LUND UNIVERSITY

New Approaches to Large-Scale Electronic Structure Calculations

Aquilante, Francesco

2007

Link to publication

Citation for published version (APA):
Aquilante, F. (2007). *New Approaches to Large-Scale Electronic Structure Calculations*. [Doctoral Thesis (compilation), Computational Chemistry]. Theoretical Chemistry, Lund University.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 21. Dec. 2025

New Approaches to Large-Scale Electronic Structure Calculations

Thesis submitted for the degree of
Doctor of Philosophy by

Francesco Aquilante

Department of Theoretical Chemistry

LUNDS
UNIVERSITET

AKADEMISK AVHANDLING för avläggande av filosofie doktorsexamen vid naturvetenskapliga fakulteten, Lunds Universitet, kommer att offentligen försvaras i

Hörsal B, Kemicentrum, fredagen den 5 Oktober 2007, kl 13.15

Fakultetens opponent är Dr. Frederick R. Manby,
School of Chemistry, University of Bristol, UK

DOKUMENTDATABLAD enl SIS 61 41 21

Organization LUND UNIVERSITY *Department of Theoretical Chemistry* *P.O.B. 124* *S-221 00 Lund* *Sweden*	Document name DOCTORAL DISSERTATION Date of issue *October 5, 2007* Sponsoring organization
Author(s) *Francesco Aquilante*	

Title and subtitle

New Approaches to Large-Scale Electronic Structure Calculations

Abstract

Recent years have witnessed a growing interest of the scientific community for the use of ab initio and density functional theory methods in theoretical studies of molecules containing many atoms. However, the ''scaling wall'' of some of the most accurate of such methods is often an obstacle for their applicability to systems of real-life interest, e.g. in biochemistry and nanotechnologies, one bottleneck being the evaluation and storage of the two-electron repulsion integrals.

We have explored the possibility to avoid the expensive evaluation of the full integral matrix by designing and implementing approximate approaches based on Cholesky decomposition techniques. The results show the general applicability of this approximation to any quantum chemical method with substantial computational savings compared to conventional implementations -- not uncommonly of 1-2 orders of magnitude. At the same time, the loss of accuracy is minimal and can be systematically reduced at the price of sustainable additional costs.

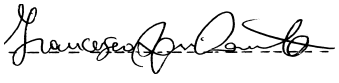
A major theoretical achievement of this study has been the reformulation of the Cholesky approximation in terms of solution to a ''density fitting'' variational problem. This has lead to an elegant formulation of the analytic derivatives of the Cholesky vectors, a long-standing lack of this technique. Moreover, this observation has paved the way to a new generation of accurate density fitting approximations free from biases and derived fully ab initio.

Key words: *Two-electron integrals, Cholesky decomposition, density fitting, resolution of identity, orbital localization, reduced-scaling, analytic gradients, electron correlation, perturbation theory*

Classification system and/or index termes (if any):

Supplementary bibliographical information:		Language *English*
ISSN and key title:		ISBN *978-91-7422-169-5*
Recipient's notes	Number of pages *196*	Price
	Security classification	

Distribution by (name and address)
I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  Date *September 11, 2007*

Alla mia cara Mamma.

Two things fill the mind
with ever new and increasing wonder
and awe - the starry sky above me
and the moral law within me.

Immanuel Kant

© **Francesco Aquilante, 2007**

Lund University, Sweden
Doctoral dissertation in Theoretical Chemistry

ISBN 978-91-7422-169-5

Written in LaTeX
Printed by Media Tryck, Lund

Contents

Preface

"Large-scale" has a twofold meaning in quantum chemistry. It can refer to the presence of many atoms in the system that we want to study, or it relates to the large number of atomic basis functions used to achieve an accurate description of the electronic structure of the system. In both cases, the effective meaning of the expression translates as a sheer increase of the computational costs for the corresponding electronic structure calculation. The only possibility to elude such bottleneck is to devise approximate methods that can cope with the increase of the computational dimensionality of the problem. As the premises are different in the two cases, the proposed solutions for the *many-atom* problem and for the *many-basis* problem do not usually overlap, although they do benefit from their mutual action in many instances.

The *many-basis* problem is more closely related to the subject of the present thesis, although the computational technology that I developed during my PhD studies is meant to be of great help for both aspects of the *large-scale* problem of electronic structure theory. The main objective of my studies has been the reduction of the computational demand associated with the storage and manipulation of the two-electron repulsion integral matrix. By exploiting the potential of the Cholesky decomposition applied to this matrix, I have shown that most quantum chemical methods can be formulated in such a way that significant computational advantages are attainable, with a minimal trade of accuracy. More importantly, this approach is unique in respect to systematic improvability of the results.

A particularly unpleasant shortcoming of all previous computer implementations of Cholesky-based quantum chemical models, known as "exchange problem", has been tackled and eventually removed. Among other new and promising insights, a major theoretical achievement has also been the reformulation of the Cholesky approximation in terms of solution to a "density fitting" variational problem. This has lead to an elegant formulation of the analytic derivatives of the Cholesky vectors, a long-standing deficiency of this technique. Moreover, this observation has paved the way to a new generation of accurate density fitting approximations free from biases and derived fully *ab initio*.

The structure of the present thesis is as follows. The first chapter introduces some basic concepts of quantum mechanics, *as I see it*. It is primarly meant to be of help to the reader who is not completely acquainted with the basic terminology and with the foundations of some of the most relevant quantum chemical methods. The issue of the computational scaling of the quantum chemical methods with the size of the system is presented in chapter 2, where I also introduce the Cholesky decomposition technique and show how it helps in reducing the computational costs of electronic structure calculations. Chapter 3 deals with the LK screening technique as a solution to the exchange problem. It also includes some basic survey on local correlation methods and a further analysis of the nature and usage of the Cholesky orbitals. The formulation of the analytical gradients for the Cholesky vectors and the hierarchies of *ab initio* density fitting approximations are discussed in chapter 4. Chapter 5 is a brief presentation of recent ideas on the possibility to use the variational formulation of the Cholesky approach in wider contexts than the ones presented earlier in the thesis. Finally, there is an appendix that collects a number of short manuscripts on various quantum chemical related topics. Mostly, they are reported exactly in the same way as I wrote them for private scientific correspondence with some of my colleagues at the department. I apologize in advance for possible bald statements. Some of the ideas enclosed in these manuscripts may hopefully be starting points for future developments, but to me they are mainly an historical memory of my *liberal studies*.

Except for the first two chapters, the form chosen to present the material of this thesis is that of personal *recollections* on the various topics. My intention has been to transmit in this way the disparate emotions that I have felt during my PhD studies — from the overwhelming desperation after yet another working hypothesis proven wrong, to the complete bliss reached when a simple idea translates into beautiful equations and these in turn lead to the *numbers* you so much longed for.

Populärvetenskaplig sammanfattning

Även om vi befinner oss i en tidsepok av informationsteknologi och superdatorer så utförs dagens forskning baserat på Galileos vetenskapliga principer. Trots det så finns det idag inget behov av att klättra upp i det lutande tornet i Pisa för att observera hur ekvationerna för rörelse, enligt den klassiska mekaniken, tar sig uttryck för fallande kroppar. Samma experiment kan idag prövas virtuellt på en dator. Det är faktiskt så de flesta forskare idag startar sina undersökningar eller prövar sina hypoteser. Sådana efterföljare till Einsteins *Gedankenexperiment*[1], så kallade datorsimuleringar, görs varje dag inom vetenskapens olika grenar, från fysik till genetik, från meteorologi till sociologi, och till en stor del inom kemin. Eftersom kemin är intimt förbunden med elektronernas egenskaper och beteende, och eftersom dessa partiklar i sin tur är icke-klassiska, så bestäms de rörelseekvationer som styr vardagliga kemiska fenomen av kvantmekanikens principer. Detta komplicerar livet för kemister. Inte så mycket från ett teoretiskt perspektiv utan snarare i termer av att beräkningstiden för kvantmekaniska datorsimuleringar oftast växer exponentiellt med antalet elektroner som skall studeras.

Dirac skrev en gång: *"Naturlagarna som behövs för den matematiska behandlingen av en stor del av fysiken och hela kemin är sålunda nu kända, och enda svårigheten är att tillämpningen av dessa lagar leder till ekvationer som är alltför komplicerade för att kunna lösas. Det är därför önskvärt att utveckla praktiska och approximativa metoder för att tillämpa kvantmekanik. Detta kan leda till att de viktigaste principerna för hur komplexa atomära system fungerar kan förklaras utan alltför svåra beräkningar."* Syftet med det forskningsarbete som summeras i denna avhandling är att bidraga, om än på en blygsam nivå, till utvecklingen av approximativa metoder som tillåter

[1]Detta är ett uttryck som kan härledas från en blandning av ett tyskt och ett latinskt ord och som bäst översätts till "tankeexperiment". Även om konceptet redan var känt för de grekiska och romerska filosoferna så myntades uttrycket i slutet på 1800-talet av den danske kemisten Hans Christian Ørsted. För vetenskapsmän och vanliga människor är dock uttrycket främst kopplat till Einsteins relativitetsteori.

teoretiska studier av molekylära modeller, vilka så bra som möjligt beskriver verkliga system. Även om det kanske för gemene man kan låta konstigt så klarar inte dagens kraftfulla datorresurser av att mätta de behov av aritmetiska och logiska operationer som krävs för noggranna kvantkemiska studier av molekyler med några hundra atomer. Under en tidsrymd motsvarande ett normalt experiment är det inte ens möjligt att utföra tillförlitliga kvantkemiska beräkningar för system av storleken under hundra atomer. Det finns inte heller några indikationer på att detta kommer att vara möjligt i framtiden, trots den positiva utveckling som Moore's lag, att datorkraften dubblas med tiden, ger vid hand.

Huvudproblemet som angrips i detta arbete är ökningen av beräkningskostnaden associerad med kvaliteten med vilken enskilda elektroner beskrivs. Den matematiska beskrivningen av system med många elektroner kan reduceras till en speciell kombination av de olika kvanttillstånden som är tillgängliga för varje enskild elektron. Dessa enskilda elektroniska kvanttillstånd (så kallade orbitaler) måste kunna beskrivas noggrant för att ge en korrekt beskrivning av ett större system. I en datorsimulering innebär detta att stora mängder data måste genereras, lagras, och sedan manipuleras under beräkningen. Det finns dock en möjlighet att denna information till stor del är redundant och därför, om man kan upptäcka denna redundans, är det möjligt att minska beräkningsbördan utan att sätta noggrannheten i resultatet av beräkningen på spel. Ett sätt att göra detta är att utnyttja en speciell matrisoperation som kallas Cholesky-uppdelning. Jag har på många sätt utnyttjat denna teknik i de arbeten som presenteras i denna avhandling och visat att den är ett effektivt redskap för att minska antalet parametrar som behövs för att beskriva den term i den kvantmekaniska energin för ett mång-elektronsystem som beskriver växelverkan mellan två elektroner. Cholesky-tekniken var känd långt före mina arbeten. Mina bidrag ligger i en undersökning av metodens noggrannhet och i en mera generell och teoretisk formulering av tekniken. Bland annat har dessa ansträngningar lett till ett antal nya insikter, vilka spänner från möjligheten att erhålla orbitalrepresentationer lokaliserade till vissa delar av en molekyl, till nya matematiskra uttryck för de krafter som styr hur atomerna konfigurerar sig i en molekylär struktur.

Popular Science Summary

In the era of information technology and supercomputing, the way of doing research is still based on the *scientific method* of Galileo. However, in today's world there is no need to climb the Leaning Tower of Pisa in order to observe the equations of motion of classical mechanics manifest themselves on a falling object. The same experiment can now be performed in the virtual reality created by a computer, and this is indeed how most contemporary researchers start their investigations or put on trial their hypothesis. Such progeny of the famous Einstein's *Gedankenexperiment*[1] are run everyday, under the more familiar name of *computer simulations*, in any branch of science, from physics to genetics, from meteorology to sociology, and to a large extent in chemistry. As the latter is intimately related to the nature and behavior of the electrons, and as in turn these are non-classical particles, the equations of motion used to rationalize many chemical phenomena of everyday life are determined by the principles of quantum mechanics. This somehow complicates the matter for chemists, not so much from the theoretical point of view, but rather in terms of the time required to perform the computation, which often grows exponentially with the number of electrons taking part in the process.

Dirac once wrote: *"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."* The objective of the research work summarized in the present thesis can be viewed as a modest contribution in the direction of developing

[1] This is a Latin-German mixed word that can be translated to English as "thought experiment". Although such concept was already known to Greek and Roman philosophers, the term was coined as late as the 19th century by the Danish chemist Hans Christian Ørsted. However, to scientists and common people, it remains mostly linked to Einstein's Theory of Relativity.

such approximate methods, which could allow theoretical studies on molecular models as representative as possible of real-life systems. As bizarre as it may appear to a novice in theoretical chemistry, in fact, even the relatively powerful computational resources available today are not sufficient to satisfy the demand of arithmetic and logical operations required by not even the most accurate quantum chemical calculations on molecules of few hundred atoms. Within a time frame compatible with routine experimental measurements, reliable quantum chemical predictions are often impossible even below the hundred-atom regime and there is no indication that things could change in the future, despite any positivist vision given by Moore's law on the doubling of computers power.

More in detail, the main problem addressed in this work is that of the increase of the computational costs with the quality of the single-particle descriptors. The mathematical description of a many-electron system, in fact, is reduced to particular superpositions of the possible quantum states of each electron. These single-electron states, so-called orbitals, need to be described very accurately in order to produce a correct representation for the whole systems. In a computer simulation, this implies that large sets of data need to be generated, stored and then manipulated during the calculations. However, there is a chance that some of this information is indeed redundant and if one is able to detect such redundancies, then one gets obviously a reduction of computational costs without jeopardizing the accuracy of the results. One way to exploit these occurrences is by employing a particular type of matrix operation known as Cholesky decomposition. I have extensively used such technique in the work described in the present thesis and shown that it is a very effective tool to reduce the number of parameters needed to describe the two-body interaction term of the quantum mechanical energy of a many-electron system. The Cholesky technology was already known before my work, however I contributed to an assessment and also to a more general theoretical formulation of it. Among other things, this effort has lead to a number of new insights, ranging from the possibility to obtain orbital representations which are localized in certain regions of the molecules, to a new formulation of mathematical expressions for the forces involved in the rearrangements of the atoms within a molecular structure.

List of Publications

I. **Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals**
F. Aquilante, T. B. Pedersen and R. Lindh, *Journal of Chemical Physics* **126** (2007) 194106–(11)

II. **Fast noniterative orbital localization for large molecules**
F. Aquilante, T. B. Pedersen, A. Sánchez de Merás and H. Koch, *Journal of Chemical Physics* **125** (2006) 174101–(7)

III. **Analytic derivatives for the Cholesky representation of the two-electron integrals**
F. Aquilante, T. B. Pedersen and R. Lindh, *manuscript*

IV. **Unbiased auxiliary basis sets for accurate two-electron integral approximations**
F. Aquilante, R. Lindh and T. B. Pedersen, *Journal of Chemical Physics*, accepted for publication

V. **Accurate *ab initio* density fitting for multiconfigurational self-consistent field methods**
F. Aquilante, T. B. Pedersen, B. O. Roos, A. Sánchez de Merás and H. Koch, *Journal of Chemical Physics*, submitted for publication

VI. **Cholesky decomposition based multiconfiguration reference perturbation theory (CD-CASPT2): Application to the spin state energetics of Co^{III}(diiminato)(NPh)**
F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh and B. O. Roos, *Journal of Physical Chemistry A*, submitted for publication

VII. **Minimal cost evaluation of the scaled opposite-spin second order canonical Møller-Plesset correlation energy**
F. Aquilante and T. B. Pedersen, *Chemical Physics Letters*, submitted for publication

Not included in this thesis:

1. **Computation of the acetone ultraviolet spectrum in gas phase and in aqueous solution by a mixed discrete/continuum model**
   F. Aquilante, M. Cossi, O. Crescenzi, G. Scalmani and V. Barone, *Molecular Physics* **101** (2003) 1945–1953

2. **The allyl radical revisited: a theoretical study of the electronic spectrum**
   F. Aquilante, K. P. Jensen and B. O. Roos, *Chemical Physics Letters* **380** (2003) 689–698

3. **A theoretical investigation of the valence and lowest Rydberg electronic states of acrolein**
   F. Aquilante, V. Barone and B. O. Roos, *Journal of Chemical Physics* **119** (2003) 12323–12334

4. **p-Benzoquinone in aqueous solution: Stark shifts in spectra, asymmetry in solvent structure**
   A. Öhrn and F. Aquilante, *Physical Chemistry Chemical Physics* **9** (2007) 470–480

5. **Magnetic coupling in binuclear μ-methoxo bridged Mn(III) complex: remarks on mappings and broken symmetry solutions**
   I. de P. R. Moreira, F. Aquilante, T. B. Pedersen, R. Lindh and F. Illas, *manuscript*

6. **Auxiliary basis sets for unbiased density fitting approximations with tunable accuracy**
   F. Aquilante, T. B. Pedersen, E. Bednarz, L. Gagliardi and R. Lindh, *manuscript*

Abstracts of the Papers

The present thesis is accompanied by seven manuscripts which collect the most relevant results of the research work described here. Some of them have already been published or accepted for publication in international scientific journals, while the remaining ones are still under peer review or subject to minor improvements. References to these manuscripts are present in the thesis whenever appropriate. The notation [**Px**] indicates that material relevant to the ongoing discussion can be found in "Paper x". In what follows, I will briefly present their content in such a way that the reader can understand also the motivations of the work which lead to these manuscripts. It will then be more clear also the logical and temporal connection between them.

Paper I introduces the Local Exchange (LK) screening technique that I have proposed and implemented in order to solve the "Exchange problem" for Cholesky-based SCF methods. The problem tackled here is the poor performance of standard Cholesky and density fitting methods in computing the exchange Fock matrix. The solution that we found to improve on this situation went through the use of rigorous upper bounds to the contribution from each occupied orbital to the exchange Fock matrix. By formulating these inequalities in terms of localized orbitals, the scaling of computing the exchange Fock matrix is reduced from quartic to quadratic with only negligible prescreening overhead and strict error control. Compared to the unscreened Cholesky SCF algorithm, the computational saving is substantial for systems of medium and large size.

Paper II. During the testing of the LK implementation, I came across an unexpected outcome: whenever using the Cholesky vectors of the density matrix decomposition as molecular orbital (MO) coefficients in input to the LK-Cholesky SCF, I observed enormous speed-ups compared to the performance of the same algorithm using canonical orbitals. This fact seemed to indicate that sparsity was introduced in the MO coefficient matrix by the Cholesky decomposition procedure. From this observation, we went on and demostrated that Cholesky decomposing the atomic orbital (AO) one-particle density matrix

is a way to generate a new set of MO coefficients that can be used to define localized orthonormal orbitals. This new localization technique has the advantage compared to the classical ones of being noniterative and extremely fast, and can be successfully applied also in localizing the virtual orbitals. In addition, since this approach does not require the use of starting orbitals, it can be very useful in local correlation treatments on top of diagonalization-free Hartree-Fock optimization algorithms.

Paper III deals with the problem of how to achieve an analytical formulation of the derivatives of the Cholesky vectors. This was a long-standing problem of the Cholesky approximation and probably the cause of the very little popularity of the method, despite its potentialities. We have shown that with a suitable choice of auxiliary basis set, the Cholesky decomposition is equivalent to a density fitting (DF) approximation of the two-electron integrals in the Coulomb metric. Since in DF the evaluation of analytical derivatives is straightforward, this complete unification of the two methods lead to a simple and general formulation of the analytical derivatives within the Cholesky approximation.

Paper IV. The analytical Cholesky derivative formulation was based on the proof that the Cholesky approximation can be formulated in terms of a particular type of density fitting variational problem. The functional which defines such problem is therefore a descriptor of the accuracy of the integral approximation. For the standard Cholesky approximation, the value of this functional is bounded by the decomposition threshold. The question then was to investigate possible variants of the standard Cholesky approximation based on the analysis of this functional and which could present computational advantages. We proposed the use of the Cholesky decomposition of the atomic two-electron integral matrix as a robust and general technique to derive fitting-free auxiliary basis for any given atomic basis set. This atomic CD (aCD) auxiliary basis set is method-independent, since it deals directly with the two-electron integrals. Moreover, the accuracy of the approximation can be systematically improved by lowering the decomposition threshold.

Paper V demonstrates the applicability of the Cholesky approximation to the CASSCF method. In light of the previous developments, we now used the terminology of *ab initio density fitting* to indicate both standard Cholesky decomposition (CD) and aCD approximation to the electron repulsion integrals. The Cholesky CASSCF algorithm implemented signficantly reduces the storage demands and computational efforts related to the basis set quality and therefore moves the limit of applicability of the CASSCF method to very large basis sets. Although the size of the active space remains the bottleneck in some cases, this algorithm paves the way for a wide range of large-scale CASSCF applications

perfomed at very reduced costs and with controllable accuracy.

Paper VI. For spectroscopic studies, the CASPT2 method is of paramount relevance. However, due to the necessity to have large atomic basis sets in order to obtain conclusive results, its applicability is still limited. We have shown that by avoiding the evaluation and MO transformation of the two-electron integrals through the use of the Cholesky approximation, the limits of applicability of CASPT2 can be moved to much larger basis set. The new method, CD-CASPT2, has been utilyzed to investigate the electronic structure and low-lying electronic states of a transition metal complex. It is shown that this approximation reproduces results obtained with the full integral set to a high accuracy, thus opening the possibility to use this approach to perform multiconfigurational wave function based quantum chemistry on much larger systems than has been possible up till now.

Paper VII is the result of an ongoing development of a new algorithm for Cholesky and density fitting canonical MP2. The fact that this new algorithm can reduce the scaling of the method from fifth to fourth order is the driven force in this case. In this formulation, the energy contribution is not computed from the MO integrals but through the use of the Cholesky vectors obtained by decomposing the MP2 amplitude matrix. The paper, however, presents this approach only for the Scaled Opposite-Spin variant of MP2. Our purpose was to show that significative computational savings are possible with this algortihm compared to the existing fourth-order algorithms that employ Laplace transforms.

1 Quantum chemistry in a nutshell

Everything is vague to a degree that you do not realize
till you have tried to make it precise.

Bertrand Russell

Quantum mechanics is one of the two cornerstone theories of contemporary physics, the other being Einstein's general relativity. Any other theory of physics conforms on the mathematical framework provided by either one or the other of these two. Simply speaking, all laws of Nature known to us are enclosed in the principles of these fundamental theories. What is striking about quantum mechanics, however, is not only its great success in explaining the behavior of the microscopic world, but also, and even more, the ring of mystery and exotism that seems to surround it. Even the language used to describe the theory is made of words, such as *second quantization, annihilation and creation operators, uncertainity principle, antisymmetry* and more, which appears almost as if they were describing some of the Eastern philosophies, rather than physics. Moreover, despite the unquestionable validity of the mathematical results of quantum mechanics, the physical interpretation of these results is very controversial. Disparate, but all valid interpretations are indeed possible — perhaps the most astonishing being the one proposed [1] by Hugh Everett in 1957 and widely used in the field of quantum computing, which invokes the existence of parallel universes enclosed in the *multiverse*.

When it comes to applications of quantum theory to chemistry, however, the point of view is much more pragmatic and the almost univocally used physical interpretation of quantum mechanics is the one attributable to Bohr and the other great scientists of the *Copenhagen school* during the early times

of the theory (years 1900-1930). Central for this interpretation is the concept of wave function Ψ, which contains all the information for the quantum state of the system. Such function, defined in the complex Hilbert space, has the following physical interpretation: $|\Psi|^2$ defines a probability density for the non-classical particles under consideration, here the electrons. Opposite to the concept of trajectory found in classical mechanics, we discover Heisenberg's uncertainity principle — "The more precise the measurement of position, the more imprecise the measurement of momentum, and vice versa." The motion of the electrons is nonetheless brought back to an equation, the Schrödinger equation, analogous to the classical wave equations

$$i\frac{\partial\Psi}{\partial t} = \hat{\mathcal{H}}\Psi \,, \tag{1.1}$$

assumed as a *postulate* of the theory. Few other postulates unravel the meaning of *measurements* on the quantum state and relate these to mathematical operators acting on Ψ, such as the *hamiltonian operator* $\hat{\mathcal{H}}$, in perfect specularity with Hamilton's formulation of classical mechanics. Finally, the fermionic nature of the electrons imposes a further restriction for Ψ being *antisymmetric with respect to interchange of any two particles* (Pauli principle).

Those aforementioned are the main contour lines to the approach chosen in quantum chemistry in order to investigate and interpret the laws underlying any chemical phenomenon. In many cases, the point of interest is not to describe the time evolution of the wave function, as given by Eq. (1.1), but rather the so-called *stationary states* of the systems. These are characterized by a definite energy E, a constant of motion as result of time-symmetry, given as eigenvalue of $\hat{\mathcal{H}}$

$$\hat{\mathcal{H}}\Psi = E\Psi \,. \tag{1.2}$$

Even this simplification is however not sufficient to reduce the complexity of finding the exact solution to a many-body problem of the type of Eq. (1.2). Any practical application of the theory must therefore aim at the pursuit of approximate solutions. For this purpose, a fundamental mathematical tool is represented by the *variational principle*: for a given *trial function* Φ that satisfies the same boundary conditions of the exact wave function for the system, we have

$$\mathcal{E}[\Phi] = \frac{<\Phi|\hat{\mathcal{H}}|\Phi>}{<\Phi|\Phi>} \geq E_0 \,, \tag{1.3}$$

where E_0 is the exact energy of the ground state of the system. The equality holds only in case the trial function Φ spans the same functional space of the exact wave function Ψ.

In the following pages, I will present a brief description of the key quantum chemical models used in electronic structure calculations. Most of the focus will be on the basic assumptions of each of these approaches and on their limits of

applicability to real systems. The whole discussion will be put in the context of the quest for electron correlation, the Holy Grail of quantum chemistry.

1.1 Hartree-Fock theory

The natural way to tackle a many-particle problem in quantum mechanics, such as the solution of the Schrödinger equation for an N-electron system, is to introduce descriptors for the state of the single particle. The resulting wave function for the whole system is then expressed in terms of such states and the first question one asks is how to obtain the set of single-particle states that is optimal according to some (variational) criteria.

For the electronic wave function, the single-particle states have the more familiar name of *spin orbitals*. Because the electron are fermions, the antisymmetry principle imposes a restriction to the possible ways in which the spin orbitals can be combined together in the N-electron wave function Ψ. A convenient functional form that fulfills the antisymmetry restriction is that of a particular type of determinat of the N spin orbitals $\{\varphi_p | p = 1, N\}$, written as

$$\Psi = \begin{vmatrix} \varphi_1(1) & \varphi_2(1) & \dots & \varphi_N(1) \\ \varphi_1(2) & \varphi_2(2) & \dots & \varphi_N(2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(N) & \varphi_2(N) & \dots & \varphi_N(N) \end{vmatrix} , \tag{1.4}$$

where the rows are labeled by electrons and the columns by spin orbitals. This is called a Slater determinant.

We can now invoke the variational principle and seek for the set of equations that the spin orbitals must satisfy to give rise to the the Slater determinant Ψ of minimal energy. To simplify the formalism, the orthonormality of the spin orbital is imposed as an external constraint :

$$\int \varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r})d\tau = < \varphi_i | \varphi_j > = \delta_{ij} . \tag{1.5}$$

As the Slater determinants are invariant to linear transformation among the spin orbitals, such orthonormality constraint is not relevant for the generality of the results. The wave function Ψ can then be normalized and written as

$$\Psi = \frac{1}{\sqrt{N!}} |\varphi_1 \varphi_2 \dots \varphi_N| , \tag{1.6}$$

where we used the compact notation $|\varphi_1 \varphi_2 \dots \varphi_N|$ to indicate the corresponding Slater determinant.

The electronic hamiltonian $\hat{\mathcal{H}}$, describing the motion of the electrons in the presence of a fixed-nuclei field, can be expressed as sum of one- and two-electron terms as

$$\hat{\mathcal{H}} = \sum_i \hat{h}_i + \sum_{i<j} \hat{g}_{ij} , \tag{1.7}$$

where

$$\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_\mu \frac{Z_\mu}{r_{i\mu}} \tag{1.8}$$

collects the kinetic energy and the electron-nuclei attraction potential, respectively, and

$$\hat{g}_{ij} = \frac{1}{r_{ij}} \tag{1.9}$$

describes the respulsion potential between electrons. By means of the Condon-Slater rules [2] for the expectation values of one- and two-electron operators between Slater determinants of the type of Eq. (1.6), we can write the corresponding energy expression, expectation value of $\hat{\mathcal{H}}$, as

$$\begin{aligned} E[\Psi] &= <\Psi|\hat{\mathcal{H}}|\Psi> \\ &= \sum_i <\varphi_i|\hat{h}|\varphi_i> + \sum_{i<j}[<\varphi_i\varphi_j|\hat{g}|\varphi_i\varphi_j> - <\varphi_i\varphi_j|\hat{g}|\varphi_j\varphi_i>] \,. \end{aligned} \tag{1.10}$$

For an infinitesimal variation of the spin orbital $\varphi_i \Rightarrow \varphi_i + \delta\varphi_i$, the corresponding variation of the wave function, $\Psi \Rightarrow \Psi + \delta\Psi$, determines the following first-order variation of the energy[1] :

$$\delta^{(1)}E[\Psi] = \sum_i <\delta\varphi_i|\hat{h}|\varphi_i> + \sum_{i<j}[<\delta\varphi_i\varphi_j|\hat{g}|\varphi_i\varphi_j> - <\delta\varphi_i\varphi_j|\hat{g}|\varphi_j\varphi_i>] \,. \tag{1.11}$$

The condition for the stationarity of $E[\Psi]$ with respect to any variation in Ψ is simply given by

$$\delta^{(1)}E[\Psi] = 0 \,, \tag{1.12}$$

and normally such stationary point will be more precisely a minimum. We now introduce the following one-electron operators, the *Coulomb*

$$\hat{\mathcal{J}}_j f(1) = \left[\int \frac{\varphi_j^*(2)\varphi_j(2)}{r_{12}} d\tau_2\right] f(1) \,, \tag{1.13}$$

and the *exchange* operator

$$\hat{\mathcal{K}}_j f(1) = \left[\int \frac{\varphi_j^*(2) f(2)}{r_{12}} d\tau_2\right] \varphi_j(1) \,, \tag{1.14}$$

[1] More rigorously, the complex conjugate of the right-hand side of Eq. (1.11) also contributes to $\delta^{(1)}E[\Psi]$, but it can be omitted from the present discussion without loss of generality.

defined by their effect when operating on a generic function. Eq. (1.12) can then have a compact expression explicitly in terms of orbital variations

$$< \delta\varphi_i|[\hat{h} + \sum_{j\neq i}\hat{\mathcal{J}}_j - \sum_{j\neq i}\hat{\mathcal{K}}_j]|\varphi_i >= 0 \; . \tag{1.15}$$

We notice that the operator in square brackets is different for every spin orbital φ_i on which it operates, due to the restriction on the summation index. However, this difficulty can be overcome easily by inspecting Eq. (1.56-1.14) and realizing that

$$[\hat{\mathcal{J}}_i - \hat{\mathcal{K}}_i]|\varphi_i >= 0 \; . \tag{1.16}$$

It is thus possible to use this result to remove the restriction on the summation index in Eq. (1.15) and define the following one-electron *Fock operator*

$$\hat{\mathcal{F}} = \hat{h} + \sum_j \hat{\mathcal{J}}_j - \sum_j \hat{\mathcal{K}}_j = \hat{h} + \hat{\mathcal{J}} - \hat{\mathcal{K}} \; . \tag{1.17}$$

With this definition, the stationary condition for the energy are finally cast into a form known as the *Hartree-Fock* (HF) equations:

$$< \delta\varphi_i|\hat{\mathcal{F}}|\varphi_i >= 0 \; . \tag{1.18}$$

The canonical form of the HF equations is obtained by considering a general variation of the spin orbitals :

$$\delta\varphi_i = \sum_j \alpha_{ij}\varphi_j + \sum_a \alpha_{ia}\varphi_a \; , \tag{1.19}$$

where φ_j and φ_a represent occupied and virtual orbitals, respectively.[2] As the first term in Eq. (1.19) is merely a rotation among occupied orbitals and therefore leaves the Ψ unchanged, the only nonredundant equations are given by

$$< \varphi_a|\hat{\mathcal{F}}|\varphi_i >= 0 \; , \tag{1.20}$$

a result also known as *Brillouin theorem.* It is easy to be convinced of the fact that Eq. (1.20) is equivalent to require that $\hat{\mathcal{F}}\varphi_i = \sum_k \lambda_{ki}\varphi_k$, where the index i runs over occupied orbitals only. Moreover, since $\hat{\mathcal{F}}$ is hermitian, the matrix λ can be reduced to a diagonal form, $\lambda_{kj} = \delta_{kj}\epsilon_k$, and the Brillouin conditions, Eq. (1.20), used to write the *canonical* HF equations :

$$\hat{\mathcal{F}}\tilde{\varphi}_j = \epsilon_j\tilde{\varphi}_j \; . \tag{1.21}$$

The particular spin orbitals $\tilde{\varphi}$ that satisfy Eq. (1.21) are called *canonical* and the set of eigenvalues ϵ are the orbital energies. The latter carry the important

[2] More precisely, the adjective "virtual" should be intended as to indicate that the orbital belongs to the complementary space of the occupied orbitals.

physical meaning of approximations to ionization potentials (occupied orbital energies) or electron affinity (virtual orbital energies), as expressed by the *Koopmans' theorem*. Finally, the energy of the HF state Ψ can be computed as expectation value of the hamiltonian operator and gives

$$E_{HF} = \sum_i \epsilon_i - \frac{1}{2} \sum_i < \varphi_i | \hat{\jmath} - \hat{\mathcal{K}} | \varphi_i > , \tag{1.22}$$

which is clearly different from just the sum of the occupied orbital energies.

The canonical HF equations represent a rather complicate set of integro-differential equations. As they stand in Eq. (1.21), one can only attempt a numerical solution, but even so, the application to standard chemical problems is not possible. Instead, if one uses a set of known functions as basis for the representation of the orbitals, then the HF equations become much simpler to solve, as they are reduced to a set of algebraic equations. This approach, historically called *linear combination of atomic orbitals*, or LCAO, is by large the most used in chemistry. Restricting the discussion to real orbital functions, and neglecting the spin part for the time being, we can express the orbitals in the LCAO formalism as

$$\phi_p(\mathbf{r}) = \sum_\mu C_{\mu p} \chi_\mu(\mathbf{r}) , \tag{1.23}$$

where the real functions, $\chi(r)$, identify the *atomic orbital* (AO) basis set. The expression for the energy of a Slater determinant, Eq. (1.10), is written as follows:

$$E[\Psi] = \sum_{\mu\nu} D_{\mu\nu}\, h_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} D_{\mu\nu} D_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu) \right] , \tag{1.24}$$

in terms of the *one-particle density matrix*

$$D_{\mu\nu} = \sum_i C_{\mu i} C_{\nu i} , \tag{1.25}$$

and the one- and two-electron integrals in AO basis:

$$h_{\mu\nu} = \int \chi_\mu(\mathbf{r}_1)\, \hat{h}\, \chi_\nu(\mathbf{r}_1) d r_1 , \tag{1.26}$$

$$(\mu\nu|\lambda\sigma) = \int \chi_\mu(\mathbf{r}_1)\, \chi_\nu(\mathbf{r}_1)\, \frac{1}{r_{12}}\, \chi_\lambda(\mathbf{r}_2) \chi_\sigma(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2 . \tag{1.27}$$

Minimizing the energy, Eq. (1.24), with respect to the orbital coefficients C_{qi}, we obtain the matrix formulation of the HF equation, thus

$$\mathbf{FC} = \mathbf{SC}\epsilon , \tag{1.28}$$

where $\mathbf{F}$ is called *Fock matrix*. This form of the HF equations is also known as the *Roothaan-Hall equations*. The matrix $\mathbf{S}$ that enters these equations is the

overlap matrix of the atomic orbital basis set. As the Fock operator, Eq. (1.17), depends on the orbitals through the Coulomb and exchange operators, the Fock matrix has also such dependence, this time through the density matrix. This implies that Eq. (1.29) must be solved through an iterative procedure, that generates a new set of orbitals at each step, until this updates produce negligible variations in the computed density matrix. This is why the HF or Roothaano-Hall equations are called *self-consistent-field* (SCF) equations. Finally, it should be pointed out that the value of the exact HF energy is obtained only if the SCF procedure uses a complete atomic orbital basis. Normally, this is not the case and therefore the HF-SCF energy is only an upper bound to the HF limit.

1.2 The electron correlation problem

The Hartree-Fock theory is of central significance in quantum chemistry, but only as the basic approximation on top of which one can build more accurate wave functions for a many-electron system. In fact, although the HF wave function is capable of recovering up to 99.5% or more of the exact total energy of atoms and molecules in typical situations, it is unfortunately the missing part that often is crucial for the correct prediction of the energetics in chemical processes.

The inaccuracy of the HF approximations has its origin in the fact that this is *mean-field* theory. Each electron is treated as nearly independent of the rest; it only "sees" the presence of the other electrons through a mean-field potential v_{HF} resulting from the classical coulombic repulsion and the non-classical exchange counterpart

$$\hat{v}_{HF}(1) = \sum_i [\hat{\mathcal{J}}_i(1) - \hat{\mathcal{K}}_i(1)] \,, \tag{1.29}$$

collected in the one-electron effective hamiltonian represented by the Fock operator. The interaction between each pair of electrons is therefore replaced by a sum of average interactions of single electrons in the field of the remaining electron cloud. As a consequence of the presence of the exchange potential, such averaged interaction is not just the classical repulsion between charged particles but includes the information about the non-classical, fermionic nature of the electrons. Following Bohr's interpretation of the electronic wave function, we can say that in the HF approximation the probabilty density of finding an electron in a given region of space is independent from the position of the other electrons. Using a more direct terminology, we can qualify the HF approximation by the absence of the *correlation* between the motion of the electrons. In real situations, such correlation is present and determines the existence of a region of space (Coulomb hole) that follows the motion of each electron and which is inaccessible to the other electrons. Such type of repulsion

between the electrons is larger than the one triggered by the mean field, and more importantly, depends on the instantaneous position of each electron.

To improve upon the Hartree-Fock model, we must take into account the instantaneous interactions among the electrons by allowing deformations to their HF "trajectories". In the orbital picture, this is achieved by exciting the electrons from occupied to virtual spin orbitals. The resulting wave function is then expressed as a linear combination of the HF wave function plus all possible (or some of) the excited Slater determinants:

$$\Phi = \sum_{ai} c_i^a \, \Psi_i^a + \sum_{abij} c_{ij}^{ab} \, \Psi_{ij}^{ab} + \sum_{abcijk} c_{ijk}^{abc} \, \Psi_{ijk}^{abc} + \dots \, , \tag{1.30}$$

where the indices $i, j, k, \dots$ and $a, b, c, \dots$ indicate occupied and virtual orbitals, respectively, and $\Psi_{ijk\dots}^{abc\cdots}$ is a shorthand notation for a Slater determinant in which the orbitals $\varphi_i, \varphi_j, \varphi_k, \dots$ have been replaced by $\varphi_a, \varphi_b, \varphi_c, \dots$. If the expansion coefficients $\mathbf{c}$ are obtained by solving the secular equations derived from the variational condition

$$E = min \frac{< \Phi | \hat{\mathcal{H}} | \Phi >}{< \Phi | \Phi >} \, , \tag{1.31}$$

then the approach takes the name of *configuration interaction* (CI) method. However, as we shall discuss later on, this variational approach is not the only way to obtain such coefficients. Moreover, the parametrization of Φ given in Eq. (1.30) can also be replaced by others. For instance, the following *exponential ansatz*

$$\Phi \; = \; exp(\hat{T}) \, \Psi \; = (1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots) \, \Psi \tag{1.32}$$

defines the quantum chemical models know as *Coupled-Cluster* (CC) hierarchy of methods. The cluster operator $\hat{T}$ is a sum of excitation operator classes comprising all single excitations ($\hat{T}_1$), double excitations ($\hat{T}_2$), etc., and they are written in second-quantization form as

$$\hat{T}_1 = \sum_{ai} t_i^a \, \hat{a}_a^\dagger \hat{a}_i \, , \qquad \hat{T}_2 = \sum_{a>b, i>j} t_{ij}^{ab} \, \hat{a}_a^\dagger \hat{a}_i \hat{a}_b^\dagger \hat{a}_j \, , \quad \dots \tag{1.33}$$

where the unknown coefficients $\mathbf{t}$ are called *cluster amplitudes*. Among others, the advantages of a parametrization of the type of Eq. (1.32) is that it guarantees the property of *size-extensivity* of the method — the calculation on a compound system AB consisting of two non-interacting systems A and B yields a total energy equal to the sum of the energies obtained in separate calculations on the two subsystems. This is true even if the cluster operator is truncated to and excitation level $m < N$, where N is the number of electrons in the system. For the CI method, this fundamental feature is missing, except in the case we have included in the expansion of Eq. (1.30) *all* possible excited determinants of the

N-electron system within the given orbital basis set. This approach takes the name of *Full CI* and leads to the evaluation of the exact energy of the system for that basis set. However, the sheer number of excited determinants with the number of electrons and the basis set size renders such approach unfeasible for nearly any practical application. Nonetheless, the interest in the *Full CI* method comes from the possibility to use it as benchmark for all other quantum chemical models.

The post-HF approaches briefly discussed up to this point rely on a purely orbital-based expansion of the wave function. One known problem arising from such a simplistic ansatz is the slow convergence of the computed energy towards the exact result. The main reason of this unpleasant behavior is the fact that the orbital picture is essentially based on functions, the orbitals, that only depend on the coordinates of a single electron. On the other hand, the electronic hamiltonian contains the term r_{ij}^{-1}, which depends instead on the coordinates of electron pairs. It can be imagined that the presence of this interelectronic term plays a central role in shaping the wave function in proximity of the coalescence point (r_{12}=0), and therefore that it strongly affects the energy. From the Schrödinger equation for the stationary states, Eq. (1.2), it can be seen that the function $\frac{\hat{\mathcal{H}}\Psi}{\Psi}$ represent a constant of motion, namely the energy of the system in a given state, and therefore cannot show any singularity. This implies that the hamiltonian $\hat{\mathcal{H}}$, besides the explicit singularity due to the interelectronic term, must hide another such singularity that can counterbalance the first one in the vicinity of r_{12}=0 . The latter resides evidently in the kinetic energy term. By imposing the balance between these two singularities, we obtain the following *cusp condition*

$$\left.\frac{\partial\Psi}{\partial r_{12}}\right|_{r_{12}=0} = \frac{1}{2}\Psi(r_{12}=0)\,. \tag{1.34}$$

This condition implies that in the neighborhood of the coalescence point, the exact wave function must be a linear function of the interelectronic distance r_{12}, as depicted in Figure 1.1. At the same time, in Figure 1.1, we observe that such property is not enjoyed by the HF wave function, as for the latter $\partial\Psi/\partial r_{12}$=0 at r_{12}=0. The cusp condition is a mathematical representation of the Coulomb hole introduced earlier, and its absence in the HF wave function implies that the probability to find two electrons close to each other is larger than it should, and so is the computed repulsion energy. In accordance with the variational principle, this translates to the fact that the HF energy is higher than the exact energy. The difference

$$E_{corr} = E_{HF} - E_{exact}\,, \tag{1.35}$$

between the HF limit and the (non relativistic) exact energy of the system, is by definition the total *correlation energy*. As the Pauli's exclusion principle imposes constraints on the relative position of electrons with the same spin (*Fermi hole*),

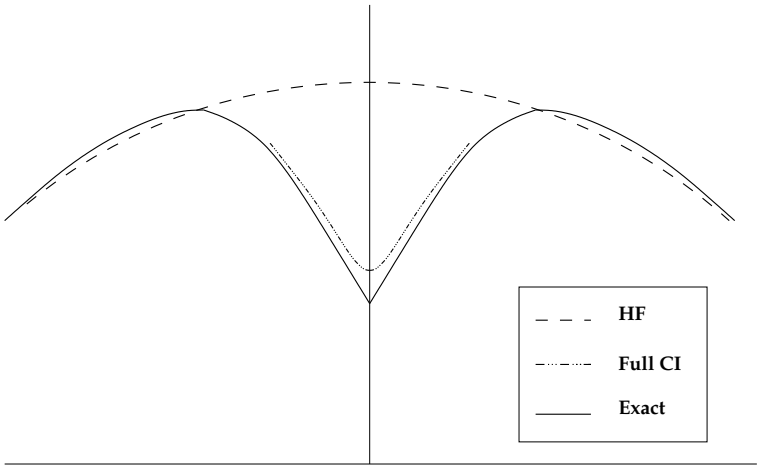


Figure 1.1: The interelectronic cusp with the characteristic Coulomb hole.

the amount of correlation energy is larger between the opposite-spin electron pairs. Moreover, in the total account of the correlation energy, the portion arising from the interaction between electron pairs is the largest. The type of short-range correlation arising from the cusp condition on the wave function is called *dynamical correlation*. In lieu of such simple terminology, often one finds in the literature a plethora of names specific for individual contribution to the total dynamical correlation. The term *radial correlation* is used to indicate that if electron 1 is far from an atomic nucleus, then electron 2 tends to be located closer to that nucleus than what predicted by the HF theory. *Left-right correlation* and *angular correlation* are instead synonyms for describing the fact that if electron 1 lies to the right of a nucleus, electron 2 will have a probability to be found on the left of the same nucleus larger than what HF theory predicts. One reason for using such more detailed terminology is to distinguish between those contributions that can be directly and completely accounted for with an orbital-based expansion of the wave function, and those that are more troublesome to get within this working assumption. Both radial and angular contributions are well described using orbital expansions, such as a truncated CI wave function. It is instead more difficult to recover the remaining portion of dynamical correlation, as this is related to the shape of the wave function at the interelectronic cusp. Large CI expansions are required to account with sufficient accuracy for such contribution, due to the fact that the resulting wave function has vanishing first derivative at $r_{12} = 0$. An approximately correct linear behavior near the coalescence point, see Figure 1.1, is then achieved only by combining a very large number of quadratic terms.

A more sensible way to introduce the short-ranged portion of the dynamical correlation is then to explicitly introduce the dependence on the interelectronic coordinate in the wave function. This simple idea was first put into practice by Hylleraas [3] as early as in 1929, who showed that this new ansatz quickly converged to the full CI wave function, also reproducing the measured ionization potential of Helium atom (24.59 eV) within 0.01 eV. A number of *explicitly correlated* methods have been developed since then and successfully applied to various problems in quantum chemistry [4, 5]. In these methods, the pairwise electron correlation, thus the most significant and slowly converging one, is described by replacing the naive HF representation of the wave function of an electron pair with a more complicated expression that also includes explicit functional dependence on r_{12}. In the most general case, we write the latter as

$$\varphi_{ij}(1,2) = \varphi_i(1)\varphi_j(2) - \varphi_j(1)\varphi_i(2) + \hat{Q}_{12}\, u_{ij}(1,2)\,, \tag{1.36}$$

where $\hat{Q}_{12}$ is a projector operator that ensures the orthogonality of the resulting wave function to the HF reference; the $u_{ij}(1,2)$ contains the implicit and explicit dependence on r_{12}, expressed through various convenient functional forms. We notice that by choosing instead $u_{ij} = \sum_{ab} c_{ij}^{ab} \varphi_{ab}$, the ansatz of Eq. (1.36) reduces to the CI-type expansion. The fast convergence with the atomic basis set size is guaranteed when using explicitly correlated wave function. However, the price to pay is the increased costs of the calculations due to the occurrence of three- and four-electron integrals. As their number grows steeply with the size of the systems, applications are still restricted to small molecules. Notwithstanding the mentioned problems, the question of whether to prefer CI-type expansions to explicitly correlated approaches, or vice versa, is still a matter of debate.

Before closing this brief survey on the central problem in quantum chemistry, thus the electron correlation problem, it is mandatory to mention another type of correlation that plays a key role in many situations and that is known with the name of *static correlation* or *non-dynamical correlation*. This type of correlation arises when one or more electronic configurations originating by excitations from the HF state are close in energy with the original HF configuration. Such situations, although not typical, are also not at all uncommon. Systems such as molecules with unfilled valencies in their electronic ground state (such as radicals, and diradicals) or containing atoms with low lying excited states (such as Li, Be, transition metals) exhibit near degeneracy problems. More generally, the same happens at the dissociation limit for chemical bonds, along reactions paths in many chemical and photochemical reactions, and often for excited electronic states. The occurrence of near degeneracy causes the HF approximation to break-down completely. As the previously discussed correlation methods use the HF as reference function, it is implicit that also these approaches are bound to fail in such situations.

The natural way to extend the HF model to these more critical cases is therefore to construct the mean-field electronic wave function from multiple

Slater determinants, preferably as many as are required to correctly describe the entire static correlation in the system. This approach results in a multi-configurational self-consistent field (MCSCF) wave function and provides the correct zeroth-order approximation for subsequent correlated electronic structure calculations. However, with the increased complexity of the MCSCF wave function compared to the HF wave function there is associated a very sizeable increase of computational costs for the method depending on the number of electronic configurations included. In general, heuristic criteria can be used to hand-pick the minimum number of such configurations required for a sufficient description of the static correlation. Alternatively, one can select a number of electrons and orbitals, thus an *active space*, and include in the MCSCF wave function all possible electronic configurations derived from distributing these active electrons into the active orbitals. This leads to the complete active space self-consistent field (CASSCF) wave function [6]. The sheer number of electronic configurations, even for a limited active space, is certainly an obstacle to the applicability of the CASSCF method to large molecules [**P5**]. However, in many cases, suitable active spaces can be devised even in extended systems while ensuring a computationally feasible size of the resulting configuration interaction (CI) expansion. On the other hand, once the correct active space has been selected, the CASSCF wave function offers maximum flexibility for a qualitative description of the correct electronic structure of even the most exotic types of chemical bonds [7]. Nonetheless, the following point should be stressed: the CASSCF wave function is based on a mean field approximation likewise the HF. Therefore, as the latter, it lacks the correct description of the dynamical correlation, and thus it cannot produce quantitative results.

In Paper V, a more detailed presentation of the CASSCF method is given, while in Paper VI the CASSCF wave function is used in a spectroscopic study as reference for perturbative treatment of the dynamical correlation.

1.3 Perturbation theory

Together with the application of the variational principle, there is a second general approach for the solution of a many-electron problem. Perturbation theory is widely used in physics and other disciplines to find an approximate solution to a problem which cannot be solved exactly, by starting from the exact solution of a related simpler problem. Perturbation theory requires the problem at hand to be formulated by adding a "small" term to the mathematical descriptors of the exactly solvable problem.

For the solution of the Schrödinger equation, we start by assuming the hamiltonian as decomposed in the sum of a zeroth-order $\hat{\mathcal{H}}_0$ term and a perturbation

$\hat{\mathcal{H}}_1$ modulated by a parameter $0 \leq \lambda \leq 1$, thus we write

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \lambda \hat{\mathcal{H}}_1 \ . \tag{1.37}$$

A second assumption is that the zeroth-order wave function Ψ_0 is an eigenfunction of $\hat{\mathcal{H}}_0$, with eigenvalue E_0. The energy E and the wave function Ψ are then expanded in series of λ as follows :

$$\Psi = \Psi_0 + \lambda \Psi_1 + \lambda^2 \Psi_2 + \dots \ , \tag{1.38}$$

$$E = E_0 + \lambda E_1 + \lambda^2 E_2 + \dots \ . \tag{1.39}$$

By replacing the expressions from Eq. (1.38) in the Schrödinger equation for the stationary states, Eq. (1.2), we arrive to an expansion over powers of λ which must be identically zero. Imposing this condition, we obtain the following set of equations (up to second order) :

$$\hat{\mathcal{H}}_0 |\Psi_0> = \ E_0 |\Psi_0> \ , \tag{1.40}$$

$$(\hat{\mathcal{H}}_0 - E_0)|\Psi_1> = \ (E_1 - \hat{\mathcal{H}}_1)|\Psi_0> \ , \tag{1.41}$$

$$(\hat{\mathcal{H}}_0 - E_0)|\Psi_2> = \ (E_1 - \hat{\mathcal{H}}_1)|\Psi_1> + E_2 |\Psi_2> \ . \tag{1.42}$$

Finally, assuming that the *intermediate normalization* holds, thus $<\Psi|\Psi_0>=1$ (or equivalently, $<\Psi_0|\Psi_i> = \delta_{i0}$), we arrive at the final expressions for the energy corrections to various order

$$E_0 = <\Psi_0|\hat{\mathcal{H}}_0|\Psi_0> \ , \tag{1.43}$$

$$E_1 = <\Psi_0|\hat{\mathcal{H}}_1|\Psi_0> \ , \tag{1.44}$$

$$E_2 = <\Psi_0|\hat{\mathcal{H}}_1|\Psi_1> \ . \tag{1.45}$$

We notice how the first-order energy correction only depends on the wave function of the unperturbed system, and that the second-order energy correction is computed from the knowledge of the first-order wave function Ψ_1. This results are completely general. In quantum chemical studies, the zeroth-order hamiltonian is usually defined from the effective potential of the mean-field approximation, thus it is chosen to be the corresponding Fock operator. This particular form of partitioning of the hamiltonian has the name of *Moller-Plesset* (MP) perturbation theory, which is among the simplest and least expensive approaches to introduce electron correlation on top of the HF approximation.

In Paper VI, MP perturbation theory is used on top of a CASSCF wave function to correct the energetics for the lack of dynamical correlation. Some relevant features of the electronic spectrum of a cobalt complex are then investigated relying on the accuracy of the perturbative correction. In Paper VII, a modified second-order MP approach is outlined, which produces further improvements of the resulting energy corrections to the HF values.

1.4 Density functional theory

Density functional theory (DFT) deals with the many-electron problem from a point of view that is completely different from what we have encountered so far in Post-HF methods. The basic idea is to move the descriptor of the system from the complicated wave function $\Psi(\mathbf{x_1} \ldots \mathbf{x_N})$, to a much simpler function, the *electron density* $\rho(\mathbf{r})$. How a function defined in R^3 could, at least in principle, contain the same amount of information as Ψ is certainly a surprising fact, however, this is indeed possible and can be proven in a relatively simple and elegant manner. For an N-electron system, the external potential $v(\mathbf{r})$ fixes the hamiltonian. We can thus affirm that N and $v(\mathbf{r})$ completely specify the ground state of the system, the energy being a functional of the external potential

$$\mathcal{E}[v] = < \Psi[v] | \hat{\mathcal{H}}_v | \Psi[v] > \ . \tag{1.46}$$

Instead of using $v(\mathbf{r})$ as fundamental variable for the characterization of the system, one can think to assign such a role to the electron density $\rho(\mathbf{r})$. These two quantities, $v(\mathbf{r})$ and $\rho(\mathbf{r})$, constitute indeed a pair of *conjugate variables*, as we shall briefly demonstrate. For this purpose, we split the hamiltonian into three obvious terms

$$\hat{\mathcal{H}}_v = \hat{\mathcal{T}} + \int \rho(\mathbf{r}) v(\mathbf{r}) + \hat{\mathcal{V}}_{ee} \ , \tag{1.47}$$

and with simple algrebra we get

$$\begin{aligned} \frac{\delta \mathcal{E}}{\delta v} &= < \frac{\delta \Psi}{\delta v} | \hat{\mathcal{H}}_v | \Psi > + < \Psi | \frac{\delta \hat{\mathcal{H}}_v}{\delta v} | \Psi > + < \Psi | \hat{\mathcal{H}}_v | \frac{\delta \Psi}{\delta v} > \\ &= E_v \frac{\delta}{\delta v} < \Psi | \Psi > + \rho(\mathbf{r}) < \Psi | \Psi > = \ \rho(\mathbf{r}) \ . \end{aligned} \tag{1.48}$$

The relation found between $v(\mathbf{r})$ and $\rho(\mathbf{r})$ allows the definition of a new functional, this time a functional of the electron density, $\hat{\mathcal{F}}[\rho]$, by operating a *Legendre transform* of the energy functional $\mathcal{E}[v]$ defined in Eq. (1.46),

$$\mathcal{F}[\rho] = \mathcal{E}[v] - \int \rho(\mathbf{r}) v(\mathbf{r}) = < \Psi | \hat{\mathcal{T}} + \hat{\mathcal{V}}_{ee} | \Psi > \ . \tag{1.49}$$

This new functional possesses the following property:

$$\frac{\delta \mathcal{F}}{\delta \rho} = \int \frac{\delta \mathcal{E}}{\delta v} \frac{\delta v}{\delta \rho} - \int \rho \frac{\delta v}{\delta \rho} - v = -v(\mathbf{r}) \ . \tag{1.50}$$

To summarize, we can say that any of the following two pairs of variables, $(v(\mathbf{r}), \mathcal{E}[v])$ and $(\rho(\mathbf{r}), \mathcal{F}[\rho])$, can be used for a full description of the system. In other words, this result proves the working hypothesis of DFT, thus that

the electron density completely determines the ground state of the system. [3] However, as it is impossible for a sytem of N interacting electrons to derive a closed-form expression for the energy functional of Eq. (1.46), the same can be said about $\mathcal{F}_{HK}[\rho]$. DFT is therefore an exact theory only in principle, but its application to real problems goes through the introduction of approximate expression for $\mathcal{F}_{HK}[\rho]$.

Although, already in 1927 Thomas and Fermi proposed a model that can be considered a predecessor of modern DFT, it was the formulation of the theory within the orbital picture by Kohn and Sham [10] in 1965 that paved the way to the tremendous success that DFT has in contemporary quantum chemistry. Kohn and Sham took the point of view of an ideal system of N non-interacting electrons. For such a system, the exact wave function is a Slater determinant, Eq. (1.6), and therefore the kinetic energy and the electron density are exactly known:

$$T_s[\rho] = \sum_i^N < \varphi_i | -\frac{1}{2}\nabla^2 | \varphi_i > , \tag{1.51}$$

$$\rho(\mathbf{r}) = \sum_i^N |\varphi_i|^2 . \tag{1.52}$$

Moreover, as the electrons are non-interacting, each electron moves under the action of a mean-field $v_s(\mathbf{r})$ due to the remaining electrons. An effective-hamiltonian eigenvalue equation can then be written for such system

$$[-\frac{1}{2}\nabla^2 + v_s(\mathbf{r})]\varphi_i = \epsilon_i \varphi_i , \tag{1.53}$$

and the energy be expressed as

$$E[\rho] = T_s[\rho] + \int v_s(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} . \tag{1.54}$$

Moving to a real system of interacting electrons, the energy is then expressable as:

$$\begin{aligned} E[\rho] &= \int v_s(\mathbf{r}) + \mathcal{T}[\rho] + \mathcal{V}_{ee}[\rho] \\ &= \int v_s(\mathbf{r}) + T_s[\rho] + J[\rho] + (\mathcal{T}[\rho] - T_s[\rho]) + (\mathcal{V}_{ee}[\rho] - J[\rho]) \\ &= \int v_s(\mathbf{r}) + T_s[\rho] + J[\rho] + E_{xc}[\rho] , \end{aligned} \tag{1.55}$$

where

$$J[\rho] = \frac{1}{2}\int\int \frac{1}{r_{12}}\rho(\mathbf{r_1})\rho(\mathbf{r_2})d\mathbf{r_1}d\mathbf{r_2} , \tag{1.56}$$

[3] A rigorous proof of this statement constitutes the core of the two famous Hohenberg and Kohn theorems [8] dated 1964. See also Ref. [9] for more details on the conditions under which these are stated.

which is the Coulomb repulsion contribution to $\mathcal{V}_{ee}[\rho]$, and

$$E_{xc}[\rho] = \mathcal{T}[\rho] - T_s[\rho] + \mathcal{V}_{ee}[\rho] - J[\rho], \tag{1.57}$$

which defines the *exchange-correlation energy*. As the name suggests, it is this quantity that carries the effects of electron correlation and the information about the fermionic nature of the particles. Comparing Eq. (1.53) with Eq. (1.55), we deduce that the problem of determining the electron density of a real system can be formulated as that of a fictitious system of non-interacting electrons satisfying the following eigenvalue equation:

$$[-\frac{1}{2}\nabla^2 + v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r})]\varphi_i = \epsilon_i \varphi_i \ , \tag{1.58}$$

where we have introduced the exchange-correlation potential

$$v_{xc}[\rho] = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \ . \tag{1.59}$$

These are the *Kohn-Sham* (KS) equations. They determine, through Eq. (1.52), the electron density of any system of interacting electrons, provided that the exact form of the functional $E_{xc}[\rho]$ is known. The KS equations are of the same type of the Hartree-Fock equations, but they differ from the latter in two important aspects. First, the effective KS hamiltonian includes only local potentials, while in HF theory there is a non-local term represented by the exchange operator. Second, as KS theory is in principle exact, it takes into account the correlation effects, provided that the functional $E_{xc}[\rho]$ is correctly chosen. The key issue in KS-DFT is therefore the determination of valid approximations to the exact exchange-correlation potential. For ideal systems, such as the uniform electron gas, the exact expression for $E_{xc}[\rho]$ can be derived [9] and this information is then used in the design of general-purpose approximate functionals. For the study of the electronic structure of molecules, it has been proven essential to introduce also the dependence from the gradient of the density in these approximate functionals. The usage of functionals of the type $E_{xc}[\rho, \nabla\rho]$ in DFT is known as Generalized Gradient Approximation (GGA). [4] The reason for that lies mainly in the necessity to take into account the fact that in molecules, electrons are distributed in the characteristic *shell structure*. In addition to the GGA approximation, it has become rather common practice to use the so-called *hybrid functionals* — obtained by mixing a certain amount of exact HF exchange into $E_{xc}[\rho]$ — for the study of many chemical problems. To legitimate such approach, we can use the following argument. We imagine a switch, represented by a real number, λ, that would smoothly introduce the electron-electron interactions in the KS reference system. The limits $\lambda = 0$ and $\lambda = 1$

[4] Functionals of solely the electron density constitute instead the Local Density Approximation (LDA). Their usage is rare in chemistry, but still significant in solid-state physics.

define the non-interacting and the real system, respectively. The Hohenberg-Kohn functional is then written in the form

$$\mathcal{F}_\lambda = min < \Psi_\rho^\lambda | \hat{\mathcal{T}} + \lambda \hat{\mathcal{V}}_{ee} | \Psi_\rho^\lambda > , \tag{1.60}$$

where Ψ_ρ^λ is the wave function that minimizes $< \hat{\mathcal{T}} + \lambda \hat{\mathcal{V}}_{ee} >$, and that gives the exact electron density ρ. The exchange-correlation energy is in this case

$$\begin{aligned} E_{xc} &= \mathcal{T}[\rho] - T_s[\rho] + \mathcal{V}_{ee}[\rho] - J[\rho] \\ &= \mathcal{F}_1[\rho] - \mathcal{F}_0[\rho] - J[\rho] \\ &= \int_0^1 \frac{\partial \mathcal{F}_\lambda}{\partial \lambda} d\lambda - J[\rho] \, . \end{aligned} \tag{1.61}$$

Using the Hellmann-Feynman theorem, the above equation becomes

$$E_{xc} = \int_0^1 < \Psi_\rho^\lambda | \hat{\mathcal{V}}_{ee} | \Psi_\rho^\lambda > d\lambda - J[\rho] \, . \tag{1.62}$$

At $\lambda = 0$, the value of E_{xc} simply reduces to the non-interacting exchange energy. This can be computed as in HF theory, but using the KS orbitals of Eq. (1.58). At $\lambda = 1$ we need to evaluate the expectation value of $\hat{\mathcal{V}}_{ee}$ for the fully interacting system, and we assume that the value, E_{xc}^{DFT}, computed through the DFT functional is a good approximation to it. We may then regard the following ansatz:

$$E_{xc} = (1 - a) \, E_{xc}^{DFT} + a \, E_x^{HF} \, , \tag{1.63}$$

where $0 \leq a < 1$ is a parameter to be optimized, as a physically justified form for the exchange correlation functional. This defines the so-called *Adiabatic Connection Method* (ACM) for the construction of hybrid functionals in DFT. There seems to be a general opinion that the hybrid functionals perform better than pure GGA functionals, and this may have something to do with the reduced *self-repulsion* energy. [5] As any deeper discussion of the variety of GGA and hybrid functionals available, and how they have been parametrized is beyond the scope of the present thesis, the interested reader is invited to consult the vaste literature on the subject (see for instance Refs. [2, 9], [**P6**] and references therein). It is however relevant to the topics treated in the present thesis the fact that hybrid DFT has become the predominant choice in quantum chemistry. In fact, as we shall see in Chapter 2 and 3, and also Paper I, the presence of HF exchange introduces serious computational difficulties for the type of approximate methods discussed here.

A concluding remark about DFT is to point the necessity to benchmark the accuracy of the method against the results of high-level correlated calculations [**P6**]. As we saw, due to the introduction of approximate expressions

[5] In Hartree-Fock theory, by virtue of Eq. (1.14), an electron does not interact with itself. The lack of balance between exchange and Coulomb interaction in DFT causes instead such anomaly called self-interaction.

for $E_{xc}[\rho]$, DFT cannot be considered an *ab intio* method, but more properly a *first principles* method. Compared to post-HF methods, DFT lacks the important property of *systematic improvability* of the results towards the correct answer. On the other hand, it certainly improves tremendously on the HF description, and this is achieved with no additional computational costs. This is the key for the success of KS-DFT in modern quantum chemistry: relatively high accuracy accompanied by sustainable costs. However, a caveat against its indiscriminate use must be given. In Paper VI, various functionals have been benchmarked against CASPT2 results for the excitation energies of a transition metal compound. The variability of the results with the choice of the functional as observed in this study indicates that DFT gives us a chance to circumvent the electron correlation problem, but certainly does not lead to its general solution.

2 When "power" is not our wish

A man provided with paper, pencil, and rubber,
and subject to strict discipline, is in effect a universal Turing Machine.

Alan Turing

The quantum chemical models described in Chapter 1 suffer of a major drawback: the slow convergence of the computed energy and properties with the size of the atomic orbital (AO) basis set. As mentioned before, the use of explicitly correlated approaches [4,5] alleviates such problem, but introduces an extra complexity to the integral calculation which makes these methods actually more expensive than the ordinary orbital-based methods. One of the bottlenecks common to any quantum chemical model based on LCAO expansions is therefore the need to compute and operate on, and often also store, a large number of electron repulsion integrals (ERIs). This problem has been addressed by several scientists over the years. Initial efforts to improve the situation focused on reducing the operation count in the evaluation of the integrals. Over the last thirty years a number of significant contributions has been presented: the Pople-Hehre approach [11], the Obara-Saika method [12], the VRR and HRR recurrence relations developed by the group around Pople [13], the McMurchie-Davidson scheme [14], the Rys-Gauss quadrature [15,16], and other more recent methods [17–19]. These approaches have been successful indeed. However, they have not addressed the critical point of any *ab initio* or density functional theory (DFT) method which explicitly relies on the ERIs: increasing the number of basis functions per atom, n, i.e. the basis set quality, the number of nonnegligible ERIs scales as heavily as $O(n^4)$. Moreover, the transformation of these integrals from AO to MO basis generates an even more severe scaling of the computational

costs. These transformations, in fact, scale with the *fifth power* of the size of the orbital basis set. This causes a sheer increase of the computational demands of even the simplest correlated methods. In light of this limitation, the best and ultimate quantum chemical method would be the one which totally avoids any explicit construction and use of ERIs integrals.

My intention is to show, from the next Section of the present thesis, that the use of the *Cholesky decomposition* technique provides a possible solution to *integral-bottleneck* problem, as it reduces the amount of information that need to be stored in order to represent the ERIs within an accuracy that is sufficient in order to achieve results comparable to those of conventional calculations. Moreover, by casting the fundamental equations of each method in terms of the Cholesky vectors, the operation count is drastically reduced.

We start by analyzing more in detail the relevant computational steps involved in Hartree-Fock (HF) calculations, as this is the starting point of any correlated method. The so-called Roothan-Hall step, Eq. (1.29), has a computational cost dependent on the diagonalization of the following Fock matrix in AO basis:

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu)] \,, \tag{2.1}$$

where the density matrix and the one- and two-electron integrals have been defined in Section 1.1. This diagonalization is a cubic step, thus its computational cost scales as N^3, where N is the number of AO basis functions. However, it is possible to reformulate this step in terms of a minimization of the energy functional $E = Tr(\mathbf{DF})$ and therefore replace the diagonalization with a conjugate gradient optimization procedure [20]. The advantage of such formulation is that it takes advantage of the possible sparsity of the density matrix [**P2**] and therefore is capable of achieving linear scaling of the computational costs with the size of the systems [21, 22]. In Chapter 3, I will discuss more in detail the situations in which sparsity is expected to be found in the density matrix and therefore linear scaling in quantum chemical models can be achieved, but here it is enough to point out that the cubic scaling of the Roothan-Hall step can be avoided. Moreover, it should be stressed the fact that even the classical diagonalization procedure, although scaling cubically, shows a small prefactor and therefore only becomes important for very large systems. We are thus left with the problem of the evaluation of the elements of the matrix **F**, and more precisely [1] with the evaluation of the ERIs present in Eq. (2.1). The total number of such integrals scales as N^4. However, due to the particular nature of these integrals, which is discussed in the next Section, the number of nonnegligible ERIs increases only quadratically rather than quartically with the size of the system. This fact, together with the aforementioned nearsightedness nature of

[1] One-electron integrals are never a problem in the context of quantum chemical models, as they are few, easy to evaluate and scale linearly with the size of the system.

the density matrix, has been extensively used to devise reduced-scaling algorithms [23,24] for the evaluation of the various contributions to the Fock matrix. The same algorithms can be used in DFT to evaluate the matrix representation of the Kohn-Sham (KS) operator, Eq. (1.58). If we fix the size of the system and increase, instead, the basis set quality — number, n, of basis functions on each atom — the nonnegligible ERIs scale as n^4. Here is where techniques like Cholesky decomposition [25, 26] and *density fitting* [27, 28] come into play. A remark is useful at this point. Cholesky and density fitting (DF) representation of the integrals are efficient techniques for the solution of the basis set quality problem. However, the reader should not be lead to think that they cannot be used for calculations on large molecules. They are indeed very helpful in reducing the computational costs of SCF and correlated calculations, especially if large AO basis sets are employed [**P6**]. The only question is that, especially for HF and DFT, the aforementioned reduced-scaling techniques, based on the conventional evaluation of the ERIs, can be very competitive in the case of calculations on large molecules and with compact basis sets [**P1**].

To the Cholesky decomposition, I will dedicate most of this Chapter, and therefore I will briefly introduce here only the DF approximation. Based on the seminal work of Whitten [27] and continued by Dunlap *et al.* [29, 30], DF [2] entered the scene as a convenient tool for approximating four-center integrals by combining two- and three-center integral in the paper by Feyereisen *et al.* [31]. The various formulations of the DF method were thoroughly discussed by Vahtras *et al.* [28] and the one based on the *Coulomb metric* was shown to be the most accurate. Nowadays, almost invariably, the notion of DF approximation of the ERIs refers to the use of the Coulomb metric, i.e., to the following inner projection

$$(\mu\nu|\lambda\sigma) = \sum_{PQ} (\mu\nu|P) G^{-1}_{PQ} (Q|\lambda\sigma) \,, \tag{2.2}$$

of the r_{12}^{-1} operator onto an atom-centered auxiliary basis set (labeled P and Q). The metric matrix collects the two-center ERI $G_{PQ} = (P|Q)$ and we have used the notation $G^{-1}_{PQ} = (G^{-1})_{PQ}$. The DF formulation as such does not prescribe the construction of the auxiliary basis set, and current implementations use auxiliary basis sets [32–35] preoptimized through data-fitting of specific energy contributions. The DF representation of the ERIs has become extremely popular in DFT calculations and in approximating ERIs for correlated calculations [36–41].

The computational benefits of the DF representation of the two-electron integrals, Eq. (2.2), are closely related to those derived by using the Cholesky representation, and will be discussed in detail in the next Section. I will show in Chapter 4 that this analogy goes beyond that, and I will prove that the two

[2] More commonly, this approximation is named "resolution of identity" (RI). For reasons that will be made clear in Chapter 4 of the present thesis, I will instead use exclusively the alternative terminology of "density fitting".

methods are indeed formally equivalent. For the time being, we will not exploit this equivalence.

2.1 Cholesky decomposition

Let us consider an atomic orbital basis of real-valued functions $\chi_\mu(\mathbf{r})$ ($k = 1, 2, \ldots, N$) and define the following two-electron integral matrix:

$$V_{IJ} = (\mu\nu|\lambda\sigma) = \int \chi_\mu(\mathbf{r}_1)\, \chi_\nu(\mathbf{r}_1)\, \frac{1}{r_{12}}\, \chi_\lambda(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2\ , \tag{2.3}$$

where the compound indices I, J identify the *orbital distributions* $\rho_{\mu\nu} = \chi_\mu\chi_\nu$. We can show that the symmetric matrix $\mathbf{V}$ is *positive definite* (PD), thus that the following inequality

$$\Omega = \sum_{IJ} X_I\, V_{IJ}\, X_J\ >\ 0 \tag{2.4}$$

holds for any non-zero vector $\mathbf{X}$ defined in the same space of the orbital distributions. In order to prove this property of $\mathbf{V}$, we can use the following Fourier transform of the r_{12}^{-1} operator:

$$\frac{1}{r_{12}} = \frac{1}{2\pi^2} \int k^{-2}\, exp[i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)]\, d\mathbf{k}\ . \tag{2.5}$$

By replacing Eq. (2.5) in Eq. (2.3), and carrying out the integration over the Cartesian coordinates, we obtain the following expression for Ω:

$$\Omega = \frac{1}{2\pi^2} \int k^{-2}\, |\omega(\mathbf{k})|^2\, d\mathbf{k}\ , \tag{2.6}$$

where we have used the following definition:

$$\omega(\mathbf{k}) = \sum_I X_I \int exp(-i\mathbf{k}\cdot\mathbf{r})\, \rho_I(\mathbf{r})\, d\mathbf{r}\ . \tag{2.7}$$

Since the integrand in Eq. (2.6) is always nonnegative, we have then proved the assertion (2.4). As a corollary, the diagonal elements, V_{II}, are positive. The two-electron integrals therefore satisfy the conditions for inner products, in a metric defined by r_{12}^{-1}. We can then apply the Schwarz inequality to obtain upper bounds to the magnitude of the off-diagonal matrix elements

$$|V_{IJ}| \leq \sqrt{V_{II} V_{JJ}}\ . \tag{2.8}$$

For the common choice of atom-centered Gaussian functions as atomic orbital basis set, Eq. (2.8) in combination with the *Gaussian product theorem* [42] explains what was already mentioned in the previous Section about the quadratic scaling with system size of the number of nonnegligible integrals.

For the purpose of the work presented in this thesis, the main consequence of the positive definiteness of the electron repulsion integral matrix $\mathbf{V}$ is the fact

that any PD matrix can be decomposed by means of the Cholesky procedure [20] into a product of a lower triangular matrix and its transpose,

$$\mathbf{V} = \mathbf{L}\mathbf{L}^{\mathbf{T}} \, . \tag{2.9}$$

The algorithm proceeds as follows. For $J = 1, 2, \ldots, M$, we perform the following diagonal update

$$\tilde{V}_{JJ} = V_{JJ} - \sum_{K=1}^{J-1} L_{JK}^2 \, , \tag{2.10}$$

and then set

$$L_{JJ} = \tilde{V}_{JJ}^{1/2} \, , \tag{2.11}$$

$$L_{IJ} = (V_{IJ} - \sum_{K=1}^{J-1} L_{IK} L_{JK})/L_{JJ} \qquad (I = J+1, \ldots, M) \, . \tag{2.12}$$

If carried through to completion, this procedure requires $\sim N^6$ floating-point operations. It is therefore not immediately clear how it can help in reducing the computational costs related to two-electron integrals in quantum chemical models. It was only by the pioneering work of Beebe and Linderberg [25] in 1977 that this was shown. They demonstrated in a number of calculations that, if one trades some of the accuracy in reproducing the matrix elements, the above procedure could be stopped at a value of M which was only a fraction of the total integral matrix dimension $N(N+1)/2$. They convincingly argued that the presence of linear dependencies among the orbital distributions $\chi_\mu \chi_\nu$ occurs in most situations, especially when more accurate atomic basis sets are employed. Moreover, the implications were that the effective rank, M, of the integral matrix would grow only linearly with the size of the basis set, as opposed to the quadratic scaling of the number of orbital distributions. They also showed that there was a one-to-one mapping between the accuracy of the computed energy and the accuracy chosen for the matrix decomposition — called also *decomposition threshold*, δ, and measured as the largest residual diagonal, Eq. (2.10), after M steps of the Cholesky algorithm.

After the work of Beebe and Linderberg, the idea of using the Cholesky decomposition (CD) of the ERI matrix did not receive much attention in the quantum chemistry community [43–45], at least not until a CD algorithm for large general basis sets and large systems was presented by Koch and coworkers [26]. (See Refs. [46–53] for an overview of the application of the CD representation of the ERIs.) It should be pointed out that, although the CD as such is among the easiest matrix operations to implement as a computer code, the situation is drastically different when trying to efficiently use it for the ERI matrix. First, due to round off errors in the integral calculation, the ERI matrix is not strictly positive definite. Its CD therefore does not enjoy

the well-known stability of the usual procedure [20]. The algorithm of Koch and co-workers makes therefore extensive use of column pivoting based on the values of the updated diagonal matrix elements. Second, the evaluation of the integral columns needed in Eq. (2.12) is necessarily performed over shell pairs, and therefore an appropriate shell qualification procedure is needed in order to minimize the chance of integral recalculations. The CD procedure starts with the calculation of the diagonal elements of the two-electron integral matrix, V_{II}. Based on the relative values of the diagonal elements, an initial preescreening is performed, which effectively reduces the range of diagonal elements, thus the largest possible value of the index J. At each step, the column identified by the largest updated diagonal is computed. The vector elements are then computed according to the recursive relation until the largest updated diagonal, Eq. (2.10), is smaller than the chosen decomposition threshold $\delta \geq 0$. From the Schwarz inequality follows that all two-electron integrals are represented with an (absolute) error of at most δ. Thus, decreasing δ, an increasingly accurate representation of the integral matrix is obtained. Further details of this algorithm are not pertinent to the scope of the present thesis and can be found in Ref. [26]. The implementation of the CD of the ERI matrix used throughout the work presented here is the one available in the MOLCAS quantum chemistry software [54] and largely based on the CD algorithm of Koch and co-workers [26].

Once the Cholesky vectors $\mathbf{L}$ are computed and stored on disk, they are ready to be used within any quantum chemical model, as they can be viewed as providing the following *Cholesky representation* of the two-electron integrals:

$$(\mu\nu|\lambda\sigma) = \sum_{J}^{M} L^{J}_{\mu\nu} L^{J}_{\lambda\sigma} \; . \tag{2.13}$$

However, it is clear that implementing a Cholesky-based SCF or correlated method by simply recomputing the integrals through Eq. (2.13) implies an increase of computational costs related to the integral evaluations from N^4 to N^4M, with a rise of the scaling from fourth to fifth order. Alternatively, one can use directly the Cholesky vectors to reformulate the equations that define the various quantum chemical models. Let us start with Hartree-Fock once again. Routinely, we define in terms of Cholesky vectors the following *Coulomb*

$$F^{c}_{\lambda\sigma} = \sum_{J} L^{J}_{\lambda\sigma} \sum_{\mu\nu} D_{\mu\nu} L^{J}_{\mu\nu} \; , \tag{2.14}$$

and *exchange*

$$F^{x}_{\lambda\sigma} = \sum_{\mu J} L^{J}_{\lambda\mu} \sum_{\nu} D_{\mu\nu} L^{J}_{\nu\sigma} \; , \tag{2.15}$$

contributions to the Fock matrix of Eq. (2.1). Both these equations are obtained by direct substitution of Eq. (2.13) in Eq. (2.1). In the above formulation, the

evaluation of the Coulomb term scales as N^2M, while the exchange contribution scales as N^3M. A straightforward way to reduce the cost of the evaluation of the exchange matrix in SCF is to rewrite the one-particle density matrix in terms of occupied molecular orbital (MO) coefficients $C_{\mu k}$, Eq. (1.25), and therefore compute the exchange contribution as

$$F^x_{\lambda\sigma} = \sum_{kJ} L^J_{\lambda k} L^J_{\sigma k}, \tag{2.16}$$

where $L^J_{\lambda k} = \sum_\mu L^J_{\lambda\mu} C_{\mu k}$ indicates the MO half-transformed vectors. The computational effort of evaluating the exchange Fock matrix from Eq. (2.16) scales as ON^2M, where O is the number of occupied orbitals. Since in usual applications $O << N$, we can immediately recognize the advantage in using Eq. (2.16) instead of Eq. (2.15) in computing the exchange term. Morever, comparing the computational costs of the Coulomb and the exchange term, we also recognize that the latter represents the bottleneck in the Cholesky-based evaluation of the Fock matrix. The Cholesky-based HF as defined through these relations is still a fourth-order method, although the better prefactor ($O << N$) reduces significantly the costs compared to conventional HF implementations. This is true unless we consider situations where the latter can benefit of a fruitful screening to bypass the evaluation of negligible Coulomb or exchange contributions to the Fock matrix [**P1**].

The implementation in MOLCAS of the Cholesky-based evaluation of the Fock and KS matrix has been the first task that I have accomplished during my PhD studies. The code has been written in FORTRAN 77 programming language, with extensive use of higly optimized Basic Linear Algebra Subprograms (BLAS) [55]. In Table 1, we observe that the use of the Cholesky representation of the ERIs introduces an error in the computed total energy that is controlled through the decomposition threshold. Throughout Paper I and Paper III-VII attached to the present thesis, the reader can find confirmation of the fact that this result is invariably shown by any application of the Cholesky integral representation to quantum chemical models. In addition to that, computed energy differences are reproduced with an accuracy that is often 1–2 orders of magnitude better than the corresponding accuracy of the total energies, which indicates that the errors introduced by the Cholesky approximation are systematic, therefore we can benefit from error cancellation. In most applications, a decomposition threshold of $\delta = 10^{-4}$ is sufficient in order to represent the ERI matrix within an accuracy suited for quantum chemical calculations. With this choice of δ, the resulting value of M is only about 3-4 times the number of atomic orbital basis functions (N). The disk space required to store the AO basis Cholesky vectors is then in the range 1-5% of the total size of the AO ERI matrix. The onset of input-output bottlenecks related to the manipulation of integral-type arrays is therefore shifted to significantly larger atomic basis sets, when using the CD approximation compared to conventional integrals.

Table 1: Linear glycines. Accuracy of the Cholesky-based Hartree-Fock at various decomposition thresholds compared to conventional calculation. Dunning's correlation consistent cc-pVDZ anc cc-pVTZ basis sets [56] have been employed.

Molecule	Absolute error in the total energy (a.u.)	
	Decomposition threshold	
	10^{-4}	10^{-6}
	cc-pVDZ	
$(Gly)_2$	$0.3 \cdot 10^{-4}$	$0.5 \cdot 10^{-6}$
$(Gly)_5$	$5 \cdot 10^{-4}$	$2 \cdot 10^{-6}$
$(Gly)_{10}$	$15 \cdot 10^{-4}$	$14 \cdot 10^{-6}$
$(Gly)_{20}$	$33 \cdot 10^{-4}$	$28 \cdot 10^{-6}$
	cc-pVTZ	
$(Gly)_2$	$0.5 \cdot 10^{-4}$	$0.4 \cdot 10^{-6}$
$(Gly)_5$	$3 \cdot 10^{-4}$	$5 \cdot 10^{-6}$
$(Gly)_{10}$	$11 \cdot 10^{-4}$	$9 \cdot 10^{-6}$

When it comes to the performance of this implementation in terms of speed-up of the calculation, the results are less exciting. As reported by Koch *et al.* [26] and as anticipated here in several occasions, the evaluation of the Fock matrix through straightforward implementation of Eq. (2.14) and Eq. (2.16) is well-suited for reducing the computational costs of calculations employing large and diffuse AO basis sets. However, for compact basis sets, integral-direct calculations can be faster in virtue of an efficient screening on the Fock matrix contributions, which avoids the evaluation of most of the integrals. An efficient implementation of a screening technique of some sort is therefore needed to reduce the costs of the exchange evaluation. I will discuss in detail this issue in Chapter 3. In Paper V, the Cholesky representation (2.13) is introduced in the CASSCF method. The formal derivation of the relevant equations can be found there, but I would like here to point out the net result of this formulation: the bottleneck of the method is moved from the MO integral transformation to the evaluation of the exchange Fock matrix. In other words, the Cholesky-based CASSCF method has a computational cost that is essentially the same as the corresponding Cholesky-based (Restricted) HF calculation, with an overall scaling reduced from fifth to fourth order[3]. In light of this result, solving the problem of the high costs of the exchange Fock matrix evaluation is therefore of paramount importance for making a step towards large-scale accurate calculations.

[3] All this is true, of course, only if the CI expansion associated with the choice of the active space is not the bottleneck of the calculation [**P5**].

As mentioned before, the DF representation of the integrals, Eq. (2.2), can be brought to a form completely identical to the Cholesky representation (2.13) by defining some "DF vectors", for example as:

$$L^P_{\mu\nu} = \sum_Q^M (\mu\nu|Q) G^{-1/2}_{PQ} \,. \tag{2.17}$$

The advantages and the bottlenecks of the CD representation are therefore directly transferable to standard DF approximations. Once again, the difficulties in exploiting such approximation for the evaluation of the exchange Fock matrix, has prevented from applying this technique to HF calculations [35]. Rather than seeking a general-purpose formulation as in the case of the CD representation, focus has then been put in the direction of developing auxiliary basis sets specific for the Coulomb contribution in DFT and for MP2 correlation energy. These auxiliary basis functions have optimized for each valence basis set by minimization of the total energy error in a set of molecular calculations [32–35]. In Chapter 4, I will show that auxiliary basis sets can indeed be devised to produce a nearly optimal fitting of the AO integrals, and therefore they show an accuracy of the resulting DF approximation unbiased towards any quantum chemical model.

For pure DFT calculations, the absence of the HF exchange term allows very fast calculations when employing CD or DF integral representation, as can be inspected by looking at the formal scaling of Eq. (2.14). Moreover, in this case the DF representation can be used without any need to construct the DF vectors. The three-center integrals of Eq. (2.2), in fact, can be contracted directly with the density matrix [32], with minimal storage demand and input-output overheads[4]. This integral-direct type of approach is obviously not possible with Cholesky, since the vectors need to be generated through the decomposition procedure, and for that, they need to be stored. Future developments in implementations of CD approximations tailored for DFT would be auspicable, but in my opinion it is unlikely that they can compete with standard DF algorithms.

Moving from SCF to the simplest of the correlated methods, canonical MP2, the use of DF or Cholesky approximation results in a more visible reduction of the computational costs compared to conventional implementations. In MP2, the need for the set of $(ai|bj)$ integrals, where i, j and a, b indicate occupied and virtual orbitals, respectively, renders the conventional calculation a fifth-order preocess with a cost of the order of ON^4. This is the amount of floating-point operations required to perform the AO to MO transformation of the integral matrix. Using Cholesky or DF, we can compute the same set of integrals as follows:

$$(ai|bj) = \sum_J^M L^J_{ai} L^J_{bj} \,, \tag{2.18}$$

[4]In addition to that, this type of DF algorithm bypasses the inversion of the metric matrix.

Table 2: MP2/cc-pVDZ calculations on linear glycines. Timings comparison between Cholesky and integral-direct implementation. AMD Opteron 2.4 GHz processor.

Molecule (N)	WALL TIME
Cholesky MP2 ($\delta = 10^{-4}$)	
$(Gly)_2$ (166)	7 sec
$(Gly)_5$ (379)	6 min
$(Gly)_{10}$ (734)	3 hr
$(Gly)_{20}$ (1444)	86 hr
	Direct MP2
$(Gly)_2$ (166)	1 min
$(Gly)_5$ (379)	30 min

where L^J_{ai} are the MO-transformed Cholesky or DF vectors. The MO transformation of these vectors is no longer a bottleneck. It requires about ON^2M operations, while the evaluation of Eq. (2.18) has a computational cost on the order of O^2V^2M, where V indicates the number of virtual orbitals. Compared to the conventional ON^4 computational requirement, once again the smaller prefactor allows substantial speed-ups, as reported for instance in Table 2. The evaluation of the integrals from Eq. (2.18) is particularly suited for canonical MP2 calculations, since in this case they can be computed on-the-fly in a batched loop, used to evaluate the energy contribution and never stored on disk. In the CASPT2 method [57, 58], there is a possibility to avoid the evaluation of the $(ai|bj)$ integrals and reformulate the method directly in terms of Cholesky vectors. However, the implementation presented in Paper VI is not optimal, since it goes through the evaluation of the $(ai|bj)$ integrals as in Eq. (2.18) and also require them to be stored on disk. Nonetheless, we want to stress that the present implementation, although of more limited applicability to large systems, allows to perform CASPT2 calculations otherwise impossible with the conventional implementation. This is achieved because it bypasses completely the AO ERIs storage bottleneck and also because it produces the needed MO integrals at reduced computational costs and input-output overheads.

Finally, in Table 3 I have reported a synoptic of how the relevant steps of the quantum chemical models treated up till now contribute to the total computational costs of the calculation. The formal and asymptotic scaling are also reported. As we shall see in Chapter 3, the fact that the exchange contribution has an asymptotically linear scaling behavior, is the key to reduce the costs of the DF and Cholesky DF methods. The formal scaling of MP2 is instead the subject

Table 3: Close-up of the computational costs and formal *vs.* asymptotic scaling behavior of the most common Cholesky-based algorithms. Comparison with other relevant steps of the calculations are reported, together with a comparison of the storage demand for the Cholesky vectors and two-electron integrals.

Computational Task	Cost	Scaling: formal → asymptotic
Diagonalization (SCF)	N^3	$\mathcal{N}^3 \rightarrow \mathcal{N}^3$
Quadrature (DFT)	N^3	$\mathcal{N}^3 \rightarrow \mathcal{N}$
$(\mu\nu\|\lambda\sigma)$	N^4	$\mathcal{N}^4 \rightarrow \mathcal{N}^2$
Cholesky-DFT	N^2M	$\mathcal{N}^3 \rightarrow \mathcal{N}^2$
Cholesky-HF	ON^2M	$\mathcal{N}^4 \rightarrow \mathcal{N}$
Cholesky-MP2	O^2V^2M	$\mathcal{N}^5 \rightarrow \mathcal{N}^3$
	Storage	
Two-el. integrals		$\mathcal{N}^4 \rightarrow \mathcal{N}^2$
Cholesky vectors		$\mathcal{N}^3 \rightarrow \mathcal{N}^2$

$\mathcal{N}$ represents any suitable measure of the system size.

of the next Section. The question is whether or not the approach adopted so far in Cholesky and DF MP2 is the best possible. I shall demonstrate that it is not, and that the scaling of the method can be reduced from fifth to fourth order, with an algorithm that is also *perfectly parallelizable*. I will only present the theoretical aspects of this new idea, since the computer implementation of the method in MOLCAS is ongoing, and not yet ready for real applications.

2.2 Fourth-order canonical MP2

The closed-shell MP2 energy in canonical orbitals is given by

$$E_2 \;=\; -\sum_{aibj} \frac{(\phi_a\phi_i|\phi_b\phi_j)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j}[2(\phi_a\phi_i|\phi_b\phi_j) - (\phi_a\phi_j|\phi_b\phi_i)] \,. \tag{2.19}$$

We now perform the Cholesky decomposition of the MP2 amplitude[5] matrix [59] :

$$t_{ai,bj} = \frac{(\phi_a\phi_i|\phi_b\phi_j)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j} = \sum_K^m R_{ai}^K R_{bj}^K \,. \tag{2.20}$$

The CD algorithm applied to this matrix requires $OVm^2/2$ operations [25,26,44] plus the evaluation of m columns of $\mathbf{t}$. The necessary two-electron integrals can be computed using Cholesky or DF representations as follows:

$$(\phi_a\phi_i|\phi_b\phi_j) = \sum_J^M L_{ai}^J L_{bj}^J \,, \tag{2.21}$$

with a total of $OVMm$ operations. It is crucial to understand how the value of m scales with system size. Although the matrix $\mathbf{t}$ has a quadratic-scaling dimension OV, its effective rank m scales only linearly with the size of the system, as a result of the way $\mathbf{t}$ is constructed [**P7**]. In fact, it is easy to realize that m is bounded by a linear scaling quantity, $n_r M$, where n_r represents the number of Cholesky vectors needed for an exact decomposition of the orbital energy denominator (OED) matrix, $\mathbf{D}_{ai,bj} = \frac{1}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j}$. It is known that such number is very small and, more importantly, *independent* of the size of the molecule [60]. Therefore the effective rank of $\mathbf{t}$ scales linearly with the size of the molecule.

We can now cast the expression for the MP2 energy (2.19) as follows:

$$\begin{aligned} E_2 &= -\sum_K \sum_{aibj} R_{ai}^K R_{bj}^K [2(\phi_a\phi_i|\phi_b\phi_j) - (\phi_a\phi_j|\phi_b\phi_i)] \\ &= -\sum_K \sum_{ij} [2(\overline{\phi}_i\phi_i|\overline{\phi}_j\phi_j) - (\overline{\phi}_i\phi_j|\overline{\phi}_j\phi_i)] \,, \end{aligned} \tag{2.22}$$

where $\overline{\phi}_i = \overline{\phi}_i(K) = \sum_a R_{ai}^K \phi_a$ is a set of nonorthogonal orbitals spanning the virtual space.[6] We notice that the expression (2.22) is *invariant under rotation of the occupied orbitals*. For instance, the occupied canonical orbitals can be localized by transformation with a proper unitary matrix $\mathbf{U}$ and the Eq. (2.22) can still be used, as long as the $\overline{\phi}_i(K)$ are backtransformed with the same $\mathbf{U}$.

The computationally most appealing way to implement Eq. (2.22) is by rewriting it in AO basis:

$$E_2 = -\sum_{\mu\nu\lambda\sigma} \Theta_{\mu\nu\lambda\sigma} [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] \,, \tag{2.23}$$

where $\Theta_{\mu\nu\lambda\sigma} = \sum_K R_{\mu\nu}^K R_{\lambda\sigma}^K$ and $R_{\mu\nu}^K$ are the backtransformed Cholesky vectors of the amplitudes. It is possible to show that the AO basis amplitudes Θ constitute

[5] More precisely, the MP2 amplitudes should have a minus sign in front, but this is irrelevant to the present discussion.

[6] Notice the analogy with local correlation methods [39,61]. Here as well, for each occupied orbital i there is a specific set of M virtual orbitals $\overline{\phi}_i(K)$, thus a "domain of excitation".

a sparse matrix [62]. Together with the sparsity of the AO two-electron integrals, this implies that for large systems an efficient screening is possible in order to reduce the costs for the evaluation of Eq. (2.22). In fact, by applying the Schwarz inequality to both Θ and the integrals, we can easily show that the exchange-type term can be computed with an effort that is asymptotically linear. However, due to the fact that the Θ needs to be computed from its Cholesky representation, most likely this step becomes quadratic scaling with a low prefactor. The same way, the Coulomb-type term will require a cubic scaling algorithm that could be probably reduced by using the multipole based integral estimates (MBIE) of Ochsenfeld and co-workers [63].

In order to efficiently implement Eq. (2.22), the exact AO two-electron integrals need to be computed in a direct fashion and not from their Cholesky or DF representation. The reason for that is the possibility to achieve efficient parallelization of the code. The Cholesky decomposition of $\mathbf{t}$ can be performed separately on each node without any communication. The amplitude matrix $\mathbf{t}$ will be different on each node and corresponds to a partial contribution to the $(\phi_a\phi_i|\phi_b\phi_j)$ integrals. On the other hand, since the evaluation of the AO two-electron integrals is at most a quadratic step, it can be performed by each node without jeopardizing the efficiency (remember, for instance, that the assembly of Θ requires a cubic step). Indeed, since the matrix Θ on each node is only partial contribution to the total matrix, the application of the Schwarz/MBIE inequalities will result in the evaluation of an even smaller (and different) set of AO integrals on each node.

The total costs of this new formulation are to be confronted to the $\sim O^2V^2M$ scaling of the usual Cholesky procedure. In Paper VII, we have shown an application of a simplified form of the present algorithm for the evaluation of the Coulomb-type term only. We have shown that by choosing $M \approx m$, the accuracy of the final energy is the same as that of the original Cholesky algorithm. Assuming $M \approx m$ in the general case, we can say that the new formulation presented here is advantageous compared to the existing ones whenever $M < OV$, a condition probably fulfilled in most cases for large molecules. More importantly, this algorithm holds the promise of being the first ever Cholesky and DF MP2 algorithm capable of quartic scaling in canonical basis. Finally, the fact that this algorithm can be parallelized with *no communication* is an extraordinary advantage over standard implementations of Cholesky and DF MP2. The presence of products of integrals in Eq. (2.19) is in fact a major obstacle for an efficient parallel implementation of any of the existing canonical MP2 formulations. Most likely, large arrays of two-electron integrals will have to be communicated in these implementations. The reduction in computing time due to the distribution of the tasks across nodes, will then be completely cancelled by the communication overheads. With massive parallelization, one can really think of moving the limit of applicability of any quantum chemical method to theoretical studies of systems in real-life scale, for example in nanotechnology

or biochemistry. Standard implementations delay such process. The one just outlined above have the potential to take fully advantage of high-performance computers to solve the "exponential scaling wall" [64] problem of quantum chemistry.

3 Localizing the electrons

The sciences do not try to explain, they hardly even try to interpret, they mainly make models.

John von Neumann

We are now entering the exciting part of this brief narration of the research work that I have undertaken during my PhD studies. We have seen in the previous Chapter that the Cholesky approximation can produce results with controlled accuracy. We saw also the potential of the method as means to speed-up correlated calculations. But at the same time, we anticipated that the performace of the method, as proposed by Beebe and Linderberg [25] and extended by Koch and co-workers [26], could not be considered satisfactory for SCF calculations. To give you a more practical example of the limits that such straightforward Cholesky implementation has, I reported in Table 4 a comparison of CPU times per SCF iteration between Cholesky and integral-direct implementation. We can easily see that with this choice of the basis set, already at about 100 atoms, the advantages of the Cholesky implementation disappear. At larger molecular sizes, the situation is even worse since the integral-direct algorithm is by far the fastest of the two. The comparison has been made with an integral-direct algorithm that uses standard screening techniques based on the Schwarz inequality. This implies that the performances of modern linear scaling integral-direct algorithms would be beyond any reach for Cholesky SCF. Table 4 shows an example of what we can call the *exchange problem* for the DF or Cholesky-based SCF methods, especially severe for electron-rich systems and large basis sets. The quartic scaling of the evaluation of the exchange Fock matrix, Eq. (2.16), quickly downgrades the capacity of the standard Cholesky

Table 4: CPU time (in seconds, AMD Opteron 2.4 GHz processor) for the exchange Fock matrix build, using the cc-pVDZ basis set. The decomposition threshold is $\delta = 10^{-4}$.

Molecule[a]	O	N	M	N_A^b	Method	
					Integral-direct	Cholesky
$(Gly)_{10}$	155	734	3556	73	1395	933
$(Gly)_{20}$	305	1444	6957	143	7059	12680
$(Gly)_{30}$	455	2154	10348	213	17019	56305

[a] α-helix geometry used for the glycines.

[b] Number of atoms. O, N, M according to the definitions given in Chapter 2.

and DF SCF implementations [26, 35] or at least renders them scarcely competitive compared to integral-direct SCF methods. Particularly unpleasant is the dependence of the computational costs ($\sim ON^2M$) on the number of occupied orbitals, as this is not the case for integral-based SCF algorithms.

At the time when I started to work on a possible solution to this problem, only a paper by Polly *et al.* [65] had explicitly discussed this issue for the DF approach. Moreover, in the literature I found no mention of attempts made to solve the problem in the framework of Cholesky-based SCF methods. As I was interested in a possible solution to the exchange problem that could work for both Cholesky and DF approximation, the approach proposed by Polly *et al.* could certainly represent a starting point. However, I soon realized that their solution, although it worked in practice, was not completely satisfactory, mainly because the procedure does not yield bounded errors — therefore the accuracy of the result cannot be controlled solely on the basis of energy thresholds. By construction, the method of Polly *et al.* is capable of achieving asymptotic linear scaling for the evaluation of the exchange Fock matrix, but this did not alleviate my concern regarding the loss of controllability of the accuracy. As discussed in the previous Chapter, it is exactly this controllability that makes the Cholesky approximation extremely appealing. With this in mind, I decided to redirect my efforts towards the setup of a screening technique that would instead closely resemble the ones used in linear scaling integral-direct algorithms. These are in fact based on exact bounds, and therefore allow a complete error controlled screening. The state-of-the-art algorithm among the ones of this sort is certainly the LinK scheme proposed by Ochsenfeld and co-workers [23, 66]. This approach is general, in the sense that it does not involve *a priori* assumptions about the size of the system or the nature of its electronic structure. Only the short-range character of the exchange interactions are required for the screening to be effective. What do we mean by short-range character of the exchange

interaction? In order to clarify this crucial point, I start by writing once more the exchange contribution [1] to the Fock matrix:

$$K_{\lambda\sigma} = \sum_{\mu\nu} P_{\mu\nu}(\lambda\mu|\nu\sigma) \; . \tag{2.24}$$

For insulators, the density matrix, **P**, is usually *sparse*, since it has been shown that the electron density $\rho(\mathbf{r},\mathbf{r}') = \sum_{\mu\nu} P_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}')$ decays exponentially with the distance $|\mathbf{r}-\mathbf{r}'|$, see e.g. Ref. [67]. The same is not true for conductors, in which the small band gap determines a delocalization of the electrons through the whole system. For the typical situations investigated, and for the domain of applicability of the techniques discussed in the present thesis, we can assume throughout that the assumption of locality of the density matrix holds. Combining this feauture with the quadratic scaling of the number of significant integrals with system size, we deduce that the exchange matrix **K** is also local. This explains the asymptotic linear scaling cost of its evaluation conjectured in Table 3. More in detail, the Schwarz inequality and Gaussian product theorem applied to the integral $(\lambda\mu|\nu\sigma)$ ensures that only the pairs of AO basis functions (λ,μ) centered on atoms sufficiently close to each other will give rise to numerically significant integrals. The same is obviously true for the pairs (ν,σ). On the other hand, the supposed locality of the density matrix implies that the only nonnegligible matrix elements are found between neighboring pairs (μ,ν). This *cross-index* relations result in the fact that the number of significant elements in the matrix **K** grows only linearly with system size. Therefore, it is possible to devise algorithm that can compute such matrix with a linear scaling computational effort, provided that an efficient screening is implemented. This nearsightdeness nature of the exchange interaction is the only requirement for the LinK scheme to achieve such scaling reduction, although the efficiency of LinK are mainly due to peculiar algorithmic choices. The details of the algorithm can be found in Ref. [23], while for the present discussion, it is only necessary to know that the spirit of the algorithm implemented for Cholesky and DF methods is in line with the ideas discussed and developed there.

[1] The choice of using new symbols, **K** and **P**, instead of maintaining the notation used ealier is mainly to follow the standards.

3.1 The LK method

As it appeared to me, for the exchange problem there was no need to aim at linear scaling. The primary goal was to reduce the costs of the evaluation of the exchange Fock matrix when using Cholesky, to a level that could make the calculations competitive with integral-direct algorithms, and possibly faster. The straightforward implementation of the LinK scheme with Cholesky integral representation would have meant the renounce to the smaller scaling prefactor gained by using the MO-based formulation, Eq. (2.16), instead of the one based on the density matrix, Eq. (2.15). I considered this option too risky: linear scaling was not the key to solve the exchange problem, since the reduction in computational costs was needed even for relatively small molecules. The onset of linear scaling is instead very system dependent, and I feared that it could have appeared perhaps outside the range of molecular sizes of interest for applications of the Cholesky method. The option was therefore to somehow include the MO formulation in a screening of the type of the LinK scheme. Likely, I came across to another way of writing Eq. (2.24), namely

$$K_{\lambda\sigma} = \sum_i (\lambda i|\sigma i) \;, \tag{2.25}$$

where I have used the definition of the density matrix in terms of MO coefficients, $P_{\mu\nu} = \sum_i^O C_{\mu i} C_{\nu i}$, in order to write the two-electron integrals $(\lambda i|\sigma i)$ in mixed AO-MO form. Not too surprisingly, this way of thinking at the exchange Fock matrix is utterly uncommon. Nobody would ever think of using such formulation in standard Hartree-Fock, as it requires fifth-order MO transformation of the integrals. Instead, this was the key that lead to the solution of the exchange problem, through the formulation of the *Local K* screening method, or simply LK. In fact, due to the invariance of the density under unitary transformations of the occupied orbitals, any such set of MOs can be used to compute the exchange Fock matrix. Among them, the use of localized orbitals can be advantageous once we consider Eq. (2.25). The integrals $(\lambda i|\sigma i)$ will show an increasing sparsity with the locality of the occupied orbitals. The possibility to screen away a substantial amount of these integrals is therefore not directly related to the size of the system; this feature is a direct consequence of the short-range nature of the exchange interaction.

In order to implement an efficient and also error bounded screening of the integrals $(\lambda i|\sigma i)$, the LK method makes use of the following chain of inequalities:

$$\begin{aligned} |K^{(i)}_{\lambda\sigma}| \;&=\; |(\lambda i|\sigma i)| \;\leq\; \sum_{\mu\nu} |C_{\mu(i)}| \;\, |(\lambda\mu|\sigma\nu)| \;\, |C_{\nu(i)}| \\ &\leq \sum_{\mu\nu} |C_{\mu(i)}| \;\, D^{1/2}_{\lambda\mu} D^{1/2}_{\sigma\nu} \;\, |C_{\nu(i)}| \;=\; Y^{(i)}_{\lambda} Y^{(i)}_{\sigma} \;, \end{aligned} \tag{2.26}$$

where $D_{\mu\nu} = (\mu\nu|\mu\nu)$ are the (exact) diagonal elements of the ERI matrix in AO

basis and the i-th vector $Y^{(i)}$ has elements $Y^{(i)}_\mu = \sum_\nu D^{1/2}_{\mu\nu}|C_{\nu(i)}|$. It is important to notice that in deriving these inequalities we have only used the fact that the ERI matrix in AO basis is positive definite and therefore satisfies the Schwarz inequality. It is straightforward to realize the importance of having a sparse set of MO coefficients, since the screening of the contribution to the exchange matrix is directly related to the number of numerically significant elements in each array $Y^{(i)}$. The sparsity of the latter is obviously related to that of the MO coefficients, as the $D_{\mu\nu}$ integrals are sparse in virtue of the Gaussian product theorem.

Employing solely Eq. (2.26), it is unlikely that an efficient screening could be achieved. A first step towards an algorithmic optimization of the LK method was to introduce the following idea. Whenever the contribution to the exchange Fock matrix from a certain number ($m < M$) of Cholesky vectors has been computed, the inequalities can be applied in the very same way to the remainder matrix. In practice this means that the $Y^{(i)}$ vectors can be computed using updated integral diagonals, namely $\tilde{D}_{\lambda\mu} = D_{\lambda\mu} - \sum_J^m (L^J_{\lambda\mu})^2$. Due to the inner projection nature of the Cholesky decomposition, these updated integral diagonals are guaranteed to remain nonnegative. Most importantly, the number of significant elements in $\tilde{D}$ will decrease, since more and more Cholesky vectors have already contributed to the exchange Fock matrix.

With this premises, an effective implementation of the LK method was possible. Its details are presented and extensively discussed in Paper I. A sketch of the LK algorithm is shown in Figure 3.1. Few comments are, however, useful at this point. First, it is natural to choose *shell pairs* as "units" on which to base the screening. A smaller unit, such as the individual matrix elements, would probably increase the screening overheads to an unacceptable level. Larger units, on the other hand, could jeopardize the efficiency of the screening, at least for the range of molecular sizes of immediate interest, $\sim 100 - 500$ atoms. Second, as shown in Figure 3.1, a step of pre-ordering of the eligible shells is present in the LK scheme. This is a very essential operation, as it allows leaving the final loop (**K**-build) as soon as the established condition is not met. Third, the need to perform an MO half-transformation of the Cholesky vectors, formally as expensive as the build of **K**, is an additional issue to consider. This has been efficiently tackled by employing estimates based on the Frobenius norm of the resulting matrices. Fourth, the LK algorithm intrinsically exploits the permutational symmetry of the ERIs, as can be seen by checking that only symmetry distinct contributions to the exchange Fock matrix are referenced in Figure 3.1. The exploitation of the permutational symmetry of the ERIs, as well as the shell pre-ordering, are clear indications that the LK algorithm is intimately related to the original LinK scheme, as these are also the key features of the latter. By the nature of the method, the accuracy of the LK screening is absolutely reliable and can be controlled by the choice of the screening thresholds. Detailed analysis of the accuracy is reported in Paper I. Here I will instead mention that

Loop over Cholesky vectors
 Loop over occupied orbitals (i)

 (1) Pre-selection and pre-ordering of the eligible shells
 Compute the $Y^{(i)}$ screening vectors from (updated) integral diagonals
 Loop over all μ-shells
  If ($MY^{(i)} \cdot max(Y_\mu^{(i)}) \geq \tau$) then
   Add the μ-shell to a list $ML^{(i)}$
  EndIf
 End Loop
 Sort the μ-shells in $ML^{(i)}$ by the value $max(Y_\mu^{(i)})$
 (2) Frobenius-norm screened MO transformation
 Loop over μ-shells in the list $ML^{(i)}$
  Loop over all ν-shells in the shell-pairs ($\mu\nu$) of the AO Cholesky vectors
   If ($||L^J_{\mu\nu}||_F \cdot ||C_\nu^{(i)}||_F \geq \tau_F$) then
    Perform the MO half-transformation $L^J_{\mu(i)} = \sum_\nu L^J_{\mu\nu} C_\nu^{(i)}$
   EndIf
  End Loop
  Remove from $ML^{(i)}$ the μ-shells that did not qualify for the MO transformation
 End Loop

 (3) Evaluation of the exchange Fock matrix $K_{\lambda\sigma}^{(i)}$
 Loop over λ-shells in $ML^{(i)}$
  Loop over the σ-shells ($\sigma \leq \lambda$) in $ML^{(i)}$
   If ($max(Y_\lambda^{(i)}) \cdot max(Y_\sigma^{(i)}) \geq \tau$) then
    Compute $K_{\lambda\sigma}^{(i)} = \sum_J L^J_{\lambda(i)} L^J_{\sigma(i)}$
   Else
    Leave σ-loop
   EndIf
  End Loop
 End Loop

 End Loop
 (4) Update integral diagonals
End Loop

Figure 3.1: Outline of the LK algorithm for reduced scaling evaluation of exchange-type Fock matrices from Cholesky integral representation.

the LK screening is indeed capable of reducing the scaling of the evaluation of the exchange Fock matrix from quartic to *quadratic*. In Figure 3.2, we can see an example of what that means in terms of performances in Cholesky SCF calculations. Currently, the LK screening is used in MOLCAS for any type of Fock matrix build based on Cholesky vectors. Particularly interesting could then be for the reader to notice also the speed-ups for CASSCF calculations, as for example the ones reported in Paper V and Paper VI.

A final remark. There is not a single notion in the LK algorithm that could distinguish between a Cholesky or a DF representation of the integrals. Consequently, all results discussed here and more extensively in Paper I, apply to both methods. In other words, the LK screening is a simple, accurate and general *solution to the exchange problem*. And as "*fortune is a woman [...] and she allows herself to be mastered by the adventurous*[2]", the solution to the exchange problem came in a purcel with an unexpected surprise, which we shall discuss next.

[2] *Machiavelli*, The Prince ("What Fortune Can Effect In Human Affairs, And How To Withstand Her")

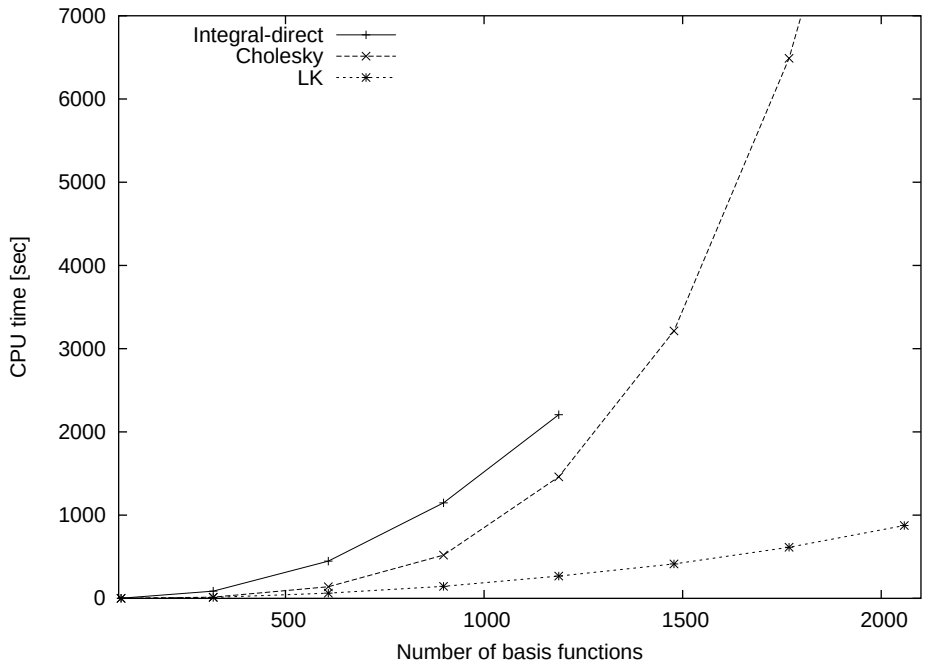


Figure 3.2: Comparison of the scaling behavior of the exchange Fock matrix build for linear alkanes using standard Cholesky, LK-Cholesky and integral-direct. Decomposition threshold $\delta = 10^{-4}$ and cc-pVTZ basis set.

3.2 A case of serendipity: Cholesky orbitals

The density matrix, as used for instance in Eq. (2.24), is obviously a positive semidefinte matrix, since it possesses a Cholesky-type representation by way of construction. Moreover, any unitary transformation of the occupied MO coefficient matrix would produce yet another of such representation. The LK algorithm can handle any of these density matrix representation. As discussed above, however, the efficiency of the screening depends on the sparsity of these MO coefficients. During the testing of the LK implementation, I came across an unexpected outcome: whenever using the Cholesky vectors of the density matrix decomposition as MO coefficients in input to the LK-Cholesky SCF, I observed enormous speed-ups compared to the performance of the same algorithm using canonical orbitals. This fact seemed to indicate that sparsity was introduced in the MO coefficient matrix by the Cholesky decomposition procedure. Many tests were performed to understand whether or not this was only accidental, but the outcome was invariably the same. From that moment we coined the term *Cholesky MOs*, to indicate the orbitals whose coefficient matrix was obtained by the decomposing the AO density matrix.

Of course, sparsity of the MO coefficients does not necessarily imply that the resulting orbitals are localized. Sparsity is in this context the property of a matrix to have relatively few, and scattered, large elements. Locality carries the notion of spatial extent of an object, which in this case indicates that the orbitals are confined to few neighboring atoms. Orbital locality implies a degree of sparsity in the MO coefficients but not *vice versa*. Canonical orbitals are extremely delocalized (cf. Section 1.1) and even in large systems they tend to span a substantial portion of the total extent of the molecule. As HF theory is invariant under rotations among the occupied and among the virtual orbitals [42], localized MOs may be obtained by a suitable unitary transformation of the canonical ones.

Localized occupied molecular orbitals are of central importance in quantum chemistry for two reasons. First, they provide the link between Hartree-Fock (HF) theory and the concept of chemical bonds formed between two atoms [68, 69] and second, they are indispensable for exploiting the short-range nature of electron correlation to achieve reductions in the computational effort [39, 61, 70–78].

Several schemes have been developed for choosing such a unitary transformation of the occupied orbitals. While the scheme usually attributed to Boys [68, 69, 79] minimizes the spatial extent of the orbitals by maximizing the distances between orbital centroids, the Edmiston-Ruedenberg [79] (ER) procedure aims at maximizing the self-repulsion energy of the orbitals, thus minimizing the exchange energy between them. The most widely used localization scheme today was introduced by Pipek and Mezey [80] in 1989. The Pipek-Mezey (PM) localization procedure seeks to minimize the number of atomic centers over which each MO extends by maximizing the sum of squares of gross

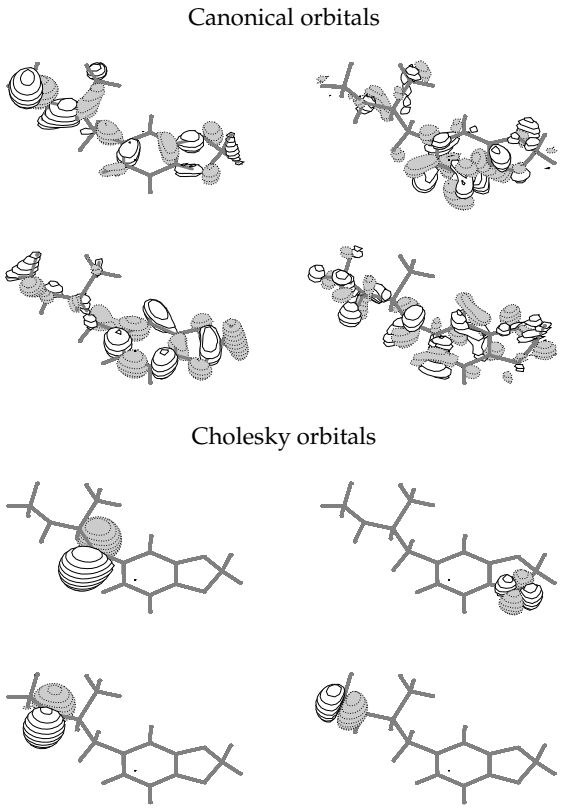


Figure 3.3: Methylenedioxymethamphetamine: MDMA (Extacy). Pictorial representation of some canonical and Cholesky occupied (left) and virtual (right) orbitals. RHF/cc-pVDZ wave function. Contour line set to 0.05 au^{-3}.

atomic Mulliken population of the MOs. In practice, the Boys, ER, and PM procedures are formulated as an optimization problem in which a localization functional, $\zeta(\phi)$, is maximized with respect to rotations among the occupied orbitals. The orbital localization thus becomes an iterative procedure [**P2**].

In light of what briefly discussed above in relation to localization techniques, it seems very unlikely that a simple Cholesky decomposition of the density matrix could produce localized orbitals. Amazingly as it may appear, reality is what Figure 3.3 shows. The acute reader may ask two preliminary questions: 1) can we be sure that we get as many Cholesky vectors as the initial occupied orbitals? 2) Are these new orbitals related to the canonical orbitals by any

unitary transformation? The answer to both questions is yes. The proof relative to the second question is lengthy and the interested reader can find it in Paper II. For the first one is enough to say that the rank of the density matrix is by definition equal to the number of occupied orbitals. A decomposition carried through to completion, will then always produce as many vectors as the number of occupied orbitals. Next question is then, how do we explain that? In paper V, the reader can find a more detailed answer to this legitimate question, in addition to a number of indirect evidences of the locality of the Cholesky MOs. Here, I will just point out that the density matrix provides a measure of the interaction between different regions of space in the molecule. The mathematical procedure that defines the Cholesky decomposition, Eq. (2.11-2.12), certainly allows the generated vectors to inherit the locality of the density matrix. This was the main reasoning that we used to support the results discussed in Paper I, although it must be stressed that the procedure actually works in the great majority of the cases even for small and medium-sized molecules.

The Cholesky localization has several computational advantages and unique features. First, Cholesky decomposition is a numerically stable and fast algorithm that can be made linear scaling [81] for matrices with linear scaling number of non-zero elements. Second, being a non-iterative procedure, complicated optimization techniques are not needed. Third, as no initial orbitals need to be given, the procedure is particularly well suited for determining local MOs directly from density matrix-based HF theory [21,81–84]. Fourth, it is the natural choice for obtaining localized orbitals to be used in connection with the LK screening. In fact, prior to the loop structure of Figure 3.1, the orbital localization must be performed. At first glance this may seem an irrelevant task but actually, depending on the type of localization employed, it can become computationally dominating. This problem was also addressed by Polly *et al.* [65] who employed Pipek-Mezey [80] localization in their algorithm for computing the exchange Fock matrix in DF. They bypassed the difficulty by performing the orbital localization only for some of the SCF iterations, thus using approximate localized orbitals in the intermediate steps. As shown in Table 5, Pipek-Mezey localization becomes prohibitively expensive for large systems due to its scaling with the number of atoms (N_A). The Cholesky localization is instead extremely fast, and therefore can be used in the LK without jeopardizing the performance of the overall SCF method.

In addition to local occupied orbitals, however, reducing the computational cost of wave function-based electron correlation models requires a set of local orbitals spanning the virtual space. It is well known [85] that the iterative localization schemes outlined above can not straightforwardly be extended to produce local virtual orbitals. Observing instead that the AO basis is local by definition, it is tempting to use these orbitals to span the virtual space. As the AO basis set spans the full space, a linearly dependent set of AOs spanning the virtual space is computed by projection onto the orthogonal complement

Table 5: CPU time (in seconds, AMD Opteron 2.4 GHz processor) for the various localization methods based on the restricted HF/cc-pVDZ optimized wave function. For Cholesky localization the time includes the calculation of the density matrix from canonical orbitals. N_A is the number of atoms.

Molecule[a]	O	N	N_A	Localization		
				Pipek-Mezey[b]	Boys[b]	Cholesky
$(Gly)_{10}$	155	734	73	78	2	0.1
$(Gly)_{20}$	305	1444	143	1432	20	1.0
$(Gly)_{30}$	455	2154	213	8895	105	3.5
$(Gly)_{40}$	605	2864	283		270	7.7

[a] Linear geometry used for the glycines.

[b] Cholesky MOs used as initial orbitals.

of the occupied space, see for instance Ref. [74]. These *Projected Atomic Orbitals* (PAOs) are used to define the so-called *orbital domains* in modern local correlation methods [39, 61, 70–78]. These methods, in fact, can achieve linear scaling of the computational costs by restricting the number of excitations from a given localized occupied orbital into the virtual space spanned by the PAOs. An orbital domain is defined by the PAOs arising from AOs centered on the atoms needed to span the localized MO with a certain accuracy according to the completeness criterium of Boughton and Pulay [70]. In short, atoms are added to the domain $[i]$ of the occupied orbital ϕ_i in the order of decreasing Mulliken charge until the function

$$f(\mathbf{a}) = \min \int (\phi_i - \phi_i')^2 \, d\tau \tag{2.27}$$

is smaller than a prescribed threshold **[P2]**. In Eq. (2.27), $\phi_i' = \sum_{\mu \in [i]} \chi_\mu a_{\mu i}$. Once the orbital domains have been defined, single excitations are obtained by exciting electrons from a local occupied orbital ϕ_i into the PAOs (approximately) centered on the atoms of the corresponding domain $[i]$. For doubles, excitations are made from both ϕ_i and ϕ_j into their pair domain, which is defined as the union of the individual domains $[i]$ and $[j]$. Moreover, based on the short range nature of the pair correlation energy, an exact correlation treatment is carried out only for pair domains composed of orbital domains spatially close (strong pairs). Different levels of approximation are then used for pair domains whose components are more and more distant from each other (weak pairs, distant pairs), and the most distant pairs are completely neglected (very distant pairs).

We can instead think of an alternative method for computing a linearly independent set of *orthonormal orbitals* spanning the virtual space. We can in

fact use Cholesky decomposition of a density-type matrix:

$$P^{v}_{\mu\nu} = \sum_{a} C_{\mu a} C_{\nu a} \ . \tag{2.28}$$

The orbitals depicted in Figure 3.3 show that this approach is perfectly suited for generating a set of localized orthonormal orbitals spanning the virtual space. In Paper I, this type of Cholesky localization of the virtual space was not investigated in detail. I assume that in the near future, the convenience of using orthonormal orbitals for the virtual space instead of (redundant) PAOs, will result in a great interest for such technique and for the Cholesky localization in general — probably the only orbital localization that works for both occupied and virtual orbitals, with and without symmetry, and produces localized orbitals directly from the density matrix, and *free of charge*.

4 A "unified" theory

An expert problem solver must be endowed with two incompatible qualities — a restless imagination and a patient pertinacity.

Howard W. Eve

Once the exchange problem had been solved, there was really no time to enjoy the success of the LK scheme and of the newly discovered Cholesky localization. Another problem plagued the Cholesky approximation to the two-electron integrals. This was possibly an even more serious one, as the same problem did not touch the DF approach. The problem can be stated very simply: the Cholesky decomposition technique as described in Chapter 2, and as adapted by Beebe and Linderberg [25] for the ERI matrix is a *numerical* procedure. This adjective has very often a negative tone in quantum chemistry, as opposite to the generally welcomed word *analytical*. The reason can be explained in few words. Most of quantum chemistry is not just about calculating the energy of a molecule including as much electron correlation as possible. Instead, the study of problems of interest in chemistry and other areas by quantum chemical methods goes through much more than that. The optimization of molecular structures, for example, is ubiquitous in real applications. The study of molecular properties, such as polarizabilities, NMR chemical shifts, etc., are often the hot-spot for accurate quantum chemical predictions. Computationally, all these investigations require the evaluation of energy derivatives. While the latter may be computed by numerical differentiation of the electronic energy, Pulay's pioneering work on the so-called "force method" [86] has made it perfectly clear that analytic derivative techniques provide a much more economic approach from the computational point of view. The superior computational efficiency

and accuracy of the analytic derivative implementations compared to finite difference calculations compensates the initial theoretical and programming difficulties. For a comprehensive description of this topic, see for example Refs. [87–89] and references therein.

It has become so crucial the use of analytic derivative techniques in studying for instance chemical reactions and molecular spectroscopy, that one can arguably say that quantum chemical approaches which do not allow an analytic formulation of the energy derivatives may be sooner or later destined for the history shelfs of quantum chemistry. This seemed to be the destiny of the Cholesky approximation, as there was no analytic formulation for the derivatives of the Cholesky vectors available yet. A major success of this research work has been indeed to find a way out from this tunnel, indeed by going back exactly to the point were we started, the original 1977 paper of Beebe and Linderberg that introduced the Cholesky approximation.

Early ideas on the evaluation of two-electron integral derivatives within the CD approximation can be found in the work of O'Neal and Simons [44]. Their point of view was essentially based on a Cholesky representation of the integral derivatives. Considering only the geomtrical gradients, this implies that the following representation is used:

$$(\mu\nu|\lambda\sigma^{(1)}) = \sum_{J}^{M} L^{J}_{\mu\nu} L^{J}_{\lambda\sigma^{(1)}} \,, \tag{2.29}$$

where (1) indicates the operation of taking the geometrical derivative of the corresponding AO function. By the properties of the Gaussian basis functions, the effect of such differential operator on a Gaussian-type orbital (GTO) is a linear combination of GTOs with angular momentum quantum numbers higher and lower than that of the initial GTO. In other words, one could think of complementing the list of orbital distributions with that of all non-zero first-order derivatives. The matrix to be Cholesky decompose is therefore defined in terms of the extended distributions $\{\chi_\mu\chi_\nu, \chi_\mu\chi^{(1)}_\nu\}$. Apart from this aspect, the Cholesky algorithm of Eqs. (2.11-2.12), can be applied in the identical manner. Indeed, O'Neal and Simons [44] were able to show that this approach is not only possible, but it also exploits the somewhat increased degree of linear dependence among the columns of the whole matrix. In practice, the effective rank of the extended matrix increases only sligthly in consequence of, and despite of the much larger number of orbital distributions. However, this way of computing the integral derivatives still orbits around the numerical Cholesky procedure, and therefore cannot be used to define analytical gradient expressions for the quantum mechanical energies. Moreover, it requires the decomposition of a very large matrix: in a molecule with N_A atoms there are $3N_A$ cartesian derivatives in principle needed, and the number of orbital distribution is N^2, where N represent the number of AO basis functions.

It is easy to realize that this route cannot take the Cholesky approximation anywhere near potential usability in geometry optimizations of even medium-sized molecules. The cost of the Cholesky decomposition itself can be quite high in many situations (see for example, Table IV in Paper 5), and this overhead could rapidly become untenable when extending the integral matrix with the differentiated GTOs. In addition to that, there is also an issue of compatibility with standard gradient code. Modern implementations of these facilities are designed to work in an integral-direct fashion. Integral derivatives are never stored on disk, but computed on-the-fly and contracted with the correponding density matrix elements. Efficient screening techniques are used in this case and the performances are remarkable, especially because any possible input-output bottleneck is eliminated.

With this state of affairs, my confidence of being able to find a solution to the problem of an analytic formulation of the derivatives of the Cholesky vectors, was not at all high. Together with the other persons who have contributed to the Cholesky implementation in MOLCAS, several possibilities were discussed. One of them was an attempt to compute the vector derivatives recursively in terms of the ERI derivatives, through differentiation of the Eqs. (2.11-2.12). Although possible in principle, this approach is clearly unfeasible. Another attempt, rather desperate indeed, is outlined in Appendix 4. For about two years, discussions on the subject of analytic derivatives of the Cholesky vectors became a constant in our group. But the general feeling was that of having reached a dead-end.

Once again, a fortunate event lead to a drastic change in this situation. This time, the source of inspiration was a work by the group around Head-Gordon and Gill. That paper [40] focuses on the fundamental issue relative to the choice of the metric in density fitting approximations to the ERIs. However, what really caught my attention was a simple detail of that paper, namely an equation of the following type:

$$(\mu\nu|\lambda\sigma) = \sum_{PQ} C^P_{\mu\nu} G_{PQ} C^Q_{\lambda\sigma} \; . \tag{2.30}$$

This was the definition that the authors of the paper used for *the density fitting representation of the integral* $(\mu\nu|\lambda\sigma)$, where P and Q represent the auxiliary basis functions and $G_{PQ} = (P|Q)$ indicates the Coulomb metric matrix. I had never seen this relation before. For me, and probably many others, the DF approximation was invariably associated to the presence of the *inverse* of the matrix $\mathbf{G}$, namely to the form reported in Eq. (2.2). [1]

Back to the problem of analytic gradients for the Cholesky vectors, this little detail became a great insight. Eq. (2.30) started to look vary familiar to me, but simply because it had the same form of the Cholesky representation, Eq. (2.13), except for the presence of the metric matrix $\mathbf{G}$, of course. I then started to think

[1] After reading next Section, the equivalence between the two formulation will be clear.

of the possibility that this was not just a formal similarity, but that it could hide a much deeper connection between DF and CD integral representations. The problem of computing analytic gradients for the Cholesky vectors had therefore become, the problem of finding *an analytic formulation of the CD approximation*. In other words, my hope was to be able to show that CD was a particular type of DF approximation: had this been proved, the solution of the analytic derivative problem in CD would have been simply a corollary to the proof. This hope was supported also by the fact that in the original CD paper of Beebe and Linderberg [25], there is a paragraph that sounds very much like a proof of this sort. The authors speak of the CD procedure as *a particular form of inner projection* of the Coulomb operator onto a specifically constructed basis set. My opinion now is that the true relevance of this statement had been overlooked for nearly thirty years. But about a year ago things started to change. The first unripe idea of an analytic formulation of the CD procedure is reported in Appendix 5. From that point, a joint effort of three people (authors of Paper III) has lead to the first analytic derivative formulation of the Cholesky integral approximation.

4.1 Analytic Cholesky decomposition

We will prove the central idea that lead to our solution to the problem of obtaining an analytic formulation of the Cholesky vector derivates. As anticipated, the idea is based on the proof that the Cholesky decomposition approach is equivalent to a particular type of density fitting. Such a "unified" theory is obviously a *sufficient* condition for an analytic formulation of the derivatives of the Cholesky vectors, as in the DF approach such formulation is trivial. We therefore start by reviewing the general theory on DF approximations. The latter aim at approximating the atomic orbital (AO) product distributions with a linear expansion on a set of auxiliary Gaussian basis functions $\chi_Q = |Q\rangle$: $\chi_\mu\chi_\nu = |\mu\nu\rangle \approx \sum_Q C^Q_{\mu\nu}\chi_Q$, and then minimize the "distance" (in Hilbert space) between the fitted distributions and the actual distributions, defined through a given positive definite metric $\hat{g}$,

$$f_{\mu\nu} = (\chi_\mu\chi_\nu - \sum_Q C^Q_{\mu\nu}\chi_Q|\hat{g}|\chi_\mu\chi_\nu - \sum_Q C^Q_{\mu\nu}\chi_Q) \,. \tag{2.31}$$

Minimization of this function leads to the set of linear equations:

$$\sum_P \langle P|Q\rangle_g C^P_{\mu\nu} = \langle Q|\mu\nu\rangle_g \,, \tag{2.32}$$

where $\langle\cdot|\cdot\rangle_g = (\cdot|\hat{g}|\cdot)$. The value of the function corresponding to the solution of Eq. (2.32) is given by

$$f^{min}_{\mu\nu} = \langle\mu\nu|\mu\nu\rangle_g - \sum_{PQ} C^P_{\mu\nu}\langle P|Q\rangle_g C^Q_{\mu\nu} \,. \tag{2.33}$$

Thus, $f_{\mu\nu}^{min}$ measures the error in representing the diagonal element $\langle\mu\nu|\mu\nu\rangle_{\hat{g}}$. We now restrict the discussion to the case of the Coulomb metric, $\hat{g} = r_{12}^{-1}$, as this is the most accurate choice for computing electronic properties [28, 40]. By solving Eq. (2.32) we obtain the more familiar inner projection representation in terms of two- and three-center integrals

$$(\mu\nu|\lambda\sigma) \approx \sum_{PQ}(\mu\nu|P)G_{PQ}^{-1}(Q|\lambda\sigma)\,, \tag{2.34}$$

where $G_{PQ}^{-1} = (G^{-1})_{PQ}$ and $G_{PQ} = (P|Q)$. Equation (2.33), which is the error in representing the two-electron integral diagonal elements, shows that in general a given auxiliary basis set does not lead to the global minimum of $f_{\mu\nu}$. As $f_{\mu\nu}$ is nonnegative, the global minimum is *zero* which is achieved by choosing the auxiliary basis set such that it spans the same space as the orbital distributions $\chi_\mu\chi_\nu$. The two-electron integrals are then represented exactly by the DF approach.

As discussed by Beebe and Linderberg [25], we can interpret the Cholesky decomposition in terms of a particular inner projection procedure onto a manifold generated from the original orbital distributions. If the Cholesky basis contains the M product functions h_J, we may introduce a set of M functions Q_J spanning the same space through the modified Gram-Schmidt procedure,

$$Q_J = \mathcal{N}_J\left[h_J - \sum_{K=1}^{J-1} Q_K(h_J|Q_K)\right]\,, \tag{2.35}$$

where the normalization constant is given by

$$\mathcal{N}_J = \left[(h_J|h_J) - \sum_{K=1}^{J-1}(h_J|Q_K)^2\right]^{-1/2}\,, \tag{2.36}$$

such that

$$(Q_J|Q_K) = \delta_{JK}\,. \tag{2.37}$$

This orthonormal (in the Coulomb metric) set of functions can of course be employed as auxiliary basis set in the DF procedure. The solution of Eq. (2.32) is then particularly simple, namely:

$$\begin{aligned} C_{\mu\nu}^J &= (\mu\nu|Q_J) \\ &= \mathcal{N}_J\left[(\mu\nu|h_J) - \sum_{K=1}^{J-1}(\mu\nu|Q_K)(h_J|Q_K)\right]\,. \end{aligned} \tag{2.38}$$

For $J = 1, 2$ we have

$$C_{\mu\nu}^1 = (\mu\nu|Q_1)$$

$$
\begin{aligned}
&= (h_1|h_1)^{-1/2}(\mu\nu|h_1) \\
&= L^1_{\mu\nu} \,, \qquad (2.39)\\
C^2_{\mu\nu} &= (\mu\nu|Q_2) \\
&= \widetilde{(h_2|h_2)}^{-1/2}\left[(\mu\nu|h_2) - L^1_{\mu\nu}L^1_2\right] \\
&= L^2_{\mu\nu} \,, \qquad (2.40)
\end{aligned}
$$

and in general,

$$C^J_{\mu\nu} = L^J_{\mu\nu} \,. \qquad (2.41)$$

Thus, DF and CD *are completely equivalent when the Cholesky basis is chosen as auxiliary basis set*. The Gram-Schmidt orthonormalization is not needed and the Cholesky basis therefore can be used directly in Eq. (2.32). The two-electron integrals can then be written as

$$(\mu\nu|\lambda\sigma) = \sum_J L^J_{\mu\nu}L^J_{\lambda\sigma} = \sum_{JK}(\mu\nu|h_J)G^{-1}_{JK}(h_K|\lambda\sigma) \,, \qquad (2.42)$$

where G is the Coulomb metric matrix in the Cholesky basis.

Using this DF reformulation of the CD representation of the integrals, it becomes straightforward to compute analytic derivatives within the CD approach. What presented here is only the basic idea behind our approach. There are a number of issues that require detailed discussion. For instance, the standard CD of the ERI matrix produces a Cholesky basis, $\{h_J\}$, that often contains product functions in which the AOs are centered on different nuclei. These two-center auxiliary functions are obviously not present in standard DF approaches. Indeed, the use of atom-centered auxiliary functions is a convenient choice, as it ensures that the resulting energy is a smooth function of the nuclear coordinates. The risk that the CD approach could produce discontinuities in the potential energy surface is a crucial point, and requires extensive studies. A preliminary investigation upon this issue is found in Paper III, together with a detailed discussion of the analytic gradient implementation for an approximate CD formulation. The latter employs a purely atom-centered Cholesky basis and is therefore called one-center CD, or 1C-CD. The use of this approximate CD approach is introduced in Paper IV, where the accuracy and the performance of the method are thoroughly analyzed.

The implementation of the analytic gradient code for the CD integral approximation in MOLCAS is still ongoing. At present, the code can only handle gradient calculations for the Coulomb contribution to the energy. It is therefore possible to run geometry optimizations for non-hybrid DFT models. Results shown in Paper III demonstrate the accuracy of the CD gradient calculation and also the improvability of the results through the choice of the decomposition threshold. The code is not yet optimal from a point of view of timings

performances. However, there is a justified expectation that the same type of cost-reduction experienced for the energy calculations, will be obtained in this case. The formulation of the gradients for the Coulomb term is reported in Paper III. As for the energy calculation, the bottleneck of the calculation is instead represented by the evaluation of the exchange contribution. In the following, a sketch of the extension of the analytic CD gradient formulation to HF and CASSCF wave function is briefly described.

Using the notation defined in Paper III, the gradients of the HF exchange energy within the CD approximation can be written as:

$$\begin{aligned} E_x^{(1)} &= 2\sum_K \sum_{\mu\nu\lambda\sigma} D_{\mu\lambda} C^K_{\mu\nu} (K|\lambda\sigma)^{(1)} D_{\nu\sigma} - \sum_{KL} \sum_{\mu\nu\lambda\sigma} D_{\mu\lambda} C^K_{\mu\nu} (K|L)^{(1)} D_{\nu\sigma} C^L_{\lambda\sigma} \\ &= 2\sum_{Kij} C^K_{ij} (K|ij)^{(1)} - \sum_{KL} (K|L)^{(1)} \sum_{ij} C^K_{ij} C^L_{ij} \\ &= 2\sum_{K\lambda\sigma} B^K_{\lambda\sigma} \, (K|\lambda\sigma)^{(1)} \; - \; \sum_{KL} (K|L)^{(1)} \, A_{KL} \, , \end{aligned} \tag{2.43}$$

where $\mathbf{X}$ is the occupied MO coefficient matrix and $B^L_{\lambda\sigma} = \sum_{ij} X_{\lambda i} X_{\sigma j} C^L_{ij}$ are the *backtransformed* $\mathbf{C}$ vectors to AO basis.

The straightforward implementation of Eq. (2.43) introduces a computational cost scaling as $\sim ON^2M + O^2M^2$ considerably higher than the one needed for the evaluation of the Coulomb term [**P3**]. However, if instead of canonical orbitals we use localized orbitals, an efficient screening can be used to reduce the costs of the MO transformations. We could assume for instance that *the orbital pairs that do not contribute considerably to the exchange energy are excluded from the evaluation of the gradients*. We know that the exchange energy contribution due to an orbital pair ij is given by the integral $(ij|ij)$. Then, the following inequality holds:

$$\begin{aligned} |E_x^{ij}| = |(ij|ij)| &= \sum_{\mu\nu\lambda\sigma} |X_{\mu i}||X_{\nu j}||(\mu\nu|\lambda\sigma)||X_{\lambda i}||X_{\sigma j}| \\ &\leq \sum_{\mu\nu\lambda\sigma} |X_{\mu i}||X_{\nu j}|(\mu\nu|\mu\nu)^{1/2}(\lambda\sigma|\lambda\sigma)^{1/2}|X_{\lambda i}||X_{\sigma j}| \\ &= (\sum_{\mu\nu} (\mu\nu|\mu\nu)^{1/2}|X_{\mu i}||X_{\nu j}|)^2 = Y_{ij}^2 \, . \end{aligned} \tag{2.44}$$

In a localized orbital basis, the use of the estimates Y_{ij} would allow to identify the set of *strong pairs* that most significantly contribute to the exchange energy. Owing to the fact that for localized orbitals the number of such pairs *increases only linearly* with the system size, the scaling of the assembly of the $\mathbf{A}$ matrix would be greatly reduced. However, the first half-transformation of the vectors and three-center integral derivatives would still scale quartically. In order to reduce the costs of this step, the same pre-screening employed for the LK method [**P1**], and based on the use of matrix norm estimates, can be used.

For the CASSCF energy, an additional term including the two-electron density matrix in the active space need to be computed:

$$\begin{aligned} E_a^{(1)} &= 2\sum_K \sum_{tuvx} P_{tuvx} C_{tu}^K (K|vx)^{(1)} - \sum_{KL} \sum_{tuvx} P_{tuvx} C_{tu}^K (K|L)^{(1)} C_{vx}^L \\ &= 2\sum_{Kp} \sum_{tu} T_{tu}^p C_{tu}^K \sum_{vx} T_{vx}^p (K|vx)^{(1)} - \sum_{KL} (K|L)^{(1)} \sum_p \sum_{tu} T_{tu}^p C_{tu}^K \sum_{vx} T_{vx}^p C_{vx}^L \\ &= 2\sum_{Kp} Q_K^p \sum_{\lambda\sigma} \Theta_{\lambda\sigma}^p \, (K|\lambda\sigma)^{(1)} - \sum_{pKL} Q_K^p \, (K|L)^{(1)} \, Q_L^p \,, \end{aligned} \tag{2.45}$$

where **T** are the vectors obtained by decomposing the (small) two-electron active density matrix, P_{tuvx}, and $\Theta_{\lambda\sigma}^p = \sum_{uv} X_{\lambda u} T_{vu}^p X_{\sigma v}$ are their AO basis backtransformed counterparts. The advantage of using its Cholesky decomposed form instead of contracting directly **P** is not a reduced computational cost but a saving in the storage of the various intermediates. The computational scaling of the required MO transformation and the evaluation of Eq. (2.45) are $\sim aN^2M + a^4M$ (where a, at most few dozens, is the number of active orbitals) and can be considered to be significantly less expensive than computing the Exchange gradients.

Finally, in order to integrate the code for the CD gradients into an existing integral-direct gradient code, the simplest solution is to construct some effective density matrices for the two-center and three-center contributions. The latter is given by:

$$P_{\lambda\sigma}^K = V^K D_{\lambda\sigma} - \frac{1}{2} B_{\lambda\sigma}^K + \sum_p Q_K^p \Theta_{\lambda\sigma}^p \,, \tag{2.46}$$

where $V^K = \sum_{\mu\nu} D_{\mu\nu} C_{\mu\nu}^K$ is defined as from Eq. (20) in Paper III, while the effective density for the two-center term is given by

$$P_{KL} = V^K V^L - \frac{1}{2} A_{KL} + \sum_p Q_K^p Q_L^p \tag{2.47}$$

4.2 *Ab initio* density fitting

Another case of serendipity enters the scene at this point. We have shown that the CD approximation is not just a numerical procedure for representing the ERI matrix, but it is better thought as a solution to a variational "density fitting" problem in the Coulomb metric. In fact, CD is not just a solution, but it is the *best possible* solution. What do we mean with that? Precisely the following. Using the Cholesky basis as auxiliary basis set, the value of $f_{\mu\nu}^{min}$, Eq. (2.33) with $\hat{g} = r_{12}^{-1}$, is bounded by the decomposition threshold,

$$f_{\mu\nu}^{min} = (\mu\nu|\mu\nu) - \sum_{J=1}^{M} \left(L_{\mu\nu}^J \right)^2 \leq \delta \,. \tag{2.48}$$

As the target functional, Eq. (2.31), of the variational problem is nonnegative, and therefore its absolute minimum is *zero*, then Eq. (2.48) indicates that the Cholesky basis is able to approximate the *exact solution* within the given threshold.

Standard auxiliary basis sets, on the other hand, are atom-centered Gaussian optimized for each valence basis set by minimization of the total energy error in a set of molecular calculations [32–35]. As such optimization is performed without any direct constraint regarding the representation of the integral diagonals $(\mu\nu|\mu\nu)$, there is no guarantee that with such auxiliary basis one can get a sufficiently accurate integral representation. Moreover, the corresponding solution of the density fitting variational problem not necessarily tends towards the exact solution when the size of the auxiliary basis set is increased. As a consequence, standard auxiliary basis sets show a strong bias towards the quantum chemical methods used to design them. On the other hand, the CD procedure can be viewed as providing a scheme for systematic improvements of the DF auxiliary basis set without bias towards a specific quantum chemical method.

With this in mind, we decided to use the CD in order to design a new generation of auxiliary basis sets, that could improve two aspects in the current status of DF technology: (a) remove the problem of bias, and (b) introduce in DF the strict error control and systematic improvability typical of the CD approximation. We chose then to define such type of auxiliary basis sets as the Cholesky basis obtained from a CD of the atomic two-electron integral matrix. This atomic CD (aCD) auxiliary basis set needs not be stored in one of the usual basis set library formats but is constructed on-the-fly and, consequently, may be applied in conjunction with any atom and valence basis set. The computational overhead of performing the aCD is insignificant, as the decomposition needs only to be done for each uniqe atom/valence basis set pair. The quality of the aCD auxiliary basis set is controlled by the aCD decomposition threshold. They constitute, therefore, a *hierarchy* of increasingly accurate auxiliary basis sets determined fully *ab initio*.

Although the value of $f_{\mu\nu}^{min}$ is no longer bounded in the case of DF approximation based on aCD auxiliary basis sets, the latter still provide increasingly accurate integral representations (see Figures 2 and 3 in Paper IV). The detailed analysis of the accuracy of the aCD auxiliary basis sets can be found in Paper IV, with additional information related to the size of such basis sets, and how this compares to the size of standard basis sets. Here I will only report an evidence of the unbiased nature of the aCD basis sets. Table 6 shows how the accuracy of the computed total energies and energy differences is completely independent on the quantum chemical method used. This can be shown to be true for any choice of the atomic orbital basis set and for any value of the decomposition threshold. The key to this fundamental feature lies in the fact that the aCD basis sets are (nearly) optimal, in terms of variational minimization

Table 6: Benzene molecule. Absolute deviations with respect to conventional calculations of the total energies at the BLYP, B3LYP, HF, and MP2 levels of theory with the SVP basis set. Density fitting with the RI-J, RI-C, and aCD auxiliary basis sets. Notation described in Paper IV.

Auxiliary basis	$\Delta E/mE_h$			
	BLYP	B3LYP	HF	MP2[a]
RI-J	0.866	19.679	95.238	101.147
RI-C	8.964	7.714	3.122	2.534
aCD-2*	5.318	1.889	11.718	12.542
aCD-3*	3.857	3.681	3.059	2.652
aCD-4*	0.014	0.016	0.151	0.208
aCD-5*	0.013	0.010	0.117	0.168
aCD-8*	0.012	0.001	0.107	0.156
aCD-2	81.708	74.928	47.712	41.683
aCD-3	27.690	27.573	26.910	24.210
aCD-4	0.487	0.285	0.257	0.932
aCD-5	0.473	0.291	0.199	0.835
aCD-8	0.116	0.009	0.417	1.015

[a]On top of DF-HF employing the same auxiliary basis.

of the "density fitting" functional. In light of that, they reproduce the whole ERI matrix with a uniform accuracy. The same is not true for standard auxiliary basis sets externally designed through data-fitting. The latter reproduce with sufficiently accuracy only specific subsets of the ERIs — and even that, without clear upper bounds for the errors.

I conclude this Chapter by remarking on how the analytic formulation of the CD approximation has not only solved the problem of the analytic CD derivatives, but has also opened a completely new window on a technology, the DF/RI approximation, that did not seem to have room for radical changes.

5 A leap in the future

Under normal conditions the research scientist is not an innovator but a solver of puzzles, and the puzzles upon which he concentrates are just those which he believes can be both stated and solved within the existing scientific tradition.

Thomas S. Kuhn

In the present Chapter, I will try to discuss the possibility to use the Cholesky approach in a completely different context than the one in which this technology has been used so far. These ideas have not been put to test yet, however they do seem to have solid foundations. I will go through the key points on which such ideas are built, but first let me explain the motivations behind them.

The computational scaling of Post-HF methods is essentially determined by the number of possible excited Slater determinants taken into account at a given level of theory. Using HF as reference functions, the correlation energy is expressed as sum of contributions deriving from single, double, triple, ... excitations; in the language of Coupled Cluster or MP Perturbation theory, this gives rise to a number of *excitation amplitudes* which consitute the unknowns of the problem. While the number of single excitation amplitudes (if present at all) is always tractable, the same cannot be said about higher-rank amplitudes. Already the number of doubles scales as $O(\mathcal{N}^4)$, where $\mathcal{N}$ is a suitable measure of the size of the system, posing severe limitations to the range of applicability of even the simplest Post-HF methods.

We have learned earlier in the present thesis that the Cholesky decomposition of the ERI is indeed a method that seeks for linear dependences in the atomic orbital product space. However, the generality of the method would

suggest that the same technique, applied to other contexts than the ERI matrix would work as well, for instance as means to recognize and eliminate redundancies in the number of parameters used to describe the wave function of the system. If this is the case, then a consistent reduction in the computational costs of any correlated method could be achieved by simply recasting the working equations in a reduced space of parameters.

5.1 Cholesky beyond the integrals

The following discussion focuses on the double excitations amplitudes, but a generalization to higher classes of excitation operators is certainly possible. In order to reduce the number of double excitation amplitudes to be computed in a correlated treatment, at least two different approaches are widely used. One possibility is to formulate the problem in a localized orbital basis and discard, by heuristics, those amplitudes connecting orbital pairs located in regions (domains) distant from each other. These so-called *local correlation methods* [39,61,71,74,78] are therefore intrinsically linear scaling but show some drawbacks, such as the fact that they are effective only for large molecules, and that the partition of the molecule in domains is somewhat arbitrary. Another approach used to shorten the list of double excitation amplitudes is that of allowing excitations from the occupied orbitals only to a limited number of virtual orbitals. These *optimized virtual orbital space* methods [90, 91] work essentially by including in the *correlating space* only the low-lying virtual orbitals up to a certain energy level. Various criteria have been established to perform this type of optimization, but in all cases the limitations of the method lie, on the one hand, in the fact that it is not guaranteed that most of the negligible excitations are excluded, and on the other hand, in the difficult *a priori* partition of the virtual space.

Is there a way to select among the full set of double excitation amplitudes only those which are truly relevant, and therefore use only those as unknowns for our problem? This is the basic question to which I would like to give a possible answer. In other words, could the (singlet) doubly excited component of the wave function generated from the Hartree-Fock state, $|HF>$,

$$\Psi_2 = \hat{T}_2|HF> = \frac{1}{2}\sum_{aibj} t_{ai,bj}\hat{E}_{ai}\hat{E}_{bj}|HF> = \frac{1}{2}\sum_{aibj} t_{ai,bj}\Phi_{ai,bj}\ , \tag{2.49}$$

be expressed in terms of a linear combination of Slater determinants smaller in number than the full set of $\Phi_{ai,bj}$ of Eq. (2.49) ?

The route we are going to pursue is that of exploiting possible linear dependences in the space of the *singly excited* determinants and use them to rewrite the $\hat{T}_2$ operator of Eq. (2.49) in a set of *nonredundant* parameters. In order to do that, we expand the singly excited determinants Φ_{ai} in a functional basis, *auxiliary Slater determinants*, Θ_R and compute the expansion coefficients $Z_{ai,R}$ in order to minimize the distance (in Hilbert space) between the exact and the approximate function, thus

$$Min\ \ f_{ai} = < \Phi_{ai} - \sum_R Z_{ai,R}\Theta_R \mid \Phi_{ai} - \sum_R Z_{ai,R}\Theta_R >_{\hat{g}}\ , \tag{2.50}$$

where $\hat{g}$ represents a given metric. Before discussing in details the choice of the metric and the solution of the variational problem of Eq. (2.50), it is worth

noticing that once the basis Θ_R has been chosen and the coefficients $Z_{ai,R}$ found, the doubly excited component of the wavefunction can be written as:

$$\Psi_2 = \frac{1}{2}\sum_{aibj} t_{ai,bj}\hat{E}_{ai}\sum_R Z_{bj,R}\Theta_R = \frac{1}{2}\sum_{aiR}\omega_{ai,R}\hat{E}_{ai}\Theta_R \ . \tag{2.51}$$

Therefore, the complexity of our problem has been moved from the determination of the $t_{ai,bj}$ unknowns to that of the $\omega_{ai,R}$, thus possibly reduced! Exactly what we were aiming to.

Back to the functional of Eq. (2.50). The choice of the metric is somewhat arbitrary but we would probably prefer to deal with a positive definite (PD) operator because then the functional would be nonnegative. For any given metric, the solution of the variational problem of Eq. (2.50) leads to the following set of linear equations

$$\sum_S < \Theta_R|\Theta_S >_{\hat{g}} Z_{ai,S} =< \Theta_R|\Phi_{ai} >_{\hat{g}} \ , \tag{2.52}$$

with the value of the functional at its minimum given by:

$$f_{ai}^{min} =< \Phi_{ai}|\Phi_{ai} >_{\hat{g}} - \sum_{RS} Z_{ai,R} < \Theta_R|\Theta_S >_{\hat{g}} Z_{ai,S} \ . \tag{2.53}$$

If the functional is nonnegative (positive definite metric), it can be proven (see Chapter 4) that its absolute minimum is obtained when these two conditions are both fulfilled:

- the basis functions Θ_R are orthonormal in the given metric;
- the coefficients are chosen to be the Cholesky vectors resulting from the decomposition of the symmetric PD matrix $M_{ai,bj} =< \Phi_{ai}|\Phi_{bj} >_{\hat{g}}$.

In order to use the standard formulation of the electron correlation methods, the most natural choice for the auxiliary Slater determinants is actually a linear combination of a *subset* of the original Φ_{ai} functions. Through the Gram-Schmidt procedure **[P3]**, an *orthonormalization matrix* **G** can be obtained. Then the auxiliary basis can be written as $|\Theta_R >= \mathbf{G}|\Phi_S >$, where only the selected Φ_S are referenced. With this choice of the auxiliary basis, Eq. (2.51) can be written in a very convenient form

$$\Psi_2 = \frac{1}{2}\sum_{aiR}\omega_{ai,R}\hat{E}_{ai}\sum_S G_{RS}\hat{E}_S|HF >= \frac{1}{2}\sum_{aiS}\theta_{ai,S}\hat{E}_{ai}\hat{E}_S|HF > \ , \tag{2.54}$$

where the index *S now just refer to a subset* $\{bj\}$ *of the full single excitation space*. It is also intended that we are only interested in the fact that the matrix **G** exists but we do not need to compute it. Now it is by the use of the Cholesky

decomposition of the metric matrix that we will answer the last (and initial) question, namely how to perform the selection of the nonredundant amplitudes.

Any positive definite metric can in principle be chosen, however we will discuss only the following three metrics:

$$\hat{g} : < \Phi_{ai} | \Phi_{bj} >_{\hat{g}} = \begin{cases} \dfrac{1}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j} & (2.55) \\ (ai|bj) & (2.56) \\ \dfrac{(ai|bj)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j} & (2.57) \end{cases}$$

expressed in their expectation value form. One might ask why from the above list the ubiquitous overlap metric is missing. The answer is very simple: in the overlap metric the Φ_{ai} are orthonormal by construction and therefore linearly independent.

The metric matrix (2.55), thus the second-rank orbital energy denominators, has the advantage of being extremely simple to decompose. Indeed, the order in which the decomposition should be carried out can be derived *without performing any Cholesky decomposition at all*. Since it is exactly this order (thus the *bj* selection) the only thing we are interested in, this choice is computationally very appealing. However, we must bear in mind that we also need to be able to relate the accuracy of the Cholesky decomposition to that of the correlation energy computed with the nonredundant amplitudes. In other words, how should the decomposition threshold — or equivalently, the number of resulting vectors — be chosen in order to ensure a given accuracy of the correlation energy? The simple answer is: we would need to experiment on that. But we shall return to this crucial point later on in the present discussion.

The second metric, Eq. (2.56) has the advantage of being defined not only in canonical orbitals but *for any choice of the orbital basis*, e.g. localized orbitals. Unfortunately, for this metric we need to really perform the Cholesky decomposition of the (large) metric matrix and this is something to take into account in the context of reduced-scaling methods. However, besides its generality, this metric presents also the nice feature of having probably an higher chance of matching the accuracy of the final energy with that of its Cholesky decomposition. Finally, the cost of the decomposition for this choice of the metric is partially recovered by the fact that the vectors can be fruitfully used in the subsequent amplitudes equations as representation of at least some of the two-electron integrals.

Now we come to the third metric of the list, thus the one defined by Eq. (2.57). It presents the following drawbacks: (a) it is only defined in canonical orbital basis, (b) needs to be explicitly decomposed, and (c) the resulting Cholesky vectors cannot be used in the subsequent amplitudes equations (except perhaps in MP2) [**P7**]. Why is it still interesting then? This metric, which at first glance

just looks like a merging of the first two, has actually a very precise meaning: it is the *matrix representation of the canonical MP2 amplitudes* (except for the sign). In other words it is probably the best candidate for exploiting the type of linear dependences we are looking for.

Now the very fundamental question is: can we somehow combine the computational simplicity of the first metric with the more solid physical nature of the third one and therefore achieve an optimal compromise between accuracy of the computed energy and computational costs? One possible answer to this question can be based on the fact that experience tells us that the decomposition of the MP2 amplitudes produces a number of vectors of the same order of magnitude [**P7**] of the one obtained when decomposing (to the same threshold) the $(ai|bj)$ integrals, which in turn is very similar to the *effective rank M* of the AO integral matrix. We can therefore rely on heuristics and adopt the following **selection criteria** to identify nonredundant amplitudes: *the size of the restricted space S is equivalent to the effective rank of the AO integral matrix, but the individual elements are sorted according to the decomposition pattern of the second-rank orbital energy denominators*. In this way, we hope to achive an accuracy for the correlation energy, obtained by solving only for the nonredundant amplitudes, similar to the AO integrals decomposition threshold, and this without any serious additional cost compared to a Cholesky based Post-HF method. We also note that while this selection certainly includes the HOMO-LUMO excitation[1] (as the first one actually), it is in general not equivalent to simply selecting the first M low-lying pairs (bj). However, it is fair to believe that most of the low-lying pairs will be selected through our procedure because they are associated with the largest diagonals of the second-rank orbital energy denominator matrix. Simply speaking, our selection criteria is at least not in conflict with physics and common sense.

We could think of how to extend this idea of selecting a nonredundant subspace of orbital parameters, to time-dependent (TD) HF and KS theory. Following the TDKS formulation of Sałek *et al.* [92], we define an effective density operator

$$\tilde{\rho}(r) = \sum_I \varphi_I(r) E_I \,, \tag{2.58}$$

in contrast to the exact density operator

$$\tilde{\rho}_{exact}(r) = \sum_{pq} \phi_p(r)\phi_q(r) E_{pq} \,. \tag{2.59}$$

Although the indices p and q refer to the full orbital space, the only relevant part for the time evolution of the electron density in DFT and HF theory is the subspace where p is a virtual orbital and q is an occupied orbital. The procedure

[1] Point group symmetry can be taken into account by performing the selection on separate matrices in which the orbital pairs match a given direct product symmetry.

described above allows for the selection of a further reduced linearly independent subspace of the latter, and therefore the index I is intended as compound index ai for such selected subspace. Eq. (2.58) can be also interpreted as derived from Eq. (2.59) when the orbital products are replaced by the "Cholesky basis". Recognizing that the density is an expectation value of the operator defined in Eq. (2.58), we deduce that an *effective density* is defined through the operator (2.58). From now on, we will focus on the time-evolution of this effective density, rather than that of the exact electron density.

We can assume that, as the KS (or HF) density is sufficiently well approximated by its effective counterpart, also the time evolution of the KS (or HF) determinant can be described within this effective subspace. In other words, the exponential parametrization of the TDKS (or TDHF) determinant can be represented through an effective $\tilde{\kappa}(t)$ operator

$$\hat{\kappa}(t) = \sum_I \kappa_I(t) E_I \, . \tag{2.60}$$

Likewise for the nth-order perturbed KS (or HF) hamiltonian

$$\hat{H}^{(n)} = \sum_I f_I^{(n)} E_I \, . \tag{2.61}$$

Following the exact same derivation as proposed by Sałek *et al.* [92], we arrive at the fundamental equation for linear response

$$(\mathbf{E} - \omega \mathbf{S}) \kappa^\omega = \mathbf{V}^\omega \, , \tag{2.62}$$

with the pivotal difference that here the dimensionality of the problem is M^2 and not O^2V^2. Does this make any sense? Imagine we are computing the excitation energies with two different basis sets which differ by the fact that the second has few additional high-angular momentum functions. The dimensionality of the problem in the second case has increased, but most likely the excitation energies up to a certain value would not differ much. In other words, within a given scope and accuracy of the solution, the second model contains redundancies. The idea of using the Cholesky subspace is indeed an attempt to find these types of redundancies.

5.2 Concluding remarks

The ideas briefly described in this Chapter, as they have not yet been tested, may reveal unsuccessful. However, the facts accounted in the present thesis have shown in several occasions that whenever "Cholesky decompositions" can play a role, the outcome is often better than what one could have hoped for. If I go three years back in my life from now, I cannot find any reference of even the

word Cholesky. Now, for almost every new development in quantum chemistry I see the potential of thinking differently, thus *thinking Cholesky*. In particular, the variational formulation of the Cholesky approximation, described in Chapter 4, is the key to exploit a range of new possibilities in quantum chemistry. The generality of that formulation has already produced one pivotal result, namely the possibility to change the future paradigm for the DF approximations — no more *yet-another data-fitting* in "ab initio quantum chemistry", and therefore no more *bona fide* accuracy of the "ab initio results".

Looking at the future, however, one should not forget that Cholesky integral approximation has a present in which it stems as mature technology for accurate and *truly ab initio* quantum chemistry applied to large-scale problems. More importantly, it is not just a way to speed up calculations; it also guarantees a controllable accuracy of the results, and that happens not by fortunate error cancellations, but simply *for the right reason*. Finally, if with Cholesky decompositions we can even localize orbitals, then we are all lead to think that there must be *more to come*.

Appendix

We are deceived by those who would have us believe that a multitude of affairs blocks their pursuit of liberal studies. As for me, Lucilius, my time is free, and wherever I am, I am master of myself.

Lucius Annaeus Seneca *(Moral Epistles)*

The pages that follow are a collection of five short manuscripts on various quantum chemical related topics. They are as much as possible adherent to the original versions of the private scientific correspondence exchanged with some of my colleagues at the department. This section is detached from the rest of the present book, therefore neither the terminology, nor the format of the manuscripts is coherent with the preceding parts. Some of the material may appear out-dated to the careful reader of the previous chapters. It is indeed my intention to show with such piece of evidence that some of the important achievements I have contributed to and described earlier in this book, went through many failures and dead-ends. For the material still of interest, I hope future developments could be built on top of those unripe ideas.

I will briefly introduce the topic of each of these manuscripts in such a way that the reader is made aware of their meaning in the context of the present thesis. The first one was written after I read a paper by Manby and Knowles [38] on "Poisson Equation in the Kohn-Sham Coulomb Problem". I realized that their approach was in part an attempt to expand the Coulomb potential in terms of atom-centered Gaussian functions. My question then was: can one use such an expansion and derive a set of conditions for the expansion coefficients to be used for a KS-DFT free from two-electron repulsion integrals?

The second manuscript focuses on a redefinition of the zeroth-order hamiltonian in MP2. This is not an attempt to re-invent local correlation methods. The idea is to add-up the reduction in computational costs arising from the use of localized orbitals with that obtained avoiding any iterative procedure to compute the MP2 amplitudes. The approach is based on using a different zeroth-order hamiltonian than the Fock operator. One problem is certainly the fact that one of the basic assumptions in perturbation theory (cf. Section 1.3), namely that the energy perturbation itself is "small" compared to the zeroth-order values, may not hold in this case.

The subject of the third manuscript is again MP2 theory, but this time I investigated a possibility to Cholesky decompose directly the compound matrices that arise from the canonical MP2 expressions. With the latter, I mean the matrices having for elements products of two-electron integrals. The approach seems sounding, but it needs to be tested, since it assumes that these matrices have the same effective rank of the original integral matrix. Mathematical reasoning does not support this assumption, but neither votes it down completely.

The fourth manuscript is really of purely historical value. It shows how many possible writhed paths the human mind can follow while searching for solutions to a problem. The theory exposed there to compute approximations to analytical derivatives of the Cholesky vectors is fortunately no longer needed.

Finally, the fifth manuscript represents what was the starting point for the idea of *ab initio* density fitting, discussed in Chapter 4. However, the focus is put on the use of Cholesky decomposition not only for the one-center orbital distributions, but also for the two-center ones. It is also mentioned the possibility to use the Poisson technique of Manby and Knowles [38] to avoid completely the evaluation of electron repulsion integrals involving more than two centers.

A.1 On a shortcut for the evaluation of exact Coulomb energies in *ab initio* quantum mechanical methods via Poisson equation

October 2005

In electronic structure theory, the electronic density is usually written in terms of the AO-basis χ used for expanding the molecular orbitals[1], thus the wavefunction. The expression for the density is

$$\rho(\mathbf{r}) = \sum_{\mu\nu} \gamma_{\mu\nu} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \,, \tag{A1}$$

where $\gamma_{\mu\nu}$ are the elements of the one-electron density matrix (in AO basis). For reasons that will be clear later on, we will refer to γ as the *wavefunction density matrix*.

The Coulomb energy is defined as:

$$J[\rho] = \frac{1}{2} \int d\mathbf{r_1} \rho(\mathbf{r_1}) \int d\mathbf{r_2} \frac{\rho(\mathbf{r_2})}{r_{12}} = \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \sum_{\lambda\sigma} \gamma_{\lambda\sigma} (\alpha\beta|\lambda\sigma) = \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} F^c_{\alpha\beta} \,, \tag{A2}$$

where $F^c_{\alpha\beta}$ are by definition the elements of the *Coulomb Fock matrix*. Equation (A2) is the usual way in which Coulomb contributions to the electronic energy are computed. The problem with this formulation however is the evaluation of the often very large number of ERIs, due to the fact that they are two-electron (mostly 4-center) integrals. Their number thus scales in general as N^4, where N is the size of the AO basis.

Now you can already see the point: if we were able to compute this energy expression by avoiding the use of ERIs, we would be happy. Probably we will have to pay a price for that, but if this price is less than computing the ERIs, we are on the right track. Let's see how we can achieve that.

The electronic density and the Coulomb potential are related by the Poisson equation:

$$\hat{P} v(\mathbf{r}) = \rho(\mathbf{r}) \,, \tag{A3}$$

where we have defined the *Poisson operator* $\hat{P} = -(4\pi)^{-1} \nabla^2$, which apart from a factor $(2\pi)^{-1}$ is identical to the *kinetic energy operator*, thus it is a *one-electron operator*. The Coulomb energy can be written in terms of the Coulomb potential as (trivial substitution):

$$J[\rho] = \frac{1}{2} \int d\mathbf{r_1} \hat{P} v(\mathbf{r_1}) \int d\mathbf{r_2} \frac{\hat{P} v(\mathbf{r_2})}{r_{12}} \,, \tag{A4}$$

[1] We will assume for simplicity all quantities to be real, the generalization being straightforward.

and we also remember a general identity of the analysis, namely

$$\int d\mathbf{r_2} \frac{\hat{P}\phi(\mathbf{r_2})}{r_{12}} = \phi(\mathbf{r_1}) \,, \tag{A5}$$

which *holds exactly for any* $\phi(\mathbf{r})$ *that vanishes more quickly than* r^{-1} *as* $r \to \infty$.

So far we have not achieved much, but we are getting closer. Without loosing in generality, we can write the Coulomb potential as

$$v(\mathbf{r}) = \sum_\alpha \bar{C}_\alpha \varphi_\alpha(\mathbf{r}) + f(\mathbf{r}) \,, \tag{A6}$$

(note that $\bar{C}_\alpha$ are not the MO coefficients!) where the $\varphi(\mathbf{r})$ constitute a *suitable basis* for the expansion of the potential in the region where the density is non zero. For semplicity we will refer to this (mathematical) region as *inside the molecule*. We can then assume the following equations being satisfied

$$f(\mathbf{r}) = 0 \qquad \text{inside the molecule} \tag{A7}$$

$$\hat{P}f(\mathbf{r}) = 0 \qquad \text{outside the molecule} \tag{A8}$$

$$\sum_\alpha \bar{C}_\alpha \hat{P}\varphi_\alpha(\mathbf{r}) = \rho(\mathbf{r}) \qquad \text{inside the molecule} \tag{A9}$$

In other words, $f(\mathbf{r})$ is chosen to be *a solution to the corresponding Laplace equation*. The previous two equations automatically imply that the Poisson equation is satisfied by $v(\mathbf{r})$. In the next section we will show that *there exist* a set of gaussian functions that satisfies *exactly* equation (A9) *if the density is given by* (A1).

We can now write the Coulomb energy as follows:

$$\begin{aligned} J[\rho] &= \frac{1}{2}\sum_{\alpha\beta} \bar{C}_\alpha \bar{C}_\beta \int d\mathbf{r_1} \hat{P}\varphi_\beta(\mathbf{r_1}) \int d\mathbf{r_2} \frac{\hat{P}\varphi_\alpha(\mathbf{r_2})}{r_{12}} \\ &+ \sum_\mu \bar{C}_\mu \int d\mathbf{r_1} \hat{P}\varphi_\mu(\mathbf{r_1}) \int d\mathbf{r_2} \frac{\hat{P}f(\mathbf{r_2})}{r_{12}} \\ &+ \frac{1}{2} \int d\mathbf{r_1} \hat{P}f(\mathbf{r_1}) \int d\mathbf{r_2} \frac{\hat{P}f(\mathbf{r_2})}{r_{12}} \\ &= \frac{1}{2}\sum_{\alpha\beta} \bar{C}_\alpha \bar{C}_\beta < \varphi_\alpha | \hat{P} | \varphi_\beta > \\ &= \frac{1}{2}\sum_{\alpha\beta} \bar{D}_{\alpha\beta} P_{\alpha\beta} \,, \end{aligned} \tag{A10}$$

where we have used the equations (A8) and (A9) together with the fact that the φ certainly satisfy the requirement of vanishing more quickly than r^{-1} as $r \to \infty$, because they *can be chosen to be* Gaussian basis functions (CGTOs).

The analogy between eq.(A2) and eq.(A10) is now more interesting: *The Poisson matrix elements* $P_{\alpha\beta}$ *play exactly the same role as the corresponding elements of*

the Coulomb Fock matrix. In other words, if we use as a definition for the Coulomb Fock matrix elements the derivative with respect to the corresponding density matrix elements of the Coulomb energy, the Poisson matrix is a representation of the Coulomb Fock matrix if we had used the following *field density matrix* $\bar{D}_{\alpha\beta} = \bar{C}_\alpha \bar{C}_\beta$ to start with. This is somewhat a result that could be guessed *a priori*, since v and ρ should be completely interchangeable for what concerns the quantum mechanical description of our system.

At this point, our problem is solved in theory but not in practice. How do we construct the field density matrix ? In other words, how do we get the expansion coefficient $\bar{C}_\alpha$? We are used to construct the wavefunction density matrix, and we certainly do not want to develop a new theory in order to get suitable approximated expansion of the potential. What can we do?

The answer is very easy and lies again in the beauty of the Poisson equation, which I now write in AO-basis

$$\sum_\alpha \bar{C}_\alpha \hat{P}\varphi_\alpha(\mathbf{r}) = \sum_{\mu\nu} \gamma_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}) \; . \tag{A11}$$

If we now multiply both sides (note that P is hermitian) by φ_β and integrate, we get our final equation

$$\sum_\alpha P_{\alpha\beta}\bar{C}_\alpha = \sum_{\mu\nu} \gamma_{\mu\nu} < \chi_\mu\chi_\nu\varphi_\beta > \tag{A12}$$

where there are only one-electron integrals! Namely, of kinetic energy and 3-center overlap type.

In conclusion, instead of computing ERIs and contract them with the wavefunction density in order to get Coulomb Fock matrix first and then the Coulomb energy, *we can construct the Poisson matrix, solve the linear system (A12) and contract the Poisson matrix with the field density matrix*. This must be done whenever the wavefunction density changes (each iteration in SCF).

So, the price to pay (if the linear system is solved in an efficient way) is relatively little compared to:

- computing the ERIs (if they are not needed elsewhere);
- reading the ERIs from disk at each iteration in conventional methods or recomputing them in a direct fashion;
- construct the Coulomb Fock matrix from the ERIs.

It is worth noticing that the linear system (A12) very likely shows a sparse structure for large molecules, which makes it very competitive in that case.

A.2 On the MBPT in localized spin-orbital basis: the $L_{ov}MP$ approach

March 2006

Any set of occupied localized orbitals $\{ \varphi^{L_o} \}$ fulfills[2] the (pseudo)eigenvalue equation of an effective Fock (or Kohn-Sham) hamiltonian defined as:

$$\hat{f}^{L_o}\varphi_i^{L_o} = [\hat{f} - \hat{\rho}\hat{f}\hat{\rho} + \hat{\rho}\hat{L}_o\hat{\rho}]\varphi_i^{L_o} = \lambda_i\varphi_i^{L_o} \,, \tag{A13}$$

where $\hat{\rho}$ is the projector onto the occupied space and $\hat{L}_o$ is a hermitian localization operator. The λ represent the eigenvalues of $\hat{f}^{L_o}$ and $\hat{L}_o$. In a similar way, a given localization operator $\hat{L}_v$ for the virtual orbitals leads to the (pseudo)eigenvalue equation

$$\hat{f}^{L_v}\varphi_a^{L_v} = [\hat{f} - \hat{\sigma}\hat{f}\hat{\sigma} + \hat{\sigma}\hat{L}_v\hat{\sigma}]\varphi_a^{L_v} = \omega_a\varphi_a^{L_v} \,, \tag{A14}$$

where $\hat{\sigma} = \hat{1} - \hat{\rho}$ is the projector onto the virtual space. Moreover, it is trivial to show that

$$\hat{f}^{L_o}\varphi_a^{L_v} = \hat{f}\varphi_a^{L_v} \,, \tag{A15}$$

$$\hat{f}^{L_v}\varphi_i^{L_o} = \hat{f}\varphi_i^{L_o} \,. \tag{A16}$$

The operators defined in (A13) and eq. (A14) if acting on the canonical orbitals give the follwing results:

$$\hat{f}^{L_o}\varphi_i^{can} = \varphi_i^{L_o} \,, \qquad \hat{f}^{L_o}\varphi_a^{can} = \hat{f}\varphi_a^{can} = \epsilon_a\varphi_a^{can} \,, \tag{A17}$$

$$\hat{f}^{L_v}\varphi_a^{can} = \varphi_a^{L_v} \,, \qquad \hat{f}^{L_v}\varphi_i^{can} = \hat{f}\varphi_i^{can} = \epsilon_i\varphi_i^{can} \,. \tag{A18}$$

In analogy to canonical MP2 theory, we can write the hamiltonian as a sum of a zeroth-order term and a fluctuation potential (besides the nuclear attraction term)

$$\hat{H} = \hat{H}_o^{L_{ov}MP} + \hat{\Phi} + \hat{h}_{nuc} \,, \tag{A19}$$

where the zeroth-order hamiltonian in the localized basis $\{ \varphi^{Lo}, \varphi^{Lv} \}$ is given by

$$\hat{H}_o^{L_{ov}MP} = \hat{f}^{L_o} + \hat{f}^{L_v} - \hat{f} = \sum_i^{occ} \lambda_i\hat{a}_i^\dagger\hat{a}_i + \sum_a^{vir} \omega_a\hat{a}_a^\dagger\hat{a}_a \,. \tag{A20}$$

Eq. (A20) indicates that these localized orbitals are the eigenfunctions of $\hat{H}_o^{L_{ov}MP}$. In addition, $\hat{H}_o^{L_{ov}MP}$ transforms the canonical orbitals $\{ \varphi_{occ}^{can}, \varphi_{vir}^{can} \}$ to the corresponding localized orbitals $\{ \varphi^{Lo}, \varphi^{Lv} \}$.

[2] L. Seijo and Z. Barandiarán, *J. Chem. Phys.*, **121** (6698) 2004.

As a result of the properties of the zeroth-order hamiltonian (A20), we can easily prove that

$$\hat{H}_o^{L_{ov}MP}|HF>= E^{(0)}_{L_{ov}MP}|HF>=\sum_i^{occ}\lambda_i|HF> , \tag{A21}$$

$$\hat{H}_o^{L_{ov}MP}|\mu>= E^{(0)}_{\mu}|\mu>=(E^{(0)}_{L_{ov}MP}+\Delta_\mu)|\mu> , \tag{A22}$$

where $|\mu>$ represents the excitation level and for example $\Delta_{\mu=2}=\omega_a-\lambda_i+\omega_b-\lambda_j$. We observe that $\hat{H}_o^{L_{ov}MP}$ possesses the same set of eigenvectors (namely $|HF>$ and $|\mu>$) of the zeroth-order hamiltonian of the usual Møller-Plesset theory, the Fock operator $\hat{f}$. Its eigenvalues, however, are different from the eigenvalues of $\hat{f}$ implying that

$$E^{(0)}_{L_{ov}MP}\neq E^{(0)}_{MP} \qquad and \qquad \Delta_\mu\neq\epsilon_\mu . \tag{A23}$$

We can now compute the first order correction to the energy

$$E^{(1)}_{L_{ov}MP}=<HF|\hat{\Phi}|HF>= E_{HF}-E^{(0)}_{L_{ov}MP}-h_{nuc} , \tag{A24}$$

where we find the important result that, although both $E^{(0)}_{L_{ov}MP}$ and $E^{(1)}_{L_{ov}MP}$ are different from the corresponding Møller-Plesset values, their sum still gives the Hartree-Fock energy. This should be expected, since in both theories the starting point is the Hartree-Fock wave function and what we are using is its property of invariance with respect to orbital rotations within the occupied space.

The first-order wave function is then given by

$$|L_{ov}MP1>=-\sum_\mu(E^{(0)}_\mu-E^{(0)}_{L_{ov}MP})^{-1}|\mu><\mu|\hat{\Phi}|HF>=-\sum_{\mu_2}|\mu_2>\Delta^{-1}_{\mu_2}<\mu_2|\hat{H}|HF> , \tag{A25}$$

and therefore it is different from the Møller-Plesset one. As a consequence, the 2nd-order correction to the energy is also not the same

$$E^{(2)}_{L_{ov}MP}=<HF|\hat{\Phi}|L_{ov}MP1>=<HF|\hat{H}|L_{ov}MP1> \neq <HF|\hat{H}|MP1>= E^{(2)}_{MP} . \tag{A26}$$

Following the usual derivation of the expressions for the energy corrections in MP theory, we come to the result for the $L_{ov}MP2$ energy in localized spin-orbital basis

$$E_{L_{ov}MP2} = E_{HF} + E^{(2)}_{L_{ov}MP} = E_{HF} - \sum_{a>b,\, i>j}\frac{|<ab||ij>|^2}{\omega_a-\lambda_i+\omega_b-\lambda_j} . \tag{A27}$$

Hybrid approaches where for instance the localized orbital basis is only used for the occupied space can be of interest as well. In the basis $\{\ \varphi^{Lo}, \varphi^{can}_{vir}\ \}$ the

zeroth-order hamiltonian reads

$$\hat{H}_o^{L_oMP} = \hat{f}^{L_o} = \sum_i^{occ} \lambda_i \hat{a}_i^\dagger \hat{a}_i + \sum_a^{vir} \epsilon_a \hat{a}_a^\dagger \hat{a}_a \; . \tag{A28}$$

with the energy expression given by

$$E_{L_oMP2} \; = \; E_{HF} \; + \; E_{L_oMP}^{(2)} \; = \; E_{HF} \; - \; \sum_{a>b,\, i>j} \frac{| < ab||ij > |^2}{\epsilon_a - \lambda_i + \epsilon_b - \lambda_j} \; . \tag{A29}$$

How are the localization operators L_o and L_v defined? *Gilbert* [3] has shown the form of some of the most common localizing operators. Following his results, we can then deduce that once the *canonical* Hartree-Fock equations have been solved, we need to perform an orbital localization of the occupied space (and in general also of the virtual space.[4] The most common localization procedures can be express as condition on the *hermiticity* of a specific matrix, as for example:

Boys

$$B_{ij} \; = < i|\mathbf{r}|i > \cdot < i|\mathbf{r}|j > \, , \tag{A30}$$

Edminston-Ruedenberg

$$R_{ij} \; = \; (ii|ij) \, , \tag{A31}$$

Pipek-Mezey

$$P_{ij} \; = \; \sum_A < i|\hat{P}_A|i >< i|\hat{P}_A|j > \; . \tag{A32}$$

The matrix relative to the chosen localization method can be then diagonalized to produce the eigenvalues and eigenvectors (almost identical to the given localized orbitals) needed in the formulae for the 2nd-order energy correction.

A.3 On reducing the scaling of Cholesky MP2

May 2006

The closed-shell MP2 energy in canonical orbitals is given by

$$E_2 \; = \; \sum_{aibj} \frac{(ai|bj)^2}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j} \; - \; \frac{1}{2} \sum_{aibj} \frac{(ai|bj)(aj|bi)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j} \; . \tag{A33}$$

[3] T. L. Gilbert, in *Molecular Orbitals in Chemistry, Physics and Biology*, edited by P. O. Löwdin and B. Pullman (Academic, New York, 1964), pp. 405–420 .

[4] Due to slow convergence of the usual localization methods for virtual orbitals, the best approach in this case is perhaps to use a spectral representation of the *Cholesky localization operator*. See also *Seijo et al.* - Sec. IIA, Eq. (20).

The MO integrals can be computed using the MO transformed Cholesky vectors as

$$(ai|bj) = \sum_{J}^{M} L_{ai}^{J} L_{bj}^{J} \,. \tag{A34}$$

We now define the following matrices

$$S_{ai,bj} = (ai|bj)^2 \,, \tag{A35}$$

$$T_{ai,bj} = (ai|bj)(aj|bi) \,, \tag{A36}$$

which are the result of the *Hadamard product* of two *positive definite* matrices, and therefore are themselves *positive definite*.

By Cholesky decomposing the $S = QQ^T$ and the $T = RR^T$ matrix we can recast the expression for the MP2 energy (A33) as follows:

$$E_2 = \sum_{\lambda I} \sum_{ai} Z_{ai}^{\lambda} Q_{ai}^{I} \sum_{bj} Z_{bj}^{\lambda} Q_{bj}^{I} - \frac{1}{2} \sum_{\lambda K} \sum_{ai} Z_{ai}^{\lambda} R_{ai}^{K} \sum_{bj} Z_{bj}^{\lambda} R_{bj}^{K} = \sum_{\lambda I} V_{\lambda I}^2 - \frac{1}{2} \sum_{\lambda K} W_{\lambda K}^2 \,, \tag{A37}$$

where ZZ^T represents the Cholesky decomposition of the *second-rank orbital energy denominators*.

The cost in evaluating the MP2 energy from the eq. (A37) is $\sim OVM_Z(M_Q + M_T)$, where M are the number of Cholesky vectors resulting from the decomposition of the corresponding matrix. We need to bear in mind that during the decompositions, an additional cost comes from computing some of the integrals, which amounts to $\sim OV(M_Q + M_T)M$. The latter is indeed the bottleneck of the method since $M_Z \ll M$. In fact, the Cholesky decomposition of the orbital energy denominators is considered to be *size-intensive* because the number of vectors only depends on the range of orbital energies and not on the number of orbitals.

The costs of this new formulation are to be confronted to the $\sim O^2V^2M$ scaling of the usual Cholesky procedure. Likely, $(M_Q + M_T) \approx M$, which implies that the new formulation is advantageous whenever $(M_Q + M_T) \leq OV$, a condition usually fulfilled.

It is important to notice that the only needed quantities are the MO transformed vectors **L** (obtained from the AO-vectors at a cost scaling as $\sim O(N + V)NM$). The **S** and **T** matrices, in fact, can be decomposed by computing the diagonal elements and the required columns along the decomposition from the vectors **L** only.

Finally, it is worth noticing that this new approach will reduce the scaling of the Cholesky-MP2 from a *fifth-order* to a *fourth-order* method, making it advantageous over the RI-MP2.

A.4 On the Γ−derivative of Cholesky vectors

December 2005

The Cholesky resolution of a diagonal element of the integral matrix is

$$(\mu\nu|\mu\nu) = \sum_{J}^{M} L^{J}_{\mu\nu} L^{J}_{\mu\nu} = \sum_{J}^{M_{\mu\nu}} L^{J}_{\mu\nu} L^{J}_{\mu\nu} \,, \tag{A38}$$

where $M_{\mu\nu}$ corresponds to the total number of non-zero elements in the vector $L_{\mu\nu}$. Taking the n^{th}-order derivative of the two expressions we obtain:

$$(\mu\nu|\mu\nu)^{(n)} = \sum_{J}^{M_{\mu\nu}} \sum_{k=0}^{n} \binom{n}{k} L^{J(n-k)}_{\mu\nu} L^{J(k)}_{\mu\nu} = \sum_{J}^{M_{\mu\nu}} \sum_{k=1}^{(n-1)} \binom{n}{k} L^{J(n-k)}_{\mu\nu} L^{J(k)}_{\mu\nu} + 2 L^{J}_{\mu\nu} L^{J(n)}_{\mu\nu} \,. \tag{A39}$$

This equation reduces to an identity if the derivatives of the Cholesky vectors are defined in the following recursive manner:

$$L^{J(n)}_{\mu\nu} = (2M_{\mu\nu})^{-1} \Gamma^{J}_{\mu\nu} \;\; [(\mu\nu|\mu\nu)^{(n)} - \sum_{J}^{M_{\mu\nu}} \sum_{k=1}^{(n-1)} \binom{n}{k} L^{J(n-k)}_{\mu\nu} L^{J(k)}_{\mu\nu}] \,, \tag{A40}$$

with $\Gamma^{J}_{\mu\nu} = (L^{J}_{\mu\nu})^{-1}$.

Even if this result is almost trivial, it represents the key point of the present formulation of analytical derivatives of Cholesky vectors. Once again, this is only a formal definition for the derivatives of the Cholesky vectors, obtained as a possible solution of the differential equation (A39) and thus it is in general different from the differentiation of the analytical form of the vector itself.

We can however give a closer look at this choice for the analytical expression of the derivative of Cholesky vectors and show that it is indeed the true derivative *for a particular choice of the vectors*.

In the hypothesis that the number of non-zero elements does not change in going from R to $(R + \Delta R)$, we obtain:

$$(\mu\nu|\mu\nu)_{R} = \sum_{J}^{M_{\mu\nu}} [L^{J}_{\mu\nu}(R)]^2 \,, \tag{A41}$$

$$(\mu\nu|\mu\nu)_{R+\Delta R} = \sum_{J}^{M_{\mu\nu}} [L^{J}_{\mu\nu}(R+\Delta R)]^2 \,, \tag{A42}$$

which gives *as a possible solution*:

$$|L^{J}_{\mu\nu}(R)| = M^{-1/2}_{\mu\nu} (\mu\nu|\mu\nu)^{1/2}_{R} \,, \tag{A43}$$

$$|L^{J}_{\mu\nu}(R+\Delta R)| = M^{-1/2}_{\mu\nu} (\mu\nu|\mu\nu)^{1/2}_{R+\Delta R} \,. \tag{A44}$$

It is clear that the vectors defined by Eq. (A43) are not the result of the decomposition (their absolute values are for instance independent of J). The vector elements defined by Eq. (A43) represent indeed an *effective Cholesky vectors elements*, namely

$$\tilde{L}^J_{\mu\nu} = \pm M^{-1/2}_{\mu\nu} \sqrt{\sum_J^{M_{\mu\nu}} (L^J_{\mu\nu})^2} \,, \tag{A45}$$

where the dependence on J, although redundant, has been introduced because the sign given to the expression can be chosen according to the one of the true vector element $L^J_{\mu\nu}$.

Applying the general definition of first derivative we get:

$$\begin{aligned}
\frac{d\tilde{L}^J_{\mu\nu}}{dR} &= \lim_{\Delta R\to 0} M^{-1/2}_{\mu\nu} \frac{(\mu\nu|\mu\nu)^{1/2}_{R+\Delta R} - (\mu\nu|\mu\nu)^{1/2}_R}{\Delta R} \\
&= M^{-1/2}_{\mu\nu} \lim_{\Delta R\to 0} \frac{(\mu\nu|\mu\nu)^{1/4}_{R+\Delta R} - (\mu\nu|\mu\nu)^{1/4}_R}{\Delta R} [(\mu\nu|\mu\nu)^{1/4}_{R+\Delta R} + (\mu\nu|\mu\nu)^{1/4}_R] \\
&= M^{-1/2}_{\mu\nu} [2(\mu\nu|\mu\nu)^{1/4}_R \frac{d}{dR} (\mu\nu|\mu\nu)^{1/4}] \\
&= M^{-1/2}_{\mu\nu} [2^{-1} (\mu\nu|\mu\nu)^{-1/2} (\mu\nu|\mu\nu)^{(1)}] \\
&= (2M_{\mu\nu})^{-1} \Gamma^J_{\mu\nu} (\mu\nu|\mu\nu)^{(1)}
\end{aligned} \tag{A46}$$

We have shown that our choice for the first-derivative of the vectors elements would be exact if we choose to represent the vectors by the effective quantity (A45). This quantity would represent exactly the diagonal part of the original integral matrix, but the error on the off-diagonal part would be in general larger than if the true vector elements were employed. We can compute the value of an off-diagonal integral obtained by using the effective Cholesky vector elements as

$$|(\mu\nu|\kappa\lambda)| = \sum_J^M |\tilde{L}^J_{\mu\nu} \tilde{L}^J_{\kappa\lambda}| = \frac{M}{\sqrt{M_{\mu\nu} M_{\kappa\lambda}}} \sqrt{(\mu\nu|\mu\nu)(\kappa\lambda|\kappa\lambda)} \leq \sqrt{(\mu\nu|\mu\nu)(\kappa\lambda|\kappa\lambda)} \tag{A47}$$

The inequality contained in eq.(A47) holds because $M = min(M_{\mu\nu}, M_{\kappa\lambda})$. The important result here is that, even if computed by means of the effective vector elements, the integral matrix is still correctly represented as a positive definite quantity. A closer look to eq.(A47) shows that the use of such an effective vector elements produces an *upper bound* to the off-diagonal integrals obtained by the initial true vector elements. On the other hand, we know that the latter produce a *lower bound* to the exact integrals. We can conclude that the use of the effective Cholesky vector elements, although producing larger error on each contribution, might still give a sensible representation of the true integral matrix.

We return now to the formulation of the derivatives as introduced in eq.(A40). It is important to notice at this point that our definition cannot be used to define the derivatives for those elements where $L^J_{\mu\nu} = 0$. As we can see from the second equality in Eq. (A39) the $n^{th}-$order derivatives of such elements (if exists and is finite), will never contribute to the diagonal integrals (the same as it is at the zeroth-order). This implies that the previous approach cannot be used to define uniquely the derivatives of the elements of the vectors which are zero at the zeroth-order. Those are all undefined as long as the diagonals only are concerned.

The generalization of the concept of derivatives of the Cholesky vectors must then go through the use of the information in the off-diagonal part of the integral matrix *as well*. The Cholesky decomposition method ensures only that *there exist at least one vector $L_{\alpha\beta}$ whose M elements are all non-zero*, otherwise the rank of the matrix would be smaller than M. Nothing is known about the number of zero-elements in the other vectors, except that each of them can contain at most $(M-1)$ zeros.

We will now show how it is possibile to devise a method able to build an analytical continuation of the definition (A40) and how this fact can be used to satisfy a larger number of physical conditions. To simplify the notation, we shall consider the first-order derivatives only. For an off-diagonal integral we have:

$$(\mu\nu|\kappa\lambda)^{(1)} = \sum_J^M L^{J(1)}_{\mu\nu} L^J_{\kappa\lambda} + L^J_{\mu\nu} L^{J(1)}_{\kappa\lambda} \ . \tag{A48}$$

For a Cholesky decomposition where all the zeroth-order vector elements are non zero, we can define the quantity:

$$< (\mu\nu|\kappa\lambda)^{(1)} >= [(2M_{\mu\nu})^{-1} \Omega^{\kappa\lambda}_{\mu\nu} (\mu\nu|\mu\nu)^{(1)} + (2M_{\kappa\lambda})^{-1} \Omega^{\mu\nu}_{\kappa\lambda} (\kappa\lambda|\kappa\lambda)^{(1)} \ ,] \tag{A49}$$

where the *first-order coupling coefficients* have been defined as:

$$\Omega^{\mu\nu}_{\kappa\lambda} = \sum_J^M \frac{L^J_{\mu\nu}}{L^J_{\kappa\lambda}} = \sum_J^M L^J_{\mu\nu} \Gamma^J_{\kappa\lambda} \qquad (L^J_{\kappa\lambda} \neq 0) \ . \tag{A50}$$

The relation (A49), even if not of great use for practical implementations, due to the need to compute ($\sim n^4 M$) and store ($\sim n^4$) the first order coupling coefficients, is of interest because it provides through Eq. (A50) yet another meaning for the Cholesky vectors elements as *interpolation coefficients* between the diagonal integral first derivatives and the off-diagonal ones. For a decomposition where there are zeros among the vectors elements, the following residual derivative can be defined:

$$R^{\kappa\lambda}_{\mu\nu} = (\mu\nu|\kappa\lambda)^{(1)} - < (\mu\nu|\kappa\lambda)^{(1)} >= \sum_J^{z_{\mu\nu}} L^J_{\kappa\lambda} L^{J(1)}_{\mu\nu} + \sum_J^{z_{\kappa\lambda}} L^J_{\mu\nu} L^{J(1)}_{\kappa\lambda} \ , \tag{A51}$$

where the summation indices run now only over the $z_{\mu\nu}$ and $z_{\kappa\lambda}$ vector elements, thus the zero elements in $L_{\mu\nu}$ and $L_{\kappa\lambda}$ respectively. It is important to notice here the fact that the residuals are in general non-zero for two reasons: one is that the definition of derivatives we have given is not exact, and another is that the derivatives of the zero-elements are not defined at all. What we do here, is indeed to determine the latter by *imposing the exact representation of the maximum possible number of off-diagonal integrals*. We shall see at this point how these residual derivatives can be used to achieve the analytical prolongation of the Cholesky vectors derivatives.

Let's imagine that we have ordered our vectors in the following way: the first vector $L_{\alpha\beta}$ is the one (and there is one) having all elements different from zero. The second vector has only one zero element, namely J^*, the third has only two (J_1 and J_2), and so on so forth until the vector having $(M-1)$ zeros. At this moment, we consider the case where there exist no more than one vector having the same number of zeros. As we shall see, the generalization of the method to the case where this restriction is not present is straightforward, provided that a particular condition is fulfilled.

For the derivative of the second vector $L_{\gamma\delta}$, we can write

$$R^{\gamma\delta}_{\alpha\beta} = L^{J^*}_{\alpha\beta} L^{J^*(1)}_{\gamma\delta} \tag{A52}$$

which provides immediately the analytical prolongation of the first derivative of the vector $L_{\gamma\delta}$ for the zero element J^*. We can now use the first two vectors to obtain the analytical prolongation for the third vector $L_{\sigma\tau}$ for the elements J_1 and J_2

$$R^{\sigma\tau}_{\alpha\beta} = L^{J_1}_{\alpha\beta} L^{J_1(1)}_{\sigma\tau} + L^{J_2}_{\alpha\beta} L^{J_2(1)}_{\sigma\tau} \,, \tag{A53}$$

$$R^{\sigma\tau *}_{\gamma\delta} = R^{\sigma\tau}_{\gamma\delta} - L^{J^*}_{\sigma\tau} L^{J^*(1)}_{\gamma\delta} = L^{J_1}_{\gamma\delta} L^{J_1(1)}_{\sigma\tau} + L^{J_2}_{\gamma\delta} L^{J_2(1)}_{\sigma\tau} \,. \tag{A54}$$

It is now clear how to proceed further: for each vector $L_{\gamma\delta}$ we need to set up the following linear equation system for each of the $(M-1)$ vectors containing zeros

$$R^{\kappa\lambda(i)*}_{\mu\nu} = \sum_{J}^{z_{\mu\nu}} L^{J}_{\kappa\lambda(i)} L^{J(1)}_{\mu\nu} \qquad i = 1, \ldots z_{\mu\nu} \,, \tag{A55}$$

each of these linear systems being of increasing dimension, equal to the number of zeros $z_{\mu\nu}$ in the corresponding vector $L_{\mu\nu}$. The residual derivatives must be updated before each vector is treated. This means that the order of solving the equation is fixed. On the other hand, this gives an easy way to generalize the procedure when there is more than one vector containing the same number of zeros. Once the first $(M-1)$ linear equation systems have been solved, we have at our disposal at least M vectors for which all derivatives are defined. The procedure can now start again, with the advantage that the order is no longer

relevant, as long as a subset of off-diagonal integrals obtained by the $(M-1)$ analytically prolounged vectors is used. *Mutatis mutandis*, this virtual procedure of constructing derivatives of the Cholesky vectors can be generalized to define the n^{th}-order derivative, but not any attempt will be done here in order to prove it explicitly.

The virtual construction devised previously cannot be used without the fulfillment of a particular condition, as shown by the following example: if we consider the case where there is only one vector with all elements different from zero and the next vector contains already two or more zeros, the definition of derivative for those elements cannot be determined uniquely. In order to state this condition in a more general way, we need some more terminology. Vectors having the same number of zero elements will be said to belong to the same *reduced set*[5]. By definition then the first reduced set will contain vectors that have no zeros at all. All the other vectors will contain a certain number of zeros according to which reduced set they belong to, and they will be called for instance 1-vector, 2-vector, ... z-vector, respectively. In other words, going from the first to the last reduced set, the length of the vectors gets shorter and shorter as the number of zeros z_k increases. Moreover, *all zero elements in a given reduced set will remain zero in each of the subsequent sets*. If R is the total number of reduced sets present in our decomposition, and h_i is the dimension (number of vectors) of the $i-$th reduced set, the *condition for $\zeta-$derivability* reads

$$z_{(k+1)} \leq \sum_{i}^{k} h_i \qquad (\forall k : 1 \leq k < R) \; . \tag{A56}$$

Finally, let's summarize the results with the following general

Definition: for a given set of non-zero elements of Cholesky vectors L_a^J obtained by decomposing the matrix H_{ab}, we define as $\Gamma-$*derivative* of order n of these elements the quantity

$$\hat{\Gamma}^{(n)} L_a^J \equiv (2M_a)^{-1} \Gamma_a^J \;\; [H_{aa}^{(n)} - \sum_{J}^{M_a} \sum_{k=1}^{(n-1)} \binom{n}{k} L_a^{J(n-k)} L_a^{J(k)}] \; , \tag{A57}$$

where M_a is the number of non-zero elements of the vector L_a, thus its length, and $\Gamma_a^J = (L_a^J)^{-1}$. Once the $\Gamma-$derivatives have been determined, an analytical prolongation of this derivative to the elements of the vectors equal to zero can be obtained, provided that the condition (A56) is fulfilled. The method to construct it goes through the solution for each vector L_a of a linear equation system of dimension equal to the number of zero elements z_a contained in it. The order in which the vectors are treated is relevant, namely the one with the smallest

[5] This kind of partitioning represents more precisely the definition of *horizontal reduced sets*. There is a corresponding definition for the *vertical reduced sets* of particular use for the implementation of the reading from disk of the Cholesky vectors.

number of zeros must be treated first and so on till the one containing the largest number of zeros. The maximun dimension of the linear equation system is

$$\zeta = Max_a \, z_a \, , \tag{A58}$$

therefore we choose to call the derivative defined by combining the definition (A57) and its analytical prolongation as $\Gamma^{(n)}_{\zeta}-$*derivative* of the Cholesky vectors.

A.5 On some integral approximations

October 2006

The two-electron integrals are defined as

$$(\mu\nu|\lambda\sigma) \; = \; \int d\mathbf{r_1} d\mathbf{r_2} \; \mu(\mathbf{r_1})\nu(\mathbf{r_1}) \; \frac{1}{r_{12}} \; \lambda(\mathbf{r_2})\sigma(\mathbf{r_2}) \, , \tag{A59}$$

where $\chi_{\mu\nu} = \mu(\mathbf{r_1})\nu(\mathbf{r_1})$ are the orbital distributions. For a molecule of fixed size, increasing the number of basis functions per atom, n, leads to the $O(n^4)$ scaling of the number of nonnegligible integrals. A workaround for this problem is that of approximating the orbital distributions with an expansion over a set of auxiliary Gaussian basis functions

$$\chi_{\mu\nu} \; = \; \mu(\mathbf{r_1})\nu(\mathbf{r_1}) \; = \; \sum_K \; C^K_{\mu\nu} \; \varphi_K \, , \tag{A60}$$

and then minimize the "distance" (in Hilbert space) between the fitted distributions and the actual distributions, defined through a given metric, usually chosen as the Coulomb metric $\frac{1}{r_{12}}$. Given an auxiliary basis φ, we define the following (non-negative) functional

$$f_{\mu\nu} \; = \; (\chi_{\mu\nu} - \sum_K \; C^K_{\mu\nu} \; \varphi_K \mid \chi_{\mu\nu} - \sum_K \; C^K_{\mu\nu} \; \varphi_K) \, , \tag{A61}$$

whose minimization leads to the set of linear equations

$$\sum_K (\varphi_K|\varphi_J) C^K_{\mu\nu} \; = \; (\mu\nu|\varphi_J) \, . \tag{A62}$$

The value of the functional in correspondence of the solution of eq. (A62) is given by the following expression

$$f^{min}_{\mu\nu} \; = \; (\chi_{\mu\nu}|\chi_{\mu\nu}) - \sum_{JK} C^J_{\mu\nu} (\varphi_J|\varphi_K) C^K_{\mu\nu} \, . \tag{A63}$$

We now assert that the same variational problem possesses a solution corresponding to the global minimum of the functional (A61) if the auxiliary basis is chosen to be *orthonormal* (in the Coulomb metric) and the expansion coefficients are simply the Cholesky vectors from the decomposition of the integral matrix. For instance, we can use the following Gram-Schmidt orthogonalization procedure (see also *Beebe and Linderberg*)[6].

$$\varphi_J = \frac{\Psi_J}{(\Psi_J|\Psi_J)^{1/2}} \;:\; \Psi_J = \{1 - \sum_K^{J-1} \hat{P}_K\}\, \chi_J \quad \textit{and} \quad C_K^J = L_K^J \,, \tag{A64}$$

where L are the Cholesky vectors from the decomposition of the integral matrix, namely $(\mu\nu|\lambda\sigma) = \sum_J^M L_{\mu\nu}^J L_{\lambda\sigma}^J$, and $\hat{P}_K\ \chi_J = (\varphi_K|\chi_J)\ \varphi_J$ projects χ_J into φ_K according to the chosen metric.

The number of vectors M is called the effective rank of the integral matrix and it is smaller than the actual dimension $N(N+1)/2$ of this matrix. The number N is equal to the number of unique distributions $\mu\nu$, thus it is itself equal to $\mathcal{N}(\mathcal{N}+1)/2$, where $\mathcal{N}$ is this time the number of AO-basis functions employed. In a typical calculation, the value of M is of the same order of magnitude of $\mathcal{N}$, let's say 3-8 times as much.

By using the definition of the Cholesky vectors and the fact that the auxiliary basis defined by Eq. (A64) is *orthonormal in the given metric* $\frac{1}{r_{12}}$, we prove the assert that the functional is at its global minimum by noting that eq. (A63) returns the value of zero[7].

The previous discussion leads to the conclusion that for a given molecule, the best approximation for the two-electron integrals is obtained by using the Cholesky vectors. On the other hand, this result has been achieved by an approximate representation of the orbital distributions $\chi_{\mu\nu}$ which minimizes the functional (A61). By this token, we can assume an atom-based approach to the problem of simplifying the evaluation of two-electron integrals.

As first step, we decompose the atomic integral matrices in order to get the Cholesky vectors and then the corresponding atomic auxiliary basis ϕ. Since the latter best represent the orbital distributions on the corresponding atom, we can compute a certain subset of two-electron integrals of the molecule by the following means:

- 1-center integrals, $(AA|AA)$

$$(\mu_A\nu_A|\lambda_A\sigma_A) = \sum_J^{M_A} L_{\mu_A\nu_A}^J L_{\lambda_A\sigma_A}^J \,; \tag{A65}$$

[6] N. H. F. Beebe and J. Linderberg, *Int. J. Quant. Chem.*, **12**, 683 (1977).

[7] Since the matrix decomposition is carried out only to a given numerical accuracy, we should rather say that the value of the functional differs from zero less than the chosen decomposition threshold

- 2-center "Coulomb" integrals, $(AA|BB)$

$$(\mu_A\nu_A|\lambda_B\sigma_B) = \sum_{JK} L^J_{\mu_A\nu_A}(\phi^J_A|\phi^K_B)L^K_{\lambda_B\sigma_B} \,. \tag{A66}$$

In order to proceed further, for the representation of the remaining two-electron integrals we need to approximate the *2-center orbital distributions*. In this case, the minimization of the functional of Eq. (A61) is achieved by choosing as expansion coefficients the Cholesky vectors *R decomposing each "Exchange" integral matrix* $(AB|AB)$. The auxiliary basis γ is similarly obtained by a Gram-Schmidt orthogonalization procedure as in Eq. (A64). The remaining set of two-electron integrals is then constructed in the following way:

- 2-center "Exchange" integrals, $(AB|AB)$

$$(\mu_A\nu_B|\lambda_A\sigma_B) = \sum_{J}^{M_{AB}} R^J_{\mu_A\nu_B}R^J_{\lambda_A\sigma_B} \,; \tag{A67}$$

- 3-center "Coulomb" integrals, $(AB|CC)$

$$(\mu_A\nu_B|\lambda_C\sigma_C) = \sum_{JK} R^J_{\mu_A\nu_B}(\gamma^J_{AB}|\phi^K_C)L^K_{\lambda_C\sigma_C} \,; \tag{A68}$$

- 4-center integrals, $(AB|CD)$

$$(\mu_A\nu_B|\lambda_C\sigma_D) = \sum_{JK} R^J_{\mu_A\nu_B}(\gamma^J_{AB}|\gamma^K_{CD})R^K_{\lambda_C\sigma_D} \,; \tag{A69}$$

The set of "atomic" Cholesky vectors can be computed once and for all and stored in a library. Moreover, the $(\phi_A|\phi_B)$ matrix can be constructed from the list of integrals over the original atomic distribution $\chi_{\mu_A\nu_A}$, which is known *a priori*. The decomposition of each of the $(AB|AB)$ matrices can be costly. However, these matrices are of short-range type and for this reason easy to handle. The same goes for the matrices $(\gamma_{AB}|\phi_C)$ and $(\gamma_{AB}|\gamma_{CD})$.

A drawback of this formulation of the density fitting is the fact that the "interatomic part" of the auxiliary basis, namely γ_{AB}, *is dependent on the molecular geometry*. Unless a complete auxiliary basis is used, thus it is not possible to correctly define the derivatives of the expressions (A67-A69). A workaround to this problem could be to assume that the γ_{AB} basis can be constructed as a linear combination of the *full set of atomic auxiliary basis* ϕ. In this way, we define an auxiliary basis to be used in order to compute the two-electron integrals of the type (A67-A69) but this time by using the general expression

$$(\mu_A\nu_B|\lambda_C\sigma_D) = \sum_{JK} C^J_{\mu_A\nu_B}(\phi^J|\phi^K)C^K_{\lambda_C\sigma_D} \,, \tag{A70}$$

where the coefficients $C^J_{\mu_A \nu_B}$ can be computed through the solution of linear equations of the type (A62). One could also think of using the *Poisson trick* for this particular density fitting, namely using instead of ϕ the basis formed by their Laplacians.

We conclude by saying that the minimization of the functional of Eq. (A61) in the full basis set for the molecule produces the well-known result stating that the best approximation to the two-electron integrals is given by their Cholesky representation. On the other hand, if we proceed with first an atomic decomposition and then decompose the 2-center "Exchange" integral matrix, we can derive a new set of formulas for the approximate evaluation of the two-electron intgrals that can be of great interest. Finally, a formulation of the energy gradients is possible if the auxiliary basis obtained from the atomic Cholesky decompositions is used in order to expand the 2-center orbital distributions in the spirit of a density fitting type approach.

References

[1] H. Everett. *Rev. Mod. Phys.*, **29**, 454, (1957).

[2] *European Summerschool in Quantum Chemistry,* Lecture Notes. Björn O. Roos and Per-Olof Widmark, Eds., Lund University, (2007).

[3] E. A. Hylleraas. *Z. Phys.*, **54**, 347, (1929).

[4] W. Kutzelnigg and W. Klopper. *J. Chem. Phys.*, **94**, 1985, (1991).

[5] H. J. Werner and F. R. Manby. *J. Chem. Phys.*, **124**, 54114, (2006).

[6] B. O. Roos. The complete active space self-consistent field method and its applications in electronic structure calculations. In K. P. Lawley, editor, *Advances in Chemical Physics; Ab Initio Methods in Quantum Chemistry - II*, chapter 69, page 399. John Wiley & Sons Ltd., Chichester, England, (1987).

[7] L. Gagliardi and B. O. Roos. *Nature*, **433**, 848, (2005).

[8] P. Hohenberg and W. Kohn. *Phys. Rev.*, **136**, B864, (1964).

[9] *Density-Functional Theory of Atoms and Molecules*. Oxford U. Press, New York, (1989).

[10] W. Kohn and L. J. Sham. *Phys. Rev.*, **140**, A1133, (1965).

[11] J. A. Pople and W. J. Hehre. *J. Comput. Phys.*, **27**, 161, (1978).

[12] S. Obara and A. Saika. *J. Chem. Phys.*, **84**, 3963, (1986).

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

[14] L. E. McMurchie and E. R. Davidson. *J. Comput. Phys.*, **26**, 218, (1978).

[15] S. F. Boys. *Proc. Roy. Soc.*, **A200**, 542, (1950).

[16] J. Rys, M. Dupuis, and H. F. King. *J. Comput. Chem.*, **4**, 154, (1983).

[17] P. M. W. Gill and J. A. Pople. *Int. J. Quantum Chem.*, **40**, 753, (1991).

[18] S. Ten-no. *Chem. Phys. Lett.*, **211**, 259, (1993).

[19] K. Ishida. *Int. J. Quantum Chem.*, **59**, 209, (1996).

[20] *Fortran Numerical Recipies, 2nd edn.* Cambridge University Press, Cambridge, (1996).

[21] H. Larsen, J. Olsen, P. Jørgensen, and T. Helgaker. *J. Chem. Phys.*, **115**, 9685, (2001).

[22] Y. Shao, C. Saravanan, M. Head-Gordon, and C. A. White. *J. Chem. Phys.*, **118**, 6144, (2003).

[23] C. Ochsenfeld, C. A. White, and M. Head-Gordon. *J. Chem. Phys.*, **109**, 1663, (1998).

[24] M. Challacombe, E. Schwegler, and J. Almlöf. *J. Chem. Phys.*, **104**, 4685, (1996).

[25] N. H. F. Beebe and J. Linderberg. *Int. J. Quantum Chem.*, **7**, 683, (1977).

[26] H. Koch, A. Sánchez de Merás, and T. B. Pedersen. *J. Chem. Phys.*, **118**, 9481, (2003).

[27] J. L. Whitten. *J. Chem. Phys.*, **58**, 4496, (1973).

[28] O. Vahtras, J. Almlöf, and M. Feyereisen. *Chem. Phys. Lett.*, **213**, 514, (1993).

[29] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin. *J. Chem. Phys.*, **71**, 3396, (1979).

[30] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin. *J. Chem. Phys.*, **71**, 4993, (1979).

[31] M. Feyereisen, G. Fitzgerald, and A. Komornicki. *Chem. Phys. Lett.*, **208**, 359, (1993).

[32] K. Eichkorn, O. Treutler, H. Öm, M. Häser, and R. Ahlrichs. *Chem. Phys. Lett.*, **240**, 283, (1995).

[33] D. E. Bernholdt and R. J. Harrison. *J. Chem. Phys.*, **109**, 1593, (1998).

[34] F. Weigend, A. Köhn, and C. Hättig. *J. Chem. Phys.*, **116**, 3175, (2002).

[35] F. Weigend. *Phys. Chem. Chem. Phys.*, **4**, 4285, (2002).

[36] A. P. Rendell and T. J. Lee. *J. Chem. Phys.*, **101**, 400, (1994).

[37] C. Hättig and F. Weigend. *J. Chem. Phys.*, **113**, 5154, (2000).

[38] F. R. Manby, P. J. Knowles, and A. W. Lloyd. *J. Chem. Phys.*, **115**, 9144, (2001).

[39] H. J. Werner, F. R. Manby, and P. J. Knowles. *J. Chem. Phys.*, **118**, 8149, (2003).

[40] Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon. *Proc. Natl. Acad. Sci. USA*, **102**, 6692, (2005).

[41] A. Sodt, J. E. Subotnik, and M. Head-Gordon. *J. Chem. Phys.*, **125**, 194109, (2006).

[42] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. Wiley, Chichester, (2000).

[43] I. Røeggen and E. Wisløff-Nielsen. *Chem. Phys. Lett.*, **132**, 154, (1986).

[44] D. W. O'Neal and J. Simons. *Int. J. Quantum Chem.*, **36**, 673, (1989).

[45] S. Wilson. *Comput. Phys. Commun.*, **58**, 71, (1990).

[46] T. B. Pedersen, A. M. J. Sánchez de Merás, and H. Koch. *J. Chem. Phys.*, **120**, 8887, (2004).

[47] I. García Cuesta, T. B. Pedersen, H. Koch, and A. M. J. Sánchez de Merás. *Chem. Phys. Lett.*, **390**, 170, (2004).

[48] T. B. Pedersen, H. Koch, L. Boman, and A. M. J. Sánchez de Merás. *Chem. Phys. Lett.*, **393**, 319, (2004).

[49] I. García Cuesta, T. B. Pedersen, H. Koch, and A. Sánchez de Merás. *ChemPhysChem*, **7**, 2503, (2006).

[50] B. Fernández, T. B. Pedersen, A. Sánchez de Merás, and H. Koch. *Chem. Phys. Lett.*, **441**, 332, (2007).

[51] I. Røeggen. *J. Chem. Phys.*, **124**, 184502, (2006).

[52] F. Aquilante, T. B. Pedersen, H. Koch, and A. Sánchez de Merás. *J. Chem. Phys.*, **125**, 174101, (2006).

[53] A. Öhrn and F. Aquilante. *Phys. Chem. Chem. Phys.*, **9**, 470, (2007).

[54] *MOLCAS 6*, (2005). University of Lund, Sweden (see http://www.teokem.lu.se/molcas/).

[55] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK Users' Guide, release 2.0*. SIAM, Philadelphia, (1994).

[56] T. H. Dunning. *J. Chem. Phys.*, **90**, 1007, (1989).

[57] K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski. *J. Phys. Chem.*, **94**, 5483, (1990).

[58] K. Andersson, P.-Å. Malmqvist, and B. O. Roos. *J. Chem. Phys.*, **96**, 1218, (1992).

[59] T. B. Pedersen, A. M. J. Sánchez de Merás, and H. Koch. *J. Chem. Phys.*, **120**, 8887, (2004).

[60] H. Koch and A. Sánchez de Merás. *J. Chem. Phys.*, **113**, 508, (2000).

[61] P. Pulay. *Chem. Phys. Lett.*, **100**, 151, (1983).

[62] G. Scuseria and P. Y. Ayala. *J. Chem. Phys.*, **111**, 8330, (1999).

[63] D. S. Lambrecht and C. Ochsenfeld. *J. Chem. Phys.*, **123**, 184101, (2005).

[64] W. Kohn. *Nobel lecture at the Centennial APS meeting, Atlanta Georgia*, (1999).

[65] R. Polly, H. J. Werner, F. R. Manby, and P. J. Knowles. *Mol. Phys.*, **102**, 2311, (2004).

[66] W. Liang, H. Shao, C. Ochsenfeld, A. T. Bell, and M. Head-Gordon. *Chem. Phys. Lett.*, **358**, 43, (2002).

[67] P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee, and M. Head-Gordon. *J. Phys. Chem., A*, **102**, 2215, (1998).

[68] S. F. Boys. *Rev. Mod. Phys.*, **32**, 296, (1960).

[69] J. M. Foster and S. F. Boys. *Rev. Mod. Phys.*, **32**, 300, (1960).

[70] J. W. Boughton and P. Pulay. *J. Comput. Chem.*, **14**, 736, (1993).

[71] S. Saebo and P. Pulay. *Annu. Rev. Phys. Chem.*, **44**, 213, (1993).

[72] E. Kapuy, F. Bogár, and C. Kozmutza. *J. Mol. Struct. (Theochem)*, **297**, 365, (1993).

[73] E. Kapuy, F. Bogár, and E. Tfirst. *Int. J. Quantum Chem.*, **52**, 127, (1994).

[74] C. Hampel and H. J. Werner. *J. Chem. Phys.*, **104**, 6286, (1996).

[75] W. Förner, R. Knab, J. Cizek, and J. Ladik. *J. Chem. Phys.*, **106**, 10248, (1997).

[76] J. Pipek and F. Bogár. *Topics in Current Chemistry*. Springer, (1999).

[77] M. Schütz, G. Hetzer, and H. J. Werner. *J. Chem. Phys.*, **111**, 5691, (1999).

[78] M. Schütz and H. J. Werner. *J. Chem. Phys.*, **114**, 661, (2001).

[79] C. Edmiston and K. Ruedenberg. *Rev. Mod. Phys.*, **35**, 457, (1963).

[80] J. Pipek and P. G. Mezey. *J. Chem. Phys.*, **90**, 4916, (1989).

[81] J. M. Millam and G. E. Scuseria. *J. Chem. Phys.*, **106**, 5569, (1997).

[82] R. McWeeny. *Rev. Mod. Phys.*, **32**, 335, (1960).

[83] M. Challacombe. *J. Chem. Phys.*, **110**, 2332, (1999).

[84] T. Helgaker, H. Larsen, J. Olsen, and P. Jørgensen. *Chem. Phys. Lett.*, **327**, 397, (2000).

[85] J. E. Subotnik and A. D. Dutoi and M. Head-Gordon. *J. Chem. Phys.*, **123**, 114108, (2005).

[86] P. Pulay. *Mol. Phys.*, **17**, 197, (1969).

[87] *Geometrical Derivatives of Energy Surfaces and Molecular Properties*. P. Jørgensen and J. Simons, eds., Reidel, Dordrecht, Netherlands, (1986).

[88] P. Jørgensen and T. U. Helgaker. *J. Chem. Phys.*, **89**, 3654, (1988).

[89] M. Schütz, H.-J. Werner, R. Lindh, and F. R. Manby. *J. Chem. Phys.*, **121**, 737, (2004).

[90] P. Neogrády, M. Pitonák, and M. Urban. *Mol. Phys.*, **103**, 2141, (2005).

[91] L. Adamowicz and R. J. Bartlett. *J. Chem. Phys.*, **86**, 6314, (1987).

[92] P. Salek, O. Vahatras, T. Helgaker, and H. Ågren. *J. Chem. Phys.*, **117**, 9630, (2002).

Acknowledgements

This is by far the most difficult task of my PhD studies. How many persons have made a contribution to my life during the long and unforgettable time spent here in Lund and at the department of Theoretical Chemistry? And what is the weight of their contribution after all? As I can't find any suitable answer to these questions, I will assume that all people I have met here were equally good to me at the zeroth-order approximation. Some of them, however, contributed with higher-order corrections, as summarized in what follows:

First, I would like to express my gratitude to **Björn**, for giving me the chance to start a PhD at Lund University and for being like a father for all of us here at the department. **Thomas**, a.k.a. "the Cholesky master": a proper thank can hardly be expressed in a closed form. **Roland**, who entered the Cholesky match during the overtime, but scored all penalty kicks! **Per-Åke**, a great teacher of quantum mechanics and also the fourth Cholesky musketeer. **Valera**, the tireless code-breaker. Without his help, the Cholesky implementation would have crashed at each sunny day in Lund! **Per-Olof**, thanks to him I don't get dizzy when looking at postscript codes or LaTeX style files. **Bo** and **Gunnar** for being so tall but still very friendly! My appreciation and high regard to the department secretaries **Eva** and **Bodil**. Among other things, they were the first who tried to use Swedish as official language of our conversations. All the present and past guests and members of the department of Theoretical Chemistry, and I would like to mention here: **Asbjørn**, **Pär**, **Jimmy**, **Anders**, **Kasper**, **Jesper**, **Magnus**, **Ulf**, **Patrick**, **Michael**, **Daniel**. **Giovanni** and **Tiziana** for their friendship. **Ajitha** and **Sasha** for the examples on quantum uncertainty of life. If I missed somebody, this is for you! The Cooking team — second greatest invention of all times — *so long, so long, so long and thanks for all the fish!*. **Kambiz** and **Magda** for the nice time spent together and for the delicious food in absence of the cooking team. **Luis** for all the squash I never played. **Lubos** a.k.a π — with a little bit of luck, we will stay friends for long. My flatmates: the blues-man-carbonara-eater-dish-washer-pianist **Juraj**, "the thing is that" **Tómas**, the funny **Claudio** and the quiet **Stefano**. My friends of various shapes and sizes at the Guest-House: in particular **Princy**, **Giorgio** and **Maria Eugenia**. People from the department of Biophysical Chemistry, especially

Houman and **Olga** for the amazing parties in the good old days; **Karim** for the long winter days spent trying to explain me philosophy — I really miss those talks. The collaborators and authors of the papers. Lund University and all the foundations (VR, Fysiografen) for the financial support. **Sweden**, a wonderful country, for receiving me and making me feel at home. This is not a goodbye, but just an *Arrivederci!* **Azita**, for being my closest friend here in Lund, a very nice girl and for cutting my hair.

Now some more people that will receive acknowledgements in Italian, as I am running out of time with respect to the already postponed appointment with **Paula** for the printing of this thesis. Thanks for the great help and patience! Un grazie a **Laura Gagliardi** per aver contribuito alla mia scelta di rimanere qui a Lund; ringrazio **Renata** che mi sta aiutando per il mio trasferimento a Ginevra. All'università di Napoli, un grazie ad **Enzo Barone** e **Maurizio Cossi**, che mi hanno introdotto alla chimica quantistica e intrapreso la collaborazione con Lund. Un saluto anche a **Roberto**, **Nadia**, e a **Gaetano F.**. Quelli di **Sartorio** e **Paduano**. In particolare **Gaetano M.** e **Mauro** de "Il Maestro". **Iberio** y **Francesc Illas** a Barcelona, per tutto quello che è stato e sarà; **Dany** por las fiestas! Mi tia **Marta** (y *Pedrito*) per avermi accolto a Barcellona con tanto amore: le parole non bastano per ringraziarti. Amici d'infanzia e adolescenza: **Luca**, **Mimmo e Maria**, **Massimo**, **Antonello**, **Vincenzo** ed il ritrovato **Antonio**. **Francesco** *'o mericano* per le tante indimenticabili vacanze. **Marco** per essere stato il mio primo grande amico qui a Lund e per tutto il tempo trascorso insieme. Un saluto anche a **Davide**, a **Teresita** e a quel personaggio che è **Nick**.

Tra le persone che ho incontrato qui a Lund, un ringraziamento speciale va certamente a **Luca** e **Kinga**. Siete dei veri amici e sarà difficile riuscire a sdebitarmi per tutto quello che avete fatto per me. Nel frattempo, comincio a farlo dicendovi *köszönöm szépen*. Ma chi più di tutti ha riempito le mie giornate qui a Lund di gioia e amore è stata certamente la mia amata **Laura**. Senza di te, tutto questo non sarebbe stato possibile. Dal profondo del mio cuore, un grazie per tutto ciò che hai fatto per me, aiutandomi a superare un duro ostacolo. Sei semplicemente da amare!

Infine, ringrazio la mia famiglia tutta. Il primo pensiero va ai miei genitori. Al mio caro **papà**, che sento sempre vicino anche se non è più qui tra noi. Alla mia dolce **mamma**, che ha superato con tanto coraggio il dolore per la mia lontananza. A lei dedico questo mio traguardo, raggiunto anche per merito suo. A **Rosetta**, per essere la migliore sorella al mondo e per tutto quello che ha fatto e fa per me. A mio cognato **Arcangelo**, per avermi trasmesso l'amore per la cultura. A mio fratello **Bernardo**, che è un punto di riferimento per tutti noi, e a sua moglie **Fortuna**. A **Sandro**, mio fratello, perchè da lui ho ereditato la passione per la musica. Infine, ai miei nipoti, che sono davvero speciali. In ordine d'età: **Pasquale**, **Katia**, **Raffaele**, **Pasqualino**, **Michela** e **Gianmarco**.

This is the last of that sort of thing.
Now I'm cleaning up and I'm moving on, going straight and choosing life.
I'm looking forward to it already.
I'm gonna be just like you. The job, the family, the [. . .] big television.
The washing machine, the car, the compact disc and electric tin opener,
good health, low cholesterol, dental insurance, mortgage, starter home, leisure
wear, luggage, three piece suite, DIY,
game shows, junk food, children, walks in the park, nine to five,
good at golf, washing the car, choice of sweaters, family Christmas, indexed
pension, tax exemption, clearing gutters,
getting by, looking ahead, the day you die.

Mark Renton (*Trainspotting*, 1996)