Explicitly Correlated Methods for Large Molecular Systems

Fabijan Pavošević

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Edward F. Valeev, Chair
T. Daniel Crawford
Diego Troya
Louis A. Madsen

December 6, 2017
Blacksburg, Virginia

Keywords: Electronic Structure, Reduced Scaling, Explicit Correlation, Green’s Functions

Copyright 2017, Fabijan Pavošević
Explicitly Correlated Methods for Large Molecular Systems

Fabijan Pavošević

(ABSTRACT)

Wave function based electronic structure methods have become a robust and reliable tool for the prediction and interpretation of the results of chemical experiments. However, they suffer from very steep scaling behavior with respect to an increase in the size of the system as well as very slow convergence of the correlation energy with respect to the basis set size. Thus these methods are limited to small systems of up to a dozen atoms. The first of these issues can be efficiently resolved by exploiting the local nature of electron correlation effects while the second problem is alleviated by the use of explicitly correlated R12/F12 methods. Since R12/F12 methods are central to this work, we start by reviewing their modern formulation.

Next, we present the explicitly correlated second-order Møller-Plesset (MP2-F12) method in which all nontrivial post-mean-field steps are formulated with linear computational complexity in system size [Pavošević et al., J. Chem. Phys. 144, 144109 (2016)]. The two key ideas are the use of pair-natural orbitals for compact representation of wave function amplitudes and the use of domain approximation to impose the block sparsity. This development utilizes the concepts for sparse representation of tensors described in the context of the DLPNO-MP2 method by Neese, Valeev and co-workers [Pinski et al., J. Chem. Phys.
Novel developments reported here include the use of domains not only for the projected atomic orbitals, but also for the complementary auxiliary basis set (CABS) used to approximate the three- and four-electron integrals of the F12 theory, and a simplification of the standard B intermediate of the F12 theory that avoids computation of four-index two-electron integrals that involve two CABS indices. For quasi-1-dimensional systems (n-alkanes) the $\mathcal{O}(N)$ DLPNO-MP2-F12 method becomes less expensive than the conventional $\mathcal{O}(N^5)$ MP2-F12 for $n$ between 10 and 15, for double- and triple-zeta basis sets; for the largest alkane, $\text{C}_{200}\text{H}_{402}$, in def2-TZVP basis the observed computational complexity is $N^{1.6}$, largely due to the cubic cost of computing the mean-field operators. The method reproduces the canonical MP2-F12 energy with high precision: 99.9% of the canonical correlation energy is recovered with the default truncation parameters. Although its cost is significantly higher than that of DLPNO-MP2 method, the cost increase is compensated by the great reduction of the basis set error due to explicit correlation.

We extend this formalism to develop a linear-scaling coupled-cluster singles and doubles with perturbative inclusion of triples and explicitly correlated geminals [Pavošević et al., J. Chem. Phys. 146, 174108 (2017)]. Even for conservative truncation levels, the method rapidly reaches near-linear complexity in realistic basis sets; e.g., an effective scaling exponent of 1.49 was obtained for n-alkanes with up to 200 carbon atoms in a def2-TZVP basis set. The robustness of the method is benchmarked against the massively parallel implementation of the conventional explicitly correlated coupled-cluster for a 20-water cluster; the
total dissociation energy of the cluster (~186 kcal/mol) is affected by the reduced-scaling approximations by only ~0.4 kcal/mol. The reduced-scaling explicitly correlated CCSD(T) method is used to examine the binding energies of several systems in the L7 benchmark data set of noncovalent interactions.

Additionally, we discuss a massively parallel implementation of the Laplace transform perturbative triple correction (T) to the DF-CCSD energy within density fitting framework. This work is closely related to the work by Scuseria and co-workers [Constans et al., *J. Chem. Phys.* **113**, 10451 (2000)]. The accuracy of quadrature with respect to the number of quadrature points has been investigated on systems of the 18-water cluster, uracil dimer and pentacene dimer. In the case of the 18-water cluster, the $\mu E_h$ accuracy is achieved with only 3 quadrature points. For the uracil dimer and pentacene dimer, 6 or more quadrature points are required to achieve $\mu E_h$ accuracy; however, binding energy of <1 kcal/mol is obtained with 4 quadrature points. We observe an excellent strong scaling behavior on distributed-memory commodity cluster for the 18-water cluster. Furthermore, the Laplace transform formulation of (T) performs faster than the canonical (T) in the case of studied systems. The efficiency of the method has been furthermore tested on a DNA base-pair, a system with more than one thousand basis functions.

Lastly, we discuss an explicitly correlated formalism for the second-order single-particle Green’s function method (GF2-F12) that does not assume the popular diagonal approxi-
mation, and describes the energy dependence of the explicitly correlated terms [Pavošević et al., J. Chem. Phys. 147, 121101 (2017)]. For small and medium organic molecules the basis set errors of ionization potentials of GF2-F12 are radically improved relative to GF2: the performance of GF2-F12/aug-cc-pVDZ is better than that of GF2/aug-cc-pVQZ, at a significantly lower cost.

This work was supported in part by the U.S. National Science Foundation and the Graduate School Doctoral Assistantship Award in Chemistry.
Explicitly Correlated Methods for Large Molecular Systems

Fabijan Pavošević

(GENERAL ABSTRACT)

Chemistry has traditionally been considered an experimental science; however, since the
dawn of quantum mechanics, scientists have investigated the possibility of predicting the
outcomes of chemical experiments via the use of mathematical models. All molecular prop-
erties are encoded in the motion of the electrons, which can be quantitatively described by
the many-body Schrödinger equation. However, the Schrödinger equation is too complicated
to be solved exactly for realistic molecular systems, and so we must rely on approximations.
The most popular way to solve the Schrödinger equation when high accuracy is required
are the coupled-cluster (CC) family of methods. These methods can provide unsurpassed
accuracy; one particularly accurate and popular method is the coupled-cluster singles and
doubles with perturbative inclusion of triples (CCSD(T)) method. The CCSD(T) method is
known as the “gold standard” of quantum chemistry, and, when combined with a high qual-
ity basis set, it gives highly accurate predictions (that is, close to the experimental results)
for a variety of chemical properties. However, this method has a very steep scaling behavior
with a computational cost of $N^7$, where $N$ is the measure of the system size. This means
that if we double the size of the system, the computation time will increase by roughly two
orders of magnitude. Another problem is that this method shows very slow convergence to
the complete basis set (CBS) limit. Thus, in order to reduce the basis set error caused by
the incompleteness of the basis set, more than 100 basis functions per atom should be used, limiting this method to small systems of up to a dozen atoms. These two issues can be efficiently resolved by exploiting the local nature of electron correlation effects (reduced-scaling techniques) and by using explicitly correlated R12/F12 methods. The main focus of this thesis is to bridge the gap between reduced-scaling techniques and the explicit correlation formalism and to allow highly accurate calculations on large molecular systems with several hundred of atoms. As our first contribution to this field, we present a linear-scaling formulation of the explicitly correlated second-order Møller-Plesset method (MP2-F12) [Pavošević et al., *J. Chem. Phys.* **144**, 144109 (2016)]. This is achieved by the use of pair-natural orbitals (PNOs) for the compact representation of the unoccupied space. The method shows near-linear scaling behavior on the linear alkane chains with a computational scaling of $N^{1.6}$ for the largest alkane, $C_{200}H_{402}$, recovering more than 99.9% of correlation energy. The MP2-F12 method is intrinsically inadequate if high accuracy is required, but our formulation of the linear-scaling MP2-F12 method lays a solid foundation for the accurate linear-scaling explicitly correlated coupled-cluster singles and doubles method with perturbative inclusion of triples (PNO-CCSD(T)-F12) [Pavošević et al., *J. Chem. Phys.* **146**, 174108 (2017)]. We have demonstrated that the PNO-CCSD(T)-F12 method shows a near-linear scaling behavior of $N^{1.5}$. The error introduced by reduce-scaling approximations is only 0.4 kcal/mol of the binding energy with respect to the canonical result in the case of a 20-water cluster which is much lower than the required chemical accuracy defined as 1 kcal/mol. Furthermore, the reduced-scaling explicitly correlated CCSD(T) method is used to examine the binding ener-
gies of large molecular systems that are far beyond the reach of the conventional CCSD(T) method. Our prediction of the binding energy for of the coronene dimer is the most accurate theoretical estimate of binding energy of the coronene dimer to this date. Such a system is an example of an organic semiconductor used for light conversion. However, the modeling of light harvesting materials requires an accurate knowledge of ionization potentials (IP) and electron affinities (EA). We describe [Pavošević et al., *J. Chem. Phys.* **147**, 121101 (2017)] how to incorporate an explicit correlation correction into the Green’s function formalism (GF2) that is used for the calculation of IPs. We show that the GF2-F12 method removes errors associated with the basis sets, allowing extremely accurate predictions of IPs to be made at a significantly lower cost than the parent GF2 method. The work presented in this thesis will set a stage for further developments in reduced-scaling explicitly correlated methods. Furthermore it will be a useful benchmarking method for parametrizing the popular DFT functionals making accurate predictions of the relative stability of different forms of pharmaceuticals. Due to the simplicity and generality of the GF2-F12 method, it has the potential to be used to augment more accurate Green’s function methods, such as NR2, allowing for the accurate prediction of IPs and EAs of large molecular and periodic systems.
Acknowledgments

There are several people I would like to thank for making graduate school at Virginia Tech the best time of my life:

- Ljiljana and Miodrag Pavošević for their unconditional support
- Ema, Ivana and Petar
- My Serbian, Croatian and Bosnian friends, Miloš, Denis and Sandro for helping me feel less homesick
- Andrey for teaching about new standards of human behavior
- My friends in the theoretical chemistry groups: Drew Lewis, Chong Peng, Xiao Wang, Varun Rishi, Florian Bischoff, Harley McAlexander, Roberto Di Remigio, Kirk Pearce, Mathew Christian, Ashutosh Kumar, Rob Chaplesky, Andrew James, Marjory Clement, Nakul Teke and Karl Pierce
- once again to Marjory Clement for her hard work on proofreading this document
• My committee members: Prof. Daniel Crawford, Prof. Louis Madsen, Prof. Diego Troya and Prof. John Morris

• Prof. Nick Mayhall for useful discussions

• Prof. Frank Neese for being an excellent collaborator

• Prof. Edward Valeev for taking me on as a graduate student and working very hard to teach me electronic structure theory, for mentoring me and for never giving up on me. This work would not be possible without him

• My wife, Helena, for always being by my side and for being awesome

• My daughter, Nora, for giving a new meaning to our lives
## Contents

1 Introduction 1

2 Many-Body Electronic Structure Methods 7

2.1 Hartree-Fock Approximation 8

2.2 Electron Correlation 11

2.2.1 Configuration Interaction Methods 12

2.2.2 Coupled-Cluster Methods 15

2.2.3 Many-Body Perturbation Theory 17

2.3 Explicitly Correlated R12/F12 Methods 19

2.3.1 Explicitly Correlated Second-Order Møller-Plesset Methods 21

2.3.2 Approximations to Many-Electron Integrals 24

2.3.3 CABS Singles Correction 28
### 2.3.4 Performance of MP2-F12 method

Page 29

### 2.3.5 Explicitly Correlated Coupled-Cluster Methods

Page 31

### 2.3.6 Approximate Explicitly Correlated Coupled-Cluster Methods

Page 33

### 2.4 Summary

Page 40

### 3 SparseMaps – A systematic infrastructure for reduced-scaling electronic structure methods. IV. Linear-scaling second-order explicitly correlated energy with pair natural orbitals.

Page 42

#### 3.1 Introduction

Page 44

#### 3.2 Formalism

Page 46

- 3.2.1 Domain approximations

Page 54

- 3.2.2 Pair screening

Page 58

- 3.2.3 Miscellany: PNO construction, pair coupling screening, integral transformation

Page 64

#### 3.3 Computational details

Page 65

#### 3.4 Results

Page 67

- 3.4.1 Selection of truncation parameters

Page 67

- 3.4.2 Accuracy of relative errors

Page 75
3.4.3 Dependence of computational cost on system size  

3.5 Summary and Perspective  

4 SparseMaps – A systematic infrastructure for reduced-scaling electronic structure methods. V. Linear-scaling explicitly correlated coupled-cluster method with pair natural orbitals.  

4.1 Introduction  

4.2 Formalism  

4.2.1 Sparsifying Approximations  

4.2.2 Efficient Evaluation of The Coupled-Cluster-Specific F12 Intermediates  

4.3 Computational details  

4.4 Results  

4.4.1 Convergence of DLPNO correlation energy with the model truncation parameters.  

4.4.2 Computational cost of DLPNO-CCSD(T)-F12  

4.4.3 Application of DLPNO-CCSD(T)-F12 to energetics of noncovalent interactions.  

4.5 Summary and Perspective  

xiii
5 Massively Parallel Laplace Transform Perturbative Triples Correction (T) to the Coupled-Cluster Singles and Doubles Method using Density Fitting (DF)  

5.1 Introduction ................................................................. 122  
5.2 Formalism ................................................................. 124  
  5.2.1 Canonical perturbative triples correction (T) .................... 124  
  5.2.2 Laplace transform .................................................. 126  
5.3 Computational details .................................................. 130  
5.4 Results ................................................................. 131  
  5.4.1 Quadrature accuracy .............................................. 131  
  5.4.2 Computational performance ..................................... 136  
5.5 Conclusions ............................................................ 142  

6 Explicitly correlated formalism for second-order single-particle Green’s function  

6.1 Introduction ................................................................. 145  
6.2 Formalism ................................................................. 146  
6.3 Results and discussion ................................................ 150
6.4 Summary and Perspective ................................. 154

7 Conclusions ................................................. 156

Publication List .............................................. 160

Bibliography .................................................. 161
List of Figures

2.1 Helium ground state wave functions with both electrons confined at the circle of radius 1\,\textit{a}_0. The wave function is of the form $\Psi_{n,l,m} = r_1^{n_l} r_2^{l_m} \exp(-1.8r_1/2 - 1.8r_2/2)$ where total angular momentum $N$ is obtained as $N = n + l + m$. The CI wave function is obtained when $m$ is an even number. The exact wave function includes all values for $m$ that satisfy the total angular momentum up to $N = 3$.

2.2 Orbital spaces

3.1 Distance decay of MP2-F12 pair energies (top) and the difference between MP2-F12 and MP2 pair energies (bottom) in n-C$_{30}$H$_{62}$ with linear geometry. The distance between orbitals $i$ and $j$ is defined as the distance between their centroids. The def2-SVP orbital basis set was used.
3.2 Distance decay of MP2-F12 pair energies (top) and the difference between MP2-F12 and MP2 pair energies (bottom) in polyene C_{30}H_{32} with linear geometry. The distance between orbitals $i$ and $j$ is defined as the distance between their centroids. The def2-SVP orbital basis set was used.

3.3 The F12 contribution to def2-SVP MP2-F12 pair energy vs. doi$(i,j)$ for linear n-C_{30}H_{62}.

3.4 The F12 contribution to def2-SVP MP2-F12 pair energy vs. doi$(i,j)$ for linear polyene sC_{30}H_{32}.

3.5 Convergence of the DLPNO-MP2 correlation energy and F12 part of the DLPNO-MP2-F12 energy, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutPNO}}$. See the text for more details.

3.6 Convergence of the DLPNO-MP2 and DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutPNO}}$. See the text for more details.

3.7 Convergence of the DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutPNO}}$. “(-SC)” indicates the omission of the semicanonical PNO truncation correction.

3.8 Convergence of the DLPNO-MP2 and DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutMKN}}$. See the text for more details.
3.9 Convergence of the DLPNO-MP2 and DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutDO}}$. $T_{\text{CutPNO}}$ and $T_{\text{CutMKN}}$ were set to $10^{-8}$ and $10^{-3}$, respectively, with the rest of truncation parameters set to 0.  

3.10 The average size of PAO/CABS atomic domains as a function of parameter $T_{\text{CutDO}}$ for the sildenafil and anthracene dimer. The total number of atoms in the sildenafil and anthracene dimer are 63 and 48, respectively. 

3.11 Wall clock time for the local DLPNO-MP2-F12 and canonical RI-MP2-F12 with the def2-SVP and def2-TZVP basis sets. The shown timings are without Hartree-Fock. 

4.1 DLPNO-CCSD and DLPNO-CCSD-F12 valence correlation energies of (H$_2$O)$_{20}$, as percentage of the respective canonical energies, as a function of PNO truncation parameter $T_{\text{CutPNO}}$. All other truncation parameters have been set to 0. 

4.2 DLPNO-CCSD and DLPNO-CCSD-F12 valence correlation energies of (H$_2$O)$_{20}$ as a function of the domain truncation parameter $T_{\text{CutDO}}$. The energies are defined as percentages of the respective DLPNO energies with $T_{\text{CutPNO}} = 10^{-7}$, $T_{\text{CutPairs}} = 10^{-5}$, and other truncation parameters set to 0.
4.3 Wall clock time (in hours) for the canonical RI-CCSD(T)-F12 and local DLPNO-CCSD(T)-F12 approach with the def2-TZVP basis set and TightPNO truncation threshold.

Timings does not include Hartree-Fock computation time. .......................... 114

5.1 Wall clock time (minutes) for (T) and LT-(T) with respect to the number of computing nodes for the 18-water cluster. LT-(T) timings are with 3 quadrature points. ................................................................. 137

5.2 Computational speedup for (T) and LT-(T) with respect to the number of computing nodes. LT-(T) are with 3 quadrature points. ................................. 138

6.1 Mean absolute error (MAE) in eV for GF2 (blue), GF2(F12) (green) and GF2-F12 (red) methods with several basis sets. Values next to the points represent maximum absolute errors (MaxAE) in eV. ................................. 153
List of Tables

2.1 Binding energies (kcal/mol) for the benzene dimer with the MP2 and MP2-F12 methods employing the cc-pVXZ-F12 and aug-cc-pVXZ (X=D,T,Q) basis sets. .......................................................... 30

2.2 Performance of various approximated CCSD-F12 methods for the calculation of reaction and atomization energies (kJ/mol). .......................................................... 35

3.1 Components of the correlation energy, expressed as percentage of the canonical values, computed with the DLPNO-MP2 and DLPNO-MP2-F12 methods for sildenafil, ATP anion and anthracene dimer. .................................................. 77

3.2 Influence of truncation parameters of the DLPNO-MP2-F12 method on the error statistics for the S66 benchmark set. The errors (kcal/mol) are relative to the conventional MP2-F12 energies. $\bar{\Delta}$ = mean error, $\Delta_{\text{std}}$ = standard deviation, $\Delta_{\text{max}}$ = maximum absolute error. The first column shows the differences relative to the default parameter values. .................................................. 78
3.3 The basis set error of the binding energy of the DLPNO-MP2-F12 for a subset of the L7 data set relative to the error due to the F12 pair screening procedure. The basis set error (kcal/mol) is defined relative to the complete basis set MP2 estimate. The F12 pair screening error (kcal/mol) is relative to the binding energy obtained without omitting any of the pairs from the F12 treatment.

3.4 Detailed breakdown of wall times (in seconds) for various steps of the DLPNO-MP2-F12 energy evaluation.

4.1 The difference between DLPNO and canonical CCSD-F12 dissociation energies of \((\text{H}_2\text{O})_{20}\), in kcal/mol as a function of the DLPNO truncation parameters. Each truncation parameter is set to zero unless otherwise stated.

4.2 Timings (in seconds) for the computation of different contributions in DLPNO-CCSD(T)-F12 method using TightPNO truncation setup.

4.3 Dissociation energy estimates (kcal/mol) for the L7/5 data set.

4.4 Breakdown of the correlation energy contributions to the DLPNO-CCSD(T)-F12 dissociation energy of coronene dimer (kcal/mol). The acronyms wp and sp stand for weak and strong pairs, respectively.

5.1 Accuracy of the Laplace transform \((T)\), \((\text{LT}-(T))\), with respect to the number of quadrature points \(n_q\) for the selected chemical systems.
5.2 Perturbative triple correction contribution (T) to the binding energy (kJ/mol) of the uracil dimer for several basis sets. ........................................ 136

5.3 Wall clock times (in minutes) of different DF-CCSD(T) contributions with respect to the number of computing nodes for 18-water cluster. LT-(T) timings are with 3 quadrature points. ........................................ 139

5.4 Wall clock times (in minutes) for CCSD, triple correction (T) and Laplace transform triple correction LT-(T) for several representative molecules. ... 140

5.5 Wall clock timings (in hours) for computation of GC-dDMP-B with the 6-311++G** basis set. ...................................................... 141

6.1 Mean absolute error (MAE) in eV and maximum absolute error (MaxAE) in eV for the IPs of the data set of 21 small molecules with respect to the CBS limit for several basis sets. ................................. 152

6.2 Mean absolute error (MAE) in eV and maximum absolute error (MaxAE) in eV for the IPs of the OAM24 data set with respect to the CBS limit for several basis sets. ........................................ 154
Chapter 1

Introduction

Computational quantum chemistry has become a routine tool for the prediction of molecular processes and properties, and it is an invaluable complement to chemical experiment. It is used to predict various chemical properties like molecular reactivity and spectra, and it can give insight into molecular structure. Most of the molecular properties come from electronic structure of a molecule and the interaction of its electrons with external fields or electromagnetic radiation. Such interactions are described by the many-body Schrödinger equation. However, the complexity of this equation prevents us from solving it exactly. Instead, we must rely on approximations, which introduce errors. There are two main theoretical approaches in use: density functional theory (DFT) and wave function based methods. Due to its low computational cost, density functional theory is extremely popular for modeling large molecular systems on the order of thousand atoms. However, there is not an obvious way how it can be systematically improved and it does not converge to the basis set limit.
Another problem is that it cannot properly describe chemically important phenomena such as non-covalent interactions or it poorly describes electronic response properties.\textsuperscript{1–4} Alternative to DFT are wave function based methods that can be systematically improved. The most popular of such methods that are routinely applied are the coupled-cluster (CC) family of methods. These methods can provide unsurpassed accuracy; one particularly accurate and popular method is the coupled-cluster singles and doubles with perturbative inclusion of triples (CCSD(T)) method, which is colloquially known as the “gold standard” of quantum chemistry.\textsuperscript{5,6} If CCSD(T) is applied to a chemical process of interest that is well described by a single Slater determinant, such as reaction barriers or non-covalent interactions, it will predict the desired quantity to within chemical accuracy, which is defined as 1 kcal/mol.\textsuperscript{7} However, this method has a very steep scaling behavior, with the computational cost growing as $N^7$ where $N$ is the measure of the system size. This means that if we double the size of the system, the computation time will increase by roughly two orders of magnitude. Another problem is that wave function based methods show very slow convergence to the complete basis set limit (CBS limit). Thus, in order to reduce the basis set error caused by the incompleteness of the basis set, more than 100 basis functions per atom should be used, limiting this method to small systems of up to a dozen atoms. These two issues can be efficiently resolved by exploiting the local nature of electron correlation effects and by using explicitly correlated R12/F12 methods.

The methods for curing unfavorable scaling with the system size can be roughly divided into two groups. The first group of methods breaks the calculation of the whole system
into smaller fragments, each of which is more computationally manageable than the whole calculation. Using these smaller calculations, the desired property of the whole system can then be reconstructed. The most prominent methods that fall into this group are the divide-and-conquer, the divide-expand-consolidate, the incremental, and the fragment molecular orbital methods. Another approach is based on the truncation of the operators or wave functions. One possible choice are projected atomic orbitals (PAOs), proposed by Pulay and Säbø. In this scheme, each unoccupied (virtual) orbital is restricted to a small region around a localized occupied orbital. Werner and co-workers further refined the PAO methodology by implementing many single-reference wave function methods in a linear-scaling fashion. The drawback of their approach is that, in attempting to achieve high accuracy, the computation time of these methods have late crossover with respect to the canonical method. Another choice are pair natural orbitals (PNOs). As shown by Neese, they represent a more compact formulation of the unoccupied space, allowing for an implementation the coupled-cluster with singles and doubles (CCSD) with $N^5$ computational cost. Yang and Chan and co-workers proposed a similar approach to PNOs, named orbital specific virtuals (OSVs), that has been applied to the MP2 and CCSD(T) methods. Riplinger and Neese expressed PNOs in the basis of PAOs leading to the domain-based local PNO-CCSD(T) method (DLPNO-CCSD(T)). This method represents the first robust and accurate near-linear-scaling CCSD(T) method that is capable of recovering more than 99.9% of the correlation energy. Riplinger and Neese were able to perform a calculation on the small protein of “crambin” with 644 atoms on single workstation.
DLPNO methodology has been reimplemented within the SparseMap infrastructure developed by Valeev and Neese. It has enabled a friendly environment for the implementation of single- and multi-reference wave function methods in a linear-scaling fashion. Werner and co-workers recently developed a similar hierarchy of approximations allowing for the efficient parallel implementation of linear-scaling methods.

The main reason for the slow convergence of conventional wave function based methods is that the Slater determinant is written as a product of one-electron functions, failing to describe the exact wave function properly at short inter-electronic distances. For conventional methods, such as CCSD(T), the basis set error for correlation energy in atoms is proportional to $O\left(\left(L_{\text{max}} + 1\right)^{-3}\right)$, where $L_{\text{max}}$ is the maximum angular momentum of the basis set.

Explicitly correlated methods increase the rate of convergence to the basis set limit. This is achieved by explicitly including the dependence on the inter-electronic distance ($r_{ij}$) in the wave function expansion. Thus, one can use a smaller basis set to achieve the same accuracy as conventional methods employed with larger basis sets at a much lower cost.

As a rule of thumb, results of $(X+2)$-level quality can be achieved with an $X$-level quality basis set, where $X$ is the basis set cardinal number. For explicitly correlated methods, the basis set error for correlation energy in atoms is proportional to $O\left((L_{\text{max}} + 1)^{-7}\right)$.

The practical explicitly correlated methods (R12/F12) were first introduced by Kutzelnigg in 1985 and further refined by others.

The next logical step was to combine local correlation methods with the explicitly correlated wave function approach. The pioneers in this field were Werner and Adler. They
implemented the first local explicitly correlated coupled-cluster method within the PAO framework.\textsuperscript{40, 41} Werner and Krause performed a comprehensive investigation of different approximations of unoccupied space (PAO, PNO and OSV) within the F12 framework, concluding that the PNO approach is superior to the others.\textsuperscript{42} Tew and Hättig extensively studied the PNO approach within the MP2-F12, MP3-F12 and, more recently, CCSD-F12 and CCSD(T)-F12 method, achieving a scaling of $O(N^4)$.\textsuperscript{43–47} At the same time as Hättig, Valeev and co-workers presented the first production-level implementation of the LPNO-CCSD-F12 method.\textsuperscript{48} In the same work, an improved version of PNOs coined the geminal-spanning orbitals (GSOs) was also investigated; the use of GSOs was shown to be highly suitable for explicitly correlated calculations.\textsuperscript{48} Furthermore, Valeev and co-workers developed production-level, linear-scaling implementations of the DLPNO-MP2-F12 and DLPNO-CCSD(T)-F12 methods using the \textit{SparseMap} infrastructure.\textsuperscript{29, 49} Around the same time, Werner and co-workers developed efficient parallel linear-scaling PNO-LMP2-F12 and PNO-LCCSD(T)-F12 methods.\textsuperscript{50, 51}

This work is structured as follows: In Chapter 2 we review standard wave function based methods. In addition, we also review the modern framework of explicitly correlated methods. Chapter 3 focuses on development of the explicitly correlated second-order Møller-Plesset perturbation theory in a linear-scaling fashion. The implementation of the linear-scaling perturbative explicit correlation correction to the CCSD(T) method is covered in Chapter 4. This is followed by Chapter 5 wherein we present the Laplace transform perturbative triple
correction (T) to the CCSD method with density fitting approximation. Lastly, in Chapter 6 we discuss the explicitly correlated formalism for single-particle Green’s functions.
Chapter 2

Many-Body Electronic Structure Methods

The central problem in the field of electronic structure methods is the time-independent Schrödinger equation

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

(2.1)

and where \( \hat{H} \) is the Hamiltonian operator while \( \Psi \) is the wave function.\textsuperscript{52-54} This equation describes the motion and interaction of all electrons and nuclei in a molecule. However, the Schrödinger equation is too complicated to be solved exactly for systems with more than one electron and nucleus, and no analytic solution exists beyond that. Major simplification can be achieved by introducing the Born-Oppenheimer approximation.\textsuperscript{54,55} The main idea is to treat the motion of the electrons and the nuclei differently since electrons are moving much
faster due to a smaller mass. This yields a simplified Hamiltonian given by

\[ \hat{H}_{\text{elec}} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} \]  

(2.2)

for which the kinetic energy of the nuclei is zero, and the nuclear-nuclear repulsion contribution is constant. The first term in the electronic Hamiltonian is the electronic kinetic energy operator, where \( N \) represents the number of electrons. The second term is the electron-nuclear attraction operator, where \( M \) and \( Z_A \) are the number of nuclei and the charge number of nucleus \( A \), respectively. \( r_{iA} \) is the distance between the \( i \)-th electron and \( A \)-th nucleus. Finally, the last term is the electron-electron repulsion operator where \( r_{ij} \) is the distance between the \( i \)-th and \( j \)-th electron. The electronic Hamiltonian in the Born-Oppenheimer approximation has a simpler form than the full Hamiltonian, but it is still too complicated for practical purposes. Further approximations have to be included, the first of which is the Hartree-Fock method. \(^{54}\)

2.1 Hartree-Fock Approximation

The Hartree-Fock method\(^ {54}\) is a starting point for other wave function based methods, and it plays an important role in the study of electronic structure theory. In this method, the state of each electron is described by a one-electron wave function (orbital). The total wave function for all electrons is constructed as a Slater determinant\(^ {56}\) of the orbitals and is given
by

\[ \Psi_0(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \ldots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \ldots & \chi_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \ldots & \chi_k(x_N) \end{vmatrix} \] (2.3)

where \( \chi_i \) is the \( i \)-th spin orbital, and \( x_i \) is the tuple of three spatial coordinates and one spin coordinate of the electron \( i \). The optimal orbitals, obtained by minimizing the Rayleigh quotient

\[ E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \] (2.4)

subject to the constraint that the spin orbitals be orthonormal to each other, ensures that the resulting energy will be greater than or equal to the exact energy due to the variational principle. This leads us to a set of integro-differential Hartree-Fock equations

\[ \hat{F} \chi_i = \epsilon_i \chi_i \] (2.5)

where \( \hat{F} \) is the Fock operator, \( \chi_i \) is the \( i \)-th canonical spin orbital and \( \epsilon_i \) is the corresponding orbital energy. The Fock operator is defined as

\[ \hat{F}(1) = \hat{h}(1) + \hat{J}(1) - \hat{K}(1) \] (2.6)

where \( \hat{h}(1) \) is the sum of the one-electron kinetic energy operator and the electron-nuclear attraction operator

\[ \hat{h}(1) = -\frac{1}{2} \nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}}. \] (2.7)
The Coulomb operator $\hat{J}(1)$ acting on the spin orbital $i$ containing electron 1 is given by

$$
\hat{J}(1)\chi_i(1) = \sum_j \int \chi_j^*(2)\chi_j(2) \frac{1}{r_{12}} \chi_i(1) d\mathbf{x}_2
$$

(2.8)

and it can be realized as the total Coulomb potential at $\mathbf{r}_1$ due to the charge density caused by all other electrons. The exchange operator $\hat{K}$ acting on the $\chi_i(1)$ spin orbital is given by

$$
\hat{K}(1)\chi_i(1) = \sum_j \int \chi_j^*(2)\chi_i(2) \frac{1}{r_{12}} \chi_j(1) d\mathbf{x}_2.
$$

(2.9)

The effect of the $\hat{K}$ operator is to exchange the labels of electrons 1 and 2. This operator does not have a classical explanation like the previous two, and its origin is from the antisymmetric nature of the Slater determinant. The form of Eq. (2.5) is not practical for use, and in order to arrive at a computationally attractive expression, we expand spin-orbitals into a set of $M$ basis functions

$$
\chi_i(\mathbf{x}) = \sum_{\mu=1}^M C_{\mu i} \phi_\mu(\mathbf{x})
$$

(2.10)

where $\phi_\mu$ are usually atom-centered functions. Inserting $\chi_i(\mathbf{x})$ into 2.5 and projecting it with the $\phi_\nu$, we obtain the matrix equation for $C_{\mu i}$. In the matrix form, this equation is called the Hartree-Fock-Roothaan equation,\textsuperscript{57} and it is given by

$$
\mathbf{F C} = \mathbf{S C} \mathbf{\epsilon}
$$

(2.11)

where $\mathbf{C}$ is the $M \times N$ matrix of molecular orbital (MO) coefficients $C_{\mu i}$. 

\textsuperscript{57}
2.2 Electron Correlation

The Hartree-Fock method is a mean-field theory, which means that each electron feels an average potential due to all other electrons in the molecule. In other words, the probability of finding an electron at a given position is independent of the positions of the other electrons. Such behavior can be seen in Figure 2.1, which illustrates the shape of different wave functions.

![Figure 2.1: Helium ground state wave functions with both electrons confined at the circle of radius 1a₀. The wave function is of the form $\Psi_{n,l,m} = r_1^n r_2^l r_{12}^m \exp(-1.8r_1/2 - 1.8r_2/2)$ where total angular momentum $N$ is obtained as $N = n + l + m$. The CI wave function is obtained when $m$ is an even number. The exact wave function includes all values for $m$ that satisfy the total angular momentum up to $N = 3$.](image)
functions when two electrons are on the same circle with one fixed and the other moving along the circle by changing $\theta_{12}$. The value of the Hartree-Fock wave function as a function of $\theta_{12}$ is a straight line, meaning that the position of electron one is completely independent of the position of the other electron. However, in reality, electrons feel each other due to the electron-electron repulsion, and the motion of all electrons has to be correlated; therefore the value of the exact wave function as a function of $\theta_{12}$ is curved. Further explanation of the plot will be given in one of the following sections. Correlation energy is defined as the difference between the exact (nonrelativistic) energy and the complete basis set Hartree-Fock energy:

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}.$$  
(2.12)

Even though the correlation energy is less than one percent of the total electronic energy, it plays an important role in predicting chemical phenomena. In the following text, we will discuss the most common methods for describing electron correlation effects.

### 2.2.1 Configuration Interaction Methods

Conceptually the simplest of all correlation methods is the configuration interaction (CI) method. In this method, the exact wave function is expanded as a linear combination of all possible Slater determinants that can be constructed from $N$ electrons and $2K$ spin
orbitals $\chi_i$ as

\[ |\text{FCI}\rangle = |\Psi_0\rangle + \sum_{i<a} C_{ai}^a |\Psi_i^a\rangle + \sum_{i<j} \sum_{a<b} C_{ab}^{ij} |\Psi_{ij}^{ab}\rangle + \sum_{i<j<k} \sum_{a<b<c} C_{abc}^{ijk} |\Psi_{ijk}^{abc}\rangle + \ldots \] (2.13)

where $|\Psi_0\rangle$ is the Hartree-Fock Slater determinant or the reference determinant; $|\Psi_i^a\rangle$ is a Slater determinant in which one electron from occupied spin orbital $i$ is excited to unoccupied orbital $a$; the $|\Psi_{ij}^{ab}\rangle$ is a Slater determinant in which electrons from occupied spin orbitals $i$ and $j$ are excited to unoccupied spin orbitals $a$ and $b$ and so on. The indices used to denote molecular orbitals are shown in Figure 2.2. In this work, we denote all occupied orbitals with $m; n; \ldots$ The set of occupied orbitals is divided into active and inactive (frozen) occupied orbitals where active occupied orbitals are denoted with $i; j; k; l; \ldots$ Unoccupied (virtual) orbitals have been denoted with $a; b; c; d; \ldots$ Combining all occupied and unoccupied orbitals gives the orbital basis space (OBS) denoted with $p; q; r; s; \ldots$ Furthermore, the OBS is extended with the complementary auxiliary basis space (CABS) that appears in explicitly correlated methods and is denoted with $a'; b'; c'; d'; \ldots$ The CABS and unoccupied orbitals form the complete unoccupied space ($\alpha, \beta; \ldots$), and the OBS and CABS form the complete basis space (CBS) where indices are denoted with $p', q', r', s'; \ldots$ The expansion coefficients $C$, which are determined by the variational method, are those which minimize the CI energy.

Eq. (2.13) is the full CI (FCI) wave function, and, with a complete basis set, it is the exact solution of the Schrödinger equation within the Born-Oppenheimer approximation. However, the number of possible excited Slater determinants grows as $\binom{2K}{N}$, which means that the computational cost of the FCI method is factorial with respect to the growth of the system size.\textsuperscript{54} In order to make the CI method applicable to larger systems, we can
discard determinants that contain excitations above a specified level. By discarding all excited determinants except the reference (|Ψ₀⟩), singly (S) (|Ψᵢ⟩), and doubly (D) (|Ψᵢⱼ⟩) excited determinants, we arrive at the CISD method. In other words, the basis for CISD are the \{|Ψ₀⟩, |Ψᵢ⟩, |Ψᵢⱼ⟩\} determinants and it scales as \(O(N^6)\), where \(N\) is the measure of the system size. However, truncation of the full CI wave function introduces problems; unlike FCI, the CISD method is no longer size-consistent and size-extensive. For a size-extensive method, the correlation energy will asymptotically grow linearly as the number of particles grow. A size-consistent method is a method that gives the same result for the energy of two molecules \(A\) and \(B\) that are infinitely separated as it gives the sum of the individual molecular energies:

\[
E_A + E_B = E_{AB}(r_{AB} = \infty). \tag{2.14}
\]
This quantity is important for chemistry due to the common need to compute such energy
differences and, therefore, the applicability of the CISD method is very limited.

2.2.2 Coupled-Cluster Methods

The coupled-cluster method\textsuperscript{61,62} is one of the most successful methods used for the predictions
of atomic and molecular electronic structure when high accuracy is required. The method is
designed to be size-consistent and size-extensive even in truncated form, which is achieved
with an exponential form of the wave function

\[ |\Psi_{CC}\rangle = \exp(\hat{T}) |\Psi_0\rangle \]  

(2.15)

where

\[ \exp(\hat{T}) = 1 + \hat{T} + \frac{1}{2!} \hat{T}^2 + \frac{1}{3!} \hat{T}^3 + \ldots + \frac{1}{n!} \hat{T}^n + \ldots \]  

(2.16)

and \( \hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots \) is the cluster operator that is composed of excitation operators of
different order.\textsuperscript{5,6} The n-particle cluster operator is used to generate excited configurations
(determinants), and it is defined by:

\[ \hat{T} = \left( \frac{1}{n!} \right) \sum_{ij\ldots ab\ldots} t_{ab\ldots}^{ij\ldots} a_a^\dagger a_b \ldots a_j a_i = \left( \frac{1}{n!} \right) \sum_{ij\ldots ab\ldots} t_{ab\ldots}^{ij\ldots} a_a^\dagger a_b \ldots . \]  

(2.17)

The operators \( a_a^\dagger \) and \( a_i \) are standard fermion creation and annihilation operators. The
operator \( a_i \) annihilates an electron from the occupied orbital \( i \), and operator \( a_a^\dagger \) creates
an electron in the unoccupied orbital \( a \). For practical reasons, the CC energy and \( t_{ab\ldots}^{ij\ldots} \)
amplitudes are determined from a reformulated Schrödinger equation:

\[ e^{-\hat{T}}\hat{H}e^{\hat{T}}|\Psi_0\rangle = \hat{H}|\Psi_0\rangle = E|\Psi_0\rangle \tag{2.18} \]

where \( \hat{H} \) is the similarity transformed Hamiltonian that can be recast by Baker-Campbell-Hausdorff (BCH) expansion as:

\[ \hat{H} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[[\hat{H}, \hat{T}], \hat{T}]] + \frac{1}{3!}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}]] + ... \tag{2.19} \]

which truncates at the fourth order due to the nature of the Hamiltonian. The amplitude equations are obtained by left projection of Eq. (2.18) with the excited determinants \( \langle \Psi_{ab...}^{ij...}| \)

\[ \langle \Psi_{ab...}^{ij...}| \hat{H} |\Psi_0\rangle = 0, \tag{2.20} \]

while energy is obtained by left-projection with the reference determinant \( \langle \Psi_0| \)

\[ \langle \Psi_0| \hat{H} |\Psi_0\rangle = E. \tag{2.21} \]

Just like the FCI method, the full CC method is the exact solution to the Schrödinger equation. Likewise, the full CC method scales factorially with the system size, so in order to make this theory practical, we truncate the cluster operator to include only single and double excitations (\( \hat{T} \equiv \hat{T}_1 + \hat{T}_2 \)), giving the CCSD method.\(^5\) The formal scaling of the CCSD method is the same as the CISD method, which is \( \mathcal{O}(N^6) \). However, in order to achieve higher accuracy, inclusion of higher than double excitations is necessary.\(^{63,64}\) Full inclusion of triple excitations defines the CCSDT method with computational cost of \( \mathcal{O}(N^8) \), which is too expensive for practical use. Alternatively, triple excitations can be approximated perturbatively as in the CCSD(T) method,\(^{65-67}\) which has a more acceptable cost and accuracy.
ratio with computational scaling of $O(N^7)$. The CCSD(T) method is known as the “gold standard” of quantum chemistry, and, when combined with a high quality basis set, it gives highly accurate predictions for various chemical properties that are close to the experimental results.\(^7\)

### 2.2.3 Many-Body Perturbation Theory

Perturbation theory is a method that is used to find an approximate solution of a problem that does not have an analytic solution by starting from a related simpler problem that has an exact solution. Perturbation theory applied to quantum mechanics is known as Rayleigh-Schrödinger perturbation theory (RSPT). In general, we want to find a solution for

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

but this task is too complex. In the RSPT, we split the Hamiltonian into two parts

$$
\hat{H}(\lambda) = \hat{H}^{(0)} + \lambda\hat{H}^{(1)}
$$

where $\hat{H}^{(0)}$ has an exact solution for $\hat{H}^{(0)}\Psi^{(0)} = E^{(0)}\Psi^{(0)}$, and $\hat{H}^{(1)}$ is the perturbative part of the Hamiltonian. Parameter $\lambda$ is the perturbation strength. The exact wave function $\Psi$ and energy $E$ can be expanded as the following infinite series

$$
\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \ldots
$$

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

\(^7\)
By inserting (2.24a) and (2.24b) into (2.22) and collecting terms that have the same power of \( \lambda \), we get the \( n \)-th order Schrödinger equation:

\[
\hat{H}^{(0)} \Psi^{(n)} + \hat{H}^{(1)} \Psi^{(n-1)} = \sum_{k=0}^{n} E^{(k)}(n-k),
\]

(2.25)

where \( E^{(n)} \) can be found by projection

\[
E^{(n)} = \langle \Psi^{(0)} | \hat{H}^{(1)} | \Psi^{(n-1)} \rangle.
\]

(2.26)

The perturbed first-order wave function is usually obtained by taking a linear combination of the zeroth-order wave functions that forms a complete set as

\[
\Psi^{(1)} = \sum_{i} c^{(1)}_{i} \Psi^{(0)}_{i}
\]

(2.27)

where \( \{c^{(1)}_{i}\} \) represents a set of undetermined coefficients.

This method has been used in quantum chemistry, too, and it is known as Møller-Plesset perturbation theory, where the zeroth-order Hamiltonian is the sum of the Fock operators

\[
\hat{H}^{(0)} = \sum_{i} \hat{F}(i).
\]

(2.28)

Then, the first-order Hamiltonian is

\[
\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)} = \sum_{i<j} \frac{1}{r_{ij}} - \sum_{i} (\hat{J}(i) - \hat{K}(i)).
\]

(2.29)

By employing (2.26), we find that

\[
E^{(0)}_{0} = \langle \Psi_{0} | \hat{H}^{(0)} | \Psi_{0} \rangle = \sum_{i=1}^{N} \epsilon_{i}
\]

(2.30a)

\[
E^{(1)}_{0} = \langle \Psi_{0} | \hat{H}^{(1)} | \Psi_{0} \rangle = \langle \Psi_{0} | \hat{H} - \hat{H}^{(0)} | \Psi_{0} \rangle = E_{\text{HF}} - E^{(0)}_{0}
\]

(2.30b)
demonstrating that \( E_{\text{HF}} \) is the sum of the zeroth- and first-order corrections. Therefore, the electron correlation effects are captured at the second-order level of the Møller-Plesset method (MP2), where the first-order correction to the wave function is given by

\[ \Psi^{(1)} = \frac{1}{4} \sum_{ijab} t_{ij}^{ab} |\Psi_{ij}^{ab}\rangle. \]  

(2.31)

In the case where canonical orbitals are employed, the second-order correction to the energy is

\[ E^{(2)} = \frac{1}{4} \sum_{ijab} \frac{|\tilde{g}_{ij}^{ab}|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \]  

(2.32)

where \( \tilde{g}_{rs}^{pq} \equiv g_{rs}^{pq} - g_{sr}^{pq} \) is the antisymmetrized Coulomb integral, and \( g_{rs}^{pq} \equiv \langle rs | r_{12}^{-1} | pq \rangle \). The MP2 method is frequently used computational model since it is both size-consistent and size-extensive and it has a lower computational scaling (\( \mathcal{O}(N^5) \)) than the CISD and CCSD methods. The MP2 method captures more than 80% of the correlation energy. Inclusion of the higher order corrections defines a series of more accurate MPn methods such as MP3 and MP4.

### 2.3 Explicitly Correlated R12/F12 Methods

In order to achieve high accuracy with the methods discussed in previous subsections, one needs to employ very large basis sets. The main reason for this is that Slater determinants are constructed from one-electron wave functions. In other words, they fail to describe the wave function properly when electrons are close to each other. The true nature of the first
derivative of the exact wave function when two electrons come close to each other (coalescence point) is

\[
\frac{\partial \Psi^{\text{sph}}}{\partial r_{12}} \bigg|_{r_{12}=0} = \frac{1}{2} \Psi(r_{12} = 0)
\]

(2.33)

where sph denotes the spherical averaging over the hypersphere where \( r_{12} = \text{const.} \) Kato suggested that extra terms have to be included in the wave function. The exact wave function for the singlet state for the helium atom is as follows

\[
\Psi_{\text{exact}}(r_1, r_2) = \left( 1 + \frac{1}{2} r_{12} + O(r_{12}^2) \right) \Psi(r_{12} = 0).
\]

(2.34)

The wave function for the helium atom in the triplet state near the electron coalescence point is

\[
\Psi_{\text{exact}}(r_1, r_2) = r_{12} \cdot w \left( 1 + \frac{1}{4} r_{12} + O(r_{12}^2) \right)
\]

(2.35)

with vector \( w = (\partial \Psi/\partial x_{12}, \partial \Psi/\partial y_{12}, \partial \Psi/\partial z_{12}) \) evaluated at the coalescence point. This means that we need to explicitly include the inter-electronic dependence \( (r_{12}) \) contribution to the wave function. Figure 2.1 from the previous section shows the plot of the exact wave function for the helium atom and convergence of the CI wave function calculated with different sizes of basis sets, where the size of the basis set is defined by \( N \). The Hartree-Fock wave function is a straight line, which means that the positions of both electrons are independent of each another. The exact wave function (cusp included) describes the electron-electron interaction near the coalescence point properly. An interesting result is obtained with CI; by increasing the size of the basis set, we approach the basis set limit, but this
approach is extremely slow, as can be seen from the plot. Also, such a wave function does not satisfy Eq. (2.33). It is obvious that in order to improve the convergence of a method to a basis set limit, one needs to include the cusp condition. This is achieved in the explicitly correlated R12/F12 methods that were introduced in quantum chemistry by Kutzelnigg.\textsuperscript{36}

In the next subsection, we describe explicitly correlated MP2 and CC methods.

### 2.3.1 Explicitly Correlated Second-Order Møller-Plesset Methods

The MP1 condition and the second-order energy expression defined by Eq. (2.25) and (2.26) can be combined to construct a special type of Lagrangian that is known as the second-order Hylleraas functional:

$$H^{(2)} = 2 \langle \psi^{(0)} | \hat{H}^{(1)} | \psi^{(1)} \rangle + \langle \psi^{(1)} | \hat{H}^{(0)} - E^{(0)} | \psi^{(1)} \rangle \geq E^{(2)}. \quad (2.36)$$

The amplitudes can be obtained variationally by satisfying the following condition

$$\frac{dH^{(2)}}{dt^{ij}_{ab}} = 0 \quad (2.37)$$

giving a set of linear equations that are solved for $t$:

$$y^{ij}_{ab} + t^{ij}_{cb} F^{c} + t^{ij}_{ac} F^{c} - t^{kj}_{ab} F^{i} = 0 \quad (2.38)$$

where $F^{p}_{q} = \langle q | \hat{F} | p \rangle$ are elements of the Fock matrix; $|\psi^{(0)}\rangle$ is the reference Hartree-Fock Slater determinant; $|\psi^{(1)}_{MP1}\rangle = \frac{1}{4} t^{ij}_{ab} a^{\dagger}_{i} a^{\dagger}_{j} a_{b} a_{a} |\psi^{(0)}\rangle$ is the first order-correction to the wave function. In this notation, $a^{\dagger}_{ij} = a^{\dagger}_{a} a^{\dagger}_{b} a_{j} a_{i}$ is a double excitation second-quantized operator, and we assume the Einstein summation convention over repeated indices. This is an alternative
yet equivalent way to derive the MP2 method that we discussed earlier, and it is a useful framework for development of the MP2-F12 method.\(^{33-35}\) In the MP2-F12 method, the standard MP1 wave function is augmented by the explicitly correlated (geminal) component:

\[
\Psi_{\text{MP1-F12}}^{(1)} = \Psi_{\text{MP1}}^{(1)} + \frac{1}{8} t_{ij}^{kl} \tilde{R}_{\alpha\beta}^{kl} a_{ij}^{\alpha\beta} |\Psi^{(0)}\rangle
\]  

(2.39)

where \(\tilde{R}_{\alpha\beta}^{kl}\) is the antisymmetrized matrix element of the geminal correlation factor \(f(r_{12})\):\(^{37}\)

\[
R_{\alpha\beta}^{ij} \equiv \langle \alpha\beta | \hat{Q}_{12} f(r_{12}) | ij \rangle.
\]

(2.40)

In addition, \(t_{ij}^{kl}\) are geminal amplitudes, and \(a_{ij}^{\alpha\beta}\) is a double excitation operator to the complete unoccupied space with indices \(\alpha\) and \(\beta\) as defined in Figure 2.2. The projector \(\hat{Q}_{12} \equiv (1 - \hat{O}_1)(1 - \hat{O}_2) - \hat{V}_1\hat{V}_2\) keeps the geminal functions orthogonal to Hartree-Fock as well as to the standard double excitations. The \(\hat{O}\) and \(\hat{V}\) operators project onto the occupied and unoccupied spaces, respectively. Insertion of Eq. (2.39) into the second-order energy equation Eq. (2.26) defines the MP2-F12 energy:

\[
E_{\text{MP2-F12}}^{(2)} = E_{\text{MP2}}^{(2)} + E_{\text{F12}}^{(2)}
\]

(2.41)

which consists of the MP2 energy and the explicit correlation correction F12 that is obtained from

\[
E_{\text{F12}}^{(2)} = -\frac{1}{4} t_{ij}^{kl} \tilde{V}_{ij}^{kl}
\]

(2.42)

with

\[
t_{ij}^{kl} = \frac{1}{2} \tilde{V}_{mn}^{ij} (\tilde{B}^{-1})_{mn}^{kl}.
\]

(2.43)
Furthermore, $V_{ij}^{kl}$ is defined in the basis of canonical orbitals as

$$V_{ij}^{kl} = V_{ij}^{kl} - \frac{1}{2} C_{ij}^{ab} \left( F_a^i + F_b^i - F_j^i - F_j^j \right), \quad (2.44)$$

while $(\tilde{B}_{(ij)})_{kl}^{mn}$ is defined as

$$(\tilde{B}_{(ij)})_{kl}^{mn} = B_{kl}^{mn} - (F_i^i + F_j^j) X_{kl}^{mn} - \frac{1}{2} C_{kl}^{ab} C_{ab}^{mn} \left( F_a^i + F_b^i - F_i^i - F_i^j \right). \quad (2.45)$$

The $V, C, B$ and $X$ terms are standard intermediates that occur in the MP2-F12 method and are defined by:

$$V_{ij}^{kl} = \frac{1}{2} g_{ij}^{\alpha \beta} R_{\alpha \beta}^{kl} \quad (2.46)$$

$$C_{ij}^{kl} = \frac{1}{2} (F_a^i R_{ab}^{kl} + F_b^i R_{ba}^{kl}) \quad (2.47)$$

$$B_{kl}^{mn} = \tilde{R}_{mn}^{\alpha \gamma} F_{\beta}^{\alpha \gamma} \tilde{R}_{\alpha \beta}^{\alpha \beta} \quad (2.48)$$

$$X_{kl}^{mn} = \frac{1}{2} \tilde{R}_{mn}^{\alpha \beta} \tilde{R}_{\alpha \beta}^{\alpha \beta}. \quad (2.49)$$

Instead of optimizing the geminal amplitudes, their values can be fixed based on the singlet and triplet cusp conditions of the exact wave function defined by Eq. (2.34) and (2.35) in the so-called SP approach. Then the F12 part of the MP1-F12 wave function becomes

$$\Psi_{F12}^{(1)} = \frac{1}{2} R_{ij}^{\alpha \beta} a_{ij}^{\alpha \beta} |\Psi^{(0)}\rangle \quad (2.50)$$

where

$$R_{ij}^{\alpha \beta} = \langle ij | \tilde{Q}_{ij} f(r_{12}) \left( \frac{1}{2} \hat{P}_0 + \frac{1}{4} \hat{P}_1 \right) |\alpha \beta\rangle. \quad (2.51)$$

The $\hat{P}_0$ and $\hat{P}_1$ operators project onto singlet and triplet spin-pair states:

$$\hat{P}_0 = |\alpha \beta\rangle_0 \langle \alpha \beta|_0 \quad (2.52)$$

$$\hat{P}_1 = |\alpha \alpha\rangle \langle \alpha \alpha| + |\alpha \beta\rangle_1 \langle \alpha \beta|_1 + |\beta \beta\rangle \langle \beta \beta| \quad (2.53)$$
with

$$|\alpha\beta\rangle_0 = \frac{1}{\sqrt{2}}(\alpha(1)\beta(2) - \beta(1)\alpha(2))$$  \hspace{1cm} (2.54)$$

$$|\alpha\beta\rangle_1 = \frac{1}{\sqrt{2}}(\alpha(1)\beta(2) + \beta(1)\alpha(2)).$$  \hspace{1cm} (2.55)$$

This choice of coefficients is appropriate when \(\lim_{r_{12} \to 0} f_{12}/r_{12} = 1\). With such a constructed wave function, the MP2-F12 correlation energy within the SP approach is obtained through the second-order Hylleraas functional. This method has the same computational scaling as the MP2 method \(O(N^5)\); however, due the need to compute the \(B, V, X\) and \(C\) intermediates, its prefactor is typically 10 times larger. This is compensated for by the fact that MP2-F12 recovers the majority of the MP2 correlation energy with a modest basis set. The intermediates \(B, V\) and \(X\) produce many-electron integrals, and in the next section, we will discuss how to deal with them in the most efficient way.

### 2.3.2 Approximations to Many-Electron Integrals

The intermediates that we introduced in the previous subsection give rise to 3- and 4-electron integrals. Their evaluation presents a computational bottleneck; therefore, they require special attention, and many approximations have been introduced for their efficient
evaluation. Here we first discuss the evaluation of the $V$ intermediate, defined by\textsuperscript{34,35}

$$V_{ij}^{kl} = \langle ij | \frac{1}{r_{12}} \hat{O}_1 f(r_{12}) | kl \rangle$$

$$= \langle ij | \frac{1}{r_{12}} \left( (1 - \hat{O}_1)(1 - \hat{O}_2) - \hat{V}_1 \hat{V}_2 \right) f(r_{12}) | kl \rangle$$

$$= \langle ij | \frac{1}{r_{12}} f(r_{12}) | kl \rangle - \langle ij | \frac{1}{r_{12}} \hat{V}_1 \hat{V}_2 f(r_{12}) | kl \rangle + \langle ij | \frac{1}{r_{12}} \hat{O}_1 \hat{O}_2 f(r_{12}) | kl \rangle$$

$$- \langle ij | \frac{1}{r_{12}} \hat{O}_1 f(r_{12}) | kl \rangle - \langle ij | \frac{1}{r_{12}} \hat{O}_2 f(r_{12}) | kl \rangle.$$  \hspace{1cm} (2.56)

The first term on the right hand side, $(gr)_{pq}^{rs} \equiv \langle pq | \frac{1}{r_{12}} f(r_{12}) | rs \rangle$, is a new type of two-electron integral that has an analytical solution if Gaussian basis sets are used. The rest of the terms can be resolved by inserting the definition of $\hat{V} = \sum_a |a\rangle \langle a|$ and $\hat{O} = \sum_i |i\rangle \langle i|$. Thus the second term is

$$\langle ij | \frac{1}{r_{12}} \hat{V}_1 \hat{V}_2 f(r_{12}) | kl \rangle = \sum_{ab} \langle ij | \frac{1}{r_{12}} |ab\rangle \langle ab| f(r_{12}) |kl \rangle.$$  \hspace{1cm} (2.57)

In addition to standard two-electron Coulomb repulsion integrals, a new type of two-electron integral arise with the correlation factor $f(r_{12})$ kernel; these integrals, which take the form

$$r_{pq}^{rs} \equiv \langle pq | f(r_{12}) | rs \rangle,$$  can be analytically solved. The third term is equivalent to the previous one, except summations over unoccupied indices $a, b$ are replaced by summations over the complete occupied indices $m, n$. The last two terms will yield three-electron integrals

$$\langle ij | \frac{1}{r_{12}} \hat{O}_1 f(r_{12}) | kl \rangle = \sum_{m} \langle ij | \frac{1}{r_{12}} f(r_{23}) | mlk \rangle$$  \hspace{1cm} (2.58)

$$\langle ij | \frac{1}{r_{12}} \hat{O}_2 f(r_{12}) | kl \rangle = \sum_{m} \langle ij | \frac{1}{r_{12}} f(r_{23}) | kml \rangle.$$  \hspace{1cm} (2.59)
The main idea is to avoid computation of such expensive integrals, and the first step toward that is to insert the resolution of the identity (RI)

$$1 = \sum_\kappa |\kappa\rangle \langle \kappa| \approx \sum_{{p'}} |p'\rangle \langle p'|$$  \hspace{1cm} (2.60)

where $\sum_{{p'}} |p'\rangle \langle p'|$ is an approximate form of the exact projector to a complete basis $|\kappa\rangle$. This becomes exact for atoms if the RI basis is saturated up to angular momentum $3L_{\text{occ}}$.

Insertion of the approximate RI will give

$$\langle ijm| \frac{1}{r_{12}} f(r_{23}) |mlk\rangle_{\text{RI}} \approx \sum_{{p'}} \langle ij| \frac{1}{r_{12}} |mp'\rangle \langle mp'| f(r_{12}) |kl\rangle.$$  \hspace{1cm} (2.61)

In the past, $p'$ is approximated by the orthogonal orbitals of the OBS space, but this required very large basis sets to be accurate enough. This was alleviated by introduction of a special auxiliary ABS designed specifically for that purpose. Valeev reformulated the projection operator as

$$\hat{Q}_{12} = 1 - \hat{P}_1 \hat{P}_2 - \hat{O}_1 \hat{V}_2' - \hat{V}_1' \hat{O}_2$$  \hspace{1cm} (2.62)

where $\hat{P}$ projects onto the OBS space while $\hat{V}'$ projects onto the space of unoccupied orbitals that are complementary to the OBS space. This approach is termed CABS, which stands for complementary auxiliary basis set or complementary ABS. Insertion of such a projector into the expression for $V$ gives

$$V_{ij}^{kl} = \langle ij| \frac{1}{r_{12}} \hat{Q}_{12} f(r_{12}) |kl\rangle$$

$$= \text{CABS} \left( \langle ij| \frac{1}{r_{12}} \left( 1 - \hat{P}_1 \hat{P}_2 - \hat{O}_1 \hat{V}_2' - \hat{V}_1' \hat{O}_2 \right) f(r_{12}) |kl\rangle \right)$$

$$= (g'p)_{ij}^{kl} - \frac{1}{2} g_j^{pq} r_{pq} - g_{ij}^{ma'} r_{ma'}^{kl}.$$
Similar considerations can be applied to intermediate $X$:

$$X_{ij}^{kl}_{\text{CABS}} = (f(r_{12})^2)_{ij}^{kl} - \frac{1}{2} \tilde{r}_{ij}^{pq} \tilde{r}_{pq}^{kl} - \tilde{r}_{ij}^{ma'} \tilde{r}_{ma'}^{kl}$$  (2.64)

giving one additional type of two-electron integral: $(f(r_{12})^2)_{pq}^{rs} \equiv \langle pq | f(r_{12})^2 | rs \rangle$. The intermediate $C$ is simply evaluated within the CABS approximation as

$$C_{ij}^{ab}_{\text{CABS}} = \tilde{r}_{aa'}^{ij} F_{b}^{a'} + \tilde{r}_{a'b}^{ij} F_{a}^{a'}.$$  (2.65)

The evaluation of the $B$ intermediate is much more involved. Since it contributes the most to the explicit correlation correction along with the $V$ intermediate, it has to be evaluated very carefully and with high accuracy. Many schemes have been proposed that are termed the A,B,C and D approximations.\textsuperscript{70–76} The contributions that arise from identity of the $\hat{Q}_{12}$ will already give rise to many-electron integrals and only these contributions should be approximated by the RI approximation. The contribution for the $B$ intermediate is

$$\langle ij | f(r_{12}) \hat{F}_{12} f(r_{12}) | kl \rangle = \frac{1}{2} \langle ij | [f(r_{12}), [\hat{F}_{12}, f(r_{12})]] | kl \rangle$$

$$+ \frac{1}{2} \langle ij | \hat{F}_{12} f(r_{12})^2 | kl \rangle + \frac{1}{2} \langle ij | f(r_{12})^2 \hat{F}_{12} | kl \rangle$$

$$= \frac{1}{2} \langle ij | [f(r_{12}), [\hat{T}_{12}, f(r_{12})]] | kl \rangle + \frac{1}{2} \langle ij | [f(r_{12}), [\hat{K}_{12}, f(r_{12})]] | kl \rangle$$

$$+ \frac{1}{2} \langle ij | \hat{F}_{12} f(r_{12})^2 | kl \rangle + \frac{1}{2} \langle ij | f(r_{12})^2 \hat{F}_{12} | kl \rangle$$

with $\hat{F}_{12} = \hat{F}_1 + \hat{F}_2$ and $\hat{K}_{12} = \hat{K}_1 + \hat{K}_2$. The first term has an analytical solution since

$$\frac{1}{2} [f(r_{12}), [\hat{T}_{12}, f(r_{12})]] = |\nabla_1 f(r_{12})|^2.$$  (2.67)
In approximation A, the commutator $[\hat{K}_{12}, f(r_{12})]$ is neglected, making this method computationally inexpensive. In approximation B, the commutator $[\hat{K}_{12}, f(r_{12})]$ is evaluated rigorously. The difference between the B and C approximations is that approximation C evaluates a single commutator between the kinetic energy and the geminal within the resolution of the identity. Both of the B and C approximations are more accurate than the simple A approximation, converging much more quickly to the basis set limit with $O[(L + 1)^{-7}]$ for the B and C approximations and $O[(L + 1)^{-5}]$ for the A approximation. The D approximation, which has been proposed by Valeev and co-workers, is a simplified version of the C approximation that omits computation of the exchange operator $\hat{K}$ over two CABS indices. Another approximation is to impose the Generalized Brillouin Condition (GBC), because these terms have a negligible effect on the explicit correlation contribution to the correlation energy. Its accuracy has been extensively validated via application to a variety of chemical processes such as noncovalent interactions, atomization energies and enthalpies of formations for both open- and closed-shell systems, in all cases yielding very small errors relative to the C approximation but at significantly lower cost. Additionally, the quality of the D approximation has been validated in the context of explicitly correlated Green’s functions.

### 2.3.3 CABS Singles Correction

Typical MP2-F12 calculations will give the Hartree-Fock energy and the correlation energy with inclusion of the basis set error due to the explicit correlation correction. Since F12 is de-
signed to recover basis set error rapidly, the main source of error becomes the HF energy, due to fact that the explicit correlation correction does not affects the singles correlation. This was realized by Adler et al. and furthermore extensively investigated by Tew and Köhn, showing that the basis set correction to HF can be efficiently recovered by employing singles excitations that couple the occupied and complementary auxiliary spaces this technique is known as CABS singles. Then the CABS singles correction is calculated by

\[ E_{(2)s} = F^a_i t^i_a + F^{a'}_i t^{i'}_{a'} \]  

(2.68)

where \( t^i_a \) and \( t^{i'}_{a'} \) are the corresponding CABS singles amplitudes that are found by solving the following system of linear equations:

\[
0 = F^i_a + F^{b'}_a t^i_b - F^{i'}_j t^j_a + F^{b'}_a t^{i'}_{b'}
\]  

(2.69)

\[
0 = F^{i'}_{a'} + F^{b'}_{a'} t^{i'}_{b'} - F^{i'}_j t^{i'}_{a'} + F^{b'}_{a'} t^i_{b'}.
\]  

(2.70)

Evaluation of the CABS singles amplitudes and energy correction requires only minor computational effort. Compared with the MP2-F12 portion, it takes only a fraction of the total computation time.

### 2.3.4 Performance of MP2-F12 method

Table 2.1 shows the accuracy of the MP2 and MP2-F12 methods in the prediction of the binding energy of the benzene dimer relative to the MP2 method at the complete basis set limit (CBS limit). Binding energies are obtained employing the aug-cc-pVXZ and cc-pVXZ-F12 basis sets with \( X = D, T, Q \). The HF/CBS limit value is taken to be the HF/aug-cc-
Explicitly correlated methods give their best results if basis set with diffuse functions are used, so the usual choice is the series of aug-cc-pVXZ basis sets where X is the basis set cardinal number. Peterson and co-workers have designed basis sets for explicit correlation calculations; these basis sets, which are designated by cc-pVXZ-F12, have more s and p functions than the aug-cc-pVXZ family for the same basis set cardinal number.

Table 2.1: Binding energies (kcal/mol) for the benzene dimer with the MP2 and MP2-F12 methods employing the cc-pVXZ-F12 and aug-cc-pVXZ (X=D,T,Q) basis sets.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>HF</th>
<th>(E_{(2)s})</th>
<th>MP2</th>
<th>MP2-F12</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ-F12</td>
<td>3.810</td>
<td>0.095</td>
<td>-5.354</td>
<td>-4.917</td>
</tr>
<tr>
<td>cc-pVTZ-F12</td>
<td>3.918</td>
<td>0.020</td>
<td>-5.028</td>
<td>-4.780</td>
</tr>
<tr>
<td>cc-pVQZ-F12</td>
<td>3.936</td>
<td>0.007</td>
<td>-4.960</td>
<td>-4.747</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>2.668</td>
<td>0.947</td>
<td>-7.572</td>
<td>-5.258</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>3.752</td>
<td>0.169</td>
<td>-5.896</td>
<td>-4.903</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td>3.892</td>
<td>0.030</td>
<td>-5.159</td>
<td>-4.817</td>
</tr>
<tr>
<td>CBS limit</td>
<td>3.948</td>
<td>0.000</td>
<td>-4.743</td>
<td>-4.743</td>
</tr>
</tbody>
</table>

We can see from Table 2.1 that the HF method, regardless of the basis set, does not predict that this system will be bound; this is due to the way in which HF treats electron correlation. In contrast, this system is qualitatively described properly by the MP2 method, which predicts the correct sign of the binding energy. We can also see the slow convergence rate of the MP2 method to the CBS limit for both basis sets, though the cc-pVXZ-F12 family
of basis sets produces smaller errors than their aug-cc-pVXZ counterparts. Inclusion of the explicit correlation correction in the MP2-F12 method improves the basis set convergence rate as expected. Already at the double-zeta basis set, the binding energy is closer to the CBS limit than the conventional MP2 value computed with quadruple-zeta basis set is. As a rule of thumb, the MP2-F12 method with an X level quality basis set will produce results on par with an MP2 calculation employing an X+2 level quality basis set, though at a much lower cost. The superiority of the cc-pVXZ-F12 basis sets is evident, and this holds for different types of chemical processes.\cite{87,88} Lastly, the CABS singles correction is negligible for the cc-pVXZ-F12 basis sets, but with the aug-cc-pVDZ basis sets it has a significant contribution of around 1 kcal/mol.

### 2.3.5 Explicitly Correlated Coupled-Cluster Methods

The MP2 binding energy of the benzene dimer is only a crude estimate of the correct binding energy. For instance, the “gold standard” of quantum chemistry CCSD(T)/CBS estimates the binding energy at -2.822 kcal/mol.\cite{81} Therefore, the next logical step is to combine the highly accurate coupled-cluster methodology with explicit correlation. This was first done by Noga et al. where they used a similar approach to the one we discussed in the context of MP2-F12, by extending the conventional cluster operator with the geminal cluster operator ($\hat{T}^\gamma$).\cite{89,90} The CC-F12 wave function ansatz is as follows:

$$\Psi_{\text{CC-F12}} = e^{\hat{T}^\gamma}$$

(2.71)
where $\hat{T}^{\gamma} \equiv \frac{1}{8} t_{ki}^{ij} p_{i}^{kl} a_{ij}^{\alpha \beta}$ has the same physical meaning as in the context of the MP2-F12 method. As was the case for the conventional CC method, we can formulate the CCSD-F12 method by including single and double excitation in the cluster operator $\hat{T}$. The energy expression is obtained by projection

$$
E_{\text{CCSD-F12}} = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle + \langle \Psi_0 | [\hat{H}, \hat{T}^{\gamma}] | \Psi_0 \rangle
$$

(2.72)

$$
= E_{\text{CCSD}} + \frac{1}{4} t_{ij}^{kl} V_{ij}^{kl}.
$$

(2.73)

where $\hat{H}$ is the similarity-transformed CCSD Hamiltonian. The amplitude equations are

$$
0 = \langle \Psi_i^q | \hat{H} + [\hat{H}, \hat{T}^{\gamma}] | \Psi_0 \rangle
$$

(2.74)

$$
0 = \langle \Psi_{ij}^{ab} | \hat{H} + [\hat{H}, \hat{T}^{\gamma}] + \frac{1}{2} [[\hat{H}, \hat{T}^{\gamma}], \hat{T}^{\gamma}] | \Psi_0 \rangle
$$

(2.75)

$$
0 = \langle \Psi_{ij}^{kl} | \hat{H} + [\hat{H}, \hat{T}^{\gamma}] + \frac{1}{2} [[\hat{H}, \hat{T}^{\gamma}], \hat{T}^{\gamma}] | \Psi_0 \rangle.
$$

(2.76)

In addition to the singles and doubles amplitude equations, the geminal amplitude equations must also be solved iteratively. This method introduces two more intermediates, $P$ and $Z$, in addition to the $B$, $X$, $V$ and $C$ intermediates. These new intermediates are defined as

$$
P_{kl}^{ij} = R_{kl}^{\alpha \beta} g_{\alpha \beta}^{\gamma \delta} R_{\gamma \delta}^{ij}
$$

(2.77)

$$
Z_{kl;m}^{ij;p} = R_{kl}^{\alpha \beta} g_{\alpha \beta}^{\gamma \delta} R_{\gamma \delta}^{ij
$$

(2.78)

where $ij;p$ is notation for antisymmetry with respect to permutation of $i$ and $j$ only. The computational bottleneck for the geminal amplitudes is of $O^3V'^3$ computational cost where $V'$ is the number of CABS orbitals; therefore, like with CCSD, the formal scaling of the CCSD-F12 method is $O(N^6)$ but with a much larger prefactor. From the implementation point of
view, the CCSD-F12 method has many more terms and more types of two-electron integrals that have to be evaluated. Within this explicitly correlated framework, the features of the original CCSD method, such as size-consistency, size-extensivity and orbital-invariance, are preserved. Due to its high complexity, the early implementations were done within the standard approximation (SA),\textsuperscript{70} which can be viewed as a simplification of the CABS approach, significantly simplifying the equations. Full implementation of the CCSD-F12 method was first done by Shiozaki et al. by employing automated derivation of equations and implementation using SMITH code. This was subsequently used to implement higher order CCSDT-F12 and CCSDTQ-F12 methods.\textsuperscript{91–93} A similar implementation was done by Köhn and co-workers with the GECKO package.\textsuperscript{94} From the present-day perspective, the exact CCSD-F12 method is only used to benchmark the accuracy of the approximate CCSD-F12 methods that will be discussed in the next section.

2.3.6 Approximate Explicitly Correlated Coupled-Cluster Methods

As was mentioned in the previous section, the high complexity of the CCSD-F12 method restricts its use to the study of molecular systems that can be investigated with the conventional CCSD method; therefore, many approximate methodologies have been developed. The methods that fall into this category are CCSD(F12),\textsuperscript{95} CCSD(F12*),\textsuperscript{96} CCSD-F12x (x = a, b),\textsuperscript{70,97} and CCSD(2)$_{F12}$.\textsuperscript{98–100} Hättig, Tew and Köhn have extensively investigated
the performance of these methods for the calculation of reaction and atomization energies.\textsuperscript{96} The first data set contains 25 reactions of the H\textsubscript{2}, F\textsubscript{2}, N\textsubscript{2}, CO\textsubscript{2} and CO molecules while the second is composed of 30 small molecules containing the H, C, N, O and F atoms. The results are summarized in Table 2.2. The Table shows the mean and maximum correlation energy error in kJ/mol (per valence electron) for each method with respect to basis set. All calculations were performed with the SP approach. In the following text, we discuss the central ideas of each of the approximate methods.

**CCSD(F12) Method**

The (F12) family of methods was proposed by Klopper and co-workers.\textsuperscript{95} Its development is based on the following approximations: only linear terms of geminal amplitudes are retained and the commutator $[\hat{H}, \hat{T}^\gamma]$ from the geminal amplitude equations is simplified to $[\hat{F}, \hat{T}^\gamma]$. This will lead to the same expression for the singles amplitude equations, while the doubles and geminal amplitude equations will have a simplified form:

\begin{align}
0 &= \langle \Psi_{ij}^{ab} | \hat{H} + [\hat{H}, \hat{T}^\gamma] | \Psi_0 \rangle \\
0 &= \langle \Psi_{ij}^{kl} | \hat{H} + [\hat{F}, \hat{T}^\gamma] | \Psi_0 \rangle
\end{align}

where $\hat{F}$ is the similarity-transformed Fock operator. The CCSD(F12) correlation energy is calculated by

$$E_{\text{CCSD(F12)}} = \langle \Psi_0 | \hat{H} + [\hat{H}, \hat{T}^\gamma] | \Psi_0 \rangle .$$
Table 2.2: Performance of various approximated CCSD-F12 methods for the calculation of reaction and atomization energies (kJ/mol).

<table>
<thead>
<tr>
<th>Method</th>
<th>Basis set</th>
<th>Reaction Energy</th>
<th>Atomization Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Error</td>
<td>Max Error</td>
</tr>
<tr>
<td>CCSD(F12)</td>
<td>cc-pVDZ-F12</td>
<td>-0.10</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td>cc-pVTZ-F12</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>cc-pVQZ-F12</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CCSD(2)</td>
<td>cc-pVDZ-F12</td>
<td>-0.13</td>
<td>-0.42</td>
</tr>
<tr>
<td></td>
<td>cc-pVTZ-F12</td>
<td>-0.02</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>cc-pVQZ-F12</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CCSD-F12a</td>
<td>cc-pVDZ-F12</td>
<td>-0.14</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>cc-pVTZ-F12</td>
<td>-0.02</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>cc-pVQZ-F12</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>CCSD-F12b</td>
<td>cc-pVDZ-F12</td>
<td>-0.16</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>cc-pVTZ-F12</td>
<td>-0.03</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>cc-pVQZ-F12</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>CCSD[F12]</td>
<td>cc-pVDZ-F12</td>
<td>-0.09</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>cc-pVTZ-F12</td>
<td>-0.01</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>cc-pVQZ-F12</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>CCSD(F12*)</td>
<td>cc-pVDZ-F12</td>
<td>-0.10</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td>cc-pVTZ-F12</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>cc-pVQZ-F12</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Originally, the authors used a linear $r_{12}$ geminal correlation factor (CCSD(R12)), but this has been replaced by an exponential function of $r_{12}$, defining the CCSD(F12) method\textsuperscript{101,102}. This method is available in the TURBOMOLE\textsuperscript{103}, GELLAN, GECKO and MPQC\textsuperscript{104} quantum chemistry packages. This method represents an excellent, accurate and efficient approximation to the rigorous CCSD-F12 method. The computational scaling of the CCSD(F12) method is $O(N^6)$; however, each iteration takes approximately 3-5 times longer than a conventional CCSD iteration due to the required of construction of intermediates that are of $O^3V^2X$ computational cost, where $O$, $V$ and $X$ are the sizes of the occupied, unoccupied and auxiliary orbital spaces, respectively. This is reduced to 2-3 times longer with respect to a conventional CCSD iteration in the SP approach. If the SP approach is used, then the energy is evaluated using the Lagrangian, introducing only a negligible loss in accuracy. Additionally, it avoids the geminal basis set superposition error and it is numerically more stable\textsuperscript{105}. Further simplifications to the CCSD(F12) method were made by Hättig and co-workers who proposed two new methods, CCSD(F12*) and CCSD[F12]. They employed a perturbation analysis to assess level of contribution of each term in the CCSD(F12) method and discarding the negligible contributions, leading to highly efficient methods that gives almost the same result as the CCSD(F12) method\textsuperscript{96}. This is demonstrated by the results given in Table 2.2.
CCSD-F12a and CCSD-F12b Methods

Another family of approximate CCSD-F12 methods has been proposed by Adler et al.\textsuperscript{79,97} In their approach, they always assume the SP ansatz or the geminal amplitudes are computed at the MP2-F12 level.\textsuperscript{38} This presents a significant computational savings since it does not require to solve geminal amplitude equations. A second approximation is that they omit $[[\hat{W}, \hat{T}_2], \hat{T}^\gamma]$ with respect to CCSD(F12); in addition, other approximations from the CCSD(F12) method have been retained, leading to the following simplified doubles amplitude equation:

$$0 = \langle \Psi_{ij} | \hat{H} + [\hat{H}, \hat{T}^\gamma] | \Psi_0 \rangle. \quad (2.82)$$

Their last approximation is obtained by omitting all terms where CABS indices occur except for those that are present in MP2-F12-specific terms. The energy is then obtained from the Lagrangian where geminal lambda ($\Lambda^\gamma$) values are set to geminal amplitudes

$$E_{\text{CCSD-F12b}} = E_{\text{CCSD}} + \langle \Psi_0 | \Lambda^\gamma (\hat{W} + [\hat{W}, \hat{T}_2] + [\hat{F}, \hat{T}^\gamma]) | \Psi_0 \rangle, \quad (2.83)$$

while for the CCSD-F12a method the commutator $[\hat{W}, \hat{T}_2]$ is further approximated by $[\hat{F}, \hat{T}_2]$. This set of methods has proved to be computationally very efficient, with a cost similar to that of the CCSD method, but since they are a subset of the CCSD(F12) family of methods, they tend to yield slightly less accurate results. The CCSD-F12a method has a tendency to overestimate the CBS limit for larger basis sets as shown in Table 2.2. These methods are available in the MOLPRO quantum chemistry package.\textsuperscript{106}
CCSD(2)_{F12} and CCSD(2)_{FI2} Methods

Valeev recognized that the explicit correlation correction is small compared to the CCSD correlation energy, so it can be determined perturbatively, similarly to how the perturbative triple correction (T) is determined in the CCSD(T) method. Unlike previous methods wherein the explicit correlation correction is determined iteratively, here it is estimated perturbatively. These methods are denoted by CCSD(2)_{F12} and CCSD(2)_{FI2}, where the later one uses the SP approach. Within this formalism the Lagrangian is given by:

\[ \mathcal{L}_{CCSD-F12} = \mathbf{L}^{\dagger} \mathbf{H} \mathbf{R} \]  

(2.84)

with \( \mathbf{L} \) and \( \mathbf{R} \) representing the left and right eigenvectors of the similarity-transformed Hamiltonian. The central idea of this formalism is to rewrite the \( \mathbf{H} \) Hamiltonian and use the Löwdin partitioning for the zeroth- and first-order Hamiltonian. In the block-matrix form, the \( \mathbf{H} \) Hamiltonian is defined by

\[ \mathbf{H} = \begin{pmatrix} \mathbf{H}_{PP} & \mathbf{H}_{PQ} \\ \mathbf{H}_{QP} & \mathbf{H}_{QQ} \end{pmatrix}, \]  

(2.85)

while with the Löwdin partitioning it is given by

\[ \mathbf{H} = \mathbf{H}^{(0)} + \mathbf{H}^{(1)} = \begin{pmatrix} \mathbf{H}_{PP} & 0 \\ 0 & \mathbf{H}_{QQ}^{(0)} \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{H}_{PQ} \\ \mathbf{H}_{QP} & \mathbf{H}_{QQ}^{(1)} \end{pmatrix}. \]  

(2.86)

The basis for the P space are the reference, singly and doubly excited determinants \( \{ |\Psi_0 \rangle, |\Psi_i^s \rangle, |\Psi_{ij}^{ab} \rangle \} \) while the basis for the external Q space is a set of geminal double substitutions \( |\Psi_{ij}^{kl} \rangle \).
Further simplifications can be obtained by employing screening approximations (ScrA) allowing the use of a second-order Hylleraas functional that does not require evaluation of the $\lambda$ amplitudes since they are set to be the same as the doubles amplitudes. Then the Lagrangian is given by

$$\mathcal{L}_{(2)_{\text{F12}}} = \frac{1}{2} V_{kl}^{ij} t_{ij}^{kl} + \frac{1}{8} t_{ij}^{ow} (\tilde{B}_{(ij)})_{ow} t_{ij}^{kl}. \quad (2.87)$$

In addition to the second term that appears in the MP2-F12 method, an extra term appears that is defined by

$$\tilde{V}_{ij}^{kl} \equiv V_{ij}^{kl} + \frac{1}{2} (V_{ab}^{kl} + C_{ab}^{kl}) t_{ab}^{ij} + V_{ia}^{kl} t_{ia}^{ij} + V_{ia}^{kl} t_{ia}^{ij}. \quad (2.88)$$

This term includes coupling to the converged singles and doubles CCSD amplitudes. This approach is an excellent approximation to the CCSD-F12 method, and it has similar accuracy to the CCSD(F12), CCSD[F12] and CCSD(F12*) methods; however, it outperforms the CCSD-F12x methods (Table 2.2). Extension of the method to include the triples correction has been considered by Valeev and Crawford showing that there is a very small coupling between geminal and triples, which allows for use of the Fock operator as the zeroth-order Hamiltonian, thus making the triples contribution completely decoupled from the F12 correction.

Then the energy correction is

$$E_{\text{CCSD(T)}_{\text{F12}}} = E_{\text{CCSD}} + E_{(T)} + E_{(2)_{\text{F12}}}. \quad (2.89)$$

This equation shows that all components of the explicitly correlated CCSD(T) method within this framework are decoupled. Implementation of this method is relatively simple since the
code for the CCSD method does not need to be modified, unlike in the case of iterative approaches, and only minor changes in the MP2-F12 code are required. This method is available in MPQC,\(^\text{104}\) MOLPRO\(^\text{106}\) and ORCA,\(^\text{109}\) as in other quantum chemistry packages.

### 2.4 Summary

Conventional quantum chemical methods, such as CCSD(T), are capable of predicting and interpreting experimental outcomes to a high degree of accuracy. However, they are plagued by slow convergence to the basis set limit. The explicitly correlated R12/F12 methods are designed to reduce the basis set error by explicit inclusion of the inter-electronic distance (\(r_{12}\)) in the wave function expansion. The first modern explicitly correlated framework was done by Kutzelnigg and Klopper, who implemented the MP2-F12 method.\(^\text{36,70,110}\) Their method has been significantly improved by the introduction of new machinery such as the RI approximation,\(^\text{39,72}\) density fitting,\(^\text{111}\) correlation factor,\(^\text{37}\) the SP ansatz,\(^\text{38}\) CABS singles,\(^\text{40,94}\) approximations of many-electron integrals,\(^\text{70-76}\) as well as basis sets optimized for explicitly correlated calculations.\(^\text{85}\) All of this has enabled this method to become a routine tool for quantum chemical calculations. Since MP2 has limited accuracy, there was a desire to combine the explicit correlation correction with the more accurate and robust coupled-cluster family of methods. The pioneers in this area were Noga and Kutzelnigg with their formulation of the CCSD-F12 method. Their formalism was an extension of the cluster operator with the geminal operator.\(^\text{89,90}\) However, this method is applicable only
to individual atoms or to molecules of up to 5 atoms. This prompted the development of the various approximate CCSD-F12 methods. Among the approximate CCSD-F12 methods, the most accurate is CCSD(F12); however, this accuracy also comes at the highest cost: 2-5 times more expensive than the conventional CCSD method. Computationally more efficient are the CCSD(F12*) and CCSD[F12] methods obtained from perturbative screening of CCSD(F12), with accuracy comparable to the more expensive CCSD(F12). Even more computationally efficient are the CCSD-F12x (x=a,b) methods, but they tend to overestimate the basis set limit. Among the noniterative CCSD-F12 methods is the method developed by Valeev, favored for its great computational performance and simplicity, as well as for its accuracy, which is comparable to that of the CCSD(F12) method. All these methods are implemented in popular quantum chemical packages that are widely available to the quantum chemistry community. However, their computational cost is same as for the CCSD method, which is limited in application to molecules of around 30 atoms. In the next chapters we describe the development of the reduced-scaling explicitly correlated methods. In addition to that, we apply explicit correlation in the context of one-particle Green’s function for the accurate calculation of ionization potentials.
Chapter 3

SparseMaps – A systematic infrastructure for reduced-scaling electronic structure methods. IV. Linear-scaling second-order explicitly correlated energy with pair natural orbitals.

Reproduced from Pavošević, F.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. The Journal of Chemical Physics 2016, 144(14), 144109., with the permission of AIP Publishing.
3.1 Introduction

Recent developments of reduced-scaling conventional many-body methods,\textsuperscript{21,22,27,41,43,44,46,48,112-115} such as coupled-cluster,\textsuperscript{6,61,116} have put within reach accurate chemical applications of such methods to molecules with hundreds of atoms on a single workstation.\textsuperscript{25,27,113} The key to these developments is the use of sparse representations for the wave functions and operators, which reduces the storage and computational complexity of such methods to linear in system size, compared to the respective $O(N^4)$ and $O(N^6)$ complexities of a naive formulation of coupled-cluster singles and doubles. Although the idea of sparse representations of many-body methods goes back many decades, we in particular note the local correlation approach of Pulay and Sæbo\textsuperscript{12,117} that generated sustained efforts\textsuperscript{14,118} to make practical reduced-scaling many-body methods. The central idea of these efforts is the use of localized occupied and unoccupied orbitals (the latter traditionally spanned by projected atomic orbitals, PAOs); then sparse structure on the wave function amplitudes is imposed by \textit{domain approximation}. The high accuracy demanded by realistic chemical applications as well as the high degree of robustness needed for practical application often requires the use of impractically large PAO domains, with 20-40 atoms or more,\textsuperscript{25} for every pair of occupied orbitals. The recent reduced-scaling methods\textsuperscript{20,21} utilize pair-natural orbitals (PNOs) – originally introduced by Edmiston and Krauss\textsuperscript{119} – to construct a representation of amplitudes with \textit{size-independent} rank for every (spatially-localized) pair of electrons. When used with localized occupied orbitals, PNO-based pair-correlated wave functions have $O(N^2)$ storage complexity in a large system, and can be reduced to linear in large systems with weak-pair
Another problem with the many-body methods – the large basis set errors of many observables, such as reaction energies – has also been solved in the recent past by the use of wave function expansions including terms with \textit{explicit} dependence on the inter-electronic distances, also known as the \textit{explicitly correlated} wave function methods. The practical R12/F12 formalism for including such terms, pioneered by Kutzelnigg\textsuperscript{36} and developed by many others,\textsuperscript{37–39,71,72,95,98,111} drastically reduces the basis set error with a modest increase of computational cost to make them generally superior to basis set extrapolation (see, for example, Ref. 120). The explicitly correlated technology was combined with the reduced-scaling coupled-cluster methods based on the domain-based PAO representation by Werner and co-workers.\textsuperscript{40,41} Several groups reported PNO-based F12 coupled-cluster methods recently. Krause and Werner simulated local CCSD-F12 methods with a variety of reduced-scaling representations using a canonical program.\textsuperscript{42} Valeev and co-workers demonstrated a production implementation of a reduced-scaling coupled-cluster method that introduced F12 terms into the CCSD method in localized PNO representation,\textsuperscript{48} and a similar effort with a different framework for the coupled-cluster F12 method was reported by Hättig and co-workers.\textsuperscript{46}

Recently Neese and co-workers started out a program to formulate practical PNO-based many-body methods with all non-trivial steps formulated in an $O(N)$ fashion. This led to a formalism we call \textit{SparseMaps} for composing sparse tensor data structures and algorithms involved in such methods.\textsuperscript{26} A SparseMaps-based implementation of the local MP2 method
in the domain based local PNO (DLPNO-MP2) representation was shown to be highly efficient for small molecules (i.e., not significantly more expensive than the state-of-the-art conventional $O(N^5)$ MP2 implementation for small and medium sized molecules) and exhibited an early onset of linear-scaling even in conjunction with a triple-zeta quality basis. Similar improvements were recently demonstrated for the DLPNO-CCSD(T) method. In this work we report a SparseMaps-based formulation of the MP2-F12 method based on the DLPNO-MP2-F12 approach.\footnote{Since our DLPNO approach as well as all other practical PNO-based methods, utilizes the density-fitting, or resolution-of-the-identity, approximation for the two-electron integrals (see Section 3.2.1), the canonical counterpart of the DLPNO-MP2-F12 method is DF, or RI, variant of MP2-F12, accessible in ORCA using RI-MP2-F12/D or F12/D-RI-MP2 keyword (with “/D” denoting the F12 intermediate formulation of Eq. (3.22)). Throughout the manuscript MP2-F12 will imply DF/RI for brevity.}

The rest of the paper is structured as follows. We describe the linear-scaling DLPNO-MP2-F12 formalism in Section 3.2, followed by the computational details in Section 3.3. Section 3.4 presents an assessment of robustness and computational scaling of the DLPNO-MP2-F12 method. Lastly, the results of this work are summarized in Section 3.5.

### 3.2 Formalism

Here we only briefly review the explicitly correlated MP2-F12 method in its iterative form; more details can be found in Refs. 33–35, 121. Consider a closed-shell single-determinant state $|0\rangle$ in which the active occupied orbitals, denoted by $i, j, \ldots$, are localized. A com-
pact representation of the MP1-F12 wave function can be obtained in terms of pair-specific unoccupied orbitals, denoted by $a_{ij}, b_{ij}, \ldots$:

$$|1\rangle \equiv |1_{\text{conv}}\rangle + |1_{\text{F12}}\rangle = \frac{1}{2} t_{a_{ij}b_{ij}}^iE_{ij}^{a_{ij}b_{ij}} |0\rangle + \frac{1}{2} \sum_{ij} \tilde{\gamma}_{ij} |0\rangle. \quad (3.1)$$

(Einstein summation convention is implied, unless noted otherwise). In (3.1) $E_{ij}^{a_{ij}b_{ij}}$ are the spin-free two-particle excitation operators expressed in terms of the spin-orbital creators/annihilators $\hat{a}_i^\dagger/\hat{a}_i$ as

$$E_{ij}^{a_{ij}b_{ij}} \equiv \sum_{\tau,\sigma=\uparrow,\downarrow} \hat{a}_{a_{ij}}^\dagger \hat{a}_{ai}^\dagger \hat{a}_j \hat{a}_i, \quad (3.2)$$

where $\tau$ and $\sigma$ denote spin-up ($m_s = 1/2$) and spin-down ($m_s = -1/2$) magnetic quantum numbers. The geminal excitation operator,

$$\tilde{\gamma}_{ij} \equiv \tilde{R}_{ij}^{\alpha_{ij} \beta_{ij}} E_{ij}^{\alpha_{ij} \beta_{ij}}, \quad (3.3)$$

allows $|1\rangle$ to satisfy the first-order electron-electron cusp condition ($\alpha_{ij}, \beta_{ij}, \ldots$ denote a complete set of pair-specific unoccupied orbitals). Eq. (3.3) is defined with the help of matrix elements of the projected geminal \cite{38}

$$R_{\alpha_{ij} \beta_{ij}}^{ij} \equiv \langle \alpha_{ij} \beta_{ij} | \hat{Q}_{ij} f_{12} | ij \rangle, \quad (3.4)$$

where $f_{12} \equiv f(r_{12})$ is a spherically-symmetric correlation factor. Due to the spin dependence of the 2-electron cusp conditions, \cite{32,122} the appropriate modulation of the cusp coefficient is obtained by using $\tilde{R}$ tensor elements in (3.3), where the tilde denotes the pair-spin projection:\cite{38,88} $\tilde{R}_{pq}^{rs} \equiv O_{pq}^{rs}(C_0 + C_1)/2 + O_{pq}^{rs}(C_0 - C_1)/2 = (3/8)O_{rs}^{pq} + (1/8)O_{rs}^{qp}$, with $C_{0,1} = 1/2, 1/4$
the cusp coefficients of (natural-parity) singlet and triplet pairs, respectively.\textsuperscript{32} This choice of coefficients is appropriate when \( \lim_{r_{12} \to 0} f_{12}/r_{12} = 1 \).

Projector \( \hat{Q}_{ij} \) in the geminal operator (3.3) ensures the orthogonality of \( |1_{\text{F12}}\rangle \) to the conventional \( ij \to a_{ij}b_{ij} \) excitations in \( |1_{\text{conv}}\rangle \) (the \textit{strong} orthogonality of the geminals is automatically satisfied as long as the pair-specific unoccupied orbitals are rigorously orthogonal to \textit{all} occupied orbitals). The standard definition of \( \hat{Q}_{ij} \) due to Werner\textsuperscript{123} reads

\[
\hat{Q}_{ij} = 1 - \frac{1}{2} \sum_{a_{ij},b_{ij}} |a_{ij}b_{ij}\rangle \langle a_{ij}b_{ij}|. \tag{3.5}
\]

We recently proposed an improved definition of \( \hat{Q}_{ij} \) in terms of \textit{geminal}-specific orbitals (GSO) that works better in the perturbative formulation of explicitly correlated PNO-based coupled-cluster;\textsuperscript{48} however, in this work, we did not utilize the GSO projector.

The MP2-F12 energy is obtained by minimization of the second-order Hylleraas functional

\[
H^{(2)} = \langle 1| \hat{F}_N |1\rangle + 2 \langle 0| \hat{W}_N |1\rangle, \tag{3.6}
\]

where we assumed real wave functions. In (3.6) \( \hat{F}_N \) and \( \hat{W}_N \) are the normal-ordered (with respect to \( |0\rangle \)) Fock and fluctuation operators, respectively:

\[
\hat{F}_N \equiv F_\lambda^\kappa \tilde{E}_\kappa^\lambda, \tag{3.7}
\]
\[
\hat{W}_N \equiv \frac{1}{2} g_{\mu\nu}^\kappa \tilde{E}_\kappa^{\mu\nu}. \tag{3.8}
\]

where \( g_{\mu\nu}^\kappa \equiv \langle \mu\nu| r_{12}^{-1} |\kappa\lambda\rangle \) and the closed-shell Fock operator is \( \hat{F} \equiv \hat{h} + 2\hat{J} - \hat{K} \). The normal order of spin-free operators \( \tilde{E} \) relative to a single-determinant reference and the associated generalized Wick theorem are defined in Ref. 124. Stationary points of \( H^{(2)} \) are found by
solving the amplitude equations:

$$
\rho_{ij}^{ab} \equiv \langle 0 | E_{a_{ij}b_{ij}}^{ij} \hat{F}_N | 1 \rangle + \langle 0 | E_{a_{ij}b_{ij}}^{ij} \hat{W}_N | 0 \rangle = 0.
$$  \hfill (3.9)

Straightforward algebra yields

$$
\rho_{ij}^{ab} = C_{ij}^{ab} + g_{ij}^{ab} + \hat{S}_{ia_{ij}j_{bij}} F_{a_{ij}}^{ij} t_{ij}^{ij} - \hat{S}_{ia_{ij}j_{bij}} F_{k_{bij}}^{kj} t_{ij}^{kj} S_{a_{ij}}^{ak_{ij}} S_{bij}^{bk_{ij}} \hfill (3.10)
$$

where $S_{a_{ij}}^{ak_{ij}}$ are the overlap integrals between PNOs of different pairs and the symmetrization operator is defined by

$$
\hat{S}_{i_1 \ldots i_n, j_1 \ldots j_n} f(i_1, \ldots , i_n, j_1, \ldots , j_n) \equiv f(i_1, \ldots , i_n, j_1, \ldots , j_n) + f(j_1, \ldots , j_n, i_1, \ldots , i_n).
$$

All but the first term on the right-hand side of (3.10) appear in the conventional PNO-MP2 amplitude equation residual; their $O(N)$ evaluation was carried out precisely as described in Ref. 26. The only contribution from the F12 terms is the first term, known as the coupling intermediate of the F12 theory:

$$
C_{ij}^{ab} \equiv \hat{S}_{ia_{ij}j_{bij}} \bar{R}_{a_{ij}a_{ij}}^{ij} F_{b_{ij}}^{ab}. \hfill (3.11)
$$

The closed-shell MP2-F12 energy is obtained by replacing the $t$ amplitudes in (3.6):

$$
E^{(2)} = \bar{t}_{ij}^{ab_{ij}} \left( g_{ij}^{ab_{ij}} + C_{ij}^{ab} + \rho_{ij}^{ab} \right) + 2V_{ij}^{ab} + B_{ij}^{ab} - X_{ij}^{ik} F_{k}^{i} - X_{ij}^{kj} F_{k}^{i}, \hfill (3.12)
$$

where the bar in $\bar{O}_{pq}^{rs} \equiv 2O_{pq}^{rs} - O_{pq}^{rs}$ denotes the closed-shell counterpart of the antisymmetrization in the spin-orbital many-body electronic structure. At convergence $\rho_{ij}^{ab} = 0$ and the expression reduces to the usual closed-shell MP2-F12 energy obtained without the assumption of canonical orbitals.

The focus of this paper is on the evaluation of the F12 intermediates ($C$, $V$, $B$, and $X$) in an $O(N)$ fashion. The standard approaches to evaluation of these intermediates approximate
sums over the complete basis set with sums over the corresponding finite-basis representations. For example, the coupling intermediate (3.11) is evaluated by inserting projector $\hat{Q}_{ij}$ first and then approximating the sum over the complete unoccupied set $\{\alpha_{ij}\}$ with a sum over a finite Complementary Auxiliary Basis Set (CABS)\(^{39}\) composed of orbitals $a', b', \ldots$, that are orthogonal to the conventional occupied and unoccupied orbitals:

$$C_{a_i b_j}^{\alpha_j} \overset{\text{Eq. (3.5)}}{=} \tilde{S}_{i a_{ij}, j b_{ij}} \tilde{r}_{ij}^{a_i a_i'} \tilde{F}_{b_{ij}}^{a_i'};$$

$$\overset{\text{CABS}}{\approx} \hat{S}_{i a_{ij}, j b_{ij}} \tilde{r}_{ij}^{a_i a_i'} \tilde{F}_{b_{ij}}^{a_i'},$$

(3.13)

where $r_{ij}^{a_i a_i'} \equiv \langle a_{ij} a_i' | f_{12} | ij \rangle$. Note that the pair-specific CABS orbitals, $\{a_i'\}$, are orthogonal to all occupied orbitals as well as to the unoccupied orbitals for the given pair $ij$ that appear in projector (3.5).

Intermediate $V$,

$$V_{ij}^{\alpha\beta} \equiv g_{\alpha_{ij}, \beta_{ij}}^{ij} \tilde{R}_{ij}^{\alpha_{ij} \beta_{ij}},$$

(3.14)

is evaluated by CABS-based resolution of the double sums over complete unoccupied orbitals:\(^{39}\)

$$V_{ij}^{\alpha\beta} \overset{\text{CABS}}{=} \left( \frac{f_{12}}{r_{12}^{ij}} \right)^{ij}_{ij} - g_{p_{ij}, q_{ij}}^{ij} \tilde{R}_{ij}^{p_{ij} q_{ij}} - \hat{S}_{m_{ij}, a_{ij}}^{ij} \tilde{R}_{ij}^{m_{ij} a_{ij}}^{ij},$$

(3.15)

where $m_{ij}, n_{ij}, \ldots,$ represent pair-specific (core and active) occupied orbitals and $\{p_{ij}\} = \{m_{ij}\} \oplus \{a_{ij}\}$ is the orbital basis set (OBS) for pair $ij$ (note that the pair-specificity of $\{m_{ij}\}$ is due to the use of domain approximation described below).
Intermediate $X$,

$$X_{ij}^{ik} \equiv \tilde{R}_{\alpha ik}^{\beta ik} \tilde{R}_{ij}^{\alpha ij\beta ij}, \quad (3.16)$$

is evaluated in a manner similar to $V$, but due to the coupling of different pairs one has to introduce overlaps between their unoccupied orbitals

$$X_{ij}^{ik} \overset{\text{CABS}}{\approx} (\tilde{f}_2^{ij})_{ik} - \tilde{R}_{pik\eta k}^{\alpha k i} \tilde{R}_{qik\eta k}^{\beta k i} S_{\eta k}^{pik} S_{\eta k}^{qik} - \left( \tilde{R}_{mk\eta m\eta}^{\alpha m i} \tilde{R}_{ij}^{\beta m ij} + \tilde{R}_{\alpha m i}^{\beta m i} \tilde{R}_{ij}^{\beta m ij} \right) S_{\eta k}^{mk} S_{\eta k}^{\alpha k}, \quad (3.17)$$

where $\tilde{O}_{ij} \equiv (3/8)\tilde{O}_{ij} + (1/8)\tilde{O}_{ji}$. In our work on the LPNO-CCSD-F12 method\textsuperscript{48} we invoked the following approximation:

$$X_{ij}^{ik} F_k^i \approx \tilde{R}_{\alpha i j}^{\beta i j} \tilde{R}_{ij}^{\alpha ij\beta ij} F_k^i \equiv X_{ij}^{ij F}, \quad (3.18)$$

where the Fock-transformed occupied orbitals $\{i_F\}$ are obtained by a linear transformation of the occupied orbitals with the occupied-occupied block of the Fock matrix. Note that expression $(3.18)$ differs from $(3.16)$ by the use of the unoccupied orbitals specific to $ij$ pairs only. The approximated intermediate is evaluated by the same strategy as $V$:

$$X_{ij}^{ij F} \overset{\text{CABS}}{\approx} (\tilde{f}_2^{ij F})_{ij} - \tilde{R}_{pij\eta ij}^{\alpha ij} \tilde{R}_{qij\eta ij}^{\beta ij} S_{\eta ij}^{pij} S_{\eta ij}^{qij} - \left( \tilde{R}_{mij\alpha m\eta ij}^{\alpha mij} \tilde{R}_{ij}^{\beta mij \alpha mij} + \tilde{R}_{\alpha mij}^{\beta mij} \tilde{R}_{ij}^{\beta mij \alpha mij} \right). \quad (3.19)$$

Due to the lack of overlaps this approach allows one to compute the F12 energy contribution for each pair independently of all others. This is the approach that we take in this work; the implementation and testing of the more rigorous approach in Eq. (3.17) is left for future work.

Among the multiple methods for evaluation of intermediate $B$ – traditionally labeled $A$, $B$, and $C$ – the least rigorous is $A$, in which the exchange operator is assumed to commute with
This simplifies approach A significantly, at the cost of slower convergence to the basis set limit in partial wave-style bases; the error due to truncation of the basis at angular momentum $L$ is $\mathcal{O}([L+1]^{-5})$ in approach A vs. $\mathcal{O}([L+1]^{-7})$ in approaches B and C. The more rigorous approaches B and C only invoke CABS and only differ by the treatment of the single commutator between the geminal and the kinetic energy. Both approaches B and C are more complicated, thus until recently all implementations of local F12 methods utilized a drastically simplified version of approach A. The recent implementations utilize the more rigorous approaches B and C: the $\mathcal{O}(N^3)$ PNO-MP2-F12 and $\mathcal{O}(N^4)$ PNO-CC-F12 methods reported by Hättig, Tew, and co-workers utilized approach B, and the $\mathcal{O}(N^3)$ LPNO-CCSD(2)$_{F12}$ method by some of us utilized approach C. Here we utilized a simplified formulation of approach C that avoids the need for computing the exchange operator in the CABS basis (we dubbed this approximation of intermediate B approach D). This approach is related to the original suggestion of Klopper (so-called hybrid method) and some approximations of Werner et al. As we demonstrate below, this simplification does not significantly affect the accuracy but greatly reduces the computational cost. Intermediate B,

$$B_{ij}^{ij} \equiv \bar{S}_{it_{\alpha_ij,j}^i} \bar{R}_{a_{ij}^i \beta_{ij}^i}^j F_{\gamma_{ij}^i}^\beta_{ij} \bar{R}_{a_{ij}^i \gamma_{ij}^i}^\beta_{ij}, \quad (3.20)$$
is evaluated in approach C\textsuperscript{72} as

\[ B_{ij}^{\text{app}\ C} \approx \left( [f_{12}|\tilde{T}, f_{12}] ] \right)_{ij} + \tilde{S}_{ij} \left( f_{12}^2 \right)_{ij} \left( \hat{h} + 2 \hat{J} \right)_{ij} \]

\[ - \tilde{r}_{ij}^p \, K_{pij} \, \tilde{r}_{ij}^q \, \tilde{q}_{ij}^r \, F_{qij}^p \, \tilde{r}_{ij}^m \, \tilde{m}_{ij}^a \]

\[ - \tilde{r}_{ij}^a \, F_{pqij}^a \, \tilde{r}_{ij}^a \, r_{ij}^a \, F_{mij}^a \, \tilde{r}_{ij}^m \, \tilde{m}_{ij}^a \]

\[ - 2 \left( \tilde{r}_{mij} \, F_{prij}^m \, \tilde{r}_{ij}^p \, \tilde{r}_{ij}^a \, F_{qij}^a \right) \]

(3.21)

The set of orbitals \( \{p'_{ij}\} \) comprises the pair-specific Resolution-of-the-Identity basis set (RIBS) and is obtained as a union of pair-specific OBS and CABS: \( \{p'_{ij}\} = \{p_{ij}\} \oplus \{a'_{ij}\} \).

To simplify Eq. (3.21) without significant errors we invoked two approximations. First, the exchange matrix in the third and fourth term of (3.21) was replaced with its OBS/OBS representation. This is a substantial simplification as this completely eliminates the need for computation of the exchange operator with both indices in RIBS. This approximation also simplifies the triple RI sum in the third term to a single RI sum. The fourth term still includes a double sum over RIBS, but it now only involves the one-electron and Coulomb parts of the Fock operator. The second approximation consists of invoking Generalized Brillouin Condition (GBC),\textsuperscript{70} \( f_i^{GBC} \neq 0 \), in the seventh term of (3.21); GBC is known to have negligible effects on the F12 energies.\textsuperscript{77} This replaces the sum over \( p'_{ij} \) with a sum over \( m_{ij} \).
and thus allows to combine it with the sixth term. The final result reads:

$$B_{ij}^{\text{app}} \approx D \left( [f_{12}[\hat{T}, f_{12}]]_{ij} + \tilde{S}_{i,j} (f_{12}^2)_{p'_i j} \left( \hat{h} + 2\hat{J} \right)_{p'_i j} \right)$$

$$- \tilde{r}_{ij}^{p'_i j} K_{q_i j}^{p'_i j} r_{q_i j}^{p'_i j} - \tilde{r}_{m_i j}^{q_i j} F_{m_i j}^{q_i j} \tilde{r}_{m_i j}^{q_i j} + \tilde{r}_{m_i j}^{p'_i j} F_{m_i j}^{q_i j} \tilde{r}_{m_i j}^{q_i j}$$

$$- \tilde{r}_{m_i j}^{p'_i j} F_{m_i j}^{a_{ij} a_{ij}} \tilde{r}_{m_i j}^{a_{ij} a_{ij}} - 2 \tilde{r}_{m_i j}^{p'_i j} F_{m_i j}^{a_{ij} a_{ij}} \tilde{r}_{m_i j}^{a_{ij} a_{ij}}$$

Note that not only is the RIBS/RIBS exchange matrix gone from (3.22), but all correlation factor integrals only carry at most 1 CABS/RIBS index. This leads to significant savings when computing the 4-center \( r \) integrals in a pair-specific representation.

The basis set error of the Hartree-Fock energy is corrected perturbatively\(^7^9\) by including single excitation correction \( E_{(2)_{s}} = 2t_{A'} F_{i}^{A'} \). The singles amplitudes are determined from linear equations \( F_{A'}^{i} = t_{A'}^{k} F_{k}^{i} - F_{A'}^{B'} t_{B'}^{i} \), in which we define \( \{A'\} = \{a'\} \oplus \{a\} \). Evaluation of \( E_{(2)_{s}} \) was simplified by omitting the exchange operator contribution to the CABS/CABS block of the Fock operator. No domain approximations were utilized due to the completely negligible expense of this correction. The numerical data on the S66 data set shows the negligible effect of this approximation on relative energies.

### 3.2.1 Domain approximations

Domain approximations are heuristics for imposing block sparsity. Pulay’s idea to impose block sparsity in the two-electron correlation operator was a domain approximation and has been the foundation of local correlation methods. The combination of the truncated...
pair natural representation with the domain approximation leads to drastic computational savings in pair correlation theories relative to either PNO or domain-based (PAO) correlation methods alone, and was first demonstrated for CCSD by Neese and Riplinger.113

Our work follows the recent revision of the domain approximations as described in Ref. 26, with some additional heuristics defined to account for the additional orbital spaces appearing in the F12 methods. Note that none of these domain approximations use geometry-based heuristics unlike those used in original the PAO-based local many-body methods.

Density fitting domains

Density fitting approximation of two-electron integrals uses domains (such use of domains was already introduced in the original local PNO work21). Consider a set of 4-index Coulomb integrals specific to particular pair $ij$, e.g. $g_{ij}^{ab} = (ia_{ij}|jb_{ij})$. Its density fitting approximation is obtained by

$$
(ia_{ij}|jb_{ij}) \approx (ia_{ij}|X_{ij})(V^{-1})_{X_{ij},Y_{ij}}(jb_{ij}|Y_{ij}),
$$

(3.23)

where $V$ is the matrix of 2-index Coulomb integrals, i.e. $V_{X_{ij},Y_{ij}} \equiv (X_{ij}|Y_{ij})$, and $\{X_{ij}\}$ is a subset of functions from the density fitting (auxiliary) basis chosen to approximate the integrals for pair $ij$. This subset is defined as a union of density-fitting domains for orbitals $i$ and $j$ defined as in Ref. 21; the density fitting domain for orbital $i$ includes all density fitting basis functions on atoms that have Mulliken charges of orbital $i$ greater than threshold $T_{\text{CutMKN}}$. The default $T_{\text{CutMKN}} = 10^{-3}$ value is used, as in Ref. 26.
MO domains

PNOs for pair \(ij\) are expanded in terms of an \(ij\)-specific set of PAOs. The PAO domain for pair \(ij\) is defined as a union of PAO domains for orbitals \(i\) and \(j\). Orbital PAO domains are defined as in Ref. 26 using the differential overlap integral (DOI), which for 2 orbitals \(\phi_p\) and \(\phi_q\) is defined by:

\[
doi(\phi_p, \phi_q) \equiv \sqrt{\int d|\phi_p(x)|^2|\phi_q(x)|^2}. \tag{3.24}
\]

DOIs is a convenient “distance” measure in Hilbert space that avoids the pitfalls of the standard overlap and, unlike the Schwartz Coulomb integrals, can be computed for any localized basis in linear effort using standard grid quadrature.\(^{26}\) The PAO domain for orbital \(i\) includes all PAOs \(\tilde{\mu}\) for which \(doi(i, \tilde{\mu})\) is greater than the \(T_{\text{CutDO}}\) threshold. As elaborated in Ref. 26 the default value of \(T_{\text{CutDO}}\) is \(10^{-2}\) is sufficient for even high-accuracy computations, and was adopted in this work as well. Also note that PAO domains are “atom-complete”, i.e. if PAO \(\tilde{\mu}\) generated by projection from an atomic orbital \(\mu\) is part of some domain then all PAOs generated from AOs on the same atom as \(\mu\) are included; this greatly increases robustness in chemical applications.

To arrive at the \(O(N)\) computational scaling of the MP2-F12 energy algorithm, in addition to the PAO domains that appeared in the MP2 methods we must also define domains for other orbital indices that appear in F12 intermediates, namely the occupied orbitals and the CABS/RIBS orbitals. The same DOI approach was used to define the occupied orbital domain for each orbital. However our initial tests indicated that using the same
value of $T_{\text{CutDO}}$ parameter as for PAOs introduced significant errors when high accuracy was sought, such as in computing energies of weak intermolecular interactions. To avoid introduction of another parameter we pegged the screening criterion for occupied orbitals to the $T_{\text{CutDO}}$ parameter as follows: the occupied domain for orbital $i$ includes orbitals $m$ such that $\text{doi}(i, m) \geq T_{\text{CutDO}}/10$.

Domains for CABS/RIBS orbitals are not defined using the DOI strategy, partially to avoid the cost of computing DOIs over the large atomic orbital basis that supports these orbitals. Also, instead of forming the analog of PAO for CABS atomic orbital basis we leave the indices that become CABS/RIBS untransformed at the 3-index transformation stage. The construction of the AO domains for the CABS basis is described in the next section.

**AO domains**

To compute integrals over localized MO bases with an $O(N)$ effort it is necessary to also screen contributions from the AO-basis integrals to the MO-basis integrals. In the sparse maps framework this is done by screening atom blocks in the MO coefficient matrices. E.g. the exact expansion of PAO $\bar{\mu}$ includes contribution from every atomic orbital $\nu$. Instead we approximate this expansion using atomic orbitals from only those atoms whose AOs have non-negligible contributions. This truncation is controlled by parameter $T_{\text{CutC}}$, set conservatively at $10^{-3}$. As explained in Ref. 26, this does not mean that AOs with coefficients $< T_{\text{CutC}}$ are omitted; this would result in unacceptably large errors. Within the linear-scaling transformation $(\mu\nu|X) \rightarrow (i\bar{\mu}|X)$ each occupied orbital/PAO is effectively expanded in a
much larger set of AOs than the naive truncation using $T_{\text{CutC}}$ would produce.\textsuperscript{26} Note that the same value of $T_{\text{CutC}}$ is used to screen both (standard and Fock-transformed, Eq. (3.18)) PAO and occupied orbital expansions.

Since for given orbitals $i$, CABS AOs are intended to span a larger space than the orbitals in the occupied and PAO domains of $i$, it is expected that CABS AO domains should include more atoms than the PAO domains. To define the CABS AO domains we utilized the following simple strategy: for orbital $i$ the CABS AO domain includes orbitals from those atoms whose PAOs satisfy the DOI criterion with threshold $T_{\text{CutDO}}/10$. This simple heuristic automatically ensures that the size of CABS AO domains increases proportionally to the PAO domain size, and does not require additional screening parameters. Tests of the CABS domains constructed using only atoms in PAO domains for the weakly-bonded benchmarks included in the S66 test set produced large maximum errors, on the order of 0.4 kcal/mol, which exceeds the 0.2 kcal/mol maximum basis set error of cc-pVDZ-F12 MP2-F12 binding energies; hence we made a very conservative choice of CABS domains to ensure small errors even for the most stringent tests of our method.

### 3.2.2 Pair screening

Linear complexity in two-body methods like MP2 and CCSD can only be attained if pair interactions are truncated to produce a linear number of electron pairs. In this work we use a two-level pair screening procedure. First, electron pairs are screened using the MP2 pair
screening procedure, as described in Ref. 26. Namely, only the pairs that survive either the DOI threshold test and dipole-dipole MP2 energy estimate test (controlled by parameters $T_{\text{CutDOij}}$ and $T_{\text{CutPre}}$, respectively) are kept in the MP1-F12 wave function computation. For the pairs that do not pass this screening test the energy contribution is estimated by a dipole-approximated MP2 energy formula and included in the total electronic energy. Only some pairs that survive the MP2 screening are described at the MP2-F12 level. This is reasonable since the F12 terms are designed to only treat the short-range part of the Coulomb correlation. Therefore the F12 contribution to pair correlation energies decay exponentially with the distance, compared to the $R^{-6}$ decay of the pair correlation energies themselves, thus for sufficiently weak pairs the F12 correction will become a negligible fraction of the MP2 energy. As illustrated for a 1-dimensional alkane in Figure 3.1, for pairs of orbitals whose centroids are separated by at least $10\,\text{Å}$ the F12 pair energy contribution becomes $< 10\%$ of the total MP2-F12 pair energy. Furthermore, the F12 pair contributions become negligible beyond $15\,\text{Å}$. However, the centroid distance alone is a poor predictor of the importance of the F12 correlation, as illustrated by the same data for a linear polyene in Figure 3.2. Because the π orbitals do not localize as efficiently, their centroids must be separated by at least $20\,\text{Å}$ for the F12 contribution to decay $< 10\%$ of the total.

The second, F12 pair screening criterion excludes pair $ij$ from the MP2-F12 treatment if its differential overlap integral, doi($i,j$), is less than threshold $T_{\text{CutDOijF12}}$. The rationale for such design is that doi($i,j$) and the F12 pair energy contributions both decay exponentially (assuming the exponential localization of the occupied orbitals) and in practice correlate
Figure 3.1: Distance decay of MP2-F12 pair energies (top) and the difference between MP2-F12 and MP2 pair energies (bottom) in n-C$_{30}$H$_{62}$ with linear geometry. The distance between orbitals $i$ and $j$ is defined as the distance between their centroids. The def2-SVP orbital basis set was used.

... as shown in Figures 3.3 and 3.4 for a linear alkane and polyene. Although it is not possible to analyze this relationship rigorously in general, it is sufficient to note that the
Figure 3.2: Distance decay of MP2-F12 pair energies (top) and the difference between MP2-F12 and MP2 pair energies (bottom) in polyene C_{30}H_{32} with linear geometry. The distance between orbitals $i$ and $j$ is defined as the distance between their centroids. The def2-SVP orbital basis set was used.

2-norm of the projector-free geminal, $f_{12} | ij \rangle$, is proportional to $d_{oi}(i, j)$. To determine the default value for $T_{\text{CutDOijF12}}$ we analyzed the energies of weak intermolecular interactions in a
subset of the L7 benchmark set, as described in Section 3.4.2. Setting $T_{\text{CutDOijF12}} = 3 \times 10^{-4}$ produced completely negligible errors in the binding energies, as well as in the absolute energies. Figures 3.3 and 3.4 suggest that this level of truncation omits F12 corrections smaller than 1 microHartree per pair. To avoid introducing the new parameter we internally peg $T_{\text{CutDOijF12}}$ to $T_{\text{CutDOij}}$ using formula $T_{\text{CutDOijF12}} = 30.0 \times T_{\text{CutDOij}}$.

![Figure 3.3: The F12 contribution to def2-SVP MP2-F12 pair energy vs. doi(i,j) for linear n-C_{30}H_{62}.](image)
The observation that the F12 pair energy decays more rapidly than the MP2 correlation energy has also been exploited recently by Ma and Werner.\textsuperscript{50} Their approach to excluding pairs from the F12 treatment is based on the MP2 pair energy criteria; only those pairs are treated at the MP2-F12 level if $E_{ij}^{MP2} > 10^{-4} E_h$. Our data suggests that this enables the F12 treatment of pairs with the F12 pair energy greater than 10 microhartrees, or at least
an order of magnitude greater cutoff than in our approach. As we show later, this can lead to significant errors in the MP2-F12 relative energies for weak intermolecular interactions.

Note that the MP2 screening scales as $O(N^2)$ whereas the cost of F12 pair screening is $O(N)$. The computational cost of the two screening procedures is completely negligible in practice.

### 3.2.3 Miscellany: PNO construction, pair coupling screening, integral transformation

The PNO construction procedure using a semicanonical amplitude guess was also fully described in Ref. 26. The alternative definition of PNOs designed specifically for F12 methods was not used in this work, since we intend to extend the DLPNO-F12 approach to the coupled-cluster with the perturbative inclusion of F12, where such choice is suboptimal.48

The PNO truncation is controlled by the $T_{\text{CutPNO}}$ parameter. For each pair the MP2 energy contribution from the omitted PNOs was estimated using the standard perturbative estimate.21,26

The pair couplings that occur via the occupied-occupied elements of the Fock matrix in the conventional part of the MP1-F12 residual are screened as in the DLPNO-MP2 method,26 namely the $F_{ij}^q$ matrix elements whose absolute magnitude is less than or equal to $F_{\text{Cut}}$ (set by default to $10^{-5}$) are set to hard zero in the evaluation of Eq. (3.10).

The linear-scaling three-index transformation described in Ref. 26 was used to compute all
three-index integrals using the sparse maps constructed with the domain approximations described in Section 3.2.1. Note that the 3-index transformations used the pair-extended PAO, CABS, and DF domains, e.g. for a given occupied orbital $i$ the transformed $(\text{occ, PAO}|\text{aux})$ integrals, $(i\tilde{\mu}_{i}|X_{i})$, included auxiliary basis functions $\{X_{i}\}$ and PAOs $\tilde{\mu}_{i}$ included in the respective $ij$ domains of all $j$ such that the $ij$ pair survived the prescreening described above.

Construction of 4-center transformed integrals using the local density fitting is completely analogous to the DLPNO-MP2 method. Evaluation of the F12 intermediates and the energy is then essentially unchanged from the standard MP2-F12 program. Aside from the addition of the coupling intermediate to residual (3.10), the amplitude equations are solved exactly as described in Ref. 26. Then the F12 contributions to the energy are evaluated according to Eq. (3.12).

### 3.3 Computational details

All computations were performed with the developmental version of the ORCA quantum chemistry package.\textsuperscript{109} The convergence of truncation parameters was tested on the same three medium-sized molecules used in Ref. 26, namely sildenafil, ATP anion and anthracene dimer. Unless noted otherwise, all computations utilized the cc-pVDZ-F12 correlation-consistent-style basis sets optimized for the explicitly correlation methods paired with the matching cc-pVDZ-F12/OptRI basis set for F12 approximations and the aug-cc-pVDZ-RI density-fitting (DF) basis, denoted for brevity as a \{cc-pVDZ-F12, cc-pVDZ-F12/OptRI,
aug-cc-pVDZ-RI basis set triplet.\textsuperscript{85,126,127}

Performance for relative energies was studied using the standard S66 benchmark\textsuperscript{81} and a subset of the newer L7 benchmark of weak intermolecular interactions.\textsuperscript{128} All computations on the S66 set were performed with the \{cc-pVDZ-F12, cc-pVDZ-F12/OptRI, aug-cc-pVTZ-RI\} basis set triplet, whereas the computations on L7 utilized the standard basis triplet.

Scaling studies of our method were performed on the linear n-alkanes with the \{def2-SVP, aug-cc-pVDZ/OptRI, def2-SVP/J, def2-SVP/C\} and \{def2-TZVP, cc-pVDZ-F12/OptRI, def2-TZVP/J, def2-TZVP/C\} basis set quartets, where the respective elements denote \{orbital, F12, Fock matrix DF, DF\} basis sets.\textsuperscript{129-131} The def2- style basis sets were used for OBS instead of the cc-pVDZ-F12 due to convergence issues in the Hartree-Fock SCF when RIJCOSX is employed in conjunction with basis sets with large metric condition numbers, such as observed for the alkane chains with 50 and more carbon atoms and the cc-pVDZ-F12 basis containing diffuse Gaussian AOs. The idealized geometry for the alkane chain was employed with the $r_{\text{C-C}} = 1.55\text{Å}$, $r_{\text{C-H}} = 1.09\text{Å}$ with all angles 109.4712°. To allow comparison of our code’s performance against that of the $\mathcal{O}(N^3)$ PNO MP2-F12 method of Schmitz et al.\textsuperscript{46} we performed computations on a 32-oligomer glycine polypeptide with the default basis set triplet.

All computations utilized a correlation factor (Eq. (3.4)) composed of six Gaussian geminals. This combination is a 6-Gaussian fit of the exponential correlation factor of Ten-no,\textsuperscript{37} $f_{12} = -\exp(-\gamma r_{12})/\gamma$ with $\gamma = 0.9\text{Bohr}^{-1}$ that was recommended for the cc-pVDZ-F12 basis set; the Gaussian geminal exponents and coefficients were \{245.3722, 44.2276, 11.7050, 3.4610,
0.9387, 0.1948 and {-0.03794, -0.0614, -0.1006, -0.1754, -0.3334, -0.3838}, respectively.\textsuperscript{77,86}

All computations utilized the frozen-core approximation. Occupied orbitals were obtained with the Foster-Boys localization procedure.\textsuperscript{132,133}

3.4 Results

3.4.1 Selection of truncation parameters

As described in Section 3.2, the DLPNO-MP2-F12 method involves several approximations whose accuracy is controlled by the corresponding truncation parameters. As these truncation parameters approach zero, the MP2-F12 energy approaches the value produced by the corresponding conventional method. Our goal for the linear-scaling DLPNO wave function methods is, at a minimum, to serve as a black-box replacements to their canonical counterparts for the purposes of computing accurate relative energies in molecular processes. Chemical accuracy in the energy is usually defined as 1 kcal/mol, and many applications will require higher accuracy (the mainstream Kohn-Sham DFT is generally incapable of attaining this level of accuracy even in relatively simple scenarios). Hence we want the DLPNO approximations to affect relative energies by at most 1 kcal/mol, and in practice by much less than that.

It is of course impossible to universally relate the accuracy in relative energies to the required accuracy in absolute energies, since the latter is size-extensive (i.e. grows linearly with the
system size), whereas the relative energies of interest to chemistry are typically — but not always — size-intensive. Our experience in chemical application of the PNO methods suggests that the approximations due to PNO and domain approximations should not affect the absolute correlation energy by more than 0.1 \%.\textsuperscript{113} This establishes the target accuracy in the absolute correlation energy of 99.9\%. Thus we will first investigate which values of truncation parameters attain this level of accuracy. For our analysis of the DLPNO approximations, the same three medium-sized molecules were selected as in Ref. 26, namely, sildenafil (63 atoms), ATP anion (46 atoms), and anthracene dimer (48 atoms).

**Convergence with respect to \( T_{\text{CutPNO}} \)**

Figure 3.5 depicts the convergence of the DLPNO-MP2 correlation energy and the F12 correction of DLPNO-MP2-F12, expressed as percentage of the corresponding canonical MP2 and F12 values, vs. the PNO truncation parameter \( T_{\text{CutPNO}} \). All other parameters were set to 0, hence the DLPNO correlation energies approach 100 \% in the limit \( T_{\text{CutPNO}} \to 0 \). The PNO truncation error of the DLPNO-MP2 correlation energy decays more rapidly than the DLPNO-F12 correction, in agreement with our experience with the LPNO-CCSD-F12 approach.\textsuperscript{48} Thus for \( T_{\text{CutPNO}} = 10^{-8} \), the DLPNO-MP2 method recovers more than 99.9 \% of the canonical MP2 correlation energy for all three systems, while in the case of the F12 contribution the percentage is above 99.5 \%. This relatively low percentage of the recovered F12 contribution compared to the MP2 correlation energy is due to fact that we have used PNOs to construct projector Eq. (3.5), which is not optimal. However, since the F12...
correction is a relative small fraction of the total energy; the truncation error of the DLPNO-MP2-F12 correlation energy is insignificantly greater than that of DLPNO-MP2, as shown separately in Figure 3.6. Both DLPNO-MP2 and DLPNO-MP2-F12 correlation energies converge with $T_{\text{CutPNO}}$ monotonically from below. With $T_{\text{CutPNO}} = 10^{-8}$ approximately 99.88% of the canonical MP2-F12 energy is recovered, which is acceptable. Thereby we set the default value for $T_{\text{CutPNO}}$ parameter at $10^{-8}$. For most chemical applications this should be sufficient, however for particularly challenging cases of weak intermolecular interaction energies smaller values of $T_{\text{CutPNO}}$ may be necessary, as pointed out later.

Note that both DLPNO-MP2 and DLPNO-MP2-F12 energies include the perturbative semicanonical energy correction\textsuperscript{21,26} which accounts for the MP2 energy contribution to the PNO truncation. As figure 3.7 demonstrates, the semicanonical correction can be substantial for loose PNO truncations ($T_{\text{CutPNO}} = 3 \times 10^{-7}$), and in all cases it reduces the error of the MP2-F12 correlation energy.

Convergence with respect to $T_{\text{CutMKN}}$

Figure 3.8 shows the variation of the MP2 and MP2-F12 correlation energies with the $T_{\text{CutMKN}}$ parameter that controls the construction of density fitting domains (see Section 3.2.1); the energies are expressed as a percentage of the DLPNO energy obtained with all parameters other than $T_{\text{CutPNO}}$ and $T_{\text{CutMKN}}$ set to 0, hence the correlation energies approach 100% in the limit $T_{\text{CutMKN}} \rightarrow 0$. 
Figure 3.5: Convergence of the DLPNO-MP2 correlation energy and F12 part of the DLPNO-MP2-F12 energy, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutPNO}}$. See the text for more details.

It is evident that the error due to the domain fitting of the MP2 correlation energy converges rapidly, and 99.99% of the canonical result has been recovered with $T_{\text{CutMKN}} = 10^{-3}$. Inclusion of the explicit correlation further reduces the error introduced by selection of domains. This suggests that more aggressive values of $T_{\text{CutMKN}}$, e.g. $10^{-2}$ are sufficient. Nevertheless, to be conservative, we set the default for $T_{\text{CutMKN}}$ at the same $10^{-3}$ value as used by default in the DLPNO-MP2 and CCSD methods.
Figure 3.6: Convergence of the DLPNO-MP2 and DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutPNO}}$. See the text for more details.

**Convergence with respect to $T_{\text{CutDO}}$**

Figure 3.9 shows the variation of the MP2 and MP2-F12 correlation energies with the $T_{\text{CutDO}}$ parameter that controls the construction of PAO and CABS domains as discussed in Section 3.2.1. The energies are reported as the percentage of the energy obtained with $T_{\text{CutPNO}}$ and $T_{\text{CutMKN}}$ set to their respective defaults and all other parameters set to 0.
Figure 3.7: Convergence of the DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutPNO}}$. “(-SC)” indicates the omission of the semicanonical PNO truncation correction.

First, note that the inclusion of explicit correlation greatly reduces the PAO domain error. This is in accordance with the observation previously made by Werner and co-workers\textsuperscript{123} that the domain error of PAO-based local MP2 method is greatly reduced by the introduction of the MP2-F12 terms. The rationale put forth by Werner and co-workers was that the F12 terms in the first-order wave function account for the parts of the wave function that were set to 0 by the domain approximation. However, as our recent work showed,\textsuperscript{48} this conjecture
Figure 3.8: Convergence of the DLPNO-MP2 and DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutMKN}}$. See the text for more details.

can not hold universally and must be an artifact of the PAO domain approximation. Namely we observed that “zeroing-out” of components of the conventional wave function by the PNO approximation is not cancelled by the F12 terms (the data in Figure 3.6 supports this as well). Indeed, PAO and PNO approximations modify the conventional part of the first-order wave function in (3.1), but they also modify the pair-specific geminal projector in (3.5) and hence the F12 part of the wave function. The two changes do not necessarily cancel out;
they appear to cancel out fairly well when a PAO domain approximation is involved, but not for the PNO approximation alone.

For the default value used in the DLPNO-MP2 method, $T_{\text{CutDO}} = 10^{-2}$, close to 100% of the MP2 and MP2-F12 energies are recovered, hence this default is used conservatively to parametrize both methods even though additional savings are possible in the MP2-F12 method.

Figure 3.10 reports the average size of PAO and CABS pair domains for sildenafil and anthracene dimer as a function of $T_{\text{CutDO}}$. The CABS pair domains are obviously larger than the PAO domains; for the default value of $T_{\text{CutDO}}$ the PAO and CABS domains span on average 30 and 50 atoms, respectively.

**Summary of truncation threshold analysis**

Table 3.1 reports the percentage of canonical MP2 and F12 energy contributions recovered with the DLPNO-MP2-F12 methods with the default thresholds for the three investigated systems. As already noted, the F12 correction alone converges slower with respect to the PNO truncation than the MP2 energy, however due to the relatively small percentage of the energy accounted for by the F12 correction the MP2-F12 energy is almost as precise as MP2. The use of perturbative semicanonical correction for the PNO truncation$^{21,26}$ decreases the error of MP2 and MP2-F12 energies and thus is enabled by default.
Figure 3.9: Convergence of the DLPNO-MP2 and DLPNO-MP2-F12 correlation energies, expressed as percentage of their canonical values, with truncation parameter $T_{\text{CutDO}}$. $T_{\text{CutPNO}}$ and $T_{\text{CutMKN}}$ were set to $10^{-8}$ and $10^{-3}$, respectively, with the rest of truncation parameters set to 0.

### 3.4.2 Accuracy of relative errors

To understand how the small errors in absolute energies induced by the DLPNO approximation affect relative energy differences we challenged the DLPNO-MP2-F12 method by computing the binding energies of the S66 benchmark set of prototypical intermolecular interactions.\textsuperscript{81} It is well known that the energetics of noncovalent interactions are highly
Figure 3.10: The average size of PAO/CABS atomic domains as a function of parameter $T_{\text{CutDO}}$ for the sildenafil and anthracene dimer. The total number of atoms in the sildenafil and anthracene dimer are 63 and 48, respectively.

sensitive to ad hoc approximations.\textsuperscript{20,21,46,48} We computed the errors of the DLPNO-MP2-F12 method relative to the canonical MP2-F12 method in conjunction with the cc-pVDZ-F12 basis set. All binding energies were computed without the counterpoise correction; in agreement with Ref. 134 we find that with the cc-pVXZ-F12 basis sets the counterpoise correction can make the binding energies worse.
Table 3.1: Components of the correlation energy, expressed as percentage of the canonical values, computed with the DLPNO-MP2 and DLPNO-MP2-F12 methods for sildenafil, ATP anion and anthracene dimer.

<table>
<thead>
<tr>
<th>Method</th>
<th>sildenafil</th>
<th>ATP anion</th>
<th>anth. dimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLPNO-MP2 (-SC)(^a)</td>
<td>99.88</td>
<td>99.92</td>
<td>99.87</td>
</tr>
<tr>
<td>DLPNO-MP2</td>
<td>99.91</td>
<td>99.94</td>
<td>99.90</td>
</tr>
<tr>
<td>F12 (DLPNO-MP2-F12)(^b)</td>
<td>99.74</td>
<td>99.64</td>
<td>99.80</td>
</tr>
<tr>
<td>DLPNO-MP2-F12 (-SC)(^a)</td>
<td>99.86</td>
<td>99.87</td>
<td>99.86</td>
</tr>
<tr>
<td>DLPNO-MP2-F12</td>
<td>99.88</td>
<td>99.89</td>
<td>99.88</td>
</tr>
</tbody>
</table>

\(^a\) With the semicanonical correction for PNO truncation\(^{21,26}\) omitted.

\(^b\) The F12 correction only.

First, we tested how the use of approach D for the F12 intermediate \(B\), Eq. (3.22), differs from approach C,\(^{72}\) (3.21), which is currently the default method in ORCA. The average error introduced by approach D is only 0.001 kcal mol\(^{-1}\), with a standard deviation of 0.01 kcal mol\(^{-1}\) and a maximum absolute error of 0.028 kcal mol\(^{-1}\). Approach D was also found to have a negligible effect on atomization energies and reaction barriers (benchmark sets HEAT\(^{63}\) and DBH24/08,\(^{135}\) respectively) when used in the context of the canonical (non-DLPNO) MP2-F12 method.

Next we used the DLPNO-MP2-F12 method to compute the binding energies of the S66 data set with the default threshold parameters: \(T_{\text{CutPNO}} = 10^{-8}\), \(T_{\text{CutMKN}} = 10^{-3}\), \(T_{\text{CutDO}} = 10^{-2}\), \(F_{\text{Cut}} = 10^{-5}\), \(T_{\text{CutPre}} = 10^{-6}\), and \(T_{\text{CutC}} = 10^{-3}\). The error statistics are reported in
the first data row of Table 3.2. As it can be seen from the table, the average error is 0.058 kcal mol\(^{-1}\), with a standard deviation of 0.038 kcal mol\(^{-1}\) and a maximum absolute error of 0.185 kcal mol\(^{-1}\). The maximum absolute error occurs for the π...π uracil dimer.

Table 3.2: Influence of truncation parameters of the DLPNO-MP2-F12 method on the error statistics for the S66 benchmark set. The errors (kcal/mol) are relative to the conventional MP2-F12 energies. \(\bar{\Delta} =\) mean error, \(\Delta_{std} =\) standard deviation, \(\Delta_{max} =\) maximum absolute error. The first column shows the differences relative to the default parameter values.

<table>
<thead>
<tr>
<th>parameter</th>
<th>(\bar{\Delta})</th>
<th>(\Delta_{std})</th>
<th>(\Delta_{max})</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>0.058</td>
<td>0.038</td>
<td>0.185</td>
</tr>
<tr>
<td>(T_{\text{CutPNO}} = 10^{-9})</td>
<td>0.031</td>
<td>0.016</td>
<td>0.085</td>
</tr>
<tr>
<td>(T_{\text{CutPNO}} = 10^{-10})</td>
<td>0.013</td>
<td>0.008</td>
<td>0.046</td>
</tr>
<tr>
<td>(T_{\text{CutMKN}} = 10^{-4})</td>
<td>0.058</td>
<td>0.039</td>
<td>0.206</td>
</tr>
<tr>
<td>(T_{\text{CutDO}} = 10^{-1})</td>
<td>0.436</td>
<td>0.394</td>
<td>1.470</td>
</tr>
<tr>
<td>(T_{\text{CutDO}} = 10^{-3})</td>
<td>0.058</td>
<td>0.038</td>
<td>0.195</td>
</tr>
<tr>
<td>(F_{\text{Cut}} = 10^{-4})</td>
<td>0.058</td>
<td>0.038</td>
<td>0.185</td>
</tr>
<tr>
<td>(F_{\text{Cut}} = 10^{-6})</td>
<td>0.058</td>
<td>0.038</td>
<td>0.185</td>
</tr>
<tr>
<td>(T_{\text{CutC}} = 10^{-2})</td>
<td>0.058</td>
<td>0.038</td>
<td>0.185</td>
</tr>
<tr>
<td>(T_{\text{CutC}} = 10^{-4})</td>
<td>0.058</td>
<td>0.038</td>
<td>0.185</td>
</tr>
<tr>
<td>(T_{\text{CutPre}} = 10^{-7})</td>
<td>0.058</td>
<td>0.038</td>
<td>0.185</td>
</tr>
</tbody>
</table>

As the rest of data in Table 3.2 indicates, the main source of the mean and max errors is
the PNO truncation: tightening other parameters does not lead to significant changes in the error statistics. By using $T_{\text{CutPNO}} = 10^{-9}$ the maximum absolute error has been reduced to $0.085 \text{ kcal mol}^{-1}$; using $T_{\text{CutPNO}} = 10^{-10}$ reduces it further to only $0.046 \text{ kcal mol}^{-1}$.

Lastly, note that the binding energies are very sensitive to the parameter $T_{\text{CutDO}}$ that controls the size of PAO and CABS domains, e.g. setting $T_{\text{CutDO}} = 10^{-1}$ results in a maximum error of $1.470 \text{ kcal mol}^{-1}$, whereas setting $T_{\text{CutDO}}$ to $10^{-2}$ (default) and $10^{-3}$ results in much smaller errors.

We have also tested the DLPNO-MP2-F12 method with the default truncation parameters on the subset of 5 molecules in the L7 data set of intermolecular interaction benchmarks that includes significantly larger systems. The results are summarized in the Table 3.3. The binding energies were compared to the MP2 binding energies at the complete basis set (CBS) limit estimated by a basis set extrapolation from the double and triple-zeta MP2 energies. As can be seen from the Table, the maximum absolute error with respect to the reference values is $0.9 \text{ kcal mol}^{-1}$. The magnitude of the error is larger than we expected, and could potentially be due to the inaccuracy of the CBS MP2 limit estimate by Sedlak et al.

Note that the Table 3.3 demonstrates the effect of F12 pair prescreening on the binding energies. The pair screening leads to completely negligible errors which do not exceed $0.04 \text{ kcal mol}^{-1}$. For reference, our tests of the more aggressive F12 pair screening procedure of Ma and Werner$^{50}$ produced much larger errors for this set of energies, as large as $1 \text{ kcal/mol}$ for the binding energy of the GCGC base pair stack.
Table 3.3: The basis set error of the binding energy of the DLPNO-MP2-F12 for a subset of the L7 data set relative to the error due to the F12 pair screening procedure. The basis set error (kcal/mol) is defined relative to the complete basis set MP2 estimate. The F12 pair screening error (kcal/mol) is relative to the binding energy obtained without omitting any of the pairs from the F12 treatment.

<table>
<thead>
<tr>
<th>system</th>
<th>basis set error</th>
<th>F12 pair screening error</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronene dimer</td>
<td>0.35</td>
<td>0.02</td>
</tr>
<tr>
<td>GCGC base pair stack</td>
<td>-0.90</td>
<td>0.02</td>
</tr>
<tr>
<td>guanine trimer</td>
<td>-0.29</td>
<td>0.01</td>
</tr>
<tr>
<td>octadecane dimer</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>phenylalanine residues trimer</td>
<td>-0.21</td>
<td>0.02</td>
</tr>
</tbody>
</table>

3.4.3 Dependence of computational cost on system size

Figure 3.11 presents the computation times for the canonical and DLPNO-MP2-F12 methods using double- and triple-zeta basis sets for n-alkane chains. The Coulomb matrix has been approximated with the standard cubically-scaling RI approach (Coulomb matrix computation is a negligible part of the overall cost), while the exchange matrices were evaluated using the linear-scaling COSX approach with default parameters.\textsuperscript{136} All computations have been performed on one core Intel Xeon E5-2680 v3 2.50 GHz (only one core was used to simplify the performance analysis; our DLPNO-MP2-F12 code is parallelized and scales reasonably well up to 16 cores).
Evaluation of the DLPNO-MP2-F12 energy in a triple-zeta basis becomes cost competitive with the canonical counterpart already for \( n\text{-C}_{10}\text{H}_{22} \) (the computation times are 1396 and 1170 seconds, respectively); whereas in the case of the double-zeta basis set this occurs at \( n\text{-C}_{15}\text{H}_{32} \). For the larger systems the scaling behavior becomes subquadratic; true linear-scaling is not achieved due to the nonlinear cost of the computing the Fock mean-field operator. The effective exponents obtained by the finite difference from the last two data points for triple-zeta and double-zeta are 1.584 and 1.578, respectively. The largest calculation performed on a single core was on \( n\text{-C}_{200}\text{H}_{402} \) using a triple-zeta basis. The corresponding sizes of the
Table 3.4: Detailed breakdown of wall times (in seconds) for various steps of the DLPNO-MP2-F12 energy evaluation.

<table>
<thead>
<tr>
<th></th>
<th>n-C_{200}H_{402}^{a}</th>
<th>(gly)$_{32}^{b}$</th>
<th>sildenafil$^c$</th>
<th>ATP anion$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of orbital/DF/CABS AOs</td>
<td>8612/21230/34266</td>
<td>4752/8596/16779</td>
<td>1269/3086/4483</td>
<td>1065/2568/3735</td>
</tr>
<tr>
<td>Fock matrix formation</td>
<td>49908</td>
<td>101765</td>
<td>557</td>
<td>303</td>
</tr>
<tr>
<td>three-index integral transformation</td>
<td>36674</td>
<td>47278.7</td>
<td>1926</td>
<td>1132</td>
</tr>
<tr>
<td>pair-specific 1-e matrices</td>
<td>15814</td>
<td>67105</td>
<td>5709</td>
<td>3490</td>
</tr>
<tr>
<td>local metric evaluation</td>
<td>767</td>
<td>885</td>
<td>1299</td>
<td>685</td>
</tr>
<tr>
<td>2-e integrals generation</td>
<td>12702</td>
<td>15821</td>
<td>5960</td>
<td>1645</td>
</tr>
<tr>
<td>construction of pair-specific CABS</td>
<td>24488</td>
<td>76888</td>
<td>5353</td>
<td>2782</td>
</tr>
<tr>
<td>evaluation of $V$</td>
<td>56</td>
<td>47</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>evaluation of $X$</td>
<td>45</td>
<td>52</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>evaluation of $B$</td>
<td>1777</td>
<td>3314</td>
<td>427</td>
<td>258</td>
</tr>
<tr>
<td>MP2-F12(total)</td>
<td>179096</td>
<td>331807</td>
<td>24290</td>
<td>12172</td>
</tr>
</tbody>
</table>

$^a$ The \{def2-TZVP, def2-TZVP/J, def2-TZVP/C, cc-pVDZ-F12/OptRI\} basis sets, 1 core.

$^b$ The \{cc-pVDZ-F12, cc-pVDZ-RI, cc-pVDZ-F12/OptRI\} basis sets, 1 core.

$^c$ The \{cc-pVDZ-F12, aug-cc-pVDZ-RI, cc-pVDZ-F12/OptRI\} basis sets, 8 cores.

orbital, density fitting, and CABS AO basis sets were 8612, 21230 and 34266 basis functions, respectively. A detailed breakdown of the DLPNO-MP2-F12 energy computation for this and some representative smaller computations are given in Table 3.4.

The bulk of effort involved in the MP2-F12 energy computation is spent in the F12-specific parts of the code; for n-C$_{200}$H$_{402}$ this comprises 85.3% of the total wall time. Some of the increase relative to the DLPNO-MP2 energy is due to the need to compute two-electron
integrals of various classes over 4 additional kernels: $f_{12}, f^2_{12}, f_{12}/r_{12}$, and $[f_{12}, \hat{T}, f_{12}]$. The density fitting approximation reduces these to 12 types of three-index two-electron integrals (5 kernels × 2 classes, one with and one without CABS spaces + 2 integrals types over the $f_{12}$ kernel involving the Fock-transformed occupied space); compare this to the single type of three-index two-electron integrals needed in the MP2 method. Moreover, the integrals that involve CABS indices are significantly more expensive. Thus the DLPNO-MP2-F12 energy is usually $> 10$ times more expensive than its MP2 counterpart. However, a more appropriate comparison should be with the cost of the $O(N)$ DLPNO-CCSD(T) method recently tested in Ref. 27. In the def2-TZVP basis the DLPNO-CCSD(T) energy of C$_{200}$H$_{402}$ can be obtained in 20 hours, which after accounting for the MP1-F12 equation cost means that the evaluation of the F12 correction (42 hours) takes about twice as long as the DLPNO-CCSD(T) energy for this system. Furthermore, there are additional optimization opportunities not exploited in this work since our focus here was on attaining linear-scaling for all non-trivial steps with as few approximations as possible. In the future we will use the method developed here as the benchmark for more approximate but faster variants of the method.

The DLPNO-MP2-F12 computation on the 32-oligomer glycine polypeptide was performed on one core with the cc-pVDZ-F12, aug-cc-pVDZ/RI and cc-pVDZ/OptRI AO basis sets. The Coulomb and exchange matrices were computed with the standard RI approach (with cubic and quartic cost, respectively). As the data in Table 3.4 indicates, the most time in this computation is spent in the construction of the Fock matrix due to quartic scaling of the exchange matrix calculation within RI approach. Nevertheless, the obtained timings
are favorable to those obtained with the $\mathcal{O}(N^3)$ PNO-MP2-F12 approach by Schmitz et al. in 2014;\cite{Schmitz2014} Figure 1 in their paper indicates that the MP2-F12 energy took more than $10^6$ seconds on a comparable CPU core.

Quasi-1-dimensional systems exhibit the earliest onset of computational savings from sparsity. However, in practice we are more interested in three dimensional systems. The timings of the F12-specific part of the code for the sildenafil and ATP anion is shown in the Table 3.4. Calculations on these two systems have been done with 8 cores Intel Xeon E5-2650 v2 2.60 GHz and with 64 GB of memory. The AO basis sets that have been employed are cc-pVDZ-F12, aug-cc-pVDZ-RI and cc-pVDZ-F12/OptRI. The Coulomb and exchange matrices were approximated with the standard RI approach. Both systems are of similar size with around a thousand basis functions.

\section{Summary and Perspective}

We presented a formulation of the explicitly-correlated second-order Møller-Plesset (MP2-F12) energy in which all nontrivial post-mean-field steps were formulated such that the computational complexity is linear in system size. The two key ideas are the use of pair-natural orbitals for compact representation of wave function amplitudes in conjunction with the domain approximation. This development utilizes the concept of sparse representation of tensors described in the context of the DLPNO-MP2 method by Neese and co-workers.\cite{Neese2015} Novel developments reported here include the use of domains not only for the projected
atomic orbitals, but also for the complementary auxiliary basis set used to express the intermediates of the F12 theory, and a simplification of intermediate $B$ of the F12 theory that avoids computation of four-index two-electron integrals that involve two CABS indices.

It is useful to compare our DLPNO-MP2-F12 method with the existing reduced-scaling MP2-F12 approaches. The almost-linear-scaling PAO-based MP2-F12 method of Werner and co-workers$^{123,125}$ used the less stringent approach A for evaluation of F12 intermediate $B$, the assumption of Extended Brillouin Condition (EBC),$^{70}$ which eliminates the coupling intermediate from Eq. (3.10), and the use of the ABS$^{73}$ rather than CABS formalism for the F12 terms. The very recent work on PNO-based MP2 approach Werner and co-workers also utilized the same F12 formulation.$^{50}$ Hättig, Tew, and co-workers have developed PNO-based MP2-F12 method with $O(N^3)$ scaling that utilized the more robust approach B without the assumption of EBC, and CABS formalism.$^{46}$ Our method has $O(N)$ cost and utilizes the state-of-the-art F12 formalism similar to that of Hättig et al. Although the more approximate F12 methods are likely to be acceptable for many applications, our experience shows that the energetics of intermolecular interactions can be very sensitive to the details of local approximations in F12 methods$^{48}$ and hence we conservatively aim for the more complete formulations.

For 1-dimensional systems (n-alkanes) the $O(N)$ DLPNO-MP2-F12 method becomes less expensive than the conventional $O(N^5)$ MP2-F12 for $n$ between 10 and 15, depending on the basis. Although its cost is significantly higher than that of the DLPNO-MP2 method, the cost increase is offset by the qualitative reduction of the basis set error by the explicitly
correlated terms. Furthermore, for the largest example considered here, linear \( \text{C}_{200}\text{H}_{400} \) in def2-TZVP basis, the computational cost of the F12 correction in DLPNO-MP2-F12 is only twice more expensive than that of the DLPNO-CCSD(T) reported recently by Neese and co-workers.\(^{27}\) Although the computational complexity is already formally linear, additional cost prefactor decreases should still be possible. These can come in the form of the use of a priori compression of the CABS AO domains along the lines of the ideas of the orbital specific auxiliary orbitals explored in the context of cubically-scaling explicitly correlated MP2 and MP3 methods by Hättig et al.\(^ {44}\) Finally, the next step is the development of the coupled-cluster F12 method based on this work and the linear-scaling DLPNO-CCSD(T) reported recently.\(^ {27}\) The work towards this goal is in progress.
Chapter 4

SparseMaps – A systematic infrastructure for reduced-scaling electronic structure methods. V.
Linear-scaling explicitly correlated coupled-cluster method with pair natural orbitals.

Reproduced from Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. The Journal of Chemical Physics 2017, 146(17), 174108., with the permission of AIP
4.1 Introduction

Recent advances in reduced-scaling electronic structure have dramatically extended the reach of chemically accurate methods like the coupled-cluster singles, doubles, and perturbative triples. Whereas the conventional state-of-the-art formulations of the CCSD(T) can be applied to systems with a few dozen atoms on a supercomputer, a few hundred atoms can be treated on a single multicore computer using the recently introduced linear-scaling formulation of the CCSD(T). To make a chemically accurate description of relative energies in systems of this size, it is mandatory to reduce or eliminate the uncertainty due to the basis set. To this end we demonstrated a linear-scaling MP2-F12 using the domain-based local pair natural orbital formulation. This work presents the extension of these ideas to develop an explicitly correlated coupled-cluster formalism with perturbative treatment of the geminal terms and state-of-the-art F12 formalism.

The current work is related to a number of recent developments of reduced-scaling explicitly correlated coupled-cluster methods. Werner and Adler were the first to introduce linear-scaling explicitly correlated coupled-cluster methods following the local correlation formalism of Pulay and Sæbo based on the use of localized occupied orbitals and projected atomic orbitals (PAO) for the unoccupied space, together with geometric locality criteria. The PAO-based formulations of many-body theories unfortunately have late crossover with...
the conventional full scaling formulations if a realistic precision is desired, due to the need to have large 20+ atom domains (i.e., hundreds of unoccupied orbitals) for each pair of electrons. The use of pair natural orbitals (PNO)\textsuperscript{17} for the unoccupied space leads to a drastically more compact representation of the doubles amplitudes, as shown in Refs. 20, 21. Werner and Krause compared several choices of the unoccupied spaces (PAOs, PNOs, and orbital-specific virtuals (OSV)\textsuperscript{22}) in the context of explicitly correlated CCSD.\textsuperscript{42} Several groups have been pursuing efficient production-quality reduced-scaling explicitly correlated coupled-cluster methods. Tew and Hättig developed polynomial scaling PNO-based formulations of MP2-F12, MP3-F12 and explicitly correlated CCSD.\textsuperscript{43-46} Werner and coworkers recently developed an efficient parallel implementation of a linear-scaling PNO-based MP2-F12 method\textsuperscript{50} and are currently working on an explicitly correlated coupled-cluster method. Valeev and co-workers presented a production level implementation of the polynomial scaling PNO-based explicitly correlated CCSD.\textsuperscript{48} Recently we also presented a linear-scaling reformulation of explicitly correlated MP2-F12 using the \textit{SparseMaps} sparse tensor infrastructure.\textsuperscript{29} This paper reports such an extension for the explicitly correlated CCSD(T) method.

We structure this paper as follows: in Sec. 4.2 we describe the theoretical background of linear-scaling explicitly correlated CCSD(T) method, followed by computational details given in Sec. 4.3. The convergence of the correlation energy to the canonical value with respect to truncation parameters, assessment of the approximations on relevant chemical systems and computational scaling with system size will be explained in Sec. 4.4. Finally,
in Sec. 4.5, we summarize our findings.

### 4.2 Formalism

In the CCSD(T)\(_{\text{F12}}\) method, a perturbative explicitly correlated extension of the coupled-cluster singles and doubles with perturbative triples\(^{88,98,99}\), the energy expression includes four contributions:

\[
E_{\text{CCSD(T)}_{\text{F12}}} = E_{\text{CCSD}} + E_{(T)} + E_{(2)s} + E_{(2)_{\text{F12}}} \tag{4.1}
\]

In addition to the standard CCSD energy and the standard \((T)\) correction of the CCSD(T) method, it includes the second-order correction due to the incompleteness of the Hartree-Fock reference, \((2)s\), and the simplified explicitly correlated second-order correction, \((2)_{\text{F12}}\). Evaluation of the first two terms within the linear-scaling DLPNO framework has been described in the second paper in this series\(^{27}\), and efficient evaluation of the third term and most of the fourth term were presented as part of the DLPNO-MP2-F12 method in the fourth paper in the series\(^{29}\). The focus of this work is computation of the coupled-cluster specific contributions to the last term, \(E_{(2)_{\text{F12}}}\), in a linear-scaling fashion.

In the traditional form of the coupled-cluster method the (right) CCSD state is obtained from a single-determinant reference by the action of an exponentiated cluster operator \(\hat{T}\):

\[
|0\rangle = \exp (\hat{T}) |0\rangle \tag{4.2}
\]
In the DLPNO-CCSD approach for closed-shell systems $\hat{T}$ is defined as

$$\hat{T} \equiv \hat{T}_1 + \hat{T}_2 = \epsilon_{a_i} E_{a_i} + \frac{1}{2} \epsilon_{a_i b_j} E_{a_i b_j}, \quad (4.3)$$

in which the spin-free one- and two-particle replacers $E^p_{pq}$ and $E^{pq}_{rs}$ are defined in terms of spin-orbital creators $\hat{a}^\dagger$ and annihilators $\hat{a}$ as

$$E^p_{pq} \equiv \sum_{\sigma=\uparrow, \downarrow} \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\sigma} \quad (4.4)$$

$$E^{pq}_{rs} \equiv \sum_{\sigma, \rho=\uparrow, \downarrow} \hat{a}_{p\sigma}^{\dagger} \hat{a}_{q\rho}^{\dagger} \hat{a}_{s\rho} \hat{a}_{r\sigma} \quad (4.5)$$

where $p_\sigma$ denotes the spin-orbital corresponding to spatial orbital $p$ and magnetic quantum number $\sigma$. Unlike the traditional CCSD operator, the singles and doubles components of Eq. (4.3) are defined in terms of the orbital- and pair-specific unoccupied indices $\{a_i\}$ and $\{a_{ij}\}$, respectively.$^A$

In the DLPNO-CCSD method orbital- and pair-specific spaces $\{a_i\}$ and $\{a_{ij}\}$ are defined as the truncated PNOs for pairs $ii$ and $ij$ (more details can be found in Section 4.2.1).

The second-order explicitly correlated correction to CCSD is evaluated as a second-order Hylleraas functional obtained from Löwdin perturbation expansion$^{108}$ of the untruncated CCSD-F12 energy$^{91,94}$ with respect to the left ($\langle 0 |$) and right ($| 0 \rangle$) CCSD states:

$$E_{(2)F12} = \langle 1_{F12} | \hat{H}^{(0)} | 1_{F12} \rangle + \langle 0 | \hat{H}^{(1)} | 1_{F12} \rangle + \langle 1_{F12} | \hat{H}^{(1)} | 0 \rangle. \quad (4.6)$$

By defining the normal ordered Fock operator, $\hat{F}_N \equiv \hat{F} - \langle 0 | \hat{F} | 0 \rangle$, as the zeroth-order Hamiltonian we ensure that the standard triples and the explicitly correlated doubles corrections to

$^A$As usual, $i, j$ and $a, b$ denote active occupied and unoccupied orbitals, $m, n$ denote all occupied orbitals, and $p, q$ refer to the full orbital basis set used to compute the Hartree-Fock state.
the CCSD energy are not coupled in Eq. (4.1).\(^{100}\) (recall that the closed shell Fock operator
\(\hat{F} \equiv \hat{h} + 2\hat{J} - \hat{K}\)). The first-order Hamiltonian in Eq. (4.6) is the CCSD similarity trans-
formed normal ordered Hamiltonian, \(H_N = \exp (-\hat{T})\hat{H}_N \exp (\hat{T})\), with \(\hat{H}_N \equiv \hat{H} - \langle 0 | \hat{H} | 0 \rangle\).

The explicitly correlated first-order wave function in Eq. (4.6) is defined for a closed-shell state as

\[ |1_{F12}\rangle \equiv \frac{1}{2} \tilde{R}_{ij}^\alpha |0\rangle . \tag{4.7} \]

where \(\tilde{R}_{ij}^\alpha\) is obtained from \(R_{ij}^\alpha\), the matrix element of the geminal correlation factor \(f(r_{12}).\(^{37}\)

\[ R_{ij}^\alpha \equiv \langle \alpha | \hat{Q}_{ij} f(r_{12}) | ij \rangle . \tag{4.8} \]

The tildeized tensors include the pair-spin projection in accordance with the singlet and
triplet cusp conditions.\(^{38,88}\)

\[ \hat{O}_{ij}^\alpha \equiv \frac{1}{2} (C_0 + C_1) O_{ij}^\alpha + \frac{1}{2} (C_0 - C_1) O_{ij}^i \]

\[ = \frac{3}{8} O_{ij}^\alpha + \frac{1}{8} O_{ij}^i \tag{4.9} \]

with \(C_{0,1} = 1/2, 1/4\) being the cusp coefficients for singlet and triplet pairs, respectively.\(^{69,122}\)

Projector \(\hat{Q}_{ij}\) in Eq. (4.8) ensures that the geminal functions are orthogonal to Hartree-Fock
as well to the standard coupled-cluster single and double excitations.\(^{123}\)

\[ \hat{Q}_{ij} = 1 - \sum_{a_i b_j} |a_i b_j\rangle \langle a_i b_j| ; \tag{4.10} \]

the alternative pair-specific projector definition explored previously\(^{48}\) is not used in this
work.
Exact evaluation of Eq. (4.6) contains a number of contributions that are smaller than
the residual basis set incompleteness error (BSIE). The approximate version of the \(2\)\(_{F12}\) correction, dubbed \(2\)\(_{F12}\)\(^{88,98,99}\) has the following energy expression:

\[
E_{(2)\text{F12}} = \sum_{i<j} \left( 2V_{ij}^{ij} + B_{ij}^{ij} \right).
\] (4.11)

Intermediates \(V\) and \(B\) are defined as

\[
V_{ij}^{ij} = V_{ij} + C_{ij}^{a_i b_j} t_{a_i b_j}^{ij} + V_{ij}^{a_j b_i} t_{a_j b_i}^{ij} + V_{ij}^{a_a i} t_{a_a i}^{ij},
\] (4.12)

\[
B_{ij}^{ij} = B_{ij} - X_{ij}^{ik} F_{k}^{i} - X_{ij}^{kj} F_{k}^{j}.
\] (4.13)

where the barred quantities are obtained by the closed-shell analog of the antisymmetrization:

\[
\bar{O}_{pq}^{rs} = 2O_{rs}^{pq} - O_{qs}^{rp}.
\]

The \(t_{a_i b_j}^{ij}\) and \(t_{a_i}^{i}\) are converged CCSD amplitudes (Eq. (4.3)), while

intermediates \(V, C, B,\) and \(X\) are defined as,

\[
V_{ij}^{p_i q_i} = g_{a_{ij} b_{ij}}^{p_i q_i} \tilde{R}_{ij}^{a_{ij} b_{ij}},
\] (4.14)

\[
C_{ij}^{p_i q_i} = F_{a_{ij}}^{p_i q_i} \tilde{R}_{ij}^{a_{ij} q_i} + F_{a_{ij}}^{q_i} \tilde{R}_{ij}^{a_{ij} p_i},
\] (4.15)

\[
B_{ij}^{ij} = \tilde{R}_{a_{ij} b_{ij}}^{ij} F_{a_{ij}}^{\gamma_{ij}} \tilde{R}_{ij}^{\gamma_{ij} b_{ij}} + \tilde{R}_{a_{ij} b_{ij}}^{ij} F_{\gamma_{ij}}^{\beta_{ij}} \tilde{R}_{ij}^{\gamma_{ij} \alpha_{ij}},
\] (4.16)

\[
X_{ij}^{ik} = \tilde{R}_{a_{ik} b_{ik}}^{ik} \tilde{R}_{ij}^{a_{ij} b_{ij}}.
\] (4.17)

Due to partitioning of the Hamiltonian into zeroth-order (\(\hat{F}_N\)) and first-order (\(\hat{H}_N\)), the
intermediates \(B\) and \(X\) that form Eq. (4.13) are free of CCSD amplitudes since they are
evaluated by resolving \(\langle 1_{F12} | \hat{F}_N | 1_{F12} \rangle\) from Eq. (4.6). Furthermore, Eq. (4.12) is obtained
by resolving the contributions from Eq. (4.6) that contains normal ordered similarity trans-
formed Hamiltonian, \(\hat{H}_N\), which in turn includes single and double amplitudes. In this work
Eq. (4.13) and the first two terms in Eq. (4.12) were evaluated exactly as in the DLPNO-MP2-F12 method;\textsuperscript{29} namely, intermediates $V$, $C$ and $X$ in these terms were evaluated within the CABS approximation,\textsuperscript{39} while for intermediate $B$ we used approximation D that avoids the need for the CABS/CABS block of the exchange operator (similar approximation was invoked in the evaluation of the CABS singles correction in Eq. (4.1)).\textsuperscript{29} Thus here we will focus on the last three terms in Eq. (4.12), namely those that involve intermediate $V$ with OSV and PNO indices. Since their contributions are smaller than other terms (but not so small that they can be neglected), they are evaluated in the so-called standard approximation (SA)\textsuperscript{70} which can be viewed as a simplification of the CABS approach used to evaluate intermediate $V$ in MP2-F12. Specifically:

\begin{equation}
V_{ij}^{a_{ij}b_{ij}} \text{SA} \left( \frac{f_{12}}{r_{12}} \right)_{ij} a_{ij} b_{ij} - g_{p_{ij}q_{ij}}^{a_{ij}b_{ij}} \tilde{R}_{ij}^{p_{ij}q_{ij}} \tag{4.18}
\end{equation}

\begin{equation}
V_{ij}^{a_{ij}} \text{SA} \left( \frac{f_{12}}{r_{12}} \right)_{ij} a_{ij} - g_{p_{ij}q_{ij}}^{a_{ij}} \tilde{R}_{ij}^{p_{ij}q_{ij}} \tag{4.19}
\end{equation}

Efficient evaluation of these terms with the help of \textit{SparseMaps} infrastructure is described in section 4.2.2. To set the stage for that discussion we start by briefly recapping the sparsifying approximations invoked in the DLPNO versions of the MP2-F12\textsuperscript{29} and CCSD(T)\textsuperscript{27} methods.

### 4.2.1 Sparsifying Approximations

A number of approximations is required to arrive at a linear-scaling DLPNO-CCSD(T)\textsubscript{F12}. Since these approximations have been described in detail previously,\textsuperscript{25–27, 29, 48, 113, 138} here we present only a brief summary.
The starting point for DLPNO methods is localization of the occupied orbitals (the Foster-Boys procedure is used by default). All unoccupied orbitals (OSVs and PNOs) are expanded in terms of the redundant set of projected atomic orbitals (PAOs). The PAOs, \{\tilde{\mu}\}, are obtained by projecting out the occupied orbitals, \{i\}, from the atomic orbitals, \{\mu\}:

\[ |\tilde{\mu}\rangle = \left(1 - \sum_i |i\rangle \langle i| \right) |\mu\rangle. \quad (4.20) \]

Next, for each occupied orbital \(i\) we assign a domain (subset) of PAOs \{\tilde{\mu}_i\} that will be used to expand its OSVs as well as PNOs for pairs \(ij, \forall j\). All SparseMaps-based DLPNO methods define the domain of orbital \(i\) to include all PAOs whose differential overlap integral (DOI) with the orbital \(i\) is less than threshold \(T_{\text{CutDO}}\). The DOI between functions \(\phi_p\) and \(\phi_q\),

\[ \text{DOI}_{pq} \equiv \sqrt{\int dx |\phi_p(x)|^2 |\phi_q(x)|^2}, \quad (4.21) \]

can be viewed as a measure of proximity of their densities, \(|\phi_p|^2\) and \(|\phi_q|^2\). DOIs can be robustly computed by numerical integration with linear effort and avoid distance-based domain definitions that could become less useful when large orbital basis sets are utilized. Note that PAO domains are atom-complete, i.e., if PAO \(\tilde{\mu}\) is included in the domain of orbital \(i\) then all PAOs generated from the AOs of the atom associated with \(\tilde{\mu}\) are included in its domain. Linear-scaling evaluation of operators in terms of localized occupied MOs and PAOs is only possible by taking advantage of the sparsity of the coefficients of these LCAOs. This is achieved by neglecting contributions to each localized LCAO from the AOs outside of its AO domain. The AO domain for LCAO \(x\) includes all AOs on atom \(I\) if there
exists an AO on \( I \) whose coefficient in \( x \) has greater magnitude than \( T_{\text{CutC}} \). Note that the effective AO domains in the three-index AO→MO transformation are much larger than this definition implies due to trading off some of the sparsity for the increased block size in linear algebra (for more details see Ref. 26). The same procedure is used to exploit the sparsity of the PAO expansion coefficients in Eq. (4.20).

The DLPNO F12 methods also require domains for occupied orbitals, \( \{m_i\} \) and CABS orbitals \( \{a'_i\} \). The occupied orbital \( m \) is included in domain \( \{m_i\} \) if DOI\(_{im} \geq T_{\text{CutDO}}/10 \). CABS domain for orbital \( i \) includes orbitals from those atoms whose PAOs satisfy the DOI criterion with threshold \( T_{\text{CutDO}}/10 \). Therefore, the size of CABS AO domains increases proportionally to the PAO domain size as the \( T_{\text{CutDO}} \) parameter is tightened.

All two-electron integrals in the DLPNO methods are computed using local density fitting (LDF), e.g., Coulomb integral \( (ia_{ij}|jb_{ij}) \) is evaluated as

\[
(ia_{ij}|jb_{ij}) \overset{\text{LDF}}{\approx} (ia_{ij}|X_{ij})(V^{-1})_{X_{ij},Y_{ij}}(jb_{ij}|Y_{ij}),
\]

where \( V \) is the matrix of 2-index Coulomb integrals, i.e., \( V_{X_{ij},Y_{ij}} \equiv (X_{ij}|Y_{ij}) \), and \( \{X_{ij}\} \) is a subset of functions from the density fitting (auxiliary) basis chosen to approximate the integrals for pair \( ij \). This subset is defined as a union of density-fitting domains for orbitals \( i \) and \( j \), denoted as \( \{X_i\} \) and \( \{X_j\} \). A density-fitting domain for orbital \( i \) includes density-fitting AOs on all atoms whose Mulliken charges for orbital \( i \) are greater than threshold \( T_{\text{CutMKN}} \).

Linear-scaling in correlated methods is straightforwardly achieved by pair prescreening. The
hierarchy of pair screening approximations of the DLPNO-CCSD method\textsuperscript{27} starts by screening the full $\mathcal{O}(N^2)$ pair list to select the $\mathcal{O}(N)$ pairs to be treated at the MP2 and higher levels. Pair $i j$ is neglected if $\text{DOI}_{ij} < T_{\text{CutDO}ij}$ (default value $= 10^{-5}$) and the maximum semicanonical dipole-dipole estimate of the MP2 pair energy does not exceed $T_{\text{CutPre}}$ (default value $= T_{\text{CutPairs}}/100$; see below for the definition of $T_{\text{CutPairs}}$).\textsuperscript{26} The energy contribution from the pairs that do not pass this screening is estimated at the semicanonical MP2 level with the dipole-dipole approximation for the Coulomb integral (Eq. (17) of Ref. 26). For the surviving $\mathcal{O}(N)$ pairs the MP2 energies are estimated by using a “crude” semicanonical MP2 computation with the domains defined by scaled truncation parameters ($2T_{\text{CutDO}}$ and $10T_{\text{CutMKN}}$).

The crude semicanonical MP2 pair energy estimates are used to sort the pairs into two categories using threshold $T_{\text{CutPairs}}$. First, the pairs with the crude MP2 energy less than $T_{\text{CutPairs}}/10$ are discarded from the calculation and their crude energies are added to the weak pair energy correction. For the reduced list of surviving pairs we evaluate the LMP2 energies without the use of semicanonical approximation (i.e., iteratively) and using “full” domains; this step involves construction of PNOs and OSVs truncated according to the $T_{\text{CutPNO}}$ and $T_{\text{CutPNOSingles}}$ thresholds, respectively, using semicanonical guesses for the T2 amplitudes:

\begin{equation}
\tilde{t}_{\tilde{\mu}\tilde{\nu}} = -\frac{(i\tilde{\mu}|j\tilde{\nu})}{\epsilon_{\tilde{\mu}} + \epsilon_{\tilde{\nu}} - F_{ii} - F_{jj}} \tag{4.23}
\end{equation}

where $F_{ii}$ and $F_{jj}$ are the diagonal elements of the Fock operator in a localized occupied
basis. From the amplitudes, we construct pair densities as:

\[ D_{\text{PNO}}^{ij} = \hat{t}^{ij} t^{ij\dagger} + \hat{t}^{ij\dagger} t^{ij} \]  

(4.24)

with \( \hat{t}^{ij}_{\mu\nu} = \frac{2}{1+b_{ij}} (2t^{ij}_{\mu\nu} - t^{ij}_{\mu\nu}) \). Diagonalization of the pair density matrix gives the PNO and OSV expansion coefficients. The perturbative PNO truncation correction to the energy\(^{21}\) is included for all LMP2 pairs. Lastly, the pairs with PNO LMP2 pair energies greater than \( T_{\text{CutPairs}} \) are included in the list of pairs used to solve the CCSD equations, as well as evaluate the perturbative (T) correction, as described in Ref. \( 27 \).

The perturbative F12 correction is evaluated for only a subset of the CCSD pairs; this is similar to how the F12 pairs were treated in the DLPNO-MP2-F12 method.\(^{29}\) Specifically, the F12 correction is evaluated for pair \( ij \) if \( DOI_{ij} > 30.0 \times T_{\text{CutDOI}} \). This ensures that the F12 terms are used for only spatially-nearby pairs, for which the exponentially-decaying F12 energy components are expected to be important, as discussed in Ref. \( 29 \).

### 4.2.2 Efficient Evaluation of The Coupled-Cluster-Specific F12 Intermediates

Let us now focus only on the F12 terms in the DLPNO-CCSD(2)\(_{\text{F12}}\) energy that do not appear in the DLPNO-MP2-F12 method, namely the last three terms in Eq. (4.12).

The first of these terms, \( \bar{V}_{ij}^{a_{ij},b_{ij}} t^{ij}_{a_{ij},b_{ij}} \), is computed straightforwardly bylooping over the F12 pairs; for each pair \( ij \) intermediate \( V_{ij}^{a_{ij},b_{ij}} \) is evaluated according to Eq. (4.18) and immediately contracted with the doubles amplitudes for this pair. Note that Eq. (4.18)
involves integrals with two and four pair-specific indices. The integral $\tilde{R}_{ij}^{p_{ij}q_{ij}}$ also appeared in DLPNO-MP2-F12 and is simply reused, whereas $\left(\frac{f_{ij}}{r_{ij}}\right)^{a_{ij}b_{ij}}$ is computed analogously using local density fitting. The lone integral with four pair-specific indices, $g_{p_{ij}q_{ij}}^{a_{ij}b_{ij}}$, did not appear in DLPNO-CCSD or DLPNO-MP2-F12. Although its evaluation is similar to that of $g_{c_{ij}d_{ij}}^{a_{ij}b_{ij}}$ that appears in DLPNO-CCSD, we briefly describe the procedure.

The local density fitting approximation to $g_{p_{ij}q_{ij}}^{a_{ij}b_{ij}}$ is expressed as

$$g_{p_{ij}q_{ij}}^{a_{ij}b_{ij}} \approx (p_{ij}a_{ij}|X_{ij})(V^{-1})_{X_{ij},Y_{ij}}(q_{ij}b_{ij}|Y_{ij}).$$  \hspace{1cm} (4.25)

Evaluation of the three-index integral $(p_{ij}a_{ij}|X_{ij})$ utilizes the $\text{SparseMaps}$-based $O(N)$ transformation,\textsuperscript{26} carried out as follows. For each DF AO $X$ a block of integrals $(\mu_X\nu_X|X)$ is computed. The set of $\mu_X$ for the given $X$ is obtained as a composition of maps $L(X \to i) \subset L(i \to j) \subset L(j \to p) \subset L(p \to \mu)$, in the notation of Ref. 26. The corresponding set of $\nu_X$ is obtained as a composite map $L(X \to i) \subset L(i \to j) \subset L(j \to \bar{\mu}) \subset L(\bar{\mu} \to \mu)$. The construction of these maps was described in Section 4.2.1.

Evaluation of the last two terms in Eq. (4.12), which involve the singles CCSD amplitudes, differs from the doubles term due to the significantly greater cost of these terms due to the much larger average rank of OSV space relative to that of PNO. Therefore, we have implemented the singles terms similarly to how the MP2-F12 energy terms containing the $X$ intermediate were computed in Ref. 29, namely by fusing the contraction with the singles amplitudes into the 3-index integral transformation. To illustrate the idea, consider $t_{\alpha_i}q_{p_{ij}q_{ij}}^{a_{ij}}$; it can be represented as $t_{\alpha_i}C_{\alpha_i}^{a_{ij}}\bar{P}_{\mu_i}g_{p_{ij}q_{ij}}^{p_{ij}}$, where $C_{\mu_i}^{a_{ij}}$ is the expansion coefficient matrix for
OSVs expressed in terms of PAOs for orbital $i$, and $\tilde{P}_\mu^{ij}$ is coefficient of AO $\mu$ in PAO $\tilde{\mu}_i$. By defining the coefficients of $t_1$-transformed occupied orbitals $i_t$ as

$$C_{\mu}^{ii} \equiv t_{a_i}^{i_t} C_{\tilde{\mu}_i}^{a_t} \tilde{P}_\mu^{ij}.$$  \hspace{1cm} (4.26)

the singles contribution to intermediate $\mathcal{V}$ are computed just like an ordinary integral with 2 pair-specific indices:

$$t_{a_i}^{i_t}, g_{p_{ij}q_{ij}}^{a_j} = C_{\mu}^{ii} g_{p_{ij}q_{ij}}^{\mu q} \approx (i_t p_{ij} | X_{ij}) (V^{-1})_{X_{ij} X_{ij}} (j q_{ij} | Y_{ij}).$$  \hspace{1cm} (4.27)

Integral $(i_t p_{ij} | X_{ij})$ is computed just like its non-$t_1$-transformed counterpart (as described in detail in Ref. 26), by utilizing map $L(i_t \rightarrow \mu)$ instead of map $L(i \rightarrow \mu)$, constructed using the parameter $T_{\text{CutC}}$.

### 4.3 Computational details

All computations with the DLPNO-CCSD(2)$_{\text{F12}}$ and CCSD(T)$_{\text{F12}}$ methods utilized a developmental version of the ORCA quantum chemistry package. The canonical CCSD(2)$_{\text{F12}}$ energies were computed with the massively parallel implementation in a developmental version of Massively Parallel Quantum Chemistry package (MPQC), version 4.

The convergence of absolute and relative errors in the correlation energy with respect to truncation parameters $\{ T_{\text{CutPNO}}, T_{\text{CutPNO}}\text{Singles}, T_{\text{CutMKN}}, T_{\text{CutDO}}, T_{\text{CutPairs}}, T_{\text{CutC}} \}$ was studied on the example of a edge-sharing pentagonal prism 20-water cluster at the optimized TIP4P geometry. The cc-pVDZ-F12 orbital basis set (OBS), a correlation-consistent
basis set optimized for the explicitly correlated calculations, accompanied by the cc-pVDZ-F12/OptRI CABS and the aug-cc-pVDZ-RI density-fitting basis set (DFBS).⁸⁵,¹²⁶,¹²⁷ We abbreviate this {OBS, CABS, DFBS} triplet as {cc-pVDZ-F12, cc-pVDZ-F12/OptRI, aug-cc-pVDZ-RI}. For simplicity, the Coulomb and exchange operators were evaluated using the DF approximation using the same DFBS as the correlated portion of the computation.

To assess the errors in a representative set of noncovalent interactions we used a subset of the L7 benchmark data set.¹²⁸ The L7 data set contains seven large molecular systems that are bound by noncovalent interaction and represent an excellent training set for assessment of accuracy of local correlation methods. Calculations have been performed with the {cc-pVDZ-F12, cc-pVDZ-F12/OptRI, aug-cc-pVDZ-RI} basis set triplet.

For the scaling studies, we have used the idealized geometry of the linear n-alkane chain (\(r_{C-C} = 1.55\text{Å}, r_{C-H} = 1.09\text{Å} \text{ with all angles } 109.4712°\)) utilizing the \{def2-TZVP, cc-pVDZ-F12/OptRI, def2-TZVP/J, def2-TZVP/C\} basis set quartet denoting \{OBS, CABS, Fock matrix DFBS, DFBS\}.¹²⁹–¹³¹ The Coulomb matrix has been approximated using standard cubic scaling density fitting approach, while exchange was treated with the linear-scaling COSX method with default truncation parameters.¹³⁶ We have used def2-TZVP basis set instead of cc-pVDZ-F12 due to poor Hartree-Fock convergence within COSX framework on alkanne chains with more than 50 carbon atoms.

All computations utilized the frozen-core approximation. A standard Slater-type correlation factor, \(f(r_{12}) = (1 - \exp(-\gamma r_{12})/\gamma), \text{ with } \gamma = 0.9 \text{Bohr}^{-1} \text{ set to the recommended value for the cc-pVDZ-F12 basis set.}⁸⁵ The occupied orbitals were localized via the Foster-Boys
4.4 Results

4.4.1 Convergence of DLPNO correlation energy with the model truncation parameters.

The goal of this work is to construct a robust and reliable reduced-scaling explicitly correlated CCSD method that will maintain the accuracy of its canonical counterpart, RI-CCSD-F12, while being computationally efficient and applicable to large molecular systems that are out of reach for the canonical computations. First we studied the convergence of the absolute and relative DLPNO-CCSD(2)F12 (for brevity denoted as DLPNO-CCSD-F12) correlation energies to the canonical RI-CCSD(2)F12 energy (denoted as RI-CCSD-F12) as a function of the truncation parameters.

The accuracy of the DLPNO method is controlled via several model truncation parameters: $T_{\text{CutPNO}}$, $T_{\text{CutPairs}}$, $T_{\text{CutDO}}$, $T_{\text{CutMKN}}$, $T_{\text{CutPNOSingles}}$, and $T_{\text{CutC}}$. Setting all parameters to 0 makes the DLPNO-CCSD-F12 energy (albeit, at a much greater computational cost due to the integral transformation) equivalent to the canonical RI-CCSD-F12 energy, verified numerically for small systems to 10 decimal places. Due to the high expense of the canonical RI-CCSD-F12 computation, the comparison between the DLPNO and canonical methods had to be limited to a modestly-sized system, a 20-molecule edge-sharing pentagonal prism.
cluster, in a realistic cc-pVDZ-F12 basis. The RI-CCSD-F12 computation took 94 minutes on 32,768 cores of IBM BlueGene/Q computer at Argonne National Laboratory. Although slightly larger computations are possible with our RI-CCSD-F12 code, this computation closely approaches the practical capability limit of standard, non-reduced-scaling coupled-cluster approaches.

Figure 4.1 demonstrates the convergence of the error in the DLPNO-CCSD (blue) and DLPNO-CCSD-F12 (red) absolute correlation energies relative to the respective canonical values, as a function of $T_{\text{CutPNO}}$ (all other truncation parameters are set to 0). Both standard and F12 DLPNO energies display monotonic convergence to the canonical value, with the explicitly correlated energies slightly less accurate than the standard counterparts when loose truncation thresholds are used. This is in agreement with our findings for the LPNO-CCSD(2)$_{\text{F12}}$ method.\textsuperscript{48} 99.9\% of correlation energy, the traditional accuracy goal of the local PNO framework,\textsuperscript{21,26} is attained for DLPNO-CCSD at $T_{\text{CutPNO}} = 10^{-7}$ whereas DLPNO-CCSD-F12 recovers about 99.86\% at this value of $T_{\text{CutPNO}}$. Thus, we set the default value for $T_{\text{CutPNO}}$ to $10^{-7}$.

Figure 4.2 illustrates how the DLPNO-CCSD and DLPNO-CCSD-F12 energies vary with the truncation parameter $T_{\text{CutDO}}$ that controls the domain construction. Both standard and explicitly correlated correlation energies converge monotonically from below. It is also evident that the domain error of DLPNO-CCSD-F12 is much smaller than the DLPNO-CCSD method, in agreement with the observation of Werner,\textsuperscript{123} and our findings for the DLPNO-MP2-F12 method.\textsuperscript{29} At $T_{\text{CutDO}} = 10^{-2}$ more than 99.99\% of the correlation energy
Figure 4.1: DLPNO-CCSD and DLPNO-CCSD-F12 valence correlation energies of $(\text{H}_2\text{O})_{20}$, as percentage of the respective canonical energies, as a function of PNO truncation parameter $T_{\text{CutPNO}}$. All other truncation parameters have been set to 0.

Note that the total correlation energy of $(\text{H}_2\text{O})_{20}$ is nearly 6 Hartrees, hence an error of 0.1% of the correlation energy, due to PNO or domain truncation, amounts to 3.7 kcal/mol! Thus it is important to analyze how the PNO and domain truncations affect relative energies (i.e., energy differences). To systematically study the effect of approximations on a relative energy we chose the dissociation energy of $(\text{H}_2\text{O})_{20}$ as the target (the optimized MP2/aug-cc-pVTZ
Figure 4.2: DLPNO-CCSD and DLPNO-CCSD-F12 valence correlation energies of \((\text{H}_2\text{O})_{20}\) as a function of the domain truncation parameter \(T_{\text{CutDO}}\). The energies are defined as percentages of the respective DLPNO energies with \(T_{\text{CutPNO}} = 10^{-7}\), \(T_{\text{CutPairs}} = 10^{-5}\), and other truncation parameters set to 0.

Monomer geometry was used. The advantage of the dissociation energy is that it is unique and more challenging to compute to chemical accuracy than relative energies of multiple cluster conformers. Of course, computing dissociation energies to chemical accuracy will be limited to relatively small systems since it is a size-extensive property, hence its systematic error will be proportional to the system size. Therefore for describing size-extensive prop-
erties the DLPNO model thresholds, like $T_{\text{CutPNO}}$ and $T_{\text{CutDO}}$, will have to be decreased as the system size increases.

Note that the errors due to the PNO truncation, domain, and pair-screening approximations are expected to affect the dissociation energy quite differently. For example, it is expected that the PNO truncation error will somewhat cancel between the cluster and isolated monomers since the intramonomer pair correlation energies will be affected by the PNO truncation to similar extent both in monomer and cluster computations. Since it appears that the PNO truncation always causes underestimation of the correlation energies, PNO truncation should decrease the dissociation energy. In contrast, the errors due to domains will vanish for monomers as small as a water molecule, thus the domain truncation error is not expected to cancel to any appreciable effect in the dissociation energy. Similarly, the pair screening should not affect the monomer energies, only the cluster energy, hence the pair screening errors are not expected to cancel in the dissociation energy.

Variation of the DLPNO-CCSD-F12 dissociation energy of $(\text{H}_2\text{O})_{20}$, relative to the canonical RI-CCSD-F12 value, as a function of the DLPNO model truncation parameters are shown in Table 4.1. The reference \{cc-pVDZ-F12, cc-pVDZ-F12/OptRI, aug-cc-pVDZ-RI\} RI-CCSD-F12 dissociation energy is 186.15 kcal/mol. The first block of data shows the dependence on the $T_{\text{CutPNO}}$ parameter only, with all other truncation parameters were set to 0. In contrast to the monotonic convergence of the absolute correlation energies (Figure 4.1), the dissociation energy converges nonmonotonically. For example, with $T_{\text{CutPNO}} = 10^{-6}$ the error is only -0.40 kcal/mol, although as Figure 4.1 illustrates this parameter setting corresponds to only about
99.1% recovery of the correlation energy (the negative sign means that the DLPNO-CCSD-F12 dissociation energy is less than the RI-CCSD-F12 counterpart, as expected). With the default value of $T_{\text{CutPNO}} = 10^{-7}$, the error is -0.64 kcal/mol, and its magnitude increases further to -0.90 kcal/mol with $T_{\text{CutPNO}} = 10^{-8}$. However, these are small errors relative to the magnitude of the dissociation energy as well as (likely) the residual basis set error, hence this performance is quite reasonable.

The second block of data in Table 4.1 illustrates variation of the dissociation energy with the $T_{\text{CutPairs}}$ parameter that controls the pair screening in DLPNO-CCSD (here we used the $T_{\text{CutPNO}} = 10^{-7}$ default, and the rest of parameters were set to 0). When $T_{\text{CutPairs}}$ approaches 0 the error should approach the value obtained with default $T_{\text{CutPNO}}$, i.e., -0.64 kcal/mol. Indeed, with $T_{\text{CutPairs}} = 10^{-6}$ the error is -0.68 kcal/mol, i.e., the effect of pair screening is only 0.04 kcal/mol. A more aggressive screening, with $T_{\text{CutPairs}} = 10^{-4}$, leads to a fortuitous error cancellation between the PNO truncation and pair screening, but it is clear that such cancellation is not systematic and cannot be always expected to occur. With the default setting of $T_{\text{CutPairs}} = 10^{-5}$ the effect of the pair screening is only 0.3 kcal/mol.

The third block of data in Table 4.1 illustrates the error of DLPNO-CCSD-F12 dissociation energy as a function of $T_{\text{CutDO}}$ parameter (with default values set for $T_{\text{CutPairs}}$ and $T_{\text{CutPNO}}$, and the rest of parameters set to 0; thus as $T_{\text{CutDO}}$ approaches zero the dissociation energy error should approach -0.31 kcal/mol). The convergence with respect to $T_{\text{CutDO}}$ is not monotonic. Note that although setting $T_{\text{CutDO}} = 10^{-1}$ introduces the error in the DLPNO-CCSD-F12 correlation energy of only 0.1%, due to the lack of the error cancellation the
resulting error in the dissociation energy is almost 3 kcal/mol! Although the domain error of the absolute correlation energy is greatly canceled by the F12 terms, there should be no reliance on the error cancellation in relative energy. Thus we conservatively set the default value for $T_{\text{CutDO}}$ in DLPNO-CCSD-F12 to $10^{-2}$, at the same value as other non-F12 methods.\textsuperscript{26,27}

The fourth block of data in Table 4.1 illustrates the error of DLPNO-CCSD-F12 dissociation energy as a function of $T_{\text{CutMKN}}$ parameter that controls the local density fitting (with default values set for $T_{\text{CutPairs}}$, $T_{\text{CutPNO}}$, and $T_{\text{CutDO}}$, and the rest of parameters set to 0; thus as $T_{\text{CutMKN}}$ approaches zero the dissociation energy error should approach -0.37 kcal/mol). The convergence with respect to $T_{\text{CutMKN}}$ is rapid and with the default $T_{\text{CutMKN}} = 10^{-4}$ value the effect of the local density fitting approximation is only 0.02 kcal/mol.\textsuperscript{48,140}

The fifth block of data in Table 4.1 illustrates the error of DLPNO-CCSD-F12 dissociation energy as a function of $T_{\text{CutPNOSingles}}$ parameter that controls the size of the OSV space used to expand $t_1$ amplitudes (with default values set for $T_{\text{CutPairs}}$, $T_{\text{CutPNO}}$, $T_{\text{CutDO}}$, and $T_{\text{CutMKN}}$, and the rest of parameters set to 0; thus as $T_{\text{CutPNOSingles}}$ approaches zero the dissociation energy error should approach -0.39 kcal/mol). The convergence is rapid and monotonic. The default $T_{\text{CutPNOSingles}} = 3 \times 10^{-9}$ value results in a OSV truncation error of only 0.03 kcal/mol.

The sixth block of data in Table 4.1 illustrates the error of DLPNO-CCSD-F12 dissociation energy as a function of $T_{\text{CutC}}$ parameter that controls the sparsity of the occupied and PAO orbital AO coefficients. There is no noticeable variation of the dissociation energy
with respect to this parameter. Thus the default is set at the same conservative value, 
$T_{\text{CutC}} = 10^{-3}$, as in the preceding DLPNO work.\textsuperscript{26,27,29}

The default set of truncation parameters that we have determined in this analysis, ($T_{\text{CutPNO}} = 10^{-7}, T_{\text{CutPNOSingles}} = 3 \times 10^{-9}, T_{\text{CutPairs}} = 10^{-5}, T_{\text{CutMKN}} = 10^{-4}, T_{\text{CutDO}} = 10^{-2}, T_{\text{CutC}} = 10^{-3}$), corresponds with the \texttt{TightPNO} keyword in ORCA that has been recommended for accurate prediction of the binding energies.\textsuperscript{140} Using the default set of truncation parameters results in a 0.42 kcal/mol underestimation of the 187 kcal/mol dissociation energy of the (H$_2$O)$_{20}$ cluster by the DLPNO-CCSD-F12 method.

### 4.4.2 Computational cost of DLPNO-CCSD(T)-F12

To illustrate the computational scaling of the explicitly correlated DLPNO CC methods we performed a series of calculations on quasilinear n-alkanes using both the canonical RI-CCSD(T)-F12 method and the DLPNO-CCSD(T)-F12 reduced-scaling method, employing the def2-TZVP basis set (Figure 4.3). To simplify the performance analysis we utilize a single core of the Intel Xeon E5-2680 v3 2.50 GHz CPU. All non-negligible post-mean-field steps in the DLPNO-CCSD(T)-F12 are formulated in linear effort; the exception is the $O(N^3)$ density fitting (DF) algorithm for the Coulomb matrix. Note that these computations used the default linear-scaling COSX algorithm for the exchange matrix.\textsuperscript{136} The default DLPNO \texttt{TightPNO} protocol was utilized.\textsuperscript{140} The largest calculation that we have performed with \texttt{TightPNO} truncation setup is on C$_{200}$H$_{402}$ n-alkane with the 8612, 31662 and 21230
Table 4.1: The difference between DLPNO and canonical CCSD-F12 dissociation energies of (H₂O)₂₀, in kcal/mol as a function of the DLPNO truncation parameters. Each truncation parameter is set to zero unless otherwise stated.

<table>
<thead>
<tr>
<th>Truncation parameter settings</th>
<th>δDₑ(DLPNO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{\text{CutPNO}} = 10^{-6} )</td>
<td>-0.40</td>
</tr>
<tr>
<td>( T_{\text{CutPNO}} = 10^{-7} )</td>
<td>-0.64</td>
</tr>
<tr>
<td>( T_{\text{CutPNO}} = 10^{-8} )</td>
<td>-0.90</td>
</tr>
<tr>
<td>( T_{\text{CutPNO}} = 10^{-9} )</td>
<td>-0.39</td>
</tr>
<tr>
<td>( a \ T_{\text{CutPairs}} = 10^{-4} )</td>
<td>0.06</td>
</tr>
<tr>
<td>( a \ T_{\text{CutPairs}} = 10^{-5} )</td>
<td>-0.31</td>
</tr>
<tr>
<td>( a \ T_{\text{CutPairs}} = 10^{-6} )</td>
<td>-0.68</td>
</tr>
<tr>
<td>( b \ T_{\text{CutDO}} = 10^{-1} )</td>
<td>-2.93</td>
</tr>
<tr>
<td>( b \ T_{\text{CutDO}} = 10^{-2} )</td>
<td>-0.37</td>
</tr>
<tr>
<td>( b \ T_{\text{CutDO}} = 10^{-3} )</td>
<td>-0.70</td>
</tr>
<tr>
<td>( c \ T_{\text{CutMKN}} = 10^{-3} )</td>
<td>-0.67</td>
</tr>
<tr>
<td>( c \ T_{\text{CutMKN}} = 10^{-4} )</td>
<td>-0.39</td>
</tr>
<tr>
<td>( c \ T_{\text{CutMKN}} = 10^{-5} )</td>
<td>-0.38</td>
</tr>
<tr>
<td>( d \ T_{\text{CutPNOSingles}} = 10^{-8} )</td>
<td>-0.50</td>
</tr>
<tr>
<td>( d \ T_{\text{CutPNOSingles}} = 3 \times 10^{-9} )</td>
<td>-0.42</td>
</tr>
<tr>
<td>( d \ T_{\text{CutPNOSingles}} = 10^{-9} )</td>
<td>-0.39</td>
</tr>
<tr>
<td>( e \ T_{\text{CutC}} = 10^{-2} )</td>
<td>-0.42</td>
</tr>
<tr>
<td>( e \ T_{\text{CutC}} = 10^{-3} )</td>
<td>-0.42</td>
</tr>
<tr>
<td>( e \ T_{\text{CutC}} = 10^{-4} )</td>
<td>-0.42</td>
</tr>
<tr>
<td>( \text{TightPNO} )</td>
<td>-0.42</td>
</tr>
</tbody>
</table>

\( a \ T_{\text{CutPNO}} = 10^{-7} \), \( b \ T_{\text{CutPNO}} = 10^{-7} \), \( T_{\text{CutPairs}} = 10^{-5} \).

\( c \ T_{\text{CutPNO}} = 10^{-7} \), \( T_{\text{CutPairs}} = 10^{-5} \), \( T_{\text{CutDO}} = 10^{-2} \).

\( d \ T_{\text{CutPNO}} = 10^{-7} \), \( T_{\text{CutPairs}} = 10^{-5} \), \( T_{\text{CutDO}} = 10^{-2} \), \( T_{\text{CutMKN}} = 10^{-4} \).

\( e \ T_{\text{CutPNO}} = 10^{-7} \), \( T_{\text{CutPairs}} = 10^{-5} \), \( T_{\text{CutDO}} = 10^{-2} \), \( T_{\text{CutMKN}} = 10^{-4} \), \( T_{\text{CutPNOSingles}} = 3 \times 10^{-9} \).
basis functions for orbital, CABS basis space and auxiliary (DFBS) respectively. As can be easily seen, a low-order scaling of the DLPNO method is rapidly achieved for the quasi-1-dimensional system. The effective scaling exponent estimated by a finite difference from the two largest computations is 1.49. The deviation from linear-scaling behavior is due to the computation of the Fock matrix in a nonlinear-scaling fashion, as well as non-reproducible performance of the I/O subsystem. There is no crossover between the reduced-scaling and canonical versions, as the DLPNO-CCSD-F12 method is found to always be more efficient than its canonical counterpart for these systems.

Table 4.2 reports detailed breakdown of the wall clock times (in seconds) for the major substeps of the DLPNO-CCSD(T)-F12 energy algorithm for the C\textsubscript{200}H\textsubscript{402} n-alkane and for coronene dimer (both computations utilized the TightPNO protocol). The “MP2-F12 specific” component includes evaluation of all contributions that are required for the F12 correction in MP2, which was discussed in detail in Ref. 29. Component “2-e integrals required for doubles” considers evaluation of the three-center integrals required for the computation of the $g_{p_iq_j}$ that enters equation (4.18).

This is followed by time required for evaluation of integrals that enter (4.19). The “computation of doubles” and “computation of singles” consists of reading of integrals form disk, transforming them to the PNO basis and contraction as defined in (4.18) and (4.19). In the case of C\textsubscript{200}H\textsubscript{402}, computation of F12 contributions that enters the CCSD-F12 correction is 37 % more time consuming than the MP2-F12 counterpart. Furthermore, the total time spent in the F12 part of the computation in the C\textsubscript{200}H\textsubscript{402} is 55 % of the entire computa-
Table 4.2: Timings (in seconds) for the computation of different contributions in DLPNO-CCSD(T)-F12 method using TightPNO truncation setup.

<table>
<thead>
<tr>
<th></th>
<th>n-C$<em>{200}$H$</em>{402}^{b}$</th>
<th>coronene dimer$^{c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of orbital/CABS AOs$^{a}$/DF</td>
<td>8612/31662/21230</td>
<td>1656/5352/4008</td>
</tr>
<tr>
<td>MP2-F12 specific</td>
<td>132 811</td>
<td>28 648</td>
</tr>
<tr>
<td>2-e integrals required for doubles</td>
<td>34 433</td>
<td>1 497</td>
</tr>
<tr>
<td>2-e integrals required for singles</td>
<td>3 243</td>
<td>746</td>
</tr>
<tr>
<td>computation of doubles</td>
<td>41 526</td>
<td>49 353</td>
</tr>
<tr>
<td>computation of singles</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>total F12</td>
<td>212 015</td>
<td>80 244</td>
</tr>
<tr>
<td>computation of CCSD</td>
<td>113 668</td>
<td>130 768</td>
</tr>
<tr>
<td>computation of (T)</td>
<td>61 970</td>
<td>781 477</td>
</tr>
<tr>
<td>CCSD(T)-F12(total)</td>
<td>387 654</td>
<td>992 489</td>
</tr>
</tbody>
</table>

$^{a}$ CABS AOs include the orbital basis AOs due to the use of the CABS+ approach.$^{39}$

$^{b}$ The \{def2-TZVP, cc-pVDZ-F12/OptRI, def2-TZVP/J, def2-TZVP/C\} basis sets, 1 core

$^{c}$ The \{cc-pVDZ-F12, cc-pVDZ-F12/OptRI, aug-cc-pVDZ-RI\} basis sets, 16 cores
Figure 4.3: Wall clock time (in hours) for the canonical RI-CCSD(T)-F12 and local DLPNO-CCSD(T)-F12 approach with the def2-TZVP basis set and TightPNO truncation threshold. Timings do not include Hartree-Fock computation time.

Hartree-Fock has been omitted from this analysis. In the case of coronene dimer, which represents a more realistic chemical system, the computation time for the F12 correction is only 9 % opposed to 14.4 and 76.6 % for computation of the CCSD and (T), respectively.
4.4.3 Application of DLPNO-CCSD(T)-F12 to energetics of non-covalent interactions.

To gauge the applicability of the explicitly correlated DLPNO coupled-cluster methods to description of noncovalent interactions we applied the DLPNO-CCSD(T)-F12 method to a 5-system subset of the L7 data set of Sedlak et al.\textsuperscript{128} comprising coronene dimer, guanine-cytosine base pair stack (GCGC), guanine trimer, octadecane dimer and phenylalanine residues trimer. Thus subset will be referred to as L7/5. In addition to the default TightPNO DLPNO-CCSD(T)-F12 protocol we also employed an improved VeryTightPNO protocol ($T_{\text{CutPNO}} = 10^{-8}, T_{\text{CutPNOSingles}} = 3 \times 10^{-9}, T_{\text{CutPairs}} = 10^{-6}, T_{\text{CutMKN}} = 10^{-4}, T_{\text{CutDO}} = 10^{-2}, T_{\text{CutC}} = 10^{-3}$). In the case of coronene dimer, the VeryTightPNO protocol is modified where PNO truncation parameter is set to $T_{\text{CutPNO}} = 3 \times 10^{-8}$. The dissociation energies were computed relative to the monomers at dimer geometry without counterpoise correction\textsuperscript{141} (Table 4.3). Recent findings by Martin and his coworkers suggests that uncorrected dissociation energies obtained by MP2-F12 and CCSD-F12b with cc-pVDZ-F12 basis set are in much better agreement with respect to CBS limit then the ones with counterpoise correction.\textsuperscript{134} Furthermore, Werner with collaborators found that dissociation energies for different configurations of benzene dimer calculated with local MP2 (LMP2) method, have very small BSSE contribution for any choice of basis set larger then double-zeta. This is due to definition of domains in local correlation methods which does not include contributions responsible for the BSSE such as double excitations from first to second monomer.\textsuperscript{142} Due to lack of canonical CCSD(T) binding energies in the CBS limit, we are unable to quantify
the errors due to the local approximations for the complete L7/5 set. Thus we limit our discussion to the comparison to the existing benchmark values for the dissociation energies.

The benchmark QCISD(T)/CBS energies reported by Sedlak et al.\textsuperscript{128} are very similar to the DLPNO-CCSD(T)-F12 results obtained with the \texttt{TightPNO} protocol for the “last” four systems, with the largest difference, $\sim 0.6$ kcal/mol, observed for octadecane dimer. However, for coronene dimer our \texttt{TightPNO} estimate is $\sim 3$ kcal/mol away from the result of Sedlak et al. To try to resolve the discrepancy, we performed computations with the improved \texttt{VeryTightPNO} set of model parameters. However, the use of tighter truncation parameters increases the discrepancy for coronene dimer to more than 5 kcal/mol. The only other reference values available to us is the DLPNO-CCSD(T) + basis set extrapolation result reported by Cui and coworkers.\textsuperscript{143} The DLPNO-CCSD(T)/CBS prediction for the coronene dimer is in a reasonable agreement with our DLPNO-CCSD(T)-F12 prediction for coronene dimer, but in a relatively poor agreement for the other 4 systems. The results of Cui et al. also significantly deviate from the Sedlak’s, 4.76 and 3.17 kcal/mol in the case of coronene dimer and GCGC base pair stack, respectively, followed by 2.16 kcal/mol in the case of phenylalanine residues trimer. It is clear that additional effort is needed to improve the confidence in the CCSD(T) CBS limit estimates for these systems.

To understand the origin of the large, 2.17 kcal/mol, difference for the dissociation energy for coronene dimer obtained with \texttt{TightPNO} and \texttt{VeryTightPNO} parameter sets we compared the contributions to the dissociation energy from the five components of the DLPNO-CCSD(T)-F12 energy (see Table 4.4). The major difference between the two predictions is attributed
Table 4.3: Dissociation energy estimates (kcal/mol) for the L7/5 data set.

<table>
<thead>
<tr>
<th>System</th>
<th>QCISD(T)/CBS&lt;sup&gt;a&lt;/sup&gt;</th>
<th>DLPNO-CCSD(T)/CBS&lt;sup&gt;b&lt;/sup&gt;</th>
<th>DPNO-CCSD(T)-F12&lt;sup&gt;c&lt;/sup&gt;</th>
<th>TightPNO</th>
<th>VeryTightPNO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronene dimer</td>
<td>24.36</td>
<td>19.60</td>
<td>21.31</td>
<td>19.14&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCGC base pair stack</td>
<td>14.37</td>
<td>11.20</td>
<td>14.51</td>
<td>13.69</td>
<td></td>
</tr>
<tr>
<td>Guanine trimer</td>
<td>2.40</td>
<td>1.60</td>
<td>2.52</td>
<td>2.36</td>
<td></td>
</tr>
<tr>
<td>Octadecane dimer</td>
<td>11.06</td>
<td>9.80</td>
<td>11.62</td>
<td>11.13</td>
<td></td>
</tr>
<tr>
<td>Phenylalanine residues trimer</td>
<td>25.76</td>
<td>23.60</td>
<td>25.61</td>
<td>25.09</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Ref. 128.

<sup>b</sup> Ref. 143.

<sup>c</sup> Using the {cc-pVDZ-F12, cc-pVDZ-F12/OptRI, aug-cc-pVDZ-RI} {OBS,CABS,DFBS} triplet.

<sup>d</sup> $T_{\text{CutPNO}} = 3 \times 10^{-8}$, $T_{\text{CutPNO} \text{Singles}} = 3 \times 10^{-9}$, $T_{\text{CutPairs}} = 10^{-6}$, $T_{\text{CutMKN}} = 10^{-4}$, $T_{\text{CutDO}} = 10^{-2}$, $T_{\text{CutC}} = 10^{-3}$

to the significantly greater numbers of pairs treated at the CCSD level and then corrected at the LMP2 level, which is controlled by the $T_{\text{CutPairs}}$ parameter. Specifically, note the significant difference in the MP2 pair corrections, 5.45 and 0.16 kcal/mol, for the TightPNO and VeryTightPNO models, respectively. This is due to a significantly increased number of pairs that are treated at the CCSD level when the VeryTightPNO model is used. It is well known that the MP2 overbinds the systems that are governed with $\pi...\pi$ interactions such as in benzene dimer and graphenes,<sup>81,144-146</sup> hence coupled-cluster level description is necessary for the majority pairs. In principle, the TightPNO dissociation energy could be improved
by treating the weak pairs using the SCS-MP2 approach\textsuperscript{147} which predicts the dissociation energy of the $\pi...\pi$ systems more accurately than the standard MP2 method,\textsuperscript{81,142} however, such empirical approaches are unlikely to work universally. The only solution seems to be to treat more pairs at the CCSD level for systems with significant $\pi - \pi$ character, by using tighter settings for the truncation parameters such as \texttt{VeryTightPNO} in such extreme circumstances.

Table 4.4: Breakdown of the correlation energy contributions to the DLPNO-CCSD(T)-F12 dissociation energy of coronene dimer (kcal/mol). The acronyms \texttt{wp} and \texttt{sp} stand for weak and strong pairs, respectively.

<table>
<thead>
<tr>
<th>System</th>
<th>TightPNO</th>
<th>VeryTightPNO</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP2 pair correction (wp)</td>
<td>5.45</td>
<td>0.16</td>
</tr>
<tr>
<td>PNO incompleteness (wp)</td>
<td>1.28</td>
<td>0.79</td>
</tr>
<tr>
<td>CCSD strong-pair energy (sp)</td>
<td>30.79</td>
<td>32.44</td>
</tr>
<tr>
<td>F12 (T)</td>
<td>-4.25</td>
<td>-4.02</td>
</tr>
<tr>
<td></td>
<td>4.90</td>
<td>5.25</td>
</tr>
</tbody>
</table>

### 4.5 Summary and Perspective

In this work we presented a reduced-scaling formulation of the explicitly correlated DLPNO-CCSD(2)$_{\text{F12}}$ method in which all nontrivial post-mean-field steps were formulated such that the computational complexity is linear in system size. The two key ideas are the use of
pair natural orbitals for compact representation of wave function amplitudes in conjunction with the domain approximation. This development utilizes the sparse tensor computation concepts and algorithms described in the context of the DLPNO-MP2, DLPNO-CCSD(T), DLPNO-NEVPT2, and DLPNO-MP2-F12 methods by us recently.\textsuperscript{26–29} In line with our prior focus on the use of minimally-invasive but rigorous F12 formulations, the present formulation introduces the explicitly correlated terms by perturbation theory such that the perturbative triples energy correction, (T), is identical to the non-F12 counterpart,\textsuperscript{100} and the treatment of F12 intermediates is a slight revision\textsuperscript{29} of the resolution-of-identity-only approach of Kedžuch \textit{et al.}\textsuperscript{72} The robustness of the DLPNO-CCSD(2)\textsubscript{F12} method was tested against the canonical RI-CCSD(2)\textsubscript{F12} counterpart applied to a large 60-atom molecular cluster in a realistic basis; the reduced-scaling computation required only 78 minutes on a 16-core workstation, whereas the canonical computation required 94 minutes on more than 32 thousand CPU cores of a state-of-the-art supercomputer.

It is useful to compare the presented DLPNO-CCSD(2)\textsubscript{F12} method to other production reduced-scaling coupled-cluster approaches. The pioneering work of Werner and Adler on LCCSD-F12a,b\textsuperscript{40,41} represented the coupled-cluster amplitudes in terms of PAOs, with the domain of PAOs for a given localized occupied orbital defined using geometric locality criteria. They also utilized a simplified formulation of some of the F12 intermediates. The formulation was formally $\mathcal{O}(N)$ scaling, but no performance data was presented to gauge the efficiency of the approach. We presented a production level implementation of the quartic scaling PNO-based CCSD\textsubscript{F12} utilized the state-of-the-art F12 technology.\textsuperscript{48} Around the
same time Tew and Hättig developed a cubic scaling PNO-based formulation of explicitly correlated CCSD,\textsuperscript{46} and demonstrated efficient applications to systems with a few dozen atoms in a realistic basis. Both perturbative and iterative inclusion of the F12 terms was achieved, and state-of-the-art formulation of the F12 intermediates was utilized. Werner and coworkers recently developed an efficient parallel implementation of a linear-scaling PNO-based MP2-F12 method\textsuperscript{50} and are currently nearing a completion of an efficient explicitly correlated coupled-cluster method.

Future work includes extension of the approach to the treatment of open-shell systems, along the lines of the recently completed work on the open-shell DLPNO-CCSD method, as well as a massively parallel implementation of the DLPNO framework.
Chapter 5

Massively Parallel Laplace Transform Perturbative Triples Correction (T) to the Coupled-Cluster Singles and Doubles Method using Density Fitting (DF)
5.1 Introduction

Wave function based methods provide an accurate and systematically improvable route for achieving chemical accuracy. Within this category, coupled-cluster (CC) hierarchy of methods are known to be robust and popular due to their appealing properties, such as size consistency, size extensivity and invariance with respect to orbital rotations.\textsuperscript{5,6} The hierarchy in the CC methods is achieved by truncation of exponential ansatz of the cluster operator. Thus, inclusion of single and double excitations in the cluster operator will give the coupled-cluster singles and doubles method (CCSD), with computational complexity of $O(N^6)$, where $N$ is the measure of the system size. However, for achieving chemical accuracy, inclusion of higher than double excitations is necessary.\textsuperscript{63,64} Full inclusion of triple excitations defines iterative CCSDT method with computational cost of $O(N^8)$. The CCSD(T) method,\textsuperscript{66} known as the “gold standard” of quantum chemistry, where triples are treated perturbatively, has the best ratio between cost and accuracy with the computational cost of $O(N^7)$. Furthermore, wave function methods are plagued by very slow convergence to the complete basis set limit (CBS limit), thus for achieving adequate accuracy, very large basis sets have to be employed. All of these combined, the unfavorable scaling exponent and the large basis sets, limit the method to systems with up to dozen atoms. The problem of slow convergence can be alleviated by employing explicitly correlated methods (R12/F12) where quality of $X+2$ basis set can be achieved with $X$ basis set ($X$ is the cardinal number of a basis set) at significantly lower cost.\textsuperscript{33–35} The method applicability can be slightly increased by designing highly optimized distributed-memory parallel algorithms\textsuperscript{104,148–155} or by employing
graphical processing units (GPU).\textsuperscript{153,156,157} Reducing the scaling exponent of the CCSD(T) method is an active field of research for the past few decades. Majority of research efforts are aimed toward exploiting the local nature of electron correlation. The pioneers in this field are Pulay and Sæbo\textsuperscript{12,13} who have introduced the concept of projected atomic orbitals (PAOs) in which unoccupied (virtual) space is restricted to small region around the localized occupied orbital.\textsuperscript{117} This has been furthermore refined by Werner and Schütz\textsuperscript{15,16,158} with their production level implementations of many methods for treating electron correlation effects such as MP2, CCSD and CCSD(T) in $\mathcal{O}(N)$ scaling fashion; however, the PAO-based methods have late crossover with respect to the canonical method if high accuracy is required.

Pair natural orbitals (PNOs)\textsuperscript{17} offer an alternative route and more compact formulation of unoccupied space as investigated by Neese.\textsuperscript{20,21,48} Neese and Riplinger have combined PAO and PNO approaches, leading to a linear-scaling implementation of the CCSD(T) method, the DLPNO-CCSD(T).\textsuperscript{25–27,29,49,113} Similar hierarchy of approximations has been used by Werner and co-workers, allowing efficient parallel implementation of linear-scaling methods.\textsuperscript{30,31,50,51} Yet another way for reduction of the scaling exponent in higher-order perturbation theories can be achieved by the Laplace transform of the denominator, allowing factorization of energy expression.\textsuperscript{159,160} This concept was introduced by Almlöf in the context of the MP2 method, where he also discussed that removal of the denominator makes MP2 energy invariant to orbital rotations.\textsuperscript{159} Scuseria with co-workers applied the Laplace transform technique on the denominator of the perturbative triple correction (T) to the CCSD method, achieving scaling reduction from $\mathcal{O}(N^7)$ to $\mathcal{O}(N^6)$.\textsuperscript{161} Their work is
central to the developments presented here; however, our work goes beyond by implementing the method in the Massively Parallel Quantum Chemistry package (MPQC) version 4 and by applying density fitting approximation (DF), allowing the method to be tested on much larger systems than previously reported. This implementation uses the TiledArray tensor library that provides distributed parallel tensor routines.\textsuperscript{162,163} Recently, Laplace transform of the (T) gained prominence in the context of local correlation methods by groups of Hättig and Kállay as an alternative to the semi-canonical (T0) method due to its invariance with orbital rotations.\textsuperscript{47,164}

The text is organized as follows: In Section 5.2, we describe theoretical background of the (T) correction of the CCSD(T) method and the Laplace transform of the (T). This is followed by the section on Computational details. Quadrature accuracy, computational performance and application to chemically relevant systems will be given in Section 5.4. In Section 5.5, we summarize our findings.

\section*{5.2 Formalism}

\subsection*{5.2.1 Canonical perturbative triples correction (T)}

The perturbative triple correction, (T),\textsuperscript{66} to the coupled-cluster singles and doubles (CCSD) consists of fourth- and fifth-order contributions:

\begin{equation}
E_{(T)} = E_{(T)}^{[4]} + E_{(ST)}^{[5]}.
\end{equation}
Their programmable equations in the spin-adapted formalism are as follows:

\[
E^{[4]}_{(T)} = \frac{1}{3} \sum_{ijk} \sum_{abc} W^{ijk}_{abc} R^{ijk}_{abc} D^{ijk}_{abc}, \quad (5.2)
\]

and

\[
E^{[5]}_{(ST)} = -\frac{1}{3} \sum_{ijk} \sum_{abc} V^{ijk}_{abc} R^{ijk}_{abc} D^{ijk}_{abc}. \quad (5.3)
\]

The six-index tensors \(W^{ijk}_{abc}\), \(V^{ijk}_{abc}\), \(R^{ijk}_{abc}\) and \(D^{ijk}_{abc}\) are defined as:

\[
W^{ijk}_{abc} = \hat{P}^{ijk}_{abc} \left( \sum_f g^{ij}_{ab} t^{kj}_{ef} - \sum_m g^{ij}_{am} t^{mk}_{bc} \right), \quad (5.4)
\]

\[
V^{ijk}_{abc} = g^{ij}_{ab} t^{jk}_{c} + g^{jk}_{bc} t^{ki}_{a} + g^{ki}_{ca} t^{ij}_{b}, \quad (5.5)
\]

\[
R^{ijk}_{abc} = 4W^{ijk}_{abc} + W^{kij}_{abc} - 2W^{kji}_{abc} - 2W^{jik}_{abc}, \quad (5.6)
\]

\[
D^{ijk}_{abc} = f^{a}_a + f^{b}_b + f^{c}_c - f^{i}_i - f^{j}_j - f^{k}_k, \quad (5.7)
\]

where \(t^{ij}_{ab}\) and \(t^{i}_a\) are converged singles and doubles CCSD cluster amplitudes, \(g^{pq}_{rs} \equiv \langle rs | r_{12}^{-1} | pq \rangle\) are the Coulomb repulsion two-electron integrals, while \(f^{p}_p\) are diagonal Fock matrix elements in the canonical basis. The \(\hat{P}^{ijk}_{abc}\) operator defines permutations of indices of a given six-index tensor \(X^{ijk}_{abc}\) as:

\[
\hat{P}^{ijk}_{abc}(X^{ijk}_{abc}) = X^{ijk}_{abc} + X^{jki}_{bca} + X^{kij}_{cab} + X^{ikj}_{acb} + X^{kji}_{cba} + X^{jik}_{bac}. \quad (5.8)
\]

As usual, occupied and unoccupied Hartree-Fock orbitals are denoted by \(i, j, k, l, \ldots\) and \(a, b, c, d, \ldots\), respectively, and \(p, q, r, s, \ldots\) refer to orbitals in the OBS space. The scaling bottleneck of the (T) comes form construction of the six-dimensional tensor \(W^{ijk}_{abc}\) with cost \(O^3V^4 + O^4V^3\), where \(O\) and \(V\) corresponds to the number of occupied and unoccupied orbitals,
respectively. For realistic basis sets, the number of occupied orbitals is much smaller than the number of unoccupied orbitals; therefore, the true bottleneck has the cost of $O^3V^4$ or $O(N^7)$. The Laplace transform of the denominator $D_{ijk}^{abc}$ allows factorization of sums in Eq. (5.2) and Eq. (5.3), which reduces the scaling as will be explained in the next section.

### 5.2.2 Laplace transform

Laplace transform can be used to rewrite the denominator in the orbital invariant form:

$$\frac{1}{D_{ijk}^{abc}} = \int_0^\infty e^{-D_{ijk}^{abc}s} ds \approx \sum_{\alpha} w_{\alpha} e^{-D_{ijk}^{abc}s_{\alpha}}, \quad (5.9)$$

where $w_{\alpha}$ and $s_{\alpha}$ are quadrature weights and roots, respectively, and $n_q$ is the number of quadrature points.Within this formalism, we can rewrite constituents of Eq. (5.1) as:

$$E_{[4]}^{(T)} = \frac{1}{3} \sum_{ijk} \sum_{abc} \int_0^\infty W_{ijk}^{abc} R_{ij}^{abc} e^{-\left(f_a^a + f_b^b + f_c^c - f_i^i - f_j^j - f_k^k\right)s} ds, \quad (5.10)$$

$$E_{[5]}^{(ST)} = -\frac{1}{3} \sum_{ijk} \sum_{abc} \int_0^\infty V_{ijk}^{abc} R_{ij}^{abc} e^{-\left(f_a^a + f_b^b + f_c^c - f_i^i - f_j^j - f_k^k\right)s} ds, \quad (5.11)$$

which enables decoupling of the denominator, since exponentials of the Fock matrix elements are free to move and to be reorganized in the most convenient way. Full expansion of Eq. (5.10) and Eq. (5.11) will produce 124 unique contributions. In the following text, we will discuss in details all possible contributions. The unique contributions from Eq. (5.10) can be categorized into three distinct groups based on the products that can be obtained from $W_{ijk}^{abc}$ and $R_{ij}^{abc}$, namely, $E_{[4]}^{(T)} \leftarrow \{ \sum_e g_{vv}^{eo} \sum_f g_{ve}^{fo} \sum_m g_{vv}^{mo} \sum_n g_{vv}^{no} \}$.
where \( o \) and \( v \) represent the general occupied and unoccupied index. We have used the same notation as in Ref. 161. The first set of constituents in the curly bracket will contribute 26 unique terms, 6 of \( OV^5 \), 16 of \( O^2V^4 \), and 4 of \( OV^4 \) cost. The second set will contribute 36 unique terms, 8 of \( O^2V^4 \), 24 of \( O^3V^3 \), and 4 of \( O^2V^3 \) cost. The third set of constituents will contribute another 24 terms, 6 of \( O^3V^3 \), 16 of \( O^4V^2 \), and 4 of \( O^3V^2 \) cost. There will be another 36 contributions that can be obtained from the expansion of Eq. (5.11) that are categorized into two groups due to the products of \( V_{abc}^{ijk} \) and \( R_{abc}^{ijk} \) giving: 

\[
E_{\text{(ST)}}^{[5]} \leftarrow \{ g_{vuv}^{o\alpha\beta} \sum_f g_{vuv}^{o\alpha\beta} f_{vuj}^{oo} g_{vv^v}^{o\alpha\beta} \sum_m g_{vmv}^{o\alpha\beta} \}.
\]

The first set of constituents in the curly bracket will contribute 18 unique terms: 4 of \( O^2V^4 \), 6 of \( O^3V^3 \), and 8 of \( O^2V^3 \) cost; and the second set will give another 18 contributions: 4 of \( O^3V^3 \), 6 of \( O^4V^2 \), and 8 of \( O^3V^2 \) cost. In total, there are 28 terms that has \( N^5 \) and 96 with \( N^6 \) computational cost. From all the listed contributions, the computational bottleneck will be observed for the terms with \( OV^5 \) cost, since the size of unoccupied orbital space is 5-10 times larger than the size of occupied space in the case of double-zeta basis sets; while for larger basis sets, this ratio is even more sever. Energy contribution of the \( OV^5 \) cost is given as:

\[
E_{\text{(T)}}^{[4]} \leftarrow \sum_{\alpha} w_{\alpha} \left( \sum_{abcdef} \sum_{ie} \alpha_{\alpha\beta}^{ive\alpha\beta} \sum_{jk} \left( \alpha t_{e\alpha}^{ja\beta} t_{f\beta}^{jb} - 2 \alpha t_{e\alpha}^{ja\beta} t_{f\beta}^{jb} \right) \right) + \sum_{abcdef} \sum_{ie} \alpha_{\alpha\beta}^{ive\alpha\beta} \sum_{jk} \left( 4 \alpha t_{e\alpha}^{ja\beta} t_{f\beta}^{jb} - 2 \alpha t_{e\alpha}^{ja\beta} t_{f\beta}^{jb} \right) - \sum_{abcdef} \sum_{ie} g_{\alpha\beta}^{ive\alpha\beta} \sum_{jk} \left( 4 \alpha t_{e\alpha}^{ja\beta} t_{f\beta}^{jb} - 2 \alpha t_{e\alpha}^{ja\beta} t_{f\beta}^{jb} \right).
\]
with $\alpha g$ and $\alpha t$ being Laplace transformed two-electron integrals and doubles amplitudes for the $\alpha$-th quadrature point, respectively, as defined by:

$$
\alpha g_{iebc} = e^{-f_{i}^{a}s_{a}/2} e^{-f_{e}^{c}s_{c}/2} g_{iebc},
$$

$$
\alpha t_{jk}^{ae} = e^{-f_{j}^{b}s_{b}/2} e^{-f_{k}^{a}s_{a}/2} t_{jk}^{ae},
$$

As it is evident from Eq. (5.12), such formulation does not require evaluation of the $W_{ijk}^{abc}$; instead, it allows construction of intermediate tensors, for example:

$$
\alpha G_{ba}^{ef} = \sum_{ia} \alpha g_{iebc} \alpha g_{ac}^{if},
$$

$$
\alpha T_{fa}^{bj} = \sum_{jk} \alpha t_{jak} \alpha t_{fa}^{jk},
$$

that are of $OV^5$ and $O^2V^4$ cost, respectively. In total, we have three such intermediates with $OV^5$ cost due to the factorized form of Eq (5.12). This can be simplified furthermore by employing density fitting approximation. Density fitting (DF) is a tensor factorization that is used to decompose the four-center two-electron repulsion integrals as a product of three-center and two-center two-electron integrals as:

$$
(\mu \nu | \rho \sigma) = \sum_{K \Lambda} (\mu \nu | K)(K | \Lambda)^{-1}(\Lambda | \rho \sigma) = B^{\Lambda}_{\mu \nu} B^{\Lambda}_{\rho \sigma},
$$

$$
B^{\Lambda}_{\mu \nu} = \sum_{K} (\mu \nu | K)(K | \Lambda)^{-\frac{1}{2}},
$$
where $K$ and $\Lambda$ come from an auxiliary basis set. Density fitting can be used to reduce last two contributions in Eq. (5.12) as follows:

\[
\left( \sum_{ic} \alpha g_{eb} g_{fa} \right) \alpha \tilde{T}_{ea} \approx \sum_{ic} \alpha \tilde{B}_{ic}^{K} \alpha \tilde{B}_{ic}^{K} \alpha \tilde{B}_{ic}^{\Lambda} \alpha \tilde{B}_{ic}^{\Lambda} \alpha \tilde{T}_{ea} = \\
\sum_{ic} \left( \alpha \tilde{B}_{eb}^{K} \alpha \tilde{T}_{ea} \right) \alpha \tilde{B}_{fa}^{\Lambda} \left( \alpha \tilde{B}_{ic}^{\Lambda} \alpha \tilde{B}_{ic}^{K} \right).
\]  

(5.19)

Similarly, the third term is given as:

\[
\left( \sum_{ic} \alpha g_{bc} g_{fa} \right) \alpha \tilde{T}_{ea} \approx \sum_{ic} \alpha g_{bc} \alpha \tilde{B}_{ic}^{\Lambda} \alpha \tilde{B}_{ic}^{\Lambda} \alpha \tilde{T}_{ea} = \\
\sum_{ic} \left( \alpha \tilde{B}_{fa}^{\Lambda} \alpha \tilde{T}_{ea} \right) \alpha g_{bc} \alpha \tilde{B}_{ic}^{\Lambda} \left( \alpha \tilde{B}_{ic}^{\Lambda} \right)
\]  

(5.20)

where $\alpha \tilde{T}_{ea} = \sum_{jk} \left( 4 \alpha \tilde{I}_{ea}^{jk} \alpha \tilde{I}_{eb}^{jk} - 2 \alpha \tilde{I}_{ea}^{jk} \alpha \tilde{I}_{fa}^{jk} \right)$ and $\alpha \tilde{B}$ are Laplace transformed three-center two-electron integrals for the $\alpha$-th quadrature point. The computational cost for Eq. (5.19) and Eq. (5.20) is at most $XV^{4}$, where $X$ is the size of auxiliary basis. Unfortunately, the same technique can not be applied to the first contribution from Eq. (5.12) since three-center two-electron integrals do not share more than one common index with tensor $\tilde{T}$. We exploit density fitting approximation on the rest of the intermediates that have large cost, reducing 24 terms of $O^{2}V^{4}$ cost at only 8 that occur from Eq. (5.10) while reducing computational cost of all 4 terms of $O^{2}V^{4}$ arising from Eq. (5.11). This tremendously increases the computational efficiency of the algorithm. To summarize, we have trade $O^{3}V^{4}$ cost of the canonical triple correction (T) with the cost of $n_{q} \times O V^{5}$. This means that as long as $O^{2}$ is greater than $n_{q} \times V$, Laplace transform (T) will have more favorable performance, which is true for the systems with large number of occupied orbitals and small basis sets that are of double-zeta quality.
Quadrature weights and roots \((w_\alpha \text{ and } s_\alpha)\) can be determined by one of the several techniques: the least squares,\(^{159,160,169}\) the Euler-McLaurin formula (radial quadrature scheme in DFT),\(^{170,171}\) the minimax scheme,\(^{172}\) the Gauss-Laguerre or the Gauss-Legendre quadrature.\(^{161}\) In this work, we employ the Gauss-Legendre integration due to its simplicity and its details are discussed in Ref. 161; however, for the clarity of presentation, we will revisit it briefly. A convenient variable substitution of \(s = -\frac{1}{\beta} \ln x\) can be inserted into Eq. \((5.9)\) giving:

\[
\int_0^\infty e^{-D_{ijk} s} ds = -\frac{1}{\beta} \int_0^1 x^{-\frac{D_{ijk}}{\beta} - 1} dx \approx -\frac{1}{\beta} \sum_{\alpha} w_\alpha x_\alpha^{-\frac{D_{ijk}}{\beta} - 1} \tag{5.21}
\]

where the integration interval over infinite limit is substituted by integration on interval 0 to 1, where parameter \(\beta\) is the triples gap, while the quadrature roots and weights, \(s_\alpha\) and \(w_\alpha\), are found using the Gaussian quadrature. The accuracy is dependent on the number of quadrature points that are used to approximate an exact integral. In the next section, we will discuss the accuracy with respect to the number of quadrature points on chemically relevant systems.

### 5.3 Computational details

All calculations have been carried out with a developmental version of massively parallel code, MPQC version 4 on the ”Blueridge” cluster at Virginia Tech. ”Blueridge” is a commodity cluster equipped with two Intel Xeon E5-2670 CPUs (332 GFLOPS) and 64 GB of RAM on each node. TiledArray and MPQC4 were compiled with GCC 5.1.0 with Intel MPI 5.0
and serial library in MKL 11.2.3. Furthermore, all calculations launched 1 MPI process per node and 16 threads per MPI process using a tile size 20. The accuracy of the method with respect to number of quadrature points was studied on the 18-water cluster, uracil dimer and pentacene dimer. For the 18-water cluster, we have used optimized TIP4P geometry employing the correlation-consistent cc-pVDZ basis set accompanied by the corresponding cc-pVDZ-RI density-fitting basis set. The calculations on uracil dimer have employed the cc-pVDZ-F12, a correlation-consistent basis set optimized for the explicitly correlated calculations along with the aug-cc-pVDZ-RI density-fitting basis set. The geometry of a uracil dimer corresponds to the complex 26 from the S66 data set. Lastly, calculations on the pentacene dimer were carried out using the 6-31G*(0.25) basis set along with the aug-cc-pVDZ-RI auxiliary basis set. Computational performance with respect to the number of nodes has been carried out on the same 18-water cluster employing the same basis set duo, cc-pVDZ/cc-pVDZ-RI. The geometry of the DNA base-pair (GC-dDMP-B) has been taken from Ref. 176.

5.4 Results

5.4.1 Quadrature accuracy

Accuracy of the method is highly dependent on the number of quadrature points required to approximate the exact integral defined by Eq. (5.9). Most common methods for finding
quadrature weights and roots are the least squares,\textsuperscript{159,160,169} Euler-McLaurin formula,\textsuperscript{170,171} minimax algorithm,\textsuperscript{172} Gauss-Laguerre and Gauss-Legendre quadrature.\textsuperscript{161} The comparison of quadrature schemes was done by Takatsuka et al. in the context of the Laplace transform MP2 method. They showed that the Gauss-Legendre is more accurate than the Gauss-Laguerre scheme, while the minimax algorithm is more favorable over the Gauss-Legendre in the case of benzene and ozone.\textsuperscript{172} The minimax algorithm minimizes the Chebyshev norm of the quadrature error on the given interval. Superiority in performance of the minimax algorithm is even more pronounced if all the electrons are correlated. Their work has been commented by Kats et al.\textsuperscript{177} showing that the least squares method has comparable accuracy to the minimax algorithm. In the context of the Laplace transform (T), Scuseria and co-workers have found that 4 quadrature points are required for achieving the maximum error of several \( \mu E_h \) on G2-1\textsuperscript{178,179} data set with various basis sets.\textsuperscript{161} The G2-1 data set is a set of very small molecular systems with up to a few atoms. They observed similar behavior on the water trimer and hexamer with the cc-pVDZ basis set where 3 quadrature points give \( \mu E_h \) accuracy with respect to the conventional (T). Moreover, they showed that in the case of water trimer, the computation time with the Laplace transform (T) over the conventional (T) is beneficial only for one quadrature point, while in the case of water hexamer, it is beneficial up to three quadrature points. More recently, Schmitz and Hättig have investigated the quadrature accuracy in the context of their reduced-scaling PNO-CCSD(T) method on 11 organic reactions containing small to medium-sized molecules. They found that the maximum discrepancy with respect to (T) of 0.1 kJ/mol is achieved by
3-4 quadrature points with triple-zeta basis set. The quadrature scheme that they have employed was the least squares method. This is in agreement with the findings of Nagy and Kállay, where they employed the minimax algorithm. They found that maximum absolute errors are in the range of 0.04-0.4 kJ/mol for NWH data set with various basis sets. Additionally, they investigated the absolute accuracy on the molecules of angiotensin, androstendion and aumin, concluding that for all these systems, the accuracy of $\mu E_h$ is obtained with 6 quadrature points. Our massively parallel DF-CCSD(T) implementation allows us to calculate and benchmark LT-(T) method for much larger systems that have been previously possible. We have limited our discussion on the double-zeta basis sets, since explicit correlation correction became a routine tool allowing the use of double-zeta basis sets only and obtaining highly accurate results. For assessment, the accuracy with respect to the number of quadrature points has been studied on both absolute and binding energies for 18-water cluster, uracil dimer and pentacene dimer. Binding energies of these systems represent chemically more relevant quantities than the absolute energies; therefore, we have included them in our analysis. Binding energy has been calculated using the monomer geometry at the dimer geometry. The results are summarized in the Table 5.1. The $\Delta_{abs}$ represents absolute error in units of $\mu E_h$, while $\Delta_{b.e.}$ stands for the binding energy error in kJ/mol. All errors are relative to the DF-CCSD(T) method. Note that with infinite number of quadrature points, these two approaches, (T) and LT-(T), will yield exactly the same results. In the case of 18-water cluster, $\mu E_h$ accuracy has been achieved by using only 3 quadrature points with error of only -5.1 $\mu E_h$. Inclusion of more quadrature points further
reduces the error. This is in agreement with previous findings by Scuseria. Binding energy error follows the same trend, showing that the error of 0.754 kJ/mol is obtained by only two quadrature points, while total (T) correction contribution to binding energy at cc-pVDZ basis set is ~47 kJ/mol, making this error relatively small. With three quadrature points, the error is furthermore reduced to sub 0.1 kJ/mol accuracy. This error is further reduced by taking more quadrature points.

Table 5.1: Accuracy of the Laplace transform (T), \((L_T-(T))\), with respect to the number of quadrature points \(n_q\) for the selected chemical systems.

<table>
<thead>
<tr>
<th>(n_q)</th>
<th>((H_2O)<em>{18}) (\Delta</em>{abs}(\mu E_h))</th>
<th>(\Delta_{b.e.}(kJ/mol))</th>
<th>Uracil dimer (\Delta_{abs}(\mu E_h))</th>
<th>(\Delta_{b.e.}(kJ/mol))</th>
<th>Pentacene dimer (\Delta_{abs}(\mu E_h))</th>
<th>(\Delta_{b.e.}(kJ/mol))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>560.5</td>
<td>0.754</td>
<td>15358.2</td>
<td>2.350</td>
<td>101843.5</td>
<td>188.1</td>
</tr>
<tr>
<td>3</td>
<td>-5.1</td>
<td>-0.017</td>
<td>2488.8</td>
<td>0.476</td>
<td>16131.2</td>
<td>38.1</td>
</tr>
<tr>
<td>4</td>
<td>-3.4</td>
<td>-0.009</td>
<td>494.4</td>
<td>0.114</td>
<td>1624.7</td>
<td>4.098</td>
</tr>
<tr>
<td>5</td>
<td>-2.0</td>
<td>-0.005</td>
<td>96.7</td>
<td>0.027</td>
<td>136.1</td>
<td>0.352</td>
</tr>
<tr>
<td>6</td>
<td>-1.3</td>
<td>-0.003</td>
<td>16.6</td>
<td>0.005</td>
<td>11.6</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Another example is a chemically more relevant system: the uracil dimer with a realistic cc-pVDZ-F12 basis set that is optimized for calculations with explicitly correlated methods, MP2-F12 and CCSD-F12. This basis set is favored for archiving greater accuracy than its aug-cc-pVDZ counterpart in the case of noncovalent interactions. Recently, we have shown that the error for such basis set in the context of the CCSD-F12 method is \(<0.1\ kcal/mol\) with respect to CCSD-F12/CBS for the binding energy of the uracil dimer. From Table 5.1,
it can be seen that absolute error of 16.6 $\mu$E$_h$ is achieved with 6 quadrature points, while more than 4 quadrature points show sub $m$E$_h$ error. Binding energy gives a satisfactory error of 0.476 kJ/mol with 3 quadrature points with (T) contribution to binding energy of $\sim$8 kJ/mol, which is around 6% error. The error in binding energy rapidly vanishes as the number of quadrature points increases. The last test case is on the pentacene dimer where geometry is optimized at the B97-D/6-31G(d,p) level of theory.$^{181,182}$ As for the basis set, we have employed the 6-31G*(0.25) basis set. The 6-31G*(0.25) basis set is obtained by changing the exponent of $d$ functions of the 6-31G* basis set from 0.8 to 0.25. It has been extensively validated on systems containing noncovalent interactions showing better or equal quality as the aug-cc-pVDZ basis set.$^{183,184}$ To gauge the accuracy of the 6-31G*(0.25) basis set, we have tested its accuracy on the uracil dimer on binding energy contributions of (T) using the aug-cc-pVXZ and cc-pVXZ-F12, with $X$=D and T. The results are summarized in Table 5.2. Both triple-zeta basis sets, namely aug-cc-pVTZ and cc-pVTZ-F12, are in close agreement that is within 0.2 kJ/mol. Furthermore, the basis set cc-pVDZ-F12, as discussed previously, outperforms the aug-cc-pVDZ basis set, yielding relatively small error of $\sim$0.5 kJ/mol with respect to the triple-zeta counterparts, while aug-cc-pVDZ accounts for $\sim$1.5 kJ/mol in difference with respect to the triple-zeta basis sets. The 6-31G*(0.25) basis set yields similar result as cc-pVDZ-F12 and is in much better agreement than the aug-cc-pVDZ basis set in the case of the uracil dimer. Results presented in Table 5.2 gives us an encouragement to employ 6-31G*(0.25) basis set for study of binding energies in large molecular systems. From Table 5.1, we can see that the accuracy of 11.6 $\mu$E$_h$ is achieved.
Table 5.2: Perturbative triple correction contribution (T) to the binding energy (kJ/mol) of the uracil dimer for several basis sets.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>(T) binding energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>aug-cc-pVDZ</td>
<td>-9.984</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>-8.416</td>
</tr>
<tr>
<td>cc-pVDZ-F12</td>
<td>-8.047</td>
</tr>
<tr>
<td>cc-pVTZ-F12</td>
<td>-8.583</td>
</tr>
<tr>
<td>6-31G*(0.25)</td>
<td>-8.040</td>
</tr>
</tbody>
</table>

with 6 quadrature points, which is similar to the uracil dimer; however, much larger error is observed in the case of 2 quadrature points that is in the order of 0.1 E\textsubscript{h}. Acceptable error in binding energy of <1 kJ/mol is achieved with 5 quadrature points. Therefore, we can conclude that more quadrature points are required for larger systems with small triple gap.

In the next section, we will discuss the performance and the cost of (T) versus LT-(T) for these three studied systems.

### 5.4.2 Computational performance

We have tested the computational performance with respect to the number of computing nodes. As a testing system, we have used 18-water cluster with cc-pVDZ basis set. Such a system is of moderate size with 432 basis functions. All calculations have been performed with 16 Intel Xeon E5-2670 2.60 GHz CPU cores. We have tested the performance on the
1, 2, 4, 8, 16 and 32 nodes for the conventional perturbative triple method (T) and newly developed Laplace transform (T). Figure 5.1 shows the total time required for computation with both (T) (blue) and LT-(T) (green) approaches as a function of the number of nodes. The wall clock timings are given in minutes. For LT-(T), we have used 3 quadrature points. It is evident from the Figure 5.1 that the computation time for the LT-(T) with 3 quadrature points is about 3 times faster than the (T) with up to 16 computing nodes, while for 32 nodes, this ratio drops to less than 2. Numerical values of timings are given in Table 5.3. The LT-

Figure 5.1: Wall clock time (minutes) for (T) and LT-(T) with respect to the number of computing nodes for the 18-water cluster. LT-(T) timings are with 3 quadrature points.
(T) (green) computation time roughly follows the straight line up to the 16 computing nodes, while there is slight drop of performance; whereas in the case of the conventional (T), the opposite behavior is observed for larger number of nodes. Figure 5.2 shows the computational speedup with respect to the number of nodes. Both methods are favored for good scaling behavior; thus, the conventional (T) method show an excellent 26.4 speedup going from 1 to 32 computing nodes, while the LT-(T) method shows a bit poorer speedup of 15.0. It is worth noting that scaling performance for the CCSD method for the same system is 14.9.

Figure 5.2: Computational speedup for (T) and LT-(T) with respect to the number of computing nodes. LT-(T) are with 3 quadrature points.
Furthermore, all results are summarized numerically in Table 5.3 and the computation time is given in minutes.

Table 5.3: Wall clock times (in minutes) of different DF-CCSD(T) contributions with respect to the number of computing nodes for 18-water cluster. LT-(T) timings are with 3 quadrature points.

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>CCSD/iter</th>
<th>(T)</th>
<th>LT-(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.7</td>
<td>1671.0</td>
<td>582.9</td>
</tr>
<tr>
<td>2</td>
<td>20.4</td>
<td>857.3</td>
<td>340.6</td>
</tr>
<tr>
<td>4</td>
<td>10.8</td>
<td>437.8</td>
<td>181.1</td>
</tr>
<tr>
<td>8</td>
<td>6.0</td>
<td>232.5</td>
<td>103.3</td>
</tr>
<tr>
<td>16</td>
<td>3.5</td>
<td>146.2</td>
<td>58.5</td>
</tr>
<tr>
<td>32</td>
<td>2.2</td>
<td>63.3</td>
<td>38.9</td>
</tr>
</tbody>
</table>

Table 5.4 contains the wall clock timings for (T) and LT-(T) of three different molecules studied in the last sections. The timings regarding LT-(T) method are for 1 quadrature point. The 18-water cluster has the ratio of active occupied and unoccupied orbitals, O/V, of 72/342. In the case of 1 node, the computation time of (T) is 1671 min, while for LT-(T) with one quadrature point, it is only 194 min. The CCSD wall clock time for each iteration takes 33 minutes. Majority of computation time, about 70%, is spent on computation of a single contribution of $OV^5$ cost. Furthermore, as discussed in the previous section, the $\mu E_h$ accuracy is achieved by only three quadrature points; therefore, LT-(T) offers a much cheaper
way for computation of the perturbative triple correction with only 35% of computation time relative to (T). In the case of uracil dimer, only with less or equal to 3 quadrature points, the

Table 5.4: Wall clock times (in minutes) for CCSD, triple correction (T) and Laplace transform triple correction LT-(T) for several representative molecules.

<table>
<thead>
<tr>
<th>System</th>
<th>Basis set</th>
<th>Atoms</th>
<th>O/V</th>
<th>Nodes</th>
<th>CCSD (T)</th>
<th>LT-(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H$<em>2$O)$</em>{18}$</td>
<td>cc-pVDZ</td>
<td>54</td>
<td>72/342</td>
<td>1</td>
<td>33</td>
<td>1671</td>
</tr>
<tr>
<td>uracil</td>
<td>cc-pVDZ-F12</td>
<td>12</td>
<td>21/247</td>
<td>1</td>
<td>0.7</td>
<td>34</td>
</tr>
<tr>
<td>uracil dimer</td>
<td>cc-pVDZ-F12</td>
<td>24</td>
<td>42/494</td>
<td>1</td>
<td>30</td>
<td>1084</td>
</tr>
<tr>
<td>pentacene</td>
<td>6-31G**(0.25)</td>
<td>36</td>
<td>51/263</td>
<td>2</td>
<td>3.7</td>
<td>126</td>
</tr>
<tr>
<td>pentacene dimer</td>
<td>6-31G**(0.25)</td>
<td>72</td>
<td>102/526</td>
<td>32</td>
<td>16</td>
<td>747</td>
</tr>
</tbody>
</table>

*Ratio between active occupied and unoccupied orbitals

*CCSD timings are for 1 iteration

*Wall clock timing for 1 quadrature point

computation time for LT-(T) is lower than the (T). This is due to the fact that the benefit from using LT-(T) comes only if the $O^2 > n_q \times V$ is satisfied, where $O$ and $V$ are the number of active occupied and unoccupied orbitals, respectively. This is indeed in agreement that if we include 4 quadrature points, calculation of LT-(T) will be longer than (T). As we discussed in the previous section, with 3 quadrature points, we recover about 0.5 mE$_m$ accuracy. This, however, does not affect binding the energy where with 3 quadrature points error is $\sim$0.5 kJ/mol. Our last example is the pentacene/pentacene dimer duo. From Table 5.4, one can see that computation of the (T) is roughly equal to computation of 8 quadrature points with
32 computing nodes. However, as discussed previously, the adequate chemical accuracy of $< 0.4 \text{ kJ/mol}$ is achieved with 5 quadrature points along with $< 0.2 \text{ mE}_h$ accuracy in absolute energy. LT-(T) with 5 quadrature points accounts for about 60% of computation time with respect to the computation time of conventional (T) correction. Additionally, it is worth comparing the computation times of CCSD and LT-(T) due to the fact that they have the same computational scaling of $\mathcal{O}(N^6)$. The results presented in Table 5.4 suggest that the LT-(T) computation time of 1 quadrature point is 5-10 iterations of the CCSD computation time.

Our last example is the DNA base pairs GC-dDMP-B, guanine-cytosine deoxydinucleotide monophosphate in B-conformation neutralized by a sodium cation, a system with 64 atoms and with 6-311++G** basis set, accounting for 103 active occupied orbitals and 1042 basis functions. This system was used by Kowalski and co-workers\textsuperscript{176} and it makes one of the largest CCSD(T) calculations. Computation time for the (T) contribution using NWChem and MPQC4 is given in the Table 5.5.

Table 5.5: Wall clock timings (in hours) for computation of GC-dDMP-B with the 6-311++G** basis set.

<table>
<thead>
<tr>
<th></th>
<th>NWChem</th>
<th>MPQC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>20 000</td>
<td>64</td>
</tr>
<tr>
<td>CCSD</td>
<td>4.0</td>
<td>13.2</td>
</tr>
<tr>
<td>(T)</td>
<td>1.4</td>
<td>47.4</td>
</tr>
<tr>
<td>LT-(T)$^a$</td>
<td>...</td>
<td>8.0</td>
</tr>
</tbody>
</table>

$^a$Wall clock timing for 1 quadrature point
The NWChem calculation used 20,000 computing nodes on Blue Waters system, with peak performances and hardware very similar to the "Blueridge" system that we have used for obtaining the data presented here. Calculation time for (T) with NWChem and on 20,000 computing nodes is 1.4 hours, while with MPQC4 is 47.4 hours on 64 computing nodes. This means that the computation time in the case of NWChem vs. MPQC4 is 34 times faster, but on roughly 300 times more computing nodes. Computation time of LT-(T) with 1 quadrature point is 8.0 hours, which means that computation time for the canonical triple correction (T) is ~6 quadrature points. As we showed in the previous examples, up to 5 quadrature points is enough to reach adequate accuracy; therefore, LT-(T) is computationally more efficient than the conventional (T).

5.5 Conclusions

In this work, we present the Laplace transform of the perturbative triple correction (LT-(T)), implemented in the MPQC4 quantum chemistry package in massively parallel fashion. The presented methodology is very similar to the one proposed by Scuseria and co-workers; however, we go beyond by inclusion of density fitting approximation (DF), making the method several times more efficient due to fewer terms of $OV^5$ and $O^2V^4$ cost that need to be evaluated. We have tested the method accuracy based on the number of quadrature points for 3 systems: 18-water cluster, uracil dimer and pentacene dimer. We have found that for larger systems, more quadrature points are required to obtain adequate accuracy. Around
5 quadrature points are required to recover < 1 kJ/mol accuracy. In this work, we have used the Gauss-Legendre quadrature for finding the quadrature roots and weights and the accuracy can be even greater with the same number of quadrature points with different quadrature algorithm such as the least squares or minimax. Additionally, we have tested the strong scaling regime behavior on the system of 18-water cluster, and found that with 32 nodes, we have about 50% of efficiency. Lastly, we have compared (T) and LT-(T) implemented in the MPQC4 package and (T) from NWChem and showed that our method allows fast and accurate computation of (T) with modest commodity cluster affordable to many research groups. Additionally, we have benchmarked the small basis set on uracil dimer, that is designed to perform well for noncovalent interactions, providing accuracy of the much larger basis set cc-pVDZ-F12 and outperforming the aug-cc-pVDZ basis set. The generality of the code allows support of different quadrature techniques with minimal programming effort. Furthermore, it can be used in the context of local methods if localized orbitals are used due to orbital invariant nature of Laplace transform of the denominator.
Chapter 6

Explicitly correlated formalism for second-order single-particle Green’s function

6.1 Introduction

The Green’s function (GF), or propagator,\textsuperscript{185} formalism of many-body quantum mechanics is a complementary approach to traditional wave function methods for computation and interpretation of electronic structure. Whereas GF formalism is dominant in condensed phase physics as a step beyond the mean-field description, it has also enjoyed a sustained, albeit a less prominent, presence in the molecular electronic structure.\textsuperscript{186,187} The single-particle GF, or electron propagator, has primarily been employed as a computationally efficient route to post-mean-field ionization potentials (IP) and electron affinities (EA) and, more generally, spectral functions necessary to interpret various photoelectron spectroscopies; recently self-consistent GF theory has been revisited as a route to quantum embedding and to finite-temperature electronic structure.\textsuperscript{188,189}

Here we present a general, \textit{explicitly correlated} formalism for computing single-particle Green’s functions. By employing many-body basis sets with explicit dependence on the inter-electronic distances, it is possible to formulate wave function methods with radically improved convergence to the analytic (complete basis set) limit.\textsuperscript{34,35,121} In this work we demonstrate how to achieve the same goal for the electron propagator. The initial validation of our approach focuses on computation of molecular ionization potentials with the second-order approximation to the self-energy (GF2). The performance of the method is tested for ionization potentials (IP) on two sets of small to medium-sized molecules.\textsuperscript{190–192} Our approach goes beyond the recent explicitly correlated correction to the GF2 ionization potentials of
Ohnishi and Ten-no\textsuperscript{193} by incorporating the energy dependence as well as extending beyond the diagonal approximation. Although the accuracy of the GF2 approximation is limited, it should be sufficient to gauge the effects of the energy dependence on the explicitly correlated contributions. Combination of our approach with higher-order nondiagonal Green’s function approaches, such as the 2ph-TDA, ADC(3) and NR2 methods,\textsuperscript{194,195} which are typically more robust than the second-order counterpart, is straightforward and will be reported elsewhere.

### 6.2 Formalism

Ionization potentials (IPs) and electron affinities (EAs) are obtained from the poles of the electron propagator. With the zeroth-order defined as usual in ab-initio molecular electronic structure by the Hartree-Fock state, the poles are obtained from the Dyson equation:

\[ [\mathbf{F} + \Sigma(E)] \mathbf{c} = E \mathbf{c}, \]  

where \( \mathbf{F} \) is the Fock operator and \( \mathbf{c} \) provides the Dyson orbitals. The energy-dependent self-energy operator \( \Sigma(E) \) incorporates the post-HF correlation and orbital relaxation effects. Within the superoperator formalism,\textsuperscript{196,197} the self-energy operator is expressed as

\[ \Sigma(E) = (\mathbf{a}|\hat{H}\mathbf{t})(\mathbf{t}|(E - \hat{H})\mathbf{t})^{-1}(\mathbf{t}|\hat{H}\mathbf{a}), \]  

where \( \mathbf{a} \equiv \{a_i\} \) is a row vector of annihilation operators and \( \mathbf{t} \equiv \{a_{p_1p_2}^{p_3}, a_{p_1p_2p_3}^{p_4}, \ldots \} \) contains the complementary two-, three-, and higher-body counterparts of \( \mathbf{a} \). The orbitals occupied in the Hartree-Fock reference state are denoted by \( i, j, \ldots \). The corresponding unoccupied
orbitals expressed in the orbital basis set (OBS) used to solve the Hartree-Fock equations are denoted by \( a, b, \ldots \). General OBS orbitals will be denoted by \( p, q, \ldots \), whereas their complement in the complete basis set (CBS) will be labeled by \( \alpha, \beta, \ldots \); the full set of CBS orbitals is denoted by \( \gamma, \delta, \ldots \). Throughout this paper, we are using the standard tensor notation for products of annihilation \( (a_p) \) and creation \( (a_p^\dagger) \) operators normal ordered with respect to the physical vacuum, i.e. \( \cdots a_p^\dagger a_p \cdots a_q^\dagger a_q = a_{p_1}^\dagger a_{p_2} \cdots a_{q_m}^\dagger a_{q_m} \). In Eq. (6.2) the braket is defined as:

\[ (X|Y) \equiv \langle 0 | [X^\dagger, Y]_+ |0 \rangle, \quad (6.3) \]

with \( |\hat{H}X \rangle \equiv [X, \hat{H}]_- |0 \rangle \), where \( |0 \rangle \) is the Hartree-Fock reference wave function. The leading order correction to the self-energy occurs at the second-order perturbation theory obtained with the usual Möller-Plesset partitioning of the Hamiltonian:

\[ \Sigma_{pq}(E) = \frac{1}{2} \sum_{ij} \frac{\bar{g}_{ij}^p g_{ij}^q}{E + \epsilon_a - \epsilon_i - \epsilon_j} + \frac{1}{2} \sum_{ab} \frac{\bar{g}_{ab}^p g_{ab}^q}{E + \epsilon_a - \epsilon_a - \epsilon_b} \quad (6.5) \]

where \( \bar{g}_{rs}^p \equiv g_{rs}^p - g_{sr}^p \) is the antisymmetrized Coulomb integral, where \( g_{rs}^p \equiv \langle rs | r_{ij}^{-1} | pq \rangle \).

The first (2-hole-1-particle, or 2h1p) term in Eq. (6.5) describes the orbital relaxation effects
and the second ($2p1h$) accounts for the electron correlation effects. As shown by Ten-no and Ohnishi, the sum over unoccupied states in the second term is slowly convergent in an atom. This can be seen immediately by recognizing that at $k$th zeroth-order poles the $2p1h$ contribution to the $k$th diagonal element of self-energy is a sum of the corresponding MP2 pair energies:

$$\sum_{kk}^{2p1h}(\epsilon_k) = \frac{1}{2} \sum_{iab} \frac{\tilde{g}_{ki}^{ab}\tilde{g}_{ki}^{ab}}{\epsilon_k + \epsilon_i - \epsilon_a - \epsilon_b} = \sum_i \epsilon_{ki}^{\text{MP2}}. \quad (6.6)$$

The slow basis set convergence is thus in direct analogy with the slowly convergent error $O[(L+1)^{-3}]$ of a truncated partial wave expansion of the atomic MP2 energy. Motivated by the close connection of the diagonal elements of the second-order self-energy at the corresponding Koopmans pole with the MP2 energy contributions, Ten-no and Ohnishi suggested an additive (energy-independent) explicitly correlated correction for the second-order self-energy using the explicitly correlated MP2-F12 pair energies.

Here we demonstrate how to go beyond a simple additive correction by including proper energy dependence for the nondiagonal second-order Green’s function method. The goal is to provide a robust reference for more approximate schemes such as that of Ten-no and Ohnishi, as well as establish an explicitly correlated approach for general GF methods. We start by augmenting the slowly convergent $t = \{a_{ab}\}$ field operator with a geminal field operator. We postulate that the form of the geminal field operator is $t^\gamma = \{\frac{1}{4} \tilde{R}_{\alpha\beta}^{\alpha\beta} \tilde{a}_i^{\alpha\beta}\}$. The $\tilde{R}_{\alpha\beta}^{\alpha\beta}$ is obtained from $R_{\alpha\beta}^{\alpha\beta}$, the antisymmetrized matrix element of the geminal correlation
factor $f(r_{12})$: \(^\text{37}\)

\[
R_{ir}^{\alpha\beta} \equiv \langle ir | \hat{Q} f(r_{12}) | \alpha\beta \rangle. \tag{6.7}
\]

The tensors with tildes include the pair-spin projection due to singlet and triplet cusp conditions: \(^\text{38, 88}\)

\[
\hat{O}_{ij}^{\alpha\beta} \equiv \frac{1}{2} (C_0 + C_1) O_{ij}^{\alpha\beta} + \frac{1}{2} (C_0 - C_1) O_{ji}^{\alpha\beta} \tag{6.8}
\]

\[
= \frac{3}{8} O_{ij}^{\alpha\beta} + \frac{1}{8} O_{ji}^{\alpha\beta}
\]

where $C_{0,1} = 1/2, 1/4$ are the cusp coefficients for singlet and triplet pairs, respectively. \(^\text{69, 122}\)

Projector $\hat{Q}$ in Eq. (6.7) ensures that the geminal functions are orthogonal to Hartree-Fock as well as to the standard double excitations.

The explicitly correlated part of the self-energy is expressed as

\[
\Sigma(E) \leftarrow \langle a | \hat{H}^{(1)} t^\gamma \rangle \langle t^\gamma | (E - \hat{H}^{(0)}) t^\gamma \rangle^{-1} \langle t^\gamma | \hat{H}^{(1)} a \rangle. \tag{6.9}
\]

Resolution of the matrix elements will give:

\[
(a_p | \hat{W} t^\gamma) = \frac{1}{4} \tilde{g}^{\gamma\delta}_{kp} \tilde{R}^{kr}_{\gamma\delta} = \frac{1}{4} V^{kr}_{kp} \tag{6.10}
\]

while

\[
(t^\gamma | \hat{W} a_q) = \frac{1}{4} \tilde{R}^{\alpha\beta}_{is} g_{\alpha\beta} = \frac{1}{4} V^{iq}_{is}. \tag{6.11}
\]

The matrix elements of the resolvent are:

\[
(t^\gamma | \hat{F}_{N} t^\gamma) = \frac{1}{4} \tilde{R}^{\alpha\beta}_{is} F_{\gamma}^{\gamma\delta} \tilde{R}^{kr}_{\gamma\delta} - \frac{1}{8} F_{i}^{i} \tilde{R}^{\alpha\beta}_{is} \tilde{R}^{kr}_{\alpha\beta} \tag{6.12}
\]

\[
= \frac{1}{4} \tilde{B}^{kr}_{is} - \frac{1}{8} F^{i} X^{kr}_{is} \tag{6.13}
\]
and

$$E(t^\gamma | t^\gamma ) = \frac{1}{8} E^\delta_i R^\rho_{is} R^\rho_{is} = \frac{1}{8} E^\gamma_i X^\rho_{is},$$  \hspace{1cm} (6.14)

where $\hat{F}_N \equiv \hat{F} - E^{(0)}$ is the normal-ordered Fock operator. The $V, X, B$ are standard F12 intermediates and their programmable expressions can be found elsewhere.$^{29,48,49,104}$ Intermediates $V$ and $X$ were evaluated in the CABS approximation,$^39$ while for intermediate $B$, we utilized approximation D.$^{29}$ This formalism has been implemented in a developmental version of Massively Parallel Quantum Chemistry package (MPQC) version 4.$^{104}$ The implementation utilizes the TiledArray tensor library, which provides distributed parallel tensor routines, and hence the implementation is massively parallel.$^{162}$ Its performance is discussed in the next section.

6.3 Results and discussion

We have assessed the performance of the new approach by computing the IPs of a set of 21 small molecules$^{190}$ and 24 medium sized organic electron accepting molecules (OAM24).$^{191}$ The basis set for the OBS that we used is aug-cc-pVXZ with the corresponding density-fitting basis set, aug-cc-pVXZ-RI, where X=D,T,Q,5 (X represents basis set cardinal number) and aug-cc-pVXZ-CABS basis set for the calculations with explicit correlations.$^{82-84,127,199}$ All computations were performed with the frozen-core approximation. The Slater-type correlation factor, $f(r_{12}) = (1 - \exp(-\gamma r_{12})/\gamma)$, with $\gamma = 1.3, 1.9$ and 2.1 Bohr$^{-1}$ for the basis sets aug-cc-pVDZ, aug-cc-pVTZ and aug-cc-pVQZ has been used.
In the following text, we denote the nondiagonal second-order Green’s function method by GF2, whereas the GF2 method with fully energy-dependent explicitly correlated correction by GF2-F12. GF2(F12) denotes a hybrid approach in which the explicitly correlated correction is computed a posteriori, by solving the Dyson equation for the pole and Dyson orbitals at the GF2 level, followed by a single-shot evaluation of the explicitly correlated contribution to the self-energy with fixed energy and orbitals. Thus the difference between the GF2(F12) and GF2-F12 results will indicate the importance of the full energy dependence of the explicitly correlated terms in the second-order self-energy.

Table 6.1 presents the statistical averages of the basis set errors of IPs for 21 small molecular systems obtained with the GF2, GF2(F12) and GF2-F12 methods. The complete basis set (CBS) limit of GF2 has been estimated using the two-point (aug-cc-pVQZ to aug-cc-pV5Z) $X^{-3}$ extrapolation scheme. Statistical analysis shows that the mean absolute errors in eV (MAE/eV) for GF2 are 0.402, 0.169, 0.078 and 0.040 with the basis sets aug-cc-pVXZ (X=D,T,Q and 5); GF2(F12) gives MAEs of 0.066, 0.013 and 0.009, while GF2-F12 produced MAEs of 0.030, 0.012 and 0.009 for the X=D,T and Q. The maximum absolute error (MaxAE/eV) in the case of the GF2 method is 0.479 (CH$_3$CH$_2$CH$_3$), 0.209 (HF), 0.100 (HF and HCl) and 0.051 (HCl) with respect to X=D,T,Q and 5; MaxAE with the GF2(F12) method is 0.194 (HF), 0.044 (HF) and 0.020 (HF), while in the case of GF2-F12 MaxAE is 0.059 (CH$_3$OH), 0.032 (CH$_3$F) and 0.023 (HF and CH$_3$F) for X=D,T and Q, respectively. These statistical parameters are shown graphically in Fig. 6.1 where the x-axis is basis set cardinal number X and the y-axis represents the mean absolute error in eV (MAE).
Table 6.1: Mean absolute error (MAE) in eV and maximum absolute error (MaxAE) in eV for the IPs of the data set of 21 small molecules with respect to the CBS limit for several basis sets.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>MAE</th>
<th>MaxAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GF2</td>
<td>GF2(F12)</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>0.402</td>
<td>0.066</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>0.169</td>
<td>0.013</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td>0.078</td>
<td>0.009</td>
</tr>
<tr>
<td>aug-cc-pV5Z</td>
<td>0.040</td>
<td>-</td>
</tr>
</tbody>
</table>

^aCH₃CH₂CH₃, ^bHF, ^cCH₃OH, ^dCH₃F, ^eHF and HCl, ^fHF and CH₃F

values next to the points show MaxAE in eV. The IPs evaluated with the GF2(F12) and GF2-F12 approaches have dramatically smaller basis set errors than their GF2 counterparts. Furthermore, the rigorous GF2-F12 approach is preferred to the simpler GF2(F12) approach with the double-zeta basis.

Note that the use of approximation D for the B intermediate introduces negligible errors. The maximum absolute error between D and the more rigorous approximation C does not exceed 0.010 eV in the case of the aug-cc-pVDZ basis set, while for the larger basis sets, the error vanishes completely.

We also tested the explicitly correlated GF2 approaches by computing IPs of medium sized molecules in the OAM24 data set. The IP basis set error statistics are given in the Table 6.2.

The CBS limit has been calculated with the GF2 method using the two-point extrapolation
The results from Table 6.2 are in agreement with those presented for the small molecules and support the same conclusion. GF2-F12 shows very small errors, giving a MaxAE of only 0.039 eV in the case of the aug-cc-pVDZ and making this method much more accurate than the more costly alternative provided by GF2 with the aug-cc-pVQZ basis. GF2-F12 is also preferred to the more approximate GF2(F12) approach, which suggests the explicitly
Table 6.2: Mean absolute error (MAE) in eV and maximum absolute error (MaxAE) in eV for the IPs of the OAM24 data set with respect to the CBS limit for several basis sets.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>MAE</th>
<th>MaxAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GF2</td>
<td>GF2(F12)</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>0.372</td>
<td>0.081</td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>0.151</td>
<td>0.021</td>
</tr>
<tr>
<td>aug-cc-pVQZ</td>
<td>0.064</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup>maleic anhydride, <sup>b</sup>naphthalenedione, <sup>c</sup>nitrobenzene, <sup>d</sup>Cl<sub>4</sub>-benzoquinone, <sup>e</sup>benzoquinone, <sup>c</sup>phenazine

correlated correction to self-energy should indeed include the energy dependence properly.

### 6.4 Summary and Perspective

We have presented and efficient, massively parallel implementation of the explicitly correlated nondiagonal energy dependent GF2-F12 method. By including geminal field operators, we account for the missing electron correlation effects due to the incompleteness of the basis set. Numerical tests on small and medium molecules suggested that the ionization potentials computed with the new explicitly correlated GF2-F12 method in conjunction with the modest aug-cc-pVDZ basis set had smaller basis set errors than their non-explicitly-correlated GF2 counterparts with a much larger aug-cc-pVQZ basis set. The computation time of the first ionization potential of pyrimidine with the GF2/aug-cc-pVQZ method takes 404 while GF2-
F12/aug-cc-pVDZ method takes only 91 seconds (using four computing nodes with 24 Intel Xeon E5-2680 v3 2.50 GHz CPU cores). The corresponding absolute basis set errors are 0.061 and 0.018 eV, respectively. The proposed GF2-F12 method is easily extensible to higher-order Green’s function approaches; such efforts will be reported elsewhere.
Chapter 7

Conclusions
Techniques for reducing the computational complexity have been successfully applied to conventional methods for describing electron correlation effects such as in the CCSD(T) method\textsuperscript{20,25,27,113} However, it suffers from the errors associated with the use of finite basis sets that can be efficiently cured by inclusion of the explicit correlation correction or R12/F12.\textsuperscript{36} The main focus of this thesis is to bridge the gap between reduced scaling techniques and explicit correlation formalism. In Chapter 2, we start our discussion by reviewing modern R12/F12 framework in the context of the MP2 and coupled-cluster methods. In addition, we discuss modern machinery associated with the explicit correlation correction such as the RI approximation,\textsuperscript{39,72} density fitting,\textsuperscript{111} correlation factor,\textsuperscript{37} the SP ansatz,\textsuperscript{38} CABS singles,\textsuperscript{40,94} approximations of many-electron integrals,\textsuperscript{70–76} as well as basis sets optimized for explicitly correlated calculations.\textsuperscript{85} In Chapter 3, we present the explicitly correlated second-order Møller-Plesset method (DLPNO-MP2-F12) in which all post-Hartree-Fock steps are formulated in a linear scaling fashion.\textsuperscript{29} This is achieved by the use of the PNOs for the compact representation of the unoccupied space and by PAOs which allows for introduction of the domains.\textsuperscript{26,27,113} The method shows near-linear scaling behavior on the linear alkane chain with the computational scaling of $N^{1.6}$, recovering more than 99.9\% of correlation energy with the default truncation parameters. The computation time of the DLPNO-MP2-F12 is $>10$ times more expensive than the DLPNO-MP2 method; however, this is compensated for by the fact that the basis set errors are greatly reduced. In addition, the DLPNO-MP2-F12 lays a solid foundation for the linear-scaling explicitly correlated coupled-cluster singles and doubles method (DLPNO-CCSD-F12) discussed in Chapter 4.\textsuperscript{49} We have demonstrated
that the DLPNO-CCSD-F12 method shows a near-linear scaling behavior of $N^{1.5}$. The error introduced by truncations is only 0.4 kcal/mol of the binding energy with respect to the canonical result in the case of 20-water cluster. Furthermore, the methods accuracy has been benchmarked against smallest 5 systems contained in L7 data-set showing an excellent agreement with the reference results.\textsuperscript{144} This method relies on (T0) correction, which is a semi-canonical approximation to the (T) approach in which only diagonal elements of the Fock matrix are used. In Chapter 5, we present a step toward in discussing an orbital invariant formalism of the (T) that is achieved by the Laplace transform of the denominator that occur in the (T) correction. We present a massively parallel implementation of the Laplace transform CCSD(T) method within the density fitting framework in which exponent of $N^7$ is reduced to $N^6$.\textsuperscript{159,161} We show that around 6 quadrature points are required for achieving $\mu E_h$ accuracy on systems with more than 50 atoms. The methods discussed in the previous chapters can be applied on systems that are far from reach of the conventional canonical methods and thus allowing for investigations of materials that are used for light conversion. However, modeling of a such systems requires an accurate knowledge of the ionization potentials (IP) and electron affinities (EA). In Chapter 6, we describe how to incorporate explicit correlation correction into the Green’s function formalism that is used for calculation of IPs.\textsuperscript{78} We show that the GF2-F12 method removes errors associated with the basis sets for small and large molecular systems showing that the performance of GF2-F12/aug-cc-pVDZ is better than that of GF2/aug-cc-pVQZ at a significantly lower cost. The work presented in this thesis will set a stage for further developments in reduced scaling explicitly correlated
methods. Furthermore it will be a useful benchmarking method for parametrizing the DFT functionals. The GF2-F12 method will open a new route due to its simplicity for augmenting more accurate Green’s function methods such as NR2, allowing for accurate prediction of IP/EA of a large molecular systems and periodic systems.
Publication List


Bibliography


81 Rezác, J.; Riley, K. E.; Hobza, P. *Journal of Chemical Theory and Computation* 2011, 7(8), 2427–2438.


126 Yousaf, K. E.; Peterson, K. A. *The Journal of Chemical Physics* **2008**, *129*(18), 184108.


143 Christensen, A. S.; Elstner, M.; Cui, Q. The Journal of Chemical Physics 2015, 143(8), 084123.


166 Dunlap, B. I.; Connolly, J.; Sabin, J. The Journal of Chemical Physics 1979, 71(8), 3396–3402.


186 Ortiz, J. V. Wiley Interdisciplinary Reviews: Computational Molecular Science 2013, 3(2), 123–142.


