Title
Short-Range Correlation Models in Electronic Structure Theory

Author
Goldey, Matthew Bryant

Publication Date
2014

Peer reviewed|Thesis/dissertation

Short-Range Correlation Models in Electronic Structure Theory

by

Matthew Bryant Goldey

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martin Head-Gordon, Chair
Professor William Miller
Professor Michael Frenklach

Spring 2014

Short-Range Correlation Models in Electronic Structure Theory

Abstract

Short-Range Correlation Models in Electronic Structure Theory

by

Matthew Bryant Goldey

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Martin Head-Gordon, Chair

Correlation methods within electronic structure theory focus on recovering the exact electron-electron interaction from the mean-field reference. For most chemical systems, including dynamic correlation, the correlation of the movement of electrons proves to be sufficient, yet exact methods for capturing dynamic correlation inherently scale polynomially with system size despite the locality of the electron cusp. This work explores a new family of methods for enhancing the locality of dynamic correlation methodologies with an aim toward improving accuracy and scalability. The introduction of range-separation into *ab initio* wavefunction methods produces short-range correlation methodologies, which can be supplemented with much faster approximate methods for long-range interactions.

First, I examine attenuation of second-order Møller-Plesset perturbation theory (MP2) in the aug-cc-pVDZ basis. MP2 treats electron correlation at low computational cost, but suffers from basis set superposition error (BSSE) and fundamental inaccuracies in long-range contributions. The cost differential between complete basis set (CBS) and small basis MP2 restricts system sizes where BSSE can be removed. Range-separation of MP2 could yield more tractable and/or accurate forms for short- and long-range correlation. Retaining only short-range contributions proves to be effective for MP2 in the small aug-cc-pVDZ (aDZ) basis. Using one range-separation parameter within either the complementary error function (erfc) or a sum of two error functions (terfc), superior behavior is obtained versus both MP2/aDZ and MP2/CBS for inter- and intra-molecular test sets. Attenuation of the long-range helps to cancel both BSSE and intrinsic MP2 errors. Direct scaling of the MP2 correlation energy (SMP2) proves useful as well. The resulting SMP2/aDZ, MP2(erfc, aDZ), and MP2(terfc, aDZ) methods perform far better than MP2/aDZ across systems with hydrogen-bonding, dispersion, and mixed interactions at a fraction of MP2/CBS computational cost.

Second, attenuated MP2 is developed within the larger aug-cc-pVTZ (aTZ) basis set for inter- and intramolecular non-bonded interactions. A single attenuation parameter is optimized on the S66 database of 66 intermolecular interactions, leading to a very large RMS error reduction by a factor of greater than 5 relative to standard MP2/aTZ. Attenuation introduces an error of opposite sign to basis set superposition error (BSSE) and overestimation of dispersion interactions in finite

basis MP2. A variety of tests including the S22 set, conformer energies of peptides, alkanes, sugars, sulfate-water clusters, and the coronene dimer establish the transferability of the MP2(terfc, aTZ) model to other inter and intra-molecular interactions. Direct comparisons against attenuation in the smaller aug-cc-pVDZ basis shows that MP2(terfc, aTZ) often significantly outperforms MP2(terfc, aDZ), although at higher computational cost. MP2(terfc, aDZ) and MP2(terfc, aTZ) often outperform MP2 at the complete basis set limit. Comparison of the two attenuated MP2 models against each other and against attenuation using non-augmented basis sets gives insight into the error cancellation responsible for their remarkable success.

Third, I present an improved algorithm for single-node multi-threaded computation of the correlation energy using resolution of the identity second-order Møller-Plesset perturbation theory (RI-MP2). This algorithm is based on shared memory parallelization of the rate-limiting steps and an overall reduction in the number of disk reads. The requisite fifth-order computation in RI-MP2 calculations is efficiently parallelized within this algorithm, with improvements in overall parallel efficiency as the system size increases. Fourth-order steps are also parallelized. As an application, I present energies and timings for several large, noncovalently interacting systems with this algorithm, and demonstrate that the RI-MP2 cost is still typically less than 40% of the underlying self consistent field (SCF) calculation. The attenuated RI-MP2 energy is also implemented with this algorithm, and some new large-scale tests of this method are reported. The attenuated RI-MP2(terfc, aug-cc-pVDZ) method yields excellent agreement with benchmark values for the L7 database (R. Sedlak et al., J. Chem. Theory Comput. 2013, 9, 3364) and 10 tetrapeptide conformers (L. Goerigk et al., Phys. Chem. Chem. Phys. 2013, 15, 7028), with at least a 90% reduction in the root-mean-squared (RMS) error versus RI-MP2/aug-cc-pVDZ.

Fourth, semi-empirical spin-component scaled (SCS) attenuated MP2 is developed for treating both bonded and nonbonded interactions. SCS-MP2 improves the treatment of thermochemistry and noncovalent interactions relative to MP2, although the optimal scaling coefficients are quite different for thermochemistry versus noncovalent interactions. This work reconciles these two different scaling regimes for SCS-MP2 by using two different length scales for electronic attenuation of the two spin components. The attenuation parameters and scaling coefficients are optimized in the aug-cc-pVTZ (aTZ) basis using the S66 database of intermolecular interactions and the W4-11 database of thermochemistry. Transferability tests are performed for atomization energies and barrier heights, as well as on further test sets for inter- and intramolecular interactions. SCS dual-attenuated MP2 in the aTZ basis, SCS-MP2(2terfc, aTZ), performs similarly to SCS-MP2/aTZ for thermochemistry while frequently outperforming MP2 at the complete basis set limit (CBS) for nonbonded interactions.

Finally, I examine the performance of attenuated MP2 for noncovalent interactions using basis sets that range as high as augmented triple (T) and quadruple (Q) zeta with TQ extrapolation towards the complete basis set (CBS) limit. By comparing training and testing performance as a function of basis set size, the effectiveness of attenuation as a function of basis set can be assessed. While attenuated MP2 with TQ extrapolation improves systematically over MP2, there are at most small improvements over attenuated MP2 in the aug-cc-pVTZ basis. Augmented functions are crucial for the success of attenuated MP2.

To my wife,
Rebecca

Contents

Contents **ii**

List of Figures **iv**

List of Tables **vi**

1 Introduction **1**
1.1 Common models 1
1.1.1 The Born-Oppenheimer Approximation 2
1.1.2 The Hartree-Fock approximation 2
1.1.3 Møller-Plesset perturbation theory 4
1.1.4 Configuration Interaction 5
1.1.5 Coupled Cluster theory 5
1.2 Choice of a finite basis 6
1.2.1 Basis set expansion 6
1.2.2 Convergence with basis set size 6
1.3 Density Functional Theory 8
1.3.1 Dispersion corrected DFT 8
1.3.2 Range-separated hybrids 9
1.4 Extending the reach of correlation methods 10
1.4.1 The resolution of the identity or density-fitting approximation 10
1.4.2 Spin-component analyses 11
1.4.3 Adjusting the treatment of long-range interactions 12
1.5 Aims of this work 13

2 Attenuating Away The Errors in Inter- and Intra-Molecular Interactions from Second Order Møller-Plesset Calculations in the Small aug-cc-pVDZ Basis Set **15**

3 Attenuated Second-Order Møller-Plesset Perturbation Theory: Performance within the aug-cc-pVTZ Basis **25**
3.1 Introduction 25
3.2 Methods 27

3.3 Parameter optimization . . . 27
3.4 Tests of transferability . . . 32
3.5 Conclusions . . . 35

4 Shared Memory Multiprocessing Implementation of Resolution-of-the-Identity Second-Order Møller-Plesset Perturbation Theory with Attenuated and Unattenuated Results for Intermolecular Interactions between Large Molecules 37
4.1 Introduction . . . 37
4.2 Algorithm . . . 39
4.3 Parallel Performance . . . 41
4.4 Applications . . . 43
4.5 Conclusions . . . 46

5 Separate Electronic Attenuation Allowing a Spin-Component Scaled Second Order Møller-Plesset Theory to Be Effective for Both Thermochemistry and Non-Covalent Interactions 47
5.1 Introduction . . . 47
5.2 Methods . . . 50
5.3 Training . . . 50
5.4 Tests . . . 53
5.5 Conclusions . . . 55

6 Convergence of attenuated MP2 to the complete basis set limit: Improving MP2 for long-range interactions without basis set incompleteness 58
6.1 Introduction . . . 58
6.2 Methods . . . 60
6.3 Training . . . 61
6.4 Transferability tests . . . 63
6.5 Conclusions . . . 63

7 Conclusion 70
7.1 Summary of attenuated MP2 methods . . . 70
7.2 Future Work . . . 71
7.2.1 Algorithm design . . . 71
7.2.2 Long-range dispersion correction . . . 71
7.2.3 Short-range correlation methods . . . 71
7.2.4 Application to weakly interacting systems . . . 72

Bibliography 73

A Performance of attenuated MP2 and other methods in the aug-cc-pVDZ basis 85

B Code for generating terf interpolation tables 96

List of Figures

1.1 The convergence of the HF and MP2 energies for the N_2 molecule with cardinal number of basis set are presented herein, reproduced from reference[1]. The correlation energy is plotted on the left in mE_h. The errors (in mE_h) for the MP2 (solid line) and HF (dashed line) energies are presented on the right versus cardinal number. 7

2.1 Performance on S66 Dataset for MP2(terfc, aDZ) with both unscaled, **I**, and scaled, **II**, variants over the range $r_0 = 0.05\text{Å} \rightarrow r_0 = 4.00\text{Å}$, which spans from the HF limit (4.0 kcal mol^{-1}) to the unattenuated MP2 limit (2.7 kcal mol^{-1}). 19

2.2 Performance on S66 Dataset for MP2(erfc, aDZ) with both unscaled, **III**, and scaled, **IV**, variants over the range $\omega = 0.01\text{Å}^{-1} \rightarrow \omega = 2.00\text{Å}^{-1}$, which spans from the unattenuated MP2 limit (2.7 kcal mol^{-1}) and approaches the HF limit of 4.0 kcal mol^{-1}. 20

2.3 Geometries from S22x5 with MP2(terfc, aDZ)(**I**), SMP2/aDZ, and MP2/aDZ. For comparison, CCSD(T)/CBS is provided. 23

3.1 The partitioning of the interelectron repulsion operator into short range and long-range components based on the long-range terf function defined in Eq. (4.1) and its short-range complement, terfc, defined in Eq. (4.2). With these definitions, $\text{terf}(r, r_0)r^{-1}$ has zero first and second derivatives in the small r limit. Therefore the short-range interelectron repulsion, $\text{terfc}(r, r_0)r^{-1}$ behaves as a smoothly shifted r^{-1}. The models developed in this paper retain only the short-range term in the MP2 energy, and optimize the single parameter r_0 to reproduce benchmark intermolecular interactions. . 28

3.2 Effect of augmented functions on root mean squared deviation of truncated MP2 methods for training set S66 with terfc-attenuation. As $r_0 \rightarrow 4.0\text{Å}$, attenuated MP2 converges to the unattenuated result. As $r_0 \rightarrow 0\text{Å}$, attenuated MP2 approaches HF results. 30

3.3 Effect of counterpoise correction on root mean squared deviation of truncated MP2 methods for training set S66 with terfc-attenuation. As $r_0 \rightarrow 4.0\text{Å}$, attenuated MP2 converges to the unattenuated result. As $r_0 \rightarrow 0\text{Å}$, attenuated MP2 approaches HF results. 31

3.4 Root mean squared deviations for MP2(terfc, aTZ) (left) and MP2(terfc, aTZ-CP) (right) versus r_0 for various subsets of the S66 database 32

4.1 Strong scaling performance of the RI-MP2 parallel algorithm presented herein for polyglycines using the cc-pVDZ AO basis set. The overall speedup is plotted on the left, whereas the speed increase for Function 4, the formation of the 4-center integrals in the MO basis, is shown on the right. 42

5.1 Weighted RMSD (kcal/mol) on S66 and W4-11 benchmark databases, as defined in Equation 5.7, evaluated as a function of the bonded attenuation length, $r_0^{(1)}$, and the non-bonded attenuation length, $r_0^{(2)}$. At each point the optimal linear coefficients are determined to obtain the value of the objective function. Note that the domain where $r_0^{(1)} \geq r_0^{(2)}$ is forbidden in Equation 5.7. The best values of $r_0^{(1)}$ and $r_0^{(2)}$ lie in a narrow valley with the minimum at $r_0^{(1)} = 0.75$Å, and $r_0^{(2)} = 1.05$Å 52
5.2 Root-mean-squared-deviations (RMSDs) in kcal/mol for MP2/aTZ, SCS-MP2/aTZ, MP2(terfc, aTZ), and SCS-MP2(2terfc, aTZ) for thermochemistry datasets 54
5.3 Root-mean-squared-deviations (RMSDs) kcal/mol for MP2/aTZ, SCS-MP2/aTZ, MP2(terfc, aTZ), SCS-MP2(2terfc, aTZ), and MP2/CBS[1] for noncovalent interaction database . . . 55
5.4 Growth of error in atomization energy (kcal/mol) as a function of alkane size 57

6.1 Root-mean-squared deviation (kcal mol^{-1}) on the 66 intermolecular interactions of the S66 dataset versus r_0/Å for attenuated MP2 with Dunning style basis sets 62

List of Tables

2.1 Root-mean-squared deviations, standard deviations of error, average, and mean unsigned errors for the S66 Dataset (kcal mol^{-1}) 18
2.2 Root-mean-squared deviations, standard deviations of error, average, and mean unsigned errors for the S22 Dataset (kcal mol^{-1}) 21
2.3 Root-mean-squared deviations for protein subsets of the P76 database (kcal mol^{-1}) . . 22
2.4 Mean absolute deviations and root-mean-squared deviations from RI-MP2/CBS on alanine tetrapeptide conformers (kcal mol^{-1}) 22

3.1 Root-mean-squared deviations(RMSD), average, and mean unsigned errors on the S66 database (kcal mol^{-1}) 29
3.2 Root-mean-squared deviations, average, and mean unsigned errors on the S22 database (kcal mol^{-1}) 33
3.3 Root-mean-squared deviations for different protein subsets of the P76 database (kcal mol^{-1}) 33
3.4 Root-mean-squared deviations and average errors on the ACONF database (kcal mol^{-1}) 33
3.5 Root-mean-squared deviations and average errors on the SCONF database (kcal mol^{-1}) 34
3.6 Root-mean-squared deviations and average errors on the CYCONF database (kcal mol^{-1}) 34
3.7 Root-mean-squared deviations for relative energies of methods on the SW49 database (kcal mol^{-1}) 35
3.8 Root-mean-squared deviations for binding energies of methods on the SW49 database (kcal mol^{-1}) 35
3.9 Binding energy of the parallel-displaced coronene dimer (kcal mol^{-1}) 36

4.1 RI-MP2 Energy Algorithm. 39
4.2 Growth of the rate-limiting step (Function 4) of RI-MP2 for polyglycines using the cc-pVDZ AO basis set. Relative cost is between Function 4 and the overall RI-MP2 time when using one core. 42
4.3 Timings for the L7 database using RI-MP2/aDZ with 64 cores. 44
4.4 Energies for the L7 database and error metrics, including root-mean-squared deviations (RMSD), mean signed errors (MSE), mean unsigned errors (MUE), and maximum deviations (MAX) in kcal/mol. 44

4.5 Timings (in minutes) for RI-MP2/aTZ on the tetrapeptide model conformers with 64 cores. 45
4.6 Energies for the tetrapeptide model conformers (relative to β_a) and root-mean-squared deviations. 45

5.1 Error statistics on the W4-11 non-multireference training set versus W4 benchmarks (in kcal/mol) with root mean-squared deviations (RMSD) for the total atomization energies (TAE), bond dissociation energies (BDE), heavy atom transfers (HAT), isomerization energies (ISO), and nucleophilic substitution reaction (SN) subsets, with total RMSD, mean-signed error (MSE), mean-unsigned error (MUE), and maximum error (MAX) 51
5.2 Error statistics on the S66 database versus CCSD(T)/CBS benchmarks (in kcal/mol) with root mean-squared deviations (RMSD) for the hydrogen-bonded (H-bonds), dispersion-bonded (disp.), and mixed subsets, with total RMSD, mean-signed error (MSE), mean-unsigned error (MUE), and maximum error (MAX) 53
5.3 Performance for MP2/aTZ variants versus L7 benchmarks (in kcal/mol) with root mean-squared deviation (RMSD), mean-signed error (MSE), mean-unsigned error (MUE), and maximum error (MAX) 56

6.1 Performance (kcal mol^{-1}) of MP2 in various basis sets for the S66 database, including root-mean-squared deviation (RMSD) for the database, the hydrogen-bonded subset, the dispersion subset, and the mixed subset, as well as mean-signed error (MSE) and mean-unsigned error (MUE). Average finite basis set-related error (FBSE) is reported for SCF and SCF+MP2 relative to the SCF/aQZ and SCF+MP2/CBS energies. Reference SCF+MP2/CBS energies were taken from the Benchmark Energy and Geometry DataBase (BEGDB.com)[2]. 65
6.2 Performance (in kcal mol^{-1}) of attenuated MP2 with optimal r_0/Å using calendar basis sets for the S66 database with overall root-mean-squared deviation (RMSD), mean-signed error (MSE) and mean-unsigned error (MUE), as well as RMSDs for the hydrogen-bonded, dispersion, and mixed interaction subsets 66
6.3 Performance (in kcal mol^{-1}) of attenuated MP2 with optimal r_0/Å using standard Dunning basis sets with T→Q extrapolated complete basis set estimates for the S66 database with overall root-mean-squared deviation (RMSD), mean-signed error (MSE) and mean-unsigned error (MUE), as well as RMSDs for the hydrogen-bonded, dispersion, and mixed interaction subsets. 66
6.4 Performance (in kcal mol^{-1}) of attenuated MP2 with optimal r_0/Å using Pople-style and Karlsruhe basis sets for the S66 database with overall root-mean-squared deviation (RMSD), mean-signed error (MSE) and mean-unsigned error (MUE), as well as RMSDs for the hydrogen-bonded, dispersion, and mixed interaction subsets 67
6.5 Root-mean-squared deviations (RMSDs) in kcal mol^{-1} for attenuated and unattenuated MP2 in the augmented Dunning basis sets on intramolecular conformational energetics databases 67

6.6 Binding energies for A24 database of attenuated and unattenuated MP2 in aDZ, aTZ, aQZ, and aTQZ basis sets with root-mean-squared deviation (RMSD), mean-signed error (MSE), and mean-unsigned error (MUE) in (kcal mol^{-1}) 68
6.7 Statistics for the performance (kcal mol^{-1}) of attenuated and unattenuated MP2 in aDZ, aTZ, aQZ, and aTQZ basis sets on the 22 intermolecular interactions defining the S22 database with root-mean-squared deviations (RMSD) for hydrogen-bonded, dispersion, and mixed subsets, as well as overall RMSD, mean-signed error (MSE), and mean-unsigned error (MUE) 69

A.1 Energetics for the S66 Hydrogen-Bonding Subset (kcal mol^{-1}) 86
A.2 Energetics for the S66 Dispersion Subset (kcal mol^{-1}) 87
A.3 Energetics for the S66 Mixed Interaction Subset (kcal mol^{-1}) 88
A.4 Energetics for the S22 Dataset (kcal mol^{-1}) 89
A.5 Energetics for phenylalanine-glycine-glycine conformers of P76 database(kcal mol^{-1}) 90
A.6 Energetics for glycine-phenylalanine-alanine conformers of P76 database(kcal mol^{-1}) . 90
A.7 Energetics for glycine-glycine-phenylalanine conformers of P76 database(kcal mol^{-1}) 91
A.8 Energetics for tryptophan-glycine conformers of P76 database(kcal mol^{-1}) 91
A.9 Energetics for tryptophan-glycine-glycine conformers of P76 database(kcal mol^{-1}) . . 92
A.10 Energetics for 27 reference alanine tetrapeptide conformers(kcal mol^{-1}) 93
A.11 S22x5 geometries for Water Dimer(kcal mol^{-1}) 94
A.12 S22x5 geometries for Parallel-Displaced Benzene Dimer(kcal mol^{-1}) 94
A.13 S22x5 geometries for T-Shaped Benzene Dimer(kcal mol^{-1}) 94
A.14 S22x5 geometries for Ammonia Dimer(kcal mol^{-1}) 95

Acknowledgments

First, I wish to thank my advisor, Martin Head-Gordon, for his long-suffering patience and sound direction, without which this work would not have happened. I am indebted to Robert DiStasio, Jr. and Paul Zimmerman for their mentorship and Adrian Mak for encouragement. I would like to thank Tony Dutoi, Evgeny Epifanovsky, and Yihan Shao for assistance in coding up different projects. Their standards of excellence for their own work have made my work and the work of others easier. I would also like to thank my parents for their years of encouragement and love. Lastly, I would not be here but for my wife, Rebecca, whose support and friendship has made all this possible.

Chapter 1

Introduction

> The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.
>
> *Paul Dirac*

The study of molecules and atoms is chemistry, which has as its theoretical groundwork the physical interactions between particles. Electronic structure theory (EST) models the properties of molecules, given the basic physical laws that constituent particles, electrons and nuclei, obey. While nuclear motion often requires quantum mechanical treatment, electrons have de Broglie wavelengths that invoke quantum mechanical effects for the simplest of cases - requiring explicit, quantum treatment of chemical systems. Full quantum mechanical treatment for molecules requires the solution of the Schrödinger equation, where the essential descriptive quantity is the wavefunction, or probability amplitude, Ψ. Given the wavefunction, all observable properties are represented as operators upon this wavefunction, which have eigenvalues corresponding to measurable properties, as the total energy, E, corresponds to the Hamiltonian, $\hat{H}$.

$$\hat{H}\Psi = E\Psi \tag{1.1}$$

A molecular Hamiltonian consists of kinetic ($\hat{T}$) and potential ($\hat{V}$) energy terms for nuclei (N) and electrons (e), according to each coordinate system, nuclear ($\vec{\mathrm{R}}$) or electronic ($\vec{\mathrm{r}}$).

$$\hat{H}(\vec{\mathrm{r}},\vec{\mathrm{R}}) = \hat{T}_N(\vec{\mathrm{R}}) + \hat{T}_e(\vec{\mathrm{r}}) + \hat{V}_{eN}(\vec{\mathrm{r}},\vec{\mathrm{R}}) + \hat{V}_{ee}(\vec{\mathrm{r}}) + \hat{V}_{NN}(\vec{\mathrm{R}}) \tag{1.2}$$

1.1 Common models

Accurate treatment of quantum mechanical systems requires the solution of the *ab initio* Schrödinger equation, which is untenable for the majority of systems of chemical interest. As such, we are constrained to use theoretical models which approximate the Schrödinger equation systematically[3].

1.1.1 The Born-Oppenheimer Approximation

The first approximation commonly used to simplify the Schrödinger equation is the Born-Oppenheimer approximation, wherein the electronic and nuclear degrees of freedom are separated[4], meaning that the wavefunction is separated into electronic and nuclear wavefunctions.

$$\Psi_{\mathrm{BO}} = \varphi(\mathrm{r};\mathrm{R})\chi(\mathrm{R}) \tag{1.3}$$

Since electronic motions occur on a time-scale much faster than the motion of nuclei such that the electronic wavefunction typically varies smoothly with R, this approximation holds for much of normal chemistry (with a notable exception being the conical intersections where different electronic states cross). The Born-Oppenheimer approximation separates the Hamiltonian as well as the wavefunction. The primary remaining problem is then the solution of the Schrödinger equation for electronic motion, based upon the electronic wavefunction and Hamiltonian, which depend parametrically on nuclear coordinates.

$$\hat{H}(\vec{\mathrm{r}};\vec{\mathrm{R}})\varphi_e(\vec{\mathrm{r}};\vec{\mathrm{R}}) = E_e\varphi_e(\vec{\mathrm{r}};\vec{\mathrm{R}}) \tag{1.4}$$

The electronic Hamiltonian is simply a function of the kinetic energy operator, the nuclear potential, and the electron-electron potential, which proves the most difficult.

$$\hat{H}(\vec{r};\vec{\mathrm{R}}) = \hat{T}_e(\vec{\mathrm{r}}) + \hat{V}_{eN}(\vec{\mathrm{r}};\vec{\mathrm{R}}) + \hat{V}_{ee}(\vec{\mathrm{r}}) \tag{1.5}$$

The Born-Oppenheimer approximation discards terms corresponding to non-adiabatic couplings between the electronic and nuclear motions due to the separation of the nuclear and electronic wavefunctions, though some research suggests that the exact wavefunction can be factorized into nuclear and electronic wavefunctions, albeit in a different manner[5].

1.1.2 The Hartree-Fock approximation

Even given the Born-Oppenheimer approximation, solving the Schrödinger equation for molecules remains impractical for all but the simplest of cases due to the difficult many-body problem of electron-electron interactions. The simplest physically meaningful wavefuction is used in the Hartree-Fock method. From chemical intuition, a reasonable basis for a wavefunction for chemicals consists of molecular orbitals or a linear combination of atomic orbitals, which can be used to construct a many-body wavefunction. Additionally, from the properties of fermions, we know that the wavefunction for a system should be antisymmetric under exchange of electrons, which can be enforced through the use of determinants. The simplest wavefunction representation of an n-electron system consists of a determinant of electronic wavefunctions, called a Slater determinant, which is represented in equation 1.7.

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n) = (n!)^{-\frac{1}{2}}\begin{vmatrix} \chi_i(\mathbf{r}_1) & \chi_j(\mathbf{r}_1) & \ldots & \chi_k(\mathbf{r}_1) \\ \chi_i(\mathbf{r}_2) & \chi_j(\mathbf{r}_2) & \ldots & \chi_k(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(\mathbf{r}_n) & \chi_j(\mathbf{r}_n) & \ldots & \chi_k(\mathbf{r}_n) \end{vmatrix} \tag{1.6}$$

$$|\Psi\rangle = |\chi_1\chi_2\cdots\chi_n\rangle \tag{1.7}$$

The Hartree-Fock ansatz approximates the many-body problem of electron-electron interactions through the generation of a "mean-field" potential. The specific electron-electron interaction is communicated through an average potential for the system, which generates a one-electron operator, $f(i)$, called the Fock operator (1.8), which in turn produces the Hartree-Fock equations (1.9).

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{\mathrm{R}_i A} + \mathrm{v}^{HF}(i) \tag{1.8}$$

$$f(i)\chi(\mathbf{r}_i) = \varepsilon\chi(\mathbf{r}_i) \tag{1.9}$$

The apparent field experienced by the individual electron averages the effects of all other electrons. This produces a nonlinear problem since these motions remain interdependent, but this is normally soluble using iterative methods. Despite the significant reduction in complexity, the Hartree-Fock potential recovers an electronic energy that often exceeds 99% of the exact answer.

The Hartree-Fock energy is formed by the expectation value of the Hamiltonian, requiring only the Fock operator, consisting of the one-electron Hamiltonian and the "mean-field" potential, as represented in the relevant matrix elements from the many-body wavefunction.

$$E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle = \sum_i \langle\chi_i|\hat{h}|\chi_i\rangle + \frac{1}{2}\sum_{ij}\langle\chi_i\chi_j||\chi_i\chi_j\rangle \tag{1.10}$$

$$\hat{h}(1)\chi_i(1) + \sum_{j\neq i}\left[\int d\mathbf{r}_2|\chi_j(2)|^2\mathrm{R}_{12}^{-1}\right]\chi_i(1) - \sum_{j\neq i}\left[\int d\mathbf{r}_2\chi_j^*(2)\chi_i(2)\mathrm{R}_{12}^{-1}\right]\chi_j(1) = \varepsilon_i\chi_i(1) \tag{1.11}$$

$$\hat{h}(1) = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{\mathrm{R}_{1A}} \tag{1.12}$$

The minimization of this energy is bound by the variational principle (1.17). Given any trial wavefunction, $\tilde{\Phi}$, we can expand it in terms of the exact solutions to our system, $\{\Phi_\alpha\}$. Since the resultant expression contains energies ε_α that are larger than the ground state ε_0 for all solutions, this requires that any trial wavefunction will have an energy that cannot be lower than the exact ground state solution.

$$\langle\tilde{\Phi}|\tilde{\Phi}\rangle = \sum_\alpha \langle\tilde{\Phi}|\Phi_\alpha\rangle\langle\Phi_\alpha|\tilde{\Phi}\rangle \tag{1.13}$$

$$\langle\tilde{\Phi}|\tilde{\Phi}\rangle = \sum_\alpha |\langle\Phi_\alpha|\tilde{\Phi}\rangle|^2 \tag{1.14}$$

$$\langle\tilde{\Phi}|\hat{H}|\tilde{\Phi}\rangle = \sum_{\alpha\beta}\langle\tilde{\Phi}|\Phi_\alpha\rangle\langle\Phi_\alpha|\hat{H}|\Phi_\beta\rangle\langle\Phi_\beta|\tilde{\Phi}\rangle \tag{1.15}$$

$$\langle\tilde{\Phi}|\hat{H}|\tilde{\Phi}\rangle = \sum_\alpha \varepsilon_\alpha|\langle\Phi_\alpha|\tilde{\Phi}\rangle|^2 \tag{1.16}$$

$$\langle\tilde{\Phi}|\hat{H}|\tilde{\Phi}\rangle \geq \sum_\alpha \varepsilon_0|\langle\Phi_\alpha|\tilde{\Phi}\rangle|^2 = \varepsilon_0 \tag{1.17}$$

The minimization of the Hartree-Fock energy corresponds to the orthogonalization of canonical molecular orbitals, represented in a specific basis using a coefficient matrix **c**.

$$\hat{H}\mathbf{c} = E\mathbf{S}\mathbf{c} \tag{1.18}$$

While the Hartree-Fock method recovers greater than 99% of the electronic energy, the remaining energetic lowering, corresponding to the correlation of electronic motions, is not recovered and is critical for describing molecules accurately. Adequately and efficiently describing the correlation energy is the preeminent challenge of electronic structure theory. Various systematic approximations which can be used to approach the exact wavefunction and energy are presented in sections 1.1.3, 1.1.4, and 1.1.5

1.1.3 Møller-Plesset perturbation theory

Since Hartree-Fock theory includes electron-electron interaction in an approximate manner, the full electronic energy is not recovered, and the wavefunction only roughly approximates the exact wavefunction. The explicit electron-electron interaction becomes the natural focus for improving the wavefunction and the resultant energy. The simplest method for improving this treatment is the inclusion of electron-electron interactions via perturbation theory.

Perturbation theory relies upon a number of approximations but most importantly assumes that the interaction between the electrons (correlation) remains small – and this interaction (the fluctuation potential corresponding to the specific $1/r$ between electrons) is used as the perturbation. While the choice of reference state results in a number of different theories with differing advantages, the most common choice is the Møller and Plesset form of Rayleigh-Schrödinger perturbation theory[6,7], which takes as its reference the Hartree-Fock energy. The perturbative terms that result from this expansion are not necessarily convergent, but the lowest order correction, second-order Møller-Plesset perturbation theory (MP2), frequently proves a useful approximation to the correlation energy. Expanding the Hamiltonian, energy, and wavefunction in terms of powers of a perturbation, the corrections to the reference energy and wavefunction are trivially obtained in mathematical form, though at ever-greater computational cost.

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{1.19}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots \tag{1.20}$$

$$|\psi_i\rangle = |\psi_i^{(0)}\rangle + \lambda|\psi_i^{(1)}\rangle + \lambda^2|\psi_i^{(2)}\rangle + \ldots \tag{1.21}$$

The first-order wavefunction, expanded in terms of the other zero-order solutions to the HF equations, generates the second-order energy, here represented as a matrix element between a doubly-excited determinant and the ground state.

$$E_i^{(2)} = \langle\psi_i^{(0)}|V|\psi_i^{(1)}\rangle \tag{1.22}$$

$$E_i^{(2)} = -\sum_{n\neq i}\frac{\langle\psi_i^{(0)}|V|\psi_n^{(0)}\rangle^2}{E_i^{(0)} - E_n^{(0)}} = \frac{1}{4}\sum_{ij}^{\text{occ}}\sum_{ab}^{\text{virt}}\left[\frac{\langle ij||ab\rangle^2}{\varepsilon_i + \varepsilon_j + \varepsilon_a - \varepsilon_b}\right] \tag{1.23}$$

1.1.4 Configuration Interaction

The most dominant direction initially explored for improving the HF wavefunction was the configuration interaction method (CI), which generates improved wavefunctions through occupied/virtual substitutions of the HF reference[8–10] , usefully conceptualized as excitations. The wavefunction that results from this expansion (Equation 1.24) reproduces the exact wavefunction and the exact energy for the electronic Schrödinger equation (within a finite basis) at the cost of examining all possible determinants, a factorial problem which grows rapidly intractable. As a result, approximate versions of CI using truncated levels of excited configurations provide a useful ansatz for chemical problems, but these methods lack size extensivity, which is to say that they fail to achieve energy additivity for a system composed of non-interacting constituents[1,11], though the rarely achieved full (untruncated) configuration interaction limit does not suffer from this problem.

$$\Psi_{\mathrm{CI}} = \Psi_0 + c_i^a \Psi_i^a + c_{ij}^{ab} \Psi_{ij}^{ab} + c_{ijk}^{abc} \Psi_{ijk}^{abc} + \dots \tag{1.24}$$

Corrections which approximate the missing terms[12] are occasionally used to remedy these systems in practice, but the CI ansätze are naturally suited to treatment of excited states[13], as well as problems where single-configurations are not a satisfactory reference[14–16].

1.1.5 Coupled Cluster theory

Coupled cluster theory (CC) constructs a wavefunction from excitations out of the HF reference using an exponential excitation operator[17,18].

$$|\psi\rangle = e^T |\phi\rangle \tag{1.25}$$

The exponentiated excitation operator constructs all possible determinants through single, double, triple, etc. excitations of the mean-field reference.

$$e^T = 1 + T + \frac{1}{2}T^2 + \frac{1}{3!}T^3 + \dots \tag{1.26}$$

$$T = T_1 + T_2 + T_3 + T_4 + \dots \tag{1.27}$$

The action of the excitation operator on the reference produces the excited determinants with corresponding amplitudes.

$$T_1|\phi\rangle = \sum_{ia} t_i^a |\phi_i^a\rangle \tag{1.28}$$

$$T_2|\phi\rangle = \frac{1}{4}\sum_{ijab} t_{ij}^{ab} |\phi_{ij}^{ab}\rangle \tag{1.29}$$

By projection onto the reference determinant, the energy expression for coupled cluster theory is generated.

$$E_{\mathrm{corr}} = \langle\phi|H_0(\frac{1}{2}T_1^2 + T_2)|\phi\rangle = \sum_{ijab}\left[\frac{1}{2}t_i^a t_j^b \langle ij||ab\rangle + \frac{1}{4}t_{ij}^{ab}\langle ij||ab\rangle\right] \tag{1.30}$$

The main challenge of coupled cluster theory, therefore, becomes the determination of the t_{ij}^{ab}, which requires the solution of the equations formed via projecting with the series of excited determinants. Similar to the necessary truncation of CI, CC theories must be truncated to a given level of excitation in practice. By design, this truncation results in an ansatz which is size-extensive at any level of theory[1].

1.2 Choice of a finite basis

The wavefunction within EST is typically represented within a basis, converting complex, integro-differential equations into matrix algebra. The cost of evaluating matrix elements depends upon the choice of the underlying basis.

1.2.1 Basis set expansion

The natural choice of basis for molecular problems remains atomic orbitals, where molecular orbitals are constructed via a linear combination of atomic orbitals. Slater type orbitals resemble hydrogenic orbitals, of the form $\phi(\mathbf{r}-\mathbf{R}) = (\frac{\zeta^3}{\pi})^{\frac{1}{2}} e^{-\zeta|\mathbf{r}-\mathbf{R}|}$ for an 's' orbital about an atom at position $\mathbf{R}$. These orbitals reproduce atomic quantities well but are computationally inefficient for large calculation. Instead, combinations of Gaussian orbitals fitted to atom-like Slater orbitals are used in practice. The equivalent 's'-type orbital form is $\phi(\mathbf{r}-\mathbf{R}) = (\frac{2\alpha}{\pi})^{\frac{3}{4}} e^{-\alpha|\mathbf{r}-\mathbf{R}|^2}$ for Gaussian orbitals. Significant amounts of effort have gone into the generation of efficient algorithms for analytically evaluating one- and two-electron matrix elements over Gaussian basis functions[19].

1.2.2 Convergence with basis set size

Any given basis has a certain amount of incompleteness associated with the representation of quantum mechanical operators and the wavefunction. This incompleteness causes a myriad of complications for model chemistries. Unless one is able to attain the complete basis set limit (CBS), the basis chosen must be held constant for comparing calculations. Correlated wavefunction calculations contain errors that scale $O(N^{-1})$ with the number of atomic orbitals, N[20]. Unfortunately, the cost of most correlation methods scales polynomially with the number of basis functions, $O(N^4)$ for MP2 and CCSD(T). Gaussian basis sets suitable for efficiently treating the electronic Schrödinger equation have been parametrized and are in common use[21–31]. Correlation consistent basis sets, *e.g.* the correlation consistent polarized valence double zeta basis set (cc-pVDZ), increase in size systematically with the cardinal number of the AO basis set. With each increase in cardinal number, another level of polarization functions is added as well as additional basis functions for all existing angular momentum numbers. For instance, by adding 1s1p1d1f to the 3s2p1d cc-pVDZ basis set (for second row atoms), the 4s3p2d1f cc-pVTZ basis set is generated. As the cardinal number is increased from X-1 to X, $(X+1)^2$ basis functions are added. Generating all AO integrals scales with the fourth power of the number of atomic orbitals, N^4, or, in this case, $(X+1)^8$. These basis sets typically provide a systematic framework for increasing the quality. By

adding more basis functions, most computed quantities such as the energy change until the basis is saturated or complete. This convergence occurs relatively quickly for HF, yet accurate description of the Coulomb cusp, which is necessary for any correlation treatment, requires substantively larger basis sets and actually converges at a significantly slower rate, as seen in figure 1.1. For SCF

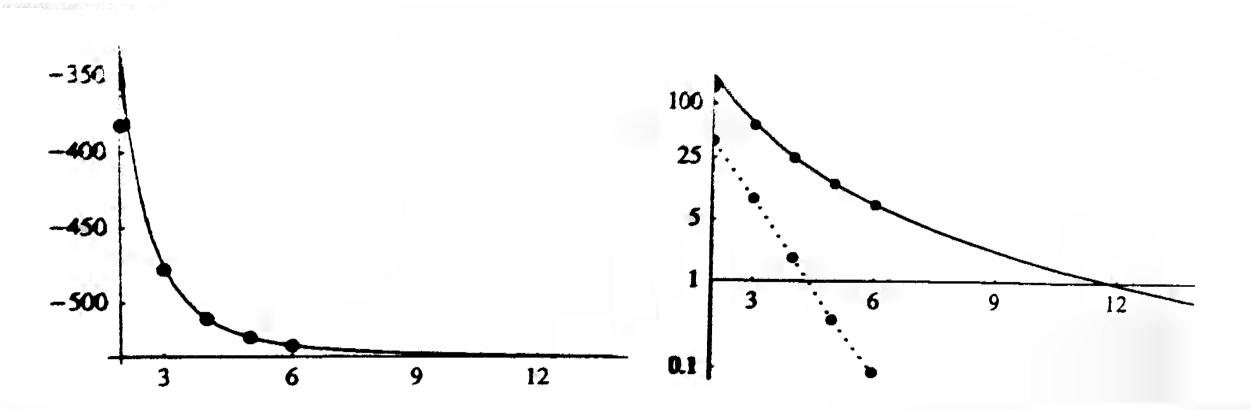


Figure 1.1: The convergence of the HF and MP2 energies for the N_2 molecule with cardinal number of basis set are presented herein, reproduced from reference[1]. The correlation energy is plotted on the left in mE_h. The errors (in mE_h) for the MP2 (solid line) and HF (dashed line) energies are presented on the right versus cardinal number.

calculations, the total energy converges roughly as $A + Be^{-cX}$ to the SCF/CBS estimate, A, with fitted parameters B and c[32–36]. The exponential convergence with cardinal number means that in practice this is normally well-converged by most triple-zeta basis sets. Correlation calculations, on the other hand, converge with the third power of cardinal number. This comparatively slow convergence means that all practical calculations will contain some amount of basis set incompleteness. Using the convergence of correlation calculation with cardinal number, extrapolation procedures can be performed[32].

$$E_{XY}^{\text{corr}} = \frac{E_X^{\text{corr}} X^3 - E_Y^{\text{corr}} Y^3}{X^3 - Y^3} \tag{1.31}$$

Given the difficulty one has in attaining the so-called complete basis set (CBS) limit, it is fortunate that the majority of chemical questions rely upon relative energies rather than absolute energies since the use of relative energies allows for significant error cancellation. Unfortunately, even relative energies are slightly (but fundamentally) inconsistent when atoms are not held fixed since the basis is tied to the atomic locations, and the problem remains of treating both sides of an equation with comparable levels of theory and basis set choice. Fictitious energy lowering, commonly called basis set superposition error (BSSE), occurs for molecules and noncovalent complexes when basis functions from neighboring fragments or atoms are used for local properties, as commonly occurs for binding energies, herein denoted with origin of the basis functions in parenthesis.

$$E_{\text{Bind}} = E_{\text{AB}}(AB) - E_{\text{A}}(A) - -E_{\text{B}}(B) \tag{1.32}$$

This phenomenon results in artificial energy-lowering relative to the atomistic or uncomplexed system. This problem is particularly pronounced when one is far from the CBS limit. One common method for partially addressing the problem is the use of the full basis set for the solution of a subsystem, which is referred to as counterpoise-correction[37]. This tends to underestimate nonbonded interactions, yet the corresponding overestimation can be catastrophic or dangerously misleading[38]. The counterpoise-corrected binding energy is shown in equation 1.33.

$$E_{\text{CP-Bind}} = E_{\text{AB}}(AB) - E_{\text{A}}(AB) - E_{\text{B}}(AB) \tag{1.33}$$

1.3 Density Functional Theory

Density functional theory (DFT) represents a recasting of the problem: instead of solving for the wavefunction, we seek the exact density and the energy as a functional of the density. The basic framework of this theory comes from the Hohenberg-Kohn theorems, which describe the correspondence between the electron density and its functional.

Hohenberg-Kohn Theorem 1. *The ground state electron density maps to a unique potential.*

$$E[n(r)] = F_{HK} + \int n(r) v_{ext} dr^3 \tag{1.34}$$

Hohenberg-Kohn Theorem 2. *Minimizing the energy yielded by a density functional produces the ground state density.*

The problem of generating a solution to the Schrödinger equation remains despite the Hohenberg-Kohn theorems. The Kohn-Sham (KS) approach addresses this through the same formalism as SCF[39] where exchange-correlation density functionals replace the Hartree-Fock exchange kernel. These functionals typically depend upon local properties of the density, either its value[40] or derivatives such as the gradient[41–44] or higher. Unfortunately, electrons within KS-DFT spuriously interact with themselves[45,46], and common KS-DFT approximations can also fail to accurately describe charge-transfer[47] as well as dispersion and other long-range interactions[48] due to the inherent locality of the DFT approximations used.

Despite the possibility for *a priori* exact functionals, parametrized DFT approximations have been necessary for chemical accuracy. Even more commonly, the fractional inclusion of SCF or correlated wavefunction-based ans atze such as MP2 has resulted in hybrid DFT methods[49–51] or double hybrid DFT methods[52,53], where Kohn-Sham orbitals are used for wavefunction correlation calculations, typically MP2.

1.3.1 Dispersion corrected DFT

Most density functionals cannot describe the attractive dispersion forces resulting from long-range electron correlation since these are inherently long-range effects and DFT approximations focus on short-range properties of the electronic density. These dispersion forces result from the interaction

of instantaneous multipoles. For closed shell subunits, this attraction starts with the induced dipole response to instantaneous charge fluctuations, which decrease in magnitude with the sixth power of the distance between the subunits with a coefficient (C_6) depending on the particular system in mind.

$$E_{\text{dispersion}} = -\frac{C_6}{\mathrm{R}^6} \tag{1.35}$$

The first description of these types of forces cast the dispersion energy in terms of ionization potentials and polarizabilities of separated systems[54]. The London formula, below, reproduces C_6 coefficients rather poorly but illustrates the conceptual dependence well.

$$E^{AB}_{\text{dispersion}} = -\frac{3}{2}\left(\frac{I_A I_B}{I_A + I_B}\right)\frac{\alpha^A \alpha^B}{\mathrm{R}^6} \tag{1.36}$$

Rigorously, C_6 coefficients come from frequency dependent polarizabilities[55] which are nontrivial to compute exactly.

$$C_6^{AB} = \frac{3}{\pi}\int_0^\infty \alpha_A(i\omega)\alpha_B(i\omega)d\omega \tag{1.37}$$

Within DFT approximations, the problem of generating these C_6 coefficients is commonly relegated to tables of experimentally or theoretically derived C_6 values[56–58] or to methods which tabulate atom-in-molecule properties[59–73] based upon Hirshfeld partitioning of the density[74] and the polarizability-volume connection ($V = \int \mathrm{r}^3 \rho(\mathrm{r}) d\mathrm{r} = \kappa\alpha$). Once computed, the dispersion energy is expressed through a simple sum over all pairs of atoms.

$$E_{\text{dispersion}} = -\sum_{A<B}\frac{C_6^{AB}}{\mathrm{R}_{AB}^6} \tag{1.38}$$

While this correction dramatically improves treatment of long-range interactions for density functionals, the reliance upon pairwise atomic contributions, which do not explicitly account for local electronic structure, proves difficult occasionally. Another approach for this problem is the design of non-local density functionals, such as VV10[75–79], which provide estimates of the interaction between two densities using an approximate non-local correlation kernel.

$$E^{\text{non-local}}_{\text{correlation}} = \frac{\hbar}{2}\int\int d\mathrm{r} d\mathrm{r}' n(\mathrm{r})\phi(\mathrm{r},\mathrm{r}')n(\mathrm{r}') \tag{1.39}$$

1.3.2 Range-separated hybrids

Accurate treatment of long-range charge-transfer excited states within DFT requires exact exchange[80], yet most hybrid functionals (those that include HF exchange) contain around 20% exact exchange, as is the case for B3LYP[49]. This fractional inclusion of HF results in a large manifold of fictitious charge-transfer excited states for time-dependent (TD) DFT calculations[81–83]. Range-separation within DFT[84–87] is used to partially remedy the charge-transfer problem and self-interaction error. In range-separated methods, the Coulomb operator is partitioned into short

and long-range operators using a distance-dependent function, as done by Gill et al.[88–90] and Savin et al.[91–94]. This function is commonly taken to be the error function, though other choices are possible.

$$\frac{1}{\mathrm{r}} = \frac{\mathrm{erfc}(\omega \mathrm{r})}{\mathrm{r}} + \frac{\mathrm{erf}(\omega \mathrm{r})}{\mathrm{r}}$$

Range-separated hybrid functionals can then be constructed from short-range DFT exchange, short-range HF exchange, and long-range HF exchange, with control over the amount of short-range exact exchange, c_{HF}, and the range-separation parameter, ω.

$$E_{\mathrm{XC}} = E_{\mathrm{C}}^{\mathrm{DFT}} + E_{\mathrm{X}}^{\mathrm{SR\text{-}DFT}} + c_{\mathrm{HF}} E_{\mathrm{X}}^{\mathrm{SR\text{-}HF}} + E_{\mathrm{X}}^{\mathrm{HF}}$$

Range-separated hybrids[52,84–87,95–102] significantly improve treatment of charge-transfer compounds and are capable of performing very well even for general chemical problems.

1.4 Extending the reach of correlation methods

1.4.1 The resolution of the identity or density-fitting approximation

The simplest (and most computationally tractable) *ab initio* description of correlation is MP2, whose scaling is determined by the transformation of atomic orbitals into the molecular orbital basis, a fifth-order process.

$$(ia|jb) = \sum_{\mu}\sum_{\nu}\sum_{\lambda}\sum_{\sigma}(\mu\nu|\lambda\sigma)C_{\mu i}C_{\nu a}C_{\lambda j}C_{\sigma b} \tag{1.40}$$

The two-electron integrals, $(\mu\nu|\lambda\sigma)$, are four-centered quantities. An auxiliary basis, $\{\phi_X\}$, can represent the space spanned by the product of two functions $(\phi_\lambda(\mathrm{R}_1)\phi_\sigma(\mathrm{R}_2))$ in a more compact manner than the full two-function basis, resulting in a different expression for forming two-electron integrals with a resolution of the identity (RI) approximation.

$$(ia|jb) = \sum_{P}\sum_{Q}(ia|P)(P|Q)^{-1}(Q|jb) = \sum_{P}\sum_{Q}\sum_{R}(ia|P)(P|Q)^{-1/2}(Q|R)^{-1/2}(R|jb) \tag{1.41}$$

Defining $B_{ia}^{Q} = \sum_{P}(ia|P)(P|Q)^{-1/2}$, we find

$$(ia|jb) = \sum_{Q} B_{ia}^{Q} B_{jb}^{Q} \tag{1.42}$$

This recasting of the equations does not ultimately solve the fifth-order cost of the two-electron MO integrals, but it does provide a reduction to O^2V^2X where O, V, and X are the number of occupied $(i,j,\ldots)$, virtual $(a,b,\ldots)$, and auxiliary functions $(P,Q,\ldots)$ employed. In practice, substantially large systems (>1500 basis functions) are required before RI-MP2 exceeds the fourth-order cost of the underlying HF calculation, and RI-MP2 calculations are now routine with minimal underlying error through careful choice (or construction) of appropriate auxiliary basis sets[103,104].

1.4.2 Spin-component analyses

Since the Hartree-Fock method incorporates the exchange of electrons, which is associated with fermions, within its wavefunction, same-spin electrons are said to be Fermi correlated. The largest correction to the Hartree-Fock method, then, is the introduction of explicit Coulomb correlation, which has its largest effect upon the opposite-spin electrons. Since MP2 provides the leading order improvement for correlation effects, the opposite-spin portion of the MP2 energy should be, and is, significantly larger than the same-spin MP2 correlation energy. The opposite-spin MP2 energy (OS-MP2) is presented below.

$$E_{\text{OS-MP2}} = -\sum_{ia}^{\alpha}\sum_{jb}^{\beta}\frac{(ia|jb)^2}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j} \tag{1.43}$$

The same-spin MP2 energy (SS-MP2) is tabulated through a similar expression.

$$E_{\text{SS-MP2}} = -\frac{1}{2}\sum_{ia}^{\alpha}\sum_{jb}^{\alpha}\frac{(ia|jb)\left[(ia|jb)-(ib|ja)\right]}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j} - \frac{1}{2}\sum_{ia}^{\beta}\sum_{jb}^{\beta}\frac{(ia|jb)\left[(ia|jb)-(ib|ja)\right]}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j} \tag{1.44}$$

Since nontrivial improvement is achieved in scaling the total correlation energy for methods[105], one possible approach for improving the MP2 correlation energy is to semi-empirically scale the resulting energies to form a spin-component scaled MP2 (SCS-MP2)[106–115],

$$E_{\text{SCS-MP2}} = c_{\text{OS}}E_{\text{OS-MP2}} + c_{\text{SS}}E_{\text{SS-MP2}} \tag{1.45}$$

In fact, spin-component scaled MP2 can be parametrized for different quantities of interest, including intermolecular interactions[116,117], and the spin-component scaled approach can be applied to higher order methods[118,119].

Notably for OS-MP2, the fifth-order computation inherent in MP2 can be avoided through the use of an auxiliary basis, where the two-electron integrals are decomposed in terms of auxiliary basis functions $(P,Q,\dots)$ spanning the necessary space[120]. Furthermore, using a Laplace transform, the OS-MP2 energy expression can be recast to eliminate the denominator.

$$E_{\text{OS-MP2}} = \sum_{ia\,jb\alpha} w_\alpha e^{-\delta_{ia}t_\alpha}e^{-\delta_{jb}t_\alpha}(ia|jb)^2 \tag{1.46}$$

$$E_{\text{OS-MP2}} = \sum_{P,\alpha} w_\alpha \left[\sum_{ia}(B^P_{ia})^T e^{-\delta_{ia}t_\alpha}B^P_{ia}\right]\left[\sum_{jb}(B^P_{jb})^T e^{-\delta_{jb}t_\alpha}B^P_{jb}\right] \tag{1.47}$$

This formula captures the opposite-spin MP2 energy exactly, subject to RI fitting and Laplace quadrature errors, and the missing same-spin energy can be approximated simply through scaling the resultant energy expression, typically by a factor of about 1.3 to generate the scaled, opposite-spin MP2 method (SOS-MP2)[120–123].

Since the difference in treatment between same- and opposite-spin correlation occurs primarily where the electron-electron distance is small, same-spin and opposite-spin correlation energies

approach each other as distances between electrons increase, as in nonbonded interactions. This convergence suggests that the optimal scaling parameter should not be distance-independent for SOS-MP2 and in fact that correlations between electrons at larger distances should be enhanced. One method of implementing this behavior is MOS-MP2, which modifies the Coulomb operator to smoothly increase with interelectronic distance[124].

$$g_\omega(\mathrm{r}) = \frac{1}{\mathrm{r}} + c_{\mathrm{MOS}} \frac{\mathrm{erf}(\omega \mathrm{r})}{\mathrm{r}} \tag{1.48}$$

The introduction of distance dependence, here a form of approximating the missing long-range interaction energy from the same-spin correlation energy, provides a tractable way for addressing noncovalent interactions with a fourth-order method.

1.4.3 Adjusting the treatment of long-range interactions

Correlated calculations capture long-range interactions through their descriptions of the frequency-dependent polarizability. MP2 qualitatively captures dispersion interactions, but it does so at an insufficient quality of theory for quantitative accuracy[125]. The MP2 interaction energy for two isolated closed shell fragments depends on fragment-local molecular orbitals.

$$E^{AB} = -4 \sum_{ia}^{\mathrm{A}} \sum_{jb}^{\mathrm{B}} \frac{|(ia|jb)|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{1.49}$$

The resulting C_6 from this interaction can be decomposed into frequency-dependent polarizabilities which depend only on the orbitals and eigenvalues of a single fragment, which are termed uncoupled.

$$C_6^{AB} = \frac{3}{\pi} \int_0^\infty \alpha_A(i\omega) \alpha_B(i\omega) d\omega \tag{1.50}$$

$$\alpha(i\omega) = 4 \sum_{ia} \frac{\varepsilon_i^a \langle i|\mathbf{r}|a\rangle^2}{(\varepsilon_i^a)^2 - (i\omega)^2} \tag{1.51}$$

The polarizability of a single fragment is not sufficient to adequately describe dispersion interactions[126]. There now exist a number of methods for improving the description of dispersion within MP2, the most direct method being that of MP2+ΔvdW[127], which constructs a C_6-level correction for MP2 from the vdW(TS) method[73] with approximate MP2 C_6s.

$$E_{\mathrm{MP2+\Delta vdW}} = E_{\mathrm{MP2}} - \sum_{AB} \frac{\Delta C_6^{AB}}{\mathrm{R}_{AB}^6} \tag{1.52}$$

An alternative approach is to correct the MP2 correlation energy using coupled response functions from time-dependent DFT. The resulting method is termed MP2C for corrected MP2[128,129]. The uncoupled HF response functions are used to calculate the intermolecular dispersion energy using well-defined fragments.

$$\chi_0(\mathrm{R}_1, \mathrm{R}_2, \omega) = 4 \sum_{ia} \frac{\varepsilon_i^a}{(\varepsilon_i^a)^2 + (\omega)^2} \phi_{ia}(\mathrm{R}_1) \phi_{ia}(\mathrm{R}_2) \tag{1.53}$$

$$E_{\text{disp}}^{AB(2)}(\text{UCHF}) = -\frac{1}{2\pi}\int_0^\infty d\omega \int d\text{R}_1 d\text{R}_2 d\text{R}_3 d\text{R}_4 \chi_0^A(\text{R}_1,\text{R}_3,\omega)\chi_0^B(\text{R}_2,\text{R}_4,\omega)\frac{1}{\text{R}_{12}}\frac{1}{\text{R}_{34}} \quad (1.54)$$

The corresponding coupled response functions are tabulated using the interelectronic interaction within a given approximation and the iterative Dyson equation.

$$W(\text{R}_1,\text{R}_2,\omega) = \frac{1}{\text{R}_1 2} + f_{xc}(\text{R}_1,\text{R}_2,\omega) \quad (1.55)$$

$$\chi_{\text{coupled}}(\text{R}_1,\text{R}_2,\omega) = \chi_0(\text{R}_1,\text{R}_2,\omega) + \int d\text{R}_3 d\text{R}_4 \chi_0(\text{R}_1,\text{R}_3,\omega)W(\text{R}_3,\text{R}_4,\omega)\chi_{\text{coupled}}(\text{R}_4,\text{R}_2,\omega) \quad (1.56)$$

These approaches have yielded dramatic improvements for intermolecular interactions[130]. Unfortunately, these methods require the full MP2 correlation energy as a starting point, and computing the long-range behavior of MP2 unsatisfactorily retains the high scaling of MP2 while eliminating all the terms that drive this scaling. Ultimately, these approaches do not exploit their full potential, and this work is a step towards new methodologies for improving the cost and accuracy of the calculation of long-range interactions.

1.5 Aims of this work

This work primarily concerns the locality of the explicit electron-electron interaction. It is not necessary or even desirable to have methods to handle long-range interactions with high cost when the accuracy is insufficient quantitatively. As such, this work explores methods of range-separation for correlation methods, using short-range correlation methods to approximately capture correlation effects and relying upon cancellation of error or explicit calculations for long-range effects. The chemical targets for these calculations are binding energies and relative energetics for equilibrium and nonequilibrium geometries for weak potential energy surfaces. The simplest biological systems rely upon the additive effect of long-range interactions for secondary structure, integrity, and functionality. Tractable, accurate methods are essential for the future of chemical inquiry into these classes of systems.

In Chapter 2, attenuated MP2 in the aug-cc-pVDZ basis is formulated and parametrized for noncovalent interactions and found to outperform complete basis set estimates of MP2 for many system types. Chapter 3 extends this ansatz to the aug-cc-pVTZ basis and finds increasing gains and more transferable performance across a wide variety of inter- and intramolecular interactions. The treatment of large systems and efficient parallelization of the RI-MP2 energy is addressed in Chapter 4, with a shared memory parallel algorithm developed and applied to system of 1000-2000 basis functions, pushing the limit of conventional RI-MP2 calculations. Along with severe examples of the failure of MP2 for large systems, attenuated MP2 in the aug-cc-pVDZ and aug-cc-pVTZ basis sets is found to transferably improve upon MP2.

I address the lack of transferability of spin-component scaled methods in Chapter 5, developing SCS-MP2(2terfc, aTZ), which provides a single set of parameters for both thermochemistry and noncovalent interactions, matching the best performance from SCS-MP2 and attenuated MP2.

Finally, estimates of the complete basis set limit of attenuated MP2 are examined in Chapter 6. I examine a series of progressively improved basis sets and show the convergence of r_0 with number of diffuse functions and overall cardinal number. The favorable error cancellation of the aug-cc-pVTZ basis set appears to have a well-tuned price/performance ratio.

Chapter 2

Attenuating Away The Errors in Inter- and Intra-Molecular Interactions from Second Order Møller-Plesset Calculations in the Small aug-cc-pVDZ Basis Set

Second order Møller-Plesset perturbation theory (MP2) is perhaps the simplest and most cost-effective wave function approach for adding dynamical correlation effects to the mean field or Hartree-Fock approximation (HF). Although density functional theory (DFT) often provides greater accuracy in bond energies and reaction barriers for less computational effort[131], MP2 is often superior for intermolecular interactions[132]. Present-day density functionals also suffer from incomplete physical descriptions leading to self-interaction errors[45,46] (that are absent in MP2) and cannot be systematically improved towards the exact density functional. By contrast, wave function theory provides a systematically improvable formal framework for electronic energies, but approaching the correct nonrelativistic limit is typically computationally prohibitive for large molecules.

For small molecules, MP2 can be corrected by use of e.g. high order coupled cluster theory, coupled with large basis sets[133–138]. Such methods are of benchmark quality, but are not generally applicable to large molecules, although this challenge is being addressed by on-going developments in explicitly correlated and local correlation methods[139,140]. Nonetheless, to be feasible for large molecules, improvements in MP2 theory must often be more heuristic in nature. An example of compensating for basis set deficiencies is to scale the correlation energy[105,141] to improve atomization energies and barrier heights. The accuracy of this approach was later greatly improved by the development of spin-component scaled (SCS)-MP2[106]. The cost of MP2 could be significantly reduced with little effect on accuracy by the scaled opposite-spin (SOS)-MP2 method[120,121]. In fact, the exploration of (SOS)-MP2 led to a 4th-order algorithm for the full MP2 energy[142]. The very strong recent interest in development of double hybrid density functionals, such as B2PLYP[143], XYG3[53], and ωB97X-2[52] represents efforts to improve the accuracy of MP2 (and DFT) by combining them together.

The focus of this paper is improving the accuracy of MP2 calculations of intermolecular inter-

actions and conformational energies in finite basis sets. This has been attempted with some success via modified SCS-MP2 parameters[116,144]. Indeed, the performance of MP2 for some types of interactions such as hydrogen bond energies is excellent, in large basis sets. However, other intermolecular interactions such as those associated with π stacking[145,146] are poorly described by MP2, even in large basis sets. Fundamentally, this is a result of MP2 long-range interactions using the erratic C_6 coefficients of uncoupled HF (UCHF) theory[125]. To address this problem, two promising approaches have recently been suggested, based on long-range corrections to MP2 theory using better C_6 coefficients. Tkatchenko et al.[147] produced a rather promising MP2+ΔvdW method that determined MP2 dispersion coefficients and replaced them, atom-wise, with improved coefficients[127]. Similarly, the MP2C method[128,129] replaces the system-wide MP2 dispersion energy with that of TD-DFT. These methods demonstrate dramatic improvement over MP2 for treating dispersion interactions, but do still rely upon possessing the full MP2 energy. This rate-determining part of the calculation is then discarded for an improved estimate of the long-range interaction energies.

The other significant issue associated with MP2 calculations is the difficulty of converging them towards the complete basis set limit. In conventional atomic orbital (AO) basis set calculations based upon the principal expansion[20], one generally obtains errors that in the most favorable case go as $O(N^{-1})$ in the number of AO's, N. At the same time, the cost of an MP2 calculation rises as the 4th power of the number of basis functions. Thus a 10-fold reduction in error requires roughly a 10,000-fold increase in computational cost. Of course such estimates are too pessimistic in practice because density-fitting approximations[148] and explicitly correlated methods[149] partially address cost and convergence with increasing basis set size. Nonetheless it is widely demonstrated that very large basis sets, and corrections for basis set superposition errors (BSSE) are required[150,151]. The BSSE corrections[37], whilst desirable for improving the accuracy of calculated intermolecular interactions in a given basis, are undesirable because they cannot be applied to the same type of interactions (stacking, H-bonds, etc.) when they occur within a given molecule.

The approach we shall employ to improve the accuracy of MP2 calculations in finite basis sets is to range-separate the correlation energy. We shall exploit a division of the Coulomb operator into short- and long-range portions, as pioneered by Gill et al.[88–90] and Savin et al.[91–94]. Range separation is most commonly accomplished using the error function and its complement in the form $\frac{1}{r} = \frac{\mathrm{erfc}(\omega r)}{r} + \frac{\mathrm{erf}(\omega r)}{r}$. It has attracted most attention for treating exchange within density functional theory[84–87], where the long-range (non-local part) is evaluated by wave function and the short-range (more local) part is treated as a density functional. The resulting range-separated functionals[52,95–102] reduce self-interaction errors, improve treatment of intermolecular interactions, and have become widely used.

Range-separation has been applied to electron correlation, for example to partition between static (long-range) and dynamic (short-range) correlation[152]. It has also been used to modify long-range opposite-spin MP2 contributions in the MOS-MP2 approach[124]. While most divisions of the Coulomb operator make use of the error function, work by Dutoi and Head-Gordon pursued a new separation using the terf function, $terf(\omega, r_0, r) = \frac{1}{2}[erf(\omega r + \omega r_0) + erf(\omega r - \omega r_0)]$, and its complement, terfc[153]. This function permits the introduction of a distance cutoff into the two-electron integrals, or the preservation of the short-range form of the operator. Thus the terfc-

attenuated Coulomb operator has the same derivative as the Coulomb operator in the short-range if the constraint, $r_0\omega = \frac{1}{\sqrt{2}}$, is applied. Additionally, the terfc-attenuated short-range portion of the MP2 correlation energy converges more rapidly to the unattenuated MP2 correlation energy as $\omega \to 0$ than the equivalent erfc-based short-range MP2 energy for the neon atom.

Since long-range contributions drive the overall computational cost of MP2 and also limit its accuracy, this paper pursues the development of a short-range MP2, targeted specifically at evaluation of inter- and intra-molecular interactions in the small augmented cc-pVDZ basis[154]. Perhaps surprisingly, we show below that the combination of unattenuated Hartree-Fock and short-range MP2 stemming from separation of the Coulomb operator improves upon unmodified MP2. In general, improvements to MP2 theory should combine an attenuated treatment of the short-range with a long-range correction, based for example on improved C_6 coefficients[56–58,127,155]. However, the relatively inadequate AO basis that we explore here will mean that in fact the results cannot be substantially improved by the addition of a long-range correction. The role of attenuation will be to remove part of the over-binding associated with BSSE in small basis sets, as well as part of the over-binding associated with MP2 itself for some types of dispersion interactions.

We shall denote a short-range MP2 method that employs erfc attenuation (in only the correlation part) in the aug-cc-pVDZ basis as MP2(erfc, aDZ). The corresponding terfc attenuated method will be denoted as MP2(terfc, aDZ). This work focuses on four short-range variants: MP2(terfc, aDZ) (**I**), scaled MP2(terfc, aDZ) (**II**), MP2(erfc, aDZ) (**III**), and scaled MP2(erfc, aDZ) (**IV**). The scaling is applied solely to the correlation energy, $E_{\text{full}} = E_{\text{HF}} + s * E_{\text{corr.}}$, akin to previous work[105,141]. The introduction of a scaling parameter allows for the possibility of correcting for systematic errors in the correlation energy due to severe truncation in the strong attenuation limit and BSSE in the weak attenuation limit. All calculations were performed within a development version of Q-Chem 4.0[156].

Parameterization of attenuated short-range MP2 requires a well-balanced set of representative molecules with established CCSD(T)/CBS energies. As we are attempting to remedy unphysical long-range behavior of MP2, the S66 database[157], consisting of hydrogen-bonding, dispersion, and mixed dimer interactions, was chosen as the training set. This training set contains a range of binding energies and system sizes. No subset-specific weighting factors were used in order to promote transferability rather than the biased treatment of any specific interaction type. The terfc-attenuated variants use the curvature constraint of $r_0\omega = \frac{1}{\sqrt{2}}$, which justifiably reduces the number of fitted parameters and preserves short-range quality. No counterpoise corrections are performed.

Figures 2.1 and 2.2 show the behavior of MP2(terfc, aDZ) and MP2(erfc, aDZ) for the S66 database. For comparison to scaled variants **II** and **IV**, scaled MP2/aDZ (SMP2) without attenuation is also optimized for this dataset. There are two limits of interest. First, the severe attenuation limit of $r_0 \to 0$ (terfc attenuation) and $\omega \to \infty$ (erfc attenuation), coincides with the HF/aDZ RMSD of 4.0 kcal mol^{-1} if no scaling is applied. This can be strikingly reduced by scaling, though the large deviation of the optimal scaling factors from unity is compensating for over-attenuation. The second limit of interest is MP2(terfc, aDZ) as $r_0 \to \infty$ and MP2(erfc, aDZ) as $\omega \to 0$. Without scaling, this limit coincides with the unattenuated MP2 result (RMSD of 2.7 kcal/mol).

Simple scaling of the MP2 correlation energy yields a striking reduction of RMS error by a

Table 2.1: Root-mean-squared deviations, standard deviations of error, average, and mean unsigned errors for the S66 Dataset (kcal mol^{-1})

RMSD	MP2/CBS[1]	MP2[2]	SMP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
H-Bonds	0.19	0.82	0.71	0.48	0.50	0.51	0.52	0.32	1.36
Disp.	1.11	3.58	0.46	0.39	0.40	0.42	0.40	1.01	4.24
Mixed	0.55	2.81	0.55	0.49	0.50	0.51	0.50	0.88	3.06
Overall	0.73	2.67	0.59	0.46	0.47	0.48	0.48	0.79	3.12
Error	MP2/CBS[1]	MP2[2]	SMP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
AVG	-0.40	-2.15	0.14	0.05	0.05	0.01	0.05	-0.61	2.62
MUE	0.48	2.15	0.49	0.34	0.35	0.36	0.36	0.64	2.62

[1] From the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

factor of 4.5 with a constant scaling factor of $s = 0.60$. While scaling the correlation energy is not a new idea[105], the very large improvement that can be obtained in intermolecular interactions using this approach for MP2/aDZ does not appear to have been appreciated. Indeed, reports aimed at atomization energies and barrier heights used scaling factors larger than one[141], whilst we find a need to significantly attenuate for non-bonded interactions with $s = 0.60$. SMP2/aDZ surprisingly surpasses MP2/aDZ with counterpoise correction, which yields a RMSD of 0.88 kcal mol^{-1}.

In between the extreme limits, even larger improvements can be obtained by consider optimal values of the attenuator. For variant **I** of MP2(terfc, aDZ), we choose $r_0 = 1.05$ Å. For **II**, $r_0 = 1.00$ Å and $s = 1.06$. For variant **III** of MP2(erfc, aDZ), we select $\omega = 0.420$ Å^{-1}, and for **IV**, $\omega = 0.420$ Å^{-1} and $s = 0.99$. Performance with these parameters is shown in Table A.3. The reduction in error relative to no correlation at all is a factor of 8.5, whilst the reduction relative to MP2/aDZ is a factor of 5.5. These methods even yield better error statistics than MP2/CBS for this S66 dataset despite requiring hundreds of times less computational effort. Furthermore, the fact that distance-dependent attenuation is more physical than simple scaling (SMP2) is consistent with the fact that one parameter attenuation out-performs one parameter scaling. These are remarkable improvements for a single parameter semi-empirical method, even given that this is training set data. None of the presented results include a long-range dispersion correction, which was found to be of minimal value for these short-range MP2 methods at the chosen attenuation parameters.

To establish transferability and thus usability, MP2(terfc, aDZ) and MP2(erfc, aDZ) have been tested against separate datasets. The S22 database[158–161] is of particular significance due to its wide usage. Table A.4 demonstrates that MP2(terfc, aDZ) and MP2(erfc, aDZ) provide significant improvement over MP2/aDZ and again performs better than MP2/CBS. The RMSD for these interaction energies has been reduced from 1.4 kcal mol^{-1} for MP2/CBS to 0.6-0.7 kcal mol^{-1} with the introduction of one parameter (or two in the case of the scaled variants, **II** and **IV**). The significant overestimation of dispersion by MP2/CBS and particularly MP2/aDZ has been reduced such that MP2(terfc, aDZ) and MP2(erfc, aDZ) perform better on these interactions (0.4-0.5 kcal mol^{-1}) than on hydrogen-bonded systems (0.8-1.0 kcal mol^{-1}). Scaling the correlation energy

Figure 2.1: Performance on S66 Dataset for MP2(terfc, aDZ) with both unscaled, **I**, and scaled, **II**, variants over the range $r_0 = 0.05$Å $\rightarrow r_0 = 4.00$Å, which spans from the HF limit (4.0 kcal mol^{-1}) to the unattenuated MP2 limit (2.7 kcal mol^{-1}).

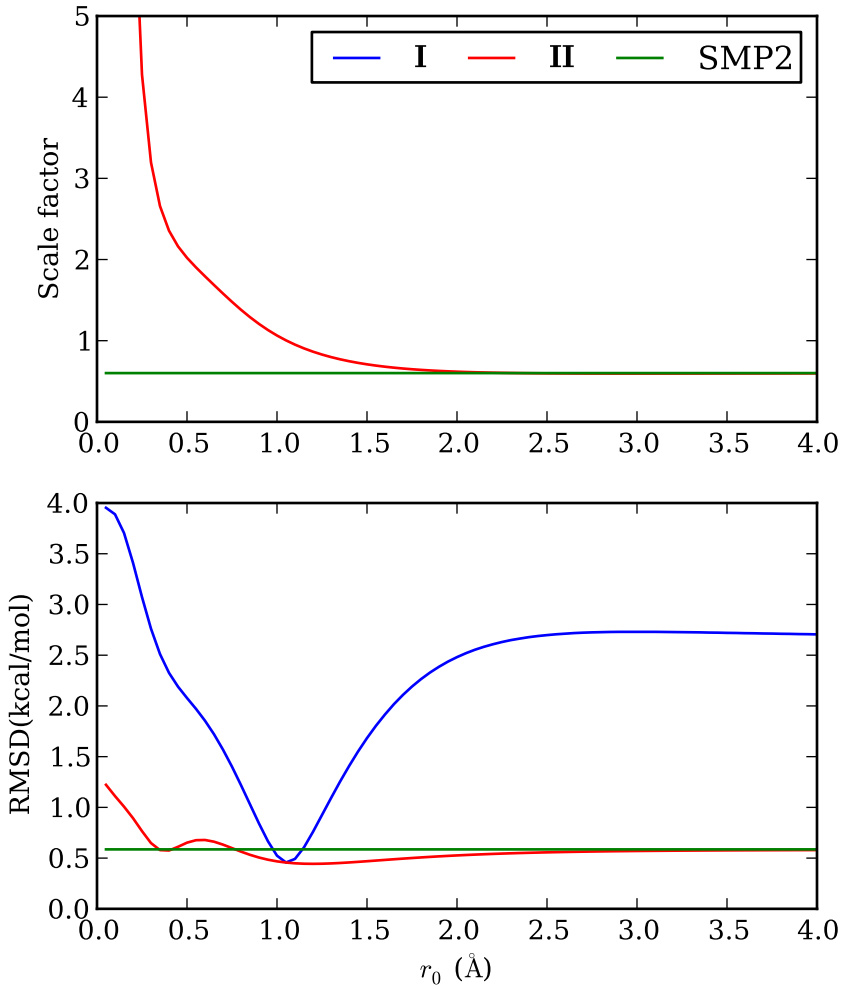


Figure 2.2: Performance on S66 Dataset for MP2(erfc, aDZ) with both unscaled, **III**, and scaled, **IV**, variants over the range $\omega = 0.01\text{Å}^{-1} \rightarrow \omega = 2.00\text{Å}^{-1}$, which spans from the unattenuated MP2 limit (2.7 kcal mol^{-1}) and approaches the HF limit of 4.0 kcal mol^{-1}.

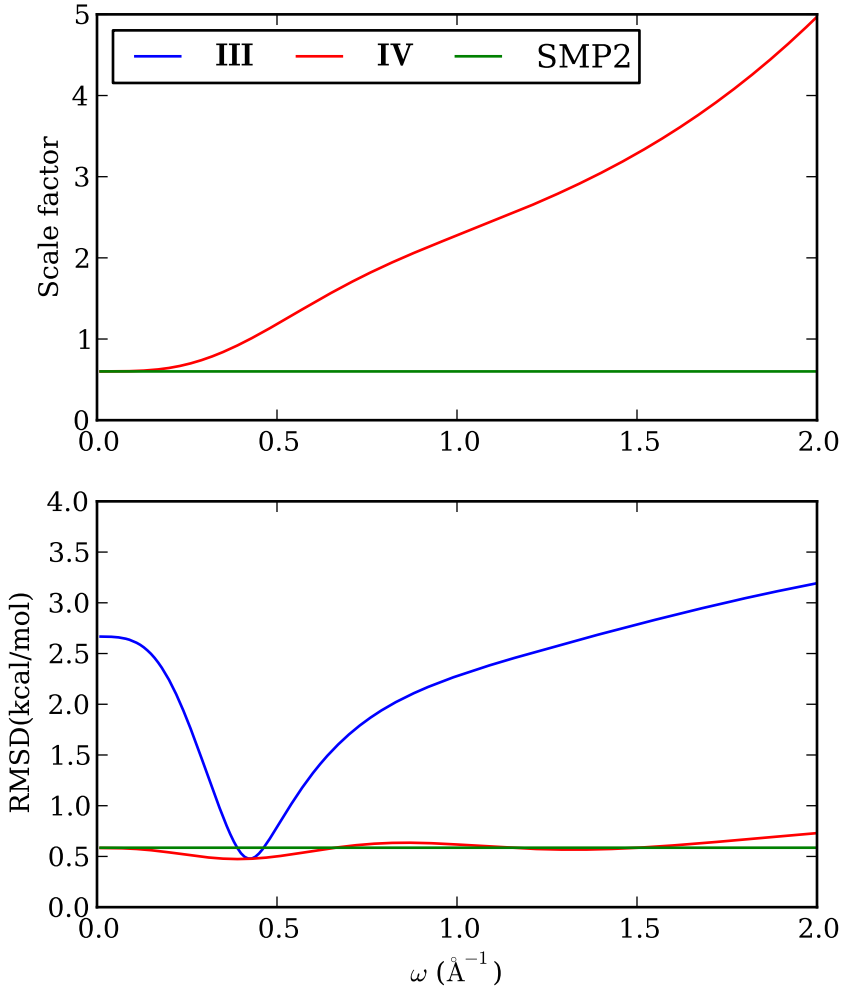


Table 2.2: Root-mean-squared deviations, standard deviations of error, average, and mean unsigned errors for the S22 Dataset (kcal mol^{-1})

RMSD	MP2/CBS[1]	MP2[2]	SMP2 [2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
H-Bonds	0.20	1.02	1.17	0.80	0.80	0.85	0.99	0.42	1.66
Disp.	1.93	4.60	0.68	0.45	0.46	0.53	0.50	0.88	4.58
Mixed	1.41	4.75	0.67	0.52	0.52	0.60	0.55	0.98	5.36
Overall	1.39	3.91	0.86	0.61	0.61	0.67	0.71	0.81	4.24
Error	MP2/CBS[1]	MP2[2]	SMP2 [2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
AVG	-0.84	-2.77	0.01	0.01	0.01	-0.04	0.03	-0.53	3.17
MUE	0.89	2.79	0.70	0.51	0.51	0.56	0.58	0.65	3.17

[1] From the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

(SMP2/aDZ) again reduces overall error by 4.5, but the RMSD is increased for hydrogen-bonding systems relative to the unscaled MP2/aDZ, which suggests the scaling parameter should be varied based upon system type, akin to (SCS)-MP2 and (SCS-MI)-MP2[116].

MP2(terfc, aDZ) and MP2(erfc, aDZ) have been parameterized without counterpoise correction; thus relative conformational energies present another metric for assessing their quality since accounting for intramolecular BSSE is nontrivial[162]. Valdes et al.[163] produced a benchmark energy and geometry database for conformers of five small peptides with aromatic side chains, which we shall refer to as P76 for the 76 conformers. The sensitivity of conformer energy ordering to quality of method across the varied noncovalent interactions makes this a potentially demanding test of the transferability of the short-range MP2 methods. The results summarized in Table 2.3 show that MP2(terfc, aDZ) and MP2(erfc, aDZ) outperform MP2/aDZ by roughly a factor of 3, and also outperform MP2/CBS, measured relative to CCSD(T)/CBS benchmarks. The error statistics also suggest that structural motifs can affect the quality of these descriptions for the GGF (glycine-glycine-phenylalanine) protein, yet MP2(terfc, aDZ) and MP2(erfc, aDZ) still significantly improve upon MP2/aDZ as well as the well-tempered M06-2X method[164]. On these systems, both terfc-attenuated variants slightly outperform the erfc-attenuated variants, particularly for the GFA (glycine-phenylalanine-alanine) protein. Both attenuated MP2 methods significantly outperform simple scaling (SMP2) in this test. Further work is necessary to fully characterize the behavior of these short-range attenuated MP2 methods based on interaction type and distance. Reduced errors are also shown for SMP2/aDZ in all cases, with particular improvement for WG (tryptophan-glycine) and WGG (tryptophan-glycine-glycine) while leaving the other peptides largely unaffected, again suggesting interaction dependence for the universal scaling of the correlation energy.

Another useful benchmark for medium-size systems is the alanine tetrapeptide system. The energetics of different conformers have pushed the limits of systems accessible for wavefunction-based correlation methods and basis set convergence[165,166]. The system of twenty-seven conformers analyzed at RI-MP2/CBS is used as a reference, and we present the deviations for various

Table 2.3: Root-mean-squared deviations for protein subsets of the P76 database (kcal mol^{-1})

Protein	MP2/CBS[1]	MP2[2]	SMP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
WG	0.35	1.15	0.53	0.19	0.22	0.19	0.19	0.48	1.63
WGG	0.59	1.49	0.52	0.38	0.38	0.40	0.40	0.72	2.23
FGG	0.44	0.98	0.81	0.46	0.44	0.48	0.50	0.61	1.71
GGF	0.19	0.57	0.51	0.33	0.34	0.32	0.32	0.49	1.14
GFA	0.41	0.89	0.81	0.25	0.24	0.32	0.32	0.30	1.10
Overall	0.42	1.06	0.65	0.33	0.33	0.35	0.36	0.54	1.61

[1] From the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ

Table 2.4: Mean absolute deviations and root-mean-squared deviations from RI-MP2/CBS on alanine tetrapeptide conformers (kcal mol^{-1})

Error[1]	MP2[2]	SMP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
MAD	0.78	0.16	0.16	0.17	0.15	0.15	0.22	1.21
RMSD	0.97	0.20	0.20	0.21	0.17	0.18	0.27	1.48

[1] These errors are relative to RI-MP2/CBS estimates[166] of these conformers, which deviates from the CCSD(T) answer significantly enough that superlative judgments of method performance cannot be made.

[2] Computed using aug-cc-pVDZ

methods in Table A.14. SMP2/aDZ, MP2(terfc, aDZ), and MP2(erfc, aDZ) present comparable behavior to RI-MP2/CBS (RMSD 0.2 kcal mol^{-1}), as well as almost fourfold smaller deviations than MP2/aDZ. This strongly suggests that attenuation of the MP2 correlation contribution in aug-cc-pVDZ is functioning effectively to remove much of the intramolecular basis set superposition error that traditionally plagues small basis set MP2 calculations of conformational energies.

Full characterization of SMP2/aDZ, MP2(terfc, aDZ), and MP2(erfc, aDZ) must include examination of behavior at equilibrium and nonequilibrium distances. Ongoing work will assess the viability of these methods for geometry optimizations. For non-equilibrium displacements, Figure 2.3 presents four selected dimers from the S22x5 database[167], which has CCSD(T) energies for contraction and extension of the S22 geometries. The behaviors of MP2(terfc, aDZ)(variant **I**), MP2/aDZ, SMP2/aDZ, and CCSD(T)/CBS are shown. Given the equivalent computational costs of MP2/aDZ, SMP2/aDZ, and MP2(terfc, aDZ), the improvement is dramatic for the introduction of only a single parameter, especially for the parallel-displaced and t-shaped benzene dimers.

With the attenuation of the Coulomb operator within MP2, MP2(terfc, aDZ) and MP2(erfc, aDZ) improve upon the description of inter- and intramolecular forces of MP2, even compared to complete basis set limit results. With excellent behavior on dispersion, hydrogen-bonded, and mixed dimer interactions, as well as protein conformations, both short-range MP2 methods perform in a transferable manner. While these methods produce comparable performance, we recommend MP2(terfc, aDZ) since its sharper attenuation parameter of $r_0 = 1.05$ Å ($\omega = 0.673$ Å^{-1}) will

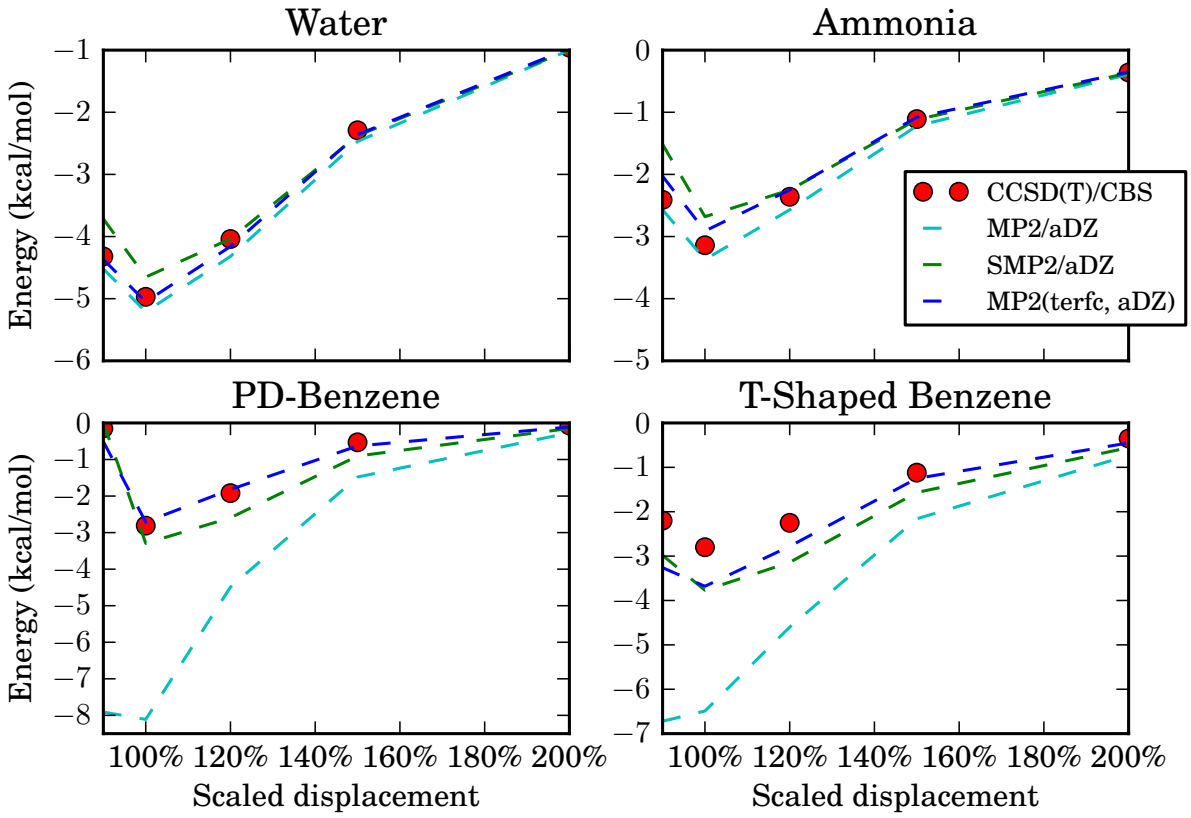


Figure 2.3: Geometries from S22x5 with MP2(terfc, aDZ)(**I**), SMP2/aDZ, and MP2/aDZ. For comparison, CCSD(T)/CBS is provided.

provide a lower prefactor for any optimized algorithm. Since integrals involving the error function are more widely available, MP2(erfc, aDZ) can be readily implemented using $\omega = 0.420$ Å^{-1}. The scaled variants are not necessary at this time, as they introduce another parameter without improving error statistics. However, they do permit shorter range truncation of the correlation contributions, and SMP2/aDZ with $s = 0.60$ provides dramatic improvements for all databases investigated. These parameters are expected to vary per basis set with degree of resulting BSSE. While parameterization could be attempted for reaction energies or electron attachment/detachment, behavior commensurate with or worse than MP2/aDZ is expected.

Relative to MP2/aDZ (and sometimes even relative to MP2/CBS), MP2(terfc, aDZ) and MP2(erfc, aDZ) show reduced deviations from benchmarks for non-bonded interactions from the S66, S22, and P76 datasets, the 27 reference alanine tetrapeptide conformers and the selected S22x5 geometries. This suggests these methods have a well-behaved and transferable compensation for BSSE, and they are thus immediately useful for this purpose. SMP2/aDZ also provides significant error reduction across most systems, which lies in accord with the understanding that MP2/aDZ, from both

BSSE and inherent MP2 exaggeration of dispersion effects, overestimates non-bonded interactions regardless of distance. By contrast, of course, MP2/aDZ underestimates bonded interactions (e.g. atomization energies) due to basis set incompleteness, which explains the very different scaling factors reported previously for bonded interactions (> 1) versus what we find here for non-bonded interactions (< 1).

In the future, MP2(terfc, aDZ) and MP2(erfc, aDZ) offer the potential for far greater computational efficiencies than MP2/aDZ because their chosen parameters attenuate the relevant two-electron integrals for correlation, reducing their spatial extent to a distance of only several bond lengths. With such limited dependence on long-range terms, there is exciting scope for low-scaling implementations of these methods that can remedy both BSSE and long-range inaccuracies within limited basis MP2.

Chapter 3

Attenuated Second-Order Møller-Plesset Perturbation Theory: Performance within the aug-cc-pVTZ Basis

3.1 Introduction

In quantum chemistry based on wave functions[168], two basic challenges must be surmounted to obtain an accurate approximation to the correlation energy, and thereby achieve accurate values of relative energies for intermolecular and intra-molecular non-bonded interactions. First is achieving a sufficiently accurate description of electron correlations to accurately approximate the full configuration interaction limit in a given basis set. Second is converging the basis expansion towards the complete basis set (CBS) limit. In practice, despite great progress, it is only possible to obtain reasonable approximations to these two limits in benchmark systems. For other cases, the computational cost of converging the correlation energy and the basis set is at present simply prohibitive.

Benchmark calculations therefore play a vital role in assessing the likely accuracy of more tractable quantum chemical models for larger systems. For intermolecular interactions, benchmark data has been evaluated for model hydrogen bonded interactions, π stacking interactions, electrostatic interactions, and interactions with mixed character. Examples of databases that contain state of the art benchmarks are the S66 set[157], and the S22 set[158–161], though there are many others. For relative conformational energies, which are largely determined by the interplay of steric effects with intramolecular H-bonding, dispersion, and electrostatic interactions, benchmark data is also available. Examples include databases of alkane conformations[169], sugar conformations[170], and cysteine conformations[171].

With respect to electron correlation, the simplest and computationally cheapest useful wave function method is second-order Møller-Plesset perturbation theory (MP2). Whilst MP2 at the CBS limit is known to be very accurate for some intermolecular interactions, such as hydrogen-bonding[172], it is also well known to yield large percentage errors for π stacking interactions[145,146].

The problem of MP2/CBS is the inaccurate description of long-range dispersion, since MP2 uses inaccurate polarizabilities from time-dependent uncoupled Hartree Fock (UCHF) for long-range interactions[125]. Recent attempts at remedying these inaccuracies have replaced the UCHF-based long-range interactions of MP2 with time-dependent DFT[128,129] or atomistic van der Waals corrections[147]. While such methods have achieved significant success, they rely upon computing the entire MP2 energy only to remove and replace the rate-limiting portion. Furthermore, they cannot be applied to intra-molecular interactions such as the important problem of relative conformational energies[173].

Even without such inherent limitations of MP2, convergence of the MP2 correlation energy to the complete basis set limit (CBS limit) is unattainable in larger chemical systems due to high computational cost[20]. There is reason for optimism about the prospects for MP2 calculations on larger molecules because of local MP2 methods[174]. Likewise, extrapolation methods[175,176] with the correlation consistent cc-pVXZ (abbreviated as XZ) basis sets[154] and explicitly correlated MP2 methods[139,140,149] are helping to more routinely approach the basis set limit. Nevertheless, the quality of relative energies from MP2 calculations in finite basis sets is degraded by basis set superposition error (BSSE) and basis set incompleteness[177]. Counterpoise (CP) correction can partially remedy BSSE[37], but this correction method cannot always be applied consistently to interactions on the same fragment or molecule. Without CP correction, however, the addition of diffuse (augmented) functions as in the aug-cc-pVXZ basis sets[31,154,178–180] (abbreviated as aXZ) which help to describe anions and polarization, also increases BSSE. In fact, for the S66 database of noncovalent interactions[157], MP2/DZ reproduces CCSD(T)/CBS estimates more accurately than MP2/aTZ, despite being roughly 100 times less computationally demanding.

Given the somewhat systematic errors of MP2 at the CBS limit (overbinding dispersion interactions), and the even more systematic behavior of BSSE in finite basis sets (overbinding all intermolecular interactions), it is natural to seek semi-empirical modifications that can remove this systematic error. Existing examples include modifying spin-component scaled MP2 (SCS-MP2)[106] for intermolecular interactions[116], as well as attempting to modify scaled opposite spin MP2 (SOS-MP2[120,124] to treat intermolecular interactions. These methods all work best in large basis sets, with the SCS approach significantly out-performing the SOS approach, as well as MP2 itself[117].

Turning to modifications of MP2 in small basis sets, we recently introduced[181] an advantageous one-parameter semi-empirical MP2 method based upon range-separating the Coulomb operator within the two-electron integrals, and keeping only the short-range portion. From results for inter- and intramolecular interactions using only the short-range portion, we designed the terfc- or erfc-attenuated MP2 within the aug-cc-pVDZ basis (aDZ), termed MP2(attenuator, aDZ). This method provided up to a five-fold improvement on unattenuated MP2/aDZ and yielded lower errors than MP2 at the complete basis set (CBS) limit for the S66 database (which was used for training) as well as for the S22 and P76 databases (which were used for testing).

This remarkable improvement raises a variety of interesting questions. First and foremost, does the improvement using attenuation in the aDZ basis carry over to larger basis sets? In this report we explore the performance of attenuated MP2 using the larger aug-cc-pVTZ (aTZ) basis and discover that it generally outperforms (albeit at greater computational cost) the attenuated aDZ model. We

also provide extensive tests to establish the extent of transferability of this model. Second, what type of error compensation is occurring to yield these improvements? We are able to gain some insight by comparing attenuated MP2 results with and without counterpoise correction in the aDZ and aTZ basis sets, relative to attenuation in the non-augmented DZ and TZ sets.

3.2 Methods

Attenuated MP2 is based on replacing the electron-electron repulsion operator, r_{12}^{-1} with an attenuated operator, $s(r_{12})\,r_{12}^{-1}$ in the evaluation of the correlation energy. The short-range function, $s(r)$, is a monotonically decreasing function which is one at $r = 0$ and tends to zero for large r. Thus $s(r)$ plus its long-range complement, $l(r)$, form a partition of unity, $1 = s(r) + l(r)$. One very suitable function is the sum of two complementary error functions, offset in such a way that the attenuated operator preserves its shape for small r, as shown in Figure 3.1. The long-range function is:

$$l(r) = \mathrm{terf}(r, r_0) = \frac{1}{2}\left\{ erf\left[\frac{(r-r_0)}{r_0\sqrt{2}}\right] + erf\left[\frac{(r+r_0)}{r_0\sqrt{2}}\right]\right\} \tag{3.1}$$

while its short-range complement is:

$$s(r) = \mathrm{terfc}(r, r_0) = 1 - \mathrm{terf}(r, r_0) \tag{3.2}$$

With the choice above, 1st and 2nd derivatives of $l(r)\,r^{-1}$ vanish exactly at $r = 0$, and approximately for small r. Therefore the attenuated Coulomb operator is merely vertically shifted in the small r regime then goes to zero smoothly (along with its derivatives) at large r.

Attenuated MP2, where r^{-1} is replaced by $terfc(r, r_0)r^{-1}$ in the second order correlation evaluation, has been implemented in the Q-Chem program[156]. Calculations within this work use the resolution-of-the-identity and frozen core approximations. Our implementation extends the original code[153] to permit the use of higher angular momentum through h functions, constructing intermediates for the terf-attenuated Coulomb integrals using 256-bit precision with the GNU multiple-precision library[182,183] and storing the resulting two-dimensional interpolation tables in 64-bit double precision on disk ($\sim$ 60 Mb).

3.3 Parameter optimization

As before[181], we chose the S66 database for training our attenuation parameter. This database contains CCSD(T)/CBS benchmarks of energies for equilibrium geometries of noncovalently bound systems. The first set of results, shown in Figure 3.2, correspond to performing the attenuated calculations without counterpoise corrections in cc-pVDZ, cc-pVTZ, aug-cc-pVDZ, and aug-cc-pVTZ basis sets. The results in this figure show that the optimal attenuation parameter, r_0, is inversely related to BSSE in the calculation. With augmented double zeta (aDZ) and triple zeta (aTZ) basis sets, attenuation can yield over 5-fold RMS error reduction. The optimal aTZ attenuation (1.35 Å) yields 40% lower RMS error than the optimal aDZ attenuation (1.05 Å).

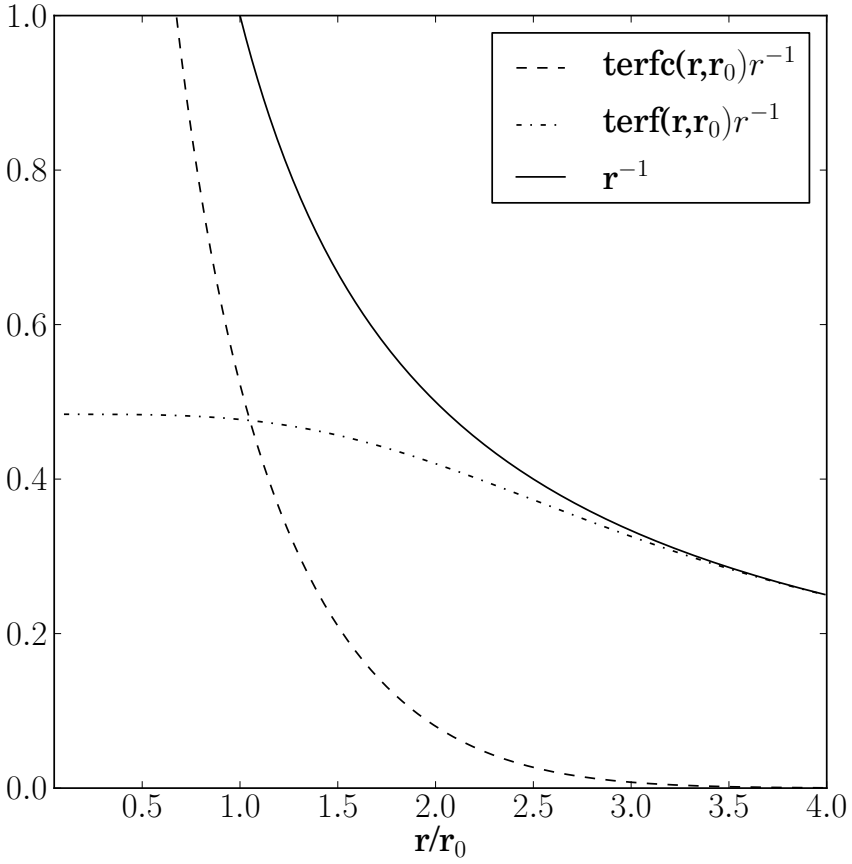


Figure 3.1: The partitioning of the interelectron repulsion operator into short range and long-range components based on the long-range terf function defined in Eq. (4.1) and its short-range complement, terfc, defined in Eq. (4.2). With these definitions, $\mathrm{terf}(r,r_0)r^{-1}$ has zero first and second derivatives in the small r limit. Therefore the short-range interelectron repulsion, $\mathrm{terfc}(r,r_0)r^{-1}$ behaves as a smoothly shifted r^{-1}. The models developed in this paper retain only the short-range term in the MP2 energy, and optimize the single parameter r_0 to reproduce benchmark intermolecular interactions.

Table 3.1: Root-mean-squared deviations(RMSD), average, and mean unsigned errors on the S66 database (kcal mol^{-1})

RMSD	MP2(terfc, aTZ)	MP2(terfc, aTZ-CP)	MP2/aTZ	MP2(terfc, aDZ)	MP2(terfc, aDZ-CP)	MP2/aDZ	MP2/CBS [a]
H-Bonds	0.18	0.62	0.51	0.48	1.22	0.82	0.19
Disp.	0.27	0.45	2.20	0.31	0.53	3.80	1.11
Mixed	0.29	0.20	1.38	0.47	0.36	2.45	0.55
Overall	0.25	0.46	1.53	0.43	0.81	2.66	0.73
AVG	-0.07	0.15	-1.23	0.05	0.38	-2.15	-0.40
MUE	0.21	0.35	1.23	0.32	0.59	2.15	0.48

[a] From the Benchmark Energy and Geometry DataBase[2]

The striking error reductions obtained with augmented basis functions cannot be replicated with the non-augmented basis sets. The attenuated DZ curve shown in Figure 3.2 shows only about 10% error reduction relative to standard MP2/DZ (large r_0). The best attenuated DZ has over 3-fold larger RMS error than the best attenuated aDZ! A larger error reduction from MP2/TZ is possible with attenuated TZ (roughly 40%) but the resulting RMS error is still more than twice that of attenuated aTZ. These comparisons show that augmented functions are essential for large improvements through attenuation. This suggests attenuated MP2 accounts for dispersion primarily through the tuned interplay of attenuation with BSSE.

Results for counterpoise (CP) correction of attenuated MP2 using augmented basis sets are shown in Figure 3.3. Attenuated MP2-CP results show strikingly less improvement than attenuated MP2 without CP correction. For instance, MP2(terfc, aDZ-CP) attains essentially no improvement (no minimum) versus MP2/aDZ-CP ($r_0 \rightarrow \infty$ limit). This suggests attenuation-based error cancellation within the aDZ basis is largely due to incomplete removal of BSSE and that this favorable cancellation disappears with counterpoise correction. Interestingly, in the larger basis, MP2(terfc, aTZ-CP) moderately outperforms MP2/aTZ-CP, suggesting that attenuation is partially removing inaccurate long-range contributions. The much larger optimal MP2(terfc, aTZ-CP) r_0 value of $1.75\mathring{A}$ vs $1.35\mathring{A}$ for MP2(terfc, aTZ) is also consistent with removing only longer range interactions. Emphasizing the importance of partial BSSE cancellation over long-range correction, MP2(terfc, aDZ) and MP2(terfc, aTZ) surpass MP2(terfc, aTZ-CP).

Results for the S66 database using basis set specific optimal r_0 parameters are presented in Table 3.1. The relatively small r_0 values for MP2(terfc, aDZ) (1.05 Å) and MP2(terfc, aTZ) (1.35 Å) cancel large BSSE for all types of interactions, which is leveraged to reduce errors in all categories quite substantially. Particularly notable is the dramatic improvement in RMSD for MP2(terfc, aTZ) over MP2(terfc, aDZ). The increase in computational cost with the larger basis is accompanied by a 41% reduction in error that appears to recover the excellent behavior of MP2 for hydrogen-bonded interactions.

Subsets of the S66 database show significant variations in resultant errors. Since attenuated MP2 converges to the unattenuated MP2 result by $r_0 \sim$4 Å, a better description of a type of interaction by the unattenuated method will lead to a more extended r_0. This extension is reflected in Figure 3.4 most clearly by the performance of MP2(terfc, aTZ-CP) on the hydrogen-bonded subset, which is optimal without attenuation. Exhibiting a different behavior, MP2(terfc, aTZ)

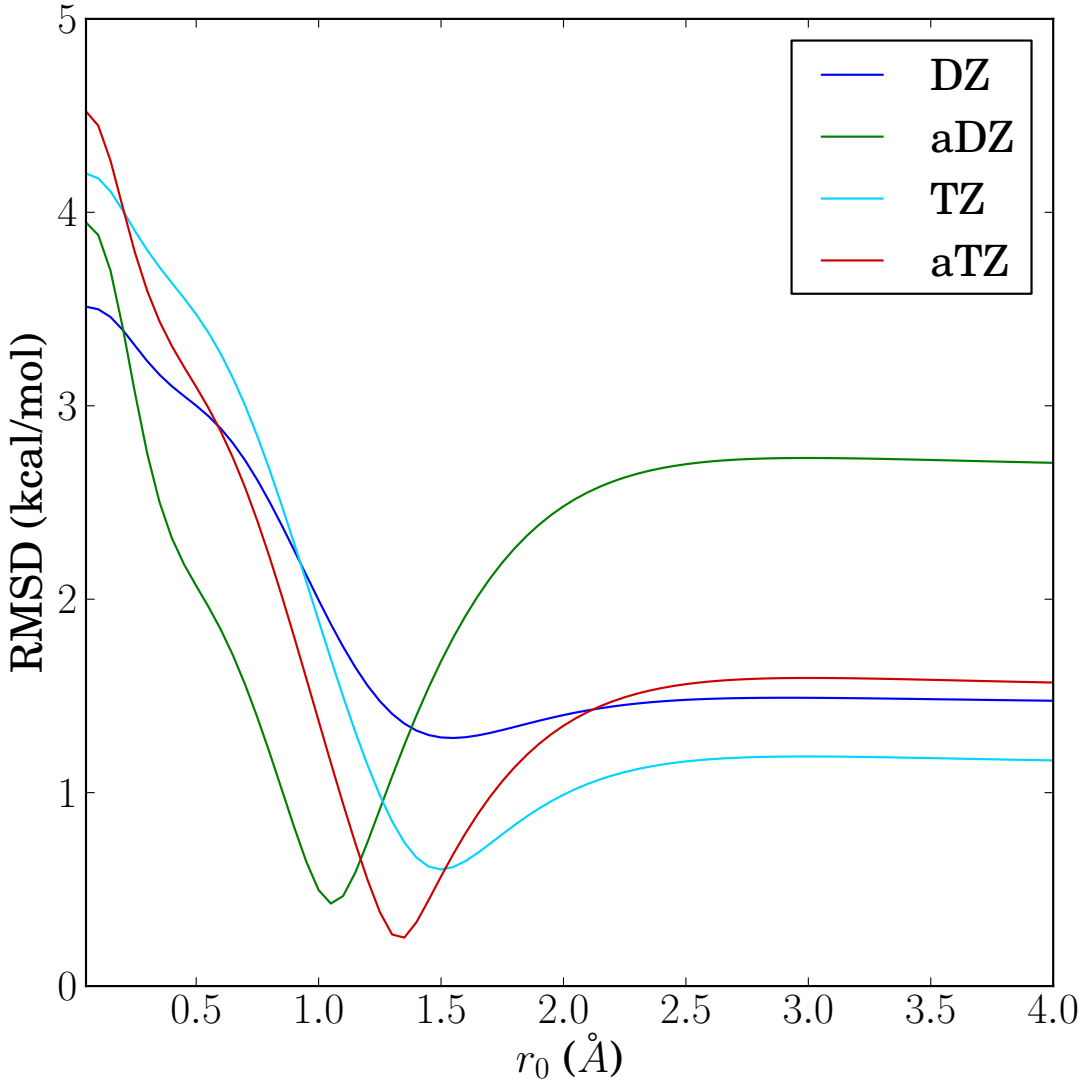


Figure 3.2: Effect of augmented functions on root mean squared deviation of truncated MP2 methods for training set S66 with terfc-attenuation. As $r_0 \to 4.0$Å, attenuated MP2 converges to the unattenuated result. As $r_0 \to 0$Å, attenuated MP2 approaches HF results.

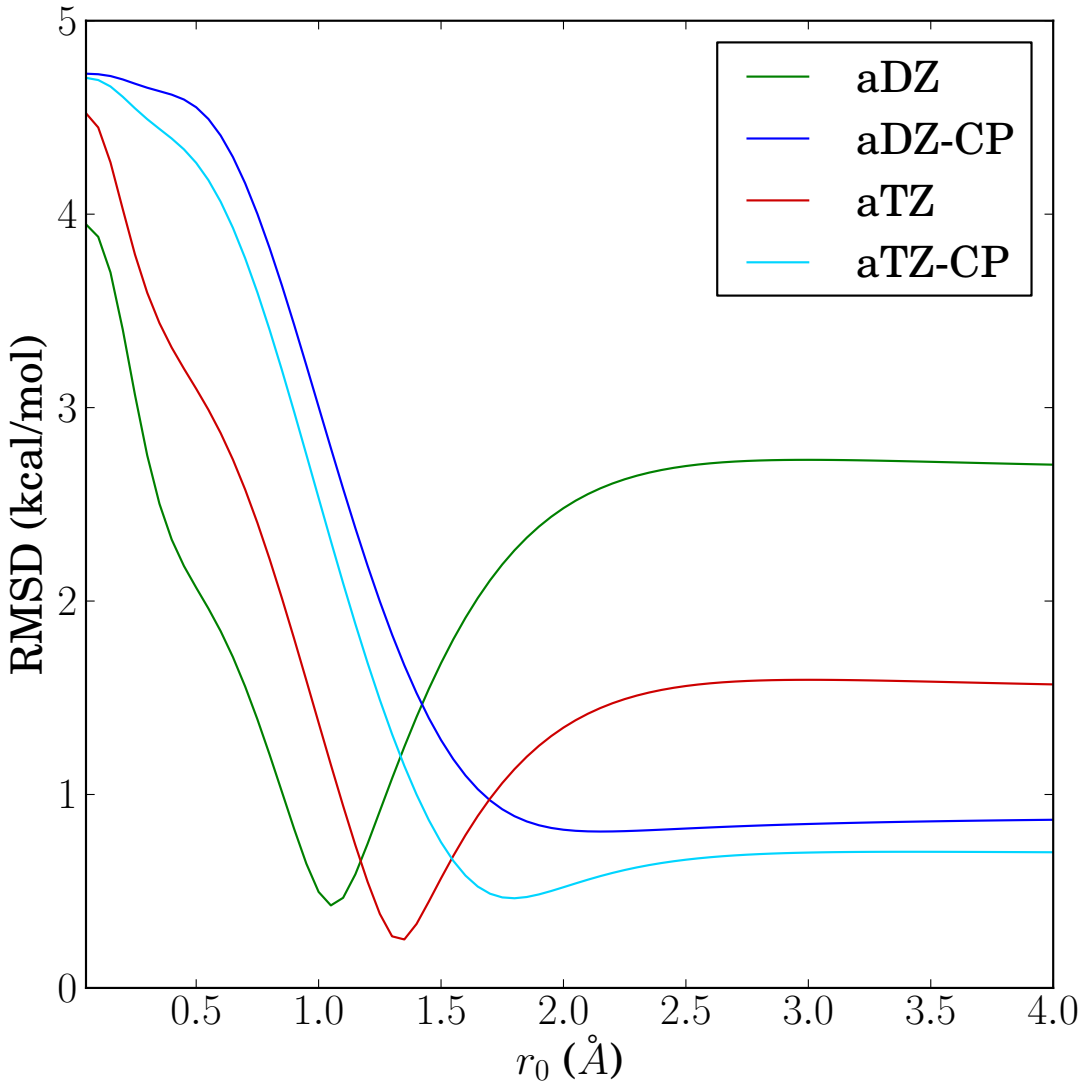


Figure 3.3: Effect of counterpoise correction on root mean squared deviation of truncated MP2 methods for training set S66 with terfc-attenuation. As $r_0 \rightarrow 4.0$Å, attenuated MP2 converges to the unattenuated result. As $r_0 \rightarrow 0$Å, attenuated MP2 approaches HF results.

shares nearly the same optimal r_0 for all types of interactions, suggesting that this parameterization is not heavily biased toward one type of interaction. This encouraging result suggests good transferability.

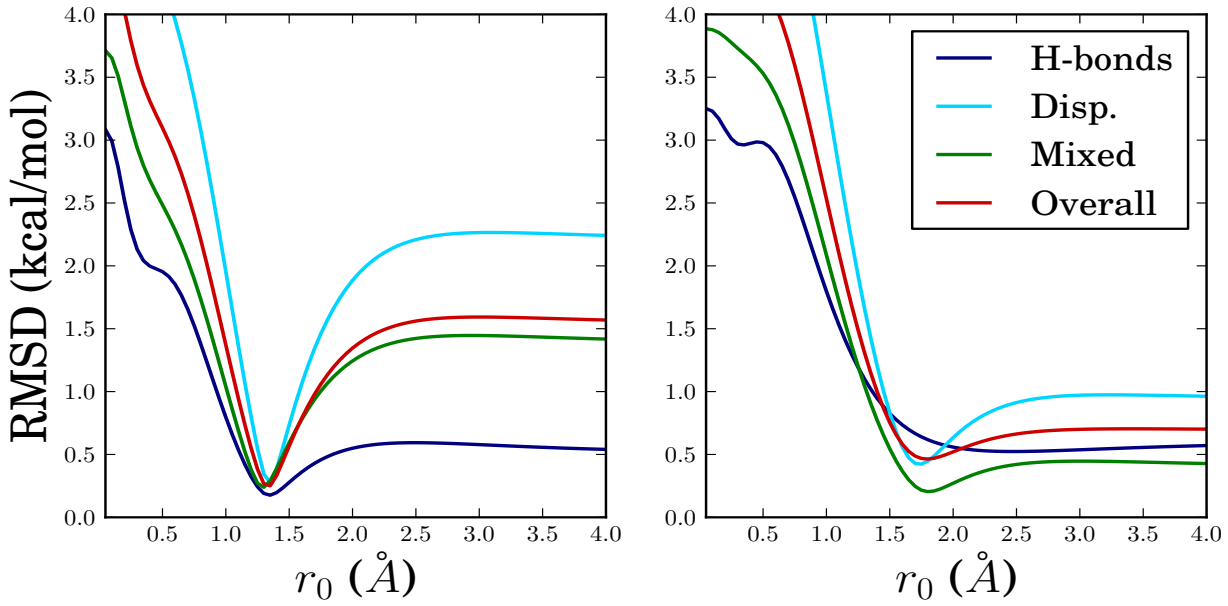


Figure 3.4: Root mean squared deviations for MP2(terfc, aTZ) (left) and MP2(terfc, aTZ-CP) (right) versus r_0 for various subsets of the S66 database

3.4 Tests of transferability

Table 3.2 presents results for terfc-attenuated MP2 for the S22 database of intermolecular interactions[145], which has recently been updated with improved estimates of the CCSD(T)/CBS energies[161]. MP2(terfc, aTZ) reduces the RMS error of standard MP2/aTZ by about 80%, which indicates a high degree of transferability from the S66 training set. Furthermore, significant improvement is shown for MP2(terfc, aTZ) over MP2(terfc, aDZ) with a 21% reduction in RMSD. The average error in MP2(terfc, aTZ) reflects a more complete recovery of the unattenuated MP2 correlation energy due to the larger r_0 in that basis. Also notable is the similarity of treatment of the dispersion and mixed subsets by MP2(terfc, aDZ) and MP2(terfc, aTZ). The main improvement in the MP2(terfc, aTZ) results relative to MP2(terfc, aDZ) is for the hydrogen-bonded subset, which is consistent with slightly reduced attenuation due to unattenuated MP2/aTZ being a somewhat better reference than MP2/aDZ.

Table 3.3 shows the behavior of attenuated MP2 for the 76 conformers of the P76 dataset[163]. Relative conformational energetics test the quality of description of intramolecular interactions in a case where CP corrections are not readily possible in conventional calculations. Relative to reference results at the extrapolated CCSD(T)/CBS limit), attenuated MP2 in both aDZ and aTZ basis sets shows similar results for overall RMSD ($\sim$0.3 kcal mol^{-1}). In the aTZ basis, this is nonetheless a 50% reduction in RMS error relative to conventional MP2. Furthermore both attenuated MP2 methods yield results that are better than the MP2/CBS limit, despite computational effort that is significantly reduced in the aTZ case, and dramatically reduced in the aDZ case.

Table 3.2: Root-mean-squared deviations, average, and mean unsigned errors on the S22 database (kcal mol^{-1})

RMSD	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/CBS[a]
H-Bonds	0.30	0.73	0.80	1.02	0.20
Disp.	0.50	3.01	0.45	4.60	1.93
Mixed	0.58	2.96	0.52	4.75	1.41
Overall	0.48	2.50	0.61	3.91	1.39
AVG	-0.26	-1.76	0.01	-2.77	-0.84
MUE	0.37	1.76	0.51	2.79	0.89

[a]From the Benchmark Energy and Geometry DataBase[2]

Table 3.3: Root-mean-squared deviations for different protein subsets of the P76 database (kcal mol^{-1})

Subset	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/CBS[a]
fgg	0.36	0.61	0.46	1.15	0.35
gfa	0.20	0.51	0.25	1.49	0.59
ggf	0.35	0.38	0.33	0.98	0.44
wg	0.16	0.58	0.19	0.57	0.19
wgg	0.40	0.80	0.38	0.89	0.41
Overall	0.31	0.59	0.33	1.06	0.42

[a]From the Benchmark Energy and Geometry DataBase[2]

The ACONF[169] database of the GMTKN30[132] presents W1h-val reference values for conformational energies of alkanes. This dataset targets intramolecular dispersion interactions. The results for terfc-attenuated MP2 on the ACONF dataset are presented in Table 3.4. MP2(terfc, aTZ) dramatically improves over both unattenuated MP2/aTZ (66% reduction in RMS error), and performs better than the MP2/CBS limit result. The reliable behavior for small alkanes here suggests that intramolecular dispersion is handled comparatively and transferably well by MP2(terfc, aTZ). By contrast, the MP2(terfc, aDZ) results are somewhat less good, although the 0.29 kcal/mol RMS error marginally improves upon the conventional MP2/aDZ RMS error of 0.31 kcal/mol.

Table 3.4: Root-mean-squared deviations and average errors on the ACONF database (kcal mol^{-1})

	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/CBS[a]
RMSD	0.08	0.24	0.29	0.31	0.11
Avg	-0.05	-0.21	0.24	-0.28	-0.08

[a]From Goerigk and Grimme[184]

The SCONF[170,185] database of the GMTKN30 comprises CCSD(T)/CBS reference values for sugar conformers, sampling different intramolecular interactions. MP2(terfc, aTZ) reduces the errors in MP2/aTZ by over 40% with a virtually identical improvement over far more computationally demanding MP2/CBS calculations. Since no similar compounds are included in the training

set, the improved behavior here also supports the transferability of attenuated MP2(terfc, aTZ). By contrast, the results with MP2(terfc, aDZ) are significantly worse, with RMS errors over 4 times larger than MP2(terfc, aTZ), and no improvement over the 0.28 kcal/mol error of MP2/aDZ.

Table 3.5: Root-mean-squared deviations and average errors on the SCONF database (kcal mol^{-1})

	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/CBS[a]
RMSD	0.12	0.22	0.52	0.28	0.21
Avg	0.03	0.08	-0.29	-0.08	-0.01

[a]From Goerigk and Grimme[184]

The CYCONF[171] database of the GMTKN30 presents CCSD(T)/CBS reference values for conformers of the amino acid cysteine. These conformers predominantly sample intramolecular hydrogen-bonds involving oxygen, sulfur, and nitrogen. This case illustrates the fact that errors in relative energies can occasionally cancel very well in otherwise poor levels of theory. The results in best agreement with the benchmark values are conventional MP2/aDZ calculations, surpassing MP2/aTZ, and even the MP2/CBS limit! As a result, MP2(terc, aDZ) slightly degrades MP2/aDZ. By contrast, MP2(terfc, aTZ) improves MP2/aTZ significantly and is also better than the MP2/CBS results.

Table 3.6: Root-mean-squared deviations and average errors on the CYCONF database (kcal mol^{-1})

	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/CBS[a]
RMSD	0.21	0.30	0.28	0.20	0.25
Avg	0.17	0.26	-0.18	0.09	0.22

[a]From Goerigk and Grimme[184]

Typically, MP2/CBS outperforms almost every lower scaling method on hydrogen-bonded systems and produces a high fidelity of agreement with CCSD(T)/CBS. This is particularly true in the case of the solvation of sulfate anions by water in the SW49 database[172,186,187]. Table 3.7 shows the behavior for terfc-attenuated MP2 for the relative energies of hydrogen bond rearrangement for the 3-6 waters solvating the sulfate anion. MP2/aTZ, MP2(terfc, aTZ), and MP2(terfc, aDZ) perform similarly for relative energies regardless of number of waters involved. For binding energies corresponding to dissociating these sulfate-water clusters, as shown in Table 3.8, MP2(terfc, aTZ) performs similarly to MP2/CBS, reflecting the removal of BSSE from this computation.

Our final test probes whether or not the good results shown above for small systems can also transfer to intermolecular interactions between larger molecules. As shown by Janowski, et al.[188], MP2 performs particularly poorly for the parallel-displaced (PD) coronene dimer; $(C_{24}H_{12})_2$ Their work showed that the overestimation of $\pi-\pi$ interactions by MP2 grows worse with larger molecular systems. We shall test the performance of the attenuated versus non-attenuated MP2 on the PD coronene dimer. Given the size of this system, we employ the dual basis approximation for our computations[189]. Optimized pairings for the aDZ and aTZ sets are available[190] which yield

Table 3.7: Root-mean-squared deviations for relative energies of methods on the SW49 database (kcal mol^{-1})

# Waters	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/a(TQ)Z[a]
3	0.34	0.32	0.40	0.49	0.07
4	0.44	0.36	0.30	0.44	0.11
5	0.30	0.28	0.42	0.63	0.08
6	0.43	0.37	0.27	0.40	0.11
Overall	0.39	0.34	0.34	0.49	0.10

[a]From Mardirossian et al.[172]

Table 3.8: Root-mean-squared deviations for binding energies of methods on the SW49 database (kcal mol^{-1})

# Waters	MP2(terfc, aTZ)	MP2/aTZ	MP2(terfc, aDZ)	MP2/aDZ	MP2/a(TQ)Z[a]
3	0.34	0.32	0.40	0.49	0.07
4	0.33	0.52	0.50	0.81	0.16
5	0.37	0.85	0.90	1.27	0.32
6	0.36	1.11	1.45	1.60	0.47
Overall	0.36	0.84	1.03	1.23	0.34

[a]From Mardirossian et al.[172]

roughly 5-10 times speedup with very small errors in binding energy. The dual basis approach is a generally useful strategy to reduce the cost of (attenuated) MP2 calculations, particularly in the larger aTZ basis.

Using Janowski et al's QCISD(T)-optimized geometry, we find that MP2/aDZ overbinds by almost 39 kcal/mol relative to QCISD(T)+ΔMP2, whilst MP2/aTZ overbinds by about 25 kcal/mol, as shown in Table 3.9. Even with counterpoise corrections, MP2/aTZ still overbinds by about 15 kcal/mol[188]. By contrast with these very poor results, attenuated MP2 in both aDZ and aTZ yields results that are in much better agreement with the benchmark. Specifically, the 4.1 kcal/mol error of MP2(terfc, aDZ) greatly improves upon the 39 kcal/mol error of MP2/aDZ. The 1.3 kcal/mol error of MP2(terfc, aTZ) yields even larger improvement over the 25 kcal/mol error of MP2/aTZ. These superior results for attenuated MP2 in both basis sets suggest that their advantages for intermolecular interactions can be retained for larger molecules.

3.5 Conclusions

In this work, we have developed a one-parameter short-range MP2 method for use in the aug-cc-pVTZ basis without counterpoise corrections. We optimized the terfc attenuator on the S66 database of intermolecular interactions to obtain the parameter $r_0 = 1.35\,\text{Å}$. This compares with our recommended value of $r_0 = 1.05\,\text{Å}$ in the aug-cc-pVDZ basis. We tested both attenuated MP2

Table 3.9: Binding energy of the parallel-displaced coronene dimer (kcal mol^{-1})

Method	Binding energy
MP2/aDZ	58.772
MP2(terfc, aDZ)	24.082
MP2/aTZ	45.031
MP2(terfc, aTZ)	21.272
QCISD(T)†	17.674
QCISD(T)+ΔMP2 †	19.981

†QCISD(T) and QCISD(T)+ΔMP2 are both from Janowski et al.[188], using cc-pVDZ with augmented functions on every other carbon atom. ΔMP2 is their estimated correction for basis set incompleteness.

methods on a variety of intermolecular interactions (the S22 dataset), and a range of conformational energies. Our main conclusions are as follows.

1. Distance-based attenuation of MP2 dramatically improves treatment of most types of inter- and intramolecular interactions in the aug-cc-pVTZ basis, The extent of improvement is as much as a 5-fold reduction of the MP2/aug-cc-pVTZ RMS error in the S22 database. All types of intermolecular interactions (hydrogen bonding, dispersion, and mixed), display similar dependence on the attenuation parameter. Transferability to the test sets is generally very encouraging in that attenuation usually significantly improves MP2/aTZ and never significantly degrades MP2/aTZ.

2. For most of the cases examined, MP2(terfc, aTZ) yields errors that are smaller than MP2/CBS. In the S22 test set, the MP2(terfc, aTZ) error is over 50% lower than the MP2/CBS RMS errror.

3. The origin of the excellent results obtained with attenuation was examined carefully in the S66 training set. We found that the benefits of attenuation are far smaller when applied to counterpoise corrected results than without correction, and the resulting CP-optimized r_0 is larger. We conclude that whilst attenuating is likely to be favorable even at the MP2/CBS limit, the excellent results obtained in the aDZ and aTZ basis sets rely upon incomplete cancellation of BSSE errors with the error associated with attenuation.

4. The results suggest that MP2(terfc, aTZ) generally out-performs MP2(terfc, aDZ), with the gap being significant enough to justify the significant additional computational cost when that is computationally feasible. The adaptation and/or development of fast algorithms to evaluate the attenuated MP2 energy appears justified and desirable.

Chapter 4

Shared Memory Multiprocessing Implementation of Resolution-of-the-Identity Second-Order Møller-Plesset Perturbation Theory with Attenuated and Unattenuated Results for Intermolecular Interactions between Large Molecules

4.1 Introduction

As the computational resources accessible to theoretical and computational chemists increases, many algorithms in electronic structure theory (EST) have been designed for high-performance massively parallel (super)computer architectures, spanning across thousands of individual nodes. While such algorithms are of significant value for large-scale calculations, many users of EST software packages are limited to a few machines and therefore a relatively moderate number of cores. Algorithms built upon the message passing interface (MPI)[191] communication protocol, a common parallelization paradigm designed for the utilization of large computer clusters, typically require either a significant amount of internode communication or duplication of computational effort. Alternatively, for shared memory systems (*i.e.*, multicore or multiprocessor architectures), shared memory multiprocessing programming using open multi processing (OpenMP)[192] for example, allows one to avoid costly internode communication and duplication of computational effort. Thus, the shared memory multiprocessing programming model can provide a useful parallelization scheme for many scientists who are limited by processing time whilst possessing only modest resources that can be devoted to a single job. In this work, we provide an algorithm

that employs a single node containing multiple shared memory cores to efficiently perform EST computations as described below.

Second-order Møller-Plesset perturbation theory[193] (MP2) provides the simplest theoretical description of electron correlation that is qualitatively correct for many phenomena, especially for noncovalent interactions, where its main competitor, density functional theory (DFT), fails without dispersion corrections[56–58,127,155]. In fact, one of the primary directions of recent DFT design and improvement has been the inclusion of second-order perturbative terms applying the MP2 ansatz to Kohn-Sham orbitals[52,53]. Although MP2 is typically qualitatively correct, significant errors can and do persist, especially for π-stacking phenomena[145,146]. Given these inaccuracies, further work has been done to improve MP2 by incorporating a more accurate treatment of dispersion[128,129,147,194].

Separately, we have recently shown[181,195] that attenuation of the Coulomb operator within MP2 theory removes long-range inaccuracies as well as basis set superposition errors (BSSE) associated with finite basis sets. This approach replaces the Coulomb operator in MP2 with a short-range operator that is parametrized for each basis set. The Coulomb operator is modified using range separation, $1 = s(r) + l(r)$, taking the terf function[153] as the long-range component,

$$l(r) = \mathrm{terf}(r, r_0) = \frac{1}{2}\left\{ erf\left[\frac{(r-r_0)}{r_0\sqrt{2}}\right] + erf\left[\frac{(r+r_0)}{r_0\sqrt{2}}\right]\right\} \tag{4.1}$$

1 whose short-range complement, terfc, is given by

$$s(r) = \mathrm{terfc}(r, r_0) = 1 - \mathrm{terf}(r, r_0). \tag{4.2}$$

Replacing r^{-1} by the attenuated Coulomb operator, $s(r)r^{-1}$, optimally preserves the short-range shape of the Coulomb operator[153]. The resulting attenuated MP2 methods, denoted MP2(terfc, aug-cc-pVDZ)[181] and MP2(terfc, aug-cc-pVTZ)[195], greatly improve treatment of noncovalent interactions at the MP2 level of theory in these basis sets without increasing the underlying scaling or changing the algorithmic mechanics. In fact, for large molecules, there are future opportunities (not considered here) for lower scaling methods, since most of the matrix elements involving this attenuated Coulomb operator become numerically insignificant and can therefore be neglected.

The computational cost associated with the MP2 energy, shown here in spin-orbital notation,

$$E_{\mathrm{MP2}} = -\frac{1}{2}\sum_{ij}\sum_{ab}\frac{(ia|jb)\left[(ia|jb) - (ib|ja)\right]}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{4.3}$$

scales with the fifth power of the system size. This scaling arises from the stepwise transformation of the four-center electron repulsion integrals (ERIs) from the atomic orbital (AO) basis $(\mu, \nu, \lambda, \sigma, \ldots)$ into the molecular orbital (MO) basis, *i.e.*,

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma}(\mu\nu|\lambda\sigma)C_{\mu i}C_{\nu a}C_{\lambda j}C_{\sigma b}. \tag{4.4}$$

The notation utilized herein employs occupied indices $i, j, \ldots \in O$, the number of occupied orbitals, and virtual indices $a, b, \ldots \in V$, the number of virtual orbitals. While the computational time

Table 4.1: RI-MP2 Energy Algorithm.

Function	Computation	Memory	Disk*
1. Form $(P\|Q)^{-1/2}$	X^3	$3X^2$	X^2
2. Form $(ia\|P) = \sum_{\mu\nu}(\mu\nu\|P)C_{\mu i}C_{\nu a}$	$O(N+V)X$	$2N^2 n_X$	OVX
3. Form† $B_{ia}^{Q} = \sum_P (ia\|P)(P\|Q)^{-1/2}$	OVX^2	$2n_O VX$	OVX
4. Form $(ia\|jb) = \sum_Q B_{ia}^{Q} B_{jb}^{Q}$	O^2V^2X	$n_B VX$	0

required by this transformation can be significantly reduced by the introduction of an auxiliary basis $(P,Q,R,\ldots)$ through the resolution-of-the-identity approximation (RI-MP2)[196] as in Equation 4.5 below,

$$\begin{aligned}(ia|jb) &= \sum_Q \left(\sum_P (ia|P)(P|Q)^{-\frac{1}{2}}\right)\left(\sum_R (Q|R)^{-\frac{1}{2}}(R|jb)\right) \\ &= \sum_Q B_{ia}^{Q} B_{jb}^{Q}, \end{aligned} \tag{4.5}$$

the fundamental fifth-order scaling is not ameliorated.

The RI-MP2 energy algorithm, as summarized in Table 4.1, requires fifth-order computational effort to form the ERIs in the MO basis. Many MPI-based RI-MP2 algorithms[197–200] require distribution of the B matrices across nodes, either through duplicated computational effort or significant internode communication costs (as much as third order in the system size). This paper pursues a different approach for tackling this asymptotically rate-limiting step using shared memory multiprocessing parallelism, which requires the computation of all precursor quantities only once. This specialized algorithm is detailed below in Section 4.2. In Section 4.3, the computational performance of this algorithm is tested on linear polypeptides, which is followed by an application of the algorithm to assess further the attenuated MP2 methods in Section 4.4. Specifically, we report attenuated MP2 calculations on the L7 database[201] of large noncovalent interactions and conformers of two model tetrapeptides[202].

4.2 Algorithm

The parallel algorithm developed in this work is shown in pseudocode in Functions 1: 2-Center Integral Formation, 2: 3-Center Integral Formation, 3: B-Matrix Formation, and 4: 4-Center Integral Formation and Energy Evaluation. The main distinguishing features of this algorithm include parallel atomic orbital (AO) to molecular orbital (MO) transformation of the three-center integrals, $(ia|P)$, parallel formation of the B matrices, and parallel construction of the $(ia|jb)$ ERIs.

The diagonalization of the two-center integrals in the auxiliary basis is straightforwardly parallelized using the Scalable Linear Algebra Package (ScaLAPACK)[203]. The transformation to the MO basis of the three-center integrals in the AO basis is discretized into a sequence of single-threaded matrix operations, each distributed to different OpenMP core. The formation of the B

matrices is similarly parallelized using a batch size determined by memory constraints and number of cores. For each occupied index i inside the batch, $(ia|P)$ is distributed to a core and B_{ia}^{Q} is computed with a single-thread.

The fifth-order computation required to form the four-center integrals in the MO basis is addressed in a similar manner. We again choose the occupied index i for batching the reading of the B_{ia}^{Q} matrices from disk and the computation of $(ia|jb)$. This choice of batched index maximizes the efficiency of matrix multiplications since the number of virtual orbitals, V, is significantly larger than that of the occupied orbitals, O. We constrain the number of B matrices to be a multiple of the number of cores.

The remaining B matrices are read from disk one at a time and all possible integrals and energetic contributions are computed through distributed matrix multiplications using OpenMP threads. By using a lopsided batching system, this reduces the overall amount of disk read operations from a theoretical maximum of $\frac{O(O+1)}{2}VX$ to $\frac{O(O+1)}{2n_B}VX$, where n_B is the number of B matrices that can be stored in memory at a given time.

This algorithm has been implemented in a development version of the Q-Chem program[204]. All calculations in this work used the frozen core approximation. Reported energies were converged to 10^{-10} Hartrees with an integral threshold of 10^{-14}. Computations on the glycine polypeptides were performed using Macintosh Pro computers containing two 2.66 GHz 6-core Intel Xeon processors with 16 GB 1333 MHz DDR3 RAM. Application work was performed using a Linux compute node containing four 2.3 GHz 16-core AMD Opteron processors with 512 GB 1600 MHz DDR3 RAM. All SCF calculations were performed using the OpenMP parallel SCF routine recently introduced in Q-Chem 4.1[204].

Data: Auxiliary basis functions (P,Q)
Result: $(P|Q)^{-1/2}$ on disk
Evaluate $(P|Q)\forall\, P,Q$;
Invert to form $(P|Q)^{-1/2}$ (ScaLAPACK[203])
Store $(P|Q)^{-1/2}$ on disk $\forall\, P,Q$

Function 1: 2-Center Integral Formation

Data: Auxiliary basis functions (P,Q), atomic orbitals (μ,ν), molecular orbitals (occupied i, virtual a), and molecular orbital coefficients $C_{\mu i}$
Result: $(ia|P)$ on disk
Identify batch size n_X given memory constraints
for $P \in X$ *in batches of* n_X **do**
    Evaluate $(\mu\nu|P)$
    Form $(i\nu|P) = \sum_\mu (\mu\nu|P)C_{\mu i}$
    Form $(ia|P) = \sum_\nu (i\nu|P)C_{\nu a}$
    Store $(ia|P)$ on disk in order $(a,P,i)\ \forall\, i,a$ and $P \in n_X$
end

Function 2: 3-Center Integral Formation

Data: Auxiliary basis functions (P,Q), molecular orbitals (occupied i, virtual a), $(ia|P)$ and $(P|Q)^{-1/2}$ on disk
Result: B_{ia}^{Q} on disk
Identify largest possible batch size n_O given memory constrains and number of cores
Read $(P|Q)^{-1/2}$ from disk $\forall\, P,Q$
for $i \in O$ *in batches of* n_O **do**
    Read $(ia|P)$ from disk $\forall\, i \in n_O, a, P$
    Form $B_{ia}^{Q} = \sum_P (ia|P)(P|Q)^{-1/2}\ \forall\, i \in n_O\ a, Q$
    Store B_{ia}^{Q} on disk in order $(a,P,i)\ \forall\, i \in n_O, a$, and P
end

Function 3: B-Matrix Formation

Data: Auxiliary basis functions (P,Q), molecular orbitals (occupied i,j, virtual a,b), B_{ia}^{Q} on disk
Determine largest possible batch size n_B given memory constraints and number of cores
for $i \in O$ *in batches of* n_B **do**
    Read $B_{ia}^{Q}\ \forall\, i \in n_B, a, Q$ from disk
    for $j \in n_B$ **do**
        Form $(ia|jb) = \sum_Q B_{ia}^{Q} B_{ib}^{Q}\ \forall\, a,b,i \in n_B, j \in n_B$
        Increment energy $\forall\, a,b,i \in n_B, j \in n_B$
    end
    for $j = O$ *decreasing until* $j = i+1$ **do**
        Read $B_{jb}^{Q}\ \forall\, a, Q$ from disk
        Form $(ia|jb) = \sum_Q B_{ia}^{Q} B_{jb}^{Q}\ \forall\, a,b,i \in n_B, j$
        Increment energy $\forall\, a,b,i \in n_B$, given j
        Store B_{jb}^{Q} for reuse if possible
    end
end

Function 4: 4-Center Integral Formation and Energy Evaluation

4.3 Parallel Performance

Since the fifth-order scaling matrix multiplication to generate the four-center integrals in the MO basis determines the overall computational cost at the asymptotic limit, the efficiency of the parallelization of this function, *i.e.* Function 4: 4-Center Integral Formation and Energy Evaluation, will determine the ultimate efficiency of this algorithm. We chose to approach this limit systematically using linear polyglycines with four, eight, sixteen, and thirty-two subunits. Performance for these systems is shown in Figure 4.1 with relative speed increases due to parallelization listed for the full RI-MP2 algorithm and the isolated fifth-order step (Function 4). Table 4.2 indicates that the fifth-order computation (Function 4) dramatically increases in relative cost with system size, but the overall parallel efficiency improves concurrently.

The relatively poor parallel efficiency of the smaller test systems indicates that the lower scaling steps are not efficiently parallelized. In particular, the MO transformation of the three-center AO integrals is computed in batches of the auxiliary index based upon shells, and the storage of these integrals is seek-bound to align with the natural atomic ordering of the auxiliary index. For the case of the 32-subunit polyglycine, where Function 4 consists of 95% of the total serial RI-MP2/cc-

pVDZ cost, this algorithm performs with significantly higher parallel efficiency. In the future, greater improvements are possible with some internal reordering of the intermediate quantities to reduce the number of seeks.

Figure 4.1: Strong scaling performance of the RI-MP2 parallel algorithm presented herein for polyglycines using the cc-pVDZ AO basis set. The overall speedup is plotted on the left, whereas the speed increase for Function 4, the formation of the 4-center integrals in the MO basis, is shown on the right.

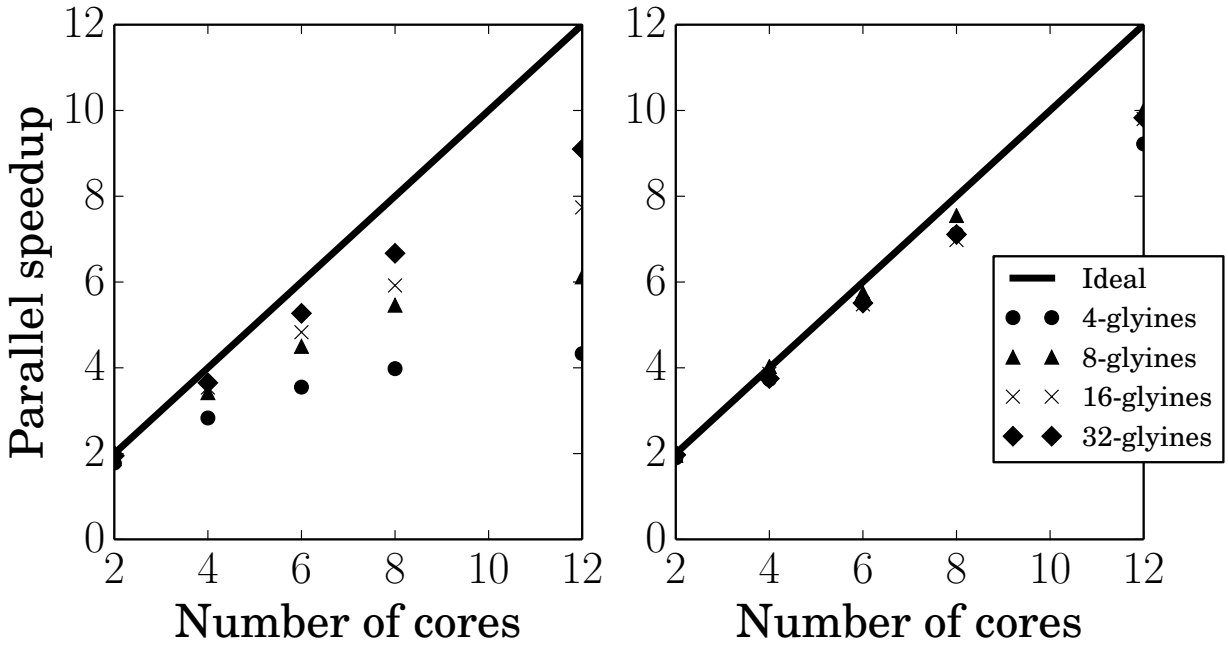


Table 4.2: Growth of the rate-limiting step (Function 4) of RI-MP2 for polyglycines using the cc-pVDZ AO basis set. Relative cost is between Function 4 and the overall RI-MP2 time when using one core.

# subunits	AO Basis functions	Relative Cost of Function 4
4	308	60%
8	592	80%
16	1160	90%
32	2296	95%

4.4 Applications

RI-MP2 remains one of the most widely used methods for treating moderate to large systems with noncovalent interactions due to its comparatively low computational scaling and qualitative accuracy. Treatment of many large systems is tenable with many current wavefunction-based methods (particularly ones that are MP2 based) in small AO basis sets. However, the cubic-scaling increase in the cost of the calculations with the number of basis functions per atom makes approaching the basis set limit computationally prohibitive for large molecules.

The L7 database[201] provides complete basis set estimates (CBS) of coupled cluster and quadratic configuration interaction with perturbative triples, CCSD(T) and QCISD(T),[205] of seven larger systems with significant dispersion interactions. These systems are as follows[201]:

- CBH: The octadecane dimer in a stacked parallel conformation.
- GGG: A π stacked guanine trimer arranged as in DNA, where the binding energy of one of the outer guanine monomers is evaluated.
- C3A: A stacked dimer of circumcoronene and adenine.
- C3GC: The binding energy between circumcoronene and a Watson-Crick hydrogen-bonded guanine-cytosine dimer.
- C2C2PD: The parallel displaced π stacked coronene dimer.
- GCGC: The binding energy of two guanine-cytosine base pairs that are arranged in a stacked Watson-Crick hydrogen-bonded arrangement as in DNA.
- PHE: The binding energy of an outer residue of a trimer of phenylalanine residues in a mixed hydrogen-bonded-stacked conformation.

In the aug-cc-pVDZ AO basis (aDZ)[154,178], these systems contain 900-2100 basis functions. Treatment within the aug-cc-pVTZ (aTZ) basis would require as many as 4000 basis functions, also causing numerical issues (such as linear dependencies) which continue to prove very problematic, as noted by the authors of the L7 database. Therefore, we restrict our analysis to the results in the aug-cc-pVDZ basis. While this basis set in known to be too small to permit generally reliable MP2 calculations, it is one of the basis sets in which we have already demonstrated greatly improved performance using attenuated MP2 for a range of small systems[181]. Therefore, the following tests on the much larger L7 systems will allow an assessment of whether the improved performance of the attenuated MP2(terfc,aug-cc-pVDZ) method relative to MP2/aug-cc-pVDZ still holds in the large-molecule limit.

Timings and energies for the L7 database are found in Tables 4.3 and 4.4 without counterpoise corrections[37] for the monomer energies. Using 64 cores, the computational cost of evaluating the RI-MP2 energies is less than 10-40% of the cost of the corresponding HF/aDZ calculations. This is somewhat surprising given the substantive size of these systems and the fifth-order scaling of

Table 4.3: Timings for the L7 database using RI-MP2/aDZ with 64 cores.

System	AO Basis Functions	SCF time (hrs)	Function 4 time (hrs)	RI-MP2 time (hrs)	% Cost RI-MP2 vs. SCF
CBH	1512	1.59	0.16	0.58	36%
C2C2PD	1320	6.45	0.10	0.46	7%
C3A	1679	13.80	0.36	1.37	10%
PHE	1413	2.84	0.18	0.61	20%
GCGC	1054	1.20	0.04	0.21	18%
GGG	894	0.61	0.02	0.10	17%
C3GC	1931	13.64	0.72	2.50	18%

Table 4.4: Energies for the L7 database and error metrics, including root-mean-squared deviations (RMSD), mean signed errors (MSE), mean unsigned errors (MUE), and maximum deviations (MAX) in kcal/mol.

System	Reference	MP2/CBS	RI-MP2(terfc, aDZ)	RI-MP2/aDZ	MP2.5/CBS
CBH	-11.06	-11.92	-10.68	-22.31	-10.88
C2C2PD	-24.36	-38.98	-24.18	-58.90	-22.80
C3A	-18.19	-27.54	-20.27	-43.46	-17.85
PHE	-25.76	-26.36	-25.63	-33.38	-25.46
GCGC	-14.37	-18.21	-15.37	-32.58	-13.41
GGG	-2.40	-4.36	-2.84	-9.81	-2.34
C3GC	-31.25	-46.02	-32.92	-72.18	-30.40
RMSD	–	8.78	1.10	24.14	0.79
MSE	–	-6.57	-0.64	-20.75	0.61
MUE	–	6.57	0.84	20.75	0.61
MAX	–	14.77	2.08	40.93	1.56

RI-MP2; however, closer examination reveals that fifth-order costs have been reduced to less than 30% of the overall RI-MP2 computational cost through efficient parallelization.

Let us now turn to the performance of the RI-MP2(terfc, aDZ) method. While RI-MP2/aDZ reproduces the sign of these interaction energies, basis set related error can be as much as 26 kcal/mol relative to the CBS estimates from the original database. By contrast, the computationally affordable RI-MP2(terfc, aDZ) method reproduces the L7 reference values quite well with a root-mean-squ deviation (RMSD) of 1.10 kcal/mol, 95% lower than that of RI-MP2/aDZ (24.1 kcal/mol) with essentially identical computational cost. The best method from the L7 database, MP2.5, has an RMSD of 0.79 kcal/mol on this database at the cost of sixth-order computation (for the MP3 energy), and was also evaluated towards the CBS limit.

Goerigk *et al.*[202] have recently reported CCSD(T)/CBS estimates for ten conformers of two model tetrapeptides, noting that limited basis MP2 frequently reorders relative conformational energetics due to basis set effects. Emphasizing the high cost of these systems, the δCCSD(T) estimates required over eight years of CPU hours. We examined these systems and report timings and energies in Tables 4.5 and 4.6 within the aDZ and aTZ AO basis sets using RI-MP2 and

Table 4.5: Timings (in minutes) for RI-MP2/aTZ on the tetrapeptide model conformers with 64 cores.

Ace-AGA-NMe [‡]	SCF time	Function 4 time	RI-MP2 time	% Cost RI-MP2 vs. SCF
β_a	120	1.5	7	5.8%
α_R	183	1.4	9	4.7%
PP-II	133	1.5	8	5.8%
α_L	183	1.5	9	4.9%
β	127	1.5	7	5.9%
Ace-ASA-NMe[§]	SCF time	Function 4 time	RI-MP2 time	% Cost RI-MP2 vs. SCF
β_a	176	2.4	11	6.4%
α_R	252	2.4	13	5.3%
PP-II	190	2.4	12	6.4%
α_L	248	2.3	13	5.2%
β	182	2.4	11	6.2%

Table 4.6: Energies for the tetrapeptide model conformers (relative to β_a) and root-mean-squared deviations.

Ace-AGA-NMe	MP2/aDZ	MP2/aTZ	MP2(terfc, aDZ)	MP2(terfc, aTZ)	MP2/CBS	CCSD(T)/CBS [¶]
β_a	0	0	0	0	0	0
α_R	-3.79	-1.81	0.37	0.28	0.10	0.57
PP-II	0.17	1.16	1.10	1.71	1.65	1.05
α_L	-2.19	-0.14	2.21	2.08	1.70	1.91
β	1.84	2.03	2.10	2.22	2.06	2.03
RMSD	3.03	1.57	0.19	0.38	0.40	–

Ace-ASA-NMe	MP2/aDZ	MP2/aTZ	MP2(terfc, aDZ)	MP2(terfc, aTZ)	MP2/CBS	CCSD(T)/CBS
β_a	0	0	0	0	0	0
α_R	-3.24	-1.37	0.73	0.63	0.53	1.05
PP-II	1.60	2.55	2.67	3.16	3.13	2.63
α_L	-2.08	-0.02	2.17	2.13	1.74	1.79
β	2.58	2.76	2.66	2.90	2.80	2.65
RMSD	2.93	1.51	0.25	0.40	0.37	–

the corresponding attenuated methods. Surprisingly, the cost of RI-MP2/aTZ is universally less than 10% of the corresponding SCF/aTZ calculation using 64 cores. The attenuated methods, RI-MP2(terfc, aDZ) and RI-MP2(terfc, aTZ), show much higher fidelity with the CCSD(T)/CBS estimates than their unattenuated counterparts, supporting that ansatz as one capable of remedying deficiencies in limited basis MP2 results. In fact, the best performing RI-MP2(terfc, aDZ) has an error that is 94% smaller than that of RI-MP2/aDZ and even outperforms MP2/CBS.

4.5 Conclusions

The shared memory multiprocessor algorithm detailed in this paper efficiently parallelizes the the evaluation of the RI-MP2 energy, with a parallel speedup that increases in efficiency with system size. Using this algorithm, we have been able to provide energies for large, noncovalently interacting systems, including the L7 database[201] and the model tetrapeptides of Goerigk *et al.*[202].

Our main conclusions follow:

1. The RI-MP2 algorithm of this work shows a parallel efficiency that increases with system size, as demonstrated by test calculations on a series of linear polyglycine chains. We recommend use of entire machines (or an entire node for multi-node systems) during application of the RI-MP2 algorithm presented herein to large molecules, in order to minimize disk read operations. Smaller systems will receive less benefits from extensive parallelization.

2. For the size regime of our application systems, we have found that RI-MP2/aDZ costs less than 40% of the underlying SCF calculations. For RI-MP2/aTZ on the tested tetrapeptides, this algorithm costs less than 10% of the underlying SCF procedure. This relative cost suggests that routine use can be made of this RI-MP2 algorithm for moderately-sized systems including 1000-2000 basis functions without any appreciable difficulty.

3. For the L7 database[201], the single-parameter attenuated RI-MP2(terfc, aDZ) shows a 95% reduction in the RMSD relative to RI-MP2/aDZ and an 87% reduction relative to MP2/CBS. On the model tetrapeptides, the single-parameter attenuated RI-MP2(terfc, aDZ) outperforms its unattenuated counterpart by 94% in RMSD, additionally outperforming MP2/CBS by over 50%. Performance comparable to MP2/CBS is attained by RI-MP2(terfc, aTZ) for this system. As a means of circumventing the high cost and inherent errors of MP2/CBS calculations, these results support the usefulness of the combination of this efficient parallel algorithm and the single-parameter attenuated MP2 methods, RI-MP2(terfc, aDZ), and RI-MP2(terfc, aTZ).

Chapter 5

Separate Electronic Attenuation Allowing a Spin-Component Scaled Second Order Møller-Plesset Theory to Be Effective for Both Thermochemistry and Non-Covalent Interactions

5.1 Introduction

Electronic structure theory pursues the solution of the electronic Schrödinger equation, which apart from relativistic and vibrational effects, is believed to be exact. However, in practice, truncations in the treatment of electron correlation and in the size of the finite basis set representation are necessary for all but the smallest of systems. While the full configuration interaction limit is usually completely intractable (although there is exciting progress towards attacking this problem[206,207]), the Møller-Plesset perturbation theory[6,7] and coupled-cluster methods[17,18] provide a systematically improvable manner for truncating the treatment of correlation.

Second order Møller-Plesset perturbation (MP2) theory provides a simple and qualitatively accurate estimate of dynamic correlation, particularly for closed shell organic and biological molecules, although it cannot be recommended for open shell systems when there is significant spin contamination[208], or an orbital instability[209]. For some intermolecular interactions, such as hydrogen-bonded clusters[172,210,211], MP2 can be exceedingly accurate, although the correlation energy exhibits only N^{-1} algebraic convergence with basis set size[212]. By contrast with hydrogen-bonding, due to its often inaccurate C_6 values[127], MP2 tends to strongly overestimate intermolecular interactions containing π-stacking motifs[145,146,213,214].

Since MP2 errors such as finite basis truncation errors appear systematic, there have been many attempts to semi-empirically modify MP2 theory to better approximate the exact, nonrelativistic limit, beginning with simply scaling the MP2 correlation energy[105,141]. It has turned out to be

far more effective to separately scale the two different spin-components of the MP2 energy, as first advocated by Grimme[106,117]. Spin-component scaling of the MP2 correlation energy (SCS-MP2) has been shown to significantly improve many types of MP2 reaction energies[107–109,215]. SCS-MP2 scales the opposite and same spin correlation components with $c_{OS} = \frac{6}{5}$ and $c_{SS} = \frac{1}{3}$ according to:

$$E_{OS} = \sum_{ia\,jb} \frac{(ia|jb)^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{5.1}$$

$$E_{SS} = \sum_{ia\,jb} \frac{(ia|jb)\left[(ia|jb) - (ib|ja)\right]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{5.2}$$

$$E_{\text{SCS-MP2}} = c_{OS} E_{OS} + c_{SS} E_{SS} \tag{5.3}$$

The SCS-MP2 approach, whilst semi-empirical in practice, can also be justified based on a redefinition of the zero order Hamiltonian[111,112]. It was also discovered that completely dropping the same-spin term, to define the scaled opposite spin MP2 (SOS-MP2) approach[120] performed essentially as well as SCS-MP2 for thermochemistry. SOS-MP2 has the advantage of requiring only fourth order computation (or less[120,123,213]) for both energy and gradient[122], rather than the standard fifth order computation of MP2 or SCS-MP2.

Further work focusing on SCS-MP2 for intermolecular interactions has shown that significantly improved performance for noncovalent interactions is possible with different parameterizations, such as the spin-component scaled MP2 for molecular interactions method, SCS(MI)-MP2[116], and alternatives[113]. These methods provide significant improvements at no additional cost, but the optimized scaling parameters (for example, in SCS(MI)-MP2, $c_{OS} = 0.40$ and $c_{SS} = 1.29$) are considerably different. The fact that the optimal SCS-MP2 parameters for thermochemistry and non-bonded interactions have values that are nearly reversed suggests that[116] "the MP2 description of bond energies contains a systematically underestimated opposite spin-component and a simultaneously overestimated same spin-component, while the reverse appears generally true for intermolecular interactions."

There have been other extensions of the SCS approach as reviewed elsewhere[110]. These include successful extensions of the SCS and SOS approaches to excited states[216,217], within the CIS(D) and CC2 frameworks[218,219]. Additionally, there has been ongoing benchmarking[144], further improvements in SCS-MP2 for intermolecular interactions[114], and the successful extension of SCS approaches to higher order coupled cluster methods[118,119], and double hybrid density functional theory[115]. However, regardless of all this progress, the problem of incompatible scaling parameters for bonded vs non-bonded interactions makes the general purpose use of SCS-MP2 methods problematical.

Attenuated MP2 is a recent development[181,195] that takes a different, complementary, approach to semi-empirically improving finite basis MP2 theory for non-covalent interactions. MP2 strongly overestimates π-stacking interactions due to its dependence on uncoupled Hartree-Fock polarizabilities. Outside of the complete basis set limit, MP2 also possesses significant basis set superposition error[177,202], which increases the overestimation of non-covalent interactions. Since both these errors have the same sign, they can be significantly compensated by attenuating the strength

of electron-electron correlations as a function of distance. Of course the attenuation protocol will be specific to a given choice of basis set. Attenuated MP2 was parametrized for the aug-cc-pVDZ (henceforth aDZ) and aug-cc-pVTZ (aTZ) basis sets[154], with reductions of several hundred percent in the RMS errors for intermolecular interactions relative to MP2 theory in the same basis set.

In detail, attenuated MP2 works by modifying the Coulomb operator within the two-electron integrals (Equation 5.4 and 6.3) such that the short-range component is preserved whilst the long-range component goes to zero. The range-separation function is chosen to be the complementary terf function (Equation 6.3), which optimally preserves the short-range behavior of the Coulomb operator[153].

$$(ia|jb) = \int\int \phi_i(r_1)\phi_a(r_1)\frac{\mathrm{terfc}(r_{12},r_0)}{r_{12}}\phi_j(r_2)\phi_b(r_2)d\tau_1 d\tau_2 \tag{5.4}$$

$$\mathrm{terfc}(r,r_0) = \frac{1}{2}\left(\mathrm{erfc}\left[\frac{(r-r_0)}{r_0\sqrt{2}}\right] + \mathrm{erfc}\left[\frac{(r+r_0)}{r_0\sqrt{2}}\right]\right) \tag{5.5}$$

The attenuation parameter for MP2(terfc, aDZ) was optimized as $r_0 = 1.05$Å, whilst for MP2(terfc, aTZ), $r_0 = 1.35$Å. Additional recent tests of the transferability of these attenuated MP2 methods to larger systems have been very encouraging[220].

Attenuated MP2 for non-covalent interactions represents the opposite of the existing scaling approaches used to correct the finite basis MP2 treatment of thermochemistry such as in scaling all correlation (SAC). For SAC-MP2, scaling factors of larger than unity are necessary to compensate for basis set incompleteness and to approximate higher order correlation effects[105,141]. As a result, attenuated MP2 methods are not expected to improve MP2 for thermochemistry. In that sense, attenuated MP2 methods have the same limitation reviewed earlier for SCS-MP2: improved accuracy for covalent interactions and non-covalent interactions require incompatible (opposite) modifications of MP2.

The purpose of this work is to propose a new method that combines spin-component scaling and electronic attenuation in such a way that the resulting method is able to inherit the good performance of SCS-MP2 for bonded interactions, and the good performance of attenuated MP2 for non-bonded interactions. The price to be paid for this step forwards is that we must increase the number of semi-empirical parameters from 2 for SCS-MP2 and 1 for attenuated MP2 to 4 for the combined method. However, this is arguably well worthwhile because the resulting method can be applied to chemical problems where energy changes involve important bonded and non-bonded contributions, without the present ambiguity of which parametrization to select.

The rest of the paper is laid out as follows. The approach we take to combine attenuated MP2 with spin-component scaling is elaborated in Section 6.2, leading to a 4-parameter form for the SCS-MP2(2terfc, aTZ) energy. The training of the 4 parameters is described in Section 6.3, which uses the S66 database of non-covalent interactions[157] and a non-multireference subset of the W4-11 benchmark dataset for thermochemistry[221]. The crucial question of the transferability of the resulting parameterized method is addressed with an extensive range of independent tests in Section 5.4, with conclusions that are generally very encouraging, as we finally summarize in Section 6.5.

5.2 Methods

Given the very promising results for non-covalent interactions obtained with attenuated MP2 with the HF/aTZ reference, we will employ that basis set. We are then confronted with the question of how attenuation can be employed to design a spin-component scaled method that performs simultaneously well for both bonded and non-bonded interactions. We have designed a relatively simple proposal that is based on the following three observations.

First, since bonded interactions occur on a shorter length-scale, we will attenuate them with a shorter length scale, $r_0^{(1)}$, than the longer attenuation length, $r_0^{(2)}$, associated with non-bonded interactions. Second, given the SCS-MP2 scaling parameters for thermochemistry ($c_{OS} = \frac{6}{5}$, $c_{SS} = \frac{1}{3}$), and the nearly equal success of SOS-MP2 for thermochemistry, we expect that the opposite-spin MP2 correlation energy can be entirely attenuated on the bonded length scale, $r_0^{(1)}$. Third, given the existing SCS(MI)-MP2 parameters for non-bonded interactions ($c_{OS} = 0.40$, $c_{SS} = 1.29$), and the nearly equal success of SSS(MI)-MP2 for non-bonded interactions[113,116], we expect that the same-spin MP2 correlation energy should be associated with the length scale, $r_0^{(2)}$ for non-bonded interactions. To accomplish this cleanly we must subtract the (smaller) same spin contribution associated with the bonded interaction length scale, to avoid double-counting contributions included in the opposite spin term. Each of the two resulting spin components will then be scaled.

The resulting method, spin-component scaled separately attenuated MP2, or, SCS-MP2(2terfc, aTZ), has two non-linear attenuation parameters ($r_0^{(1)}, r_0^{(2)}$), which enter the two-electron integrals in E_{OS} and E_{SS} through Eqs. 5.4 and 6.3. Additionally there are two linear coefficients scaling the separately attenuated same and opposite spin correlation energies. Thus the 4-parameter SCS-MP2(2terfc, aTZ) model is:

$$E = c_{OS} E_{OS}(r_0^{(1)}) + c_{SS}\left[E_{SS}(r_0^{(2)}) - E_{SS}(r_0^{(1)})\right] \tag{5.6}$$

The spin-component scaling approach described above has been implemented in a development version of Q-Chem[156,204], which was used for all calculations reported here. SCF calculations are converged to 10^{-10} Hartree using integral thresholds of 10^{-14}. Correlation calculations use the frozen core and resolution of the identity approximations.

5.3 Training

We choose as training sets the S66 database of noncovalent interactions[157] and a non-multireference subset of the W4-11 benchmark dataset for thermochemistry[221], including atomization energies, bond dissociation energies, heavy-atom transfers, isomerization energies, and nucleophilic substitution reactions. We employ an objective function constructed from root-mean-squared deviations (RMSDs), as shown in Equation 5.7 below, on the S66 and W4-11 databases as weighted by the average interaction energy of the two databases:

$$\mathrm{RMSD}_{\mathrm{Weighted}} = \frac{|E|_{\text{W4-11}}\mathrm{RMSD}_{\text{S66}} + |E|_{\text{S66}}\mathrm{RMSD}_{\text{W4-11}}}{|E|_{\text{W4-11}} + |E|_{\text{S66}}} \tag{5.7}$$

We determine the optimal non-linear attenuation lengths, $r_0^{(1)}$ and $r_0^{(2)}$, simultaneously to a resolution of 0.05Å based on explicitly evaluating the energies on a 2-d grid of that spacing. We report the linear spin component scaling coefficients to two significant figures. The dependence of our objective function upon the attenuation parameters, $r_0^{(1)}$ and $r_0^{(2)}$, is shown in Figure 5.1. In this figure, optimal spin-components scaling coefficients are determined separately at each grid point. The optimal attenuation parameters were determined to be $r_0^{(1)} = 0.75$Å, and $r_0^{(2)} = 1.05$Å while the optimal scaling coefficients were found to be $c_{\mathrm{OS}} = 1.27$ and $c_{\mathrm{SS}} = 4.05$ for opposite and same-spin correlation energies. The high same-spin scaling coefficient stems from the removal of the short-range ($r_0^{(1)}$) same-spin correlation energy in Equation 5.6.

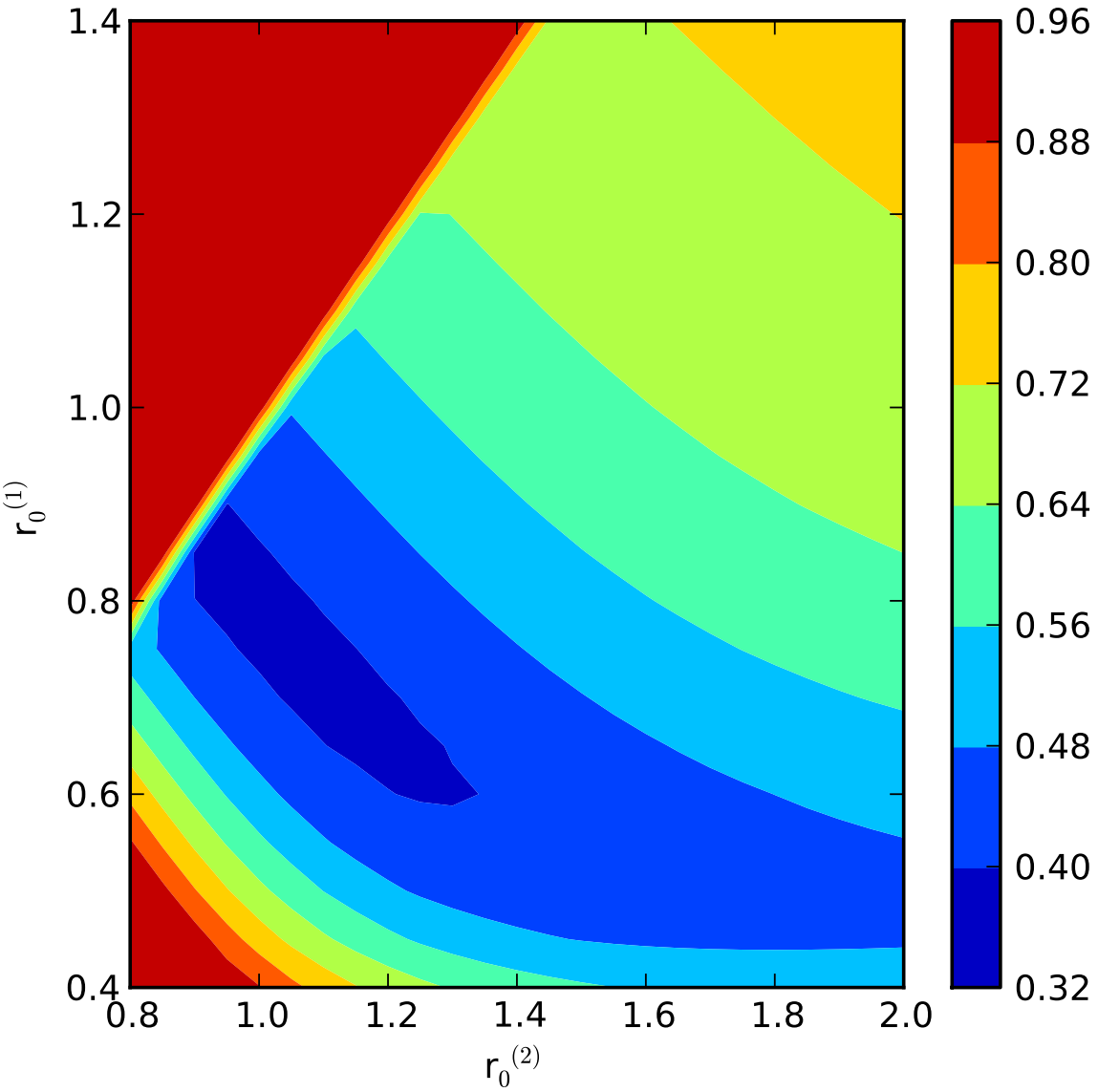
The results for SCS-MP2(2terfc, aTZ) on the W4-11 non-multireference training set are shown in Table 5.1. SCS-MP2(2terfc, aTZ) performs best, with an RMS error roughly one third lower than regular MP2/aTZ. This result is just slightly better than the improvement seen with the standard (unfitted) SCS-MP2/aTZ method. SCS-MP2(2terfc, aTZ) outperforms SCS-MP2/aTZ on the atomization, isomerization, and bond dissociation subsets, while the error is increased on the nucleophilic substitution subset. By contrast, and more or less as expected, MP2(terfc, aTZ) degrades MP2/aTZ for atomization energies, though it yields a very slight improvement of 0.3 kcal/mol in the overall RMS error relative MP2/aTZ.

Table 5.1: Error statistics on the W4-11 non-multireference training set versus W4 benchmarks (in kcal/mol) with root mean-squared deviations (RMSD) for the total atomization energies (TAE), bond dissociation energies (BDE), heavy atom transfers (HAT), isomerization energies (ISO), and nucleophilic substitution reaction (SN) subsets, with total RMSD, mean-signed error (MSE), mean-unsigned error (MUE), and maximum error (MAX)

	MP2/aTZ	SCS-MP2/aTZ	MP2(terfc, aTZ)	SCS-MP2(2terfc, aTZ)
TAE	8.33	5.96	8.59	4.80
BDE	7.79	5.92	6.68	5.54
HAT	6.89	4.75	6.41	4.86
ISO	3.32	1.88	3.02	1.76
SN	4.57	0.87	4.80	2.02
Total	7.29	5.16	6.97	4.79
MSE	-1.69	0.10	-1.33	-0.63
MUE	5.59	3.57	5.46	3.38
MAX	25.73	22.15	24.34	20.09

The performance for SCS-MP2(2terfc, aTZ) on the S66 training set is shown in Table 5.2. It is evident that the design we have chosen for SCS-MP2(2terfc, aTZ) is capable of slightly bettering the already outstanding performance of MP2(terfc, aTZ), which has an RMS error roughly 6 times smaller than unmodified MP2/aTZ. SCS-MP2(2terfc, aTZ) performs equally well on all the subsets examined, reducing overall root mean-squared deviation, mean signed error, mean unsigned error, and maximum error relative to MP2(terfc, aTZ). SCS-MP2/aTZ itself has an RMS error roughly

Figure 5.1: Weighted RMSD (kcal/mol) on S66 and W4-11 benchmark databases, as defined in Equation 5.7, evaluated as a function of the bonded attenuation length, $r_0^{(1)}$, and the non-bonded attenuation length, $r_0^{(2)}$. At each point the optimal linear coefficients are determined to obtain the value of the objective function. Note that the domain where $r_0^{(1)} \geq r_0^{(2)}$ is forbidden in Equation 5.7. The best values of $r_0^{(1)}$ and $r_0^{(2)}$ lie in a narrow valley with the minimum at $r_0^{(1)} = 0.75$Å, and $r_0^{(2)} = 1.05$Å



2.5 times smaller than MP2/aTZ, but it is between 2 and 3 times larger than MP2(terfc, aTZ) and SCS-MP2(2terfc, aTZ).

Table 5.2: Error statistics on the S66 database versus CCSD(T)/CBS benchmarks (in kcal/mol) with root mean-squared deviations (RMSD) for the hydrogen-bonded (H-bonds), dispersion-bonded (disp.), and mixed subsets, with total RMSD, mean-signed error (MSE), mean-unsigned error (MUE), and maximum error (MAX)

	MP2/aTZ	SCS-MP2/aTZ	MP2(terfc, aTZ)	SCS-MP2(2terfc, aTZ)
H-Bonds	0.506	0.585	0.176	0.174
Disp.	2.197	0.765	0.274	0.235
Mixed	1.380	0.503	0.293	0.270
Total	1.533	0.632	0.251	0.228
MSE	-1.229	-0.138	-0.068	-0.015
MUE	1.229	0.481	0.208	0.182
MAX	3.665	1.462	0.521	0.516

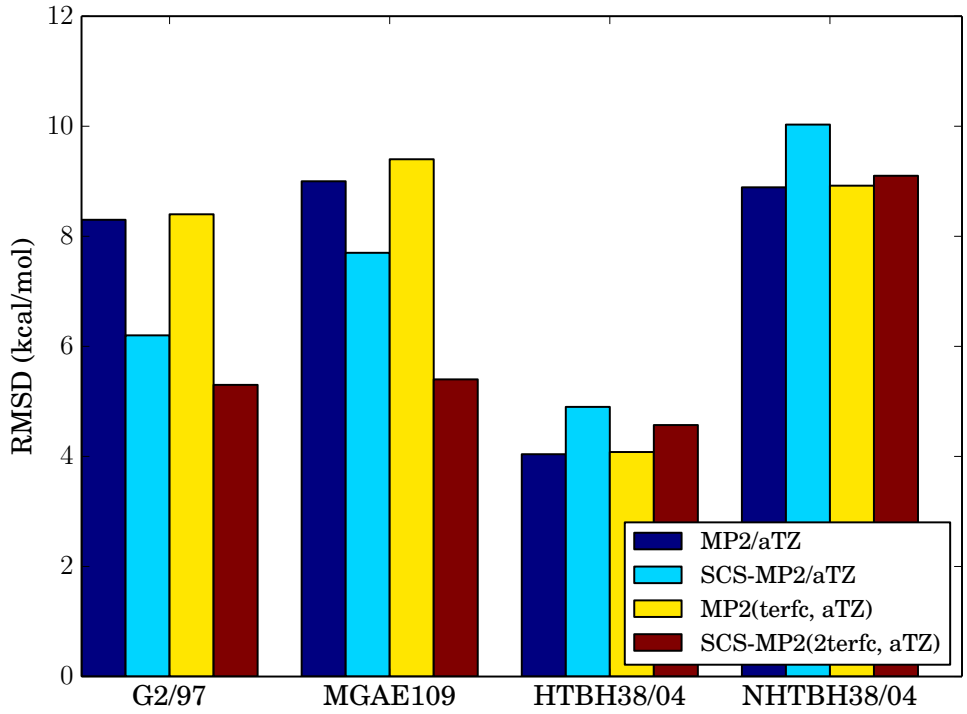
5.4 Tests

Since this spin-component scaled method is based upon an ansatz originally designed for long-range interactions, capturing the performance of spin-component scaled MP2 for thermochemistry is a necessary starting point for transferability tests. Figure 5.2 presents the behavior of MP2/aTZ, SCS-MP2/aTZ, MP2(terfc, aTZ) and SCS-MP2(2terfc, aTZ) for the G2/97[222] and MGAE109[131,223] sets of atomization energies and the HTBH38/08[131,223] and NHTBH38/08[131,223] sets of barrier height energies. For the G2/97 and MGAE109 sets, where spin-component scaling significantly improves MP2/aTZ, SCS-MP2(2terfc, aTZ) outperforms SCS-MP2/aTZ and MP2/aTZ. For the barrier height datasets, where SCS-MP2/aTZ slightly degrades MP2/aTZ, we find slight degradation relative to MP2/aTZ but to a lesser extent for SCS-MP2(2terfc, aTZ). These results suggest SCS-MP2(2terfc, aTZ) exhibits a similar level of transferability as SCS-MP2 for thermochemistry for similar reasons.

The behavior of SCS-MP2(2terfc, aTZ) for noncovalent interactions is shown in Figure 5.3. The databases presented are the S22 database of intermolecular interactions[145,161], the relative energetics of 76 conformers of small tripeptides (denoted herein P76)[163], several relative conformational energetics databases from the GMTKN30[132], including alkanes (ACONF)[169], cysteine (CYCONF)[171], and sugars (SCONF)[170,185], and sulfate-water cluster conformers with both relative and binding energies, SW49(rel) and SW49(bind)[172,186,187].

For non-covalent databases where SCS-MP2/aTZ outperforms MP2/aTZ (the S22, P76, ACONF, and SW49(rel) databases), SCS-MP2(2terfc, aTZ) exceeds or matches SCS-MP2/aTZ. When MP2(terfc, aTZ) significantly outperforms SCS-MP2/aTZ (the S22, ACONF, SCONF, and SW49(bind) databases), SCS-MP2(2terfc, aTZ) matches this behavior. SCS-MP2(2terfc, aTZ) is

Figure 5.2: Root-mean-squared-deviations (RMSDs) in kcal/mol for MP2/aTZ, SCS-MP2/aTZ, MP2(terfc, aTZ), and SCS-MP2(2terfc, aTZ) for thermochemistry datasets

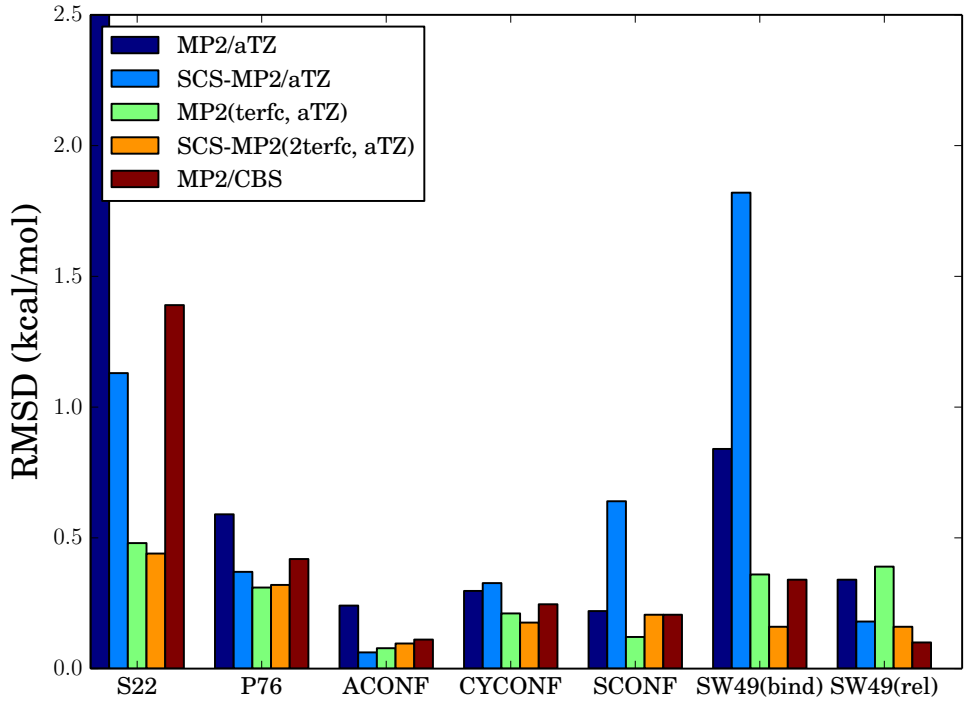


the best method for the S22, CYCONF, and SW49(bind) databases. The SCONF database shows a low RMSD for all methods ($\leq$ 0.5 kcal/mol) except for SCS-MP2/aTZ, which appears to be quite unfavorable. In this instance, MP2(terfc, aTZ) performs best while SCS-MP2(2terfc, aTZ) deviates slightly. When spin-component scaling degrades MP2/aTZ for the SW49(bind) databases, SCS-MP2(2terfc, aTZ) also deviates from MP2(terfc, aTZ), though in a favorable manner.

The error in the MP2 estimate of binding energies for noncovalent interactions grows non-linearly with system size. As a test of this behavior, we examined the L7 database[201], which contains seven large dispersion-bound complexes which were examined at the CCSD(T)/CBS or QCISD(T)/CBS level of theory. These include the octadecane dimer (CBH), the guanine trimer (GGG), the circumcoronene adenine dimer (C3A), the circumcoronene Watson-Crick guanine-cytosine dimer (C3GC), the parallel-displaced coronene dimer (C2C2PD), stacked Watson-Crick guanine-cytosine dimers (GCGC), and the phenylalanine trimer (PHE). Using the resolution of the identity and dual basis approximations[224], these systems were tabulated at the aug-cc-pVTZ level with results summarized in Table 5.3. The high error of MP2/aTZ is reduced through attenuation and spin-component scaling. It is noteworthy that SCS-MP2(2terfc, aTZ) reduces the RMS errors of both SCS-MP2 and SCS(MI)-MP2 by approximately a factor of two.

SCS-MP2(2terfc, aTZ) does not reproduce the L7 benchmarks as reliably as MP2(terfc, aTZ), due primarily to a systematic relative underbinding (compare the mean-signed error). The un-

Figure 5.3: Root-mean-squared-deviations (RMSDs) kcal/mol for MP2/aTZ, SCS-MP2/aTZ, MP2(terfc, aTZ), SCS-MP2(2terfc, aTZ), and MP2/CBS*for noncovalent interaction database



derbinding likely stems from the harsher attenuation of the same-spin correlation within SCS-MP2(2terfc, aTZ) (where $r_0^{(2)} = 1.05$Å) than in MP2(terfc, aTZ) (where $r_0 = 1.35$Å). This suggests that a long-range correction to the SCS-MP2(2terfc, aTZ) method might be a useful addition in the future.

The atomization energies of linear alkane chains are poorly treated by MP2 in a limited basis set relative to W4/quasi-W4 estimates[225]. Errors in total atomization energies for MP2 and SCS-MP2 in the aug-cc-pVTZ and aug-cc-pVQZ (aTZ and aQZ) basis sets, MP2(terfc, aTZ), and SCS-MP2(2terfc, aTZ) are plotted in Figure 5.4. Neither attenuated nor spin-component scaling alone ameliorates the increase in error with system size, but encouragingly, SCS-MP2(2terfc, aTZ) exhibits behavior much more consistent with MP2/aQZ and SCS-MP2/aQZ.

5.5 Conclusions

This work reported a spin-component scaled separately attenuated MP2 method within the aug-cc-pVTZ basis, denoted as SCS-MP2(2terfc, aTZ). We optimized the attenuation parameters and scaling coefficients using the W4-11 database of thermochemistry reactions and S66 database of noncovalent interactions to find attenuation parameters of 0.75 and 1.05Å and scaling coefficients of 1.27 (c_{OS}) and 4.05 (c_{SS}). We have tested this method against MP2/aTZ, SCS-MP2/aTZ, and

Table 5.3: Performance for MP2/aTZ variants versus L7 benchmarks (in kcal/mol) with root mean-squared deviation (RMSD), mean-signed error (MSE), mean-unsigned error (MUE), and maximum error (MAX)

System	Reference[a]	MP2/CBS[a]	MP2/aTZ	SCS-MP2/aTZ	SCS-MI-MP2/aTZ[b]	MP2(terfc, aTZ)	SCS-MP2(2terfc, aTZ)
CBH	-11.06	-11.92	-15.71	-11.83	-10.95	-8.39	-7.94
C2C2PD	-24.36	-38.98	-45.03	-33.79	-33.72	-21.27	-18.94
C3A	-18.19	-27.54	-32.85	-25.18	-25.00	-17.11	-15.69
PHE	-25.76	-26.36	-29.65	-26.25	-27.44	-24.82	-24.60
GCGC	-14.37	-18.21	-24.83	-18.59	-17.32	-14.63	-13.85
GGG	-2.40	-4.36	-6.99	-4.66	-3.65	-2.65	-2.23
C3GC	-31.25	-46.02	-54.95	-41.66	-41.60	-28.86	-26.65
RMSD	0.00	8.78	14.00	6.21	6.03	1.87	3.12
MSE	0.00	-6.57	-11.80	-4.94	-4.61	1.38	2.50
MUE	0.00	6.57	11.80	4.94	4.65	1.52	2.50
MAX	0.00	14.77	23.70	10.41	10.35	3.09	5.42

[a]Reference and MP2/CBS values obtained from the Benchmark Energy and Geometry DataBase[2]
[b]Obtained using $c_{\mathrm{OS}} = 0.29$ and $c_{\mathrm{SS}} = 1.46$

MP2(terfc, aTZ) on a range of thermochemistry datasets and intermolecular and intramolecular interaction datasets. Our conclusions from these tests are as follows.

1. SCS-MP2(2terfc, aTZ) performs favorably when spin-component scaling improves MP2/aTZ for thermochemistry. When SCS-MP2/aTZ degrades MP2/aTZ results, SCS-MP2(2terfc, aTZ) outperforms SCS-MP2/aTZ, which suggests that SCS-MP2(2terfc, aTZ) exceeds SCS-MP2/aTZ in transferability.

2. For noncovalent interactions, SCS-MP2(2terfc, aTZ) typically matches MP2(terfc, aTZ) quality. On all but the SW49(rel) database, SCS-MP2(2terfc, aTZ) reduces MP2/CBS RMSDs for noncovalent interactions at a fraction of the cost.

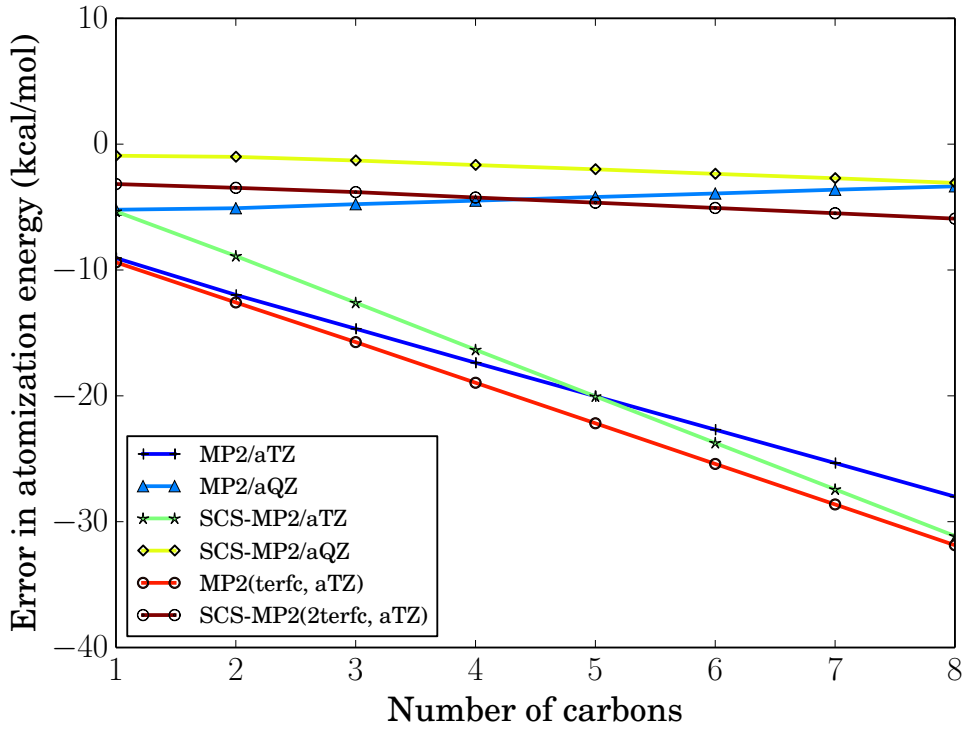
3. SCS-MP2(2terfc, aTZ) and MP2(terfc, aTZ) reproduce benchmark values for the L7 database of large, noncovalent interactions with significantly higher fidelity than MP2/aTZ and MP2/CBS, surpassing MP2/CBS RMSDs by at least 5 kcal/mol.

4. The poor behavior of MP2 for total atomization energies of linear alkanes in a limited basis (aTZ) is not ameliorated by spin-component scaling or attenuation, though SCS-MP2(2terfc, aTZ) performs similarly to MP2/aQZ results.

5. For limited basis studies of mixed interactions and chemical problems, SCS-MP2(2terfc, aTZ) reproduces the improvements of SCS-MP2 for thermochemistry while frequently matching or outperforming MP2/CBS results for noncovalent interactions.

6. There are a variety of interesting possible future developments. The formulation in terms of attenuated MP2 components permits the development of lower-scaling algorithms; and investigation of either long-range corrections, and/or development of a double hybrid density functionals based upon this approach appear interesting.

Figure 5.4: Growth of error in atomization energy (kcal/mol) as a function of alkane size



Chapter 6

Convergence of attenuated MP2 to the complete basis set limit: Improving MP2 for long-range interactions without basis set incompleteness

6.1 Introduction

Systematically approximating the electronic Schrödinger equation to generate a chemical model[3] requires truncation by level of excitation (i.e. number of occupied-virtual substitutions) as well as use of a finite basis set capable of efficiently representing the wavefunction or density[1]. The simplest correction to the Hartree-Fock reference is second-order Møller-Plesset perturbation theory[6,7] (MP2). While MP2 in large basis sets can be impressively accurate for many systems such as hydrogen bonded complexes[172,210,211], slow convergence of the MP2 correlation energy to the complete basis set (CBS) limit, $O(N^{-1})$ for N atomic basis functions[212], can make attaining the MP2/CBS limit difficult if not computationally untenable[201]. Exciting progress toward solving this problem has been made using local correlation schemes and explicitly correlated wavefunctions[139,140], and adequately addressing basis set incompleteness and related effects on finite-basis correlation calculations remains an area of active inquiry[158,173,177,201,202,226].

The inaccurate physics encoded in MP2 for long-range dispersion-dominated interactions through poor C_6 coefficients[125,127] means that MP2 treats many π-stacking and $\pi-\pi$ complexes extremely poorly[145,146,213,214]. These systematic overestimations can be partially corrected through semi-empirical scaling[105,141], and other inaccuracies are addressed through spin-component scaling of the MP2 correlation energy[106–109,111,112,117,120,122,123,213,215]. However different spin-component scaling parameters result when they are optimized for intermolecular interactions[113,114,116]. Further improvements have been gained through mixing of density functional theory (DFT) exchange and correlation functionals with HF exchange and second order perturbation theory (PT2) correlation to produce double hybrid density functionals[52,53,143], which occasionally incorporate spin-

component scaled PT2 contributions[115].

The fundamental inaccuracies of finite-basis MP2 calculations stem from overestimation of long-range interactions due to errors in the effective C_6 coefficients[125] and from finite basis effects which require the use of correction schemes, most commonly the counterpoise correction scheme of Boys and Bernard[227]. There is some dispute as to whether this is optimal[226], and other schemes such as averaging the counterpoise corrected energy and uncorrected energy are in common use[228]. An alternative approach for BSSE in HF and DFT is the geometric counterpoise correction (gCP) of Kruse et al[162,229], which tabulates a parametrized correction for basis set superposition error. This method is particularly useful for intramolecular BSSE, which has no trivial, formally exact correction. Together with the -D3 dispersion correction[58], the composite method B3LYP-gCP-D3/6-31G* has produced promising results for limited basis studies of large systems[229].

The convergence of the HF energy with basis set is approximately exponential, with triple-zeta quality basis sets capturing reasonable portions of the CBS limit in practice. Correlation energies, on the other hand, converge only as N^{-1} for N atomic basis functions. The most popular Gaussian basis sets, the Pople-style basis sets[21], are commonly augmented with diffuse[22,23] and polarization[24] functions to improve the quality of the basis for molecular energies and properties. Correlation consistent polarized valence basis sets, styled cc-pVXZ (hereafter XZ) for cardinal number X, from Dunning, et al[25–31] are designed to systematically approach the complete basis set limit, allowing the use of basis set extrapolation schemes[32,230].

$$E_{XY}^{\text{corr}} = \frac{E_X^{\text{corr}} X^3 - E_Y^{\text{corr}} Y^3}{X^3 - Y^3} \tag{6.1}$$

The Dunning style basis sets also are commonly augmented with diffuse functions, denoted aug-cc-pVXZ (hereafter aXZ). Similarly, the latest generation Karlsruhe basis sets[231], such as def2-SVPD or def2-TZVPPD, are designed for efficient reproduction of atomic polarizabilities, with a select number of diffuse functions added and tuned appropriately. Since different chemical motifs and desired accuracies require different basis sets, the cardinal number and number of diffuse functions are chosen per problem and method. For calculations involving ions, the response to electric or magnetic fields, or energies and structures of van der Waals complexes, diffuse basis functions are essential for correlation calculations. Since these functions significantly increase the cost of the overall calculation —common correlation methods scale $O(N^4)$ with N atomic basis functions —in practice many computations use mixed basis sets, only including diffuse functions on heavy atoms[232] or on every other heavy atom[188]. One systematic approach to this increase in diffuse functions is that of Papajak et. al.[233], who generate a series of diminishingly augmented basis sets from the standard Dunning-style basis sets through the removal of diffuse functions. These "calendar" basis sets allow selective and systematic inclusion of diffuse basis functions for calculations balancing cost and performance.

One recent methodological development for addressing both sources of error for finite basis MP2 is attenuated MP2[181,195]. Attenuated MP2 partitions the Coulomb operator of two-electron integrals into short- and long-range portions, retaining only the short-range contributions to the correlation energy. This partitioning resembles the range-separation as used in the complete attenuated Schrødinger equation[88–90] and range-separated hybrid density functionals[84,85]. By only

preserving short-range correlation, attenuated MP2 removes the long-range errors of finite basis MP2 (BSSE and over-estimated C_6 coefficients), as well as all true long-range correlation.

Perhaps remarkably, attenuated MP2 is very effective. The single attenuation length, r_0, has been parametrized for the aDZ[181] and aTZ[195] basis sets. The resulting methods are denoted as MP2(terfc, aDZ) and MP2(terfc, aTZ), since the r_0 parameter derives from terfc attenuation[153] of the correlation energy. They often outperform MP2/CBS estimates of intermolecular and intramolecular interactions. For example, tests for large systems show MP2(terfc, aDZ) and MP2(terfc, aTZ) reduce MP2 errors of 20-30 kcal mol^{-1} on the coronene dimer[195,220,234] to within 2-4 kcal mol^{-1} of the best available calculations[188,201,214].

An extension has defined a transferable spin-component scaled, attenuated MP2 for bonded and nonbonded interactions, SCS-MP2(2terfc, aTZ)[234], and further work has paired attenuated MP2 with the long-range dispersion energy from time-dependent Kohn-Sham density functional theory to form the attenuated MP2C method[235], which has significant promise for modeling intermolecular interactions with high accuracy for comparatively low cost. Additionally, it has recently been discovered that attenuated MP2, despite completely omitting long-range dispersion, correctly describes the long-range correlation contributions of most noncovalent complexes of dipolar molecules, including the water-dimer[236]. This is because the dominant long-range correlation contribution is the correction of mean-field overestimates of the dipole-dipole interaction, which attenuated MP2 does capture.

Following these developments in finite basis attenuated MP2 methods, this work examines the behavior of attenuated MP2 as a function of improvements in basis set quality, towards the complete basis set (CBS) limit. As the CBS limit is approached, it becomes possible to assess the balance between the overestimation of dispersion inherent in MP2/CBS calculations and attenuation of the Coulomb operator, without interference from the presence of BSSE in the HF or MP2 energies. On the other hand, BSSE is already known to play a significant role in the success of attenuated MP2, as attenuated MP2 works far less well when counterpoise corrections to remove BSSE are performed than when they are not. We will also examine the effect of augmented functions on the success of attenuated MP2 methods in some detail.

6.2 Methods

Attenuated MP2 partitions the electron-electron interaction, r_{12}^{-1}, using smooth, range-dependent short-range functions, $s(r_{12})$ and $l(r_{12})$, such that $1 = s(r) + l(r)$. As in previous work[181,195], this function is chosen to be a combination of two error functions, terfc[153], with a single parameter, r_0.

$$\frac{1}{r} = \frac{\mathrm{terf}(r, r_0)}{r} + \frac{\mathrm{terfc}(r, r_0)}{r} \tag{6.2}$$

$$\mathrm{terfc}(r, r_0) = \frac{1}{2}\left(\mathrm{erfc}\left[\frac{(r - r_0)}{r_0\sqrt{2}}\right] + \mathrm{erfc}\left[\frac{(r + r_0)}{r_0\sqrt{2}}\right]\right) \tag{6.3}$$

This construction defines a switching distance, r_0, around which the attenuated Coulomb operator, $\frac{\mathrm{terfc}(r,r_0)}{r}$, decays.

All calculations in this work utilize a developmental version of Q-Chem 4.2[204]. MP2 energies are computed using the resolution of the identity (RI) approximation[237] and the frozen core approximation. Additionally, the dual basis approximation[238–241] was employed for all quadruple zeta basis sets. For complete basis set estimates, quadruple zeta HF is not extrapolated, but correlation energies are extrapolated using cardinal number[230]. For consistency, dual basis calculations were performed for triple-zeta correlation energies for T→Q extrapolation. No counterpoise corrections are performed for any interactions, unless explicitly indicated.

6.3 Training

As in previous work, we chose the S66 database[157] for training attenuated MP2 methods. This database contains CCSD(T)/CBS reference values for a variety of sizes and strengths of intermolecular interactions in non-covalently bound complexes at their equilibrium geometries. Before turning to attenuation of MP2 theory, it is useful to assess the performance of the unmodified MP2 calculations across a range of basis sets to explore the relative importance of basis set incompleteness errors, and inaccurate physics within MP2 itself. Results for unmodified MP2 are presented in Table 6.1 for a wide range of basis sets. No counterpoise corrections are included, since we would like to be able to directly transfer the methods (and conclusions) to non-bonded intramolecular interactions where counterpoise corrections are not possible.

Several interesting points can be made. First, if we compare the first and last lines of Table 6.1, we see that the overall improvement in accuracy between 6-31G* and aTQZ (i.e. augmented TQ extrapolation) is minimal. The relatively modest performance of aTQZ indicates the significant intrinsic errors associated with MP2 theory for calculating intermolecular interactions (particularly dispersion interactions). Despite very large errors at the SCF level, the reasonable performance of MP2/6-31G* indicates fortuitous cancellation between basis set incompleteness effects at the SCF and correlated levels, also particularly for dispersion interactions.

The second main point is that there is significant reduction in finite basis set error for SCF calculations with any inclusion of diffuse functions. However, for small basis sets (e.g. 6-31+G* or def2-SVPD or aug-cc-pVDZ) this significantly *increases* the error at the MP2 level when counterpoise corrections are not used. Only for very large basis sets (e.g. extrapolated aTQZ) are the statistics significantly better. Similiarly, the use of intermediate level of diffuse functions, via the calendar basis sets of Papajak et al.[233] leads to better overall performance than full augmentation. Thus little or no augmentation is preferable if counterpoise corrections cannot be performed.

Exploring the behavior of attenuated MP2 as a function of basis set size is the main purpose of this paper. Therefore we have used the S66 dataset to optimize the attenuation parameter, r_0 as function of basis set size for a range of regular and augmented Dunning basis sets, and the intermediately augmented calendar basis sets of Papajak et al. The optimized results without extrapolation are summarized in Table 6.2, and for TQ extrapolation, in Table 6.3. Figure 6.1 shows the behavior for attenuated MP2 as a function of r_0 for the DZ, aDZ, TZ, aTZ, QZ, aQZ, TQZ, and aTQZ basis sets. There is much information in this figure and these tables, which we shall discuss in the following paragraphs.

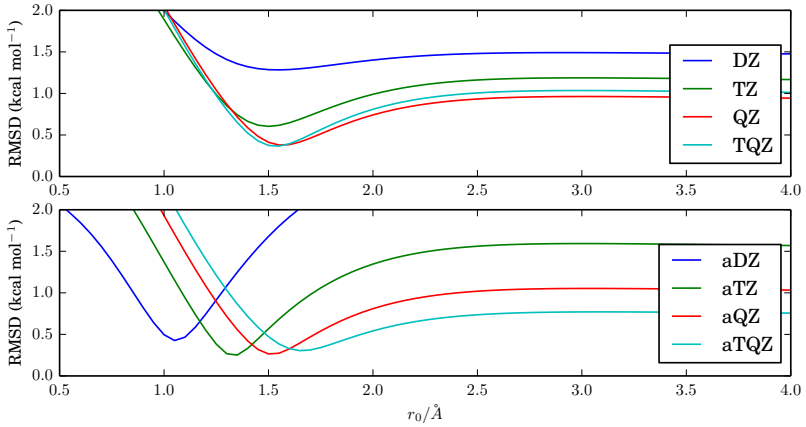


Figure 6.1: Root-mean-squared deviation (kcal mol^{-1}) on the 66 intermolecular interactions of the S66 dataset versus r_0/Å for attenuated MP2 with Dunning style basis sets

The first main point is the behavior of the RMS error as a function of basis set size augmentation. With the augmented basis sets, there is essentially no reduction in RMS error beyond the aTZ basis, with both aQZ and aTQZ showing slightly larger errors. Evidently some component of BSSE is essential for the remarkable success of attenuated MP2 in the aTZ basis. Still, it is interesting to observe that even at the aTQZ level of theory, the error without attenuation is 240% larger than with optimal attenuation. So even as the CBS limit is approached, substantial improvements in MP2 theory are possible with attenuation of the PT2 correction.

By contrast, attenuation in the non-augmented basis sets show significant reduction in RMS error as basis set is improved. However at all levels the results are much poorer than for attenuation with augmented functions. For example, MP2(terfc, QZ) has an RMS error that is still more than 40% larger than MP2(terfc, aQZ). While the intermediate calendar augmentations are superior to no augmentation at all, they fall short of the results using full augmentation at each cardinal number. The best method on this training data is attenuation in the aTZ basis: MP2(terfc, aTZ).

The second point is that r_0 behaves differently for augmented and non-augmented basis sets. For the augmented Dunning basis sets, r_0 increases monotonically from aDZ (1.05Å) to aTZ (1.35Å) to aQZ (1.50Å) to aTQZ (1.65Å), consistent with reduced attenuation being favored as BSSE is diminished with increasing basis set size. However, there is no such clear trend in the dependence of r_0 on basis set size for the non-augmented (regular) Dunning basis sets. The intermediate calendar augmentations show intermediate behavior.

We were also curious about whether MP2 in other systematic sequences of basis sets could be usefully attenuated as well. Results for a number of similar double and triple zeta quality basis sets are shown in Table 6.4. Comparing against Table 6.2, it is evident that the Dunning style basis sets generate the best performing attenuated MP2 models. Attenuated MP2 in the Karlsruhe and Pople-style basis sets yields RMS errors that are comparable to the most similar calendar basis sets. The relatively short attenuation parameter for the def2-SVPD basis set ($r_0 = 0.75$Å) stems

from poor performance for underlying MP2/def2-SVPD, which has an RMSD of 4.3 kcal mol^{-1} on the training set. The optimal attenuation parameters for def2-TZVPPD and 6-311++G** match that of aTZ (1.35Å), suggesting similar underlying error cancellation. However the RMS error is nearly 300% larger at the 6-311++G** level and is still nearly 150% larger in def2-TZVPPD.

6.4 Transferability tests

The performance of attenuated MP2 for the ACONF[169], CYCONF[171], and SCONF[170] databases is presented in Table 6.5. These databases probe the relative energies of different conformers of alkanes, cysteine, and sugars, sampling a variety of intramolecular interactions, with CCSD(T)/CBS or W1h reference values. MP2(terfc, aQZ) performs slightly less well than MP2(terfc, aTZ) with RMSDs of 0.1 to 0.2 kcal mol^{-1}, across these different systems. MP2(terfc, aTQZ) shows a slight further degradation relative to MP2(terfc, aQZ), and closely resembles MP2/aTQZ without attenuation.

Second, we examine the A24 dataset of 24 small non-covalently bound dimers, with reference CCSDT(Q)/CBS estimates of binding energies at CCSD(T)/CBS-optimized geometries[242]. The binding energies obtained by attenuated MP2 and regular MP2 in the aDZ, aTZ, aQZ, and aT→QZ basis sets are shown in Table 6.6. MP2(terfc, aTZ) matches the performance of MP2/CBS, as reported previously. In this case, MP2(terfc, aQZ) and MP2(terfc, aTQZ) outperform all other methods shown, with root-mean-squared deviations (RMSDs) of 0.137 and 0.138 kcal/mol. The improvements of MP2(terfc, aQZ) and MP2(terfc, aTQZ) relative to MP2(terfc, aTZ) are primarily found in reducing overbinding for a few systems, most notably the HCN dimer, which is overbound by 0.65 kcal/mol by MP2/aTZ and 0.55 kcal mol^{-1} by MP2(terfc, aTZ).

Finally, we assess attenuated MP2 on the S22[145,161] database of intermolecular interactions in Table 6.7. Since the error in MP2 binding energies grows with system size, significant overestimation of these MP2 binding energies occurs, with mean-signed errors between -2.77 (aDZ) and -0.83 (aTQZ) kcal mol^{-1}. The attenuated MP2 methods provide substantial error reductions relative to regular MP2 in all basis sets considered. MP2(terfc, aQZ) and MP2(terfc, aTQZ) performs similarly to MP2(terfc, aTZ), with an improved value of the mixed interaction RMSD, even relative to MP2(terfc, aTZ). MP2(terfc, aTQZ) reduces the RMS error of MP2/aTQZ by 62% and the MSE by 82%, illustrating again that attenuated MP2 outperforms conventional MP2 as the basis set limit is approached.

6.5 Conclusions

This work examines the behavior of attenuated MP2 as a function of basis set size, and level of augmentation with diffuse functions. Our results go as far as T→Q extrapolation of the correlation energy towards the CBS limit. Our main conclusions are as follows:

1. Systematic progression towards the complete basis set limit suggests an optimal MP2(terfc, aTQZ) attenuation parameter of 1.65Å, which is on a slightly longer length scale than the

aDZ (1.05Å), aTZ (1.35Å) or aQZ (1.50Å) results, as anticipated by the removal of long-range charge transfer-like BSSE.

2. Attenuated MP2 shows well-behaved convergence with cardinal number and level of augmentation. Full inclusion of diffuse functions is clearly advantageous relative to use of non-augmented Dunning basis sets. Minimally augmented triple zeta basis sets perform appreciably better than fully augmented double zeta basis sets.

3. The cancellation of MP2/CBS errors by attenuation transfers well across a number of different system types, including intramolecular and intermolecular interactions. Considering both training, and particularly test cases, MP2(terfc, aQZ) and MP2(terfc, aTQZ) perform roughly comparably in a statistical sense to MP2(terfc, aTZ), and significantly better than MP2/CBS. MP2(terfc, aTZ) is recommended due to its far lower computational cost, and if still not viable, then MP2(terfc, aDZ) is still a tremendous improvement of regular MP2 in the same basis.

Table 6.1: Performance (kcal mol^{-1}) of MP2 in various basis sets for the S66 database, including root-mean-squared deviation (RMSD) for the database, the hydrogen-bonded subset, the dispersion subset, and the mixed subset, as well as mean-signed error (MSE) and mean-unsigned error (MUE). Average finite basis set-related error (FBSE) is reported for SCF and SCF+MP2 relative to the SCF/aQZ and SCF+MP2/CBS energies. Reference SCF+MP2/CBS energies were taken from the Benchmark Energy and Geometry DataBase (BEGDB.com)[2].

Basis	RMSD	HB RMSD	DISP RMSD	MIX RMSD	MSE	MUE	SCF FBSE	MP2 FBSE
6-31g*	1.093	1.554	0.659	0.819	-0.793	0.941	-1.493	-0.388
6-31+g*	1.535	1.216	2.023	1.170	-1.064	1.203	-0.660	-0.659
6-31++g**	1.701	0.993	2.357	1.424	-1.264	1.365	-0.652	-0.860
6-311++g**	1.796	0.833	2.558	1.525	-1.225	1.397	-0.597	-0.820
def2-SVPD	4.318	2.161	5.892	4.029	-3.767	3.767	-1.293	-3.362
def2-TZVPD	1.677	0.367	2.501	1.389	-1.177	1.247	-0.038	-0.772
def2-TZVPPD	1.555	0.282	2.328	1.287	-1.111	1.128	-0.036	-0.707
DZ	1.456	2.013	1.006	1.082	-1.182	1.264	-1.454	-0.778
jun-DZ	1.312	0.642	1.918	0.985	-0.716	0.927	-0.460	-0.312
jul-DZ	1.899	0.337	2.892	1.468	-1.253	1.320	-0.415	-0.848
aDZ	2.667	0.823	3.807	2.454	-2.155	2.155	-0.626	-1.750
TZ	1.137	0.970	1.412	0.944	-0.977	0.980	-0.502	-0.572
may-TZ	0.920	0.198	1.401	0.699	-0.542	0.604	-0.083	-0.138
jun-TZ	1.205	0.176	1.841	0.927	-0.777	0.814	-0.051	-0.372
jul-TZ	1.244	0.215	1.887	0.978	-0.844	0.859	-0.054	-0.439
aTZ	1.533	0.506	2.197	1.380	-1.229	1.229	-0.095	-0.824
QZ	0.912	0.494	1.277	0.769	-0.721	0.721	-0.181	-0.316
apr-QZ	0.806	0.129	1.225	0.630	-0.501	0.528	-0.012	-0.096
may-QZ	0.872	0.136	1.330	0.673	-0.548	0.577	0.003	-0.143
jun-QZ	0.938	0.151	1.430	0.724	-0.609	0.633	0.003	-0.205
jul-QZ	0.917	0.163	1.397	0.708	-0.595	0.622	0.006	-0.190
aQZ	1.000	0.250	1.482	0.840	-0.742	0.742	–	-0.337
TQZ	0.979	0.463	1.388	0.839	-0.774	0.774	-0.181	-0.370
aTQZ	0.730	0.143	1.119	0.543	-0.457	0.479	–	-0.052

Table 6.2: Performance (in kcal mol^{-1}) of attenuated MP2 with optimal r_0/Å using calendar basis sets for the S66 database with overall root-mean-squared deviation (RMSD), mean-signed error (MSE) and mean-unsigned error (MUE), as well as RMSDs for the hydrogen-bonded, dispersion, and mixed interaction subsets

	r_0	RMSD	HB RMSD	DISP RMSD	MIX RMSD	MSE	MUE
DZ	1.55	1.283	1.933	0.743	0.709	-0.571	0.986
jun-DZ	1.50	0.687	0.784	0.772	0.403	0.118	0.510
jul-DZ	1.25	0.644	0.670	0.797	0.351	0.219	0.484
aDZ	1.05	0.426	0.483	0.311	0.469	0.051	0.325
TZ	1.50	0.604	0.826	0.520	0.326	-0.202	0.465
may-TZ	1.60	0.369	0.311	0.494	0.238	0.064	0.288
jun-TZ	1.45	0.388	0.334	0.526	0.223	0.122	0.296
jul-TZ	1.45	0.378	0.270	0.542	0.221	0.053	0.296
aTZ	1.35	0.251	0.176	0.274	0.293	-0.068	0.208
QZ	1.55	0.379	0.419	0.433	0.240	-0.049	0.305
apr-QZ	1.65	0.301	0.198	0.429	0.208	-0.003	0.237
may-QZ	1.60	0.309	0.214	0.442	0.197	0.029	0.243
jun-QZ	1.55	0.313	0.235	0.441	0.192	0.062	0.244
jul-QZ	1.55	0.315	0.251	0.437	0.187	0.077	0.245
aQZ	1.50	0.265	0.208	0.357	0.187	0.035	0.210

Table 6.3: Performance (in kcal mol^{-1}) of attenuated MP2 with optimal r_0/Å using standard Dunning basis sets with T→Q extrapolated complete basis set estimates for the S66 database with overall root-mean-squared deviation (RMSD), mean-signed error (MSE) and mean-unsigned error (MUE), as well as RMSDs for the hydrogen-bonded, dispersion, and mixed interaction subsets.

	r_0	RMSD	HB RMSD	DISP RMSD	MIX RMSD	MSE	MUE
TQZ	1.55	0.366	0.376	0.421	0.274	-0.101	0.306
aTQZ	1.65	0.304	0.214	0.440	0.174	0.032	0.237

Table 6.4: Performance (in kcal mol^{-1}) of attenuated MP2 with optimal r_0/Å using Pople-style and Karlsruhe basis sets for the S66 database with overall root-mean-squared deviation (RMSD), mean-signed error (MSE) and mean-unsigned error (MUE), as well as RMSDs for the hydrogen-bonded, dispersion, and mixed interaction subsets

	r_0	RMSD	HB RMSD	DISP RMSD	MIX RMSD	MSE	MUE
6-31g*	1.75	1.063	1.558	0.707	0.605	-0.482	0.873
6-31+g*	1.45	0.916	1.155	0.923	0.507	-0.135	0.747
6-31++g**	1.35	0.720	0.938	0.655	0.453	-0.029	0.585
6-311++g**	1.35	0.741	0.952	0.693	0.466	0.036	0.586
def2-SVPD	0.75	0.493	0.422	0.473	0.584	-0.075	0.407
def2-TZVPD	1.30	0.439	0.577	0.397	0.268	0.138	0.324
def2-TZVPPD	1.35	0.375	0.340	0.479	0.256	0.050	0.294

Table 6.5: Root-mean-squared deviations (RMSDs) in kcal mol^{-1} for attenuated and unattenuated MP2 in the augmented Dunning basis sets on intramolecular conformational energetics databases

Database	MP2/aDZ	MP2/aTZ	MP2/aQZ	MP2/aTQZ
ACONF	0.305	0.241	0.152	0.100
CYCONF	0.198	0.297	0.295	0.312
SCONF	0.282	0.220	0.313	0.130
Database	MP2(terfc, aDZ)	MP2(terfc, aTZ)	MP2(terfc, aQZ)	MP2(terfc, aTQZ)
ACONF	0.289	0.078	0.088	0.092
CYCONF	0.277	0.211	0.249	0.270
SCONF	0.519	0.121	0.129	0.140

Table 6.6: Binding energies for A24 database of attenuated and unattenuated MP2 in aDZ, aTZ, aQZ, and aTQZ basis sets with root-mean-squared deviation (RMSD), mean-signed error (MSE), and mean-unsigned error (MUE) in (kcal mol^{-1})

Dimer	Name	CCSDT(Q)/CBS	MP2/aDZ	MP2/aTZ	MP2/aQZ	MP2/aTQZ	MP2(terfc, aDZ)	MP2(terfc, aTZ)	MP2(terfc, aQZ)	MP2(terfc, aTQZ)
1	water-ammonia	-6.49	-6.95	-6.76	-6.71	-6.69	-6.68	-6.75	-6.75	-6.76
2	water-dimer	-4.99	-5.24	-5.19	-5.12	-5.08	-5.06	-5.21	-5.17	-5.14
3	hcn-dimer	-4.74	-5.63	-5.39	-5.11	-5.03	-5.26	-5.29	-5.10	-5.08
4	hf-dimer	-4.56	-4.62	-4.68	-4.64	-4.60	-4.59	-4.73	-4.70	-4.66
5	ammonia-dimer	-3.14	-3.39	-3.25	-3.22	-3.20	-2.93	-3.13	-3.18	-3.21
6	hf-methane	-1.66	-1.86	-1.85	-1.76	-1.72	-1.61	-1.81	-1.77	-1.75
7	ammonia-methane	-0.77	-1.15	-0.81	-0.76	-0.73	-0.83	-0.67	-0.69	-0.70
8	water-methane	-0.67	-0.92	-0.75	-0.68	-0.64	-0.67	-0.65	-0.64	-0.63
9	formaldehyde-dimer	-4.48	-4.83	-4.72	-4.67	-4.68	-4.02	-4.45	-4.55	-4.63
10	water-ethene	-2.56	-3.21	-3.08	-2.89	-2.80	-2.68	-2.91	-2.82	-2.79
11	formaldehyde-ethene	-1.62	-2.19	-1.92	-1.80	-1.74	-1.39	-1.55	-1.58	-1.62
12	ethyne-dimer	-1.53	-2.35	-1.94	-1.76	-1.68	-1.74	-1.64	-1.58	-1.58
13	ammonia-ethene	-1.38	-2.05	-1.76	-1.61	-1.53	-1.48	-1.52	-1.47	-1.47
14	ethene-dimer	-1.11	-1.97	-1.54	-1.39	-1.30	-0.96	-1.02	-1.05	-1.09
15	methane-ethene	-0.51	-1.01	-0.71	-0.61	-0.56	-0.57	-0.48	-0.46	-0.47
16	borane-methane	-1.51	-1.73	-1.58	-1.50	-1.44	-1.02	-1.37	-1.41	-1.43
17	methane-ethane	-0.84	-1.40	-0.98	-0.87	-0.80	-0.67	-0.65	-0.67	-0.69
18	methane-ethane	-0.61	-1.11	-0.70	-0.61	-0.56	-0.58	-0.43	-0.44	-0.46
19	methane-dimer	-0.54	-0.95	-0.61	-0.54	-0.50	-0.48	-0.39	-0.41	-0.43
20	ar-methane	-0.41	-0.56	-0.53	-0.48	-0.47	-0.21	-0.36	-0.39	-0.43
21	ar-ethene	-0.37	-0.45	-0.52	-0.49	-0.48	-0.04	-0.28	-0.33	-0.39
22	ethene-ethyne	0.79	0.13	0.33	0.39	0.41	1.22	0.90	0.77	0.65
23	ethene-dimer	0.91	0.13	0.47	0.56	0.60	1.29	1.06	0.94	0.83
24	ethyne-dimer	1.08	0.46	0.62	0.63	0.62	1.49	1.18	1.02	0.88
RMSD		0.000	0.524	0.307	0.215	0.185	0.261	0.183	0.137	0.138
MSE		0.000	-0.464	-0.256	-0.164	-0.121	0.093	-0.018	-0.030	-0.056
MUE		0.000	0.464	0.256	0.167	0.142	0.206	0.143	0.106	0.110

Table 6.7: Statistics for the performance (kcal mol^{-1}) of attenuated and unattenuated MP2 in aDZ, aTZ, aQZ, and aTQZ basis sets on the 22 intermolecular interactions defining the S22 database with root-mean-squared deviations (RMSD) for hydrogen-bonded, dispersion, and mixed subsets, as well as overall RMSD, mean-signed error (MSE), and mean-unsigned error (MUE)

Error metric	MP2/aDZ	MP2/aTZ	MP2/aQZ	MP2/aTQZ
H-bonds	1.02	0.73	0.37	0.31
Dispersion	4.60	3.01	2.27	1.86
Mixed	4.75	2.96	2.03	1.52
Overall RMSD	3.909	2.497	1.782	1.406
MSE	-2.77	-1.76	-1.16	-0.83
MUE	2.79	1.76	1.18	0.90
Error metric	MP2(terfc, aDZ)	MP2(terfc, aTZ)	MP2(terfc, aQZ)	MP2(terfc, aTQZ)
H-bonds	0.98	0.30	0.45	0.50
Dispersion	0.40	0.50	0.49	0.64
Mixed	0.43	0.58	0.42	0.46
Overall RMSD	0.649	0.479	0.451	0.536
MSE	0.25	-0.26	-0.12	-0.15
MUE	0.51	0.37	0.31	0.34

Chapter 7

Conclusion

7.1 Summary of attenuated MP2 methods

For second-order Møller-Plesset perturbation theory (MP2), small and moderate-sized basis sets are plagued not only by basis set superposition error, but also by fundamental long-range inaccuracies in the MP2 energy expression. The cost of complete basis set (CBS) limit calculations dramatically restricts the regime of applicability of MP2 computations, but even then, MP2/CBS often lacks quantitative accuracy. Attenuated MP2 directly addresses these problems through preserving only short-range correlation. The previous chapters demonstrate the applicability of attenuated MP2 for efficiently describing intramolecular and intermolecular interactions.

The cancellation of finite basis set error and methodological inaccuracies by attenuation performs well for the majority of noncovalent interactions, especially in augmented, triple-zeta basis sets. Attenuated MP2 in any augmented basis reduces MP2/CBS errors on intermolecular interactions by 60-80%, with the improvement growing more dramatic in more extended systems, especially those involving π-stacking or other van der Waals phenomena. Improvement of MP2/CBS is more difficult for intramolecular phenomena, but attenuated MP2 is perfectly suited for finite basis study of these systems, especially when basis set superposition error differs between conformations, rendering finite-basis MP2 woefully inadequate.

As basis set quality increases, the removal of finite basis set error extends the range of the attenuated correlation ansatz. Using spin-component scaling, both noncovalent and covalent bonds are transferably treated with high fidelity, though improving MP2 semi-empirically is fundamentally limited by neglect of higher order excitations and inadequacies of the underlying reference.

Much work remains to take advantage of the improvements demonstrated by these theories, namely low-scaling MP2 variants using the increased sparsity of attenuated MP2, as well as double hybrid density functionals based upon spin-component scaled attenuated MP2. The increased sparsity of integrals should advantageously be affected by the use of the terfc attenuator, which more drastically removes long-range terms due to its construction. Despite maintaining the current scaling of MP2 with system size, the ability to use small basis sets without counterpoise correction results in cost savings of up to 80% with respect to complete basis set estimates.

7.2 Future Work

7.2.1 Algorithm design

Given the enhanced sparsity of two-electron integrals included in attenuated MP2, algorithms can be designed to have improved scaling relative to the fifth-order cost of MP2. A number of possible directions forward exist, including localized orbitals, atomic-orbital ansätze, and Laplace-transformed methods. Work should also be done to assess the sparsity of attenuated integrals based on different range-separation functions and the resulting efficiency in recovering the correlation energy.

7.2.2 Long-range dispersion correction

The clearest direction forward for improving attenuated MP2 is the inclusion of long-range dispersion. This correction should result in a more compact attenuated MP2 when paired with one of the many adequate long-range dispersion corrections. Interesting paths for generating accurate long-range dispersion energies include VV10, atom-wise dispersion corrections (e.g. XDM, Grimme, or Tkatchenko-Scheffler), or long-range RPA correlation energies. The principal challenge is the design of a compatible short-range damping function.

7.2.3 Short-range correlation methods

Alternatively, other short-range correlation methods should be designed and compared. Attenuated MP2 can be viewed as the perturbation theory resulting from a short-range electron-electron interaction. Clear analogies to perturbation theory using a range-separated perturbation are possible, both in terms of attenuated third-order and fourth-order Møller-Plesset perturbation theory, as well as attenuated coupled cluster theory.

Separating the Coulomb operator into short- and long-range portions, $\frac{1}{r} = \frac{s(r)}{r} + \frac{l(r)}{r}$, short-range and long-range perturbations, $V_1 = \frac{s(r)}{r}$ and $V_2 = \frac{l(r)}{r}$, trivially define double perturbation theory in terms of different ranges of electronic interactions.

$$H = H_0 + \lambda V_1 + \mu V_2 \tag{7.1}$$

The energies are determined based upon the order of the underlying perturbations (which can differ) in operator or wavefunction, here (λ, μ).

$$\begin{aligned} E^{(2,0)} &= \langle \psi^{(0,0)} | V_1 | \psi^{(1,0)} \rangle \\ E^{(0,2)} &= \langle \psi^{(0,0)} | V_2 | \psi^{(0,1)} \rangle \\ E^{(1,1)} &= \langle \psi^{(0,0)} | V_2 | \psi^{(1,0)} \rangle + \langle \psi^{(0,0)} | V_1 | \psi^{(0,1)} \rangle \end{aligned} \tag{7.2}$$

Thus attenuated MP2 is not a unique choice, not only due to the ambiguity of choice of attenuator, but also in terms of which terms to preserve to define a short-range MP2. Currently, attenuated MP2 is defined solely as $E^{(2,0)}$, but easily implementable are variants such as $E^{(2,0)} + \frac{1}{2}E^{(1,1)}$,

which contains the entire first-order short-range correction to the wavefunction. For MP2, four contributions to the energy occur for a given range-separation function. For MP3, each expression included in the energy now has eight possible combinations of short- and long-range perturbations. Since any MPn will contain 2^n possible contributions for each term in the energy, a simplified approach is clearly needed, and ongoing work is examining the possible short-range correlation methods for suitability in modeling covalent and noncovalent compounds. These methods present the most natural directions for directly improving the short-range correlation energies while still preserving the locality and simplicity of the method.

7.2.4 Application to weakly interacting systems

Weak interactions in biomolecules frequently are poorly treated by small basis calculations with correlation methods[173,177,243]. For all but the most minuscule systems, accurate benchmarks for structure (even just along critical coordinates) or relative energetics are intractable. Using attenuated MP2, more trustworthy studies can and should be done for moderate sized biomolecules.

Bibliography

[1] T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, Ltd., New York, NY, 2000.

[2] J. Řezáč, P. Jurecka, K. E. Riley, J. Cerny, H. Valdes, K. Pluhackova, K. Berka, T. Řezáč, M. Pitoňák, J. Vondrasek and P. Hobza, *Collect. Czech. Chem. C.*, 2008, **73**, 1261–1270.

[3] J. Pople, *Rev. Mod. Phys.*, 1999, **71**, 1267–1274.

[4] M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **84**, 457–484.

[5] L. S. Cederbaum, *J. Chem. Phys.*, 2013, **138**, –.

[6] C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 0618–0622.

[7] D. Cremer, *WIREs Comput. Mol. Sci.*, 2011, **1**, 509–530.

[8] P. J. Knowles and N. C. Handy, *Chem. Phys. Lett.*, 1984, **111**, 315–321.

[9] P. E. M. Siegbahn, *Chem. Phys. Lett.*, 1984, **109**, 417–423.

[10] J. Olsen, B. O. Roos, P. J. rgensen and H. J. rgen Aa. Jensen, *J. Chem. Phys.*, 1988, **89**, 2185–2192.

[11] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, Inc., Mineola, New York, 1982.

[12] S. R. Langhoff and E. R. Davidson, *Int. J. Quantum Chem.*, 1974, **8**, 61–72.

[13] J. B. Foresman, M. Head-Gordon, J. A. Pople and M. J. Frisch, *J. Phys. Chem.*, 1992, **96**, 135–149.

[14] P. M. Zimmerman, F. Bell, M. Goldey, A. T. Bell and M. Head-Gordon, *J. Chem. Phys.*, 2012, **137**, 164110.

[15] F. Bell, P. M. Zimmerman, D. Casanova, M. Goldey and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2013, **15**, 358–366.

[16] N. J. Mayhall, M. Goldey and M. Head-Gordon, *J. Chem. Theory Comput.*, 2013.

[17] T. D. Crawford and H. F. Schaefer, *Rev. Comput. Chem.*, 2000, **14**, 33–136.

[18] R. Bartlett and M. Musial, *Rev. Mod. Phys.*, 2007, **79**, 291–352.

[19] M. Head-Gordon and J. A. Pople, *J. Chem. Phys.*, 1988, **89**, 5777.

[20] W. Klopper, K. L. Bak, P. Jørgensen, J. Olsen and T. Helgaker, *J. Phys. B-At. Mol. Opt.*, 1999, **32**, R103.

[21] R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650–654.

[22] T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.

[23] P. M. W. Gill, B. G. Johnson, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett.*, 1992, **197**, 499.

[24] M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265.

[25] T. H. Dunning, Jr., *J. Chem. Phys.*, 1989, **90**, 1007–1023.

[26] R. A. Kendall and T. H. Dunning, Jr., *Chem. Phys. Lett.*, 1992, **96**, 6796.

[27] D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.*, 1993, **98**, 1358.

[28] D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.*, 1995, **103**, 4572.

[29] D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.*, 1994, **100**, 2975.

[30] A. K. Wilson, T. van Mourik and T. H. Dunning, Jr., *J. Mol. Struct. Theochem*, 1996, **388**, 339.

[31] D. E. Woon and J. Thom H. Dunning, *J. Chem. Phys.*, 1993, **98**, 1358–1371.

[32] T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639.

[33] T. Helgaker, J. Gauss, P. Jørgensen and J. Olsen, *J. Chem. Phys.*, 1997, **106**, 6430.

[34] K. Bak, P. Jørgensen, T. Helgaker and W. Klopper, *J. Chem. Phys.*, 2000, **112**, 9229.

[35] D. Feller, *J. Chem. Phys.*, 1992, **96**, 6104–6114.

[36] D. Feller, *J. Chem. Phys.*, 1993, **98**, 7059.

[37] S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.

[38] T. van Mourik and R. J. Gdanitz, *J. Chem. Phys.*, 2002, **116**, 9620–9623.

[39] W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

[40] P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

[41] D. C. Langreth and J. P. Perdew, *Phys. Rev. B*, 1980, **21**, 5469–5493.

[42] J. P. Perdew and Y. Wang, *Phys. Rev. B*, 1986, **33**, 8800–8802.

[43] J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.

[44] D. C. Langreth and M. J. Mehl, *Phys. Rev. B*, 1983, **28**, 1809–1834.

[45] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 194112.

[46] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2007, **126**, 104102.

[47] A. Dreuw, J. L. Weisman and M. Head-Gordon, *J. Chem. Phys.*, 2003, **119**, 2943–2946.

[48] S. Kristyàn and P. Pulay, *Chem. Phys. Lett.*, 1994, **229**, 175–180.

[49] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

[50] R. H. Hertwig and W. Koch, *Chem. Phys. Lett.*, 1997, **268**, 345.

[51] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

[52] J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2009, **131**, 174105.

[53] Y. Zhang, X. Xu and W. A. Goddard, *P. Natl. Acad. Sci. USA*, 2009, **106**, 4963–4968.

[54] F. London, *Trans. Faraday Soc.*, 1937, **33**, 8b–26.

[55] J. F. Stanton, *Phys. Rev. A*, 1994, **49**, 1698–1703.

[56] S. Grimme, *Journal of Computational Chemistry*, 2004, **25**, 1463–1473.

[57] S. Grimme, *Journal of Computational Chemistry*, 2006, **27**, 1787–1799.

[58] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[59] J. G. Angyán, *J. Chem. Phys.*, 2007, **127**, 024108.

[60] A. Becke and M. Roussel, *Phys. Rev. A*, 1989, **39**, 3761–3767.

[61] A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 154108.

[62] A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **122**, 154104.

[63] A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 124108.

[64] A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2006, **124**, 14104.

[65] A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **123**, 154101.

[66] A. D. Becke, A. a. Arabi and F. O. Kannemann, *Can J Chemistry*, 2010, **88**, 1057–1062.

[67] L. A. Burns, A. Vázquez-Mayagoitia, B. G. Sumpter and C. D. Sherrill, *J. Chem. Phys.*, 2011, **134**, 84107.

[68] E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2005, **123**, 24101.

[69] E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2006, **124**, 174104.

[70] F. O. Kannemann and A. D. Becke, *J. Chem. Theory Comput.*, 2010, **6**, 1081–1088.

[71] J. Kong, Z. Gan, E. Proynov, M. Freindorf and T. Furlani, *Phys. Rev. A*, 2009, **79**, 1–10.

[72] T. Sato and H. Nakai, *J. Chem. Phys.*, 2009, **131**, 224104.

[73] A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.

[74] F. Hirshfeld, *Theor. Chem. Acc.*, 1977, **44**, 129–138.

[75] O. Vydrov and T. Van Voorhis, *Phys. Rev. A*, 2010, **81**, 1–6.

[76] O. Vydrov and T. Van Voorhis, *J. Chem. Phys.*, 2010, **133**, 244103.

[77] O. Vydrov and T. Van Voorhis, *Phys. Rev. Lett.*, 2009, **103**, 7–10.

[78] O. Vydrov and T. Van Voorhis, *J. Chem. Theory Comput.*, 2012.

[79] O. Vydrov, Q. Wu and T. Van Voorhis, *J. Chem. Phys.*, 2008, **129**, 014106.

[80] A. Dreuw, J. L. Weisman and M. Head-Gordon, *J. Chem. Phys.*, 2003, **119**, 2943.

[81] A. Lange and J. M. Herbert, *J. Chem. Theory Comput.*, 2007, **3**, 1680.

[82] A. W. Lange, M. A. Rohrdanz and J. M. Herbert, *J. Phys. Chem. B*, 2008, **112**, 6304.

[83] A. W. Lange and J. M. Herbert, *J. Am. Chem. Soc.*, 2009, **131**, 124115.

[84] P. M. W. Gill, R. D. Adamson and J. A. Pople, *Mol. Phys.*, 1996, **88**, 1005–1009.

[85] T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51 – 57.

[86] M. J. G. Peach, A. J. Cohen and D. J. Tozer, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4543–4549.

[87] A. J. Cohen, P. Mori-Sanchez and W. Yang, *J. Chem. Phys.*, 2007, **126**, 191109.

[88] A. M. Lee, S. W. Taylor, J. P. Dombroski and P. M. W. Gill, *Phys. Rev. A*, 1997, **55**, 3233–3235.

[89] P. M. Gill, *Chem. Phys. Lett.*, 1997, **270**, 193 – 195.

[90] J. P. Dombroski, S. W. Taylor and P. M. W. Gill, *J. Phys. Chem.*, 1996, **100**, 6272–6276.

[91] J. Toulouse, F. Colonna and A. Savin, *Phys. Rev. A*, 2004, **70**, 062505.

[92] J. Toulouse, A. Savin and H.-J. Flad, *Int. J. Quantum Chem.*, 2004, **100**, 1047–1056.

[93] K. Sharkas, J. Toulouse and A. Savin, *J. Chem. Phys.*, 2011, **134**, 064113.

[94] P. Gori-Giorgi and A. Savin, *Phys. Rev. A*, 2006, **73**, 032506.

[95] H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540–3544.

[96] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2004, **120**, 8425–8433.

[97] J.-W. Song, D. Peng and K. Hirao, *J. Comput. Chem.*, 2011, **32**, 3269–3275.

[98] J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.

[99] E. Weintraub, T. M. Henderson and G. E. Scuseria, *J. Chem. Theory Comput.*, 2009, **5**, 754–762.

[100] B. G. Janesko, T. M. Henderson and G. E. Scuseria, *Phys. Chem. Chem. Phys.*, 2009, **11**, 443–454.

[101] R. Haunschild and G. E. Scuseria, *J. Chem. Phys.*, 2010, **132**, 224106.

[102] R. Peverati and D. G. Truhlar, *The Journal of Physical Chemistry Letters*, 2011, **2**, 2810–2817.

[103] F. Weigend, A. Kóhn and C. Háttig, *J. Chem. Phys.*, 2002, **388**, 3175.

[104] C. Háttig, *available for download at ftp://ftp.chemie.uni-karlsruhe.de/pub/cbasen*.

[105] M. Gordon and D. Truhlar, *J. Am. Chem. Soc.*, 1986, **108**, 5412–5419.

[106] S. Grimme, *J. Chem. Phys.*, 2003, **118**, 9095–9102.

[107] S. Grimme, *J. Phys. Chem. A*, 2005, **109**, 3067–3077.

[108] M. Gerenkamp and S. Grimme, *Chem. Phys. Lett.*, 2004, **392**, 229–235.

[109] I. Hyla-Kryspin and S. Grimme, *Organometallics*, 2004, **23**, 5581–5592.

[110] S. Grimme, L. Goerigk and R. F. Fink, *WIREs Comput. Mol. Sci.*, 2012, **2**, 886–906.

[111] A. Szabados, *J. Chem. Phys.*, 2006, **125**, 214105.

[112] R. F. Fink, *J. Chem. Phys.*, 2010, **133**, 174113.

[113] J. G. Hill and J. A. Platts, *J. Chem. Theor. Comput.*, 2007, **3**, 80–85.

[114] I. Grabowski, E. Fabiano and F. Della Sala, *Phys. Chem. Chem. Phys.*, 2013, **15**, 15485–15493.

[115] S. Kozuch and J. Martin, *J. Comput. Chem.*, 2013, **34**, 2327–2344.

[116] R. A. DiStasio Jr. and M. Head-Gordon, *Mol. Phys.*, 2007, **105**, 1073–1083.

[117] J. Antony and S. Grimme, *J. Phys. Chem. A*, 2007, **111**, 4862–4868.

[118] T. Takatani, E. G. Hohenstein and C. D. Sherrill, *J. Chem. Phys.*, 2008, **128**, 124111.

[119] M. Pitonak, J. Rezac and P. Hobza, *Phys. Chem. Chem. Phys.*, 2010, **12**, 9611–9614.

[120] Y. Jung, R. C. Lochan, A. D. Dutoi and M. Head-Gordon, *J. Chem. Phys.*, 2004, **121**, 9793–9802.

[121] R. C. Lochan, Y. Shao and M. Head-Gordon, *J. Chem. Theor. Comput.*, 2007, **3**, 988–1003.

[122] R. C. Lochan, Y. H. Shao and M. Head-Gordon, *J. Chem. Theor. Comput.*, 2007, **3**, 988–1003.

[123] Y. S. Jung, Y. H. Shao and M. Head-Gordon, *J. Comput. Chem.*, 2007, **28**, 1953–1964.

[124] R. C. Lochan, Y. Jung and M. Head-Gordon, *The Journal of Physical Chemistry A*, 2005, **109**, 7598–7605.

[125] A. Szabo and N. S. Ostlund, *J. Chem. Phys.*, 1977, **67**, 4351–4360.

[126] P. W. Langhoff, M. Karplus and R. P. Hurst, *J. Chem. Phys.*, 1966, **44**, 505–&.

[127] A. Tkatchenko, R. A. DiStasio, Jr., M. Head-Gordon and M. Scheffler, *J. Chem. Phys.*, 2009, **131**, 094106.

[128] A. Hesselmann, *J. Chem. Phys.*, 2008, **128**, 144112.

[129] M. Pitoňák and A. Heßelmann, *J. Chem. Theory Comput.*, 2010, **6**, 168–178.

[130] Y. Huang, Y. Shao and G. J. O. Beran, *J. Chem. Phys.*, 2013, **138**, –.

[131] J. Zheng, Y. Zhao and D. G. Truhlar, *J. Chem. Theor. Comput.*, 2007, **3**, 569–582.

[132] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.

[133] L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **126**, 084108.

[134] J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.*, 1999, **111**, 1843–1856.

[135] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay and J. Gauss, *J. Chem. Phys.*, 2004, **120**, 4129–4141.

[136] A. Tajti, P. G. Szalay, A. G. Csaszar, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez and J. F. Stanton, *J. Chem. Phys.*, 2004, **121**, 11599–11613.

[137] Y. J. Bomble, J. Vazquez, M. Kallay, C. Michauk, P. G. Szalay, A. G. Csaszar, J. Gauss and J. F. Stanton, *J. Chem. Phys.*, 2006, **125**, 064108.

[138] M. E. Harding, J. Vazquez, B. Ruscic, A. K. Wilson, J. Gauss and J. F. Stanton, *J. Chem. Phys.*, 2008, **128**, 114111.

[139] T. B. Adler, H.-J. Werner and F. R. Manby, *J. Chem. Phys.*, 2009, **130**, 054106.

[140] T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2009, **130**, 241101.

[141] P. L. Fast, J. Corchado, M. L. Sanchez and D. G. Truhlar, *J. Phys. Chem. A*, 1999, **103**, 3139–3143.

[142] F. Aquilante and T. B. Pedersen, *Chem. Phys. Lett.*, 2007, **449**, 354 – 357.

[143] S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.

[144] K. E. Riley, J. A. Platts, J. Rezac, P. Hobza and J. Hill, *J. Phys. Chem. A*, 2012, **116**, 4159–4169.

[145] P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.

[146] S. M. Cybulski and M. L. Lytle, *J. Chem. Phys.*, 2007, **127**, 141102.

[147] A. Tkatchenko, J. Robert A. DiStasio, M. Head-Gordon and M. Scheffler, *J. Chem. Phys.*, 2009, **131**, 094106.

[148] D. R. A., R. P. Steele, Y. M. Rhee, Y. Shao and M. Head-Gordon, *J. Comput. Chem.*, 2007, **28**, 839–856.

[149] W. Klopper, F. R. Manby, S. Ten-No and E. F. Valeev, *Int. Rev. Phys. Chem.*, 2006, **25**, 427–468.

[150] C. D. Sherrill, T. Takatani and E. G. Hohenstein, *J. Phys. Chem. A*, 2009, **113**, 10146–10159.

[151] T. Van Mourik, *J. Phys. Chem. A*, 2008, **112**, 11017–11020.

[152] R. D. Adamson, J. P. Dombroski and P. M. Gill, *Chem. Phys. Lett.*, 1996, **254**, 329 – 336.

[153] A. D. Dutoi and M. Head-Gordon, *J. Phys. Chem. A*, 2008, **112**, 2110–2119.

[154] T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007–1023.

[155] A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 154108.

[156] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio Jr, R. C. Lochan, T. Wang, G. J. Beran, N. A. Besley, J. M. Herbert, C. Yeh Lin, T. Van Voorhis, S. Hung Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. Min Rhee, J. Ritchie, E. Rosta, C. David Sherrill, A. C. SimmOnett, J. E. Subotnik, H. Lee Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172–3191.

[157] J. Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427–2438.

[158] P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.

[159] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.

[160] R. Podeszwa, K. Patkowski and K. Szalewicz, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5974–5979.

[161] M. S. Marshall, L. A. Burns and C. D. Sherrill, *J. Chem. Phys.*, 2011, **135**, 194102.

[162] H. Kruse and S. Grimme, *J. Chem. Phys.*, 2012, **136**, 154101.

[163] H. Valdes, K. Pluhackova, M. Pitoňák, J. Řezáč and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2747–2757.

[164] Y. Zhao and D. Truhlar, *Theor. Chim. Acta.*, 2008, **120**, 215–241.

[165] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren and R. A. Friesner, *J. Am. Chem. Soc.*, 1997, **119**, 5908–5920.

[166] R. A. DiStasio, Jr., Y. Jung and M. Head-Gordon, *J. Chem. Theory Comput.*, 2005, **1**, 862–876.

[167] L. Gráfová, M. Pitoňák, J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2010, **6**, 2365–2376.

[168] J. A. Pople, *Angew. Chem. Int. Ed.*, 1999, **38**, 1894–1902.

[169] D. Gruzman, A. Karton and J. M. L. Martin, *J. Phys. Chem. A*, 2009, **113**, 11974–11983.

[170] G. I. Csonka, A. D. French, G. P. Johnson and C. A. Stortz, *J. Chem. Theory Comput.*, 2009, **5**, 679–692.

[171] J. J. Wilke, M. C. Lind, H. F. Schaefer, A. G. Csaszar and W. D. Allen, *J. Chem. Theory Comput.*, 2009, **5**, 1511–1523.

[172] N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas and M. Head-Gordon, *J. Chem. Theory Comput.*, 2013, **9**, 1368–1380.

[173] L. F. Holroyd and T. van Mourik, *Chem. Phys. Lett.*, 2007, **442**, 42 – 46.

[174] S. Saebo and P. Pulay, *Ann.Rev. Phys. Chem.*, 1993, **44**, 213–236.

[175] D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45 – 48.

[176] F. Neese and E. F. Valeev, *J. Chem. Theor. Comput.*, 2011, **7**, 33–43.

[177] A. E. Shields and T. van Mourik, *J. Phys. Chem. A.*, 2007, **111**, 13272–13277.

[178] R. A. Kendall, J. Thom H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.

[179] D. Feller, *J. Comput. Chem.*, 1996, **17**, 1571–1586.

[180] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.

[181] M. Goldey and M. Head-Gordon, *J. Phys. Chem. Lett.*, 2012, **3**, 3592–3598.

[182] T. Granlund and the GMP development team, *GNU MP: The GNU Multiple Precision Arithmetic Library*, 5th edn., 2012.

[183] GMPY Development Team, *GMPY: Multiple-precision arithmetic for Python*, 1st edn., 2012.

[184] L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670–6688.

[185] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **6**, 107–126.

[186] D. S. Lambrecht, G. N. I. Clark, T. Head-Gordon and M. Head-Gordon, *J. Phys. Chem. A*, 2011, **115**, 11438–11454.

[187] D. S. Lambrecht, L. McCaslin, S. S. Xantheas, E. Epifanovsky and M. Head-Gordon, *Mol. Phys.*, 2012, **110**, 2513–2521.

[188] T. Janowski, A. R. Ford and P. Pulay, *Mol. Phys.*, 2010, **108**, 249–257.

[189] R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong and M. Head-Gordon, *J. Chem. Phys.*, 2006, **125**, 074108.

[190] R. P. Steele, R. A. DiStasio, Jr. and M. Head-Gordon, *J. Chem. Theor. Comput.*, 2009, **5**, 1560–1572.

[191] Message Passing Interface Forum, *MPI: A Message-Passing Interface Standard: Version 3.0*, 3rd edn., 2012.

[192] OpenMP Architecture Review Board, *OpenMP Application Program Interface*, 3rd edn., 2008.

[193] C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.

[194] Y. Huang, Y. Shao and G. J. O. Beran, *J. Chem. Phys.*, 2013, **138**, 224112.

[195] M. Goldey, A. Dutoi and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2013, 15869–15875.

[196] M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359 – 363.

[197] D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.*, 1996, **250**, 477 – 484.

[198] M. Katouda and S. Nagase, *Int. J. Quant. Chem.*, 2009, **109**, 2121–2130.

[199] C. Hattig, A. Hellweg and A. Kohn, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1159–1169.

[200] M. Katouda and T. Nakajima, *J. Chem. Theory Comput.*, In Press.

[201] R. Sedlak, T. Janowski, M. Pitonak, J. Rezac, P. Pulay and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 3364–3374.

[202] L. Goerigk, A. Karton, J. M. L. Martin and L. Radom, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7028–7031.

[203] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker and R. C. Whaley, *ScaLAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[204] A. I. Krylov and P. M. Gill, *WIREs Comput Mol Sci*, 2013, **3**, 317–326.

[205] K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chemical Physics Letters*, 1989, **157**, 479 – 483.

[206] U. Schollwock, *Rev. Mod. Phys.*, 2005, **77**, 259–315.

[207] G. K. L. Chan and S. Sharma, *Annu. Rev. Phys. Chem.*, 2011, **62**, 465–481.

[208] D. Stuck, T. A. Baker, P. Zimmerman, W. Kurlancheek and M. Head-Gordon, *J. Chem. Phys.*, 2011, **135**, 194306.

[209] W. Kurlancheek and M. Head-Gordon, *Mol. Phys.*, 2009, **107**, 1223–1232.

[210] S. S. Xantheas and E. Apra, *J. Chem. Phys.*, 2004, **120**, 823–828.

[211] B. Temelso, K. Archer and G. Shields, *J. Phys. Chem. A*, 2011, **115**, 12034–12046.

[212] T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639–9646.

[213] Y. Jung and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2006, **8**, 2831–2840.

[214] T. Janowski and P. Pulay, *J. Am. Chem. Soc.*, 2012, **134**, 17520–17525.

[215] T. P. M. Goumans, A. W. Ehlers, K. Lammertsma, E. U. Wurthwein and S. Grimme, *Chem. Eur. J.*, 2004, **10**, 6468–6475.

[216] Y. M. Rhee and M. Head-Gordon, *J. Phys. Chem. A*, 2007, **111**, 5314–5326.

[217] A. Hellweg, S. A. Grun and C. Hattig, *Phys. Chem. Chem. Phys.*, 2008, **10**, 4119–4127.

[218] M. Head-Gordon, R. J. Rico, M. Oumi and T. J. Lee, *Chem. Phys. Lett.*, 1994, **219**, 21–29.

[219] O. Christiansen, H. Koch and P. Jorgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.

[220] M. Goldey, R. A. DiStasio, Jr., Y. Shao and M. Head-Gordon, *Mol. Phys.*, 2014, **112**, (in press).

[221] A. Karton, S. Daon and J. M. Martin, *Chem. Phys. Lett.*, 2011, **510**, 165 – 178.

[222] R. Haunschild and W. Klopper, *J. Chem. Phys.*, 2012, **136**, 164102.

[223] R. Peverati and D. G. Truhlar, *J. Chem. Phys.*, 2011, **135**, 191102.

[224] R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong and M. Head-Gordon, *J. Chem. Phys.*, 2006, **125**, 074108.

[225] A. Karton, D. Gruzman and J. M. L. Martin, *J. Phys. Chem. A*, 2009, **113**, 8434–8447.

[226] Å. M. Mentel and E. J. Baerends, *J. Chem. Theory Comput.*, 2014, **10**, 252–267.

[227] S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.

[228] L. A. Burns, M. S. Marshall and C. D. Sherrill, *J. Chem. Theory Comput.*, 2014, **10**, 49–57.

[229] H. Kruse, L. Goerigk and S. Grimme, *J. Org. Chem.*, 2012, **77**, 10824–34.

[230] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett.*, 1998, **286**, 243.

[231] D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, –.

[232] E. Papajak, H. R. Leverentz, J. Zheng and D. G. Truhlar, *J. Chem. Theory Comput.*, 2009, **5**, 1197–1202.

[233] E. Papajak, J. Zheng, X. Xu, H. R. Leverentz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 3027–3034.

[234] M. Goldey and M. Head-Gordon, *J. Phys. Chem. B*, 2014, (in press).

[235] Y. Huang, M. Goldey, M. Head-Gordon and G. Beran, *J. Chem. Theory Comput.*, 2014, Accepted.

[236] J. Thirman and M. Head-Gordon, *J. Phys. Chem. Lett.*, 2014, **5**, 1380–1385.

[237] F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175–3183.

[238] K. Wolinski and P. Pulay, *J. Chem. Phys.*, 2003, **118**, 9497–9503.

[239] S. Havriliak and H. F. King, *J. Am. Chem. Soc.*, 1983, **105**, 4–12.

[240] R. Jurgens-Lutovsky and J. Almlöf, *Chem. Phys. Lett.*, 1991, **178**, 451.

[241] R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong and M. Head-Gordon, *J. Chem. Phys.*, 2006, **125**, 074108.

[242] J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 2151–2155.

[243] D. Toroz and T. van Mourik, *Mol. Phys.*, 2006, **104**, 559–570.

Appendix A

Performance of attenuated MP2 and other methods in the aug-cc-pVDZ basis

Definitions of **I**, **II**, etc. are taken from Chapter 2.

Table A.1: Energetics for the S66 Hydrogen-Bonding Subset (kcal mol^{-1})

System	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
1	-5.01	-4.96	-5.21	-5.04	-5.05	-4.99	-4.97	-5.18	-4.64
2	-5.70	-5.69	-6.07	-5.69	-5.71	-5.63	-5.62	-5.86	-4.99
3	-7.04	-7.08	-7.50	-7.10	-7.14	-7.03	-7.01	-7.25	-6.76
4	-8.22	-8.07	-8.53	-7.93	-7.92	-7.88	-7.87	-8.77	-7.21
5	-5.85	-5.84	-6.36	-5.74	-5.75	-5.70	-5.68	-5.82	-4.82
6	-7.67	-7.73	-8.51	-7.59	-7.62	-7.54	-7.51	-8.01	-6.68
7	-8.34	-8.18	-8.91	-7.88	-7.88	-7.86	-7.83	-8.60	-6.80
8	-5.09	-5.03	-5.39	-5.05	-5.06	-5.01	-4.99	-5.13	-4.48
9	-3.11	-3.06	-3.77	-2.92	-2.92	-2.93	-2.91	-3.17	-1.95
10	-4.22	-4.29	-5.15	-4.01	-4.01	-3.99	-3.96	-4.68	-2.69
11	-5.48	-5.53	-6.75	-5.04	-5.02	-5.05	-5.02	-6.17	-3.05
12	-7.40	-7.52	-8.08	-7.55	-7.60	-7.47	-7.45	-7.90	-6.83
13	-6.28	-6.32	-7.40	-6.14	-6.14	-6.14	-6.12	-6.55	-4.48
14	-7.56	-7.68	-9.12	-7.55	-7.57	-7.55	-7.51	-8.02	-5.85
15	-8.72	-8.67	-10.30	-8.38	-8.37	-8.41	-8.38	-9.16	-6.27
16	-5.20	-5.15	-5.89	-5.25	-5.25	-5.24	-5.23	-5.38	-4.30
17	-17.45	-17.17	-18.65	-16.59	-16.55	-16.58	-16.54	-17.14	-15.74
18	-6.98	-7.07	-7.68	-7.16	-7.22	-7.10	-7.07	-6.90	-6.52
19	-7.51	-7.68	-8.52	-7.50	-7.55	-7.45	-7.43	-7.31	-6.48
20	-19.42	-19.00	-19.41	-18.55	-18.55	-18.45	-18.43	-19.81	-18.22
21	-16.53	-16.12	-16.78	-15.46	-15.42	-15.40	-15.37	-16.44	-14.93
22	-19.78	-19.40	-20.26	-18.86	-18.83	-18.80	-18.78	-19.77	-18.31
23	-19.47	-19.10	-20.08	-18.42	-18.36	-18.40	-18.37	-19.35	-17.82

[1] Extrapolated to the complete basis set limit with counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.2: Energetics for the S66 Dispersion Subset (kcal mol^{-1})

System	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
24	-2.72	-4.70	-6.52	-3.62	-3.64	-3.70	-3.66	-3.44	0.11
25	-3.80	-6.01	-6.70	-3.86	-3.87	-3.94	-3.90	-4.06	-0.41
26	-9.75	-11.14	-15.71	-9.56	-9.50	-9.75	-9.67	-11.32	-1.88
27	-3.34	-5.43	-6.91	-4.06	-4.08	-4.14	-4.10	-3.92	-0.28
28	-5.59	-7.54	-11.75	-5.97	-5.96	-6.16	-6.08	-7.03	1.25
29	-6.70	-8.63	-12.52	-6.91	-6.89	-7.10	-7.02	-7.78	0.10
30	-1.36	-2.33	-3.55	-0.91	-0.90	-1.01	-0.98	-2.38	1.44
31	-3.33	-4.01	-5.82	-3.07	-3.05	-3.17	-3.13	-4.30	0.20
32	-3.69	-4.41	-5.75	-3.25	-3.21	-3.35	-3.31	-4.56	-0.40
33	-1.81	-2.83	-4.09	-1.42	-1.41	-1.51	-1.47	-2.71	1.17
34	-3.76	-3.97	-6.96	-3.34	-3.35	-3.42	-3.36	-5.31	0.67
35	-2.60	-2.68	-5.21	-2.63	-2.67	-2.69	-2.65	-3.38	0.34
36	-1.76	-1.74	-3.99	-2.07	-2.11	-2.13	-2.09	-2.17	0.13
37	-2.40	-2.49	-4.92	-2.44	-2.48	-2.50	-2.46	-3.14	0.33
38	-2.99	-3.14	-5.64	-2.73	-2.76	-2.80	-2.75	-3.57	0.39
39	-3.51	-4.58	-7.88	-3.98	-4.00	-4.09	-4.04	-4.70	0.78
40	-2.85	-3.60	-6.57	-3.46	-3.48	-3.55	-3.51	-3.63	0.52
41	-4.81	-5.44	-9.23	-4.74	-4.76	-4.86	-4.79	-6.39	0.94
42	-4.09	-4.70	-8.26	-4.22	-4.24	-4.32	-4.26	-4.97	1.08
43	-3.69	-4.05	-7.15	-3.90	-3.93	-3.98	-3.93	-4.54	0.37
44	-1.99	-2.15	-3.40	-1.51	-1.50	-1.56	-1.53	-2.63	0.54
45	-1.72	-2.10	-3.19	-1.42	-1.42	-1.49	-1.46	-2.29	0.56
46	-4.26	-4.51	-7.52	-4.01	-4.02	-4.09	-4.03	-5.91	0.28

[1] Extrapolated to the complete basis set limit with counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.3: Energetics for the S66 Mixed Interaction Subset (kcal mol^{-1})

System	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
47	-2.83	-3.75	-7.56	-2.73	-2.71	-2.92	-2.87	-4.23	1.87
48	-3.51	-4.39	-8.78	-3.79	-3.78	-3.98	-3.92	-5.08	1.21
49	-3.29	-4.18	-8.29	-3.36	-3.35	-3.55	-3.49	-4.80	1.49
50	-2.86	-3.46	-5.61	-3.87	-3.87	-3.93	-3.90	-3.54	-0.95
51	-1.54	-1.66	-2.35	-1.74	-1.74	-1.77	-1.76	-1.66	-1.03
52	-4.73	-5.25	-7.14	-4.17	-4.15	-4.27	-4.23	-4.76	-0.01
53	-4.41	-4.72	-6.31	-4.32	-4.30	-4.38	-4.35	-4.87	-1.82
54	-3.29	-3.57	-4.73	-3.58	-3.57	-3.60	-3.57	-3.93	-1.43
55	-4.17	-4.76	-6.68	-4.51	-4.51	-4.55	-4.52	-4.94	-1.11
56	-3.20	-3.84	-5.86	-3.53	-3.52	-3.58	-3.55	-3.99	-0.12
57	-5.26	-6.20	-9.30	-5.91	-5.92	-6.00	-5.95	-6.37	-1.12
58	-4.24	-4.37	-5.81	-4.15	-4.17	-4.19	-4.16	-4.18	-2.54
59	-2.93	-2.87	-3.52	-3.14	-3.12	-3.16	-3.16	-3.24	-2.79
60	-4.97	-5.03	-5.42	-4.41	-4.39	-4.43	-4.41	-5.42	-3.57
61	-2.91	-3.03	-5.30	-2.80	-2.81	-2.87	-2.83	-3.82	0.33
62	-3.53	-3.66	-5.81	-3.01	-3.00	-3.08	-3.03	-4.50	0.24
63	-3.75	-4.56	-7.20	-5.07	-5.07	-5.10	-5.07	-5.32	-1.66
64	-3.00	-3.17	-4.42	-2.59	-2.57	-2.65	-2.62	-3.57	-0.55
65	-4.10	-4.21	-5.33	-4.40	-4.40	-4.43	-4.42	-4.28	-3.87
66	-3.97	-4.55	-6.00	-3.84	-3.84	-3.87	-3.83	-4.54	-1.16

[1] Extrapolated to the complete basis set limit with counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.4: Energetics for the S22 Dataset (kcal mol^{-1})

System	Type	CCSD(T)[1]	MP2[2]	MP2[3]	**I**	**II**	**III**	**IV**	M06-2X[3]	B3LYP[3]
1	HB	-3.13	-3.20	-3.37	-2.91	-2.91	-2.89	-2.86	-3.43	-2.37
2	HB	-4.99	-5.03	-5.21	-5.03	-5.05	-4.97	-4.92	-5.20	-4.64
3	HB	-18.75	-18.60	-18.56	-17.90	-17.88	-17.80	-17.62	-19.39	-17.75
4	HB	-16.06	-15.86	-16.16	-15.01	-14.96	-14.96	-14.81	-16.22	-14.65
5	HB	-20.64	-20.61	-21.72	-19.68	-19.63	-19.68	-19.48	-20.23	-18.82
6	HB	-16.93	-17.37	-18.96	-16.32	-16.35	-16.27	-16.11	-16.59	-14.81
7	HB	-16.66	-16.54	-18.38	-15.51	-15.60	-15.43	-15.28	-16.06	-13.87
8	D	-0.53	-0.51	-0.92	-0.48	-0.48	-0.50	-0.50	-0.85	0.06
9	D	-1.47	-1.62	-2.10	-1.01	-0.99	-1.04	-1.03	-2.00	0.06
10	D	-1.45	-1.86	-3.28	-1.84	-1.84	-1.87	-1.86	-1.79	0.40
11	D	-2.65	-4.95	-8.11	-2.73	-2.71	-2.91	-2.88	-4.04	2.82
12	D	-4.26	-6.90	-9.87	-4.51	-4.52	-4.66	-4.61	-5.02	1.67
13	MX	-9.81	-11.39	-15.57	-9.53	-9.47	-9.73	-9.63	-11.23	-2.09
14	D	-4.52	-8.12	-12.83	-4.88	-4.86	-5.13	-5.08	-6.01	3.63
15	MX	-11.73	-14.93	-21.59	-12.41	-12.33	-12.71	-12.58	-13.72	-0.29
16	MX	-1.50	-1.69	-2.53	-1.86	-1.86	-1.89	-1.87	-1.73	-1.04
17	MX	-3.28	-3.61	-4.67	-3.55	-3.54	-3.58	-3.54	-3.86	-1.51
18	MX	-2.31	-2.72	-3.97	-2.65	-2.64	-2.68	-2.66	-2.77	-0.49
19	MX	-4.54	-5.16	-6.94	-5.28	-5.26	-5.35	-5.29	-5.29	-2.39
20	D	-2.72	-3.62	-6.49	-3.65	-3.68	-3.72	-3.68	-3.22	0.26
21	MX	-5.63	-7.03	-10.37	-6.48	-6.51	-6.57	-6.50	-6.31	-1.55
22	MX	-7.10	-7.76	-10.07	-7.29	-7.31	-7.31	-7.23	-7.32	-3.64

[1] Extrapolated to the complete basis set limit with counterpoise correction, from Marshall et al[161]

[2] Extrapolated to the complete basis set limit with counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[3] Computed using aug-cc-pVDZ without counterpoise correction

Table A.5: Energetics for phenylalanine-glycine-glycine conformers of P76 database(kcal mol^{-1})

Label	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
fgg114	-0.02	-0.75	-1.25	-0.10	-0.13	-0.09	-0.06	-0.79	1.57
fgg215	-0.76	-0.77	-0.30	-0.17	-0.24	-0.07	-0.05	-0.18	-0.85
fgg224	0.38	0.33	0.31	0.55	0.47	0.62	0.63	0.60	-0.03
fgg252	0.68	0.92	1.10	0.41	0.48	0.31	0.29	1.09	0.37
fgg300	1.07	1.93	1.60	-0.29	-0.21	-0.34	-0.38	0.11	-1.93
fgg357	-0.87	-1.57	-1.73	-0.65	-0.68	-0.61	-0.58	-1.17	0.50
fgg366	-0.53	0.15	1.29	-0.99	-0.92	-1.00	-1.05	0.06	-2.65
fgg380	0.72	0.74	0.95	0.70	0.60	0.81	0.82	0.87	-0.08
fgg412	0.31	0.04	-0.94	0.61	0.67	0.47	0.48	-0.46	2.61
fgg444	-1.36	-1.22	-0.51	-0.99	-1.08	-0.84	-0.83	-0.36	-2.35
fgg470	0.47	0.49	0.73	0.52	0.55	0.46	0.44	0.35	-0.15
fgg55	0.99	1.07	0.72	0.98	0.92	1.05	1.06	1.36	0.61
fgg691	0.31	0.81	1.87	0.32	0.38	0.27	0.24	1.10	-1.13
fgg80	0.66	0.16	-0.23	0.58	0.54	0.59	0.61	0.19	1.68
fgg99	-2.05	-2.32	-3.62	-1.46	-1.36	-1.62	-1.61	-2.77	1.83

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.6: Energetics for glycine-phenylalanine-alanine conformers of P76 database(kcal mol^{-1})

Label	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
gfa01	0.69	0.12	-0.19	0.33	0.39	0.16	0.15	0.57	1.44
gfa02	0.26	-0.06	-0.46	0.29	0.37	0.11	0.10	-0.02	1.18
gfa03	0.56	0.00	-0.34	0.20	0.26	0.02	0.01	0.35	1.29
gfa04	0.31	0.35	0.46	0.19	0.28	0.04	0.02	0.16	0.37
gfa05	0.38	0.44	0.53	0.26	0.35	0.11	0.09	0.08	0.38
gfa06	-0.02	0.50	1.59	0.05	-0.04	0.19	0.18	0.48	-2.35
gfa07	-0.57	-0.19	0.61	-0.46	-0.54	-0.34	-0.34	-0.11	-2.11
gfa08	0.02	0.31	1.12	0.50	0.44	0.63	0.64	0.29	-1.19
gfa09	-0.53	-0.98	-1.40	-0.43	-0.44	-0.43	-0.41	-0.72	0.77
gfa10	-0.62	-1.08	-1.50	-0.52	-0.53	-0.52	-0.50	-0.91	0.73
gfa11	-0.06	0.20	0.94	0.36	0.30	0.50	0.51	0.37	-1.28
gfa12	-0.31	-0.12	0.17	-0.45	-0.53	-0.32	-0.31	0.00	-1.57
gfa13	0.09	0.58	0.12	0.12	0.20	0.04	0.02	0.40	0.81
gfa14	-0.02	0.72	0.62	0.25	0.35	0.19	0.17	-0.17	0.62
gfa15	-0.87	-1.10	-1.77	-1.05	-1.11	-0.91	-0.88	-1.17	-0.35
gfa16	0.69	0.31	-0.52	0.35	0.27	0.52	0.56	0.39	1.24

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.7: Energetics for glycine-glycine-phenylalanine conformers of P76 database(kcal mol^{-1})

Label	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
ggf01	1.08	0.69	-0.14	0.09	0.07	0.14	0.15	0.30	1.06
ggf02	0.93	0.87	0.86	1.30	1.34	1.23	1.23	0.92	1.33
ggf03	0.75	0.73	1.70	0.68	0.74	0.57	0.54	0.56	-0.72
ggf04	0.65	0.73	0.31	0.35	0.34	0.32	0.32	0.09	0.74
ggf05	0.60	0.31	-0.32	0.88	0.95	0.80	0.81	-0.54	3.81
ggf06	0.58	0.60	0.43	0.63	0.57	0.71	0.72	1.06	0.61
ggf07	0.51	0.65	0.53	0.37	0.37	0.33	0.33	0.64	-0.45
ggf08	0.49	0.31	0.31	0.74	0.78	0.68	0.67	0.44	1.00
ggf09	0.30	0.17	0.30	0.67	0.72	0.59	0.59	0.16	1.21
ggf10	-0.11	-0.03	-0.01	-0.24	-0.24	-0.29	-0.29	0.20	-0.73
ggf11	-0.61	-0.54	0.20	-0.57	-0.60	-0.47	-0.48	-0.40	-1.98
ggf12	-0.78	-0.52	-0.88	-0.83	-0.75	-0.90	-0.91	-0.67	0.33
ggf13	-1.09	-1.04	-0.99	-1.02	-1.09	-0.91	-0.90	-0.71	-1.45
ggf14	-1.45	-1.46	-1.29	-1.30	-1.38	-1.17	-1.16	-0.75	-1.80
ggf15	-1.84	-1.46	-0.99	-1.74	-1.82	-1.62	-1.61	-1.29	-2.95

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.8: Energetics for tryptophan-glycine conformers of P76 database(kcal mol^{-1})

Label	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
wg01	-1.53	-1.03	0.44	-1.43	-1.51	-1.27	-1.29	-0.79	-3.90
wg02	-1.13	-1.06	-1.55	-1.32	-1.36	-1.27	-1.26	-0.66	-0.56
wg03	-0.63	-0.64	-0.94	-0.59	-0.63	-0.52	-0.50	-0.73	-0.16
wg04	-0.27	0.15	1.30	-0.43	-0.51	-0.27	-0.29	0.43	-3.01
wg05	-0.27	0.53	2.50	-0.26	-0.31	-0.10	-0.14	0.23	-3.91
wg06	-0.21	-0.12	-0.47	-0.28	-0.31	-0.23	-0.22	0.33	0.33
wg07	-0.01	-0.45	-0.61	0.42	0.47	0.31	0.32	0.05	1.34
wg08	0.53	0.67	0.85	0.13	0.06	0.24	0.24	0.55	-0.67
wg09	0.07	0.02	-0.64	-0.07	0.00	-0.22	-0.24	-0.45	0.77
wg10	-0.01	-0.36	-1.12	-0.02	0.05	-0.18	-0.18	-0.47	1.38
wg11	0.49	0.28	0.22	0.67	0.71	0.62	0.63	0.31	0.97
wg12	0.92	0.88	0.87	1.01	0.98	1.10	1.11	0.92	1.03
wg13	0.50	0.05	-0.45	0.72	0.79	0.59	0.60	-0.14	2.26
wg14	0.68	0.55	0.12	0.73	0.80	0.58	0.57	0.22	1.45
wg15	0.88	0.53	-0.53	0.72	0.76	0.63	0.64	0.19	2.68

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.9: Energetics for tryptophan-glycine-glycine conformers of P76 database(kcal mol^{-1})

Label	CCSD(T)[1]	MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
wgg01	-2.42	-1.85	0.08	-2.06	-2.09	-1.93	-1.95	-1.56	-5.42
wgg02	-2.16	-2.28	-1.69	-2.34	-2.35	-2.27	-2.28	-2.28	-3.12
wgg03	-1.33	-0.04	0.14	-0.26	-0.27	-0.26	-0.28	0.36	-1.46
wgg04	-0.33	-0.23	-0.29	-0.15	-0.15	-0.13	-0.12	-0.46	0.02
wgg05	-0.71	-0.82	-2.57	-0.77	-0.65	-0.96	-0.95	-1.66	2.73
wgg06	0.11	0.28	0.48	0.39	0.38	0.37	0.37	0.66	-0.83
wgg07	-0.05	-0.91	-2.01	-0.20	-0.21	-0.20	-0.17	-0.84	2.46
wgg08	0.54	0.85	1.17	0.65	0.63	0.64	0.62	1.22	-0.62
wgg09	0.36	0.53	0.57	-0.36	-0.37	-0.37	-0.38	0.32	-0.95
wgg10	0.94	1.41	2.80	0.76	0.72	0.85	0.83	1.29	-2.35
wgg11	0.92	0.76	0.77	0.68	0.77	0.53	0.51	0.65	0.80
wgg12	1.41	0.51	-0.53	1.50	1.49	1.50	1.53	0.67	4.59
wgg13	1.82	1.27	0.28	1.60	1.60	1.58	1.61	1.18	3.96
wgg14	-0.04	-0.91	-2.00	-0.19	-0.21	-0.20	-0.16	-0.83	2.48
wgg15	0.95	1.43	2.80	0.77	0.73	0.86	0.83	1.29	-2.28

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.10: Energetics for 27 reference alanine tetrapeptide conformers(kcal mol^{-1})

Label	RI-MP2[1]	MP2[2]	**I**	**II**	**III**	**IV**	M06-2X[2]	B3LYP[2]
1	0.40	2.79	0.50	0.52	0.55	0.51	0.40	-2.18
2	0.46	2.26	0.37	0.37	0.42	0.39	0.53	-1.93
3	-3.16	-4.00	-3.20	-3.22	-3.21	-3.20	-2.73	-2.70
4	2.00	3.36	1.74	1.72	1.75	1.72	2.12	-0.11
5	1.53	3.00	1.72	1.74	1.70	1.67	1.82	0.19
6	-0.84	-0.86	-0.67	-0.65	-0.71	-0.71	-1.02	-0.34
7	2.93	2.29	2.99	3.00	2.96	2.97	2.28	3.79
8	0.91	-0.08	0.80	0.82	0.77	0.78	0.61	2.47
9	4.19	4.29	3.85	3.81	3.96	3.96	4.07	2.26
10	4.06	3.65	4.12	4.13	4.17	4.18	3.69	4.73
11	-3.73	-4.87	-3.57	-3.55	-3.62	-3.60	-3.60	-1.91
12	-3.44	-4.67	-3.05	-3.05	-3.12	-3.10	-3.76	-0.74
13	-0.08	0.97	-0.31	-0.33	-0.26	-0.28	-0.03	-1.89
14	0.95	1.59	1.10	1.11	1.08	1.07	0.95	0.50
15	-1.54	-3.06	-1.49	-1.48	-1.53	-1.50	-1.90	1.45
16	-0.18	-0.70	-0.20	-0.21	-0.25	-0.24	0.25	0.28
17	-0.31	-1.54	-0.50	-0.49	-0.55	-0.53	-0.55	0.59
18	-1.82	-1.18	-2.11	-2.14	-2.03	-2.04	-1.88	-3.35
19	0.08	0.57	-0.21	-0.25	-0.12	-0.12	-0.05	-1.65
20	-1.98	-1.69	-2.02	-2.03	-2.00	-2.00	-1.89	-2.33
21	-0.81	-1.59	-1.14	-1.12	-1.20	-1.20	-1.05	-0.49
22	2.09	1.74	2.36	2.37	2.33	2.34	2.19	2.48
23	2.08	1.66	2.22	2.23	2.20	2.21	2.12	2.49
24	0.24	-0.02	0.42	0.42	0.40	0.40	0.47	1.14
25	-1.24	-1.05	-1.16	-1.17	-1.14	-1.15	-0.94	-1.10
26	-3.06	-3.30	-2.84	-2.84	-2.86	-2.85	-2.52	-1.77
27	0.29	0.41	0.28	0.27	0.31	0.32	0.44	0.11

[1] Computed at the aug-cc-pV(T→Q)Z level without counterpoise correction, from DiStasio et al[166]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.11: S22x5 geometries for Water Dimer(kcal mol^{-1})

Scaling	CCSD(T)[1]	MP2[2]	**I**	**II**	**III**	**IV**
90%	-4.32	-4.52	-4.33	-4.37	-4.23	-4.22
100%	-4.97	-5.21	-5.03	-5.05	-4.97	-4.96
120%	-4.04	-4.32	-4.16	-4.16	-4.16	-4.15
150%	-2.29	-2.47	-2.37	-2.36	-2.38	-2.38
200%	-0.96	-1.00	-0.97	-0.97	-0.97	-0.98

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.12: S22x5 geometries for Parallel-Displaced Benzene Dimer(kcal mol^{-1})

Scaling	CCSD(T)[1]	MP2[2]	**I**	**II**	**III**	**IV**
90%	-0.15	-7.91	-0.47	-0.51	-0.55	-0.42
100%	-2.81	-8.11	-2.73	-2.71	-2.91	-2.84
120%	-1.92	-4.49	-1.82	-1.82	-1.95	-1.93
150%	-0.53	-1.48	-0.61	-0.63	-0.62	-0.62
200%	-0.07	-0.27	-0.11	-0.11	-0.10	-0.10

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.13: S22x5 geometries for T-Shaped Benzene Dimer(kcal mol^{-1})

Scaling	CCSD(T)[1]	MP2[2]	**I**	**II**	**III**	**IV**
90%	-2.20	-6.72	-3.21	-3.26	-3.24	-3.18
100%	-2.80	-6.49	-3.65	-3.68	-3.72	-3.68
120%	-2.25	-4.60	-2.77	-2.78	-2.84	-2.82
150%	-1.12	-2.16	-1.25	-1.25	-1.28	-1.27
200%	-0.35	-0.73	-0.44	-0.45	-0.45	-0.44

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Table A.14: S22x5 geometries for Ammonia Dimer(kcal mol^{-1})

Scaling	CCSD(T)[1]	MP2[2]	**I**	**II**	**III**	**IV**
90%	-2.41	-2.57	-2.02	-2.03	-1.94	-1.92
100%	-3.14	-3.37	-2.91	-2.91	-2.89	-2.87
120%	-2.36	-2.57	-2.26	-2.25	-2.28	-2.27
150%	-1.11	-1.22	-1.08	-1.08	-1.09	-1.09
200%	-0.36	-0.39	-0.35	-0.35	-0.35	-0.35

[1] Extrapolated to the complete basis set limit without counterpoise correction, from the Benchmark Energy and Geometry DataBase(BEGDB.com)[2]

[2] Computed using aug-cc-pVDZ without counterpoise correction

Appendix B

Code for generating terf interpolation tables

The following is a python script for generating the interpolation tables required to form the primitive terf integrals. The resulting interpolation tables are provided with any copy of Q-Chem, but the interpolation tables are truncated to a finite maximum angular momentum, currently including 'h' functions. The inherent numerical noise of interpolation tables (here minimized using 256-bit floating point numbers) or the desire to do 5Z calculations may require the refinement or extension of these interpolation tables at some future point. For further information about the implementation, please consult the derivation of the terf primitives done by Dutoi and Head-Gordon[153].

```
#!/usr/bin/python
import os, sys
import math, sys, time
import pp
from math import *
from scipy import *
from numpy import *
from scipy.special import *
from gmpy import *
import numpy, gmpy, scipy, scipy.special

usage = "usage: %s S s interval" % os.path.basename(sys.argv[0])
print usage
print """
Needed files include
4 2 16
10 5 8
20 20 4
20 80 2
"""
if len(sys.argv)<3:
```

```
    sys.exit(0)

def gs1(x,i):
    tmp=gmpy.mpf(math.exp(-x),256)
    for j in range(i):
        tmp=tmp*gmpy.mpf(x,256)/gmpy.mpf((j+1),256)
    return tmp

def df(x):
    if x<=0.0:
        return gmpy.mpf(1.0,256)
    if x==1.0:
        return gmpy.mpf(.5,256)
    else:
        return (gmpy.mpf(x,256)/gmpy.mpf(x+1,256))*
            gmpy.mpf(df(x-2.0),256)
dimi=500
dimm=24
dimn=12
interval=1.000/int(sys.argv[3])
Sstart=0.00
Send=float(sys.argv[1])+interval
deltaS=interval
sstart=0.00
send=float(sys.argv[2])+interval
deltas=interval

Srange=numpy.arange(Sstart,Send,deltaS)
srange=numpy.arange(sstart,send,deltas)
print "Setup now running"
G=[[]]
for S in Srange:
    for s in srange:
        G[Srange.searchsorted(S)].append([])
    G.append([])
ppservers = ()
job_server = pp.Server(ppservers=ppservers)
print "Starting pp with", job_server.get_ncpus(), "workers"
start_time = time.time()


def dosrange(S,s,dimi,dimm,dimn):
```

```
        gS=[[],[]]
        for i in numpy.arange(dimi):
            tmp=gmpy.mpf(0,256)
            gS[1].append(gs1(S,i))
            for j in numpy.arange(i+1):
                tmp=tmp+gS[1][j]
            gS[0].append(tmp)
        for k in numpy.arange(2,dimm,1):
            gS.append([])
            for i in numpy.arange(dimi):
                if i>0:
                    gS[k].append(gS[k-1][i]-gS[k-1][i-1])
                else:
                    gS[k].append(gS[k-1][i])
        gs=[[],[]]
        for i in numpy.arange(dimi):
            tmp=gmpy.mpf(0,256)
            gs[1].append(gs1(s,i))
            for j in numpy.arange(i+1):
                tmp=tmp+gs[1][j]
            gs[0].append(tmp)
        for k in numpy.arange(2,dimn,1):
            gs.append([])
            for i in numpy.arange(dimi):
                if i>0:
                    gs[k].append(gs[k-1][i]-gs[k-1][i-1])
                else:
                    gs[k].append(gs[k-1][i])
        Gmn=[]
        for k in numpy.arange(dimm):
            for j in numpy.arange(dimn):
                tmp=gmpy.mpf(0,256)
                for i in range(dimi):
                    tmp2=df(gmpy.mpf(2,256)*gmpy.mpf(i,256))
#strictly, this would be gS[k][i+1],
#but TD wanted to generalize this
#for the hypergeometric function that was at the root
                    tmp3=gS[k][i]*gs[j][i]
                    tmp=tmp+tmp2*tmp3
                Gmn.append(tmp)
        return Gmn
```

```
print "Code executing"
jobs = [((S,s), job_server.submit(dosrange,(S,s,dimi,dimm,dimn),
         (df,gs1),("math","numpy","gmpy")))
         for s in tuple(srange) for S in tuple(Srange)]
for (S,s), job in jobs:
    print "S %f s %f" %(S,s)
    G[Srange.searchsorted(S)][srange.searchsorted(s)]=job()

print "Time elapsed: ", time.time() - start_time, "s"
job_server.print_stats()

output=open(sys.argv[3]+"_"+sys.argv[1]+"_"+sys.argv[2]+".txt", 'w')
size=dimm*dimn*((Send-Sstart)/deltaS)*((send-sstart)/deltas)
output.write('%d' %size)
for i in G:
    for j in i:
        for k in j:
            output.write('%+.18e' %k)
output.close()
```