</td>
<td>[84]</td>
</tr>
<tr>
<td>min-aug-cc-pVQZ</td>
<td>maug-pVQZ</td>
<td>H-Ar</td>
<td>[84]</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>aug-pVDZ</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>aug-pVTZ</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td>aug-pVQZ</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>aug-cc-pV5Z</td>
<td>aug-pV5Z</td>
<td>H-Kr</td>
<td>[76–81]</td>
</tr>
<tr>
<td>def2-SVP</td>
<td>SVP</td>
<td>H-Rn</td>
<td>[71]</td>
</tr>
<tr>
<td>def2-TZVP</td>
<td>TZVP</td>
<td>H-Rn</td>
<td>[71]</td>
</tr>
<tr>
<td>def2-QZVP</td>
<td>QZVP</td>
<td>H-Rn</td>
<td>[71]</td>
</tr>
<tr>
<td>def2-SVPD</td>
<td>SVPD</td>
<td>H-Rn</td>
<td>[75]</td>
</tr>
<tr>
<td>def2-TZVPD</td>
<td>TZVPD</td>
<td>H-Rn</td>
<td>[75]</td>
</tr>
<tr>
<td>def2-QZVPD</td>
<td>QZVPD</td>
<td>H-Rn</td>
<td>[75]</td>
</tr>
<tr>
<td>pc-1</td>
<td>pc-1</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>pc-2</td>
<td>pc-2</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>pc-3</td>
<td>pc-3</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>pc-4</td>
<td>pc-4</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>aug-pc-1</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>aug-pc-2</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>aug-pc-3</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>aug-pc-4</td>
<td>H-Ar</td>
<td>[85–90]</td>
</tr>
</tbody>
</table>

Table 6.1: Basis sets and abbreviated names used in this Chapter
Table 6.2: Basis sets used in this Chapter: names and types, cardinality, diffuse functions, and contracted basis functions for hydrogen and first-row elements. Omitted basis functions are indicated with curly brackets.
basis sets to reproduce numerical Hartree-Fock repulsion energies of rare-gas diatomics at their equilibrium separation, using HF theory and also the PW86 exchange GGA. The effect of BSSE is studied in detail through the comparison of counterpoise-corrected and uncorrected repulsion energies. In Section 6.4, the empirical parameters in the damping function of the PW86PBE-XDM DFA (Chapters 3, 4, and 5) are fitted to reproduce reference binding energies of rare-gas diatomics, using a basis set close to the CBS limit. Finally, the basis-set dependence of rare-gas binding energies is analyzed in Section 6.5.

6.2 Computational Details

Calculations of rare-gas interactions using Hartree-Fock and density-functional theory were performed self-consistently with Gaussian 09 [95], using the keywords IOp(3/74=800) for PW86 and IOp(3/74=809) for PW86PBE. Pople-type orbital basis sets [69] and the W06 density fitting basis set [234] were used as implemented in Gaussian 09. All other basis sets were obtained from the EMSL basis set exchange [96, 97]. The “ultrafine” numerical integration grid (99 radial shells, 590 angular points) was used throughout as implemented in Gaussian 09.

The XDM dispersion energy [50–56] was computed non-self-consistently using an interface program called “POSTG” [215]. The POSTG program uses Gaussian 09 wfn files (containing the Kohn-Sham MOs) to calculate XDM dispersion energies, but is limited to basis functions up to $f$ angular momentum. Therefore, basis functions of $g$ and higher angular momentum were discarded by using the keyword IOp(3/86=4) in Gaussian 09. This only affects quadruple-zeta and larger basis sets, and is expected to have a negligible influence on binding energies since DFT is known to be relatively insensitive to the presence of higher angular momentum polarization functions in the basis set.

The POSTG program is currently not equipped to handle the “ghost” basis functions used in the counterpoise scheme [211]. CP-corrected PW86PBE-XDM binding energies are thus obtained by adding the XDM dispersion energy calculated with the POSTG program (without CP correction, using PW86PBE orbitals) to the CP-corrected PW86PBE binding energy calculated with Gaussian 09. As the XDM dispersion energy is very insensitive to basis set effects including BSSE, this scheme is not expected to introduce a significant error.

Rare-gas interaction (i.e., repulsion and binding) energies are given in microhartzees ($1 \mu E_h = 4.3597 \times 10^{-24} \text{ J} = 0.00062751 \text{ kcal/mol}$), dispersion coefficients
\( C_n \) are given in atomic units (\( E_h a_0^n = e^2 a_0^{n-1}/4\pi\epsilon_0 \)), and internuclear separations are given in angstrom (1 Å = 10^{-10} \text{ m}).

### 6.3 Basis-Set Effects On Rare-Gas Repulsion Energies

The performance of DFAs for NCIs is commonly benchmarked [40, 46–48, 195, 200, 201, 232, 233, 235–255] against databases of reference NCI energies [212, 213, 231, 256] computed using highly accurate wavefunction methods. In this approach, the effect of the employed basis set is difficult to separate from the intrinsic accuracy of the DFA so that relatively large (triple/quadruple zeta) basis sets and/or counterpoise corrections are commonly used to minimize the errors due to the basis set. This is problematic for two reasons:

1. The (semi-)empirical dispersion damping function (or equivalent damping mechanism) may compensate for basis set effects unless parametrized strictly in the CBS limit

2. The underlying (semilocal) exchange(-correlation) functional has a large effect on calculated NCI energies, and is difficult to separate from the inherent accuracy of the dispersion correction

Both effects contribute to the deviation in computed NCI energies in addition to basis set errors, which may be exacerbated or cancelled out. To study the basis-set-dependence of DFT-derived NCI energies, it is advantageous to use an alternative reference independent of these influences.

Rare-gas interactions can be cleanly decomposed into a repulsive contribution (“exchange repulsion”) and an attractive dispersion contribution. As shown later in detail for the XDM model, density-functional models for the dispersion contribution have a negligible dependence on the basis set, or are completely basis-set-independent as in the case of empirical (-D) dispersion corrections. The basis-set dependence of DFT-based NCI energies arises chiefly from the underlying exchange-correlation functional. The exchange repulsion contribution can be calculated exactly using Hartree-Fock theory in the CBS limit.

Ideally, we would like to accurately reproduce Hartree-Fock repulsion energies using semilocal DFAs for exchange, but in practice the deviations can be enormous [99, 189, 191]. Recent developments have led to empirical “dispersionless”
Table 6.3: Equilibrium separations $r^{eq}$ (Å), reference binding energies $B E^{ref}$ ($\mu E_h$), and Hartree-Fock binding energies $B E^{HF}$ ($\mu E_h$) of rare-gas diatomics.

<table>
<thead>
<tr>
<th></th>
<th>$r^{eq}$ [210]</th>
<th>$B E^{ref}$ [210]</th>
<th>$B E^{HF}$ [189]</th>
</tr>
</thead>
<tbody>
<tr>
<td>He$_2$</td>
<td>2.97</td>
<td>34.8</td>
<td>-28</td>
</tr>
<tr>
<td>He-Ne</td>
<td>3.05</td>
<td>66</td>
<td>-47</td>
</tr>
<tr>
<td>He-Ar</td>
<td>3.50</td>
<td>93.7</td>
<td>-85</td>
</tr>
<tr>
<td>Ne$_2$</td>
<td>3.09</td>
<td>134</td>
<td>-96</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>3.48</td>
<td>211</td>
<td>-190</td>
</tr>
<tr>
<td>Ar$_2$</td>
<td>3.76</td>
<td>454</td>
<td>-435</td>
</tr>
<tr>
<td>reference</td>
<td>He₂</td>
<td>He-Ne</td>
<td>He-Ar</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>31(d)</td>
<td>3</td>
<td>-17</td>
<td>18</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>8</td>
<td>-33</td>
<td>50</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>-15</td>
<td>-310</td>
<td>30</td>
</tr>
<tr>
<td>31+(d)</td>
<td>3</td>
<td>-80</td>
<td>-97</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>8</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>16</td>
<td>-65</td>
<td>63</td>
</tr>
<tr>
<td>pVDZ</td>
<td>2</td>
<td>-11</td>
<td>18</td>
</tr>
<tr>
<td>pVTZ</td>
<td>20</td>
<td>-13</td>
<td>58</td>
</tr>
<tr>
<td>pVQZ</td>
<td>25</td>
<td>18</td>
<td>77</td>
</tr>
<tr>
<td>pV5Z</td>
<td>27</td>
<td>43</td>
<td>81</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>2</td>
<td>-87</td>
<td>-92</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>20</td>
<td>36</td>
<td>66</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>25</td>
<td>39</td>
<td>77</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>14</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>26</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>26</td>
<td>38</td>
<td>79</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>27</td>
<td>44</td>
<td>81</td>
</tr>
<tr>
<td>SVP</td>
<td>0</td>
<td>-13</td>
<td>-12</td>
</tr>
<tr>
<td>TZVP</td>
<td>10</td>
<td>14</td>
<td>46</td>
</tr>
<tr>
<td>QZVP</td>
<td>27</td>
<td>45</td>
<td>80</td>
</tr>
<tr>
<td>SVPD</td>
<td>-96</td>
<td>-152</td>
<td>-188</td>
</tr>
<tr>
<td>TZVPD</td>
<td>-13</td>
<td>-5</td>
<td>14</td>
</tr>
<tr>
<td>QZVPD</td>
<td>27</td>
<td>44</td>
<td>80</td>
</tr>
<tr>
<td>pc-1</td>
<td>12</td>
<td>-3</td>
<td>48</td>
</tr>
<tr>
<td>pc-2</td>
<td>26</td>
<td>37</td>
<td>71</td>
</tr>
<tr>
<td>pc-3</td>
<td>28</td>
<td>46</td>
<td>82</td>
</tr>
<tr>
<td>pc-4</td>
<td>28</td>
<td>46</td>
<td>82</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>-37</td>
<td>-84</td>
<td>-10</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>9</td>
<td>-29</td>
<td>33</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>27</td>
<td>43</td>
<td>79</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>28</td>
<td>46</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 6.4: HF repulsion energies of rare-gas diatomics ($\mu E_h$)
(i.e., deviations are always negative). Interaction energies of the wrong (negative) sign are found in multiple cases, corresponding to an artificial attraction caused by the basis set.

Figure 6.1 shows the percent deviation for each system as a function of basis set. Data points between 0 and -100% indicate underestimated repulsion, while deviations below -100% (i.e., more negative) indicate artifactually attractive interaction energies.

Relative repulsion energy deviations for many common basis sets are enormous (> 100%), due in part to the very small absolute repulsion energy values (on the order of microhartrees). On the other hand, the systematic families of basis sets (i.e., all basis sets except Pople-style ones) show decreasing deviations with increasing basis set size, and deviations are found to be converged for the largest basis set of each type. The effectiveness of convergence to the CBS limit, however, varies greatly among different basis set families:

1. The standard cc basis sets are converged to the CBS limit only at the pV5Z level. Neon-containing systems are the most problematic for these basis sets, and the deviations for He-Ne and Ne-Ne surprisingly increase when going from pVDZ to pVTZ, while all other deviations decrease with increasing basis set size.

2. Adding diffuse functions to the cc basis sets greatly reduces deviations to \( \leq 50\% \), most effectively for aug-pVDZ and aug-pVTZ. Despite a comparable composition in terms of (contracted) basis functions (c.f. Table 6.2), the aug-pVQZ basis set is not as accurate as unaugmented pV5Z, indicating BSSE caused by diffuse functions (vide supra).

3. Truhlar’s minimally augmented cc basis sets are much more accurate than unaugmented pVnZ, and slightly more accurate than fully augmented aug-pVnZ (despite being of smaller size, having less diffuse functions). However, large deviations are found for helium-containing systems with maug-pVDZ. While none of the maug basis sets includes diffuse functions on helium, this alone cannot explain the large deviations with maug-pVDZ since both maug-pVTZ and maug-pVQZ are highly accurate for helium-containing systems. In fact, maug-pVTZ is more accurate than the comparable pVQZ and aug-pVTZ basis sets, and maug-pVQZ is slightly more accurate than aug-pVQZ despite the lack of diffuse functions on helium.
Figure 6.1: HF repulsion energy % deviations of rare-gas diatomics
4. Ahlrichs’ SVP and TZVP basis sets have deviations comparable to pVDZ and pVTZ, respectively. QZVP is as accurate as pV5Z and aug-pV5Z but of much smaller size (c.f. Table 6.2).

5. Adding diffuse functions to Ahlrichs’ basis sets greatly increases deviations for helium- and neon-containing systems while reducing deviations for Ar2. This surprising result is further investigated below. Nonetheless, QZVPD is as accurate as QZVP and more accurate than (maug-/aug-)pVQZ.

6. Jensen’s pc-1 basis set performs comparable to pVDZ and SVP. The unaugmented pc sets converge much faster to the CBS limit than the cc basis sets. In particular, the triple-zeta pc-2 basis set is very effective, surpassing all other triple zeta basis sets (including augmented ones) in accuracy. The CBS limit is reached with the quintuple-zeta pc-3 basis set, which is also more accurate than the comparable pV5Z.

7. Without exception, adding diffuse functions to Jensen’s pc basis sets is highly detrimental to the accuracy of rare-gas repulsion energies. aug-pc-1 fails similarly as SVPD, and aug-pc-2 performs worse than any triple-zeta basis set mentioned so far. Nonetheless, the CBS limit is reached with aug-pc-3 and aug-pc-4, although they are less accurate than the unaugmented pc-3 and pc-4 basis sets, respectively.

The behaviour of Pople-style basis sets for HF repulsion energies is rather unsystematic and unsatisfactory. The double-zeta 31(d) basis has (large) errors comparable to the other double zeta basis sets. Adding diffuse functions on heavy atoms to give 31+(d) worsens deviations on average, and for He-Ne and He-Ar in particular. The triple-zeta 311(2dfp) basis improves results over 31(d) but is worse than other triple-zeta basis sets. The larger 311(3df3pd) basis is even more disastrous with a mean absolute deviation of 236% and a maximum deviation of -760% (He-Ne). Adding diffuse functions to give 311+(2dfp) and 311++(3df3pd) reduces deviations, but both basis sets are still very inaccurate and not much better than the simple 31(d) basis.

Summarizing the data in Table 6.4 and Figure 6.1, we find that rare-gas repulsion energies at the HF level are strongly basis-set dependent. Double-zeta basis sets are generally inadequate, but aug-pVDZ performs exceptionally well with a MAPD of 33%. The accuracy of (augmented) triple zeta basis sets varies widely: pc-2 and maug-pVTZ are very accurate (MAPD < 20%), aug-pVTZ is
slightly worse, and all other TZ basis sets are inadequate due to MAPDs > 50%. Quadruple zeta and larger basis sets give accurate repulsion energies, approaching the CBS limit. Only pVQZ cannot be recommended, as its MAPD of 31% is similar to the much smaller aug-pVDZ and worse than well-performing TZ basis sets.

The accuracy of cc basis sets is greatly improved by the addition of diffuse functions, using either the original full augmentation scheme or the more economical “minimally augmented” scheme of Truhlar and coworkers. Conversely, diffuse functions are detrimental for Ahlrichs-type and pc-type basis sets.

### 6.3.2 Counterpoise Correction And Basis-Set Superposition Error

The basis set errors studied thus far in general arise from two distinct effects. **Basis set incompleteness error** (BSIE) exists due to the deviation of basis sets from the CBS limit. **Basis set superposition error** (BSSE) arises in interaction energy calculations due to the difference between the dimer basis set and the basis set describing the isolated monomers. In He-Ne, for example, the helium-centered basis functions are supplemented by (unoccupied) basis functions centered on neon (and vice versa), leading to an energy lowering in the diatomic molecule due to basis set superposition. In the calculation of the isolated He and Ne atoms, no such artificial energy lowering arises. This imbalance between the basis sets of the monomers and the dimer leads to BSSE, which typically appears as overbinding in the interaction energy.

For repulsive interactions, BSSE leads to underestimated repulsion or even artificial attraction, as is seen for all basis sets in Table 6.4. To quantify the effect of BSSE on rare-gas interaction energies, we now employ the counterpoise (CP) correction scheme. In this scheme, the monomer energies are calculated using the full dimer basis set [211], therefore eliminating the basis set imbalance and BSSE.

Counterpoise-corrected HF repulsion energies for the six rare-gas diatomics are given in Table 6.5. Figure 6.2 shows the percent deviation for each system as a function of basis set. Data points above and below 0% indicate overestimated and underestimated repulsion energies, respectively.

Without exception, CP correction improves the accuracy of HF repulsion energies, in many cases by an order of magnitude or more. The artificially attractive
<table>
<thead>
<tr>
<th></th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31(d)</td>
<td>25</td>
<td>32</td>
<td>79</td>
<td>28</td>
<td>171</td>
<td>428</td>
<td>-21.9</td>
<td>21.9</td>
<td>-1.7</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>27</td>
<td>45</td>
<td>79</td>
<td>66</td>
<td>188</td>
<td>413</td>
<td>-8.7</td>
<td>8.7</td>
<td>-0.8</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>25</td>
<td>39</td>
<td>77</td>
<td>61</td>
<td>177</td>
<td>416</td>
<td>-14.2</td>
<td>14.2</td>
<td>-4.4</td>
</tr>
<tr>
<td>31+(d)</td>
<td>25</td>
<td>48</td>
<td>90</td>
<td>97</td>
<td>193</td>
<td>431</td>
<td>0.1</td>
<td>3.4</td>
<td>5.9</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>27</td>
<td>48</td>
<td>81</td>
<td>102</td>
<td>188</td>
<td>420</td>
<td>-0.7</td>
<td>3.5</td>
<td>5.8</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>28</td>
<td>47</td>
<td>81</td>
<td>97</td>
<td>184</td>
<td>424</td>
<td>-1.6</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td>pVDZ</td>
<td>25</td>
<td>35</td>
<td>76</td>
<td>44</td>
<td>165</td>
<td>418</td>
<td>-19.7</td>
<td>19.7</td>
<td>-4.0</td>
</tr>
<tr>
<td>pVTZ</td>
<td>27</td>
<td>48</td>
<td>80</td>
<td>84</td>
<td>182</td>
<td>411</td>
<td>-4.8</td>
<td>5.4</td>
<td>1.9</td>
</tr>
<tr>
<td>pVQZ</td>
<td>27</td>
<td>46</td>
<td>81</td>
<td>90</td>
<td>183</td>
<td>426</td>
<td>-3.6</td>
<td>3.6</td>
<td>-1.7</td>
</tr>
<tr>
<td>pV5Z</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>93</td>
<td>185</td>
<td>429</td>
<td>-2.4</td>
<td>2.4</td>
<td>-1.4</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>25</td>
<td>51</td>
<td>91</td>
<td>101</td>
<td>195</td>
<td>432</td>
<td>1.9</td>
<td>5.6</td>
<td>8.6</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>27</td>
<td>48</td>
<td>81</td>
<td>98</td>
<td>185</td>
<td>417</td>
<td>-1.6</td>
<td>2.9</td>
<td>2.3</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>27</td>
<td>46</td>
<td>81</td>
<td>92</td>
<td>184</td>
<td>429</td>
<td>-3.0</td>
<td>3.0</td>
<td>-1.4</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>29</td>
<td>48</td>
<td>83</td>
<td>98</td>
<td>188</td>
<td>419</td>
<td>-0.1</td>
<td>2.2</td>
<td>3.1</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>94</td>
<td>183</td>
<td>424</td>
<td>-2.6</td>
<td>2.6</td>
<td>-1.3</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>93</td>
<td>185</td>
<td>430</td>
<td>-2.2</td>
<td>2.2</td>
<td>-0.7</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>94</td>
<td>186</td>
<td>430</td>
<td>-1.7</td>
<td>1.7</td>
<td>-0.6</td>
</tr>
<tr>
<td>SVP</td>
<td>25</td>
<td>35</td>
<td>57</td>
<td>46</td>
<td>108</td>
<td>344</td>
<td>-31.2</td>
<td>31.2</td>
<td>-11.6</td>
</tr>
<tr>
<td>TZVP</td>
<td>28</td>
<td>46</td>
<td>80</td>
<td>91</td>
<td>179</td>
<td>408</td>
<td>-4.4</td>
<td>4.4</td>
<td>-0.7</td>
</tr>
<tr>
<td>QZVP</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>94</td>
<td>186</td>
<td>429</td>
<td>-2.2</td>
<td>2.2</td>
<td>-1.4</td>
</tr>
<tr>
<td>SVPD</td>
<td>20</td>
<td>47</td>
<td>83</td>
<td>90</td>
<td>184</td>
<td>440</td>
<td>-6.5</td>
<td>7.0</td>
<td>1.2</td>
</tr>
<tr>
<td>TZVPD</td>
<td>25</td>
<td>46</td>
<td>82</td>
<td>94</td>
<td>185</td>
<td>420</td>
<td>-3.9</td>
<td>3.9</td>
<td>-1.3</td>
</tr>
<tr>
<td>QZVPD</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>94</td>
<td>186</td>
<td>430</td>
<td>-1.9</td>
<td>1.9</td>
<td>-1.1</td>
</tr>
<tr>
<td>pc-1</td>
<td>31</td>
<td>49</td>
<td>85</td>
<td>74</td>
<td>190</td>
<td>394</td>
<td>-2.7</td>
<td>8.2</td>
<td>11.7</td>
</tr>
<tr>
<td>pc-2</td>
<td>26</td>
<td>44</td>
<td>78</td>
<td>89</td>
<td>179</td>
<td>415</td>
<td>-6.3</td>
<td>6.3</td>
<td>-4.5</td>
</tr>
<tr>
<td>pc-3</td>
<td>28</td>
<td>47</td>
<td>83</td>
<td>95</td>
<td>187</td>
<td>433</td>
<td>-1.2</td>
<td>1.2</td>
<td>-0.3</td>
</tr>
<tr>
<td>pc-4</td>
<td>28</td>
<td>46</td>
<td>82</td>
<td>94</td>
<td>186</td>
<td>429</td>
<td>-1.7</td>
<td>1.7</td>
<td>-0.7</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>30</td>
<td>49</td>
<td>84</td>
<td>94</td>
<td>189</td>
<td>413</td>
<td>0.2</td>
<td>3.2</td>
<td>6.9</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>28</td>
<td>46</td>
<td>83</td>
<td>94</td>
<td>185</td>
<td>431</td>
<td>-1.7</td>
<td>1.7</td>
<td>-0.4</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>28</td>
<td>46</td>
<td>83</td>
<td>94</td>
<td>187</td>
<td>431</td>
<td>-1.5</td>
<td>1.5</td>
<td>-0.7</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>28</td>
<td>46</td>
<td>83</td>
<td>94</td>
<td>186</td>
<td>429</td>
<td>-1.7</td>
<td>1.7</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

Table 6.5: CP-corrected HF repulsion energies of rare-gas diatomics ($\mu E_h$)
Figure 6.2: CP-corrected HF repulsion energy % deviations of rare-gas diatomics
interaction energies caused by BSSE in Section 6.3.1 are completely eliminated after applying CP correction. Most basis sets display a residual under-repulsive tendency, but some of the smaller (DZ and TZ) basis sets become over-repulsive, most likely due to an over-correcting of BSSE in the CP scheme. Out of the 32 basis sets tested only four (31(d), 311(3df3pd), pVDZ and SVP) have a MAPD > 10%, while 21 basis sets are very accurate with a MAPD < 5%. The CBS limit is clearly reached for the basis sets aug-pV5Z, QZVPD, pc-3, pc-4, aug-pc-2, aug-pc-3 and aug-pc-4 with MAPDs < 2%. This excellent agreement on the order of microhartrees (10^{-4} kcal/mol) between the largest basis sets and the numerical HF reference repulsion energies validates our computational procedure and the accuracy of the Numol reference data.

CP-corrected and uncorrected MAPDs are shown on the same scale in Figure 6.3, demonstrating how BSSE dominates the total basis set error for all but the very largest basis sets. In extreme cases such as 311++(3df3pd), SVPD and aug-pc-1, CP-correction lowers the MAPD by almost two orders of magnitude. Even for the largest basis sets BSSE contributes significantly to the total deviation, and CP-correction halves the MAPD.

An analysis of Figures 6.2 and 6.3 by basis set family leads to the following conclusions:

1. Unaugmented Pople-type basis sets show rather large deviations for the He-Ne and Ne-Ne systems, and adding a single set of diffuse functions on non-hydrogenic atoms as in 31+(d) improves the MAPD by almost one order of magnitude.

2. Unaugmented cc basis sets now converge smoothly to the CBS limit. He-Ne and Ne-Ne are still the most difficult systems (c.f. Section 6.3.1) indicating that diffuse functions on neon are essential for accurate rare-gas interaction energies.

3. Augmentation of the cc basis sets significantly improves accuracy at the DZ and TZ levels. Minimal augmentation (maug-pVnZ) is sufficient at the TZ and QZ levels, but full augmentation (aug-pVnZ) is necessary to approach convergence with DZ basis sets. Comparison with the uncorrected data from Section 6.3.1 shows that BSSE caused by diffuse functions is effectively eliminated through CP correction, making maug-pV[T,Q]Z and aug-pV[D,T,Q,5]Z as accurate as pV5Z.
Figure 6.3: CP-corrected and uncorrected (noCP) HF repulsion energy MAPD by basis set
CP correction also removes the notably large deviations for helium-containing systems with maug-pVDZ. The better performance of uncorrected maug-pVTZ over aug-pVTZ found in Section 6.3.1 vanishes after applying CP correction, and can be attributed to fortuitous error cancellation.

4. Ahlrichs’ SVP basis set has the largest repulsion energy deviations but is still comparable to pVDZ; TZVP and QZVP are slightly more accurate than pVTZ and pVQZ, respectively. QZVP has the smallest BSSE of all QZ basis sets, validating the use of CP-uncorrected QZVP in the parametrization and benchmarking of DFT-D methods by Grimme and coworkers.

5. CP-corrected, diffuse-augmented Ahlrichs-type basis sets (SVPD, TZVPD and QZVPD) are very accurate for rare-gas repulsion energies, while without CP correction the SVPD and TZVPD basis sets performed worse than the corresponding unaugmented Ahlrichs-type basis sets. In other words, the property-optimized diffuse functions of Rappoport and Furche [75] cause significant BSSE (especially at the SVPD level) and can only be recommended for interaction energy calculations if BSSE is removed through CP correction.

6. A similar interplay between BSSE and diffuse functions is found for Jensen’s pc-type basis sets. aug-pc-1, aug-pc-2, and (to a much lesser extent) aug-pc-3 performed worse than the corresponding unaugmented pc basis sets without CP correction, but become very accurate when CP correction is applied.

7. The pc-n and aug-pc-n series (n=1-4) perform very similar to the pVnZ and aug-pVnZ (n=D,T,Q,5) series, respectively. The CBS limit for rare-gas repulsion energies is reached at the quintuple-zeta level (pV5Z/aug-pc-3) without diffuse functions and at the triple-zeta level (aug-pVTZ/aug-pc-2) with diffuse functions.

Regarding the effectiveness of diffuse functions and their effect on BSSE, we conclude that for the rare-gas interactions investigated here, cc-type basis sets always benefit from diffuse functions, whether CP correction is applied or not. In other words, adding diffuse functions to cc-type basis sets does not cause increased BSSE that would deteriorate accuracy, the only exception being maug-pVDZ for He-Ne and He-Ar. Conversely, diffuse functions cause strong BSSE with Ahlrichs- and pc-type basis sets. They are detrimental if no CP correction is applied, but improve accuracy over the corresponding unaugmented basis sets after CP correction.
Figure 6.4: CP-corrected and uncorrected (noCP) HF repulsion energy MAPD ordered by basis set size
Figure 6.4 shows the same data as Figure 6.3 with the basis sets ordered by increasing size (c.f. Table 6.2). It allows a direct comparison between basis sets of comparable size but different family. When no CP correction is applied, aug-pVDZ is the only DZ basis set that yields semi-accurate rare-gas repulsion energies. Improved accuracy over aug-pVDZ at the triple-zeta level is found for maug-pVTZ & aug-pVTZ, but the pc-2 basis set performs even better despite its smaller size and lack of diffuse functions. Quantitative accuracy is only reached at the QZ level with Ahlrichs’ QZVP basis set.

Applying CP correction improves accuracy for all basis sets, and even the worst performer (SVP) has a MAPD lower than uncorrected aug-pVDZ. On average, diffuse-augmented DZ basis sets perform better than unaugmented TZ basis sets. The CBS limit is essentially reached at the augmented TZ level, further increasing the basis set size does not significantly improve accuracy. The CP-corrected 31+(d), aug-pVDZ and aug-pc-1 basis sets give extraordinarily accurate rare-gas repulsion energies (close to the CBS limit) despite their small size.

6.3.3 Basis-Set Dependence Of DFT (PW86) Repulsion Energies

In rare-gas systems, DFAs for exchange should reproduce HF repulsion energies as closely as possible. Unfortunately, this is not the case with most common exchange DFAs. As detailed in Chapter 3, the nonempirical exchange GGA of Perdew and Wang known as PW86 most closely reproduces HF repulsion in rare-gas systems (MAPD of 10%) and small intermolecular complexes. PW86 in conjunction with various DF models for dispersion (XDM, VV10 and vdW-DF2) has also been shown to provide accurate NCI energies. However, these studies did not investigate the basis-set-dependence of PW86-based interaction energies.

A preliminary investigation of basis-set effects on PW86 interaction energies was performed in Chapter 5, where the aug-pVDZ and aug-pVTZ basis sets were tested for their ability to reproduce HF repulsion for ten rare-gas diatomics (He to Kr) without CP correction. With MAPDs of 30% and 13% for aug-pVDZ and aug-pVTZ, respectively, both basis sets were deemed accurate enough to warrant a parametrization of the XDM dispersion damping function, and were shown to give accurate NCI energies on a benchmark set of 65 intermolecular complexes.

Table 6.6 lists HF/aug-pVDZ and HF/aug-pVTZ repulsion energies in addition to the PW86 values from Table 5.1, revealing that the rather accurate PW86/aug-pVTZ repulsion energies are a result of error cancellation between the density
Table 6.6: HF and PW86 repulsion energies of ten rare-gas diatomics ($\mu$Eh)

<table>
<thead>
<tr>
<th></th>
<th>HF/CBS numerical</th>
<th>PW86/CBS aug-pV5Z CP</th>
<th>HF/CBS aug-pVTZ noCP</th>
<th>PW86 aug-pVTZ noCP</th>
<th>HF aug-pVDZ noCP</th>
<th>PW86 aug-pVDZ noCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>He-He</td>
<td>28</td>
<td>29</td>
<td>26</td>
<td>33</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>He-Ne</td>
<td>47</td>
<td>52</td>
<td>25</td>
<td>40</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>He-Ar</td>
<td>85</td>
<td>86</td>
<td>75</td>
<td>95</td>
<td>56</td>
<td>75</td>
</tr>
<tr>
<td>He-Kr</td>
<td>103</td>
<td>97</td>
<td>87</td>
<td>103</td>
<td>62</td>
<td>85</td>
</tr>
<tr>
<td>Ne-Ne</td>
<td>96</td>
<td>115</td>
<td>50</td>
<td>111</td>
<td>63</td>
<td>18</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>190</td>
<td>197</td>
<td>150</td>
<td>201</td>
<td>148</td>
<td>120</td>
</tr>
<tr>
<td>Ne-Kr</td>
<td>227</td>
<td>220</td>
<td>182</td>
<td>211</td>
<td>168</td>
<td>140</td>
</tr>
<tr>
<td>Ar-Ar</td>
<td>435</td>
<td>503</td>
<td>394</td>
<td>524</td>
<td>357</td>
<td>486</td>
</tr>
<tr>
<td>Ar-Kr</td>
<td>538</td>
<td>625</td>
<td>495</td>
<td>632</td>
<td>432</td>
<td>619</td>
</tr>
<tr>
<td>Kr-Kr</td>
<td>674</td>
<td>796</td>
<td>627</td>
<td>783</td>
<td>533</td>
<td>801</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxAPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW86</td>
<td>7.9</td>
<td>-19.4</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td>9.7</td>
<td>19.4</td>
<td>48.4</td>
</tr>
<tr>
<td></td>
<td>12.8</td>
<td>30.7</td>
<td>20.6</td>
</tr>
<tr>
<td></td>
<td>-30.7</td>
<td>30.6</td>
<td>50.9</td>
</tr>
<tr>
<td></td>
<td>-21.5</td>
<td>-80.9</td>
<td></td>
</tr>
</tbody>
</table>

functional and basis set approximations. In the CBS limit, PW86 overestimates rare-gas repulsion energies by 8% on average (MPD), but since HF/aug-pVTZ underestimates repulsion energies with a MPD of -19.4%, the resulting MPD for PW86/aug-pVTZ is only 8.5% and the MAPD is similarly reduced compared to HF/aug-pVTZ. Less obviously, PW86/aug-pVDZ repulsion energies also benefit from error cancellation resulting in a MPD of -21.5% and a MAPD of 30.6%.

These findings motivate a more systematic and detailed investigation of PW86 exchange repulsion energies in order to understand the interplay between basis set and DFA error, identify further cases of error cancellation, and to gauge the basis-set dependence of PW86 in general. Since maug- and pc-type basis sets are not available for krypton, the test set is limited to the six rare-gas diatomics of helium, neon, and argon. In order to quantify BSSE, we follow the same approach as in Sections 6.3.1 and 6.3.2, analyzing both CP-corrected and uncorrected interaction energies.

PW86 repulsion energies (no CP correction) are shown in Table 6.7 and include error statistics wrt the reference (HF) repulsion energies.

Figure 6.5 shows the percent deviation for each system as a function of basis set. Data points above and below 0% indicate overestimated and underestimated repulsion energies, respectively, while deviations below -100% (i.e., more
<table>
<thead>
<tr>
<th>Reference</th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>31(d)</td>
<td>8</td>
<td>-32</td>
<td>-18</td>
<td>-119</td>
<td>-364</td>
<td>163</td>
<td>-156</td>
<td>156</td>
<td>-62.4</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>-2</td>
<td>-92</td>
<td>15</td>
<td>-123</td>
<td>-449</td>
<td>364</td>
<td>-178</td>
<td>178</td>
<td>-16.3</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>-149</td>
<td>-890</td>
<td>-77</td>
<td>-202</td>
<td>-780</td>
<td>313</td>
<td>-611</td>
<td>611</td>
<td>-28.0</td>
</tr>
<tr>
<td>31+(d)</td>
<td>8</td>
<td>-421</td>
<td>-674</td>
<td>-62</td>
<td>90</td>
<td>541</td>
<td>-359</td>
<td>367</td>
<td>24.4</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>-2</td>
<td>-93</td>
<td>-131</td>
<td>-9</td>
<td>118</td>
<td>483</td>
<td>-132</td>
<td>136</td>
<td>11.1</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>36</td>
<td>-20</td>
<td>86</td>
<td>-32</td>
<td>73</td>
<td>472</td>
<td>-50.0</td>
<td>62.5</td>
<td>27.7</td>
</tr>
<tr>
<td>pVDZ</td>
<td>7</td>
<td>-21</td>
<td>-1</td>
<td>-51</td>
<td>197</td>
<td>195</td>
<td>-122</td>
<td>122</td>
<td>-55.1</td>
</tr>
<tr>
<td>pVTZ</td>
<td>12</td>
<td>-94</td>
<td>33</td>
<td>-178</td>
<td>-183</td>
<td>344</td>
<td>-153</td>
<td>153</td>
<td>-21.0</td>
</tr>
<tr>
<td>pVQZ</td>
<td>14</td>
<td>-58</td>
<td>56</td>
<td>-124</td>
<td>-54</td>
<td>447</td>
<td>-110</td>
<td>111</td>
<td>2.8</td>
</tr>
<tr>
<td>pV5Z</td>
<td>20</td>
<td>24</td>
<td>71</td>
<td>61</td>
<td>152</td>
<td>485</td>
<td>-23.2</td>
<td>27.0</td>
<td>11.5</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>7</td>
<td>-516</td>
<td>-658</td>
<td>25</td>
<td>139</td>
<td>558</td>
<td>-370</td>
<td>380</td>
<td>28.2</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>12</td>
<td>-1</td>
<td>18</td>
<td>149</td>
<td>225</td>
<td>540</td>
<td>-23.5</td>
<td>56.1</td>
<td>55.3</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>14</td>
<td>26</td>
<td>38</td>
<td>137</td>
<td>207</td>
<td>500</td>
<td>-13.9</td>
<td>36.1</td>
<td>42.8</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>28</td>
<td>13</td>
<td>75</td>
<td>18</td>
<td>120</td>
<td>486</td>
<td>-32.2</td>
<td>36.1</td>
<td>11.6</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>33</td>
<td>40</td>
<td>95</td>
<td>111</td>
<td>201</td>
<td>524</td>
<td>9.8</td>
<td>14.4</td>
<td>20.6</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>29</td>
<td>50</td>
<td>84</td>
<td>127</td>
<td>200</td>
<td>496</td>
<td>10.1</td>
<td>10.6</td>
<td>32.1</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>29</td>
<td>48</td>
<td>84</td>
<td>110</td>
<td>193</td>
<td>500</td>
<td>5.8</td>
<td>6.4</td>
<td>14.9</td>
</tr>
<tr>
<td>SVP</td>
<td>4</td>
<td>-24</td>
<td>-19</td>
<td>-63</td>
<td>-102</td>
<td>-156</td>
<td>-135</td>
<td>135</td>
<td>-84.8</td>
</tr>
<tr>
<td>TZVP</td>
<td>5</td>
<td>-25</td>
<td>16</td>
<td>-112</td>
<td>6</td>
<td>306</td>
<td>-110</td>
<td>110</td>
<td>-29.7</td>
</tr>
<tr>
<td>QZVP</td>
<td>22</td>
<td>36</td>
<td>73</td>
<td>83</td>
<td>170</td>
<td>486</td>
<td>-11.6</td>
<td>15.5</td>
<td>11.8</td>
</tr>
<tr>
<td>SVPD</td>
<td>-308</td>
<td>-711</td>
<td>-949</td>
<td>-169</td>
<td>-144</td>
<td>115</td>
<td>-759</td>
<td>759</td>
<td>-73.6</td>
</tr>
<tr>
<td>TZVPD</td>
<td>-138</td>
<td>-233</td>
<td>-269</td>
<td>115</td>
<td>187</td>
<td>458</td>
<td>-263</td>
<td>272</td>
<td>19.3</td>
</tr>
<tr>
<td>QZVPD</td>
<td>8</td>
<td>19</td>
<td>48</td>
<td>104</td>
<td>192</td>
<td>503</td>
<td>-24.6</td>
<td>33.0</td>
<td>17.3</td>
</tr>
<tr>
<td>pc-1</td>
<td>22</td>
<td>-16</td>
<td>47</td>
<td>-71</td>
<td>-135</td>
<td>385</td>
<td>-92.9</td>
<td>92.9</td>
<td>-11.6</td>
</tr>
<tr>
<td>pc-2</td>
<td>27</td>
<td>26</td>
<td>73</td>
<td>55</td>
<td>151</td>
<td>448</td>
<td>-20.6</td>
<td>21.6</td>
<td>2.9</td>
</tr>
<tr>
<td>pc-3</td>
<td>30</td>
<td>50</td>
<td>87</td>
<td>115</td>
<td>200</td>
<td>510</td>
<td>9.4</td>
<td>9.4</td>
<td>19.3</td>
</tr>
<tr>
<td>pc-4</td>
<td>31</td>
<td>51</td>
<td>88</td>
<td>115</td>
<td>202</td>
<td>514</td>
<td>10.8</td>
<td>10.8</td>
<td>19.7</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>38</td>
<td>-6</td>
<td>96</td>
<td>-123</td>
<td>82</td>
<td>535</td>
<td>-54.4</td>
<td>78.2</td>
<td>35.7</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>30</td>
<td>47</td>
<td>75</td>
<td>94</td>
<td>186</td>
<td>455</td>
<td>-0.7</td>
<td>4.7</td>
<td>7.4</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>29</td>
<td>47</td>
<td>87</td>
<td>109</td>
<td>195</td>
<td>511</td>
<td>7.0</td>
<td>7.0</td>
<td>17.5</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>30</td>
<td>50</td>
<td>85</td>
<td>114</td>
<td>198</td>
<td>510</td>
<td>9.1</td>
<td>9.1</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Table 6.7: PW86 repulsion energies of rare-gas diatomics ($\mu E_h$)
Figure 6.5: PW86 repulsion energy % deviations of rare-gas diatomics
negative) indicate artifactualy attractive interaction energies. Compared to the HF repulsion energies of Figure 6.1, PW86 deviations are generally larger and less systematic. More specifically, PW86 yields enormously large interaction energies of the wrong sign (i.e., attractive) for He-containing systems with certain basis sets (31+(d), 311(3df3pd), maug-pVDZ, SVPD, and TZVPD) which show up as deviations between -2000% and -400% in Figure 6.5.

To directly compare the basis-set dependence of HF and PW86 repulsion energies, mean absolute percent deviations (MAPD) for both HF and PW86 are shown in Figure 6.6. PW86 performs well (and similar to HF) for rare-gas repulsion energies only in conjunction with the aug-pVnZ (n=D,T,Q,5) and (aug-)pc-n (n=2,3,4) basis sets. Error cancellation (i.e., decrease in MAPD from HF to PW86) occurs only in very few cases (311++(3df3pd), aug-pVTZ, aug-pVQZ, aug-pc-1, and aug-pc-2), and is most significant for aug-pc-2 with an exceptionally low MAPD of 4.7%.

The CBS limit for PW86 repulsion energies cannot be deduced with certainty from the data in Table 6.7 and Figures 6.5 and 6.6, as general basis set errors, BSSE, and the intrinsic error of the PW86 GGA combine in an unsystematic way. In contrast to the systematically convergent HF repulsion deviations, the PW86 MAPD increases from pVDZ to pVTZ, from pc-3 to pc-4, and from aug-pc-2 to aug-pc-3 to aug-pc-4 despite increasing basis set size. We therefore employ the CP correction in order to eliminate BSSE and approach the CBS limit in a more systematic manner.

CP-corrected PW86 repulsion energies are shown in Table 6.8, and percent deviations by basis set are shown in Figure 6.7. The deviations in Figure 6.7 behave rather unsystematically but are confined to a much narrower range than the uncorrected data in Figure 6.5. The CBS limit is only reached at the augmented quintuple-zeta or higher level (aug-pV5Z, aug-pc-3, and aug-pc-4) with mean absolute % deviations around 10%. This remaining error characterizes the intrinsic accuracy of PW86 compared to (exact) HF repulsion and is consistent with earlier investigations (Chapters 3 and 5). In the CBS limit, PW86 overestimates repulsion energies by 10% on average, with maximum deviations up to 20% (Ne2 and Ar2).

CP correction decreases the MAPD of PW86 repulsion energies by up to an order of magnitude as shown in Figure 6.8. In general, PW86 repulsion energies are much more sensitive to BSSE than HF as evidenced by the different scales of Figures 6.3 and 6.8. With HF, CP correction always decreases the MAPD (c.f. Figure 6.3). This is not always the case with PW86. For the aug-pVnZ (n=T,Q,5),
Figure 6.6: HF and PW86 repulsion energy mean absolute percent deviations (MAPD)
<table>
<thead>
<tr>
<th>Reference</th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>31(d)</td>
<td>51</td>
<td>53</td>
<td>117</td>
<td>41</td>
<td>157</td>
<td>597</td>
<td>15.7</td>
<td>40.7</td>
<td>81.1</td>
</tr>
<tr>
<td>31(2dip)</td>
<td>54</td>
<td>77</td>
<td>116</td>
<td>109</td>
<td>205</td>
<td>541</td>
<td>39.8</td>
<td>39.8</td>
<td>93.6</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>19</td>
<td>-24</td>
<td>80</td>
<td>88</td>
<td>129</td>
<td>516</td>
<td>-35.1</td>
<td>41.3</td>
<td>18.7</td>
</tr>
<tr>
<td>31+(d)</td>
<td>51</td>
<td>27</td>
<td>109</td>
<td>80</td>
<td>194</td>
<td>598</td>
<td>14.8</td>
<td>34.7</td>
<td>81.1</td>
</tr>
<tr>
<td>311+(2dip)</td>
<td>54</td>
<td>61</td>
<td>108</td>
<td>121</td>
<td>202</td>
<td>524</td>
<td>33.9</td>
<td>33.9</td>
<td>93.6</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>43</td>
<td>48</td>
<td>104</td>
<td>106</td>
<td>181</td>
<td>522</td>
<td>17.1</td>
<td>18.7</td>
<td>52.1</td>
</tr>
<tr>
<td>pVDZ</td>
<td>50</td>
<td>60</td>
<td>119</td>
<td>69</td>
<td>187</td>
<td>617</td>
<td>26.6</td>
<td>36.7</td>
<td>79.1</td>
</tr>
<tr>
<td>pVTZ</td>
<td>48</td>
<td>86</td>
<td>119</td>
<td>152</td>
<td>241</td>
<td>556</td>
<td>51.2</td>
<td>51.2</td>
<td>83.9</td>
</tr>
<tr>
<td>pVQZ</td>
<td>36</td>
<td>76</td>
<td>93</td>
<td>151</td>
<td>224</td>
<td>511</td>
<td>32.2</td>
<td>32.2</td>
<td>62.3</td>
</tr>
<tr>
<td>pV5Z</td>
<td>32</td>
<td>58</td>
<td>90</td>
<td>123</td>
<td>208</td>
<td>513</td>
<td>16.6</td>
<td>16.6</td>
<td>27.8</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>50</td>
<td>20</td>
<td>111</td>
<td>82</td>
<td>197</td>
<td>633</td>
<td>14.5</td>
<td>38.5</td>
<td>79.1</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>48</td>
<td>66</td>
<td>115</td>
<td>156</td>
<td>230</td>
<td>550</td>
<td>42.7</td>
<td>42.7</td>
<td>70.6</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>36</td>
<td>64</td>
<td>93</td>
<td>148</td>
<td>216</td>
<td>505</td>
<td>26.4</td>
<td>26.4</td>
<td>54.7</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>37</td>
<td>40</td>
<td>106</td>
<td>77</td>
<td>185</td>
<td>577</td>
<td>8.8</td>
<td>21.1</td>
<td>32.7</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>35</td>
<td>52</td>
<td>101</td>
<td>133</td>
<td>215</td>
<td>542</td>
<td>21.9</td>
<td>21.9</td>
<td>38.2</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>30</td>
<td>58</td>
<td>88</td>
<td>141</td>
<td>214</td>
<td>503</td>
<td>18.4</td>
<td>18.4</td>
<td>47.4</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>29</td>
<td>52</td>
<td>86</td>
<td>115</td>
<td>197</td>
<td>503</td>
<td>9.0</td>
<td>9.0</td>
<td>19.7</td>
</tr>
<tr>
<td>SVP</td>
<td>49</td>
<td>60</td>
<td>89</td>
<td>70</td>
<td>141</td>
<td>445</td>
<td>9.6</td>
<td>27.1</td>
<td>75.7</td>
</tr>
<tr>
<td>TZVP</td>
<td>58</td>
<td>94</td>
<td>126</td>
<td>161</td>
<td>248</td>
<td>556</td>
<td>63.3</td>
<td>63.3</td>
<td>107</td>
</tr>
<tr>
<td>QZVP</td>
<td>34</td>
<td>60</td>
<td>92</td>
<td>125</td>
<td>208</td>
<td>514</td>
<td>19.1</td>
<td>19.1</td>
<td>30.0</td>
</tr>
<tr>
<td>SVPD</td>
<td>-9</td>
<td>2</td>
<td>20</td>
<td>60</td>
<td>135</td>
<td>472</td>
<td>-60.6</td>
<td>63.4</td>
<td>8.4</td>
</tr>
<tr>
<td>TZVPD</td>
<td>19</td>
<td>61</td>
<td>94</td>
<td>146</td>
<td>229</td>
<td>531</td>
<td>16.8</td>
<td>27.8</td>
<td>52.0</td>
</tr>
<tr>
<td>QZVPD</td>
<td>27</td>
<td>49</td>
<td>86</td>
<td>109</td>
<td>196</td>
<td>506</td>
<td>5.8</td>
<td>7.2</td>
<td>16.3</td>
</tr>
<tr>
<td>pc-1</td>
<td>68</td>
<td>90</td>
<td>149</td>
<td>120</td>
<td>239</td>
<td>651</td>
<td>68.5</td>
<td>68.5</td>
<td>142</td>
</tr>
<tr>
<td>pc-2</td>
<td>39</td>
<td>80</td>
<td>103</td>
<td>162</td>
<td>246</td>
<td>525</td>
<td>41.6</td>
<td>41.6</td>
<td>69.8</td>
</tr>
<tr>
<td>pc-3</td>
<td>31</td>
<td>53</td>
<td>90</td>
<td>118</td>
<td>203</td>
<td>515</td>
<td>12.7</td>
<td>12.7</td>
<td>22.9</td>
</tr>
<tr>
<td>pc-4</td>
<td>31</td>
<td>52</td>
<td>90</td>
<td>115</td>
<td>202</td>
<td>516</td>
<td>11.9</td>
<td>11.9</td>
<td>19.9</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>49</td>
<td>49</td>
<td>127</td>
<td>77</td>
<td>206</td>
<td>623</td>
<td>26.9</td>
<td>33.6</td>
<td>76.5</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>40</td>
<td>74</td>
<td>102</td>
<td>163</td>
<td>233</td>
<td>517</td>
<td>38.7</td>
<td>38.7</td>
<td>70.0</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>30</td>
<td>51</td>
<td>90</td>
<td>114</td>
<td>201</td>
<td>517</td>
<td>10.6</td>
<td>10.6</td>
<td>19.0</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>30</td>
<td>51</td>
<td>88</td>
<td>114</td>
<td>199</td>
<td>513</td>
<td>10.2</td>
<td>10.2</td>
<td>18.5</td>
</tr>
</tbody>
</table>

Table 6.8: CP-corrected PW86 repulsion energies of rare-gas diatomics ($\mu E_h$)
Figure 6.7: CP-corrected PW86 repulsion energy % deviations of rare-gas diatomics
Figure 6.8: CP-corrected and uncorrected (noCP) PW86 repulsion energy MAPD
QZVP, (aug-)pc-2 and (aug-)pc-3 basis sets the MAPD increases when CP correction is applied. This counterintuitive result arises from error cancellation between BSSE (which causes underestimated repulsion energies) and the intrinsic PW86 error (which is over-repulsive). The effect is generally rather small but is responsible for the unexpectedly accurate PW86 repulsion energies with aug-pVDZ and aug-pVTZ found earlier (Chapter 5) as well as the exceptionally low MAPD of 4.7% with aug-pc-2. It should be emphasized that such error cancellation is completely fortuitous and applies only in very specific cases. In general, uncorrected PW86 repulsion energies are highly inaccurate (especially with small basis sets) and applying CP correction leads to much more reliable results.

A direct comparison of CP-corrected HF and PW86 repulsion energies (Figure 6.9) shows that CP-corrected PW86 repulsion energies are up to one order of magnitude less accurate than the corresponding HF values. Error cancellation between BSIE and DFA occurs only with the SVP basis set. In all other cases, PW86 repulsion energies are less accurate compared to HF. Due to the intrinsic error of PW86, MAPDs converge to around 10% in the CBS limit. The QZVPD basis set has a fortuitously low MAPD of 7.2%.

6.3.4 Summary

Concluding the investigation of basis set effects in rare-gas repulsion energies, error statistics for HF and PW86 (both CP-corrected and uncorrected) are aggregated in Table 6.9.

Figures 6.10 and 6.11 allow a comparison of repulsion energy mean absolute % deviations by basis set family and by basis set size, respectively. Both Figures use a compressed scale above 100% in order to represent the full range of MAPD values in a single plot.

From Figure 6.10 it is evident that the augmented cc basis set family (aug-pVnZ) is the only one which (1) converges efficiently to the CBS limit, (2) gives semi-accurate double-zeta repulsion energies, (3) has similar accuracy for HF and PW86, (4) benefits from error cancellation for non-CP-corrected PW86 repulsion energies, (5) gives highly accurate CP-corrected HF repulsion energies at all (DZ, TZ, QZ & 5Z) levels, and (6) yields MAPDs below 10% (both HF and PW86, both CP-corrected and uncorrected) for the largest (aug-pV5Z) basis set.

As shown in Figure 6.11, all double zeta basis sets except aug-pVDZ suffer from large BSSE and cannot be recommended without CP correction. Even when
Figure 6.9: CP-corrected HF and PW86 repulsion MAPD
<table>
<thead>
<tr>
<th>Method</th>
<th>HF CP</th>
<th>noCP</th>
<th>PW86 CP</th>
<th>noCP</th>
<th>MAPD HF CP</th>
<th>noCP</th>
<th>PW86 CP</th>
<th>noCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>31(d)</td>
<td>21.9</td>
<td>-112</td>
<td>15.7</td>
<td>-156</td>
<td>21.9</td>
<td>112</td>
<td>40.7</td>
<td>156</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>-8.7</td>
<td>-102</td>
<td>39.8</td>
<td>-178</td>
<td>8.7</td>
<td>102</td>
<td>39.8</td>
<td>178</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>-14.2</td>
<td>-236</td>
<td>-35.1</td>
<td>-611</td>
<td>14.2</td>
<td>236</td>
<td>41.3</td>
<td>611</td>
</tr>
<tr>
<td>31+(d)</td>
<td>0.1</td>
<td>-132</td>
<td>14.8</td>
<td>-359</td>
<td>3.4</td>
<td>132</td>
<td>34.7</td>
<td>367</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>-0.7</td>
<td>-87.9</td>
<td>33.9</td>
<td>-132</td>
<td>3.5</td>
<td>87.9</td>
<td>33.9</td>
<td>136</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>-1.6</td>
<td>-110</td>
<td>17.1</td>
<td>-50.0</td>
<td>2.2</td>
<td>110</td>
<td>18.7</td>
<td>62.5</td>
</tr>
<tr>
<td>pVDZ</td>
<td>-19.7</td>
<td>-98.9</td>
<td>26.6</td>
<td>-122</td>
<td>19.7</td>
<td>98.9</td>
<td>36.7</td>
<td>122</td>
</tr>
<tr>
<td>pVTZ</td>
<td>-4.8</td>
<td>-73.0</td>
<td>51.2</td>
<td>-153</td>
<td>5.4</td>
<td>73.0</td>
<td>51.2</td>
<td>153</td>
</tr>
<tr>
<td>pVQZ</td>
<td>-3.6</td>
<td>-31.4</td>
<td>32.2</td>
<td>-110</td>
<td>3.6</td>
<td>31.4</td>
<td>32.2</td>
<td>111</td>
</tr>
<tr>
<td>pV5Z</td>
<td>-2.4</td>
<td>-5.4</td>
<td>16.6</td>
<td>-23.2</td>
<td>2.4</td>
<td>5.4</td>
<td>16.6</td>
<td>27.0</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>1.9</td>
<td>-107</td>
<td>14.5</td>
<td>-370</td>
<td>5.6</td>
<td>107</td>
<td>38.5</td>
<td>380</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>-1.6</td>
<td>-19.7</td>
<td>42.7</td>
<td>-23.5</td>
<td>2.9</td>
<td>19.7</td>
<td>42.7</td>
<td>56.1</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>-3.0</td>
<td>-10.6</td>
<td>26.4</td>
<td>-13.9</td>
<td>3.0</td>
<td>10.6</td>
<td>26.4</td>
<td>36.1</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>-0.1</td>
<td>-33.3</td>
<td>8.8</td>
<td>-32.2</td>
<td>2.2</td>
<td>33.3</td>
<td>21.1</td>
<td>36.1</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>-2.6</td>
<td>-23.9</td>
<td>21.9</td>
<td>9.8</td>
<td>2.6</td>
<td>23.9</td>
<td>21.9</td>
<td>14.4</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>-2.2</td>
<td>-11.1</td>
<td>18.4</td>
<td>10.1</td>
<td>2.2</td>
<td>11.1</td>
<td>18.4</td>
<td>10.6</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>-1.7</td>
<td>-4.5</td>
<td>9.0</td>
<td>5.8</td>
<td>1.7</td>
<td>4.5</td>
<td>9.0</td>
<td>6.4</td>
</tr>
<tr>
<td>SVP</td>
<td>-31.2</td>
<td>-121</td>
<td>9.6</td>
<td>-135</td>
<td>31.2</td>
<td>121</td>
<td>27.1</td>
<td>135</td>
</tr>
<tr>
<td>TZVP</td>
<td>-4.4</td>
<td>-51.2</td>
<td>63.3</td>
<td>-110</td>
<td>4.4</td>
<td>51.2</td>
<td>63.3</td>
<td>110</td>
</tr>
<tr>
<td>QZVP</td>
<td>-2.2</td>
<td>-4.4</td>
<td>19.1</td>
<td>-11.6</td>
<td>2.2</td>
<td>4.4</td>
<td>19.1</td>
<td>15.5</td>
</tr>
<tr>
<td>SVPD</td>
<td>-6.5</td>
<td>-254</td>
<td>-60.6</td>
<td>-759</td>
<td>7.0</td>
<td>254</td>
<td>63.4</td>
<td>759</td>
</tr>
<tr>
<td>TZVPD</td>
<td>-3.9</td>
<td>-64.7</td>
<td>16.8</td>
<td>-263</td>
<td>3.9</td>
<td>64.7</td>
<td>27.8</td>
<td>272</td>
</tr>
<tr>
<td>QZVPD</td>
<td>-1.9</td>
<td>-4.3</td>
<td>5.8</td>
<td>-24.6</td>
<td>1.9</td>
<td>4.3</td>
<td>7.2</td>
<td>33.0</td>
</tr>
<tr>
<td>pc-1</td>
<td>-2.7</td>
<td>-76.3</td>
<td>68.5</td>
<td>-92.9</td>
<td>8.2</td>
<td>76.3</td>
<td>68.5</td>
<td>92.9</td>
</tr>
<tr>
<td>pc-2</td>
<td>-6.3</td>
<td>-16.0</td>
<td>41.6</td>
<td>-20.6</td>
<td>6.3</td>
<td>16.0</td>
<td>41.6</td>
<td>21.6</td>
</tr>
<tr>
<td>pc-3</td>
<td>-1.2</td>
<td>-2.5</td>
<td>12.7</td>
<td>9.4</td>
<td>1.2</td>
<td>2.5</td>
<td>12.7</td>
<td>9.4</td>
</tr>
<tr>
<td>pc-4</td>
<td>-1.7</td>
<td>-2.3</td>
<td>11.9</td>
<td>10.8</td>
<td>1.7</td>
<td>2.3</td>
<td>11.9</td>
<td>10.8</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>0.2</td>
<td>-199</td>
<td>26.9</td>
<td>-54.4</td>
<td>3.2</td>
<td>199</td>
<td>33.6</td>
<td>78.2</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>-1.7</td>
<td>-98.3</td>
<td>38.7</td>
<td>-0.7</td>
<td>1.7</td>
<td>98.3</td>
<td>38.7</td>
<td>4.7</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>-1.5</td>
<td>-6.3</td>
<td>10.6</td>
<td>7.0</td>
<td>1.5</td>
<td>6.3</td>
<td>10.6</td>
<td>7.0</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>-1.7</td>
<td>-3.1</td>
<td>10.2</td>
<td>9.1</td>
<td>1.7</td>
<td>3.1</td>
<td>10.2</td>
<td>9.1</td>
</tr>
</tbody>
</table>

Table 6.9: HF and PW86 repulsion energy error statistics
Figure 6.10: CP-corrected and uncorrected HF and PW86 repulsion MAPD
Figure 6.11: CP-corrected and uncorrected HF and PW86 repulsion MAPD
CP correction is applied, DZ PW86 repulsion energies are rather inaccurate. To a slightly lesser degree, this is still the case at the triple zeta level where aug-pVTZ is the overall most accurate basis set.

In general, augmented pc-type basis sets perform worse than the corresponding aug-cc basis sets at the double zeta (aug-pc-1) and triple zeta (aug-pc-2) levels. However, aug-pc-2 yields a fortuitously low MAPD with CP-uncorrected PW86 (4.7%). The next largest pc-type basis sets, pc-3 and aug-pc-3, are already quintuple zeta and have similar accuracy as aug-pV5Z. The minimally augmented cc basis set family (maug-pVnZ) is significantly less accurate than aug-pVnZ with PW86. Double zeta (SVP) and triple zeta (TZVP) Ahlrichs-type basis sets are very inaccurate when no CP correction is applied (especially with diffuse functions), and also perform much worse with PW86 compared to HF.

At the QZ level, aug-pVQZ performs best overall, but QZVP is more accurate for HF repulsion energies. Quintuple zeta and larger basis sets (except pV5Z) approach the CBS limit for both HF and PW86 repulsion energies.

6.4 Parametrization Of The XDM Damping Function In The CBS Limit

So far, the investigation of basis-set effects has been focused on the repulsive (exchange) component of rare-gas interactions as described by HF theory and semi-local DFAs. Inclusion of an attractive dispersion component is necessary to fully describe rare-gas interactions. Universally applicable DFAs also contain a component accounting for dynamical electron correlation, often termed simply “correlation”. In Chapters 3, 4, and 5 we have combined the PW86 exchange GGA with the PBE dynamical correlation GGA and the XDM method for dispersion to yield the PW86PBE-XDM method. The two empirical parameters in the XDM damping function were fitted to the binding energies of rare-gas diatomics and intermolecular complexes in Chapters 3 and 4, respectively. These fits were performed at the CBS limit but used LDA orbitals for technical reasons. This “post-LDA” approach has a limited utility since it does not allow for the computation of forces and hence geometry optimizations. In Chapter 5, the XDM damping parameters were fitted to intermolecular binding energies using GGA (PW86PBE) orbitals and the aug-pVDZ and aug-pVTZ basis sets. This “post-GGA” approach allows computation of forces but conflates basis set effects with the error inherent in the PW86PBE-XDM method.
Table 6.10: Rare-gas binding energies at various levels of theory (μEh)

<table>
<thead>
<tr>
<th></th>
<th>(BE^{HF}_{\text{aug-pV5Z CP}})</th>
<th>(BE^{PW86}_{\text{aug-pV5Z CP}})</th>
<th>(\Delta E^{PBE}_{\text{aug-pV5Z CP}})</th>
<th>(BE^{PW86PBE}_{\text{aug-pV5Z CP}})</th>
<th>(\Delta E^{XDM}_{\text{aug-pV5Z CP}})</th>
<th>(BE^{PW86PBE−XDM}_{\text{aug-pV5Z CP}})</th>
<th>(BE^{ref}_{\text{reference}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>He-He</td>
<td>-27.8</td>
<td>-29.2</td>
<td>20.0</td>
<td>-9.2</td>
<td>42.2</td>
<td>33.0</td>
<td>34.8</td>
</tr>
<tr>
<td>He-Ne</td>
<td>-46.4</td>
<td>-51.6</td>
<td>42.6</td>
<td>-9.0</td>
<td>77.7</td>
<td>68.7</td>
<td>66.0</td>
</tr>
<tr>
<td>He-Ar</td>
<td>-82.4</td>
<td>-85.9</td>
<td>66.8</td>
<td>-19.1</td>
<td>116.9</td>
<td>97.8</td>
<td>93.7</td>
</tr>
<tr>
<td>Ne-Ne</td>
<td>-94.4</td>
<td>-114.9</td>
<td>99.6</td>
<td>-15.3</td>
<td>147.4</td>
<td>132.1</td>
<td>134</td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>-185.9</td>
<td>-196.8</td>
<td>175.3</td>
<td>-21.5</td>
<td>239.9</td>
<td>218.4</td>
<td>211</td>
</tr>
<tr>
<td>Ar-Ar</td>
<td>-429.7</td>
<td>-502.9</td>
<td>417.7</td>
<td>-85.2</td>
<td>515.5</td>
<td>430.3</td>
<td>454</td>
</tr>
</tbody>
</table>

In this Chapter, we take the more systematic approach of fitting the XDM damping function in the CBS limit using PW86PBE orbitals. In section 6.3, we found that the CBS limit for PW86 rare-gas repulsion energies is reached at the quintuple-zeta level. Specifically, the aug-pVnZ family of basis sets converges very systematically to the CBS limit, with aug-pV5Z yielding the most accurate repulsion energies of all tested basis sets. We therefore employ the aug-pV5Z basis set for the parametrization of the XDM damping function in the CBS limit.

The \(a_1\) and \(a_2\) parameters in the XDM damping function (Eq. 3.4 in Chapter 3) are determined by minimizing the RMSPD of CP-corrected PW86PBE-XDM/aug-pV5Z binding energies \((BE^{PW86PBE−XDM})\) wrt reference binding energies \(BE^{ref}\) of the six rare-gas diatomics. The optimized damping parameters are \(a_1 = 0.65\) and \(a_2 = 1.68\) Å yielding a RMSPD of 4.2%.

The resulting BEs are shown in Table 6.10. Also shown are HF, PW86 and PW86PBE BEs (without dispersion), as well as the contribution of the PBE correlation functional to the PW86PBE binding energy

\[
\Delta E^{PBE} = BE^{PW86PBE} - BE^{PW86}
\]

and the XDM dispersion contribution to the PW86PBE-XDM binding energy

\[
\Delta E^{XDM} = BE^{PW86PBE−XDM} - BE^{PW86PBE}
\]

Binding energies are defined as positive quantities, negative values indicate an (artificially) repulsive interaction at a particular level of theory. The (exchange-only) HF and PW86 binding energies are just the negative of the HF/aug-pV5Z and PW86/aug-pV5Z repulsion energies from Tables 6.5 and 6.8, respectively. The semi-local PBE DFA for dynamical correlation provides significant binding at the equilibrium separation of rare-gas diatomics, but the resulting PW86PBE BEs are
still repulsive. Accurate BEs close to the reference values are found only when the explicit dispersion component $\Delta^{\text{EXDM}}$ is included. For the helium-containing systems, $\Delta^{\text{EPBE}}$ is about half of $\Delta^{\text{EXDM}}$, but for the heavier diatomics $\Delta^{\text{EPBE}}$ approaches the magnitude of the XDM dispersion contribution.

### 6.4.1 Effect Of Dynamical Correlation On Rare-Gas Binding Energies, Total Energies And Dispersion Coefficients

The magnitude of the PBE dynamical correlation component raises the question of whether this contribution is physically valid. By comparison with exact HF rare-gas repulsion energies, we were able to show that the binding provided by various exchange DFAs (including PBE) is a spurious overlap effect [189]. No such comparison is possible for semilocal correlation DFAs, as a rigorous separation of electron correlation into “dynamical correlation” and “dispersion correlation” is not possible.

From a fundamental standpoint that views dispersion solely as the asymptotic attraction between separated, non-overlapping (closed-shell) systems, it is known that semilocal DFAs are fundamentally unable to capture this asymptotic attraction [7, 10]. Indeed, typical correlation GGAs are constructed to capture electron correlation effects very different from dispersion, such as correlation in the slowly varying electron gas (e.g., PBE [29]) or short-ranged intra-atomic electron correlation (e.g., LYP [31]). On the other hand, studies of “medium-range” electron correlation and the good performance of highly empirical semilocal DFAs for NCI suggest that at finite interatomic separations, semilocal DFAs do capture at least part of the electron correlation due to dispersion, even if fortuitously.

To provide further insight into the importance of dynamical correlation, it is instructive to investigate the effect of PBE correlation on total energies and dispersion coefficients. Total energies $E^{\text{total}}$ of rare-gas atoms (atomic units) and total-energy deviations (kcal/mol) at the HF, PW86 and PW86PBE levels of theory using the aug-pV5Z basis set are shown in Table 6.11. HF and PW86 total energies are systematically too high due to lack of dynamical electron correlation. PW86PBE significantly improves total energies over the exchange-only HF and PW86 values. Therefore, inclusion of PBE correlation is necessary to yield a universal functional accurate not only for rare-gas binding energies but also for total energies, thermochemistry and other molecular properties.

Table 6.12 elucidates the effect of including dynamical electron correlation on
Table 6.11: Total energies $E_{\text{total}}$ (hartrees, $E_h$) and total-energy deviations (kcal/mol) of rare-gas atoms at various levels of theory

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>He-He</td>
<td>$r_{eq}$</td>
<td>1.42</td>
<td>1.47</td>
<td>1.65</td>
<td>14</td>
<td>20</td>
<td>18</td>
<td>183.6</td>
<td>183</td>
<td>320</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>100Å</td>
<td></td>
<td>1.46</td>
<td>1.42</td>
<td>1.75</td>
<td>14</td>
<td>20</td>
<td>18</td>
<td>183.6</td>
<td>183</td>
<td>320</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>He-Na</td>
<td>$r_{eq}$</td>
<td>3.029</td>
<td>2.82</td>
<td>3.62</td>
<td>36</td>
<td>53</td>
<td>48</td>
<td>541.5</td>
<td>520</td>
<td>943</td>
<td>810</td>
<td></td>
</tr>
<tr>
<td>100Å</td>
<td></td>
<td>9.538</td>
<td>8.20</td>
<td>9.98</td>
<td>185</td>
<td>245</td>
<td>218</td>
<td>3701</td>
<td>4036</td>
<td>6195</td>
<td>5206</td>
<td></td>
</tr>
<tr>
<td>He-Ar</td>
<td>$r_{eq}$</td>
<td>5.62</td>
<td>7.48</td>
<td>6.94</td>
<td>88</td>
<td>134</td>
<td>121</td>
<td>1433</td>
<td>1425</td>
<td>2691</td>
<td>2284</td>
<td></td>
</tr>
<tr>
<td>100Å</td>
<td></td>
<td>6.383</td>
<td>5.61</td>
<td>7.50</td>
<td>90.34</td>
<td>88</td>
<td>136</td>
<td>1536</td>
<td>1425</td>
<td>2691</td>
<td>2284</td>
<td></td>
</tr>
<tr>
<td>Na-Ne</td>
<td>$r_{eq}$</td>
<td>6.23</td>
<td>5.67</td>
<td>6.17</td>
<td>69.04</td>
<td>88</td>
<td>136</td>
<td>1536</td>
<td>1425</td>
<td>2691</td>
<td>2284</td>
<td></td>
</tr>
<tr>
<td>100Å</td>
<td></td>
<td>19.50</td>
<td>16.45</td>
<td>20.51</td>
<td>188.5</td>
<td>390</td>
<td>507</td>
<td>9335</td>
<td>10067</td>
<td>16007</td>
<td>13381</td>
<td></td>
</tr>
<tr>
<td>Ne-Ar</td>
<td>$r_{eq}$</td>
<td>54.29</td>
<td>63.10</td>
<td>57.86</td>
<td>1902</td>
<td>2365</td>
<td>2088</td>
<td>62233</td>
<td>85695</td>
<td>71874</td>
<td>62233</td>
<td></td>
</tr>
<tr>
<td>100Å</td>
<td></td>
<td>64.30</td>
<td>54.38</td>
<td>63.52</td>
<td>58.00</td>
<td>1623</td>
<td>2401</td>
<td>2105</td>
<td>49060</td>
<td>87901</td>
<td>72912</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.12: Rare-gas dispersion coefficients at various levels of theory (atomic units)

XDM dispersion coefficients. C6, C8 and C10 values are calculated at both the equilibrium separation ($r_{eq}$) of the corresponding rare-gas diatomic and at a separation of 100Å. We note in passing that dispersion coefficients are largely independent of separation, allowing the computation of "post-GGA" dispersion forces without including the XDM potential in the SCF calculation.

PW86PBE-XDM gives (on average) slightly more accurate C6 values than HF-XDM and PW86-XDM. HF-XDM underestimates C6s especially for Ar-containing systems, while PW86-based C6s are overestimated especially for He- and Ne-containing systems. C8 and C10 coefficients are much more accurate if based on HF orbitals compared to the DFT-based ones. PW86 significantly overestimates C8s and C10s, and inclusion of PBE dynamical correlation partly reduces this overestimation. However, for all three types of dispersion coefficients the difference
between HF-XDM and PW86-XDM values is larger than the difference between PW86-XDM and PW86PBE-XDM values. Thus, dynamical correlation has only a minor effect on dispersion coefficients in rare-gas diatomics, smaller than the deviation between HF and PW86 exchange. Unlike the BEs of Table 6.10 which depend on empirical damping function parameters, the dispersion coefficients of Table 6.12 are completely nonempirical.

We conclude that PBE correlation yields significant binding at the equilibrium separation of rare-gas diatomics, significantly improves total energies of rare-gas atoms compared to exchange-only HF and PW86, and has a small beneficial effect on XDM dispersion coefficients. We therefore proceed with the PW86PBE-XDM functional used previously and now parametrized in the CBS limit. The basis-set dependence of PW86PBE-XDM rare-gas BEs will be investigated in the following section, and comprehensive benchmarking will be performed in the following Chapters.

6.5 Basis-Set Effects On Rare-Gas Binding Energies

With the CBS-limit damping parameters \(a_1 = 0.65\) and \(a_2 = 1.68\) Å determined in the previous section, we now proceed to calculate rare-gas binding energies with the PW86PBE-XDM functional. Note that the same damping parameters are used throughout this section, i.e. no refitting was performed. PW86PBE-XDM rare-gas BEs without counterpoise correction are listed in Table 6.13, and percent deviations from the reference BEs [210] for each diatomic are shown in Figure 6.12. Negative percent deviations (on the left side of Figure 6.12) indicate underbinding, and positive percent deviations indicate overbinding. All tested basis sets give binding energies of the correct sign. As was the case with repulsion energies, the 31+(d), 311(3df3pd), maug-pVDZ, SVPD, and TZVPD basis sets completely fail to describe binding energies of helium-containing diatomics, with deviations exceeding 200%. Again, only the aug-pVnZ series converges smoothly to the CBS limit while yielding quite accurate binding energies already at the DZ (22.2% MAPD) and TZ (8.3% MAPD) levels. The CBS limit is reached at the aug-pVQZ level with a MAPD of 5.6% and a maximum deviation of less than 10%. These deviations characterizes the intrinsic accuracy of PW86PBE-XDM for rare-gas binding energies and are also attained with the larger aug-pV5Z, (aug-)pc-3, and (aug-)pc-4 basis sets.

In general, rare-gas binding energies have a very similar basis-set dependence compared to the repulsion energies of Section 6.3. Indeed, Figure 6.12 is almost
<table>
<thead>
<tr>
<th>reference</th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>31(d)</td>
<td>30</td>
<td>91</td>
<td>133</td>
<td>219</td>
<td>541</td>
<td>526</td>
<td>50.4</td>
<td>54.8</td>
<td>156.5</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>49</td>
<td>156</td>
<td>133</td>
<td>238</td>
<td>652</td>
<td>482</td>
<td>85.2</td>
<td>85.2</td>
<td>209.1</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>164</td>
<td>816</td>
<td>197</td>
<td>311</td>
<td>942</td>
<td>523</td>
<td>351.8</td>
<td>351.8</td>
<td>1136.7</td>
</tr>
<tr>
<td>31+(d)</td>
<td>30</td>
<td>422</td>
<td>653</td>
<td>288</td>
<td>309</td>
<td>416</td>
<td>212.7</td>
<td>219.8</td>
<td>596.5</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>49</td>
<td>167</td>
<td>255</td>
<td>247</td>
<td>289</td>
<td>473</td>
<td>82.0</td>
<td>82.0</td>
<td>172.5</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>32</td>
<td>143</td>
<td>110</td>
<td>273</td>
<td>340</td>
<td>490</td>
<td>50.0</td>
<td>52.6</td>
<td>117.3</td>
</tr>
<tr>
<td>pVDZ</td>
<td>32</td>
<td>82</td>
<td>141</td>
<td>273</td>
<td>340</td>
<td>490</td>
<td>15.4</td>
<td>28.0</td>
<td>38.3</td>
</tr>
<tr>
<td>pVTZ</td>
<td>42</td>
<td>163</td>
<td>120</td>
<td>305</td>
<td>450</td>
<td>492</td>
<td>17.0</td>
<td>22.2</td>
<td>38.3</td>
</tr>
<tr>
<td>pVQZ</td>
<td>42</td>
<td>141</td>
<td>113</td>
<td>288</td>
<td>381</td>
<td>451</td>
<td>58.4</td>
<td>58.6</td>
<td>114.7</td>
</tr>
<tr>
<td>pV5Z</td>
<td>39</td>
<td>86</td>
<td>106</td>
<td>165</td>
<td>243</td>
<td>436</td>
<td>14.8</td>
<td>16.1</td>
<td>30.1</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>32</td>
<td>490</td>
<td>638</td>
<td>171</td>
<td>242</td>
<td>401</td>
<td>207.5</td>
<td>214.4</td>
<td>642.2</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>42</td>
<td>100</td>
<td>148</td>
<td>98</td>
<td>198</td>
<td>433</td>
<td>15.4</td>
<td>28.0</td>
<td>58.0</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>42</td>
<td>87</td>
<td>132</td>
<td>124</td>
<td>217</td>
<td>433</td>
<td>14.2</td>
<td>18.1</td>
<td>40.8</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>29</td>
<td>91</td>
<td>106</td>
<td>180</td>
<td>266</td>
<td>483</td>
<td>17.0</td>
<td>22.2</td>
<td>38.3</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>31</td>
<td>80</td>
<td>98</td>
<td>140</td>
<td>226</td>
<td>451</td>
<td>4.1</td>
<td>8.3</td>
<td>21.6</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>33</td>
<td>72</td>
<td>101</td>
<td>135</td>
<td>225</td>
<td>439</td>
<td>2.8</td>
<td>5.6</td>
<td>9.8</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>33</td>
<td>72</td>
<td>100</td>
<td>138</td>
<td>222</td>
<td>433</td>
<td>2.6</td>
<td>5.5</td>
<td>9.4</td>
</tr>
<tr>
<td>SVP</td>
<td>34</td>
<td>85</td>
<td>113</td>
<td>170</td>
<td>282</td>
<td>605</td>
<td>23.2</td>
<td>24.2</td>
<td>33.7</td>
</tr>
<tr>
<td>TZVP</td>
<td>43</td>
<td>112</td>
<td>127</td>
<td>281</td>
<td>318</td>
<td>502</td>
<td>49.9</td>
<td>49.9</td>
<td>109.5</td>
</tr>
<tr>
<td>QZVP</td>
<td>37</td>
<td>76</td>
<td>103</td>
<td>148</td>
<td>229</td>
<td>434</td>
<td>7.6</td>
<td>9.1</td>
<td>16.0</td>
</tr>
<tr>
<td>SVPD</td>
<td>294</td>
<td>650</td>
<td>880</td>
<td>359</td>
<td>481</td>
<td>700</td>
<td>470.1</td>
<td>470.1</td>
<td>884.7</td>
</tr>
<tr>
<td>TZVPD</td>
<td>156</td>
<td>276</td>
<td>358</td>
<td>143</td>
<td>239</td>
<td>491</td>
<td>162.8</td>
<td>162.8</td>
<td>348.8</td>
</tr>
<tr>
<td>QZVPD</td>
<td>48</td>
<td>91</td>
<td>124</td>
<td>138</td>
<td>218</td>
<td>428</td>
<td>18.3</td>
<td>20.1</td>
<td>38.2</td>
</tr>
<tr>
<td>pc-1</td>
<td>27</td>
<td>92</td>
<td>103</td>
<td>199</td>
<td>392</td>
<td>481</td>
<td>27.7</td>
<td>35.3</td>
<td>85.8</td>
</tr>
<tr>
<td>pc-2</td>
<td>36</td>
<td>87</td>
<td>106</td>
<td>177</td>
<td>246</td>
<td>454</td>
<td>16.3</td>
<td>16.3</td>
<td>32.2</td>
</tr>
<tr>
<td>pc-3</td>
<td>32</td>
<td>69</td>
<td>96</td>
<td>130</td>
<td>212</td>
<td>419</td>
<td>-1.9</td>
<td>4.4</td>
<td>-8.3</td>
</tr>
<tr>
<td>pc-4</td>
<td>32</td>
<td>69</td>
<td>97</td>
<td>132</td>
<td>215</td>
<td>428</td>
<td>-0.9</td>
<td>4.4</td>
<td>-8.5</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>34</td>
<td>127</td>
<td>116</td>
<td>361</td>
<td>351</td>
<td>537</td>
<td>61.3</td>
<td>62.3</td>
<td>169.6</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>38</td>
<td>90</td>
<td>125</td>
<td>196</td>
<td>269</td>
<td>518</td>
<td>27.5</td>
<td>27.5</td>
<td>46.1</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>33</td>
<td>73</td>
<td>97</td>
<td>137</td>
<td>219</td>
<td>427</td>
<td>1.6</td>
<td>5.3</td>
<td>10.5</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>32</td>
<td>70</td>
<td>98</td>
<td>132</td>
<td>216</td>
<td>430</td>
<td>-0.1</td>
<td>4.5</td>
<td>-7.3</td>
</tr>
</tbody>
</table>

Table 6.13: PW86PBE-XDM binding energies of rare-gas diatomics ($\mu E_h$)
Figure 6.12: PW86PBE-XDM binding energy percent deviations of rare-gas diatomics
a perfect mirror image of Figure 6.5, although the scales of both figures differ by a factor of two. A notable exception is the aug-pc-2 basis set which gave highly accurate repulsion energies (4.7% MAPD) due to fortuitous error cancellation, but yields much less accurate binding energies (27.5% MAPD).

6.5.1 Basis-Set Dependence Of XDM Dispersion Energies

To quantify the supposition that the basis-set dependence of rare-gas BEs arises primarily from the underlying DFA, damped XDM dispersion energies are listed in Table 6.14 and plotted in Figure 6.13. No reference values or error statistics are given as no rigorous reference for damped dispersion energies at finite interatomic separation exists.

The largest basis-set effects on damped XDM dispersion energies range from a few $\mu E_h$ for He$_2$ up to 100$\mu E_h$ for Ar$_2$. Significant basis-set dependence is found only with certain unaugmented DZ and TZ basis sets, especially 31(d), pVDZ and SVP. Diffuse-augmented basis sets yield essentially constant (i.e., CBS limit) dispersion energies. Overall, XDM dispersion energies are much less sensitive to the employed basis set than repulsion and binding energies, even though they are of the same magnitude (c.f. Tables 6.3 and 6.10). Diffuse functions are necessary to yield accurate XDM dispersion energies with small (DZ and TZ) basis sets.

6.5.2 Counterpoise Correction And Basis-Set Superposition Error

The rare-gas repulsion energies of Section 6.3 were strongly affected by BSSE, and could be improved by up to an order of magnitude for many common basis sets by employing the CP correction scheme. We therefore apply the CP correction to PW86PBE-XDM rare-gas binding energies, which are listed in Table 6.15. Percent deviations for each diatomic are shown in Figure 6.14.

Uncorrected PW86PBE-XDM rare-gas binding energies showed an overbinding tendency most pronounced with small basis sets due to significant BSSE. Conversely, CP-corrected PW86PBE-XDM tends to underestimate BEs. In fact, the 31(d), 31+(d), pVDZ, maug-pVDZ, SVP, TZVP and pc-1 basis sets yield CP-corrected BEs of the wrong sign (i.e., repulsive interactions) for He$_2$, corresponding to percent deviations $\leq -100\%$ in Figure 6.14. This qualitatively incorrect behaviour could be prevented by fitting the XDM damping function for each individual basis set. However, this approach conflates basis set effects with the intrinsic error of
<table>
<thead>
<tr>
<th></th>
<th>He2</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne2</th>
<th>Ne-Ar</th>
<th>Ar2</th>
</tr>
</thead>
<tbody>
<tr>
<td>31(d)</td>
<td>36</td>
<td>61</td>
<td>100</td>
<td>108</td>
<td>194</td>
<td>471</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>40</td>
<td>68</td>
<td>110</td>
<td>122</td>
<td>213</td>
<td>502</td>
</tr>
<tr>
<td>311(3df3pd)</td>
<td>40</td>
<td>68</td>
<td>111</td>
<td>122</td>
<td>214</td>
<td>503</td>
</tr>
<tr>
<td>31+(d)</td>
<td>36</td>
<td>72</td>
<td>107</td>
<td>147</td>
<td>239</td>
<td>515</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>40</td>
<td>76</td>
<td>113</td>
<td>148</td>
<td>240</td>
<td>518</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>43</td>
<td>78</td>
<td>117</td>
<td>148</td>
<td>240</td>
<td>518</td>
</tr>
<tr>
<td>pVDZ</td>
<td>36</td>
<td>61</td>
<td>99</td>
<td>110</td>
<td>196</td>
<td>468</td>
</tr>
<tr>
<td>pVTZ</td>
<td>41</td>
<td>72</td>
<td>112</td>
<td>131</td>
<td>221</td>
<td>500</td>
</tr>
<tr>
<td>pVQZ</td>
<td>42</td>
<td>75</td>
<td>115</td>
<td>139</td>
<td>231</td>
<td>510</td>
</tr>
<tr>
<td>pV5Z</td>
<td>42</td>
<td>77</td>
<td>116</td>
<td>146</td>
<td>237</td>
<td>513</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>36</td>
<td>72</td>
<td>107</td>
<td>146</td>
<td>240</td>
<td>515</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>41</td>
<td>77</td>
<td>115</td>
<td>148</td>
<td>242</td>
<td>519</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>42</td>
<td>77</td>
<td>116</td>
<td>148</td>
<td>240</td>
<td>516</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>42</td>
<td>77</td>
<td>117</td>
<td>146</td>
<td>240</td>
<td>515</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>42</td>
<td>78</td>
<td>118</td>
<td>148</td>
<td>242</td>
<td>519</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>148</td>
<td>240</td>
<td>516</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>147</td>
<td>240</td>
<td>516</td>
</tr>
<tr>
<td>SVP</td>
<td>36</td>
<td>61</td>
<td>93</td>
<td>110</td>
<td>185</td>
<td>418</td>
</tr>
<tr>
<td>TZVP</td>
<td>40</td>
<td>74</td>
<td>109</td>
<td>140</td>
<td>228</td>
<td>494</td>
</tr>
<tr>
<td>QZVP</td>
<td>42</td>
<td>77</td>
<td>116</td>
<td>146</td>
<td>238</td>
<td>513</td>
</tr>
<tr>
<td>SVVPD</td>
<td>36</td>
<td>72</td>
<td>107</td>
<td>144</td>
<td>239</td>
<td>516</td>
</tr>
<tr>
<td>TZVPD</td>
<td>40</td>
<td>76</td>
<td>113</td>
<td>148</td>
<td>240</td>
<td>515</td>
</tr>
<tr>
<td>QZVPD</td>
<td>42</td>
<td>77</td>
<td>116</td>
<td>146</td>
<td>239</td>
<td>515</td>
</tr>
<tr>
<td>pc-1</td>
<td>41</td>
<td>71</td>
<td>113</td>
<td>128</td>
<td>221</td>
<td>509</td>
</tr>
<tr>
<td>pc-2</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>147</td>
<td>238</td>
<td>514</td>
</tr>
<tr>
<td>pc-3</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>147</td>
<td>240</td>
<td>515</td>
</tr>
<tr>
<td>pc-4</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>147</td>
<td>240</td>
<td>515</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>44</td>
<td>79</td>
<td>120</td>
<td>150</td>
<td>246</td>
<td>530</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>43</td>
<td>79</td>
<td>118</td>
<td>150</td>
<td>243</td>
<td>520</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>147</td>
<td>240</td>
<td>516</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>42</td>
<td>78</td>
<td>117</td>
<td>147</td>
<td>240</td>
<td>515</td>
</tr>
</tbody>
</table>

Table 6.14: Damped XDM dispersion energies of rare-gas diatomics ($\mu E_h$)
Figure 6.13: Damped XDM dispersion energies of rare-gas diatomics
<table>
<thead>
<tr>
<th>reference</th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>31(d)</td>
<td>-8</td>
<td>15</td>
<td>22</td>
<td>73</td>
<td>101</td>
<td>240</td>
<td>-70.1</td>
<td>70.1</td>
<td>-122.2</td>
</tr>
<tr>
<td>311(2dfp)</td>
<td>3</td>
<td>11</td>
<td>59</td>
<td>32</td>
<td>117</td>
<td>376</td>
<td>-58.2</td>
<td>58.2</td>
<td>-92.1</td>
</tr>
<tr>
<td>311(3d3f3pd)</td>
<td>34</td>
<td>100</td>
<td>88</td>
<td>54</td>
<td>192</td>
<td>399</td>
<td>-6.2</td>
<td>23.1</td>
<td>-60.1</td>
</tr>
<tr>
<td>31+(d)</td>
<td>-8</td>
<td>71</td>
<td>56</td>
<td>137</td>
<td>200</td>
<td>359</td>
<td>-29.9</td>
<td>33.0</td>
<td>-122.2</td>
</tr>
<tr>
<td>311+(2dfp)</td>
<td>3</td>
<td>49</td>
<td>76</td>
<td>104</td>
<td>197</td>
<td>428</td>
<td>-28.7</td>
<td>28.7</td>
<td>-92.1</td>
</tr>
<tr>
<td>311++(3d3f3pd)</td>
<td>24</td>
<td>68</td>
<td>89</td>
<td>120</td>
<td>221</td>
<td>435</td>
<td>-7.3</td>
<td>10.0</td>
<td>-32.4</td>
</tr>
<tr>
<td>pVDZ</td>
<td>-7</td>
<td>10</td>
<td>16</td>
<td>50</td>
<td>69</td>
<td>199</td>
<td>-79.2</td>
<td>79.2</td>
<td>-120.4</td>
</tr>
<tr>
<td>pVTZ</td>
<td>14</td>
<td>17</td>
<td>63</td>
<td>30</td>
<td>120</td>
<td>372</td>
<td>-50.8</td>
<td>50.8</td>
<td>-78.0</td>
</tr>
<tr>
<td>pVQZ</td>
<td>26</td>
<td>39</td>
<td>90</td>
<td>79</td>
<td>179</td>
<td>418</td>
<td>-22.4</td>
<td>22.4</td>
<td>-41.2</td>
</tr>
<tr>
<td>pV5Z</td>
<td>30</td>
<td>62</td>
<td>93</td>
<td>124</td>
<td>207</td>
<td>422</td>
<td>-5.9</td>
<td>5.9</td>
<td>-13.6</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>-7</td>
<td>75</td>
<td>52</td>
<td>127</td>
<td>193</td>
<td>332</td>
<td>-31.8</td>
<td>36.6</td>
<td>-120.4</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>14</td>
<td>52</td>
<td>76</td>
<td>89</td>
<td>192</td>
<td>420</td>
<td>-24.8</td>
<td>24.8</td>
<td>-58.2</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>26</td>
<td>59</td>
<td>92</td>
<td>112</td>
<td>208</td>
<td>427</td>
<td>-10.3</td>
<td>10.3</td>
<td>-25.2</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>24</td>
<td>70</td>
<td>80</td>
<td>134</td>
<td>210</td>
<td>397</td>
<td>-8.8</td>
<td>11.1</td>
<td>-31.0</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>29</td>
<td>68</td>
<td>90</td>
<td>116</td>
<td>209</td>
<td>426</td>
<td>-6.4</td>
<td>7.4</td>
<td>-16.6</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>32</td>
<td>64</td>
<td>96</td>
<td>120</td>
<td>211</td>
<td>430</td>
<td>-3.9</td>
<td>4.8</td>
<td>-10.8</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>33</td>
<td>69</td>
<td>98</td>
<td>132</td>
<td>218</td>
<td>430</td>
<td>0.1</td>
<td>4.0</td>
<td>-5.2</td>
</tr>
<tr>
<td>SVP</td>
<td>-6</td>
<td>11</td>
<td>20</td>
<td>49</td>
<td>70</td>
<td>107</td>
<td>-81.2</td>
<td>81.2</td>
<td>-117.5</td>
</tr>
<tr>
<td>TZVP</td>
<td>-1</td>
<td>19</td>
<td>46</td>
<td>75</td>
<td>148</td>
<td>353</td>
<td>-53.5</td>
<td>53.5</td>
<td>-102.2</td>
</tr>
<tr>
<td>QZVP</td>
<td>28</td>
<td>60</td>
<td>91</td>
<td>120</td>
<td>204</td>
<td>420</td>
<td>-8.8</td>
<td>8.8</td>
<td>-20.0</td>
</tr>
<tr>
<td>SVVPD</td>
<td>47</td>
<td>95</td>
<td>127</td>
<td>152</td>
<td>228</td>
<td>377</td>
<td>20.1</td>
<td>25.8</td>
<td>44.6</td>
</tr>
<tr>
<td>TZVPD</td>
<td>35</td>
<td>60</td>
<td>90</td>
<td>115</td>
<td>199</td>
<td>425</td>
<td>-6.7</td>
<td>6.7</td>
<td>-14.3</td>
</tr>
<tr>
<td>QZVPD</td>
<td>35</td>
<td>69</td>
<td>96</td>
<td>133</td>
<td>215</td>
<td>426</td>
<td>0.4</td>
<td>2.9</td>
<td>-6.2</td>
</tr>
<tr>
<td>pc-1</td>
<td>-12</td>
<td>-1</td>
<td>25</td>
<td>28</td>
<td>86</td>
<td>320</td>
<td>-79.6</td>
<td>79.6</td>
<td>-134.4</td>
</tr>
<tr>
<td>pc-2</td>
<td>28</td>
<td>50</td>
<td>90</td>
<td>105</td>
<td>184</td>
<td>412</td>
<td>-15.4</td>
<td>15.4</td>
<td>-24.7</td>
</tr>
<tr>
<td>pc-3</td>
<td>31</td>
<td>67</td>
<td>94</td>
<td>129</td>
<td>211</td>
<td>417</td>
<td>-3.5</td>
<td>3.9</td>
<td>-9.9</td>
</tr>
<tr>
<td>pc-4</td>
<td>32</td>
<td>68</td>
<td>95</td>
<td>132</td>
<td>215</td>
<td>427</td>
<td>-1.6</td>
<td>3.8</td>
<td>-8.7</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>16</td>
<td>64</td>
<td>71</td>
<td>143</td>
<td>202</td>
<td>409</td>
<td>-14.7</td>
<td>17.0</td>
<td>-54.8</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>26</td>
<td>55</td>
<td>89</td>
<td>107</td>
<td>202</td>
<td>423</td>
<td>-13.3</td>
<td>13.3</td>
<td>-26.8</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>33</td>
<td>69</td>
<td>95</td>
<td>132</td>
<td>214</td>
<td>423</td>
<td>-1.3</td>
<td>3.7</td>
<td>-6.8</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>32</td>
<td>69</td>
<td>96</td>
<td>132</td>
<td>216</td>
<td>428</td>
<td>-1.0</td>
<td>3.8</td>
<td>-7.5</td>
</tr>
</tbody>
</table>

Table 6.15: CP-corrected PW86PBE-XDM binding energies of rare-gas diatomics (μEₖ)
Figure 6.14: CP-corrected PW86PBE-XDM binding energy percent deviations of rare-gas diatomics
the PW86PBE-XDM method and would lead to a proliferation of damping parameter fits, complicating implementation of the PW86PBE-XDM method in standard electronic structure codes.

The CP-corrected PW86PBE-XDM BE deviations of Figure 6.14 are generally much smaller than the uncorrected values of Figure 6.12, demonstrating the effectiveness of CP correction for rare-gas BEs. The CBS limit is reached for the aug-pVQZ, aug-pV5Z, QVZPD, (aug-)pc-3 and (aug-)pc-4 basis sets with MAPDs below 5% and maximum percent deviations around 10%. The aug-pVnZ series yields the most accurate BEs and converges most effectively to the CBS limit.

The BE deviations in Figure 6.14 are comparable in magnitude to the CP-corrected PW86 repulsion energy deviations of Figure 6.7. When CP correction is applied, small basis sets such as 31(d), 31+(d), (maug-)pVDZ, SVP and pc-1 strongly overestimate the repulsion energy of He₂ (c.f. Figure 6.7). Combined with the underestimation of the attractive XDM dispersion energy (c.f. Figure 6.13), this results in an artificially repulsive interaction of the helium dimer.

A plot of CP-corrected and uncorrected PW86PBE-XDM binding energy MAPDs is given in Figure 6.15, which may be compared to Figure 6.8 for PW86 repulsion energies. Noting the different scales of both figures, it is evident that BEs are less sensitive to BSSE than repulsion energies. This is due to the significant contribution of the dispersion energy, which is rather insensitive to basis set effects, including BSSE.

Due to fortuitous error cancellation, the 31(d), pVDZ, SVP, TZVP and pc-1 basis sets give lower MAPD without CP correction than with CP correction. Since these basis sets rely on artificial attraction due to BSSE to give the correct sign of He₂ binding energies (vide supra), they cannot be recommended for the description of rare-gas interactions with PW86PBE-XDM or similar DFT-based methods for dispersion interactions. As was the case for repulsion energies, PW86PBE-XDM binding energies with small basis sets are rather inaccurate in general, and CP correction improves accuracy for most of the tested basis sets.

Figure 6.16 gives a comparison of PW86PBE-XDM binding energy MAPDs ordered by increasing basis set size. This Figure can be directly compared to the repulsion energy MAPDs of Figure 6.11, leading to very similar conclusions. aug-pVDZ, aug-pVTZ and aug-pVQZ are the most accurate basis sets at their respective levels, with or without CP correction. TZVPD, 311++(3df3pd), and QZVPD yield comparable accuracy only when CP correction is applied.
Figure 6.15: CP-corrected and uncorrected PW86PBE-XDM binding energy MAPD
Figure 6.16: CP-corrected and uncorrected PW86PBE-XDM binding energy MAPD
6.5.3 Rare-Gas Binding Energies With Pople-Type Basis Sets

Pople-type basis sets showed a rather unsystematic behaviour for repulsion energies and binding energies of rare-gas diatomics. These basis sets differ from the other basis set families in that the user may freely combine a given valence basis set (denoted as 3-21G, 6-31G, or 6-311G) with various polarization (denoted as e.g., (d), (d,p), (2df,p), (2df,2p), or (3df,3pd)) and/or diffuse functions (denoted as + or ++). The disadvantage to this flexibility is that the resulting basis set may be unbalanced in the interplay between valence, polarization, and diffuse functions. So far, we have tested the widely used 6-31G(d) (also known as 6-31G*) and 6-311G(3df,3pd) basis sets as well as the intermediate 6-311G(2df,p) basis set and their diffuse-augmented variants 6-31+G(d), 6-311+G(2df,p), and 6-311++G(3df,3pd).

Tables 6.17 and 6.18 present uncorrected and CP-corrected rare-gas BEs for a larger selection of Pople-type basis sets. Figure 6.17 shows the percent deviation for each diatomic, using a compressed scale above 200%. The basis sets are divided into three groups:

1. No diffuse functions (top)
2. “+”: sp-type diffuse functions on Ne, and Ar; no diffuse functions on He (middle)
3. “++”: s-type diffuse function on He; sp-type diffuse functions on Ne and Ar (bottom)

Within each group, basis sets are listed in order of increasing size. Table 6.16 lists the composition of each basis set in terms of contracted Gaussian basis functions. Note that groups (1) and (2) represent identical basis sets for the helium atom, therefore yielding identical He$_2$ binding energies. The 6-311G(3df,p) and 6-311G(3df,3pd) are the same basis sets for the neon and argon atoms, yielding identical BEs for the Ne$_2$, Ne-Ar and Ar$_2$ systems. This is also the case for 6-311+G(3df,p), 6-311++G(3df,p), 6-311+G(3df,3pd), and 6-311++G(3df,3pd).

When no CP correction is applied, increasing the size of Pople-type basis sets does not generally improve the accuracy of rare-gas BEs. In fact, the small 6-31G and the intermediate 6-311++G(2df,p) basis set are the most accurate double-zeta and triple-zeta Pople-type basis sets with MAPDs of 41.1% and 30.4%, respectively. The 6-311(2df,p), 6-311(3df,p), and 6-311G(3df,3pd) basis sets suffer from
Figure 6.17: Uncorrected (no counterpoise) and CP-corrected PW86PBE-XDM binding energy percent deviations of rare-gas diatomics with Pople-type basis sets
Table 6.16: Pople-type basis sets for rare-gas elements

<table>
<thead>
<tr>
<th>basis set</th>
<th>diffuse functions</th>
<th>contracted basis functions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>He</td>
<td>Ne, Ar</td>
</tr>
<tr>
<td>3-21G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-31G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-31G(d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(2df,p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(3df,p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(3df,3pd)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-21+G</td>
<td>sp</td>
<td></td>
</tr>
<tr>
<td>6-31+G</td>
<td>sp</td>
<td></td>
</tr>
<tr>
<td>6-31+G(d)</td>
<td>sp</td>
<td></td>
</tr>
<tr>
<td>6-311+G(2df,p)</td>
<td>sp</td>
<td></td>
</tr>
<tr>
<td>6-311+G(3df,p)</td>
<td>sp</td>
<td></td>
</tr>
<tr>
<td>6-311+G(3df,3pd)</td>
<td>sp</td>
<td></td>
</tr>
<tr>
<td>6-31++G(d)</td>
<td>s</td>
<td>sp</td>
</tr>
<tr>
<td>6-311++G(2df,p)</td>
<td>s</td>
<td>sp</td>
</tr>
<tr>
<td>6-311++G(3df,p)</td>
<td>s</td>
<td>sp</td>
</tr>
<tr>
<td>6-311++G(3df,3pd)</td>
<td>s</td>
<td>sp</td>
</tr>
</tbody>
</table>

large BSSE due to an excess of polarization functions while lacking diffuse functions. Similarly, 6-311G+(3df,3pd) is very inaccurate for helium-containing systems. Adding diffuse functions on helium to yield 6-311G++(3df,3pd) significantly improves accuracy but is still worse than the more balanced 6-311++G(2df,p) basis set. Conversely, the 3-21+G basis set contains sp-type diffuse functions on Ne and Ar but has an inadequate valence basis set, resulting in extreme artificial overbinding due to BSSE for all diatomics except He2 (where it is identical to 3-21G). To a lesser degree, the same effect is observed with the 6-31+G and 6-31+G(d) basis sets.

When BSSE is eliminated by employing the counterpoise correction, Pople-type basis sets yield less erratic and more accurate BEs. Specifically, the largest basis sets (6-311G+(3df,3pd) and 6-311G++(3df,3pd)) are also the most accurate with MAPDs of 5.7% and 10%, respectively. The better performance of 6-311G+(3df,3pd) over 6-311G++(3df,3pd) is perhaps accidental, as (+)-augmented Pople-type basis sets are generally less accurate for helium-containing systems (especially He2) than (++)-augmented ones, due to the lack of diffuse functions on helium.

In conclusion, CP correction is necessary to achieve acceptable accuracy for rare-gas BEs with Pople-type basis sets. Without CP correction, adding diffuse
<table>
<thead>
<tr>
<th>reference</th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-21G</td>
<td>25</td>
<td>53</td>
<td>134</td>
<td>89</td>
<td>430</td>
<td>720</td>
<td>21.0</td>
<td>47.6</td>
<td>104.0</td>
</tr>
<tr>
<td>6-31G</td>
<td>30</td>
<td>81</td>
<td>131</td>
<td>150</td>
<td>521</td>
<td>507</td>
<td>36.7</td>
<td>41.1</td>
<td>146.9</td>
</tr>
<tr>
<td>6-31G(d)</td>
<td>30</td>
<td>91</td>
<td>133</td>
<td>219</td>
<td>541</td>
<td>526</td>
<td>50.4</td>
<td>54.8</td>
<td>156.5</td>
</tr>
<tr>
<td>6-311(2df,p)</td>
<td>49</td>
<td>156</td>
<td>133</td>
<td>238</td>
<td>652</td>
<td>482</td>
<td>85.2</td>
<td>85.2</td>
<td>209.1</td>
</tr>
<tr>
<td>6-311G(3df,p)</td>
<td>49</td>
<td>160</td>
<td>154</td>
<td>311</td>
<td>942</td>
<td>523</td>
<td>123.7</td>
<td>123.7</td>
<td>346.2</td>
</tr>
<tr>
<td>6-311G(3df,3pd)</td>
<td>164</td>
<td>816</td>
<td>197</td>
<td>311</td>
<td>942</td>
<td>523</td>
<td>351.8</td>
<td>351.8</td>
<td>1136.7</td>
</tr>
<tr>
<td>3-21+G</td>
<td>25</td>
<td>963</td>
<td>1711</td>
<td>881</td>
<td>1107</td>
<td>1948</td>
<td>728.1</td>
<td>737.2</td>
<td>1726.2</td>
</tr>
<tr>
<td>6-31+G</td>
<td>30</td>
<td>415</td>
<td>656</td>
<td>237</td>
<td>282</td>
<td>442</td>
<td>203.9</td>
<td>209.3</td>
<td>599.8</td>
</tr>
<tr>
<td>6-31+G(d)</td>
<td>30</td>
<td>422</td>
<td>653</td>
<td>288</td>
<td>309</td>
<td>416</td>
<td>212.7</td>
<td>219.8</td>
<td>596.5</td>
</tr>
<tr>
<td>6-311+G(2df,p)</td>
<td>49</td>
<td>167</td>
<td>255</td>
<td>247</td>
<td>289</td>
<td>473</td>
<td>82.0</td>
<td>82.0</td>
<td>172.5</td>
</tr>
<tr>
<td>6-311+G(3df,p)</td>
<td>49</td>
<td>172</td>
<td>277</td>
<td>273</td>
<td>340</td>
<td>490</td>
<td>95.0</td>
<td>95.0</td>
<td>195.6</td>
</tr>
<tr>
<td>6-311+G(3df,3pd)</td>
<td>164</td>
<td>240</td>
<td>309</td>
<td>273</td>
<td>340</td>
<td>490</td>
<td>172.7</td>
<td>172.7</td>
<td>369.7</td>
</tr>
<tr>
<td>6-311++G(d)</td>
<td>10</td>
<td>86</td>
<td>73</td>
<td>288</td>
<td>309</td>
<td>416</td>
<td>14.7</td>
<td>49.1</td>
<td>115.0</td>
</tr>
<tr>
<td>6-311++G(2df,p)</td>
<td>21</td>
<td>76</td>
<td>95</td>
<td>247</td>
<td>289</td>
<td>473</td>
<td>17.0</td>
<td>30.4</td>
<td>84.2</td>
</tr>
<tr>
<td>6-311++G(3df,p)</td>
<td>21</td>
<td>78</td>
<td>99</td>
<td>273</td>
<td>340</td>
<td>490</td>
<td>26.2</td>
<td>39.6</td>
<td>103.5</td>
</tr>
<tr>
<td>6-311++G(3df,3pd)</td>
<td>32</td>
<td>143</td>
<td>110</td>
<td>273</td>
<td>340</td>
<td>490</td>
<td>50.0</td>
<td>52.6</td>
<td>117.3</td>
</tr>
</tbody>
</table>

Table 6.17: PW86PBE-XDM binding energies of rare-gas diatomics for Pople-type basis sets ($\mu E_h$)

<table>
<thead>
<tr>
<th>reference</th>
<th>He₂</th>
<th>He-Ne</th>
<th>He-Ar</th>
<th>Ne₂</th>
<th>Ne-Ar</th>
<th>Ar₂</th>
<th>MPD</th>
<th>MAPD</th>
<th>MaxPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-21G</td>
<td>8</td>
<td>36</td>
<td>22</td>
<td>80</td>
<td>75</td>
<td>168</td>
<td>-61.3</td>
<td>61.3</td>
<td>-78.3</td>
</tr>
<tr>
<td>6-31G</td>
<td>-8</td>
<td>11</td>
<td>22</td>
<td>54</td>
<td>94</td>
<td>219</td>
<td>-74.6</td>
<td>74.6</td>
<td>-122.2</td>
</tr>
<tr>
<td>6-31G(d)</td>
<td>-8</td>
<td>15</td>
<td>22</td>
<td>73</td>
<td>101</td>
<td>240</td>
<td>-70.1</td>
<td>70.1</td>
<td>-122.2</td>
</tr>
<tr>
<td>6-311(2df,p)</td>
<td>3</td>
<td>11</td>
<td>59</td>
<td>32</td>
<td>117</td>
<td>376</td>
<td>-58.2</td>
<td>58.2</td>
<td>-92.1</td>
</tr>
<tr>
<td>6-311G(3df,p)</td>
<td>3</td>
<td>14</td>
<td>69</td>
<td>54</td>
<td>192</td>
<td>399</td>
<td>-46.4</td>
<td>46.4</td>
<td>-92.1</td>
</tr>
<tr>
<td>6-311G(3df,3pd)</td>
<td>34</td>
<td>100</td>
<td>88</td>
<td>54</td>
<td>192</td>
<td>399</td>
<td>-6.2</td>
<td>23.1</td>
<td>-60.1</td>
</tr>
<tr>
<td>3-21+G</td>
<td>8</td>
<td>151</td>
<td>90</td>
<td>215</td>
<td>264</td>
<td>293</td>
<td>16.1</td>
<td>55.2</td>
<td>128.7</td>
</tr>
<tr>
<td>6-31+G</td>
<td>-8</td>
<td>68</td>
<td>56</td>
<td>134</td>
<td>193</td>
<td>334</td>
<td>-32.3</td>
<td>33.5</td>
<td>-122.2</td>
</tr>
<tr>
<td>6-31+G(d)</td>
<td>-8</td>
<td>71</td>
<td>56</td>
<td>137</td>
<td>200</td>
<td>359</td>
<td>-29.9</td>
<td>33.0</td>
<td>-122.2</td>
</tr>
<tr>
<td>6-311+G(2df,p)</td>
<td>3</td>
<td>49</td>
<td>76</td>
<td>104</td>
<td>197</td>
<td>428</td>
<td>-28.7</td>
<td>28.7</td>
<td>-92.1</td>
</tr>
<tr>
<td>6-311+G(3df,p)</td>
<td>3</td>
<td>52</td>
<td>85</td>
<td>120</td>
<td>221</td>
<td>435</td>
<td>-22.0</td>
<td>23.6</td>
<td>-92.1</td>
</tr>
<tr>
<td>6-311+G(3df,3pd)</td>
<td>34</td>
<td>74</td>
<td>91</td>
<td>120</td>
<td>221</td>
<td>435</td>
<td>-3.7</td>
<td>5.7</td>
<td>11.7</td>
</tr>
<tr>
<td>6-31++G(d)</td>
<td>9</td>
<td>58</td>
<td>58</td>
<td>137</td>
<td>200</td>
<td>359</td>
<td>-24.5</td>
<td>25.3</td>
<td>-73.1</td>
</tr>
<tr>
<td>6-31++G(2df,p)</td>
<td>15</td>
<td>45</td>
<td>80</td>
<td>104</td>
<td>197</td>
<td>428</td>
<td>-22.9</td>
<td>22.9</td>
<td>-56.6</td>
</tr>
<tr>
<td>6-31++G(3df,p)</td>
<td>15</td>
<td>47</td>
<td>84</td>
<td>120</td>
<td>221</td>
<td>435</td>
<td>-17.6</td>
<td>19.2</td>
<td>-56.6</td>
</tr>
<tr>
<td>6-31++G(3df,3pd)</td>
<td>24</td>
<td>68</td>
<td>89</td>
<td>120</td>
<td>221</td>
<td>435</td>
<td>-7.3</td>
<td>10.0</td>
<td>-32.4</td>
</tr>
</tbody>
</table>

Table 6.18: CP-corrected PW86PBE-XDM binding energies of rare-gas diatomics for Pople-type basis sets ($\mu E_h$)
and/or polarization function often worsens performance due to basis set imbalance and increased BSSE. Augmentation of small basis sets such as 3-21G or 6-31G with diffuse functions is especially detrimental, leading to massive artificial overbinding from BSSE.

### 6.5.4 Rare-Gas Binding Energies With The Density Fitting Approximation

As shown in Figure 6.16, basis sets of at least augmented triple-zeta quality (e.g., aug-pVDZ) are needed to achieve quantitative accuracy (MAPD below 10% and maximum percent deviations below 20%) for rare-gas binding energies, and approaching the CBS limit requires at least augmented quadruple-zeta basis sets. Such basis sets are easily applied to rare-gas diatomics but become prohibitively expensive for larger intermolecular complexes and other systems with noncovalent interactions of chemical interest, when used in a canonical manner. However, recent improvements in algorithms and computational methods have enabled the application of extended basis sets to systems containing hundreds of atoms. The density fitting approximation [5], also known as “resolution of the identity” for Coulomb integrals (RI-J), is such a method.

Density fitting greatly reduces the cost of calculating the Coulomb part of the total electronic energy (i.e., the classical electron-electron repulsion energy) which is the most computationally expensive part of typical DFT calculations using pure (i.e., non-hybrid) functionals. This is achieved by expanding the electron density \( \rho(r) \) in a set of \( N \) auxiliary basis functions \( \phi_i \)

\[
\rho(r) \approx \bar{\rho}(r) = \sum_{i} c_i \phi_i(r)
\]  

(6.3)

with the fitting coefficients \( c_i \). Through this expansion, four-center Coulomb integrals are approximated by two- and three-center integrals, reducing the formal scaling of the Coulomb part from \( N^4 \) to \( N^3 \) and reducing the computational cost by an order of magnitude or more for typical DFT calculations of molecular systems.

The basis functions \( \phi_i \) are usually atom-centered Gaussian functions just as in standard orbital basis sets. The auxiliary basis set \( \{\phi_i\} \) is typically larger than the orbital basis set in order to minimize the error introduced by the density fitting approximation. It can be automatically constructed [259] from the orbital basis set employed in the calculation, but a number of auxiliary basis sets optimized for use with specific orbital basis sets exist.
Recently, Weigend showed that a single carefully optimized auxiliary basis set can be used in conjunction with orbital basis sets up to quadruple-zeta quality while introducing negligible error \[234\]. This is due to the rather weak basis-set dependence of the electron density, in contrast to the very strong basis-set dependence of the total energy. The W06 auxiliary basis set was constructed to yield errors at least one order of magnitude smaller than the basis-set incompleteness error of the def2-QZVP orbital basis set, corresponding to an average error of less than 50\(\mu E_h\) per atom in the total energies of a set of 300 small molecules (covering the elements H-Rn). It was further shown to introduce negligible error in atomization energies, geometries and molecular properties, and has been employed in computational studies of noncovalent interactions by Grimme and coworkers \[260–264\].

However, the accuracy of the density fitting approximation in conjunction with the W06 auxiliary basis set has not been benchmarked for dispersion interactions in general, or for rare-gas systems in particular. Since rare-gas interaction energies are of the same magnitude (\(\approx 100\mu E_h\)) as the reported W06 total energy deviations for small molecules, a thorough benchmarking of W06 density fitting for rare-gas systems is necessary.

Per-atom deviations in total energies of rare-gas atoms \(\Delta E_{total}^{W06} = E_{total}^{W06} - E_{total}\) and rare-gas diatomics \(\Delta E_{total}^{W06} = E_{total}^{W06} - E_{total}/2\) are shown in Table 6.19. The deviations are of the same magnitude as the per-atom deviations in total energies reported by Weigend for a set of 300 molecules (Table 2 in Ref.). Averaged over basis sets, the largest deviations are found for the neon atom and neon-containing diatomics. However, the largest absolute deviations occur with the SVP basis set for Ar (117\(\mu E_h\)) and Ar_2 (126\(\mu E_h\) per atom), the latter being slightly larger than the worst case (P_4 molecule, def2-SV(P) basis set) found by Weigend (123.2\(\mu E_h\) per atom). Density fitting errors in total energies are rather strongly basis-set dependent, being largest for small basis sets. With triple-zeta and larger basis sets, the deviations are always less than 20\(\mu E_h\) per atom. Notably, W06 combined with the SVP and SVPD basis sets yields by far the largest deviations (for Ne and Ar in particular), whereas the combination of W06 with the comparable 31+(d), pVDZ and aug-pVDZ basis sets is much more accurate.

The deviations in Table 6.19 are significant compared to rare-gas binding energies, potentially jeopardizing the usefulness of the density fitting approximation for rare-gas systems. However, the similarity of the deviations for atoms and diatomics suggests that rare-gas binding energies will benefit from substantial error cancellation. This is easy to rationalize from a chemical viewpoint: As no chemical
<table>
<thead>
<tr>
<th>$\Delta E^{total}$ per atom</th>
<th>atoms</th>
<th>rare-gas diatomics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>He</td>
<td>Ne</td>
</tr>
<tr>
<td>31+(d)</td>
<td>-3</td>
<td>-37</td>
</tr>
<tr>
<td>311+(3dfp)</td>
<td>-3</td>
<td>-17</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>-3</td>
<td>-17</td>
</tr>
<tr>
<td>pVDZ</td>
<td>-3</td>
<td>-52</td>
</tr>
<tr>
<td>pVTZ</td>
<td>-1</td>
<td>-8</td>
</tr>
<tr>
<td>pVQZ</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>pV5Z</td>
<td>-4</td>
<td>-6</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>-3</td>
<td>-42</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>-1</td>
<td>-8</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>-3</td>
<td>-4</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>-13</td>
<td>-42</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>-3</td>
<td>-8</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>-4</td>
<td>-4</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>-4</td>
<td>-7</td>
</tr>
<tr>
<td>SVP</td>
<td>-3</td>
<td>-77</td>
</tr>
<tr>
<td>TZVP</td>
<td>-3</td>
<td>-6</td>
</tr>
<tr>
<td>QZVP</td>
<td>-3</td>
<td>-9</td>
</tr>
<tr>
<td>SVPD</td>
<td>-3</td>
<td>-59</td>
</tr>
<tr>
<td>TZVPD</td>
<td>-3</td>
<td>-5</td>
</tr>
<tr>
<td>QZVPD</td>
<td>-3</td>
<td>-10</td>
</tr>
<tr>
<td>pc-1</td>
<td>-6</td>
<td>-62</td>
</tr>
<tr>
<td>pc-2</td>
<td>-4</td>
<td>-7</td>
</tr>
<tr>
<td>pc-3</td>
<td>-4</td>
<td>-8</td>
</tr>
<tr>
<td>pc-4</td>
<td>-4</td>
<td>-7</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>-10</td>
<td>-51</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>-5</td>
<td>-9</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>-4</td>
<td>-8</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>-4</td>
<td>-7</td>
</tr>
</tbody>
</table>

Table 6.19: Per-atom deviation ($\Delta E^{total}$) of W06 density-fitted PW86PBE-XDM rare-gas total energies from canonical total energies ($\mu E_h$).
bond is formed, the electron density of rare-gas diatomics at equilibrium separation is expected to be very similar to the density of the isolated atoms. Therefore, deviations in total energies due to the density fitting approximation can be expected to systematically cancel out for binding energies.

This hypothesis is proven in Table 6.20, listing the deviations $\triangle BE = BE^{W06} - BE$ of W06 density-fitted PW86PBE-XDM rare-gas binding energies ($BE^{W06}$) from canonical binding energies ($BE$). Significant deviations ($> 2\mu E_h$) are observed only with the smallest (double-zeta) basis sets, in particular pVDZ, aug-pVDZ, SVP, QZVP and pc-1. Since these basis sets (except aug-pVDZ) give rather inaccurate rare-gas BEs without density fitting (c.f. Figure 6.16), the deviation due to the W06 approximation is negligible. The $6\mu E_h$ deviation in He$_2$ with aug-pVDZ is perhaps most noteworthy, but since aug-pVDZ underestimates the He$_2$ BE by $11\mu E_h$ (c.f. Table 6.14), density fitting (fortuitously) improves the agreement with the reference BE. In fact, by comparing the MAPD values with and without density fitting (the last 2 columns in Table 6.20), it can be seen that density fitting yields a slightly better average accuracy for almost all tested basis sets. However, this effect is rather negligible as well as being completely accidental. Either way, for all but the smallest basis sets, the effect of the density fitting approximation on rare-gas binding energies is negligible ($\leq 2\mu E_h$). The deviations of Tables 6.19 and 6.20 were computed with counterpoise correction. W06 total energies, BEs and deviations have also been calculated without CP correction, but are very similar and are therefore omitted.

In conclusion, the density fitting approximation employing the W06 auxiliary basis set introduces noticeable error in total energies but has a negligible effect on the binding energies of rare-gas diatomics. As found by Weigend for total energies, atomization energies and other properties, deviations are largest with the smallest basis sets. This observation reflects one of the design principles of the W06 set, as deviations due to density fitting with double-zeta basis sets are multiple orders of magnitude smaller than the corresponding basis set incompleteness errors, and can therefore be accommodated without introducing additional error.
<table>
<thead>
<tr>
<th>△(BE)</th>
<th>(\text{He}_2)</th>
<th>(\text{He-Ne})</th>
<th>(\text{He-Ar})</th>
<th>(\text{Ne}_2)</th>
<th>(\text{Ne-Ar})</th>
<th>(\text{Ar}_2)</th>
<th>\ MAPD</th>
<th>\ MAPD W06</th>
</tr>
</thead>
<tbody>
<tr>
<td>31+(d)</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>3</td>
<td>-2</td>
<td>2</td>
<td>33.0</td>
<td>33.1</td>
</tr>
<tr>
<td>311+(3dfp)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>-1</td>
<td>0</td>
<td>23.6</td>
<td>23.3</td>
</tr>
<tr>
<td>311++(3df3pd)</td>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>3</td>
<td>-1</td>
<td>0</td>
<td>10.0</td>
<td>8.0</td>
</tr>
<tr>
<td>pVDZ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>79.2</td>
<td>78.0</td>
</tr>
<tr>
<td>pVTZ</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>50.8</td>
<td>50.6</td>
</tr>
<tr>
<td>pVQZ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>22.4</td>
<td>22.1</td>
</tr>
<tr>
<td>pV5Z</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5.9</td>
<td>5.4</td>
</tr>
<tr>
<td>maug-pVDZ</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>36.6</td>
<td>36.1</td>
</tr>
<tr>
<td>maug-pVTZ</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>24.8</td>
<td>24.2</td>
</tr>
<tr>
<td>maug-pVQZ</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>10.3</td>
<td>9.7</td>
</tr>
<tr>
<td>aug-pVDZ</td>
<td>6</td>
<td>-2</td>
<td>2</td>
<td>1</td>
<td>-2</td>
<td>1</td>
<td>11.1</td>
<td>7.0</td>
</tr>
<tr>
<td>aug-pVTZ</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7.4</td>
<td>6.1</td>
</tr>
<tr>
<td>aug-pVQZ</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4.8</td>
<td>4.1</td>
</tr>
<tr>
<td>aug-pV5Z</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4.0</td>
<td>3.6</td>
</tr>
<tr>
<td>SVP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>81.2</td>
<td>77.7</td>
</tr>
<tr>
<td>TZVP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>53.5</td>
<td>53.4</td>
</tr>
<tr>
<td>QZVP</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8.8</td>
<td>8.3</td>
</tr>
<tr>
<td>SVPD</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>25.8</td>
<td>25.9</td>
</tr>
<tr>
<td>TZVPD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6.7</td>
<td>6.5</td>
</tr>
<tr>
<td>QZVPD</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.9</td>
<td>3.3</td>
</tr>
<tr>
<td>pc-1</td>
<td>0</td>
<td>-2</td>
<td>-1</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>79.6</td>
<td>78.8</td>
</tr>
<tr>
<td>pc-2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.4</td>
<td>14.4</td>
</tr>
<tr>
<td>pc-3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.9</td>
<td>3.5</td>
</tr>
<tr>
<td>pc-4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.8</td>
<td>3.4</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td>3</td>
<td>-3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>17.0</td>
<td>16.5</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>13.3</td>
<td>11.5</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.7</td>
<td>3.3</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3.8</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table 6.20: Deviation (△\(BE\)) of W06 density-fitted PW86PBE-XDM rare-gas binding energies from canonical binding energies (\(\mu E_h\)).
6.6 Conclusions

In the first part of this Chapter, the basis-set dependence of rare-gas repulsion energies was investigated using both HF theory and the PW86 exchange GGA. In general, basis sets of at least quadruple-zeta quality are necessary to achieve converged repulsion energies. Double- and triple-zeta basis sets tend to give incorrectly attractive interaction energies due to basis set superposition error. The counterpoise correction scheme is very effective at eliminating BSSE, and converged repulsion energies are obtained with CP-corrected aug-cc-pVDZ or augmented triple-zeta basis sets.

In the second part of this Chapter, the two parameters in the XDM dispersion damping function were fitted to reproduce reference binding energies of rare-gas diatomics at equilibrium separation, using the large aug-cc-pV5Z basis set. Using those damping parameters, the basis-set dependence of rare-gas binding energies calculated with the PW86PBE-XDM method was studied and found to be very similar to the case of repulsion energies. This was shown to be a consequence of the insensitivity of XDM dispersion energies to basis-set effects. XDM dispersion energies are essentially constant (i.e., converged) over a wide range of basis sets, except for certain unaugmented double- and triple-zeta basis sets that yield a slight underestimation.

With a converged basis set, the PW86PBE-XDM method yields rare-gas binding energies within 10% of the reference values. Overall, counterpoise correction and the aug-cc-pVnZ basis sets are recommended for calculations of rare-gas interaction energies using dispersion-corrected DFT methods such as PW86PBE-XDM. Pople-type basis sets behave erratically and cannot be recommended.

Finally, the density fitting approximation was shown to be applicable for rare-gas interactions due to favourable error cancellation, yielding negligible ($\leq 2\mu E_h$) deviations on the binding energies of rare-gas diatomics for all but the smallest basis sets.
Chapter 7

Benchmarking Of The PW86PBE-XDM Method For
Intermolecular Interactions

7.1 Introduction

In the previous Chapter, the PW86PBE-XDM DFA was applied to the dispersion interactions in rare-gas diatomics. Systematic benchmarking of standard Gaussian basis sets revealed that highly accurate repulsion and binding energies can be achieved with basis sets of at least augmented triple-zeta quality. However, good performance in rare-gas systems does not automatically translate to accurate inter- or intramolecular dispersion interactions, as was shown in the past for empirical functionals fitted to rare-gas binding energies [265].

Unlike highly empirical DFAs for NCIs, the PW86PBE-XDM DFA was constructed and parametrized with a minimum of empiricism, and with careful attention to each individual functional component:

1. The PW86 exchange functional has the correct asymptotic behaviour for the exchange enhancement factor \( F(\chi) \) at large reduced gradient \( \chi \) (relevant for the correct description of the vDW region in atomic and molecular systems), and was found to give the best reproduction of exact (HF) exchange repulsion in rare-gas diatomics.

2. The PBE dynamical correlation functional is an appropriate match to PW86 exchange, yielding improved total energies and dispersion coefficients compared to exchange-only methods.

3. The XDM dispersion functional provides highly accurate dispersion coefficients and noncovalent interaction energies when paired with an appropriate exchange-correlation functional.

All three functional components are completely nonempirical by themselves. The two empirical parameters in the dispersion damping function were determined by fitting to accurate reference rare-gas binding energies in the complete basis
set limit, yielding the currently most physically sound, systematically constructed semilocal DFA for NCIs. In previous work using a fully numerical “post-LDA” computational procedure (Chapter 4), we found that rare-gas optimized damping parameters are transferable to other noncovalent interactions, yielding a MAPD of 14.2% on a benchmark set of 65 diverse intermolecular complexes. This is only slightly inferior than a direct fit to the 65 complexes (MAPD of 12.6%). Furthermore, the rare-gas optimized damping parameters yield a significantly better accuracy on the S22 benchmark set of biomolecular interactions (MAD of 0.31 kcal/mol) than the fit to the 65 complexes (MAD 0.46 kcal/mol, see Table 4.4 in Chapter 4).

It is therefore of interest to evaluate the “post-GGA” parametrization of Chapter 6 on standard benchmark sets for noncovalent interactions. The widely-used S22 benchmark set has recently been superseded by the S66 set from the same authors [231]. S66 contains three times as many complexes as S22, and is better balanced with respect to various noncovalent interaction types. S22 is biased towards interactions between aromatic monomers, whereas S66 contains a meaningful number of aliphatic-aliphatic and π-aliphatic interaction motifs. Also, the reference geometries and binding energies of the S66 set are more accurate and consistent than for S22. Finally, the S66x8 [231] and S66a8 [266] sets allow the benchmarking of intermolecular dissociation curves and angular-displaced geometries.

This Chapter is organized as follows: First, the overall accuracy of PW86PBE-XDM for intermolecular binding energies is assessed using the S22 and S66 benchmark sets. Basis set effects are studied in detail, and the method is compared to competing dispersion-corrected DFT methods for noncovalent interactions. In the second part, non-equilibrium geometries are investigated using the S66x8 (intermolecular dissociation curves) and S66a8 (angular-displaced geometries) sets.

### 7.2 Computational Details

All calculations were performed with GAUSSIAN 09 [95] (for the PW86PBE part, self-consistently) and the POSTG interface program (for the XDM part, non-self-consistently). Pople-type orbital basis sets [69] and the W06 density fitting basis set [234] were used as implemented in GAUSSIAN 09. All other basis sets were obtained from the EMSL basis set exchange [96, 97]. The “ultrafine” numerical integration grid (99 radial shells, 590 angular points) was used throughout as implemented in GAUSSIAN 09.
The XDM dispersion energy [50–56] was computed non-self-consistently using an interface program called "POSTG" [215]. The POSTG program uses GAUSSIAN 09 wfn files (containing the Kohn-Sham MOs) to calculate XDM dispersion energies, but is limited to basis functions up to \( f \) angular momentum. Therefore, basis functions of \( g \) and higher angular momentum were discarded by using the keyword `IOp(3/86=4)` in GAUSSIAN 09. This only affects quadruple-zeta and larger basis sets, and is expected to have a negligible influence on binding energies since DFT is known to be relatively insensitive to the presence of higher angular momentum polarization functions in the basis set.

The POSTG program is currently not equipped to handle the “ghost” basis functions used in the counterpoise scheme [211]. CP-corrected PW86PBE-XDM binding energies are thus obtained by adding the XDM dispersion energy calculated with the POSTG program (without CP correction, using PW86PBE orbitals) to the CP-corrected PW86PBE binding energy calculated with GAUSSIAN 09. As the XDM dispersion energy is very insensitive to basis set effects including BSSE, this scheme is not expected to introduce a significant error. Intermolecular binding energies are given in kcal/mol (1 kcal/mol = 4.184 kJ/mol).

### 7.3 Intermolecular Interactions At Equilibrium Geometries

#### 7.3.1 Overall Accuracy Of PW86PBE-XDM For The S22 And S66 Benchmark Sets

Standard semilocal DFAs fail to describe the broad range of noncovalent interactions consistently. For hydrogen bonds the PW86PBE GGA is qualitatively correct but systematically underestimates the interaction strength, as shown in Tables 7.1 and 7.6 (Appendix) for the S22 and S66 benchmark sets, respectively. Strong hydrogen bonds involving water, alcohols, carboxylic acids and nucleic acid base pairs are underbound by up to 16%, whereas the strength of hydrogen bonds involving Methylamine (S66 #9–12) and a peptide model (N-methylacetamide, S66 #13–16) is underestimated by 20-40%. Addition of the XDM dispersion correction yields very accurate hydrogen bond energies, with typical deviations from the reference below 5% (MAPD = 2–3.7%) and a maximum deviation of 10%. Since the empirical parameters in the XDM damping function were not fitted to any hydrogen-bonded systems, this is very gratifying. Apparently, XDM captures very accurately the dispersion contribution to the hydrogen bond strength, which—although small—is
<table>
<thead>
<tr>
<th>S22</th>
<th>$BE^{ref}$</th>
<th>$BE^{PP}$</th>
<th>$%\Delta$</th>
<th>$BE^{PPX}$</th>
<th>$%\Delta$</th>
<th>$BE_{CP}^{PPX}$</th>
<th>$%\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydrogen bonding</td>
<td>MA%D</td>
<td>10.8</td>
<td>2.6</td>
<td>2.0</td>
<td>3.2</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>1 (NH$_3$)$_2$</td>
<td>3.17</td>
<td>2.66</td>
<td>-16.0</td>
<td>3.29</td>
<td>3.7</td>
<td>3.28</td>
<td>3.3</td>
</tr>
<tr>
<td>2 (H$_2$O)$_2$</td>
<td>5.02</td>
<td>4.78</td>
<td>-4.9</td>
<td>5.23</td>
<td>4.3</td>
<td>5.21</td>
<td>3.8</td>
</tr>
<tr>
<td>3 Formic acid dimer</td>
<td>18.80</td>
<td>17.62</td>
<td>-6.3</td>
<td>19.27</td>
<td>2.5</td>
<td>19.04</td>
<td>1.3</td>
</tr>
<tr>
<td>4 Formamide dimer</td>
<td>16.12</td>
<td>14.50</td>
<td>-10.0</td>
<td>16.24</td>
<td>0.7</td>
<td>16.04</td>
<td>-0.5</td>
</tr>
<tr>
<td>5 Uracil dimer</td>
<td>20.69</td>
<td>18.35</td>
<td>-11.3</td>
<td>20.62</td>
<td>-0.3</td>
<td>20.24</td>
<td>-2.2</td>
</tr>
<tr>
<td>6 2-pyridoxine···2-aminopyridine</td>
<td>17.00</td>
<td>15.04</td>
<td>-11.5</td>
<td>17.74</td>
<td>4.3</td>
<td>17.41</td>
<td>2.4</td>
</tr>
<tr>
<td>7 Adenine···thymine (base pair)</td>
<td>16.74</td>
<td>14.15</td>
<td>-15.5</td>
<td>17.09</td>
<td>2.1</td>
<td>16.66</td>
<td>-0.5</td>
</tr>
<tr>
<td>dispersion</td>
<td>MA%D</td>
<td>110.4</td>
<td>8.4</td>
<td>6.6</td>
<td>MA%D</td>
<td>55.2</td>
<td>4.8</td>
</tr>
<tr>
<td>8 (CH$_4$)$_2$</td>
<td>0.53</td>
<td>0.00</td>
<td>-111.7</td>
<td>0.60</td>
<td>14.0</td>
<td>0.58</td>
<td>10.2</td>
</tr>
<tr>
<td>9 (C$_2$H$_4$)$_2$</td>
<td>1.50</td>
<td>0.19</td>
<td>-87.1</td>
<td>1.56</td>
<td>4.0</td>
<td>1.49</td>
<td>-0.9</td>
</tr>
<tr>
<td>10 Benzene···CH$_4$</td>
<td>1.45</td>
<td>-0.04</td>
<td>-102.6</td>
<td>1.61</td>
<td>10.8</td>
<td>1.51</td>
<td>4.0</td>
</tr>
<tr>
<td>11 Benzene dimer ($\pi$–$\pi$)</td>
<td>2.62</td>
<td>-1.78</td>
<td>-167.8</td>
<td>3.16</td>
<td>20.6</td>
<td>2.95</td>
<td>12.7</td>
</tr>
<tr>
<td>12 Pyrazine dimer ($\pi$–$\pi$)</td>
<td>4.20</td>
<td>-0.64</td>
<td>-115.2</td>
<td>4.19</td>
<td>-0.2</td>
<td>3.96</td>
<td>-5.7</td>
</tr>
<tr>
<td>13 Uracil dimer ($\pi$–$\pi$)</td>
<td>9.74</td>
<td>2.91</td>
<td>-70.2</td>
<td>9.43</td>
<td>-3.2</td>
<td>9</td>
<td>-7.6</td>
</tr>
<tr>
<td>14 Indole···benzene ($\pi$–$\pi$)</td>
<td>4.59</td>
<td>-1.98</td>
<td>-143.2</td>
<td>5.00</td>
<td>8.9</td>
<td>4.69</td>
<td>2.1</td>
</tr>
<tr>
<td>15 Adenine···thymine ($\pi$–$\pi$)</td>
<td>11.66</td>
<td>1.72</td>
<td>-85.2</td>
<td>11.04</td>
<td>-5.3</td>
<td>10.52</td>
<td>-9.8</td>
</tr>
<tr>
<td>mixed interactions</td>
<td>MA%D</td>
<td>55.2</td>
<td>4.8</td>
<td>6.6</td>
<td>MA%D</td>
<td>55.2</td>
<td>4.8</td>
</tr>
<tr>
<td>16 Ethene···ethyne</td>
<td>1.51</td>
<td>1.10</td>
<td>-27.4</td>
<td>1.83</td>
<td>21.5</td>
<td>1.74</td>
<td>15.0</td>
</tr>
<tr>
<td>17 Benzene···H$_2$O (OH···$\pi$)</td>
<td>3.29</td>
<td>1.85</td>
<td>-43.7</td>
<td>3.31</td>
<td>0.5</td>
<td>3.21</td>
<td>-2.3</td>
</tr>
<tr>
<td>18 Benzene···NH$_3$ (NH···$\pi$)</td>
<td>2.32</td>
<td>0.80</td>
<td>-65.7</td>
<td>2.37</td>
<td>2.0</td>
<td>2.28</td>
<td>-1.9</td>
</tr>
<tr>
<td>19 Benzene···HCN (CH···$\pi$)</td>
<td>4.55</td>
<td>2.67</td>
<td>-41.4</td>
<td>4.48</td>
<td>-1.5</td>
<td>4.3</td>
<td>-5.5</td>
</tr>
<tr>
<td>20 Benzene dimer (TS)</td>
<td>2.71</td>
<td>0.06</td>
<td>-97.7</td>
<td>2.74</td>
<td>1.3</td>
<td>2.54</td>
<td>-6.1</td>
</tr>
<tr>
<td>21 Indole···benzene (TS)</td>
<td>5.62</td>
<td>1.92</td>
<td>-65.9</td>
<td>5.46</td>
<td>-2.9</td>
<td>5.18</td>
<td>-7.9</td>
</tr>
<tr>
<td>22 Phenol dimer</td>
<td>7.09</td>
<td>3.92</td>
<td>-44.7</td>
<td>6.81</td>
<td>-4.0</td>
<td>6.56</td>
<td>-7.5</td>
</tr>
<tr>
<td>full S22 set</td>
<td>MD</td>
<td>-2.78</td>
<td>0.10</td>
<td>-0.12</td>
<td>MAD</td>
<td>2.78</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>MA%D</td>
<td>-61.1</td>
<td>3.8</td>
<td>-0.2</td>
<td>MA%D</td>
<td>61.1</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>Max%D</td>
<td>-167.8</td>
<td>21.4</td>
<td>#16</td>
<td>#15</td>
<td>#16</td>
<td>#7</td>
</tr>
</tbody>
</table>

Table 7.1: Binding energies (kcal/mol), percent deviations ($\%\Delta$), and error statistics for the S22 benchmark set with PW86PBE/aug-cc-pVTZ ($BE^{PP}$), PW86PBE-XDM/aug-cc-pVTZ ($BE^{PPX}$) and CP-corrected PW86PBE-XDM/aug-cc-pVTZ ($BE_{CP}^{PPX}$).
needed for quantitative accuracy.

PW86PBE fails completely for dispersion-bound complexes, yielding negative binding energies (i.e., repulsion) for all dispersion interactions involving aliphatic molecules and most \( \pi-\pi \) interactions. PW86PBE-XDM corrects this qualitative failure. \( \pi-\pi \) interactions (#9, #11–15 in S22, #24–33 in S66) are accurately described on average, with the most significant deviations occurring for the stacked (parallel displaced) benzene dimer, benzene-ethene, and pyridine-ethene complexes. Aliphatic-aliphatic (#8 in S22, #34–38 in S66) and \( \pi \)-aliphatic (#10 in S22, #39–46 in S66) interactions are systematically overestimated, but the consistently improved agreement with reference values upon counterpoise correction suggests that this is partly a BSSE effect.

Finally, the “mixed” and “other” categories of the S22 and S66 sets feature important interaction motifs such as XH-\( \pi \) (X=C,N,O) and T-shaped (TS) complexes, having varying contributions to the total interaction strength from electrostatics, induction, and dispersion. The strength of these mixed interactions is strongly underestimated with the PW86PBE GGA by itself. Including XDM dispersion again yields quantitative accuracy, with most deviations well below 10%. Notable outliers are the ethene-ethyne (S22 #16) and ethyne-ethyne (S66 #51) complexes, whose large percent deviations (\( \approx 20\% \)) are partly due to the weak interaction strength and partly due to BSSE, being reduced to \( \approx 15\% \) upon counterpoise correction.

BSSE effects are also reflected in the positive mean deviations and mean percent deviations of the uncorrected PW86PBE-XDM/aug-cc-pVTZ method (column 6 in Tables 7.1 and 7.6). With counterpoise correction these values are closer to zero or negative, demonstrating that BSSE has a significant effect (i.e., systematic overbinding) even with the relatively large aug-cc-pVTZ basis set.

Overall, PW86PBE-XDM accurately and consistently reproduces the intermolecular binding energies of the S22 and S66 benchmark sets. Better than “chemical accuracy” (1 kcal/mol) is achieved for all systems except the stacked adenine-thymine base pair in the S22 set, with average deviations around 0.25 kcal/mol (=1 kJ/mol). Compared to the PW86PBE GGA, both MAD and MA%D are reduced by an order of magnitude.

The method is compared to various other DFAs for noncovalent interactions in Table 7.2 and Figure 7.1. Among the dispersion-corrected functionals in Table 7.2, VV10 and vdW-DF2 stand out as genuine nonlocal correlation functionals [40].
Table 7.2: Mean absolute deviations (kcal/mol) of various DFAs for the S22 and S66 benchmarks sets, and the three S66 subcategories (hydrogen bonding, “HB”; dispersion, “disp”; and other interactions, “other”). *: S22 MADs (S66 values not reported)

vdW-DF2 [46] is a seamless (i.e., no explicit damping function) functional without empirically fitted parameters, although the gradient correction in the nonlocal vdW-DF2 correlation kernel was adjusted to obtain more accurate binding energies for the S22 set compared to the original vdW-DF [43–45] functional. VV10 [48] is a modification of vdW-DF that features an explicit damping mechanism, with an empirical parameter fitted to the S22 set.

The -D3 method of Grimme and coworkers is a more empirical dispersion correction that employs high-level ab initio precomputed dispersion coefficients and cutoff radii. Two empirical parameters fitted to a set containing S22, conformational energies, alkane and rare-gas dimers define the -D3 damping function [229]. The -D3 method has been parametrized for a wide range of DFAs,
Figure 7.1: S66 and S22 mean absolute deviations (MAD) of various DFAs for non-covalent interactions ranging from simple GGAs and hybrids to the most recent range-separated hybrid (RSH) and double-hybrid (involving MP2-like correlation) functionals. The \(-D3(BJ)\) method is a variation of \(-D3\) using the XDM damping function of Becke & Johnson, found to be beneficial for noncovalent interactions and thermochemistry involving “medium-range” correlation [263].

The “TS” method [185] derives effective dispersion coefficients from free-atom reference values, the empirical parameter in the damping function is fitted to S22 data. The M0x [214, 270] functionals of Truhlar and coworkers, their \(-D3\) corrected variants [229], and the \(\omega\)B97X-D2 [195, 247] method of Chai and Head-Gordon are included as recent examples of empirically parametrized DFAs for noncovalent interactions. Finally, B3LYP-DCP [269] (dispersion-correcting potential) uses an empirical pseudopotential to mimic dispersion and was fitted to a set of small intermolecular complexes partly taken from the S66x8 set.
The data in Table 7.2 are graphically represented in Figure 7.1, where the length of each colored bar represents the MAD for a subset of approximately 22 intermolecular complexes, thereby giving equal weight to the S22 and S66 sets. PW86-PBE-XDM is one of the most accurate DFAs for noncovalent interactions. Among the GGA-based functionals, only BLYP-D3(BJ) is slightly better, at the cost of a higher level of empiricism. Other functionals that perform similar to PW86PBE-XDM are of the hybrid meta-GGA, range-separated hybrid, and double hybrid type, and come with a higher computational cost. In general, Grimme’s -D3 correction and Truhlar’s M06-2X functional yield accurate intermolecular binding energies. vdW-DF2, M06-L and M05-2X are significantly less accurate and cannot be recommended.

Interestingly, rPW86PBE-VV10 performs slightly worse than PW86PBE-XDM despite being parametrized on the S22 set, specifically overbinding hydrogen bonds [268]. No such overbinding is seen with LC-ωPBE-VV10, leading Vydrov and van Voorhis to attribute this deficiency to the rPW86 exchange component [268]. However, our PW86PBE-XDM method is as accurate for hydrogen bonds as LC-ωPBE-VV10 (despite being parametrized on rare-gas systems only), but less accurate than both VV10 methods for dispersion interactions. This finding contradicts Vydrov and van Voorhis’ reasoning, and points to a delicate balance between the underlying exchange-correlation functional and the dispersion component in the PW86PBE-XDM and rPW86PBE-VV10 methods.

7.3.2 Basis Set Effects

The benchmarking of the PW86PBE-XDM method in the previous section utilized the aug-cc-pVTZ basis set, which is generally known to be adequate for semilocal DFT calculations [271], and has been used extensively in the development, benchmarking, and application of DFT methods for noncovalent interactions [213, 268, 269, 272, 273]. Since this thesis aims to produce an essentially nonempirical density-functional approximation, it is necessary to further investigate the basis set dependence of PW86PBE-XDM binding energies, and to assess the method in the complete basis set (CBS) limit. This will be the focus of the current section, using the basis sets listed in Tables 2.1–2.3.

Mean absolute deviations for the S66 set are plotted in Figure 7.2 by basis set. MADs both with and without counterpoise correction are shown, the difference representing the BSSE for a given basis set averaged over the S66 complexes. The
Figure 7.2: CP-corrected and uncorrected PW86PBE-XDM binding energy MAD (S66)
order of basis sets follows Tables 2.1–2.3, except that heavy- and fully augmented basis sets are shown together due to their similar MADs. From top to bottom, basis sets are grouped in the following categories: 1. unaugmented double zeta, 2. diffuse-augmented double zeta, 3. unaugmented triple zeta, 4. diffuse-augmented triple zeta, 5. unaugmented quadruple zeta and larger, and 6. diffuse-augmented quadruple zeta and larger.

Looking first at CP-corrected MADs, it is evident that unpolarized basis sets (i.e., 3-21G, 6-31G, and diffuse-augmented variants) fail with MADs $> 0.7$ kcal/mol. Apparently, at least a set of “d-type” polarization functions on first-row atoms is necessary to properly describe the polarization of electron density due to noncovalent interactions. With such polarization functions, unaugmented double zeta basis sets yield reasonable MADs around 0.5 kcal/mol. Augmentation of double zeta basis sets with diffuse functions on non-hydrogen atoms (e.g., 6-31+G(d)) yields a further decrease in MAD to 0.35 kcal/mol, although this accuracy is also achieved (perhaps accidentally) with the pc-1 basis set.

Note that diffuse (e.g., 6-31++G(d)) and polarization (e.g., 6-31+G(d,p)) functions on hydrogen atoms have no discernable effect on the S66 MAD, indicating that polarization of hydrogen atom electron densities in noncovalent interactions is either negligible or sufficiently captured by basis functions on neighbouring atoms. However, inclusion of two sets of “d-type” polarization functions on first-row atoms (e.g., 6-31+G(2d, 2p)) noticeably increases accuracy. Indeed, the 6-31+G(2d, 2p), def2-SVPD, (maug/haug/aug)-cc-pVDZ, and aug-pc-1 basis sets have MADs of $< 0.3$ kcal/mol, close to the CBS limit of 0.24 kcal/mol. Triple zeta and larger basis sets are essentially converged to the CBS limit with CP correction, as long as 2d polarization functions are present on heavy atoms.

It is obvious from Figure 7.2 that basis set convergence to the CBS limit is much slower without counterpoise correction, demonstrating the large influence of BSSE on intermolecular interaction energies. Contrary to literature reports claiming a sometimes detrimental effect of CP correction on S22 interaction energies [237], we find that CP correction always improves accuracy, often dramatically. With very few exceptions, double zeta and unaugmented triple zeta basis sets suffer from significant BSSE and cannot be recommended for noncovalent interactions without CP correction. Only 6-31+G(2d,2p), maug-cc-pVDZ, haug-cc-pVDZ, def2-TZVP and pc-2 have a reasonable MAD around or below 0.5 kcal/mol, but even here

\footnote{The cited paper benchmarked standard semilocal DFAs without dispersion corrections in conjunction with the 6-31+G(d,p) and 6-311+G(2df,2p) basis sets on the S22 set.}
accuracy can be improved by a factor of two with CP correction. Note also that
diffuse functions on hydrogen atoms increase BSSE for double zeta basis sets, as
can be seen by comparing 6-31+G(2d,2p) to 6-31++G(2d,2p) or (maug/haug)-cc-
pVDZ to aug-cc-pVDZ. Diffuse-augmented triple zeta and larger basis sets have
negligible BSSE and are converged to the CBS limit even without CP correction,
exceptions being the Pople-type 6-311+G basis sets and cc-pVQZ.

For completeness, S22 mean absolute deviations by basis set are shown in Fig-
ure 7.18. They are very similar to the S66 values but slightly larger in magnitude
and less sensitive to BSSE, both artifacts of the dominance of strongly-bound dou-
bble hydrogen bonds and nucleic acid complexes in S22. In contrast, the S66 set
has a better balance of binding energies and features more aliphatic-aliphatic and
π-aliphatic dispersion interactions which were essentially excluded from S22.

Overall, PW86PBE-XDM yields very accurate intermolecular binding energies
with augmented double zeta (CP-corrected) and triple zeta or larger (with or with-
out CP correction) basis sets. This is achieved without compensating for basis set
effects in the parametrization of the XDM damping function and without fitting to
the S22 or S66 sets. With dispersion damping parameters optimized on rare-gas diatomics
only, the PW86PBE-XDM method is applicable to the full range of intermolecular
noncovalent interactions, including weak and strong hydrogen bonds, dispersion
(π-π, π-aliphatic, aliphatic-aliphatic) and interactions of mixed electrostatic/induc-
tion/dispersion character. The CBS limit damping parameters are applicable with
a wide range of Gaussian basis sets. Due to the good performance of many stan-
dard basis sets, it is difficult to recommend any particular one on the basis of the
S66 MAD alone.

To make such recommendations, it is helpful to examine the correlation be-
tween accuracy and computational expense for the tested basis sets. Table 7.3 lists
the number of contracted Gaussian basis functions and number of Gaussian prim-
itives per dimer complex (averaged over the S66 set), total CPU time for the single-
point energy calculation of the S66 set (dimers and monomers) with a given basis
set, and the mean absolute deviation (without and with counterpoise correction).
Excluded from Table 7.3 are fully augmented (++), 6-31G(d,p) and 6-31G(3df,3pd)
Pople-type basis sets, due to their inferior accuracy. The basis sets are ordered as
in Figure 7.2.

As the number of basis functions increases, so does the CPU time. Formally,
density-fitted DFT calculations scale as $O(N^3)$, where $N$ is the number of basis
functions [234, 259]. Due to the relatively small size of the S66 complexes and
<table>
<thead>
<tr>
<th>basis functions</th>
<th>type</th>
<th>Gaussian primitives</th>
<th>CPU time</th>
<th>MAD [kcal/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-21G</td>
<td>DZ</td>
<td>101</td>
<td>164</td>
<td>1</td>
</tr>
<tr>
<td>6-31G</td>
<td>DZ</td>
<td>101</td>
<td>237</td>
<td>2</td>
</tr>
<tr>
<td>6-31G(d)</td>
<td>DZ</td>
<td>153</td>
<td>289</td>
<td>3</td>
</tr>
<tr>
<td>def2-SVP</td>
<td></td>
<td>178</td>
<td>296</td>
<td>5</td>
</tr>
<tr>
<td>cc-pVDZ</td>
<td></td>
<td>178</td>
<td>366</td>
<td>5</td>
</tr>
<tr>
<td>pc-1</td>
<td></td>
<td>178</td>
<td>331</td>
<td>5</td>
</tr>
<tr>
<td>3-21+G</td>
<td>aug-DZ</td>
<td>136</td>
<td>199</td>
<td>3</td>
</tr>
<tr>
<td>6-31+G</td>
<td>DZ</td>
<td>136</td>
<td>272</td>
<td>4</td>
</tr>
<tr>
<td>6-31+G(d)</td>
<td></td>
<td>188</td>
<td>324</td>
<td>6</td>
</tr>
<tr>
<td>6-31+G(2d,2p)</td>
<td></td>
<td>308</td>
<td>443</td>
<td>10</td>
</tr>
<tr>
<td>def2-SVP</td>
<td></td>
<td>268</td>
<td>395</td>
<td>12</td>
</tr>
<tr>
<td>min-aug-cc-pVDZ</td>
<td></td>
<td>213</td>
<td>401</td>
<td>8</td>
</tr>
<tr>
<td>heavy-aug-cc-pVDZ</td>
<td></td>
<td>257</td>
<td>454</td>
<td>11</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td></td>
<td>301</td>
<td>498</td>
<td>16</td>
</tr>
<tr>
<td>aug-pc-1</td>
<td></td>
<td>301</td>
<td>463</td>
<td>16</td>
</tr>
<tr>
<td>6-311G(d,p)</td>
<td>TZ</td>
<td>224</td>
<td>369</td>
<td>7</td>
</tr>
<tr>
<td>6-311G(2df,p)</td>
<td></td>
<td>329</td>
<td>508</td>
<td>13</td>
</tr>
<tr>
<td>6-311G(2df,2pd)</td>
<td></td>
<td>418</td>
<td>609</td>
<td>21</td>
</tr>
<tr>
<td>def2-TZVP</td>
<td></td>
<td>338</td>
<td>535</td>
<td>15</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td></td>
<td>418</td>
<td>719</td>
<td>24</td>
</tr>
<tr>
<td>pc-2</td>
<td></td>
<td>418</td>
<td>681</td>
<td>19</td>
</tr>
<tr>
<td>6-311+G(d,p)</td>
<td>aug-TZ</td>
<td>259</td>
<td>404</td>
<td>9</td>
</tr>
<tr>
<td>6-311+G(2df,p)</td>
<td></td>
<td>364</td>
<td>543</td>
<td>21</td>
</tr>
<tr>
<td>6-311+G(2df,2pd)</td>
<td></td>
<td>453</td>
<td>644</td>
<td>25</td>
</tr>
<tr>
<td>def2-TZVPD</td>
<td></td>
<td>427</td>
<td>633</td>
<td>26</td>
</tr>
<tr>
<td>min-aug-cc-pVTZ</td>
<td></td>
<td>453</td>
<td>679</td>
<td>24</td>
</tr>
<tr>
<td>heavy-aug-cc-pVTZ</td>
<td></td>
<td>558</td>
<td>818</td>
<td>39</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td></td>
<td>659</td>
<td>1005</td>
<td>61</td>
</tr>
<tr>
<td>aug-pc-2</td>
<td></td>
<td>659</td>
<td>968</td>
<td>59</td>
</tr>
<tr>
<td>def2-QZVP</td>
<td>QZ</td>
<td>754</td>
<td>1097</td>
<td>49</td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td></td>
<td>737</td>
<td>1180</td>
<td>56</td>
</tr>
<tr>
<td>pc-3</td>
<td></td>
<td>860</td>
<td>1284</td>
<td>58</td>
</tr>
<tr>
<td>def2-QZVPD</td>
<td>aug-QZ</td>
<td>844</td>
<td>1195</td>
<td>64</td>
</tr>
<tr>
<td>min-aug-cc-pVQZ</td>
<td></td>
<td>772</td>
<td>1094</td>
<td>53</td>
</tr>
<tr>
<td>heavy-aug-cc-pVQZ</td>
<td></td>
<td>877</td>
<td>1355</td>
<td>86</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td></td>
<td>1055</td>
<td>1579</td>
<td>131</td>
</tr>
<tr>
<td>aug-pc-3</td>
<td></td>
<td>1179</td>
<td>1682</td>
<td>137</td>
</tr>
<tr>
<td>cc-pV5Z</td>
<td>5Z</td>
<td>1055</td>
<td>1662</td>
<td>101</td>
</tr>
<tr>
<td>pc-4</td>
<td></td>
<td>1302</td>
<td>1831</td>
<td>121</td>
</tr>
<tr>
<td>aug-cc-pV5Z</td>
<td>aug-5Z</td>
<td>1374</td>
<td>2060</td>
<td>202</td>
</tr>
<tr>
<td>aug-pc-4</td>
<td></td>
<td>1621</td>
<td>2230</td>
<td>220</td>
</tr>
</tbody>
</table>

Table 7.3: Basis functions, Gaussian primitives, CPU time and S66 MAD (see text)
integral screening techniques employed by GAUSSIAN 09 [95], less than quadratic scaling with the number of basis functions, $O(N^{1.6})$, is actually observed (Figure 7.3).

Counterpoise correction roughly doubles the required CPU time with a given basis set, as the monomer systems are calculated in the full dimer basis set, as opposed to the much smaller monomer basis set without CP correction. For basis sets up to triple zeta quality this increase in computational cost is clearly warranted, since the S66 MAD is reduced by a factor of 2–6. For augmented triple zeta basis sets this factor is between 1 and 2, and approaches 1 for even larger basis sets, making counterpoise correction uneconomical and redundant.

As the computational expense (as measured by CPU time) increases, the MAD for the S66 set decreases. However, this inverse correlation is highly nonlinear as shown in Figure 7.4. Of the Pople-type basis sets listed in Table 7.3, only 6-31+G(d) and 6-31+G(2d,2p) are shown in Figure 7.4, as these have the best accuracy:expense ratio among Pople-type basis sets. On the left side of Figure 7.4 the double zeta basis sets (without CP correction) form a discernable cluster of widely varying, but generally large, MAD values. The bottom right contains quadruple zeta and CP-corrected triple zeta basis sets, having low MAD values at high computational cost. Basis sets that are both accurate and computationally inexpensive appear in the bottom left corner, and are shown in more detail in the inset. Counterpoise-corrected double zeta basis sets have the overall best accuracy:expense ratio, in particular pc-1, 6-31+G(d), 6-31+G(2d,2p), min-aug-cc-pVDZ (maDZ), and heavy-aug-cc-pVDZ (haDZ). As counterpoise correction is not always practical or possible, the triple zeta basis sets pc-2, def2-TZVPD, min-aug-cc-pVTZ (maTZ), and heavy-aug-cc-pVTZ (haTZ) are recommended without CP correction. Note that these are much cheaper than the def2-QZVP basis recommended by Grimme or aug-cc-pVTZ (aTZ, used in Chapter 5), at a comparable level of accuracy.

### 7.3.3 Basis Set Effects By Interaction Type

So far, only mean absolute deviations have been considered in the analysis of basis set effects for the S66 and S22 benchmark sets. MAD values are a good measure of overall accuracy, but are insensitive to outliers and do not allow an analysis of basis set effects by interaction type. With the widespread adoption of the S22 benchmark set for noncovalent interactions, it has become customary [46,199,245,256,274–279] to plot binding energy deviations ($BE - BE_{ref}$) against the S22 complex number,
Figure 7.3: Correlation between CPU time and number of basis functions
Figure 7.4: Correlation between S66 MAD and CPU time
enabling a detailed intuitive assessment of a method’s accuracy.

Deviations of PW86PBE-XDM binding energies from S22 reference values are shown in Figure 7.5 for Pople-type basis sets, and Figure 7.6 for cc-, def2, and pc-type basis sets. Equivalent plots for the S66 set are shown in Figures 7.19–7.22 in the Appendix. All graphs in this subsection use the following conventions:

- unaugmented basis sets are shown with dashed lines and open symbols
- diffuse-augmented basis sets are shown with continuous lines and filled symbols
- double zeta basis sets are shown in orange, red or magenta
- triple zeta basis sets are shown in blue, cyan or steel-blue
- the 6-311++G(3df,3pd) basis set is shown with a blue dotted line and cross symbol

The necessity of “d-type” polarization functions for an accurate description of non-covalent interactions arises chiefly in hydrogen-bonded complexes, as can be seen by the large (≥1 kcal/mol) binding energy deviations with the unpolarized 6-31G and 6-31+G basis sets (orange) in Figures 7.5, 7.19, and 7.20. When corrected for BSSE—which causes artificial overbinding especially with 6-31G(d), as seen in the top panel of Figures 7.5 and 7.19—the polarized 6-31G(d) and 6-31+G(d) basis sets (red) yield reasonable hydrogen bond energies within 1 kcal/mol of the reference. However, two sets of “d-type” polarization functions on first-row atoms as in 6-31+G(2d,2p) (magenta) further reduce deviations for hydrogen bonds by a factor of two, approaching the CBS limit. Diffuse functions are of minor importance, yielding small improvements.

For dispersion-dominated interactions, the opposite basis-set dependence is found: Lack of diffuse functions causes significant underbinding (1–3 kcal/mol), most severely for π-π-stacked systems (#11-15 in S22, #24-33 in S66); see Figure 7.20 and the bottom panel of Figure 7.5. This systematic error is proportional to the interaction strength, leading to deviations >2 kcal/mol for the stacked uracil-uracil (#26 in S66, #13 in S22) and adenine-thymine (#15 in S22) nucleobase complexes. It occurs when counterpoise correction is applied. Without CP correction, fortuitous error cancellation due to lack of diffuse functions (causing underbinding) and BSSE (causing overbinding) results in superficially accurate π-π-stacking energies with the otherwise inappropriate 6-31G and 6-31G(d) basis sets (see Figure
Figure 7.5: Deviation of PW86PBE-XDM binding energies from S22 reference for Pople-type basis sets
7.19 and the top panel of Figure 7.5). Less fortunate error cancellation for aliphatic-aliphatic and \( \pi \)-aliphatic complexes leads to systematic overbinding of 1–2 kcal/mol with these basis sets (Figure 7.19). Polarization functions have a negligible effect on dispersion-bound complexes, as can be seen by the similar performance of 6-31+G and 6-31+G(d) for these systems. This is in accordance with the traditional view of dispersion as a purely nonlocal correlation phenomenon, having negligible effect on the electron density of the interacting fragments.

“Mixed” and “other” interactions are less basis-set-dependent but generally benefit from the presence of diffuse functions, highlighting the importance of dispersion. The phenol dimer (#22 in S22) and ethyne-water complex (#59 in S66) behave like hydrogen bonds.

Due to these two distinct basis set effects—polarization functions necessary for hydrogen bonds, diffuse functions necessary for dispersion-bound complexes—6-31+G(d) is the smallest Pople-type basis sets that yields reasonable accuracy for both types of noncovalent interactions (maximum deviation from reference \( \pm 1.2 \) kcal/mol with CP correction, MAD \( \approx 0.4 \) kcal/mol). An additional set of \( \text{“d-type”} \) polarization functions on first-row atoms further improves hydrogen bond energies. The essentially identical deviations with the 6-31+G(2d,2p) and 6-311++G(3df,3pd) basis sets across all S22 and S66 complexes show that diffuse and polarization functions on hydrogen atoms are unnecessary.

Looking at the CP-corrected cc-, def2, and pc-type basis sets (all of which include polarization function by default) in Figure 7.22 and the bottom panel of Figure 7.6, it is found that triple zeta and augmented double zeta basis sets yield virtually identical binding energies within \( \pm 0.5 \) kcal/mol of the reference values, and are essentially converged to the CBS limit. The unaugmented cc-pVDZ and def2-SVP basis sets underestimate XDM dispersion energies (see Table 6.14 and Figure 6.13), leading to systematic underbinding, most noticeably for \( \pi-\pi \) dispersion-bound complexes. The pc-1 basis set performs much better, which may be traced to its reasonable accuracy for XDM dispersion energies despite the lack of diffuse functions (Figure 6.13). This analysis reinforces and validates the earlier conclusions based on the basis set dependence of S66 MAD values (Figure 7.2), and provides an explanation for the exceptional accuracy of pc-1 compared to other unaugmented double zeta basis sets.

Without counterpoise correction, the (aug-)cc-pVDZ, def2-SVP(D), and (aug-)pc-1 double zeta basis sets show systematic overbinding due to BSSE (Figure 7.21 and top panel of Figure 7.6). The largest BSSE effects (2–6 kcal/mol) are seen with
Figure 7.6: Deviation of PW86PBE-XDM binding energies from S22 reference for cc-, def2, and pc-type basis sets
unaugmented basis sets for hydrogen bonds, whereas dispersion-bound systems are seemingly less affected. However, recalling the underestimation of dispersion energies with these basis sets, this is clearly a result of fortuitous error cancellation. The augmented double zeta basis sets, which performed very accurately for dispersion interactions with CP correction (see Figure 7.22 and the bottom panel of Figure 7.6), show overbinding of 1–2.5 kcal/mol due to BSSE.

The augmented triple zeta basis sets aug-cc-pVTZ, def2-TZVPD, and aug-pc-2 yield virtually identical results for all complexes in the S66 and S22 sets. Without diffuse functions, residual BSSE results in overbinding for hydrogen bonding and “mixed” interactions. The cc-pVTZ basis set is most, def2-TZVP less, and pc-2 is least affected by BSSE. Again, this behaviour was already seen in the S66 MAD values of Figure 7.2 but is now much better understood, in particular the excellent performance of pc-2 without CP correction.

7.4 Intermolecular Interactions At Non-Equilibrium Geometries

The results presented in the previous section show that the PW86PBE-XDM method gives accurate noncovalent interaction energies at the potential energy minima of intermolecular complexes. For practical applications including geometry optimization and molecular dynamics simulation it is important to have accurate interaction energies also at non-equilibrium geometries [231]. Ideally, a method should have consistent accuracy across the entire potential energy surface. Secondly, for complex biomolecular and condensed-matter systems even the global minimum-energy structure may contain local regions of intermolecular interaction that are not at the potential energy minimum for the respective interaction motif. Finally, the asymptotic behaviour of a method becomes important in large systems due to the quadratic increase in pair interactions with the number of atoms.

Highly accurate dissociation curves for selected intermolecular complexes have been available for a few years [177, 280], but large-scale benchmark sets of intermolecular interactions at non-equilibrium geometries have appeared only very recently [231, 266, 281]. The S66x8 set of Rezac and coworkers features the intermolecular complexes of the S66 set at 8 points along the dissociation curve, which
were obtained by scaling the intermolecular distance\textsuperscript{2} of MP2-optimized equilibrium structures\textsuperscript{3} by factors of 0.9, 0.95, 1, 1.05, 1.1, 1.25, 1.5, and 2. The reference (CCSD(T)/CBS) S66x8 interaction energies have an error estimated to be below 3%, corresponding to an error below 0.12 kcal/mol for a typical interaction energy of 4 kcal/mol [231]. Note that the actual S66 equilibrium geometries were obtained by interpolation of the first five points of the S66x8 dissociation curves, and are hence of CCSD(T)/CBS-quality, too.

### 7.4.1 Overall Accuracy Of Intermolecular Dissociation Curves

First, the overall accuracy of PW86PBE-XDM for intermolecular dissociation curves is assessed on the S66x8 set and compared to the DFT-D3 methodology of Grimme and coworkers [229, 263], which is to date the only other DFT method for which S66x8 data are available in the literature [242]. For consistency with the data in [242], the def2-QZVP basis set is employed without CP correction. We only\textsuperscript{4} show results for the “BJ” variant [263] of Grimme’s -D3 scheme which uses the XDM damping function of Becke & Johnson, and was found to perform slightly better for the S66x8 set with most functionals than the original -D3 scheme [242].

The data are presented as mean (absolute) percent deviations (MPD/MAPD) averaged over the 66 complexes for a given distance multiplier. Note that each complex has a different degree of dissociation for a given distance multiplier, due to the difference in equilibrium intermolecular distance, different noncovalent interaction types, and varying definitions of the intermolecular dissociation coordinate [242]. Still, analysis of MPD and MAPD values by distance multiplier enables an intuitive assessment of a method’s accuracy for compressed, near-equilibrium, and stretched geometries.

Mean absolute percent deviations (MAPD) for PW86PBE-XDM as well as -D3 and M0X-type functionals are shown in Figure 7.7. Due to the large number of functionals benchmarked by Goerigk et al., individual Figures are shown for GGA, hybrid, M0X, and double hybrid functionals. All Figures use the same scale, and legend entries are ordered by decreasing MAPD for large (1.5 and 2) distance multipliers in order to facilitate interpretation of the plots.

---

\textsuperscript{2}defined as the closest intermolecular distance in the complex along an axis defined by the centers of mass of the monomers, except hydrogen-bonded systems, where the hydrogen and hydrogen-acceptor atoms define the intermolecular axis.

\textsuperscript{3}optimized at the MP2/cc-pVTZ level with CP correction

\textsuperscript{4}except for M0X-type functionals, for which only the original -D3 parametrization is available
Figure 7.7: S66x8 mean absolute percent deviation (MAPD) by distance multiplier for various methods with the def2-QZVP basis set.
PW86PBE-XDM is clearly one of the most accurate methods for intermolecular dissociation curves. At compressed and near-equilibrium geometries, it has similar accuracy as D3-corrected (meta-)GGA and M0X functionals, but is much more accurate at stretched geometries. Hybrid and double-hybrid functionals are slightly better than PW86PBE-XDM at compressed and near-equilibrium geometries but become consistently less accurate with increasing intermolecular distance. At twice the equilibrium distance (multiplier 2), PW86PBE-XDM has a MAPD of 6% for the 66 complexes whereas -D3-corrected functionals have MAPDs of 13–38%. Truhlar’s M0X-type functionals without dispersion correction⁵ fail completely for stretched geometries with MAPDs of 63-77% (off scale). Thus, PW86PBE-XDM is at present the only DFT method that can describe the dissociation of intermolecular complexes with consistent accuracy at both equilibrium and stretched geometries, with MAPDs < 10%. Only for the most compressed geometries (multiplier 0.9) does its accuracy deteriorate significantly (MAPD 19%). This is also seen with all other (meta-)GGA based methods, whereas hybrid and double-hybrid functionals fare better. The origin of this effect is investigated below, and can be traced to a small number of outliers among the S66 complexes.

Goerigk et al. have ascribed the decreasing accuracy of DFT-D3 methods with increasing separation to double-counting effects, as both dispersion and “medium-range” correlation modeled by the underlying DFT functional contribute to the binding energy in this regime. This can be seen in plots of the signed mean percent deviation (MPD, Figure 7.8), where most dispersion-corrected functionals are found to be systematically overbound at near-equilibrium and stretched geometries. PW86PBE-XDM overbinds by 4–7%. Grimme’s hybrid- and (meta-)GGA-based methods overbind more strongly by up to 30%, having a curious peak in the MPD curve at multiplier 1.5. Exceptions are the BLYP-D3(BJ), B3LYP-D3(BJ), and double-hybrid *LYP-D3(BJ) methods. Goerigk et al. ascribed this to the presence of LYP correlation, but did not provide an explanation. It is worth noting that LYP correlation yields binding in rare-gas dimers even with a repulsive exchange component, and despite the lack of dispersion physics [282]. Therefore, one would expect dispersion-corrected methods based on LYP correlation to be especially susceptible to double-counting effects. However, a second commonality of the LYP-containing functionals used by Goerigk et al. is the presence of the B88 exchange functional (“B”). B88 exchange is notoriously overrepulsive for van der

⁵which are noticeably less accurate than dispersion-corrected methods at compressed and near-equilibrium geometries
Figure 7.8: S66x8 mean percent deviation (MPD) by distance multiplier for various methods with the def2-QZVP basis set. The same color scheme as in Figure 7.7 is used.
Waals interactions even at large separations (up to 4 Å, see Figure 3.1 in Chapter 3), apparently offsetting the double-counting of correlation by LYP and -D3 dispersion. Overall, this leads to fortuitously accurate B98LYP-D3 S66x8 dissociation curves with only a small overbinding effect near equilibrium (see Figure 7.8). Still, these methods have at least twice the MAPD of PW86PBE-XDM at multiplier 2. Note that the other DFT-D3 methods tested by Goerigk et al. contain neither B88 exchange nor LYP correlation. The exchange functional is also likely responsible for the “spike” in MPD at multiplier 1.5 with PW6B95-D3(BJ) and PWPB95-D3(BJ), as the PW91 exchange component is known to yield massive artificial binding in rare-gas systems, leading to overbinding in stretched intermolecular complexes when combined with the -D3 dispersion correction. Finally, the large MAPDs and much smaller MPDs at multiplier 2—especially with hybrid and double hybrid functionals—indicate that the error of -D3 methods at large separation is unsystematic, and cannot be easily improved.

PW86PBE-XDM was specifically constructed to contain physically reasonable, nonempirical functionals for exchange, dynamical correlation, and dispersion. It achieves consistently high accuracy without relying on empiricism or fortuitous error cancellation, as is the case with Grimme’s D3 methods. A small systematic
overbinding is observed for distance multipliers 0.95–1.5, which is partly a BSSE effect as shown in Figure 7.9. With counterpoise correction the overbinding is reduced to below 5%, close to the estimated 3% accuracy of S66x8 reference binding energies. The remaining effect can be attributed to double counting of electron correlation between the PBE and XDM functionals.

7.4.2 Detailed Assessment

The preceding analysis establishes PW86PBE-XDM as the only DFT method yielding highly accurate intermolecular dissociation curves on the basis of MPD and MAPD values for the whole S66x8 set. For multiple reasons a more detailed analysis is warranted: Firstly, S66x8 covers a wide variety of intermolecular interactions including hydrogen bonding and various types of dispersion interaction for which the method may perform differently. Secondly, S66x8 contains important prototypical complexes that merit detailed examination of their dissociation curves. Thirdly, the rather large MAPD at multiplier 0.9 needs investigation.

To set the stage, S66x8 reference binding energies are plotted in Figure 7.10. Binding energies at the potential energy minimum range from 1 to 10 kcal mol\(^{-1}\) with an average of 5.5 kcal/mol. A rapid decrease in binding energy occurs with increasing intermolecular separation, the magnitude of which depends on the interaction motif but is quite constant for a given type of interaction. This is more clearly shown in Figure 7.11. Hydrogen bond energies show the least distance-dependence with around 70% (5 kcal/mol), 45% (3 kcal/mol), and 15% (1 kcal/mol) of the equilibrium binding energy remaining at multipliers 1.25, 1.5, and 2, respectively. Dispersion and other interactions fall off more rapidly, with only around 60% (2 kcal/mol), 30% (1 kcal/mol), and 10% (0.3 kcal/mol) of the equilibrium binding energy remaining at multipliers 1.25, 1.5, and 2. Complexes 24 (stacked benzene dimer), 33 (pyridine–ethene), and 45 (ethyne–pentane) have binding energies \(\leq 0.1\) kcal/mol at twice the equilibrium separation, within the error bar of the \(\Delta\text{CCSD}(T)\) scheme used to calculate reference binding energies. Benzene–ethene (%30) has an erroneously repulsive reference interaction energy of -0.009 kcal/mol at this separation.

\(\pi-\pi\)-stacked complexes also show a noticeable decrease in binding energy at

---

\(^{6}\) e.g., water dimer, hydrogen-bonded and stacked uracil-uracil base pairs, stacked and T-shaped benzene dimers, aliphatic hydrocarbon dimers etc.

\(^{7}\) except double hydrogen bonds

\(^{8}\) excluding double hydrogen bonds
Figure 7.10: S66x8 reference binding energies by distance multiplier [kcal/mol]. Compressed geometries shown in yellow-orange, near-equilibrium geometries in red, and stretched geometries in shades of purple and blue.
Figure 7.11: S66x8 reference binding energies as percent fraction of the S66 reference equilibrium binding energy [%]
compressed geometries (multiplier 0.9). Complexes 24 (stacked benzene dimer) and 30 (benzene–ethene) are almost repulsive at multiplier 0.9 with binding energies around 0.25 kcal/mol, and complexes 25 (stacked pyridine dimer), 27 (benzene–pyridine) and 33 (pyridine–ethene) have binding energies around 1 kcal/mol, less than 50% of the equilibrium value. In other words, π-π-stacked complexes experience steep exchange repulsion at compressed geometries, due to their large contact surface and short equilibrium distance [231].

The large range of binding energies in the S66x8 set necessitates the use of relative (percent) deviations as employed in Figures 7.7–7.9. However, relative deviations tend to exaggerate the error inherent to a method when the reference binding energy becomes very small. This is the case for the asymptotic tail of dissociation energy curves but also in the repulsive region where such curves cross zero, as shown in Figure 1 of reference [283]. Various fixes such as the use of median relative errors [241] or a complicated weighting scheme [283] have been proposed. The latter scheme essentially amounts to the quantity \( \frac{(BE - BE_{ref})}{BE_{ref}} \). Such schemes have merit in a purely statistical assessment but are rather nonstandard and difficult to interpret. Furthermore, there are only a handful of systems in the S66x8 set where the reference binding energy at compressed geometry (repulsive region) does get close to 0, and these are listed above. At long range, the “exaggeration of errors” by the use of percent deviations is in fact insightful as it allows to discriminate whether a method has the correct asymptotic \( R^{-6} \) behaviour.

Hence, percent deviations of PW86PBE-XDM/heavy-aug-cc-pVTZ binding energies from reference values are shown in Figure 7.12 for the full S66x8 set. The heavy-aug-cc-pVTZ basis set (no CP correction) was chosen as it yields essentially CBS-limit accuracy for the S66 set at reasonable computational cost (see Figure 7.4), and due to its widespread use for calculation of noncovalent interaction energies [266, 283]. PW86PBE-XDM is highly accurate for the binding energies of hydrogen-bonded complexes at all 8 separations. Aliphatic-aliphatic dispersion, π-aliphatic dispersion, and other interactions are also very accurately described at equilibrium and stretched separations. A small overbinding tendency is observed at these geometries. Noticeable deviations occur for some complexes at distance multiplier 0.9. π-aliphatic, T-shaped, and XH-π interactions are generally underbound while aliphatic-aliphatic complexes are overbound at this compressed geometry.

π-π-stacked complexes show clear outliers with large relative deviations at distance multipliers 0.9 and 2. The seemingly large deviations at multiplier 2 for
Figure 7.12: Percent deviation of PW86PBE-XDM/heavy-aug-cc-pVTZ binding energies from S66x8 reference by distance multiplier [%].
complexes 24 (stacked benzene dimer) and 33 (pyridine–ethene) can be attributed to the limited accuracy of the reference binding energies. Note that the percent deviation for complex 30 at multiplier 2 was set to 0 due to the questionable reference binding energy of -0.009 kcal/mol, which would otherwise result in a percent deviation of -344%. No removal of outliers was performed in the computation of the MPDs and MAPDs plotted in Figures 7.7–7.9.9 Benzene–ethene (#30) is the most difficult system among the whole S66x8 set, and is consistently overbound by 40–50% at near-equilibrium and stretched geometries.

Very large relative deviations are also seen at multiplier 0.9, and can be attributed to the steep repulsion in $\pi$-$\pi$-stacked complexes. The most obvious outliers — complexes 24, 27, and 30 — have reference binding energies between 0 and 1 kcal/mol and are responsible for the large MAPD of 19% at multiplier 0.9 seen in Figure 7.7.

Looking at binding energy deviations in units of kcal/mol (Figure 7.13), similar trends are observed. Most PW86PBE-XDM binding energies are within $\pm 1$ kcal/mol of the reference values. Outliers are the stacked uracil dimer (#26), benzene–uracil (#28), and pyridine–uracil (#29), which are underbound by more than 1 kcal/mol at compressed geometries (multiplier 0.9). These are by far the most strongly-bound dispersion complexes, having reference binding energies $> 3$ kcal/mol at multiplier 0.9.

### 7.4.3 Exemplary Potential Energy Curves And Basis Set Effects

To show the impact of these deviations on actual potential energy curves, selected examples are plotted in Figure 7.14. Note that these are worst cases for the PW86PBE-XDM method, and yet the potential energy curves have the correct shape, well depth, and long-range behaviour. The potential energy minimum for the stacked benzene dimer is slightly deeper and at larger separation compared to the reference curve. As both the repulsive wall and the long-range asymptote are accurately described, the large percent errors of Figure 7.12 are clearly an artifact of the near-zero reference binding energy at distance multipliers 0.9 and 2. PW86PBE-XDM yields accurate well depths and dissociation behaviour for $\pi$-$\pi$-stacked complexes involving uracil (26, 28, 29), but is too repulsive in the short-range region.

---

9The def2-QZVP basis set yields a binding energy of -0.011 kcal/mol for complex 30 at multiplier 2. Negative PW86PBE-XDM binding energies at multiplier 2 are only observed in the case of complex 30, and with the following basis sets (with or without CP correction): cc-pVnZ (n=D,T,Q,5), def2-*, pc-1 and pc-2.
Figure 7.13: Deviation of PW86PBE-XDM/heavy-aug-cc-pVTZ binding energies from S66x8 reference by distance multiplier [kcal/mol].
Figure 7.14: Selected potential energy curves from the S66x8 set ($\pi$-$\pi$-stacked complexes). Y-axis: interaction energy (kcal/mol). X-axis: intermolecular distance multiplier. PW86PBE (thin lines) and PW86PBE-XDM (thick lines) curves are shown for three different basis sets that were found to be accurate and economical for the S66 set (see Figure 7.4). The long-range asymptote of selected potential energy curves is magnified in the inset.
Complexes of ethene and an aromatic ring (30, 33) are systematically overbound at near-equilibrium and stretched separations with the heavy-aug-cc-pVTZ basis set. Even in these systems, the PW86PBE-XDM curves become too repulsive at compressed geometries. This is a deficiency of semi-local DFT itself, as can be seen in Figure 3.1, where all common exchange GGAs are found to be overly repulsive at short range. Figures 7.7 and 7.8 suggest that the hybrid and double-hybrid functionals benchmarked by Goerigk et al. are less susceptible to over-repulsion, but this cannot be said with certainty on the basis of MAPD or MPD values alone.

Figure 7.14 also shows that accurate PW86PBE-XDM potential energy curves can be obtained with small double-zeta basis sets such as 6-31+G(d) and 6-31+G(2d, 2p) in conjunction with CP correction. At all geometries except the most stretched one, the remaining small basis set differences are already present in the PW86PBE curve. Only at multiplier 2 does the basis-set dependence of XDM dispersion energies become noticeable (see insets), resulting in a small overbinding compared to the larger heavy-aug-cc-pVTZ basis set, which is very close to the reference curve. Hence, augmented triple zeta or larger basis sets are recommended for the calculation of properties that are sensitive to the long-range asymptote of dissociation energy curves.

In order to demonstrate the typical performance of PW86PBE-XDM in less pathological cases, potential energy curves for prototypical noncovalent interactions are shown in Figures 7.15 and 7.16. Hydrogen bonding in the water dimer (OH–O) is already well described at the PW86PBE level, and addition of the XDM dispersion term leads to slight overbinding. On the other hand, the dispersion correction is essential to capture the full strength of NH–O hydrogen bonds such as in the “peptide” dimer (N-methylacetamide) and the base-paired uracil dimer. The 6-31+G(d) basis set is inadequate for hydrogen bonds involving water due to the lack of a second set of d polarization functions. For molecules larger than water, basis functions on neighbouring atoms compensate for this deficiency, making 6-31+G(d) comparable to the larger 6-31+G(2d,2p) basis set.

PW86PBE-XDM is very accurate for the dispersion interaction between linear (pentane dimer) and cyclic (cyclopentane dimer) aliphatic hydrocarbons. Interactions between branched hydrocarbons (neopentane dimer) are overestimated by 0.5 kcal/mol near the potential minimum. Basis set effects are negligible for these complexes. π-aliphatic (benzene–cyclopentane), T-shaped (benzene dimer), OH–π (benzene–water), and NH–π (benzene–peptide) interactions are also very
Figure 7.15: Selected potential energy curves from the S66x8 set (hydrogen bonds and aliphatic-aliphatic dispersion interactions). Y-axis: interaction energy (kcal/mol). X-axis: intermolecular distance multiplier.
Figure 7.16: Selected potential energy curves from the S66x8 set (π-aliphatic, T-shaped, XH-π and other interactions). Y-axis: interaction energy (kcal/mol). X-axis: intermolecular distance multiplier.
accurately described. Worst cases in these categories are the ethyne dimer (end-
on CH-π) and pyridine–ethyne (end-on CH-N), which are both overbound with
PW86PBE-XDM.

7.4.4 Angular-Displaced Geometries

The S66a8 set samples intermolecular angular degrees of freedom by rotating frag-
ments away from the potential energy minimum. Each fragment (1 or 2) is rotated
by −30° and +30° within (y) and perpendicular to (z) the “principal” molecular
plane [266]. In the nomenclature used by Rezac et al. and adopted here, these
eight angular-displaced geometries are labelled as 1y-30, 1y+30, 1z-30, 1z+30, 2y-
30, 2y+30, 2z-30, and 2z+30. The angular-displaced structures have binding en-
ergies that are on average 77% (hydrogen bonds and other complexes) and 67%
(dispersion-dominated interactions) of the S66 equilibrium binding energy [266].

Deviations of PW86PBE-XDM/heavy-aug-cc-pVTZ binding energies from the
S66a8 reference values are shown in Figure for each angular-displaced geome-
try. All nonequilibrium binding energies are within -0.3 kcal/mol and +0.8 kcal/-
mol of the reference values. Binding energy deviations at angular-displaced ge-
ometries tend to follow the deviations at equilibrium geometries, confirming that
PW86PBE-XDM is applicable to noncovalent interactions at equilibrium and non-
equilibrium geometries. The largest spread with respect to the S66 equilibrium
deviations is seen for hydrogen bonds.

Apart from the benchmarking of ab initio methods in the original publication
[266]—for which only an overall root mean square deviation is reported—no other
methods have been benchmarked on the S66a8 set as of 2012. Note that Rezac et al.
an anticipate significant deviations for more approximate methods such as dispersion-
corrected DFT. However, the PW86PBE-XDM is found to be very accurate for the
whole S66a8 set.

7.5 Conclusions

In this Chapter, the PW86PBE-XDM method was benchmarked for noncovalent
interactions using the S22 and S66 data sets. The method predicts binding energies
with "chemical accuracy" (1 kcal/mol) for a wide range of noncovalent interaction
motifs. Average deviations of 0.25 kcal/mol (10%) are found with respect to ac-
curate ab initio (extrapolated CCSD(T)/CBS) reference values. This is comparable
Figure 7.17: Deviation of PW86PBE-XDM/heavy-aug-cc-pVTZ binding energies from S66a8 reference values [kcal/mol]. Angular-displaced geometries within the principal molecular plane (y) are shown in shades of blue, perpendicular-displaced geometries (z) are shown in shades of orange.
to other dispersion-corrected functionals, but unlike those, our method was not fitted to S22, S66, or similar sets of intermolecular complexes, instead having been parametrized in Chapter 6 on rare-gas interactions only.

Basis sets of triple-zeta quality, augmented with diffuse functions, are necessary to achieve converged binding energies for the S22 and S66 intermolecular complexes. With counterpoise correction, already double-zeta basis sets such as 6-31+G(2d,2p) yield converged results. Diffuse and polarization functions on non-hydrogen atoms are shown to be essential for the correct description of dispersion and hydrogen bonding interactions, respectively.

The performance of PW86PBE-XDM for intermolecular dissociation curves was benchmarked on the S66x8 data set. Unlike other dispersion-corrected functionals that show increasing deviations with increasing intermolecular separation, the PW86PBE method is consistently accurate (MAPD below 10%, binding energies within ±1 kcal/mol of reference values) in the near-equilibrium and stretched regions of noncovalent dissociation curves. For π-π stacked complexes, the method underestimates interaction energies at compressed geometries. This is due to the overly repulsive behaviour of semilocal density functional approximations at short internuclear separations, as already shown in Chapter 3 for rare-gas systems.

Finally, the method was tested on the angular-displaced geometries of the S66a8 data set, where fragments are rotated away from the potential energy minimum in order to sample angular degrees of freedom. Binding energies are found to be within -0.3 kcal/mol and +0.8 kcal/mol of ab initio reference values, confirming that the method successfully describes noncovalent interactions both at equilibrium and non-equilibrium geometries.
### 7.6 Appendix

<table>
<thead>
<tr>
<th>S66</th>
<th>( BE^{\text{ref}} )</th>
<th>( BE^{\text{PP}} )</th>
<th>( %\Delta )</th>
<th>( BE^{\text{PPX}} )</th>
<th>( %\Delta )</th>
<th>( BE^{\text{PPX}} )</th>
<th>( %\Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydrogen bonding</td>
<td>MA%D</td>
<td>15.2</td>
<td>3.7</td>
<td>( BE^{\text{CP}} )</td>
<td>3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>water···water</td>
<td>5.01</td>
<td>4.74</td>
<td>-5.5</td>
<td>5.19</td>
<td>3.6</td>
<td>5.17</td>
</tr>
<tr>
<td>2</td>
<td>water···MeOH</td>
<td>5.70</td>
<td>5.01</td>
<td>-12.1</td>
<td>5.77</td>
<td>1.2</td>
<td>5.72</td>
</tr>
<tr>
<td>3</td>
<td>water···MeNH2</td>
<td>7.04</td>
<td>6.95</td>
<td>-1.2</td>
<td>7.72</td>
<td>9.7</td>
<td>7.68</td>
</tr>
<tr>
<td>4</td>
<td>water···peptide</td>
<td>8.22</td>
<td>7.13</td>
<td>-13.3</td>
<td>8.22</td>
<td>0.0</td>
<td>8.15</td>
</tr>
<tr>
<td>5</td>
<td>MeOH···MeOH</td>
<td>5.85</td>
<td>4.93</td>
<td>-15.8</td>
<td>5.96</td>
<td>1.9</td>
<td>5.89</td>
</tr>
<tr>
<td>6</td>
<td>MeOH···MeNH2</td>
<td>7.67</td>
<td>6.93</td>
<td>-9.7</td>
<td>8.34</td>
<td>8.8</td>
<td>8.26</td>
</tr>
<tr>
<td>7</td>
<td>MeOH···peptide</td>
<td>8.34</td>
<td>7.01</td>
<td>-15.9</td>
<td>8.62</td>
<td>3.3</td>
<td>8.50</td>
</tr>
<tr>
<td>8</td>
<td>MeOH···water</td>
<td>5.09</td>
<td>4.64</td>
<td>-8.9</td>
<td>5.26</td>
<td>3.4</td>
<td>5.22</td>
</tr>
<tr>
<td>9</td>
<td>MeNH2···MeOH</td>
<td>3.11</td>
<td>2.08</td>
<td>-33.3</td>
<td>3.30</td>
<td>6.0</td>
<td>3.24</td>
</tr>
<tr>
<td>10</td>
<td>MeNH2···MeNH2</td>
<td>4.22</td>
<td>2.92</td>
<td>-30.9</td>
<td>4.43</td>
<td>4.9</td>
<td>4.37</td>
</tr>
<tr>
<td>11</td>
<td>MeNH2···peptide</td>
<td>5.48</td>
<td>3.29</td>
<td>-40.0</td>
<td>5.53</td>
<td>0.9</td>
<td>5.42</td>
</tr>
<tr>
<td>12</td>
<td>MeNH2···water</td>
<td>7.40</td>
<td>6.90</td>
<td>-6.7</td>
<td>7.81</td>
<td>5.5</td>
<td>7.76</td>
</tr>
<tr>
<td>13</td>
<td>peptide···MeOH</td>
<td>6.28</td>
<td>4.45</td>
<td>-29.2</td>
<td>6.15</td>
<td>-2.0</td>
<td>6.01</td>
</tr>
<tr>
<td>14</td>
<td>peptide···MeNH2</td>
<td>7.56</td>
<td>5.94</td>
<td>-21.4</td>
<td>7.95</td>
<td>5.1</td>
<td>7.81</td>
</tr>
<tr>
<td>15</td>
<td>peptide···peptide</td>
<td>8.72</td>
<td>6.31</td>
<td>-27.7</td>
<td>8.80</td>
<td>0.8</td>
<td>8.59</td>
</tr>
<tr>
<td>16</td>
<td>peptide···water</td>
<td>5.20</td>
<td>4.21</td>
<td>-18.9</td>
<td>5.09</td>
<td>-2.1</td>
<td>5.00</td>
</tr>
<tr>
<td>17</td>
<td>uracil···uracil (base pair)</td>
<td>17.45</td>
<td>15.44</td>
<td>-11.5</td>
<td>17.73</td>
<td>1.6</td>
<td>17.33</td>
</tr>
<tr>
<td>18</td>
<td>water···pyridine</td>
<td>6.98</td>
<td>6.71</td>
<td>-3.7</td>
<td>7.63</td>
<td>9.3</td>
<td>7.54</td>
</tr>
<tr>
<td>19</td>
<td>MeOH···pyridine</td>
<td>7.51</td>
<td>6.80</td>
<td>-9.5</td>
<td>8.20</td>
<td>9.1</td>
<td>8.07</td>
</tr>
<tr>
<td>20</td>
<td>AcOH···AcOH</td>
<td>19.42</td>
<td>18.19</td>
<td>-6.3</td>
<td>19.98</td>
<td>2.9</td>
<td>19.70</td>
</tr>
<tr>
<td>21</td>
<td>AcNH2···AcNH2</td>
<td>16.53</td>
<td>14.87</td>
<td>-10.0</td>
<td>16.73</td>
<td>1.2</td>
<td>16.48</td>
</tr>
<tr>
<td>22</td>
<td>AcOH···uracil</td>
<td>19.78</td>
<td>18.08</td>
<td>-8.6</td>
<td>20.07</td>
<td>1.5</td>
<td>19.75</td>
</tr>
<tr>
<td>23</td>
<td>AcNH2···uracil</td>
<td>19.47</td>
<td>17.48</td>
<td>-10.2</td>
<td>19.55</td>
<td>0.4</td>
<td>19.25</td>
</tr>
</tbody>
</table>

| dispersion | MA\%D | 106.0 | 14.1 | 11.0 |
| 24 | benzene···benzene (\( \pi \rightarrow \pi \)) | 2.72 | -1.10 | -140.2 | 3.37 | 23.6 | 3.17 | 16.4 |
| 25 | pyridine···pyridine (\( \pi \rightarrow \pi \)) | 3.80 | -0.40 | -110.4 | 4.24 | 11.5 | 4.03 | 6.1 |
| 26 | uracil···uracil (\( \pi \rightarrow \pi \)) | 9.75 | 2.74 | -72.0 | 9.36 | -4.0 | 8.92 | -8.6 |
| 27 | benzene···pyridine (\( \pi \rightarrow \pi \)) | 3.34 | -0.71 | -121.2 | 3.86 | 15.5 | 3.66 | 9.5 |
| 28 | benzene···uracil (\( \pi \rightarrow \pi \)) | 5.59 | -0.16 | -102.8 | 5.58 | -0.3 | 5.26 | -6.0 |
| 29 | pyridine···uracil (\( \pi \rightarrow \pi \)) | 6.70 | 0.96 | -85.7 | 6.60 | -1.5 | 6.29 | -6.1 |
| 30 | benzene···ethene | 1.36 | -0.70 | -151.2 | 1.95 | 43.3 | 1.85 | 35.6 |
| 31 | uracil···ethene | 3.33 | 0.41 | -87.8 | 3.49 | 4.8 | 3.32 | -0.3 |
| 32 | uracil···ethyne | 3.69 | 1.08 | -70.7 | 3.86 | 4.5 | 3.72 | 0.8 |
| 33 | pyridine···ethene | 1.81 | -0.39 | -121.3 | 2.32 | 28.5 | 2.21 | 22.5 |

Continued on next page
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>BE_{ref}</th>
<th>BE_{PP}</th>
<th>Δ</th>
<th>BE_{PPX}</th>
<th>Δ</th>
<th>BE_{PPX}</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>pentane···pentane</td>
<td>3.76</td>
<td>-0.42</td>
<td>-111.2</td>
<td>4.34</td>
<td>15.4</td>
<td>4.17</td>
<td>10.7</td>
</tr>
<tr>
<td>35</td>
<td>neopentane···pentane</td>
<td>2.60</td>
<td>-0.19</td>
<td>-107.3</td>
<td>3.11</td>
<td>19.5</td>
<td>2.98</td>
<td>14.5</td>
</tr>
<tr>
<td>36</td>
<td>neopentane···neopentane</td>
<td>1.76</td>
<td>-0.11</td>
<td>-106.0</td>
<td>2.34</td>
<td>32.8</td>
<td>2.24</td>
<td>26.9</td>
</tr>
<tr>
<td>37</td>
<td>cyclopentane···neopentane</td>
<td>2.40</td>
<td>-0.22</td>
<td>-109.1</td>
<td>2.93</td>
<td>22.4</td>
<td>2.80</td>
<td>17.0</td>
</tr>
<tr>
<td>38</td>
<td>cyclopentane···cyclopentane</td>
<td>2.99</td>
<td>-0.14</td>
<td>-104.6</td>
<td>3.54</td>
<td>18.4</td>
<td>3.38</td>
<td>13.2</td>
</tr>
<tr>
<td>39</td>
<td>benzene···cyclopentane</td>
<td>3.51</td>
<td>-0.22</td>
<td>-106.2</td>
<td>3.96</td>
<td>12.7</td>
<td>3.74</td>
<td>6.4</td>
</tr>
<tr>
<td>40</td>
<td>benzene···neopentane</td>
<td>2.85</td>
<td>-0.07</td>
<td>-102.4</td>
<td>3.30</td>
<td>15.9</td>
<td>3.12</td>
<td>9.4</td>
</tr>
<tr>
<td>41</td>
<td>uracil···pentane</td>
<td>4.81</td>
<td>-0.40</td>
<td>-108.2</td>
<td>4.88</td>
<td>1.4</td>
<td>4.56</td>
<td>-5.3</td>
</tr>
<tr>
<td>42</td>
<td>uracil···cyclopentane</td>
<td>4.09</td>
<td>-0.46</td>
<td>-111.3</td>
<td>4.19</td>
<td>2.5</td>
<td>3.91</td>
<td>-4.4</td>
</tr>
<tr>
<td>43</td>
<td>uracil···neopentane</td>
<td>3.69</td>
<td>-0.02</td>
<td>-100.5</td>
<td>3.82</td>
<td>3.6</td>
<td>3.58</td>
<td>-3.0</td>
</tr>
<tr>
<td>44</td>
<td>ethene···pentane</td>
<td>1.99</td>
<td>-0.11</td>
<td>-105.4</td>
<td>2.27</td>
<td>13.7</td>
<td>2.17</td>
<td>9.0</td>
</tr>
<tr>
<td>45</td>
<td>ethyne···pentane</td>
<td>1.72</td>
<td>-0.02</td>
<td>-100.9</td>
<td>2.14</td>
<td>24.7</td>
<td>2.07</td>
<td>20.7</td>
</tr>
<tr>
<td>46</td>
<td>peptide···pentane</td>
<td>4.26</td>
<td>-0.03</td>
<td>-100.7</td>
<td>4.43</td>
<td>4.0</td>
<td>4.21</td>
<td>-1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>other interactions</td>
<td>MA%D</td>
<td>59.5</td>
<td>4.7</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>benzene···benzene (TS)</td>
<td>2.83</td>
<td>0.17</td>
<td>-93.8</td>
<td>2.94</td>
<td>4.1</td>
<td>2.75</td>
<td>-2.9</td>
</tr>
<tr>
<td>48</td>
<td>pyridine···pyridine (TS)</td>
<td>3.51</td>
<td>0.72</td>
<td>-79.6</td>
<td>3.48</td>
<td>-0.9</td>
<td>3.30</td>
<td>-5.9</td>
</tr>
<tr>
<td>49</td>
<td>benzene···pyridine (TS)</td>
<td>3.29</td>
<td>0.58</td>
<td>-82.5</td>
<td>3.34</td>
<td>1.5</td>
<td>3.14</td>
<td>-4.6</td>
</tr>
<tr>
<td>50</td>
<td>benzene···ethyne (CH···N)</td>
<td>3.86</td>
<td>1.16</td>
<td>-59.5</td>
<td>3.92</td>
<td>2.2</td>
<td>3.75</td>
<td>-3.8</td>
</tr>
<tr>
<td>51</td>
<td>ethyne···ethyne (TS)</td>
<td>1.54</td>
<td>1.15</td>
<td>-25.3</td>
<td>1.84</td>
<td>19.8</td>
<td>1.76</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>benzene···AcOH (OH···π)</td>
<td>4.73</td>
<td>2.18</td>
<td>-53.9</td>
<td>4.54</td>
<td>-3.9</td>
<td>4.32</td>
<td>-8.5</td>
</tr>
<tr>
<td>53</td>
<td>benzene···AcNH2 (NH···π)</td>
<td>4.41</td>
<td>2.21</td>
<td>-49.8</td>
<td>4.35</td>
<td>-1.3</td>
<td>4.19</td>
<td>-5.0</td>
</tr>
<tr>
<td>54</td>
<td>benzene···water (OH···π)</td>
<td>3.29</td>
<td>1.80</td>
<td>-45.3</td>
<td>3.30</td>
<td>0.3</td>
<td>3.20</td>
<td>-2.6</td>
</tr>
<tr>
<td>55</td>
<td>benzene···MeOH (OH···π)</td>
<td>4.17</td>
<td>1.72</td>
<td>-58.8</td>
<td>4.27</td>
<td>2.4</td>
<td>4.10</td>
<td>-1.6</td>
</tr>
<tr>
<td>56</td>
<td>benzene···MeNH2 (NH···π)</td>
<td>3.20</td>
<td>0.69</td>
<td>-78.4</td>
<td>3.35</td>
<td>4.6</td>
<td>3.21</td>
<td>0.2</td>
</tr>
<tr>
<td>57</td>
<td>benzene···peptide (NH···π)</td>
<td>5.26</td>
<td>1.61</td>
<td>-69.3</td>
<td>5.30</td>
<td>0.9</td>
<td>5.06</td>
<td>-3.7</td>
</tr>
<tr>
<td>58</td>
<td>pyridine···pyridine (CH···N)</td>
<td>4.24</td>
<td>2.52</td>
<td>-40.5</td>
<td>4.12</td>
<td>-2.8</td>
<td>3.94</td>
<td>-7.1</td>
</tr>
<tr>
<td>59</td>
<td>ethyne···water (CH···O)</td>
<td>2.93</td>
<td>2.55</td>
<td>-12.8</td>
<td>3.00</td>
<td>2.4</td>
<td>2.90</td>
<td>-1.0</td>
</tr>
<tr>
<td>60</td>
<td>ethyne···AcOH (OH···π)</td>
<td>4.97</td>
<td>4.02</td>
<td>-19.1</td>
<td>5.33</td>
<td>7.3</td>
<td>5.22</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>pentane···AcOH</td>
<td>2.91</td>
<td>-0.10</td>
<td>-103.6</td>
<td>3.15</td>
<td>8.5</td>
<td>2.96</td>
<td>1.8</td>
</tr>
<tr>
<td>62</td>
<td>pentane···AcNH2</td>
<td>3.53</td>
<td>0.26</td>
<td>-92.5</td>
<td>3.82</td>
<td>8.3</td>
<td>3.63</td>
<td>3.0</td>
</tr>
<tr>
<td>63</td>
<td>benzene···AcOH</td>
<td>3.75</td>
<td>0.49</td>
<td>-86.9</td>
<td>3.82</td>
<td>1.9</td>
<td>3.60</td>
<td>-4.0</td>
</tr>
<tr>
<td>64</td>
<td>peptide···ethene</td>
<td>3.00</td>
<td>0.88</td>
<td>-70.7</td>
<td>3.15</td>
<td>4.8</td>
<td>3.02</td>
<td>0.7</td>
</tr>
<tr>
<td>65</td>
<td>pyridine···ethyne</td>
<td>4.10</td>
<td>3.68</td>
<td>-10.4</td>
<td>4.61</td>
<td>12.4</td>
<td>4.42</td>
<td>7.6</td>
</tr>
<tr>
<td>66</td>
<td>MeNH2···pyridine</td>
<td>3.97</td>
<td>1.67</td>
<td>-58.0</td>
<td>4.13</td>
<td>4.2</td>
<td>4.03</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>S66</th>
<th>$BE^{\text{ref}}$</th>
<th>$BE^{PP}$</th>
<th>$% \Delta$</th>
<th>$BE^{PPX}$</th>
<th>$% \Delta$</th>
<th>$BE^{PPX}_{\text{CP}}$</th>
<th>$% \Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>full S66 set</td>
<td>MD</td>
<td>-2.31</td>
<td>0.24</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAD</td>
<td>2.31</td>
<td>0.27</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MaxD</td>
<td>-7.02</td>
<td>#26</td>
<td>0.69</td>
<td>#19</td>
<td>-0.84</td>
<td>#26</td>
</tr>
<tr>
<td></td>
<td>M%D</td>
<td>-60.3</td>
<td>7.1</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA%D</td>
<td>60.3</td>
<td>7.6</td>
<td>6.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max%D</td>
<td>-151.2</td>
<td>#30</td>
<td>43.3</td>
<td>#30</td>
<td>35.7</td>
<td>#30</td>
</tr>
</tbody>
</table>

Table 7.4: Binding energies (kcal/mol), percent deviations ($\% \Delta$), and error statistics for the S66 benchmark set with PW86PBE/aug-cc-pVTZ, PW86PBE-XDM/aug-cc-pVTZ and CP-corrected PW86PBE-XDM/aug-cc-pVTZ.
Figure 7.18: CP-corrected and uncorrected PW86PBE-XDM binding energy MAD (S22)
Figure 7.19: Deviation of PW86PBE-XDM binding energies from S66 reference for Pople-type basis sets (no CP)
Figure 7.20: Deviation of PW86PBE-XDM binding energies from S66 reference for Pople-type basis sets (CP)
Figure 7.21: Deviation of PW86PBE-XDM binding energies from S66 reference for cc-, def2, and pc-type basis sets (no CP)
Figure 7.22: Deviation of PW86PBE-XDM binding energies from S66 reference for cc-, def2, and pc-type basis sets (CP)
Chapter 8

Atomic Volumes And Polarizabilities

© Copyright 2012 American Institute of Physics

Abstract

Becke and Johnson introduced an ad hoc definition of atomic volume [J. Chem. Phys. 124, 014204 (2006)] in order to obtain atom-in-molecule polarizabilities from free-atom polarizabilities in their nonempirical exchange-hole dipole moment (XDM) model of dispersion interactions. Here we explore the dependence of Becke-Johnson atomic volumes on basis sets and density-functional approximations and provide reference data for all atoms H-Lr. A persuasive theoretical foundation for the Becke-Johnson definition is also provided.

8.1 Introduction

Recent nonempirical dispersion models in density-functional theory depend on ratios of effective atom-in-molecule volumes to free atomic volumes. In the exchange-hole dipole moment (XDM) dispersion model of Becke and Johnson [52,54,56], the static dipole polarizability $\alpha_i$ of atom $i$ in a molecule is calculated from the corresponding free-atom polarizability $\alpha_{i,\text{free}}$ according to [51,56]

$$\alpha_i = \frac{V_{i,\text{eff}}}{V_{i,\text{free}}} \alpha_{i,\text{free}}$$

(8.1)

The dispersion model of Tkatchenko and Scheffler [185] derives effective atom-in-molecule dispersion coefficients $C_{6,i}$ from free-atom dispersion coefficients according to

$$C_{6,i} = \left( \frac{V_{i,\text{eff}}}{V_{i,\text{free}}} \right)^2 C_{6,i,\text{free}}$$

(8.2)
The use of volume ratios to calculate effective polarizabilities and dispersion coefficients from the corresponding free-atom values exploits the proportionality between polarizability and volume known from classical electromagnetic theory, from the statistical theory of atoms [284], and from empirical validations on real atoms [285] and molecules of varying geometries and electronic structures [286].

Becke and Johnson proposed a measure of effective atom-in-molecule volume as [51, 56]

\[
V_{i,\text{eff}} = \left\langle r^3 \right\rangle_{i,\text{eff}} = \int r^3 w_i (r) \rho (r) \, dr
\]  

(8.3)

where \( \rho (r) \) is the total electron density of the molecule and \( w_i (r) \) is the Hirshfeld atomic partitioning weight [60] of atom \( i \) in the molecule. The equivalent expression for free atomic volume is [51, 56]

\[
V_{i,\text{free}} = \left\langle r^3 \right\rangle_{i,\text{free}} = \int r^3 \rho_{i,\text{free}} (r) \, dr
\]  

(8.4)

where \( \rho_{i,\text{free}} (r) \) is the ground-state electron density of the free atom \( i \). Atomic units are used throughout, with polarizabilities and volumes given in \( a_0^3, a_0 \) being the Bohr radius. Based on dimensionality and intuition, the definition of Becke-Johnson atomic volume is essentially \textit{ad hoc}. Interestingly, in the statistical theory of atoms [287–289] a formula for polarizability can be derived [284, 289], which reduces to Eq. (8.4) at large \( r \) where we expect the dominant contributions to the polarizability to arise. This is discussed in the Appendix.

Previous applications of the XDM dispersion model by Becke and Johnson used fully-numerical local-spin-density approximation (LSDA) values for the free atomic volumes \( V_{i,\text{free}} \) [57, 189, 209, 226, 290]. In principle, however, the free atomic volumes of Eq. (8.4) should be calculated with the same model chemistry (i.e., same density-functional approximation and basis set) that is used in the calculation of the effective atom-in-molecule volumes of Eq. (8.3). This is impractical due to the many possible combinations of functional and basis set. Furthermore, atomic SCF calculations on d- and f-block elements can be difficult to converge to the correct ground state wavefunction, making on-the-fly calculation of free atomic volumes impractical.

A universal set of free atomic volumes that can be used with any model chemistry is desirable. This requires that the dependence of the free atomic volumes on the basis set and functional is negligible.

The purpose of this paper is to investigate this dependence, and to provide a
universal reference set of Becke-Johnson free atomic volumes for the elements H-Lr. Additionally, a comprehensive analysis of the relation between atomic polarizability and Becke-Johnson volume across the whole periodic table is presented.

8.2 Computational Methods

The local-spin-density approximation/complete basis set (LSDA/CBS) free atomic volumes in Table 8.1 for atoms H-Ar and Kr were calculated with the fully-numerical basis-set-free *Numol* program of Becke and Dickson [64–66] using the Perdew-Wang LSDA correlation functional [20]. All other calculations employed Gaussian 09 [95] with the SVWN5 [19] LSDA and the PW86PBE [26,29] generalized gradient approximation (GGA), using the keyword IOp(3/74=809) for the latter. Dunning’s aug-cc-pVDZ (aDZ) and aug-cc-pVTZ (aTZ) basis sets [76,78,79,81], and the universal Gaussian basis set (UGBS) of de Castro, Jorge and coworkers [291–293], were used as implemented in Gaussian 09.

The G09 calculations in Table 8.1 used the “ultrafine” numerical integration grid (99 radial x 590 angular points). The ground-state multiplicities of the free atoms were taken from NIST [294]. Open-shell atoms were treated within the unrestricted Kohn-Sham formalism. The G09 free-atom densities were numerically integrated to obtain the free-atom volumes according to Eq. 8.4, using a grid of 194 angular points and from 40 to 160 (depending on Z) radial points. Extension of the grid to 302 angular points and twice as many radial points has a negligible effect (< 1%) on the calculated volumes.

The LSDA/UGBS free atomic volumes in Table 8.2 for atoms H through Lr were also calculated with Gaussian 09 using the SVWN5 LSDA functional. To ensure accurate numerical integration for the heavy elements, a larger unpruned grid of 499 radial x 590 angular points was used throughout. Scalar relativistic effects were included for all elements in Table 8.2 with the second-order Douglas-Kroll-Hess (DKH) Hamiltonian [295–297].

8.3 Results And Discussion

Free atomic volumes ($a_0^3$) of elements H-Ar and Kr are shown in Table 8.1 for various model chemistries, along with the percent deviation from the LSDA/CBS values. The LSDA/aDZ volumes have deviations up to 3% from the CBS values. The LSDA/aTZ volumes are within 1%, and LSDA/UGBS has negligible deviations...
Table 8.1: Free atomic volumes ($a_0^3$) of elements H-Ar and Kr with various model chemistries, and % deviation from LSDA/CBS.

From the CBS limit (with the exception of Al). This rather small basis-set dependence suggests that a universal reference set of free atomic volumes can be generated with appropriately large basis sets. The best choice is the universal Gaussian basis set (UGBS), as it is close to the CBS limit and available for the whole periodic table. Also, since UGBS consists of uncontracted Gaussian primitive functions, relativistic effects can be accounted for without the need for a relativistic recontruction of the basis set.

To investigate the dependence of the free atomic volumes on the choice of density-functional approximation, additional calculations were performed with the PW86PBE GGA. This functional was chosen because PW86 is an optimal exchange GGA for dispersion interactions [99, 189, 191], and is commonly used for dispersion energy calculations in combination with nonempirical dispersion models such as XDM [189, 209, 226], vdW-DF2 [46] and VV10 [47, 48]. The GGA free atomic volumes in Table 8.1 are rather close to the corresponding LSDA values. Significant deviations are only found for the first five elements H, He, Li, Be and B. With the aTZ and UGBS basis sets, the GGA volumes of the remaining elements are within 1% of the LSDA/CBS values (with the exception of Al). In particular, the GGA free atomic volumes of the elements C, N and O, which are expected to provide the main contribution to the dispersion energy in typical organic molecules, are very close to the LSDA/CBS values.

Thus, the choice of density-functional approximation (i.e., LSDA or GGA) for
<table>
<thead>
<tr>
<th>Z</th>
<th>Symbol</th>
<th>Polarizability</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H</td>
<td>0.666793</td>
<td>9.194</td>
</tr>
<tr>
<td>2</td>
<td>He</td>
<td>0.2050522</td>
<td>4.481</td>
</tr>
<tr>
<td>3</td>
<td>Li</td>
<td>24.33</td>
<td>91.96</td>
</tr>
<tr>
<td>4</td>
<td>Be</td>
<td>5.60</td>
<td>61.36</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>3.03</td>
<td>49.81</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>1.76</td>
<td>36.73</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>1.10</td>
<td>27.63</td>
</tr>
<tr>
<td>8</td>
<td>O</td>
<td>0.802</td>
<td>23.52</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>0.557</td>
<td>19.32</td>
</tr>
<tr>
<td>10</td>
<td>Ne</td>
<td>0.3956</td>
<td>15.95</td>
</tr>
<tr>
<td>11</td>
<td>Na</td>
<td>24.11</td>
<td>109.4</td>
</tr>
<tr>
<td>12</td>
<td>Mg</td>
<td>10.6</td>
<td>103.1</td>
</tr>
<tr>
<td>13</td>
<td>Al</td>
<td>6.8</td>
<td>120.4</td>
</tr>
<tr>
<td>14</td>
<td>Si</td>
<td>5.38</td>
<td>104.2</td>
</tr>
<tr>
<td>15</td>
<td>P</td>
<td>3.63</td>
<td>86.78</td>
</tr>
<tr>
<td>16</td>
<td>S</td>
<td>2.90</td>
<td>77.13</td>
</tr>
<tr>
<td>17</td>
<td>Cl</td>
<td>2.18</td>
<td>66.37</td>
</tr>
<tr>
<td>18</td>
<td>Ar</td>
<td>1.6411</td>
<td>57.34</td>
</tr>
<tr>
<td>19</td>
<td>K</td>
<td>43.4</td>
<td>203.1</td>
</tr>
<tr>
<td>20</td>
<td>Ca</td>
<td>22.8</td>
<td>212.2</td>
</tr>
<tr>
<td>21</td>
<td>Sc</td>
<td>17.8</td>
<td>183.1</td>
</tr>
<tr>
<td>22</td>
<td>Ti</td>
<td>14.6</td>
<td>162.3</td>
</tr>
<tr>
<td>23</td>
<td>V</td>
<td>12.4</td>
<td>143.2</td>
</tr>
<tr>
<td>24</td>
<td>Cr</td>
<td>11.6</td>
<td>108.2</td>
</tr>
<tr>
<td>25</td>
<td>Mn</td>
<td>9.4</td>
<td>123.1</td>
</tr>
<tr>
<td>26</td>
<td>Fe</td>
<td>8.4</td>
<td>105.7</td>
</tr>
<tr>
<td>27</td>
<td>Co</td>
<td>7.5</td>
<td>92.94</td>
</tr>
<tr>
<td>28</td>
<td>Ni</td>
<td>6.8</td>
<td>83.79</td>
</tr>
<tr>
<td>29</td>
<td>Cu</td>
<td>6.2</td>
<td>75.75</td>
</tr>
<tr>
<td>30</td>
<td>Zn</td>
<td>5.75</td>
<td>81.18</td>
</tr>
<tr>
<td>31</td>
<td>Ga</td>
<td>8.12</td>
<td>118.4</td>
</tr>
<tr>
<td>32</td>
<td>Ge</td>
<td>6.07</td>
<td>116.3</td>
</tr>
<tr>
<td>33</td>
<td>As</td>
<td>4.31</td>
<td>107.5</td>
</tr>
<tr>
<td>34</td>
<td>Se</td>
<td>3.77</td>
<td>103.2</td>
</tr>
<tr>
<td>35</td>
<td>Br</td>
<td>3.05</td>
<td>95.11</td>
</tr>
<tr>
<td>36</td>
<td>Kr</td>
<td>2.4844</td>
<td>87.61</td>
</tr>
<tr>
<td>37</td>
<td>Rb</td>
<td>47.3</td>
<td>248.8</td>
</tr>
<tr>
<td>38</td>
<td>Sr</td>
<td>27.6</td>
<td>273.7</td>
</tr>
<tr>
<td>39</td>
<td>Y</td>
<td>22.7</td>
<td>249.2</td>
</tr>
<tr>
<td>40</td>
<td>Zr</td>
<td>17.9</td>
<td>223.8</td>
</tr>
<tr>
<td>41</td>
<td>Nb</td>
<td>15.7</td>
<td>175.8</td>
</tr>
<tr>
<td>42</td>
<td>Mo</td>
<td>12.8</td>
<td>156.8</td>
</tr>
<tr>
<td>43</td>
<td>Tc</td>
<td>11.4</td>
<td>160.0</td>
</tr>
<tr>
<td>44</td>
<td>Ru</td>
<td>9.6</td>
<td>136.7</td>
</tr>
<tr>
<td>45</td>
<td>Rh</td>
<td>8.6</td>
<td>127.8</td>
</tr>
<tr>
<td>46</td>
<td>Pd</td>
<td>4.8</td>
<td>97.02</td>
</tr>
<tr>
<td>47</td>
<td>Ag</td>
<td>7.2</td>
<td>112.8</td>
</tr>
<tr>
<td>48</td>
<td>Cd</td>
<td>7.36</td>
<td>121.6</td>
</tr>
<tr>
<td>49</td>
<td>In</td>
<td>10.2</td>
<td>167.9</td>
</tr>
<tr>
<td>50</td>
<td>Sn</td>
<td>7.7</td>
<td>172.0</td>
</tr>
<tr>
<td>51</td>
<td>Sb</td>
<td>6.6</td>
<td>165.5</td>
</tr>
<tr>
<td>52</td>
<td>Te</td>
<td>5.5</td>
<td>163.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Z</th>
<th>Symbol</th>
<th>Polarizability</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>I</td>
<td>5.35</td>
<td>154.0</td>
</tr>
<tr>
<td>54</td>
<td>Xe</td>
<td>4.044</td>
<td>146.1</td>
</tr>
<tr>
<td>55</td>
<td>Cs</td>
<td>59.42</td>
<td>342.0</td>
</tr>
<tr>
<td>56</td>
<td>Ba</td>
<td>39.7</td>
<td>385.8</td>
</tr>
<tr>
<td>57</td>
<td>La</td>
<td>31.1</td>
<td>343.4</td>
</tr>
<tr>
<td>58</td>
<td>Ce</td>
<td>29.6</td>
<td>350.3</td>
</tr>
<tr>
<td>59</td>
<td>Pr</td>
<td>28.2</td>
<td>334.9</td>
</tr>
<tr>
<td>60</td>
<td>Nd</td>
<td>31.4</td>
<td>322.2</td>
</tr>
<tr>
<td>61</td>
<td>Pm</td>
<td>30.1</td>
<td>310.3</td>
</tr>
<tr>
<td>62</td>
<td>Sm</td>
<td>28.8</td>
<td>299.5</td>
</tr>
<tr>
<td>63</td>
<td>Eu</td>
<td>27.7</td>
<td>289.6</td>
</tr>
<tr>
<td>64</td>
<td>Gd</td>
<td>23.5</td>
<td>216.1</td>
</tr>
<tr>
<td>65</td>
<td>Tb</td>
<td>25.5</td>
<td>268.9</td>
</tr>
<tr>
<td>66</td>
<td>Dy</td>
<td>24.5</td>
<td>259.8</td>
</tr>
<tr>
<td>67</td>
<td>Ho</td>
<td>23.6</td>
<td>251.3</td>
</tr>
<tr>
<td>68</td>
<td>Er</td>
<td>22.7</td>
<td>243.2</td>
</tr>
<tr>
<td>69</td>
<td>Tm</td>
<td>21.8</td>
<td>235.5</td>
</tr>
<tr>
<td>70</td>
<td>Yb</td>
<td>21.0</td>
<td>228.3</td>
</tr>
<tr>
<td>71</td>
<td>Lu</td>
<td>21.9</td>
<td>229.6</td>
</tr>
<tr>
<td>72</td>
<td>Hf</td>
<td>16.2</td>
<td>210.0</td>
</tr>
<tr>
<td>73</td>
<td>Ta</td>
<td>13.1</td>
<td>197.5</td>
</tr>
<tr>
<td>74</td>
<td>W</td>
<td>11.1</td>
<td>183.2</td>
</tr>
<tr>
<td>75</td>
<td>Re</td>
<td>9.7</td>
<td>174.7</td>
</tr>
<tr>
<td>76</td>
<td>Os</td>
<td>8.5</td>
<td>164.1</td>
</tr>
<tr>
<td>77</td>
<td>Ir</td>
<td>7.6</td>
<td>150.4</td>
</tr>
<tr>
<td>78</td>
<td>Pt</td>
<td>6.5</td>
<td>135.8</td>
</tr>
<tr>
<td>79</td>
<td>Au</td>
<td>5.8</td>
<td>125.3</td>
</tr>
<tr>
<td>80</td>
<td>Hg</td>
<td>5.02</td>
<td>131.3</td>
</tr>
<tr>
<td>81</td>
<td>Tl</td>
<td>7.6</td>
<td>185.8</td>
</tr>
<tr>
<td>82</td>
<td>Pb</td>
<td>6.8</td>
<td>195.7</td>
</tr>
<tr>
<td>83</td>
<td>Bi</td>
<td>7.4</td>
<td>193.0</td>
</tr>
<tr>
<td>84</td>
<td>Po</td>
<td>6.8</td>
<td>189.1</td>
</tr>
<tr>
<td>85</td>
<td>At</td>
<td>6.0</td>
<td>185.9</td>
</tr>
<tr>
<td>86</td>
<td>Rn</td>
<td>5.3</td>
<td>181.1</td>
</tr>
<tr>
<td>87</td>
<td>Fr</td>
<td>48.60</td>
<td>357.8</td>
</tr>
<tr>
<td>88</td>
<td>Ra</td>
<td>38.3</td>
<td>407.3</td>
</tr>
<tr>
<td>89</td>
<td>Ac</td>
<td>32.1</td>
<td>383.1</td>
</tr>
<tr>
<td>90</td>
<td>Th</td>
<td>32.1</td>
<td>362.1</td>
</tr>
<tr>
<td>91</td>
<td>Pa</td>
<td>25.4</td>
<td>346.6</td>
</tr>
<tr>
<td>92</td>
<td>U</td>
<td>24.9</td>
<td>332.5</td>
</tr>
<tr>
<td>93</td>
<td>Np</td>
<td>24.8</td>
<td>319.6</td>
</tr>
<tr>
<td>94</td>
<td>Pu</td>
<td>24.5</td>
<td>308.1</td>
</tr>
<tr>
<td>95</td>
<td>Am</td>
<td>23.3</td>
<td>297.4</td>
</tr>
<tr>
<td>96</td>
<td>Cm</td>
<td>23.0</td>
<td>300.6</td>
</tr>
<tr>
<td>97</td>
<td>Bk</td>
<td>22.7</td>
<td>275.8</td>
</tr>
<tr>
<td>98</td>
<td>Cf</td>
<td>20.5</td>
<td>266.3</td>
</tr>
<tr>
<td>99</td>
<td>Es</td>
<td>19.7</td>
<td>257.4</td>
</tr>
<tr>
<td>100</td>
<td>Fm</td>
<td>23.8</td>
<td>209.7</td>
</tr>
<tr>
<td>101</td>
<td>Md</td>
<td>18.2</td>
<td>203.2</td>
</tr>
<tr>
<td>102</td>
<td>No</td>
<td>17.5</td>
<td>230.2</td>
</tr>
<tr>
<td>103</td>
<td>Lr</td>
<td>na</td>
<td>236.9</td>
</tr>
</tbody>
</table>

Table 8.2: Free atomic polarizabilities (10^{-24}cm^3) and LSDA/UGBS free atomic volumes (a_0^3) of elements H-Lr in the ground state.
the calculation of reference free atomic volumes is essentially arbitrary, but the LSDA is the simplest choice. A reference set of free atomic volumes for the elements H-Lr, calculated with LSDA/UGBS and including scalar relativistic effects, is presented in Table 8.2. Relativistic effects account for the small discrepancies in LSDA/UGBS volumes between Tables 8.1 and 8.2.

Table 8.2 also lists ground-state free atomic polarizabilities taken from the literature [61]. These reference polarizabilities were determined experimentally or through relativistic calculation, and are required in the XDM dispersion model via Eq. 8.1.

Figure 8.1 shows volumes and polarizabilities plotted against atomic number Z on the same vertical scale, using the conversion factor $\alpha (\text{cm}^3) = 0.148184 \times 10^{-24} \alpha (a_0^3)$ [61]. Both quantities follow the expected trends across the periodic table: volume and polarizability decrease within a horizontal period, and increase down a vertical group. The elements 46 (Pd) and 64 (Gd) have exceptionally small volumes and polarizabilities compared to their neighbouring elements.
A qualitative proportionality between polarizability and volume is apparent from Figure 8.1 across the whole periodic table. The only exception is found for the group 1 elements, the volumes of which are systematically too low. Furthermore, starting with the K/Ca pair, the volume of each group 1 element is lower than the volume of its neighbouring group 2 element. This is in contradiction to the progression of polarizability between groups 1 and 2, as well as most definitions of atomic radius [285, 298]. This discrepancy has been reported previously for the K/Ca pair [285], and is now shown to extend throughout the periodic table.

In addition to the LSDA/UGBS free atomic volumes, Figure 8.1 also shows the corresponding GGA/UGBS values. The close agreement between the two density-functional approximations across the entire periodic table confirms that the LSDA/UGBS volumes of Table 8.2 are indeed universally applicable.

8.4 Conclusions

1. The dependence of Becke-Johnson free atomic volume on basis set and density-functional approximation is shown to be negligible, and a universal set of reference volumes at the scalar relativistic LSDA/UGBS level is given for the elements H-Lr.

2. The Becke-Johnson free atomic volumes follow the expected systematic trends across the periodic table, and are roughly proportional to the polarizabilities of free ground-state atoms.

Note that while conclusion (2) is a satisfying observation regarding the volume-polarizability relationship, it is not a necessary condition for the usefulness of volume ratios to calculate effective atom-in-molecule polarizabilities or dispersion coefficients, provided the free-atom and atom-in-molecule volumes are consistent. The free atomic volumes of Table 8.2 are expected to provide a useful reference set for density-functional dispersion models that involve volume ratios, such as Eqs. (8.1) and (8.2), regardless of basis set and density-functional approximation. Finally, we provide in the Appendix a theoretical foundation for the Becke-Johnson atomic volume definition based on the statistical theory of atoms.
8.5 Appendix: Statistical Theory Of Atomic Polarizability

In their work on atomic response functions in the statistical theory of atoms [287–289], Brandt and Lundqvist derived a formula for the dynamic (frequency-dependent) atomic dipole polarizability \( \alpha(\omega) \) [288]

\[
\alpha(\omega) = -\frac{4\pi}{\omega^2} \int_0^\infty \int_0^\infty \int_0^\infty d\rho_0(r') \phi(qr) \phi(qr') e^{-1}(q, \omega, \rho_0(r))
\]

(8.5)

where \( \rho_0(r) \) is the ground-state electron density at distance \( r \) from the nucleus; \( q \) is the wave number; \( \phi(qr) = \sqrt{2/\pi} \sin qr \); and \( e(q, \omega, \rho_0(r)) \) is the local dielectric function. This expression was derived assuming a linear response of the electron density to the external field, and neglecting effects due to the inhomogeneity of the electron density in atoms (ie., invoking the zero-gradient or local-density approximation).

Using the Drude approximation to the local dielectric function

\[
e(\mathbf{r}) = 1 - \frac{\omega_0^2(\mathbf{r})}{\omega^2}
\]

(8.6)

(ie, neglecting spatial dispersion in the atom’s response to the external field), the dynamic dipole polarizability becomes

\[
\alpha(\omega) = \int \frac{\rho_0(\mathbf{r})}{\omega_0^2(\mathbf{r}) - \omega^2} d\mathbf{r}
\]

(8.7)

and the static dipole polarizability is given by [284]

\[
\alpha = \int \rho_0(\mathbf{r}) \omega_0^{-2}(\mathbf{r}) d\mathbf{r} = \int_0^\infty \frac{4\pi \rho_0(\mathbf{r}) \omega_0^{-2}(\mathbf{r})}{\rho_0^2(\mathbf{r})} dr
\]

(8.8)

assuming that a single, position-dependent characteristic frequency \( \omega_0(\mathbf{r}) \) can be associated with each point \( \mathbf{r} \) in an atom [287].

In the statistical theory of atoms, \( \omega_0(\mathbf{r}) \) is chosen to be the local plasma frequency of a free electron gas of density \( \rho_0 \): \( \omega_0(\mathbf{r}) = \sqrt{4\pi \rho_0(\mathbf{r})} \). Thus the polarizability becomes

\[
\alpha = \int_0^{r_0} r^2 dr = \frac{1}{3} r_0^3
\]

(8.9)
where \( r_0 \) is the atomic boundary radius in the statistical theory [284, 299]. This choice of \( \omega_0(r) \) amounts to a local long-wavelength screening of the free-electron-gas response, replacing the single-particle response with a collective response characterized by the local plasma frequency [288]. Unsurprisingly, the zero-gradient approximation and the use of screened electron-gas linear response, combined with the neglect of spatial dispersion in \( \varepsilon \) and contributions due to exchange and correlation, leads to an expression for atomic polarizability equivalent to the classical uniform-density- or conducting-sphere models of polarizability.

Alternatively, Brandt and coworkers have defined the characteristic atomic frequency in Eq. (8.8) as the single-particle frequency \( \omega_{sp}(r) \) related to the strength of the atomic electrostatic field \( E(r) \) and the local potential \( V(r) \) [289, 299, 300]

\[
\omega_{sp}^2(r) = -\frac{E(r)}{r} = \frac{1}{r} \frac{dV(r)}{dr} \tag{8.10}
\]

In the statistical theory of atoms [289],

\[
\frac{dV(r)}{dr} = \frac{N_{out}(r)}{r^2} \left[ 1 + \frac{1}{N_{out}(r)} - \frac{1}{Z} \right] \tag{8.11}
\]

with \( N_{out}(r) = \int_{r'}^\infty \rho_0(r')d^3r' \). In the core regions deep inside an atom \( (r \to 0) \), where \( N_{out} \approx Z \), the last two terms in Eq. (8.11) cancel out, so that [289]

\[
\frac{dV(r)}{dr} = \frac{N_{out}}{r^2}; \quad \omega_{sp}^2(r) = r^{-3}N_{out}(r) \tag{8.12}
\]

and

\[
\alpha = \int r^3 \rho_0(r)N_{out}^{-1}(r)dr \tag{8.13}
\]

Conversely, in the outer regions of an atom \( (r \to \infty) \) where \( N_{out} \approx 0 \), only the second term in Eq. (8.11) contributes, so that

\[
\frac{dV(r)}{dr} = r^{-2}; \quad \omega_{sp}^2(r) = r^{-3} \tag{8.14}
\]

and

\[
\alpha = \int r^3 \rho_0(r)dr \tag{8.15}
\]

This result is also obtained from the exact asymptotic limit of the local (Kohn-Sham) potential, \( V(r \to \infty) = -r^{-1} \), without resorting to the approximate \( V(r) \) of Eq. (8.11) from statistical theory.
Eqs. (8.13) and (8.15) are limits for the polarizability in the inner and outer regions of an atom, respectively. The asymptotic polarizability of Eq. (8.15) is equivalent to Becke and Johnson’s definition of free atomic volume, Eq. (8.4).
Chapter 9

Conclusion

This thesis documents the development and testing of a semilocal density-functional approximation for noncovalent interactions, including a proper treatment of dispersion. The method was developed systematically and with a minimum of empiricism in order to ensure its theoretical soundness and robustness for a wide variety of chemical problems. First, standard density-functional approximations (that do not account for dispersion) were applied to rare-gas diatomics to find functionals that are suitable for combination with an explicit dispersion functional (Chapter 3). PW86 exchange and PBE correlation were identified as suitable functionals, and were subsequently combined with the XDM dispersion model to yield the PW86PBE-XDM method.

This method was applied to rare-gas binding energy curves (Chapter 3) and a newly-compiled set of 65 van der Waals complexes (Chapter 4) using a fully numerical, basis-set-free methodology. Excellent agreement with high-level reference data was found in both cases. The “Becke-Roussel” (BR) variant of the XDM model was found to perform better for intermolecular noncovalent interactions than the “exact exchange” (XX) variant, which was rationalized in terms of the different exchange holes of the two variants. The dispersion damping parameters optimized on rare-gas diatomics were found to be transferable to other noncovalent interactions including hydrogen bonding, electrostatic, stacking and “mixed” interactions.

In Chapter 5, the PW86PBE-XDM method was parametrized and tested using two standard Gaussian basis sets, aug-cc-pVDZ and aug-cc-pVTZ. The larger aug-cc-pVTZ basis set was found to have an accuracy comparable to the fully numerical approach, whereas the smaller aug-cc-pVDZ basis set was found to be slightly inferior. In Chapter 6, basis-set effects in rare-gas interactions were assessed comprehensively on repulsion energies, dispersion energies, and binding energies. All common types of Gaussian basis set were tested, and the augmented correlation-consistent basis sets aug-cc-pV\(n\)Z \((n = D, T, Q, 5)\) were found to be the most accurate basis sets for rare-gas interactions. The dispersion damping function was
reparametrized using the large aug-cc-pV5Z basis set and including counterpoise corrections, in order to separate basis-set effects from the inherent performance of the PW86PBE-XDM method.

The improved parametrization of PW86PBE-XDM from Chapter 6 was then applied to a wide variety of noncovalent interactions including hydrogen bonding, electrostatic, \( \pi-\pi \)-stacking, aliphatic, T-shaped, and other interactions in Chapter 7. Using the standard S22 and S66 benchmark sets, the method was found to be highly competitive compared to other DFAs for noncovalent interactions, even though it was not specifically optimized for these benchmark sets. Basis-set effects were investigated in detail, and 6-31+G(d), 6-31+G(2d,2p) and heavy-aug-cc-pVTZ were identified as accurate and efficient basis sets for the calculation of noncovalent interaction energies. The method was also applied to intermolecular dissociation curves, which are reproduced with high accuracy, and to angular-displaced geometries of intermolecular complexes. Compared to the latest empirical DFT-D methods, PW86PBE-XDM is found to be much more accurate at large intermolecular separation, and can reproduce whole dissociation curves with high precision.

On the methodological side, a consistent set of free atomic volumes for elements H–Lr was generated and adopted as new reference volumes for the XDM dispersion model (Chapter 8). Using the statistical theory of atoms, the Becke-Johnson definition of atomic volume as used in the XDM model was shown to be equivalent to the asymptotic polarizability of an atom.
Appendix A

Copyright Permissions
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
This is a License Agreement between Felix Kannemann ("You") and NRC Research Press ("NRC Research Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by NRC Research Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>2983250121370</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Sep 06, 2012</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>NRC Research Press</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Canadian Journal of Chemistry</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Nonempirical density-functional theory for van der Waals interactions</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Axel D. Becke, Alya A. Arabi, Felix O. Kannemann</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Nov 1, 2010</td>
</tr>
<tr>
<td>Volume number</td>
<td>88</td>
</tr>
<tr>
<td>Issue number</td>
<td>11</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>Author (original work)</td>
</tr>
<tr>
<td>Format</td>
<td>Print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Full article</td>
</tr>
<tr>
<td>Order reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Development and Benchmarking of a Semilocal Density-Functional Approximation including Dispersion</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2012</td>
</tr>
<tr>
<td>Estimated size(pages)</td>
<td>200</td>
</tr>
<tr>
<td>Total size</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Terms and Conditions</td>
<td>General Terms &amp; Conditions</td>
</tr>
</tbody>
</table>

Permission is granted upon the requester's compliance with the following terms and conditions:

1. A credit line will be prominently placed in your product(s) and include: for books the author, book title, editor, copyright holder, year of publication; for journals the author, title of article, title of journal, volume number, issue number, and the inclusive pages. The credit line must include the following wording: "© 2008 Canadian Science Publishing or its licensors. Reproduced with permission," except when an author of an original article published in 2009 or later is reproducing his/her own work.
2. The requester warrants that the material shall not be used in any manner that may be derogatory to the title, content, or authors of the material or to Canadian Science Publishing, including but not limited to an association with conduct that is fraudulent or otherwise illegal.

3. Permission is granted for the term (for Books/CDs-Shelf Life; for Internet/Intranet-In perpetuity; for all other forms of print-the life of the title) and purpose specified in your request. Once term has expired, permission to renew must be made in writing.

4. Permission granted is nonexclusive, and is valid throughout the world in English and the languages specified in your original request. A new permission must be requested for revisions of the publication under current consideration.

5. Canadian Science Publishing cannot supply the requester with the original artwork or a "clean copy."

6. If the Canadian Science Publishing material is to be translated, the following lines must be included: The authors, editors, and Canadian Science Publishing are not responsible for errors or omissions in translations.

v1.4

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500852960.**

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 061
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

**Gratia licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**
American Institute of Physics License

Terms and Conditions

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 2983241050835
Order Date Sep 06, 2012
Publisher American Institute of Physics
Publication Journal of Chemical Physics
Article Title Atomic volumes and polarizabilities in density-
functional theory
Author Felix O. Kannemann, Axel D. Becke
Online Publication Date Jan 18, 2012
Volume number 136
Issue number 3
Type of Use Thesis/Dissertation
Requestor type Author (original article)
Format Print and electronic
Portion Excerpt (> 800 words)

Will you be translating? No
Title of your thesis / dissertation Development and Benchmarking of a Semilocal Density-Functional Approximation including Dispersion
Expected completion date Dec 2012
Estimated size (number of pages) 200
Total 0.00 USD

Terms and Conditions

American Institute of Physics -- Terms and Conditions: Permissions Uses

American Institute of Physics ("AIP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: “Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], American Institute of Physics.” For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.

2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP, with permission, from another source. Under no circumstances does AIP, purport or intend to grant permission to reuse material to which it does not hold copyright.

3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed
translation rights. You may not use the Material for promotional purposes.
AIP reserves all rights not specifically granted herein.
4. The foregoing license shall not take effect unless and until AIP or its agent,
Copyright Clearance Center, receives the Payment in accordance with
Copyright Clearance Center Billing and Payment Terms and Conditions,
which are incorporated herein by reference.
5. AIP or the Copyright Clearance Center may, within two business days of
granting this license, revoke the license for any reason whatsoever, with a
full refund payable to you. Should you violate the terms of this license at
any time, AIP, American Institute of Physics, or Copyright Clearance
Center may revoke the license with no refund to you. Notice of such
revocation will be made using the contact information provided by you.
Failure to receive such notice will not nullify the revocation.
6. AIP makes no representations or warranties with respect to the Material.
You agree to indemnify and hold harmless AIP, American Institute of
Physics, and their officers, directors, employees or agents from and
against any and all claims arising out of your use of the Material other
than as specifically authorized herein.
7. The permission granted herein is personal to you and is not transferable or
assignable without the prior written permission of AIP. This license may
not be amended except in a writing signed by the party to be charged.
8. If purchase orders, acknowledgments or check endorsements are issued
on any forms containing terms and conditions which are inconsistent with
these provisions, such inconsistent terms and conditions shall be of no
force and effect. This document, including the CCC Billing and Payment
Terms and Conditions, shall be the entire agreement between the parties
relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws
of the State of New York. Both parties hereby submit to the jurisdiction of the
courts of New York County for purposes of resolving any disputes that may
arise hereunder.

If you would like to pay for this license now, please remit this license
along with your payment made payable to "COPYRIGHT CLEARANCE
CENTER" otherwise you will be invoiced within 48 hours of the license
date. Payment should be in the form of a check or money order
referencing your account number and this invoice number
RLNK50852956.
Once you receive your invoice for this order, you may pay your invoice
by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink
Customer Support: customercare@copyright.com or +1-877-622-5543
(toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please
retain this printable license for your reference. No payment is
required.
Bibliography


