203
Topics in Current Chemistry

Topics in Current Chemistry
Now Also Available Electronically

For all customers with a standing order for Topics in Current Chemistry we offer the electronic form via LINK free of charge. Please contact your librarian who can receive a password for free access to the full articles by registration at:

http://link.springer.de/series/tcc/reg_form.htm

If you do not have a standing order you can nevertheless browse through the table of contents of the volumes and the abstracts of each article at:

http://link.springer.de/series/tcc

There you will also find information about the

- Editorial Board
- Aims and Scope
- Instructions for Authors

Springer

Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Correlation and Localization

Volume Editor: P. R. Surján

With contributions by
R. J. Bartlett, F. Bogár, D. L. Cooper, B. Kirtman,
W. Klopper, W. Kutzelnigg, N. H. March, P. G. Mezey,
H. Müller, J. Noga, J. Paldus, J. Pipek, M. Raimondi,
I. Røeggen, J.-Q. Sun, P. R. Surján, C. Valdemoro,
S. Vogtner

Springer

This series presents critical reviews of the present position and future trends in modern chemical research. It is addressed to all research and industrial chemists who wish to keep abreast of advances in the topics covered.

As a rule, contributions are specially commissioned. The editors and publishers will, however, always be pleased to receive suggestions and supplementary information. Papers are accepted for "Topics in Current Chemistry" in English.

In references Topics in Current Chemistry is abbreviated Top. Curr. Chem. and is cited as a journal.

Springer WWW home page: http://www.springer.de
Visit the TCC home page at http://link.springer.de/series/toc

ISSN 0340-1022
ISBN 3-540-65754-1
Springer-Verlag Berlin Heidelberg New York

Library of Congress Catalog Card Number 74-644622

Printed in Germany

Cover design: Friedhelm Steinen-Broo, Barcelona; MEDIO, Berlin
Typesetting with LATEX: PTP-Berlin, Stefan Sossna

SPIN: 10681133 02/3020 – 5 4 3 2 1 0 – Printed on acid-free paper

Volume Editor

Prof. Dr. Péter R. Surján
Department of Theoretical Chemistry
Eötvös University, Budapest
Pazmany Peter setany 1/A
H-1117 Budapest, Hungary
E-mail: surjan@para.chem.elte.hu

Editorial Board

A Tribute to Ede Kapuy

Preface

Development in science depends on several factors. Among these, the role of individual scientists is perhaps not the most important one. Science is typically a body of collective knowledge and any increase in the amount of this knowledge is certainly due to strong interaction among scientists. Even in the past, it happened quite rarely that a single person, without any aid of others, discovered something fundamental or opened a new chapter in science. Great figures of science history have, in most cases, had rather a summarizing and synthesizing role.

This is especially valid over the last few decades. On one hand, the amount of information necessary to achieve new discoveries, has increased tremendously. On the other hand, improvement of technical facilities has increased the speed of information exchange. These factors resulted in a degree of specialization in science that had never seen before. Most of us are experts and specialists rather than scientists in the classical sense. My personal feeling is that, even nowadays, there is a strong need for professionals with a broad knowledge and comprehensive mind, although they may not be competitive in the number of their publications or the sizes of their grants. Every time I have met such a person (I can count these cases on my fingers) I have become deeply influenced by his or her strong intellect.

One of the most knowledgeable quantum chemists I have ever met is certainly Professor Ede Kapuy, to whom this volume is dedicated on the occasion of his 70th birthday. Apart from being a good researcher, he is known to us as a professor with an extraordinary breadth of knowledge in all aspects of quantum chemistry, as well as in some loosely related fields such as the theory of relativity, particles and fields, etc. Before his retirement, he regularly sat in the library for hours, several times a week, read all the important journals from A to Z, and was able to memorize almost everything he had read. We often turned to him with difficult questions for which we got answers much more relevant than we could have obtained from the best database. He filtered the information and always pointed out the essence of the problem. I, personally, was never a student of his but benefited from his knowledge by many informal discussions, and his impact on my scientific thinking was almost as strong as that of my supervisors.

Ede Kapuy is not only an outstanding scientist but also a person of great general erudition. History and geography are among his hobbies. I shall never forget a conversation I had with him during our preparation for the WATOC '96 congress in Israel. His illness prevented him from joining us but he gave us advice

with pertinent details about important historical sights in various small cities some of which I had never heard of. He has himself never been to the Holy Land, however.

The scientific work of Ede Kapuy covers basically two strongly related issues. He started his carrier with studying two-electron functions (geminals) which were among the main targets of investigation from the late 1950s; he joined this project almost in statu nascendi and published more than a dozen keynote papers, some of which are still being cited in recent publications. The theory of geminals is strongly related to the field of electron localization, and Ede Kapuy soon realized that an effective account of the correlation energy emerges if the electronic wave function is written in terms of localized quantities. This understanding led to him setting up a project in which he studied many body perturbation theory in terms of localized orbitals. These are the two issues that constitute the topics of the present Festschrift Volume, which is naturally entitled "Correlation and Localization".

In organizing the volume, we felt that the best way of celebrating Kapuy's birthday was to produce a book which is really useful for a young generation of quantum chemists. Accordingly, we wanted to give a cross-section of modern quantum chemistry along the lines connected to Kapuy's life work; this is indicated by the title. Unfortunately, space limitations did not allow us to invite a larger number of authors, neither for including extensive, voluminous reviews, thus the book remained necessarily incomplete. We start with a paper by Paldus and Li giving a short introduction into modern correlation theory for small systems. Then, in an article with Kutzelnigg as the senior author, we start to discuss the intimate connection between localization and correlation. Pipek and Bogár, former students of Ede Kapuy, review versions of many-body perturbation theory based on localized molecular orbitals. The theory of geminals is our next target: a didactic survey is written by the volume editor about the basis, followed by Roggen's paper on modern extensions of geminal models. As another alternative to MO-based approaches, recent developments in ab inito valence bond theory constitute the topic of the following article. Then we turn to extended systems: Sun and Bartlett apply strict periodicity using k-space methods where Wannier functions may be invoked to describe localization, while locally perturbed inherently delocalized systems (e.g., metal surfaces) are treated by Kirtman. Density and density matrices are considered in the three following papers: Mezey discusses the properties of the molecular electron density function and some of its approximate construction schemes, Valdemoro reviews recent results on the theory of reduced density matrices, while the issue is completed by a thorough analysis by March of electron localization in density functional theory.

The volume editor hopes sincerely that this book will contribute to a better understanding of the difficult problem of correlation and localization, that it will stimulate further discussions in this subject, and that many young scientists will enter this research area. Then, the enormous work Ede Kapuy started more than 40 years ago will certainly not have been wasted.

Budapest, April 1999 Péter Surján

Contents

Contents of Volume 180

Density Functional Theory I
Functionals and Effective Potentials

Volume Editor: R. F. Nalewajski
ISBN 3-540-61091-X

Contents of Volume 181

Density Functional Theory II
Relativistic and Time Dependent Extensions

Volume Editor: R. F. Nalewajski
ISBN 3-540-61092-8

Contents of Volume 182

Density Functional Theory III
Interpretation, Atoms, Molecules and Clusters

Volume Editor: R. F. Nalewajski
ISBN 3-540-61132-0

Contents of Volume 183

Density Functional Theory IV
Theory of Chemical Reactivity

Volume Editor: R. F. Nalewajski
ISBN 3-540-61131-2

Electron Correlation in Small Molecules: Grafting CI onto CC

Josef Paldus* · Xiangzhu Li

Department of Applied Mathematics, University of Waterloo,
Waterloo, Ontario, Canada N2L 3G1.
E-mail: paldus@theochem.uwaterloo.ca, xli@cauga.uwaterloo.ca

This paper is dedicated to Professor Ede Kapuy — a long time cherished friend and colleague.

Among the post-Hartree-Fock methods, those based on the coupled cluster (CC) ansatz for the electronic wave function proved to be extremely valuable in quantum chemical computations of the molecular electronic structure, being capable of attaining chemical accuracy for many molecular properties of interest. While the widely exploited single reference (SR) singles and doubles CC method (CCSD) is remarkably efficient in handling dynamic correlation, a proper account of nondynamic correlation, which becomes essential in the presence of the quasidegeneracy, requires multireference (MR) formalism. In view of the complexity and computational demands of the available MR CC methods, it is highly desirable to design SR CCSD-type approaches that are capable of accommodating both types of correlation effects. One avenue to achieve this goal is offered by the so-called externally corrected (ec) CCSD methods, which exploit some independent source of higher than pair clusters — whose importance rises with the increasing quasidegeneracy — to correct the standard CCSD equations. In view of the complementarity of SR CC and MR configuration interaction (CI) methods in their ability to describe the dynamic and nondynamic correlation effects, a particularly suitable and affordable external source proved to be an MR CISD wave function, based on a small active or model space, leading to the so-called reduced multireference (RMR) CCSD approach. Following a brief outline of the origins and of the *status quo* of the ecCCSD and

* Also at: Department of Chemistry and Guelph-Waterloo Center for Graduate Work in Chemistry, Waterloo Campus, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; visiting at: Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 87540 Garching bei München, Germany.

Topics in Current Chemistry, Vol. 203

RMR CCSD methodologies, their performance is illustrated by a few examples, and their potential and relationship with other approaches is discussed.

Keywords: many-electron correlation problem, post-Hartree-Fock methods, coupled cluster approaches, configuration interaction, externally corrected coupled cluster methods, reduced multireference coupled cluster method

1 Introduction

Since the pioneering work of Heitler and London [124] appeared and the first quantum chemistry textbook was written by Hellmann a decade later [2], the need to explain chemical bonding phenomena from first principles became recognized by molecular physicists and chemists alike. Both the valence bond (VB) and molecular orbital (MO) methodologies proved to be of enormous value in understanding and interpreting molecular electronic spectra, various molecular properties, as well as chemical reactivity, even though quantitative predictions had to await the arrival of the digital computer.

In view of the sheer mathematical complexity that we face when trying to solve the Schrödinger equation for even the simplest of molecular systems, the main emphasis of quantum chemical methodology has been on the design of computationally manageable, yet reliable, approximation schemes, based on various model Hamiltonians. Even when ignoring the relativistic effects and freezing the nuclear motion by relying on the Born-Oppenheimer approximation [3], the problem is still too formidable for any system having more than two electrons. For this very reason, almost all molecular applications are based on finite dimensional models.

A very important conceptual step within the MO framework was achieved by the introduction of the independent particle model (IPM), which reduces the N-electron problem effectively to a one-electron problem, though a highly nonlinear one. The variation principle based IPM leads to Hartree–Fock (HF) equations [4, 5] (cf. also [6, 7]) that are solved iteratively by generating a suitable self-consistent field (SCF). The numerical solution of these equations for the one-center atomic problems became a reality in the fifties, primarily owing to the earlier efforts by Hartree and Hartree [8]. The fact that this approximation yields well over 99% of the total energy led to the general belief that SCF wave functions are sufficiently accurate for the computation of interesting properties of most chemical systems. However, once the SCF solutions became available for molecular systems, this hope was shattered.

Although the availability of numerical solutions of HF equations is still restricted to at most two-center (or linear) systems, the development of suitable basis sets enabled the computation of SCF solutions within the Roothaan linear combination of atomic orbitals (LCAO) SCF formalism [9]. Generation of such solutions, even for systems with several hundreds of electrons, is no-

wadays routine, although the handling of general open shell states can still be frustrating at times due to the possible multiplicity of various SCF solutions.

More importantly, however, it became abundantly clear during the late sixties that in spite of the conceptual importance of HF solutions and their utility in supplying quantitative, or at least semiquantitative, information about some molecular properties, their capability to provide a reliable description of the electronic structure in general was rather limited. The first dramatic indication of the inadequacy of SCF solutions emerged from Wahl's study of the F_2 molecule [10]. Using a carefully optimized basis set, Wahl showed that at the SCF level of approximation, F_2 is not bound: the negative SCF dissociation energy he obtained was as large as 1.6 eV!

The shortcomings of IPM are nowadays well recognized. Even though the HF total energies are very accurate, say within 0.1% of their exact value, they are not accurate enough to describe many chemical phenomena or properties of interest. For example, the total energies of first row diatomics are of the order of 10^2 hartree, so we need at least two orders of magnitude higher precision to achieve the so-called chemical accuracy of $\sim$1 millihartree. The situation is even more critical when considering nonenergetic properties. For this reason, all present day quantitative studies account, in one way or another, for the many-electron correlation effects that are lacking in the IPM descriptions.

The generally applicable post-Hartree-Fock methods that are currently widely used are basically of two types: variational and perturbative. The former ones are typified by various configuration interaction (CI) methods (also referred to as shell model), employing a linear ansatz for the wave function in the spirit of Ritz variation principle (cf., e.g. [11]). Since the dimension of the CI problem grows rapidly with the increasing size of the system and the size of the atomic orbital (AO) basis set employed, it is necessary to rely on truncated CI expansions, in spite of the fact that these expansions are slowly convergent, even when based on optimal natural orbitals (NOs). Consequently, such truncated expansions (usually at the doubly excited level relative to the IPM reference, resulting in the CISD method) are not only unable to properly account for the so-called *dynamic* correlation due to higher than doubly excited configurations, but also lack size-extensivity.

To a large extent, both of these shortcomings can be eschewed by relying on multireference (MR) CI approaches, which account for configurations up to a chosen excitation (usually single and double) level (MR CISD method). This approach is particularly effective when handling a manifold of near lying states (that invariably arise when exploring the entire potential energy surfaces (PESs) or curves (PECs) near the dissociation limit), since it is capable to properly account for the so-called *nondynamic* correlation arising in quasidegenerate situations.

However, even the MR CISD methods cannot properly handle dynamic correlation, since this would require a large reference space and thus the *N*-

electron spaces of too large a dimension. Consequently, the MR CISD results are subsequently corrected both for the lack of dynamic correlation (usually via low order perturbation theory) and size extensivity (via various versions of semi-empirical Davidson-type corrections).

The many-body perturbation theory (MBPT) based methods [12, 13], whether of finite or an infinite order, are in many respects complementary to the variational ones, being size-extensive but, at least in their single reference (SR) version, unable to properly account for nondynamic correlation as soon as the state considered becomes quasidegenerate with some higher lying state. The finite order SR MBPT methods are limited to the third, or at most the fourth, order in view of the rapidly increasing computational demands. Although the fourth order results often suffice, there are many instances where the selective higher order terms make a significant contribution. For this reason one usually relies on the coupled cluster (CC) approaches [14, 15, 16], based on the exponential ansatz for the wave function, in which certain classes of the MBPT terms are automatically summed to an infinite order by virtue of solving the energy independent CC equations. These equations may be viewed as recursion formulas for the generation of higher order MBPT contributions of a certain type on the basis of the lower order ones.

For nondegenerate closed shell (CS) ground states, or even for high spin open shell (OS) states, the CC approach at the SD level of truncation (SR CCSD method) generally provides very reliable and precise results that are fully size-extensive. In order to achieve the chemical accuracy of $\sim$1 kcal/mol, at least tri-excited clusters should be accounted for as well. Although a proper account of higher than pair clusters is again computationally too demanding, except for relatively small systems, an approximate account can be achieved perturbatively via the CCSD(T) [17] method (keeping in mind, however, that such an approach will invariably break down when considering highly stretched geometries). Indeed, CCSD and CCSD(T) approaches are presently very popular in view of their high accuracy and reliability [18].

In quasidegenerate situations, the role played by higher-than-pair clusters in the SR type CC approaches can neither be ignored nor accounted for perturbatively [via, e.g., CCSD(T)], thus requiring an MR-type approach. Although much theoretical work has been devoted to this problem during the past two decades (cf., e.g., [16, 19, 20]) and two *bona fide* genuine MR CC methodologies have been developed, their generic computer implementation has yet to be carried out. Both types of methods, the so-called *valence universal* (VU) or Fock space, and *state universal* (SU) or Hilbert space, approaches are based on the effective Hamiltonian formalism and generalized Bloch equations (cf., e.g., [16, 19, 20]), and just as the finite dimensional MR MBPT approaches are plagued by the intruder state problems, not to mention multiple solution problems and computational complexities involved. For this very reason, much attention continues to be devoted to the so-called *state selective* or *state speci-*

fic (SS) MR CC approaches that focus on one state at a time, and essentially represent modified SR-type CC approaches that account for higher than pair clusters in some nonstandard manner.

The complementarity of variational- and perturbative-type approaches, specifically of CI and CC methods, should now be obvious: While the former ones can simultaneously handle a multitude of states of an arbitrary spin multiplicity, accounting well for nondynamic correlation in cases of quasidegeneracy, they are not size-extensive and are unable to properly describe dynamic correlation effects unless excessively large dimensions can be handled or afforded. On the other hand, CC approaches are size-extensive at any level of truncation and very efficiently account for dynamic correlation, yet encounter serious difficulties in the presence of significant nondynamic correlation effects. In view of this complementarity, a conjoint treatment, if at all feasible, would be highly desirable.

Our recently developed reduced multireference (RMR) CCSD method [16, 21, 22, 23, 24, 25] represents such a combined approach. In essence, this is a version of the so-called externally corrected CCSD method [26, 27, 28, 29, 30, 31, 32, 33, 34] that uses a low dimensional MR CISD as an external source. Thus, rather than neglecting higher-than-pair cluster amplitudes, as is done in standard CCSD, it uses approximate values for triply and quadruply excited cluster amplitudes that are extracted by the cluster analysis from the MR CISD wave function. The latter is based on a small active space, yet large enough to allow proper dissociation, and thus a proper account of dynamic correlation. It is the objective of this paper to review this approach in more detail and to illustrate its performance on a few examples.

We first outline the basic idea and origins of the externally corrected CCSD methods in Sect. 2, followed by the formulation and discussion of its special version, the RMR CCSD method, in Sect. 3. In Sect. 4 we present a few illustrative examples and summarize the general conclusions in Sect. 5.

2
Externally Corrected CCSD Methods

In the standard SR CC approach, the exact (nonrelativistic) N-electron wave function $|\Psi\rangle$ for the state of interest (assumed to be energetically the lowest state of a given symmetry species) is represented by the so-called *cluster expansion* relative to some IPM wave function $|\Phi_0\rangle$. This expansion is concisely expressed via the exponential cluster ansatz

$$|\Psi\rangle = e^T|\Phi_0\rangle, \quad \langle\Phi_0|\Phi_0\rangle = \langle\Psi|\Phi_0\rangle = 1\,, \tag{1}$$

with the cluster operator T given by the sum of its i-body components T_i,

$$T = \sum_{i=1}^{N} T_i\,. \tag{2}$$

In the simplest case of a nondegenerate ground state of a closed shell system, one employs the HF wave function as a reference $|\Phi_0\rangle$ and represents the i-body cluster operators T_i as linear combinations of the i-fold excitation operators $G_j^{(i)}$,

$$T_i = \sum_j t_j^{(i)} G_j^{(i)} , \tag{3}$$

where $t_j^{(i)}$ are the unknown cluster amplitudes and $|\Phi_j^{(i)}\rangle = G_j^{(i)}|\Phi_0\rangle$ are the i-times excited configurations (relative to $|\Phi_0\rangle$) spanning the i-times excited subspace of the N-electron space considered.

The exponential character of the cluster expansion warrants the size-extensivity of the resulting formalism regardless of the truncation scheme employed, as implied by a comparison with the standard linear CI expansion of (intermediately normalized) $|\Psi\rangle$,

$$|\Psi\rangle = C|\Phi_0\rangle = \sum_{i=0}^{N} C_i|\Phi_0\rangle, \quad C_0 = 1 , \tag{4}$$

which yields

$$\begin{aligned} C_1 &= T_1, \\ C_2 &= T_2 + \tfrac{1}{2}T_1^2, \\ C_3 &= T_3 + T_1T_2 + \tfrac{1}{6}T_1^3, \\ C_4 &= T_4 + \tfrac{1}{2}T_2^2 + \tfrac{1}{2}T_1^2T_2 + T_1T_3 + \tfrac{1}{24}T_1^4, \quad \text{etc.} \end{aligned} \tag{5}$$

Setting, for simplicity, $T_1 = 0$ (which is equivalent to employing Brueckner's maximum overlap orbitals), we have that $C_1 = 0, C_2 = T_2, C_3 = T_3$ and $C_4 = T_4 + \frac{1}{2}T_2^2$. Thus, even when we approximate T solely by its pair cluster component, $T \approx T_2$ (i.e., $T_i = 0, i \neq 2$), we obtain contributions from all relevant even number of times excited configurations, since $C_{2n} \approx (1/n!)T_2^n$, the most important one being from $C_4 \approx \frac{1}{2}T_2^2$. Since the connected tetra-excited cluster contribution of T_4 is usually negligible in comparison with its disconnected counterpart $\frac{1}{2}T_2^2$, we see that the CCD approach, employing the same number of unknowns as CID, also accounts for a large part of quadruples that are entirely missed by CID.

The unknown cluster amplitudes $t_j^{(i)}$ are determined by the energy-independent CC equations, obtained by projecting the Schrödinger equation, premultiplied with e^{-T}, onto the excited state manifold, i.e.

$$\langle\Phi_j^{(i)}|e^{-T}He^{T}|\Phi_0\rangle = 0 , \tag{6}$$

while the energy results by projecting onto the reference $|\Phi_0\rangle$,

$$E = \langle\Phi_0|e^{-T}He^{T}|\Phi_0\rangle = \langle\Phi_0|He^{T}|\Phi_0\rangle, \tag{7}$$

so that

$$E - \langle\Phi_0|H|\Phi_0\rangle = \langle\Phi_0|H(T_1 + T_2 + \tfrac{1}{2}T_1^2)|\Phi_0\rangle . \tag{8}$$

Thus, the energy is fully determined by one- and two-body clusters in view of the fact that the Hamiltonian contains at most two-body interactions. For the same reason, Eq. (6) implies the following chain of CC equations

$$\begin{aligned}
&\langle\Phi_j^{(1)}|H[1+T_1+(T_2+\tfrac{1}{2}T_1^2)+(T_3+T_1T_2+\tfrac{1}{6}T_1^3)]|\Phi_0\rangle_C=0\,,\\
&\langle\Phi_j^{(2)}|H[1+T_1+(T_2+\tfrac{1}{2}T_1^2)+(T_3+T_1T_2+\tfrac{1}{6}T_1^3)\\
&\qquad\qquad+(T_4+T_1T_3+\tfrac{1}{2}T_1^2T_2+\tfrac{1}{24}T_1^4)]|\Phi_0\rangle_C=0\,,\\
&\langle\Phi_j^{(3)}|H[T_1+(T_2+\tfrac{1}{2}T_1^2)\qquad+\cdots+(T_5+T_1T_4+\cdots)]|\Phi_0\rangle_C=0,\ \text{etc.}\,,
\end{aligned}\tag{9}$$

the subscript C indicating that only connected components are to be retained. This structure reflects the block diagonal structure of the corresponding CI chain,

$$\langle\Phi_j^{(i)}|H(C_{i-2}+C_{i-1}+C_i+C_{i+1}+C_{i+2})|\Phi_0\rangle=E\langle\Phi_j^{(i)}|C_i|\Phi_0\rangle,\ (i=0,1,2,\cdots,N)\tag{10}$$

where $C_i=0$ if $i<0$ or $i>N$. Thus, formally, the CC chain is obtained by setting to zero the left hand side of the CI chain and by replacing the C_i excitation operators by their cluster analogues, Eq. (5), while retaining only the connected terms, the disconnected ones being cancelled by the right hand side energy term in CI Eqs. (10) [cf., e.g., [35] for details]. Clearly, without truncation, both chains are equivalent, providing the full CI (FCI) or full CC (FCC) result, representing the exact solution for the given finite dimensional ab-initio model (as defined by the AO basis set).

Since FCI or FCC can only be carried out for relatively small model systems, all practical applications rely on truncated schemes. Thus, setting $T_3=T_4=0$, the CC chain decouples after the second equation in the chain (9), resulting in the CCSD method. As already pointed out, CCSD provides an excellent approximation that can be further improved by a perturbative account of T_3, as in the CCSD(T) method. Unfortunately, when considering stretched geometries or, generally, quasidegenerate states, T_3 and T_4 are no longer small enough to be neglected or handled perturbatively [in fact, in such cases, CCSD is preferable to CCSD(T)]. The prominence of these higher than pair clusters stems from the inadequacy of the SR formalism. Clearly, if we are able to obtain at least approximate values of T_3 and T_4 clusters from some independent source, the validity of the CCSD approximation could be extended to significantly stretched geometries. This is precisely the idea of the *externally corrected* CCSD (ecCCSD).

Considering the general algebraic structure of CCSD equations (for simplicity we drop the superscript indicating the excitation order),

$$a_i+\sum_j b_{ij}t_j+\sum_{j<k}c_{ijk}t_jt_k+\cdots=0\,,\tag{11}$$

it is easy to realize that the T_3 and T_4 terms will only modify the absolute terms a_i, while the T_1T_3 term will contribute to the linear b_{ij} terms associated with monoexcited $t_j^{(1)}$ cluster amplitudes. In fact, even this latter term can

be treated noniteratively using the approximate T_1 clusters provided by the external source [33].

Several possible choices of an external source have been tested so far. The basic requirement is that such a source must provide a reasonable approximation of the most important three- and four-body clusters that are missing in the SR CCSD approach. At the very least, we require it to describe the essential nondynamic correlation effects. The practical aspects require that it be easily accessible. The first attempts in this direction exploited the unrestricted Hartree Fock (UHF) wave function [of different orbitals for different spins (DODS) type]. Its implicit exploitation lead to the so-called ACPQ (approximate coupled pairs with quadruples) method [26, 27]. Recently, its explicit version was also developed and implemented [31]. Although in many cases this source enables one to reach the correct dissociation channel, its main shortcoming is the fact that for the CS systems it can only provide T_4 clusters, since the T_3 contribution vanishes due to the spin symmetry of the DODS wave function. Nonetheless, the ACPQ method enabled an effective handling of extended linear systems (at the semi-empirical level), which are very demanding, since the standard CCSD method completely breaks down in this case [27].

Another potentially suitable source is represented by simple VB-type wave functions, which again correctly describe bond breaking situations. This possibility was tested at the semi-empirical level with considerable success [28, 30].

At the ab-initio level, the most obvious possibility is offered by CAS SCF or CAS FCI (i.e., CI within the CAS or, equivalently, CAS SCF without the orbital reoptimization based on RHF orbitals, cf. [33, 34]) wave functions based on the smallest possible active-space that warrants the correct description of the dissociation channel at hand. This option was also suggested by Stolarczyk [29], although we are not aware of any concrete implementation. Our testing proved to be very encouraging [33, 34], particularly for open shell systems, in which case we employed the spin-adapted CCSD based on the unitary group approach (UGA) [16, 36]. Even in the case of triple bond breaking, the applicability of the CCSD approximation can be significantly extended, as will be shown in Sect. 4. Most recently, we have explored the MR CISD wave function as an external source, as described in the next section.

3
Reduced Multi-Reference (RMR) CCSD

Relying on the above discussed complementarity of the SR CCSD and MR CISD ansätze, it seems particularly attractive to employ the latter as an external source of T_3 and T_4 corrections. In order to explicitly illustrate this complementarity and the scope of the formalism involved, let us consider a minimal 2-reference case, i.e. let us assume that a given SCF reference becomes quasidegenerate with another configuration. For a CS system this case arises when the one-electron active-space involves only two MOs, each belonging to a different

symmetry species. Designating these active orbitals by i and j, the N-electron model (or reference) space will be spanned by configurations $|\Phi_0\rangle = |(\text{core})i\bar{i}\rangle$ and $|\Phi_1\rangle = |(\text{core})j\bar{j}\rangle$. The corresponding two reference (2R) CISD will thus involve single and double excitations relative to $|\Phi_0\rangle$ and $|\Phi_1\rangle$, the latter being triply and quadruply excited relative to $|\Phi_0\rangle$.

Focusing, for the sake of simplicity, on an even-number-of-times excited states only, and labeling core and virtual orbitals by letters from the beginning $(a, b, c, \cdots)$ and the end $(r, s, t, \cdots)$ of the alphabet, respectively, we realize that only core–virtual excitations will produce a distinct set of states; ignoring the details of spin couplings, we can symbolically represent them as follows: $|\Phi_0(a, b \rightarrow r, s)\rangle = |(\{\text{core}\}\backslash\{a, b\})i\bar{i}r\bar{s}\rangle$ and $|\Phi_1(a, b \rightarrow r, s)\rangle = |(\{\text{core}\}\backslash\{a, b\})j\bar{j}r\bar{s}\rangle$. Clearly, the former ones are doubles and the latter ones quadruples relative to $|\Phi_0\rangle$. All other double excitations lead to identical sets of doubles, namely the active to virtual ones to $|\Phi_0(i, i \rightarrow r, s)\rangle = |(\text{core})r\bar{s}\rangle = |\Phi_1(j, j \rightarrow r, s)\rangle$, and core to active ones to $|\Phi_0(a, b \rightarrow j, j)\rangle = |(\{\text{core}\}\backslash\{a, b\})i\bar{i}j\bar{j}\rangle = |\Phi_1(a, b \rightarrow i, i)\rangle$. We thus see that the 2R CISD involves the same number of quadruples as there are core–virtual doubles, both proportional to $n_c^2 n_v^2$ in number (we designate the number of core and virtual orbitals by n_c and n_v, respectively). The number of remaining doubles (core–active and active–virtual) is much smaller, being proportional to n_c^2 and n_v^2, respectively.

Although the MR CISD, based on a small reference space, will involve only a small subset of quadruple excitations relative to the leading configuration $|\Phi_0\rangle$, these quadruples will clearly be those that are primarily responsible for a proper account of dynamic correlation and that guarantee the correct description of the dissociation channel at hand. Indeed, the number of core–virtual quadruples is proportional to $n_c^4 n_v^4$. The majority of these will, of course, primarily contribute to the dynamic correlation and may thus be effectively represented by their disconnected components arising through the CCSD cluster ansatz. We can thus expect the low-dimensional MR CISD to supply us with the most important connected triples and quadruples, while relying on the CC ansatz to represent the remaining higher-than-pair clusters.

A similar situation to that just described for the 2R CISD will clearly arise in the general case, the number of quadruples involved being proportional to the dimension of the model space. Nonetheless, these quadruples will represent a very small subset of all possible quadruples, their number being essentially a multiple of the number of doubles by the factor $(M - 1)$, M being the number of reference configurations. Moreover, these quadruples will be used only once to correct the absolute term of CCSD equations.

In general, the MR CISD wave function based on an M-dimensional reference space has the form

$$|\Psi_0\rangle = \sum_{p=0}^{M-1} C^{[p]} |\Phi_p\rangle \,, \tag{12}$$

where $C^{[p]} = C_1^{[p]} + C_2^{[p]}$ designates the CI-type SD excitation operator relative to the reference $|\Phi_p\rangle$. Clearly, the fact that the same configuration may arise by the excitation from different references must be properly taken into account. Designating the leading reference configuration in $|\Psi_0\rangle$ as $|\Phi_0\rangle$, the SD core–valence excitations from the remaining references, namely $|\Phi_1\rangle, |\Phi_2\rangle, \cdots, |\Phi_{M-1}\rangle$, will represent higher-than-double excitations relative to $|\Phi_0\rangle$, so that we can rewrite the MR CISD expansion (12) in the SR CI form, using $|\Phi_0\rangle$ as a reference,

$$|\Psi_0\rangle = \left(1 + \sum_{i=1}^{N} C_i^{(0)}\right) |\Phi_0\rangle\,, \tag{13}$$

distinguishing the relevant i-fold excitation operators by the superscript "(0)". Applying standard cluster analysis [cf. Eqs. (5)], it is then straightforward to rewrite this wave function in the SR CC form, namely

$$\begin{aligned} |\Psi_0\rangle &= e^{T_1^{(0)} + T_2^{(0)} + \Delta} |\Phi_0\rangle \\ &= e^{T_1^{(0)} + T_2^{(0)} + \{T_3^{(0)} + T_4^{(0)} + \cdots\}_{\text{subset}}} |\Phi_0\rangle\,, \end{aligned} \tag{14}$$

where Δ represents a subset of higher than pair cluster operators arising through the MR CISD ansatz.

Assuming now that the $T_3^{(0)}$ and $T_4^{(0)}$ amplitudes so obtained represent a reasonable approximation of actual T_3 and T_4 clusters, we can employ the RMR CCSD ansatz

$$|\Psi\rangle = e^{T_1 + T_2 + T_3^{(0)} + T_4^{(0)}} |\Phi_0\rangle\,, \tag{15}$$

leading to a special case of the general ecCCSD. Clearly, the decoupling of the SR CC chain of equations, Eq.(9), using approximate values of T_3 and T_4, should provide a superior approximation to the standard one which simply assumes that T_3 and T_4 vanish. This will be particularly the case in quasidegenerate situations, when T_3 and T_4 are no longer negligible, in fact not even accessible to a perturbative treatment. On the other hand, the SD amplitudes defining $T_1^{(0)}$ and $T_2^{(0)}$, although representing a reasonable approximation to the actual SD amplitudes, do not properly reflect the effect of a large number of higher excited (disconnected) clusters that are entirely absent in the low-dimensional MR CISD approach, thus missing the essential part of the dynamic correlation. It is thus worthwhile to recalculate these amplitudes using ecCCSD, relying on ansatz (15). Since only T_1 and T_2 involve the unknown cluster amplitudes, ecCCSD is no more demanding computationally than standard CCSD, once the initial $T_3^{(0)}$ and $T_4^{(0)}$ corrections are evaluated.

To summarize, the RMR CCSD method involves the following three steps: (i) We choose a suitable reference space and compute the corresponding MR CISD. Next, (ii) we compute $T_3^{(0)}$ and $T_4^{(0)}$ clusters by cluster analyzing the MR CISD wave function of step (i), and finally (iii) we use these amplitudes to generate and solve ecCCSD equations. The details of the actual implementation of the RMR CCSD method for various types of reference spaces can be found in our earlier papers [21, 22, 23, 24, 25].

4 Illustrative Examples

Although the existing applications of RMR CCSD, and of other versions of the ecCCSD method, have shown considerable promise, much work remains to be done in order to establish optimal sources of higher than pair clusters that would be both reliable and computationally affordable, as well as to determine the limits of applicability of this type of approaches. Here we shall only present a few typical examples that illustrate the potential of this technique, drawing on both the existing applications and recently generated new results.

Both most extensively tested ecCCSD approaches employ the CI-type wave functions as a source of TQ clusters. Those using active-space triples and quadruples (ASTQ) are based on CAS SCF, or CAS FCI, or even truncated CAS FCI wave functions [32, 33, 34], while the RMR CCSD method relies on a small reference space MR CISD wave function [21, 22, 23, 24, 25]. Thus, in the former case, only those three- and four-body clusters are taken into account that are obtainable through the excitations within the set of active (spin)orbitals. Consequently, this type of ecCCSD approaches requires a large enough active-space in order to obtain a meaningful subset of triply and quadruply excited cluster amplitudes. Note, for example, that the minimal 2-electron/2-orbital active-space will not provide any such cluster amplitudes. We thus generally refer to these type of approaches as the ASTQ CCSD methods. In contrast, the RMR CCSD method employs a small reference space, but considers excitations involving virtual orbitals as well. Thus, while the reference (or active) space may not provide any information concerning the three- and four-body clusters in this case (as is the case, e.g., with the 2R-RMR CCSD method, employing the minimal 2-electron/2-orbital space), the SD excitations involving virtuals will. In the following, we shall briefly compare the performance of these two types of ecCCSD approaches.

So far, our primary focus was on the computation of potential energy surfaces (PESs), or curves (PECs), and a few derived properties, such as molecular geometries and harmonic vibrational frequencies, and only little attention was devoted to other properties (e.g., various multipolar moments, etc. [16]). As explained above, the main benefit of the ecCCSD approaches pertains to quasi-degenerate situations, most often encountered for highly stretched geometries, or in states having a genuine multi-reference character, since for nondegenerate systems in their equilibrium geometries the standard CCSD or CCSD(T) approaches work extremely well [15, 16, 18]. Thus, when exploring, for example, various static properties, the main interest in ecCCSD approaches will be when generating property surfaces. Since these are seldom directly accessible experimentally, it will be important to test these methods on systems where the desired data are significantly influenced by the shape of the property surface away from the equilibrium geometry (c.f., e.g., [37]).

Although the comparison with the experimental data, and their reliable prediction and rational interpretation, represent the ultimate objective of any theory, such an undertaking involves a multitude of unknown factors that can play a significant role, starting with the adequacy of the model employed (as defined by the AO basis set in our case), possible nonadiabatic or relativistic effects, etc., including the shortcomings in the analysis of the experimental data. It has thus been long recognized that the most dependable and least biased test of the performance of approximate methods is the comparison with the FCI data, representing the exact result for a chosen ab-initio model. Unfortunately, the FCI results can be generated for only a relatively small model system, leaving open the question of the basis-set-dependence of the investigated performance. Nonetheless, our experience indicates that the methods performing well for relatively small model systems also do well when applied to realistic models. In any case, the comparison with FCI results is nowadays a sine qua non for any new method or approximation.

As explained in the Introduction (Sect. 1), the performance of standard CCSD directly depends on the role played by higher than pair clusters. It is thus important to test the methods that are designed to correct — at least approximately — this inadequacy for nonequilibrium geometries, even though they may be beneficial for equilibrium geometries as well [as is the case, for example, for the CCSD(T) method which, unfortunately, breaks down in quasidegenerate situations]. For this reason, our primary focus is on the performance for highly stretched geometries. Clearly, the difficulty dramatically increases with the multiplicity of the bond being stretched.

A typical example when stretching a single bond is given in Table 1 for the $X^2\Pi$ state of the OH radical at both the double zeta (DZ) and DZ plus polarization (DZP) levels of approximation. Since for most applications the absolute energies are of little importance, we also present the so-called "nonparallelism error" (NPE), defined as the maximal difference between the exact FCI and approximate energies for a given range of geometries considered. Thus, when both PESs or PECs are "parallel" (i.e., separated by the same energy difference for all geometries), we have that NPE $= 0$. Clearly, when the differences between the FCI and approximate energies do not depend monotonically on the varied geometry parameter (i.e., the internuclear separation R in our case), the given NPE values are only approximate, since they are based on the results for a finite number of geometries considered. Ideally, NPE is given by the sum of absolute values of the maximal positive and negative energy differences. An even better measure would be, of course, provided by the integral

$$\int_{\mathcal{D}} |\Delta E(\mathbf{R}) - \Delta E(\mathbf{R}_x)|\mu(\mathbf{R}) d|\mathbf{R}|\,,$$

where $\Delta E(\mathbf{R}) = E^{(\mathrm{FCI})}(\mathbf{R}) - E^{(X)}(\mathbf{R})$, X designates the approximation employed, $\mathcal{D}$ represents the parameter domain of geometries considered, $\mathbf{R}_x = \mathbf{R}_e$ or $\mathbf{R}_\infty$ or other suitable reference geometry, and $\mu(\mathbf{R})$ is the optional weight factor.

Table 1. Comparison of the SR CCSD, RMR CCSD, and ASTQ CCSD energies with the exact FCI result for the $X^2\Pi$ state of OH at three internuclear separations R, $R = R_e = 1.832$ bohr, $R = 1.5R_e$, and $R = 2R_e$. Except for the SCF and FCI total energies, which are reported as $-(E + 75)$ (in hartree), the energy differences (in millihartree) relative to the FCI result are given in all cases. The nonparallelism error (NPE) for the interval $R \in [R_e, 2R_e]$ (in millihartree) is also given for easier comparison (see the text for details)

Method [a]	Dimension[b]	R_e	$1.5R_e$	$2R_e$	NPE	Reference
DZ basis set						
SCF	1	0.384893	0.294586	0.192836	84.595	
FCI	3460(CSF)	0.481223	0.423691	0.373761	0.000	[33]
SR CCSD	112(CSF)	1.195	2.705	5.648	4.453	[33]
2R-RMR CCSD	362(Det)	1.048	1.591	3.048	2.000	This work
5D-RMR CCSD	607(Det)	0.915	1.222	0.993	0.307	This work
CCSD-(7,8)FCI		0.511	1.431	1.558	1.047	[33]
CCSD-(7,5/3)SOCI	268(CSF)	0.75	1.90	2.32	1.57	[34]
CCSD-(7,5/5)SOCI	656(CSF)	0.71	1.08	0.14	0.94	[34]
DZP basis set						
SCF	1	0.406319	0.302993	0.195050	75.899	
FCI	441792(CSF)	0.568015	0.492100	0.432645	0.000	[33]
SR CCSD	540(CSF)	2.440	5.046	12.699	10.259	[33]
2R-RMR CCSD	2031(Det)	2.141	2.896	5.568	3.427	This work
5D-RMR CCSD	3705(Det)	1.880	2.090	1.648	0.442	This work
CCSD-(7,8)FCI		1.702	3.885	9.016	7.314	[33]
CCSD-(7,5/3)SOCI	268(CSF)	1.97	4.27	9.80	7.83	[34]
CCSD-(7,10)FCI	3460(CSF)	1.382	2.759	7.020	5.638	[33]
CCSD-(7,5/5)SOCI	656(CSF)	1.92	3.68	8.56	6.63	[34]

[a] Except for RMR CCSD, the ecCCSD methods exploiting an ASTQ source are designated as CCSD-source.

[b] The dimension is given either by the number of configurations (CSF) or by the number of Slater determinants (Det) involved. In case of RMR CCSD and ASTQ CCSD, the indicated dimension is that of the relevant CI (the dimension of ecCCSD being the same as that of SR CCSD).

Table 1 is typical of several other examples that have been examined so far, including a simultaneous stretching of several single bonds [21, 22, 33, 34]. We see that the standard SR CCSD method represents an excellent approximation, being only about 2 mhartree away from the FCI energy at the equilibrium geometry ($R = R_e$) for both DZ and DZP basis sets. Nonetheless, this error steadily increases with the increasing internuclear separation R (measured in units of $R_e = 1.832$ a.u.). The magnitude of this error, and its rate of increase with R, are larger for a larger basis set, resulting in a considerably larger NPE (10.3 vs. 4.5 mhartree), even though the energy lowering due to the increase in the basis set size decreases with R (from 86.8 to 58.9 mhartree). However, since the RHF or SCF wave function is used as a reference, the correlation energy rapidly increases with R (from 96.3 to 180.9 mhartree in the DZ case and from

161.7 to 237.6 mhartree for a DZP basis set), as does the SR CCSD error as a percentage of the correlation energy (from 1.2 to 3.1% and from 1.5 to 5.3% for DZ and DZP basis sets, respectively). Consequently, even though SCF NPE is slightly smaller at the DZP level (yet, still enormous), the SR CCSD NPE at the DZP level is larger by about a factor of two relative to the DZ level. This indicates that the need for a proper account of nondynamic correlation effects increases with the increasing basis set size.

The $X^2\Pi$ state of the OH radical has been examined with both the RMR CCSD and ASTQ CCSD methods [33, 34], thus providing a simple example for the comparison of both types of approaches. The ASTQ CCSD results are based on different CAS FCI related sources. The symbol (n, m) FCI indicates a CAS FCI external source using n-electron/m-orbital active-space. A less demanding source employs the second order CI (SOCI), designated as $(n, m_1/m_2)$ SOCI, involving $(n, m_1 + m_2)$ active-space, in which the m active orbitals are subdivided into the m_1 internal and m_2 external ones. Thus, for example, (7,5/3) SOCI employs a reference space involving 7 electrons distributed over 5 internal orbitals, together with all singly and doubly excited configurations into the 3 external orbitals. We recall that $(n, m_1/m_2)$ SOCI usually represents a good approximation to $(n, m_1 + m_2)$ FCI, as also indicated by our results in Table 1.

The CAS FCI or SOCI corrected CCSD results represent a definite improvement over the standard SR CCSD results. Nonetheless, for a given choice of the active-space, the performance of the ASTQ CCSD approaches deteriorates with the increasing size of the basis set. This is easily understood, since the external source does not include excitations out of the active-space, which are more important for larger basis sets. We should also mention that computationally much more demanding CAS SCF wave functions are about as effective as CAS FCI ones (cf., e.g., Tables II and III of [33]).

For the sake of comparison, we also present two RMR CCSD results in Table 1. Here, the basic guiding principle for the choice of the reference space is to consider suitable bonding and *corresponding* antibonding orbitals in the vicinity of the Fermi level as active orbitals. In the present case, these are the doubly occupied σ_{OH} and unoccupied σ^*_{OH} orbitals. The minimal reference space is then spanned by the RHF configuration and by the doubly excited configuration involving these two orbitals, and the correspondiung RMR CCSD is referred to as 2R-RMR CCSD. This type of reference space is very similar to that employed in the study of the HF and F_2 molecules [22]. In the case of the $X^2\Pi$ state of the OH radical, the situation is slightly different in that its HOMO is a singly occupied π orbital. When the singly excited σ to σ^* configuration is also included, the reference space is spanned by 5 determinants (D), and the relevant RMR CCSD is designated as 5D-RMR CCSD.

It is immediately apparent from the results in Table 1 that the RMR CCSD strategy is yet more effective for reasons given in Sect. 3. Already the 2R version

Table 2. Comparison of various limited SR CI, SR CCSD, and selected RMR CCSD and corresponding MR CISD energies with the exact FCI result for the N_2 molecule, obtained with a DZ basis for three internuclear separations R, $R = R_e = 2.068$ bohr, $R = 1.5R_e$, and $R = 2R_e$ [23]. Except for the FCI total energy, which is reported as $-(E + 108)$ (in hartree), the energy differences (in millihartree) relative to the FCI result are given in all cases. The nonparallelism error (NPE) for the intervals $R \in [R_e, 1.5R_e]$ and $R \in [R_e, 2R_e]$ (in millihartree) is given for easier comparison (see the text for details)

Method	R_e	$1.5R_e$	$2R_e$	NPE ($1.5R_e$)	NPE ($2R_e$)
FCI	1.105115	0.950728	0.868239	0	0
SR CISD	22.637	90.113	210.896	66.48	187.26
SR CISDT	18.170	74.297	181.989	56.13	163.82
SR CISDTQ	1.379	13.938	39.361	12.56	37.98
SR CCSD	8.289	33.545	−69.917	25.26	103.46
symmetry-non-adapted reference space[a]					
CAS(2,2)-MR CISD	23.413	75.600	145.285	52.19	121.87
CAS(4,4)-MR CISD	8.457	16.122	30.047	7.67	21.59
CAS(6,6)-MR CISD	6.798	6.423	6.168	1.12	1.37
CAS(2,2)-RMR CCSD	8.086	22.385	−77.701	14.30	100.09
CAS(4,4)-RMR CCSD	2.129	4.667	−15.018	2.54	19.69
CAS(6,6)-RMR CCSD	1.417	1.987	2.775	0.57	1.36
S_z and point-group symmetry-adapted reference space[b]					
(MS9)-8D-MR CISD	13.174	21.787	31.113	8.61	17.94
(MS8)-16D-MR CISD [c]	11.848	17.755	31.618	5.91	19.77
(MS1)-48D-MR CISD [c]	8.290	8.634	7.438	1.61[d]	2.46[d]
(MS9)-8D-RMR CCSD	4.123	5.829	−38.638	1.71	44.47
(MS8)-16D-RMR CCSD	2.982	6.202	−8.711	3.22	14.91
(MS1)-48D-RMR CCSD	2.146	2.704	3.004	0.56	0.86

[a] The symmetry-non-adapted reference space is spanned by determinants obtained by distributing the active electrons over the active spin orbitals in all possible ways without imposing any symmetry restrictions.

[b] The definition of model spaces (MSi) is given in [23]. The number n of determinants (D) spanning the reference space is also indicated by a prefix nD.

[c] These MR CISD results are not reported in [23].

[d] These NPEs are only approximate, and cannot be obtained from the data given in the table, since the largest deviation from FCI was found at $1.25R_e$ (see [23]).

decreases the NPE by a factor of two or more, while the 5D version[1] gives by far the best results. Most importantly, the 5D-RMR CCSD performs equally well in both DZ and DZP cases.

To illustrate how these methods work in the most challenging case when dissociating a triple bond, we provide a brief summary of our earlier results

[1] We note here that since RMR CCSD is based on a special version of MR CISD, our codes are not yet spin-adapted so that the corresponding dimensions are given in terms of the number of Slater determinants rather than CSFs. Note also that, on average, the number of CSFs is about one-third of the number of determinants or less.

for the ground state of N_2 [23] in Table 2. In this case, NPE is given for two regions (based on the energies for $R = R_e, 1.25R_e, 1.5R_e$, $1.75R_e$ and $2.0R_e$). Only the DZ basis is considered, since no FCI results are available for larger basis sets (for RMR CCSD data obtained with a DZP basis, see [23]).

A rapid increase in the importance of higher-than-pair clusters is clearly illustrated by the sequence of SR CI results. Even the SR CISDTQ NPEs amount to 12.6 and 38 mhartree. This clearly indicates the role played by higher-than-4-body (both connected and disconnected) clusters as $R \to \infty$.

We see again that RMR CCSD provides a dramatic improvement over SR CCSD, even though larger and larger MRCIs are required if we wish to correct at larger R values. In this case we also present the energy values obtained with MR CISD employed as the external source. The RMR CCSD results invariably provide significantly closer energies to FCI, although the improvements in the NPE may be less significant. Clearly, the larger the CI used, the smaller the difference between both energies, since when using FCI as an external source, the ecCCSD method would simply recover the FCI result.

In addition to CAS(n, n) RMR CCSD results, we also present those based on incomplete active-spaces. These model spaces (MSi's) are obtained by imposing certain restrictions on the orbital occupancies of active orbitals (see [23]). Only three typical spaces of the latter type, namely MSi for $i = 1, 8$ and 9, involving 8, 16 and 48 determinants, respectively, are included.

In concluding this section, let us note that the largest system considered so far was the ozone molecule at the DZP level of approximation. Here we studied the geometry and the harmonic force field for the ground state, obtaining excellent results [25].

5 Conclusions

The above given results, together with those published earlier [21, 22, 23, 24, 25, 30, 32, 33, 34], clearly indicate the usefulness of incorporating the corrections for higher-than-pair clusters, obtained from some external source, in the standard SR CCSD method, and thus extending the generally excellent performance of the latter into the regions where a strong quasidegeneracy sets in. Although the search for the optimal source, or sources, of 3- and 4-body clusters should be continued, there is little doubt that it is the RMR CCSD strategy that currently represents such an optimal choice, at least within the class of sources based on CI-type wave functions. RMR CCSD also represents a state selective (SS) MR CC approach, whose essential features are very close to those of genuine MR CC approaches.

The CI-type sources that have been explored so far to generate external corrections are CAS SCF, CAS FCI, and various truncated versions of CAS FCI (such as AS CISD, etc.). Since all these sources account for excitations involving

only active orbitals, we refer to them as active-space triples and quadruples (ASTQ) approaches.

Interestingly enough, there is usually a little difference between the CAS SCF and CAS FCI ecCCSD energies. In fact, in some cases CAS SCF corrections yield (slightly) worse results than much more affordable CAS FCI corrections (cf., e.g., the results for the $^2\Pi$ state of OH, especially Tables II and III of [33]). This may seem surprising at first glance, since the CAS SCF energies generally represent a very significant improvement over the CAS FCI ones. The principal reason is the fact that the leading configuration in the CAS SCF wave function, which is used as a reference, represents an inferior approximation to the RHF (or ROHF) wave function, which by definition is the best single configurational wave function from the energy point of view. One could, of course, employ CAS SCF corrections while relying on the RHF reference in the CCSD part. This option, which unfortunately requires an additional orbital transformation, is being investigated by Toboła (see also [29]) as an alternative option.

In any case, even severely truncated ASTQ approaches invariably provide a significant improvement over the energies associated with the wave function used as a source of TQ corrections, as well as over the standard SR CCSD. We mention that a particularly effective truncation can be designed, as very recently shown by Peris et al. [38], using the CIPSI-type selection of truncated active-space configurations. Nonetheless, since in all these approaches the excitations are restricted to the active orbitals, one has to employ large enough active-space in order to obtain reasonable T_3 and T_4 corrections, as already stated (for example, the minimal 2-electron/2-orbital active-space does not provide any T_3 and T_4 corrections in this approach). On the other hand, the CAS FCI and related truncated ASTQ approaches have the advantage that the cost of extracting T_3 and T_4 clusters does not increase with the size of the basis set. Unfortunately, for a chosen active space, the performance of these ASTQ approaches deteriorates when larger and larger basis sets are employed, indicating the importance of excitations involving virtual orbitals (not included in the one-electron active space). However, since the computational effort that is required to generate CAS FCI, or truncated CAS CI, wave functions is rather modest when compared with an overall cost (assuming that we use a moderately large active space), and the improvements in the results in the quasidegenerate region are significant, it is worthwhile to exploit this approach should other versions prove to be excessively demanding.

The RMR CCSD approach certainly represents the optimal ecCCSD method relying on the CI-type wave function. In particular, we must emphasize its conceptual simplicity and the fact that the method is unambiguously defined by the choice of the reference space, whose dimension can be very small indeed. Already a 2-dimensional 2R-RMR CCSD often provides excellent results. This is important, since the computational cost, in contrast to ASTQ approaches, is proportional to the number of references employed. On the other hand, in view of the fact that in computing a subset of $T_3^{(0)}$ and $T_4^{(0)}$ amplitudes we

automatically generate $T_1^{(0)}$ and $T_2^{(0)}$ clusters, it is beneficial to exploit the latter as the initial approximation when solving the ecCCSD equations. In fact, it was shown recently, that one can achieve excellent results already at the linear level of approximation (equivalent to the first iteration in the standard algorithm for solving CCSD equations), in the framework of the so-called almost linear (AL) CCSD approach of Jankowski et al. [39].

In concluding this section, let us briefly mention two other techniques aiming in the same direction as RMR CCSD and having at least a conceptual relationship with it. One possibility is, of course, to consider the same subset of T_3 and T_4 amplitudes as does RMR CCSD, but to determine them from a set of suitably truncated SR CCSDTQ equations, in the spirit of the SS CC approach by Adamowicz et al. [40]. This method, referred to as CCSDtq, was recently proposed by Piecuch et al. [41], and is being implemented computationally. The attractive feature of this approach is undoubtedly the fact that it does not depart from the CC formalism. On the other hand, the dimensionality of the relevant CCSDtq equations will still be quite formidable for larger systems and/or basis sets. Even more importantly, at the conceptual level, this approach does not account for higher-than-T_4 clusters in contrast to ecCCSD. Here we recall the fact that even though at most T_4 clusters are required in the ecCCSD approaches, they implicitly account for all higher excited clusters that are present in the MR CISD or other wave function used as an external source. Indeed, when we use FCI T_3 and T_4 cluster amplitudes, ecCCSD returns the exact FCI energy.

The other possibility is to focus on the MR CISD wave function and exploit the $T_1^{(0)}$ and $T_2^{(0)}$ clusters it provides to account for the dynamic correlation due to disconnected triples and quadruples that are absent in the MR CISD wave function. This approach, recently proposed and tested by Meissner and Grabowski [42], may thus be characterized as a CC-ansatz-based Davidson-type correction to MR CISD. The duplication of contributions from higher-than-doubly excited configurations that arise in MR CISD as well as through the CC exponential ansatz is avoided by a suitable projection onto the orthogonal complement to the MR CISD N-electron space. The results are very encouraging, particularly in view of their affordability, though somewhat inferior to RMR CCSD.

On the whole, the strategy of ecCCSD is already providing useful results and its further pursuit, in whichever form it may take, should enable us to substantially extend the range of applicability of the standard SR CCSD or CCSD(T) methods that serve us so well in nondegenerate situations.

Acknowledgements. Continued support by the National Sciences and Engienering Research Council of Canada (J.P.) is acknowledged. The senior author also gratefully acknowledges an Alexander von Humboldt Research Award, as well as the hospitality of Professor G.H.F. Diercksen and the Max-Planck-Institute for Astrophysics in Garching during his visit.

6 References

1. Heitler W, London F (1927) Z Phys 44:455
2. Hellmann H (1937) Einführung in die Quantenchemie. Franz Deuticke, Leipzig, Wien
3. Born M, Oppenheimer JR (1927) Ann Phys (Leipzig) 84:457
4. (a) Hartree DR (1927) Proc Cambridge Phil Soc 24:89. (b) idem (1948) ibid 45:230
5. (a) Fock V (1930) Z Phys 61:126, 62:795. (b) idem (1932) ibid 75:622
6. Hartree DR (1957) The Calculation of Atomic Structures. Wiley, New York
7. Klobukowski M, Carbó R (eds) (1990) Self-Consistent Field: Theory and Applications. Elsevier, Amsterdam
8. Hartree DR, Hartree W (1936) Proc Roy Soc (London) A154:588
9. Roothaan CCJ (1951) Rev Mod Phys 23:69
10. Wahl AC (1964) J Chem Phys 41:2600
11. Shavitt I (1977) The Method of Configuration Interaction. In: Schaefer III HF (ed) Methods of Electronic Structure Theory. Plenum, New York, London, p 189
12. Wilson S (1984) Electron Correlation in Molecules. Clarendon, Oxford
13. Harris FE, Monkhorst HJ, Freeman DL (1992) Algebraic and Diagrammatic Methods in Many-Fermion Theory. Oxford University Press, New York, Oxford
14. (a) Čížek J (1996) J Chem Phys 45:4256. (b) idem (1969) Adv Chem Phys 14:35
15. (a) Bartlett RJ (1989) J Phys Chem 93:1697. (b) idem (1995) Coupled-Cluster Theory: An Overview of Recent Developments. In: Yarkony DR (ed) Modern Electronic Structure Theory, Part II, Advanced Series in Physical Chemistry, Vol. 2. World Scientific, Singapore, p 1047
16. Paldus J, Li X (1999) Adv Chem Phys 110: (in press)
17. (a) Raghavachari K (1985) J Chem Phys 82:4607. (b) Urban M, Noga J, Cole SJ, Bartlett RJ (1985) ibid 83:4041. (c) Raghavachari K, Trucks GW, Pople JA, Head-Gordon M (1989) Chem Phys Lett 157:479
18. Lee TJ, Scuseria GE (1995) Achieving Chemical Accuracy with Coupled-Cluster Theory. In: Langhoff SR (ed) Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy. Kluwer, Dordrecht, p 47
19. Paldus J (1992) Coupled Cluster Theory. In: Wilson S, Diercksen GHF (eds) Methods in Computational Molecular Physics, NATO ASI Series, Series B: Physics, Vol 293. Plenum, New York, p 99
20. Paldus J (1994) Algebraic Approach to Coupled Cluster Theory. In: Malli GL (ed) Relativistic and Correlation Effects in Molecules and Solids, NATO ASI Series, Series B: Physics, Vol 318. Plenum, New York, p 207
21. Li X, Paldus J (1997) J Chem Phys 107:6257
22. Li X, Paldus J (1998) J Chem Phys 108:637
23. Li X, Paldus J (1998) Chem Phys Lett 286:145
24. Li X, Paldus J (1998) Coll Czech Chem Commun 63:1381
25. Li X, Paldus J (1999) J Chem Phys 110:2844
26. Paldus J, Čížek J, Takahashi M (1984) Phys Rev A 30:2193
27. Paldus J, Takahashi M, Cho RWH (1984) Phys Rev B 30:4267
28. Paldus J, Planelles J (1994) Theor Chim Acta 89:13
29. Stolarczyk LZ (1994) Chem Phys Lett 217:1
30. (a) Planelles J, Paldus J, Li X (1994) Theor Chim Acta 89:33. (b) idem (1994) ibid 89:59
31. Piecuch P, Toboła R, Paldus J (1996) Phys Rev A 54:1210
32. Peris G, Planelles J, Paldus J (1997) Intern J Quantum Chem 62:137
33. Li X, Peris J, Planelles J, Rajadell F, Paldus J (1997) J Chem Phys 107:90

34. Peris G, Rajadell F, Li X, Planelles J, Paldus J (1998) Mol Phys 94:235
35. (a) Čížek J, Paldus J (1971) Intern J Quantum Chem 5:359. (b) Paldus J, Čížek J, Jeziorski B (1989) J Chem Phys 90:4356
36. Li X, Paldus J (1994) J Chem Phys 101:8812
37. (a) Kondo AE, Piecuch P, Paldus J (1995) J Chem Phys 102:6511. (b) idem (1996) ibid 104:8566. (c) Piecuch P, Kondo AE, Špirko V, Paldus J (1996) ibid 104:4699. (d) Špirko V, Piecuch P, Kondo AE, Paldus J (1996) ibid 104:4716. (e) Piecuch P, Špirko V, Paldus J (1996) ibid 105:11068. (f) Piecuch P, Špirko V, Kondo AE, Paldus J (1998) Mol Phys 94:55. (g) Piecuch P, Špirko V, Paldus J (1998) Polish J Chem 72:1635
38. Peris G (1998) Ph.D. Thesis, Universitat Jaume I, Castello de la Plana, Spain
39. (a) Jankowski K, Kowalski K (1996) Chem Phys Lett 256:141. (b) Jankowski K, Grabowski I, Kowalski K (1998) J Chem Phys 109:6255. (c) Li X, Grabowski I, Jankowski K, Paldus J (1999) Adv Quantum Chem (in press).
40. (a) Oliphant N, Adamowicz L (1991) J Chem Phys 94:1229. (b) idem (1992) ibid 96:3739. (c) Piecuch P, Oliphant N, Adamowicz L (1993) ibid 99:1875. (d) Piecuch P, Adamowicz L (1994) Chem Phys Lett 221:121. (e) idem (1994) J Chem Phys 100:5792, 5857.
41. Piecuch P, Kucharski SA, Bartlett RJ (1999) J Chem Phys 110:6103
42. Meissner L, Grabowski I (1999) Chem Phys Lett 300:53

Extremal Electron Pairs — Application to Electron Correlation, Especially the R12 Method

Wim Klopper[1] · Werner Kutzelnigg[2] · Hendrik Müller[2] · Jozef Noga[3] · Stefan Vogtner[2]

[1]Theoretical Chemistry Group, Debye Institute, Utrecht University, Padualaan 14, NL-3584 CH Utrecht, The Netherlands.
[2] Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany.
[3] Institute of Inorganic Chemistry, Slovak Academy of Sciences, SK-84236 Bratislava, Slovakia.

The concept of extremal electron pairs is discussed in the context of coupled-cluster theory and the MP2-R12 method. Using extremal pairs the numerical stability of R12-methods is considerably improved, which is demonstrated for CCSD(T)-R12 calculations of the molecules F_2, N_2, and Be_2.

Keywords: Electron correlation, Electron pairs, Perturbation theory, Coupled clusters, R12 method

Topics in Current Chemistry, Vol. 203

1 Introduction

A good first approach to a quantum mechanical system is often to consider one-electron functions only, associating one such function, a 'spin-orbital', with one electron. Most popular are the one-electron functions which minimize the energy in the sense of Hartree–Fock theory. Alternatively one can start from a post-HF wave function and consider the 'strongly occupied natural spin orbitals' (i.e. the eigenfunctions of the one-particle density matrix with occupation numbers close to 1) as the best one-electron functions. Another possibility is to use the Kohn–Sham orbitals, although their physical meaning is not so clear.

If the wave function that one considers, is a single Slater determinant Φ, the spin orbitals φ_i from which Φ, is constructed, are not uniquely determined, but rather there is an infinity of equivalent sets of $\{\varphi_i\}$ related by unitary transformations. To some extent one can make the φ_i unique if one requires either that they are canonical (diagonalize the Fock operator) and are symmetry-adapted, or localized (e.g. according to the criteria of Edmiston and Ruedenberg or Foster and Boys [1–3]). The localized spin orbitals have some advantages both for the chemical interpretation and for the computation of correlation corrections.

An apparent next step is to describe the quantum mechanical state in terms of electron-pair functions, rather than one-electron functions [4]. In fact the concept of electron pairs plays an important role in the theory of the chemical bond.

Again there are two ways to arrive at pair functions. One can either start from a given n-electron wave function and construct pair functions from it, or one can formulate an ansatz for the wave function in terms of pair functions and obtain these by a variational or quasi-variational procedure.

One may construct pair functions from a single Slater determinant Φ or from a post-Hartree–Fock wave function Ψ. At first glance the former choice does not look very useful, since for Φ all information is based on the one-electron functions, so why use a more complicated description in terms of 2-electron functions? Nevertheless the study of pair functions at the independent particle level has both an intrinsic interest and is useful for the preparation of their use at a correlated level [5]. Like the spin-orbitals φ_i associated with the Slater determinant Φ to corresponding pair functions ψ_μ are not uniquely defined and one can transform a given set $\{\varphi_\mu\}$ by any unitary transformation to an equivalent set. One can establish uniqueness — or at least reduce the arbitrariness — by requiring certain extremal properties of the pair functions. This led us to the concept of extremal electron pairs, studied in detail in Part 1 of this series [5].

One possible criteron for extremal pairs is that they maximize the inverse interelectronic distance and hence the Coulomb repulsion within a pair.

$$< \psi_\mu(1,2)|r_{12}^{-1}|\psi_\mu(1,2) >= \max . \tag{1}$$

Another one is that they minimize the square of the interelectronic distance

$$< \psi_\mu(1,2)|r_{12}^{2}|\psi_\mu(1,2) >= \min . \tag{2}$$

These two criteria appear to be generalizations of the localization criteria named after Edmiston–Ruedenberg and Boys respectively [1, 2, 3]. Nevertheless the pair functions ψ_μ, constructed by either of these criteria, are generally symmetry-adapted (in spite of having localized features as well); unlike for the one-electron case there is no strict alternative symmetry-adapted vs. localized [5].

In Paper I of this series [5], the extremal pair functions for the systems He_2, Ne, F^-, HF, H_2O, NH_3 and CH_4 were analyzed. We now follow a different line of thought that was also opened in Paper I, namely to use extremal pairs for the construction of correlated wave functions. We have already pointed out that there is a special set of extremal pair functions associated with MP2 (Møller–Plesset perturbation theory of second order). In fact we have shown, that there are two choices for which the Hylleraas functional of MP2 decomposes exactly into a sum of pair contributions. One choice is the conventional one of pairs of canonical spin orbitals, the other one the use of first-order pairs with extremal norm

$$< \psi_\mu^{(1)}|\psi_\mu^{(1)} >= \text{extremum} \tag{3}$$

which are pairwise orthogonal to each other

$$< \psi_\mu^{(1)}|\psi_\nu^{(1)} >= \delta_{\mu\nu} < \psi_\mu^{(1)}|\psi_\mu^{(1)} > . \tag{4}$$

We could further show that the extremal pairs in the sense of MP2 are, to the leading order of perturbation theory of electron correlation, equal to the first $\binom{N}{2}$ strongly occupied natural spin-geminals, N being the number of electrons.

We begin this paper by shortly recapitulating the concept of extremal pair functions (Sect. 2). Then we consider extremal pair functions in the context of Møller–Plesset perturbation theory (Sect. 3) and coupled-cluster theory (Sect. 4). We then come to the main topic of this paper, the use of extremal pairs in R12-methods. To this end we formulate a new access to R12-theory starting with two-electron systems (Sect. 5) and generalizing it to n-electron systems (Sect. 6). We show then how extremal pairs arise in a natural way in R12-methods (Sect. 7). We finish (Sect. 8) by giving numerical examples which demonstrate the gain in numerical stability by using extremal pairs in R12-calculations.

2 Extremal Pair Functions

Let a Slater determinant Φ, built up from the spin-orbitals φ_i, be given. Consider further some totally symmetric hermitean two-electron operator $\Omega(1,2)$. We wish to determine pair functions of the type (if no summation sign is used, the Einstein summation convention over repeated indices is implied)

$$\psi_\mu(1,2) = \sum_{i<j} b_{ij}^\mu \varphi_{ij}(1,2) = \frac{1}{2} b_{ij}^\mu a^j a^i |0> \tag{5}$$

(with $|0>$ the vacuum state and a^j the creation operator for the spin-orbital φ_j) from the requirement

$$<\psi_\mu(1,2)|\Omega(1,2)|\psi_\mu(1,2)> = \text{extremum} \tag{6}$$

with the subsidiary condition that the ψ_μ are normalized to unity. Condition for this stationarity is the eigenvalue equation

$$\Omega(1,2)\psi_\mu(1,2) = \lambda_\mu \psi_\mu(1,2)\,. \tag{7}$$

The ψ_μ which satisfy (7) can be chosen orthonormal, and they diagonalize Ω, i.e.

$$<\psi_\mu|\psi_\nu> = \delta_{\mu\nu} \tag{8}$$

$$<\psi_\mu|\Omega|\psi_\nu> = \lambda_\mu \delta_{\mu\nu}\,. \tag{9}$$

Assuming the φ_i orthonormal, the b_{ij}^μ in (5) obey the orthogonality relation

$$b_{ij}^\mu b_\mu^{kl} = \sum_\mu b_{ij}^\mu b_\mu^{kl} = \delta_{ij}^{kl};\ \sum_{i<j} b_\mu^{ij} b_{ij}^\nu = \frac{1}{2} b_\mu^{ij} b_{ij}^\nu = \delta_\mu^\nu \tag{10}$$

$$\delta_{ij}^{kl} = \delta_i^k \delta_j^l - \delta_j^k \delta_i^l\,. \tag{11}$$

Although only $i < j$ is needed, it is sometimes advantageous to use also coefficients with $i > j$ (with $b_\mu^{ij} = -b_\mu^{ji}$) and to sum independently over i and j.

The expectation value of Ω is, of course, given as

$$<\Omega> = <\Phi|\Omega|\Phi> = \sum_{\mu=1}^{\binom{N}{2}} <\psi_\mu|\Omega|\psi_\mu> = \sum_{\mu=1}^{\binom{N}{2}} \lambda_\mu = \sum_{i<j=1}^{n} <\varphi_{ij}|\Omega|\varphi_{ij}> \tag{12}$$

If one wants to approximate this expectation value by a sum of less than $\binom{N}{2}$ terms, one gets the best approximation of this kind for $<\Omega> > 0$, if one sums over the extremal pairs with the largest eigenvalues λ_μ.

For Ω spin-independent — which will usually be the case — the ψ_μ will be automatically either singlet or triplet functions. The singlet and triplet blocks can be diagonalized independently.

One possible choice for Ω is

$$\Omega(1,2) = r_{12}^2 \tag{13}$$

The ψ_μ corresponding to this Ω make the square of the interelectronic distance extremal (including the maximum and the minimum). A small λ_μ means a small average interelectronic distance in ψ_μ. Such a pair will much demand the inclusion of electron correlation effects, whereas for a large λ_μ the electrons in ψ_μ may be so distant from each other that there is little need for the inclusion of correlation effects, at least as far as dynamic correlation is concerned.

Other choices of Ω like r_{12}^{-1} were discussed in Paper I [5].

3
Extremal Pair Functions in Møller–Plesset Perturbation Theory

Extremal pairs in the context of Møller–Plesset perturbation theory were already discussed in Paper I of this series [5]. We recapitulate what we will need in this paper.

We use a Fock space notation in terms of excitation operators and matrix elements of operators (using spin-orbitals):

$$a_p^q = a^q a_p = a_q^\dagger a_p;\ a_{pq}^{rs} = a^r a^s a_q a_p = a_r^\dagger a_s^\dagger a_q a_p \tag{14}$$

$$h_q^p = < \varphi_q | h | \varphi_p > \tag{15}$$

$$g_{rs}^{pq} = < \varphi_r(1)\varphi_s(2)|g_{12}|\varphi_p(1)\varphi_q >;\quad g_{12} = r_{12}^{-1} \tag{16}$$

$$\bar{g}_{rs}^{pq} = g_{rs}^{pq} - g_{rs}^{qp}\,. \tag{17}$$

The Fock operator H_0 and the electron interaction $\hat{g}$ are then:

$$H_0 = f_q^p a_p^q;\ f_q^p = h_q^p + g_{iq}^{ip} - g_{qi}^{ip};\ \hat{g} = \sum_{i<j} g_{ij} = \frac{1}{2} g_{rs}^{pq} a_{pq}^{rs}\,. \tag{18}$$

We start from the Hylleraas functional

$$F_2 = < \Psi^{(1)} | H_0 - E_0 | \Psi^{(1)} > + 2\ \mathrm{Re}\ < \Psi^{(1)} | \hat{g} | \Phi > \tag{19}$$

with Φ the Hartree–Fock function, and the first-order function

$$\Psi^{(1)} = \sum_{i<j} \Phi_{ij} = \frac{1}{4} d_{ab}^{ij} a_{ij}^{ab} \Phi\,. \tag{20}$$

In Fock space notation F_2 is

$$\begin{aligned} F_2 \ =\ & \frac{1}{16} d_{kl}^{cd} d_{ab}^{ij} < \Phi | a_{cd}^{kl} (H_0 - E_0) a_{ij}^{ab} | \Phi > \\ & + \frac{1}{2} \mathrm{Re}\ d_{ab}^{ij} < \Phi | \hat{g}\ a_{ij}^{ab} | \Phi > \end{aligned} \tag{21}$$

where Φ is the Hartree–Fock Slater determinant and H_0 the Fock operator. The labels $i, j...$ refer to spin-orbitals occupied in Φ, $a, b...$ to virtual (unoccupied) spin orbitals, and $p, q..$ to arbitrary ones. The Einstein summation convention is implied. If the basis is chosen such that

$$H_0 = \varepsilon_p a_p^p \tag{22}$$

one can rewrite (21) as

$$F_2 = \frac{1}{4} d_{ij}^{ab} d_{ab}^{ij} (\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j) + \frac{1}{2} \mathrm{Re}\; d_{ab}^{ij} \bar{g}_{ij}^{ab} \tag{23}$$

Condition of stationarity with respect to variation of the d_{ij}^{ab} is

$$d_{ab}^{ij} (\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j) + \bar{g}_{ab}^{ij} = 0 \tag{24}$$

and the 2^{nd} order MP energy is

$$E_2 = \frac{1}{4} \mathrm{Re}\; d_{ab}^{ij} \bar{g}_{ij}^{ab} \,. \tag{25}$$

Rather than taking excitations from the pairs $\{i, j\}$ we now consider excitations from general pairs μ and rewrite (20) as

$$\Psi^{(1)} = \sum_{\mu} \Phi_{\mu} \tag{26}$$

$$\Phi_{\mu} = \frac{1}{2} d_{ab}^{\mu} a_{\mu}^{ab};\; a_{\mu}^{ab} = \frac{1}{2} b_{\mu}^{ij} a_{ij}^{ab};\; d_{ab}^{\mu} = \frac{1}{2} d_{ab}^{ij} b_{ij}^{\mu} \tag{27}$$

where b_{ij}^{μ} represents a unitary transformation between the old and new pairs.

In terms of the new pairs, (21) becomes

$$\begin{aligned} F_2 \;=\; & \frac{1}{4} d_{\mu}^{cd} d_{ab}^{\nu} < \Phi | a_{cd}^{\mu} (H_0 - E_0) a_{\nu}^{ab} | \Phi > \\ & + \frac{1}{2} d_{ab}^{\mu} < \Phi | \hat{g} a_{\mu}^{ab} | \Phi > + \frac{1}{2} d_{\mu}^{ab} < \Phi | a_{ab}^{\mu} \hat{g} | \Phi > \\ =\; & \frac{1}{2} d_{\mu}^{ab} d_{ab}^{\mu} (\varepsilon_a + \varepsilon_b) - \frac{1}{2} d_{\mu}^{ab} \varepsilon_{\nu}^{\mu} d_{ab}^{\nu} + \mathrm{Re}\; d_{ab}^{\mu} \bar{g}_{\mu}^{ab} \end{aligned} \tag{28}$$

with

$$\varepsilon_{\nu}^{\mu} = \frac{1}{2} b_{ij}^{\mu} (\varepsilon_i + \varepsilon_j) b_{\nu}^{ij} \,. \tag{29}$$

Condition for stationarity with respect to variation of the d-coefficients is

$$d_{ab}^{\mu} (\varepsilon_a + \varepsilon_b) - \varepsilon_{\nu}^{\mu} d_{ab}^{\nu} + \bar{g}_{ab}^{\mu} = 0 \tag{30}$$

and the MP2 energy is

$$E_2 = \frac{1}{2} \mathrm{Re}\; d_{\mu}^{ab} \bar{g}_{ab}^{\mu} = \sum_{\mu} E_2^{(\mu)} = \varepsilon_{\nu}^{\mu} d_{ab}^{\nu} d_{\mu}^{ab} - d_{ab}^{\mu} d_{\mu}^{ab} (\varepsilon_a + \varepsilon_b) \,. \tag{31}$$

This is similar to (25), with the difference that there is a coupling between the pairs via ε_{ν}^{μ}, as it has been known for the formulation of MP2 in terms of localized molecular orbitals (LMOs).

There are two ways to ensure that E_2 becomes a sum of pair contributions. One is to make ε_{ν}^{μ} diagonal, which is the case for canonical pairs. The other is to impose

$$d_{ab}^{\nu} d_{\mu}^{ab} = \delta_{\mu}^{\nu} D_{\mu} \tag{32}$$

which defines extremal pairs in the sense of MP2 theory. The relation of the extremal pairs satisfying this criterion, to the natural geminals has been shown in Paper I [5].

4
Extremal Pair Functions in Coupled-Cluster Theory

We make the coupled cluster ansatz

$$\Psi = e^S \Phi;\ S = S_1 + S_2 + S_2 + ... \tag{33}$$

for the exact wave function, with Φ a single Slater determinant reference function and S_n the n-particle cluster operators

$$S_1 = S_a^i a_i^a;\ S_2 = \frac{1}{4} S_{ab}^{ij} a_{ij}^{ab}\ etc. \tag{34}$$

It is possible to 'absorb' S_1 into the reference function, by choosing the Brueckner Φ rather than the Hartree–Fock Φ. Alternatively one can 'absorb' S_1 into the Hamiltonian by choosing $\tilde{H} = e^{-S_1} H e^{S_1}$. We assume that the Brueckner Φ has been chosen. Then it is easily seen that

$$E = < \Phi|H|\Phi > + < \Phi|HS|\Phi >= E_0 + \sum_{i<j} e_{ij} \tag{35}$$

$$e_{ij} = \frac{1}{2} < \Phi|HS_{ab}^{ij} a_{ij}^{ab}|\Phi >= \frac{1}{2} S_{ab}^{ij} \bar{g}_{ij}^{ab} \tag{36}$$

i.e. that the correlation energy can exactly be expressed as a sum of pair correlation energies, and that the cluster amplitudes S_{ij}^{ab} etc. can be obtained from a hierarchy of equations, the leading term of which is the non-linear equation

$$< \Phi|a_{ab}^{ij} \left\{ H + [H, S_2] + \frac{1}{2}[[H, S_2], S_2] \right\} |\Phi >= 0 \tag{37}$$

Obviously the pair-cluster function S_2 is invariant with respect to a two-particle transformation, i.e. alternatively to (34) we can express S_2 as

$$S_2 = \frac{1}{2} S_{ab}^{\mu} a_{\mu}^{ab} \tag{38}$$

where μ counts the pairs. As a counterpart of (35, 36) we then get

$$E = E_0 + \sum_{\mu} e_{\mu};\ e_{\mu} = \frac{1}{2} < \Phi|HS_{ab}^{\mu} a_{\mu}^{ab}|\Phi >= \frac{1}{2} S_{ab}^{\mu} \bar{g}_{\mu}^{ab}\,. \tag{39}$$

The correlation energy is again expressed as a sum of pair energies. The counterpart of (37) becomes then

$$< \Phi|a_{ab}^{\mu} \left\{ H + [H, S_2] + \frac{1}{2}[[H, S_2], S_2] \right\} |\Phi >= 0 \tag{40}$$

or

$$< \Phi|a_{ab}^{\mu} H|\Phi > + \frac{1}{2} \sum_{\nu,c,d} S_{cd}^{\nu} < \Phi|a_{ab}^{\mu}[H, a_{\nu}^{cd}]|\Phi >$$

$$+ \frac{1}{8} \sum_{\nu,c,d} \sum_{\kappa,e,f} S_{cd}^{\nu} S_{ef}^{\kappa} < \Phi|a_{ab}^{\mu}[[H, a_{\nu}^{cd}], a_{\kappa}^{ef}]|\Phi >= 0\,. \tag{41}$$

This is a non-linear system of equations for the determination of the amplitudes S_{cd}^{ν}, very similar to the ordinary coupled-pair equations.

As in MP theory it is appropriate to define extremal electron pairs by requiring that the coupling between different pairs in (41) is minimized. As in MP2 theory, a first guess is to require that

$$S_{\mu}^{ab} S_{ab}^{\nu} = \tilde{D}_{\mu} \cdot \delta_{\mu}^{\nu} \tag{42}$$

is diagonal, i.e. that the μ-th and ν-th pair correlation functions are orthogonal to each other — or (equivalently) have extremal norm.

Let us tentatively assume that after introducing extremal pairs the couplings of the different pairs are negligibly small. If there is a justification for an 'independent pair theory' in which one neglects the coupling between different pairs, this will probably be in terms of these extremal pairs. If we neglect the coupling between different pairs in (41) we get

$$\begin{aligned} &< \Phi | a_{ab}^{\mu} H | \Phi > \; + \; \frac{1}{2} \sum_{cd} S_{cd}^{\mu} < \Phi | a_{ab}^{\mu} [H, a_{\mu}^{cd}] | \Phi > \\ &\qquad + \; \frac{1}{8} \sum_{c,d} \sum_{e,f} S_{cd}^{\mu} S_{ef}^{\mu} < \Phi | a_{ab}^{\mu} [[H, a_{\mu}^{cd}] a_{\mu}^{ef}]] | \Phi > \approx 0 \end{aligned} \tag{43}$$

This defines an independent electron pair approximation in terms of extremal pairs, which can be regarded as a generalization of the independent electron pair approximation (IEPA) [4, 8] in terms of pairs (ij) constructed form (preferably) localized orbitals. As in the discussion in Paper I for MP2 [5], one can show that the extremal pairs defined in this section are related to approximate natural geminals corresponding to the coupled-cluster wave function.

As in the case of MP2, the problem is that one must first perform a conventional CCD calculation in order to construct the extremal pairs by applying the criterion (42). Alternatively one can first perform an MP2 calculation, which is much cheaper, transform the canonical pairs to extremal ones in the MP2 sense and then use them for a CCD calculation.

Using extremal pairs in CCD, even if one does not take the optimal ones in the sense of CCD, one gains two things, (a) the couplings between different pairs are diminished, (b) the number of relevant pairs is reduced. Pairs that have only a marginal correlation at the MP2 level do not require a full CCD treatment. This is especially important for large molecules, where the number of relevant pairs scales with N rather than N^2 as the number of canonical pairs, where N is the number of atoms.

5
The Basic Ideas of MP2-R12 for two Electron States

The most important practical use of extremal pairs so far has been in the context of the R12-methods [9]. To explain this we must first outline the essential

features of these. An exact wave function must satisfy the correlation cusp relation

$$\lim_{r_{ij}\to 0}\left(\frac{\partial\Psi}{\partial r_{ij}}\right)_{av} = \frac{1}{2}\Psi(r_{ij}=0) \tag{44}$$

where *av* stands for spherical averaging.

Let us first consider a *two-electron* system.

If one treats electron correlation by perturbation theory, with λ as the formal perturbation parameter, one has

$$\Psi = \Phi^{(0)} + \lambda\Psi^{(1)} + O(\lambda^2) \tag{45}$$

and the cusp relation (for $\Phi^{(0)}$ cuspfree) reads

$$\lim_{r_{12}\to 0}\left(\frac{\partial\Psi^{(1)}}{\partial r_{12}}\right)_{av} = \frac{1}{2}\Phi^{(0)}(r_{12}=0)\,. \tag{46}$$

The ansatz for $\Psi^{(1)}$ to enforce (46) is [10]

$$\Psi^{(1)} = \Phi^{(1)} + c(1-P)r_{12}\Phi^{(0)} \tag{47}$$

where $\Phi^{(1)}$ is a conventional first order wave function of the CI type which satisfies

$$P\Psi^{(1)} = P\Phi^{(1)} = \Phi^{(1)} \tag{48}$$

and where P is the projector onto the CI-space. To fulfill (46) one should choose $c = \frac{1}{2}$, at least for a singlet pair. It is preferable not to fix c too early but to choose it such as to optimize the wave function, in the sense of variation perturbation theory to be discussed now.

Let us insert (47) into the Hylleraas functional (19) (with H_0 the two-electron Fock operator, $\hat{g} = H - H_0$, and $\Phi^{(0)}$ the Hartree–Fock wave function)

$$\begin{aligned} F_2 &= <\Psi^{(1)}|H_0 - E_0|\Psi^{(1)}> + 2\mathrm{Re}\ <\Phi^{(0)}|\hat{g}|\Psi^{(1)}> \\ &= F_{2,conv} + F_{2,R12} \end{aligned} \tag{49}$$

$$F_{2,conv} = <\Phi^{(1)}|H_0 - E_0|\Phi^{(1)}> + 2\mathrm{Re} <\Phi^{(0)}|\hat{g}|\Phi^{(1)}> \tag{50}$$

$$\begin{aligned} F_{2,R12} &= c^2 <\Phi^{(0)}|r_{12}(1-P)(H_0-E_0)r_{12}|\Phi^{(0)}> \\ &\quad + 2c\mathrm{Re}\ <\Phi^{(0)}|g_{12}(1-P)r_{12}|\Phi^{(0)}>\ . \end{aligned} \tag{51}$$

In (49 to 51) we have assumed that $[H_0, P] = 0$, i.e. that the eigenstates of H_0 can with arbitrary accuracy be obtained in the CI space (i.e. that the CI basis is sufficiently saturated at the one-electron level). This implies that the 'mixed term' between 'conv.' and 'R12' vanishes

$$c <\Phi^{(0)}|r_{12}(1-P)(H_0-E_0)|\Phi^{(1)}> = c <\Phi^{(0)}|r_{12}(H_0-E_0)(1-P)|\Phi^{(1)}> = 0 \tag{52}$$

Of course, $F_{2,conv}$ is just the conventional Hylleraas functional, minimized by $\Phi^{(1)}$ while $F_{2,R12}$ is an 'R12-correction'. (51) can be reformulated to

$$\begin{aligned} F_{2,R12} &= c^2 <\Phi^{(0)}|r_{12}(1-P)[T-K, r_{12}]|\Phi^{(0)}> \\ &\quad + c^2 <\Phi^{(0)}|r_{12}(1-P)r_{12}(H_0-E_0)|\Phi^{(0)}> \\ &\quad + 2c\mathrm{Re} <\Phi^{(0)}|g_{12}(1-P)r_{12}|\Phi^{(0)}> \end{aligned} \tag{53}$$

where T is the operator of the kinetic energy and K the exchange operator.

The second term on the r.h.s. of (53) vanishes if $H_0\Phi^{(0)} = E_0\Phi^{(0)}$. We assume this to be the case [as we have already done in deriving (49) to (51)] and we note that

$$[T, r_{12}] = -2g_{12} + 2\hat{U}_{12} \tag{54}$$

$$\hat{U}_{12} = -\frac{1}{2} g_{12}\vec{r}_{12} \cdot (\nabla_1 - \nabla_2)\,. \tag{55}$$

We further neglect the commutator $[K, r_{12}]$, which is justified in the so-called standard approximation A [11]. (For a two-electron closed-shell singlet state a formulation in which K does not arise is always exactly possible [10]). This allows us to rewrite (53) as

$$\begin{aligned} F_{2,R12} \;&=\; -2(c^2 - c)\mathrm{Re} < \Phi^{(0)}|r_{12}(1-P)g_{12}|\Phi^{(0)} > \\ &\quad +\; 2c^2\mathrm{Re} < \Phi^{(0)}|r_{12}(1-P)\hat{U}_{12}|\Phi^{(0)} > \,. \end{aligned} \tag{56}$$

Minimization of this R12-contribution with respect to c leads to

$$\begin{aligned} 2c\Big\{ &-\mathrm{Re}\; < \Phi^{(0)}|g_{12}(1-P)r_{12}|\Phi^{(0)} > \\ &+\mathrm{Re}\; < \Phi^{(0)}|r_{12}(1-P)\hat{U}_{12}|\Phi^{(0)} > \Big\} \\ &+\mathrm{Re}\; < \Phi^{(0)}|g_{12}(1-P)r_{12}|\Phi^{(0)} > \;=\; 0\,. \end{aligned} \tag{57}$$

Equation (57) can be solved for c. The result depends on P. If the basis for the CI is complete, one gets $c = 0/0$, i.e. c is undetermined. If we consider an atomic state and the CI basis is saturated up to some angular momentum quantum number $l-1$, then the first and the last term in (57) — which involve identical matrix elements — go as $(l+\frac{1}{2})^{-4}$, the second one as $(l+\frac{1}{2})^{-6}$ [11]. For a sufficiently large value of l the second term is negligible and from the variational treatment we get $c = \frac{1}{2}$ as required by the cusp condition. This exact result appears here as a consequence of variation-perturbation theory.

The R12-contribution can be interpreted as a basis-incompleteness correction. The key expression bounded between 0 and 1 is

$$\mathrm{Re}\; < \Phi^{(0)}|g_{12}(1-P)r_{12}|\Phi^{(0)} >= 1 - \mathrm{Re}\; < \Phi^{(0)}|g_{12}Pr_{12}|\Phi^{(0)} > \,. \tag{58}$$

If this is large, the R12-correction will matter a lot; if it is small, there will not be much need for an R12 correction.

One has actually to deal with three types of unperturbed two-electron functions $\Phi^{(0)}$.

(a) 'Compact' pairs as in the ground state of He, or H_2 at small internuclear distance. There is a large probability that r_{12} will be small and there is a large dynamical correlation. The variational ansatz (47) is effective even without the projector P. The exponential decay of the wave function takes care that the unphysical behaviour of r_{12} when it is large does not matter. The 'residual interaction' $\hat{U}_{12}$ is 'small' compared to g_{12}. The large-l limit $c = 1/2$ is reached already for small l values. The R12 correction is very effective.

(b) 'Distant' pairs such as for a high Rydberg state of the He atom, or a state of a $1\sigma_g 1\sigma_u$ configuration of H_2 for large internuclear distances. Such a pair is rather well described by the reference function. The probability of two electrons coming close to each other is small. There is neither large dynamical nor non-dynamical correlation. The ansatz (47) without a projector is not effective at all in a variational treatment, because the region of large r_{12} would get too a great weight. Correspondingly, for a small CI space the residual potential $\hat{U}_{12}$ is not small compared to g_{12}. A variationally determined factor c will come out much smaller than 1/2 (for a singlet state). The R12-correction is, so to speak, *not accepted*. Only for a very large CI space, i.e. for very large maximum l, will the limit $c = 1/2$ be reached, but even then the R12 correction will not be very important.

(c) 'Diffuse' pairs such as a σ_g^2 configuration of H_2 at a large internuclear distance. Here non-dynamic correlation effects matter and quasi-degenerate perturbation theory should be applied, i.e. it is more important to include the σ_u^2 configuration rather than an r_{12} term. The R12 correction becomes effective only if one has first taken care over the correct mixing of σ_g^2 and σ_u^2.

6 MP2-R12 for *n*-Electron States

The generalization of MP2-R12 from two-electron to many-electron states is straightforward, at least as long as we use a *single coefficient c* for the entire 'cusp correction'. Instead of (47) we now have

$$\Psi^{(1)} = \Phi^{(1)} + c(1-P)\hat{r}\Phi^{(0)} \tag{59}$$

$$\hat{r} = \sum_{i<j} r_{ij}\,. \tag{60}$$

The counterpart of (51) is

$$\begin{aligned} F_{2,R12} \;=\; & c^2 < \Phi^{(0)}|\hat{r}(1-P)(H_0 - E_0)\hat{r}|\Phi^{(0)} > \\ & +2c\mathrm{Re} < \Phi^{(0)}|\hat{g}(1-P)\hat{r}|\Phi^{(0)} > \,. \end{aligned} \tag{61}$$

The reformulation, similar to that from (53) to (56), is again possible. Hence

$$\begin{aligned} F_{2,R12} \;=\; & -2c^2\mathrm{Re}\; < \Phi^{(0)}|\hat{r}(1-P)\hat{g}|\Phi^{(0)} > \\ & +2c^2\mathrm{Re}\; < \Phi^{(0)}|\hat{r}(1-P)\hat{U}|\Phi^{(0)} > \\ & +c^2\;\mathrm{Re}\; < \Phi^{(0)}|\hat{r}(1-P)[K,\hat{r}]|\Phi^{(0)} > \\ & +2c\mathrm{Re}\; < \Phi^{(0)}|\hat{g}(1-P)\hat{r}|\Phi^{(0)} > \end{aligned} \tag{62}$$

$$\hat{U} = \sum_{i<j} U_{ij}\,. \tag{63}$$

We can minimize (62) with respect to variation of c. We come to this later in this section, after we have expressed (62) in a Fock space formulation. Although both the resulting c and the minimum of $F_{2,R12}$ are of some practical relevance,

we cannot expect that the use of a single coefficient c for all pairs is very effective.

Of course, electrons are indistinguishable, and so are pairs of electrons. Only orbitals or pairs of orbitals are distinguishable. Before we can think of using different coefficients for different pairs we must switch from a representation in terms of the coordinates of the electrons (configuration space) to one in terms of orbitals labels (Fock) space. We now make this change to prepare the use of different c-coefficients, to which we shall only come in Sect. 7.

The formulation of (59,60) in Fock space is (for $\Phi^{(0)}$ a Slater determinant built up from the spin orbitals φ_i)

$$\Psi^{(1)} = \Phi^{(1)} + \frac{c}{4}\bar{R}^{ij}_{k\lambda}a^{k\lambda}_{ij}\Phi^{(0)} = \Phi^{(1)} + \frac{c}{4}(\bar{r}^{ij}_{k\lambda}a^{k\lambda}_{ij} - \bar{r}^{ij}_{pq}a^{pq}_{ij})\Phi^{(0)} \tag{64}$$

where a^{pq}_{ij} is defined by (14) and $\bar{r}^{ij}_{pq}$ is an antisymmetrized matrix element of r_{12}.

$$r^{ij}_{pq} = < \varphi_p(1)\varphi_q(2)|r_{12}|\varphi_i(1)\varphi_j(2) >,\ \bar{r}^{ij}_{pq} = r^{ij}_{pq} - r^{ji}_{pq}\,. \tag{65}$$

The φ_p, φ_q... symbolize an orthonormal basis corresponding to the conventional CI-type approach, while $\varphi_\kappa, \varphi_\lambda$ refer to a complete set. Of course, this notation is meaningful only if sums over complete sets can be performed in closed form.

We use here a formulation in terms of spin-orbitals, which is formally easier, although in actual calulations a formulation in terms of spinfree orbitals, distinguishing between singlet and triplet pairs, is chosen.

Within the *standard approximation* we assume that the given basis is saturated at the one-electron level, which implies that the Hartree–Fock equations are solved exactly and that a Brillouin theorem holds even for virtual orbitals that are not represented in the given basis. Then one can replace (64) by the equivalent expression

$$\Psi^{(1)} = \Phi^{(1)} + \frac{c}{4}\bar{R}^{ij}_{\alpha\beta}a^{\alpha\beta}_{ij}\Phi^{(0)} = \Phi^{(1)} + \frac{c}{4}\left\{\bar{r}^{ij}_{\alpha\beta}a^{\alpha\beta}_{ij} - \bar{r}^{ij}_{ab}a^{ab}_{ij}\right\}\Phi^{(0)}\,. \tag{66}$$

Now α, β, ... refer to virtual orbitals in the complete basis, a, b, c, ... in the given basis. For (62) we get

$$\begin{aligned} F_{2,R12} \;=\; & \frac{c}{2}\mathrm{Re}\bar{R}^{ij}_{\alpha\beta} < \Phi^{(0)}|\hat{g}\; a^{\alpha\beta}_{ij}|\Phi^{(0)} > \\ & + \frac{c^2}{16}\bar{R}^{\alpha\beta}_{ij}\bar{R}^{kl}_{\gamma\delta} < \Phi^{(0)}|a^{ij}_{\alpha\beta}(H_0 - E_0)a^{\gamma\delta}_{kl}|\Phi^{(0)} > \,. \end{aligned} \tag{67}$$

To express (67) in a compact form, it is convenient to define the following quantities [9, 12]. To evaluate these we explicitly use for example $gr = 1$; i.e. $\frac{1}{2}g^{\kappa\lambda}_{kl}r^{ij}_{\kappa\lambda} = \delta^{ij}_{kl}$. We define

$$V^{ij}_{kl} = \frac{1}{2}\bar{g}^{\kappa\lambda}_{kl}\bar{R}^{ij}_{\kappa\lambda} = \frac{1}{2}\bar{g}^{\kappa\lambda}_{kl}\bar{r}^{ij}_{\kappa\lambda} - \frac{1}{2}\bar{g}^{pq}_{kl}\bar{r}^{ij}_{pq} = \delta^{ij}_{kl} - \frac{1}{2}\bar{g}^{pq}_{kl}\bar{r}^{ij}_{pq} \tag{68}$$

$$X^{ij}_{kl} = \frac{1}{2}\bar{R}^{\kappa\lambda}_{kl}\bar{R}^{ij}_{\kappa\lambda} = (r^2)^{ij}_{kl} - \frac{1}{2}\bar{r}^{pq}_{kl}\bar{r}^{ij}_{pq} \tag{69}$$

$$\mathcal{U}_{kl}^{ij} = \frac{1}{2}\bar{R}_{kl}^{\kappa\lambda}\bar{U}_{\kappa\lambda}^{ij} = \frac{3}{2}\delta_{kl}^{ij} - \frac{1}{2}\bar{r}_{kl}^{pq}\bar{U}_{pq}^{ij} \tag{70}$$

$$U_{pq}^{kl} =< \varphi_p(1)\varphi_q(2)|\hat{U}_{12}|\varphi_k(1)\varphi_l(2) >;\ \bar{U}_{pq}^{kl} = U_{pq}^{kl} - U_{pq}^{lk} \tag{71}$$

$$\mathcal{F}_{kl}^{ij} = X_{kl}^{mj} f_m^i + X_{kl}^{im} f_m^j \tag{72}$$

$$Q_{kl}^{ij} = X_{kl}^{rj} g_{rm}^{mi} + X_{kl}^{ir} g_{rm}^{mj} \tag{73}$$

$$\mathcal{B}_{kl}^{ij} = \bar{R}_{kl}^{\alpha\beta} f_\beta^\gamma \bar{R}_{\alpha\gamma}^{ij} \tag{74}$$

$$\tilde{\mathcal{B}}_{kl}^{ij} = \mathcal{B}_{kl}^{ij} - \mathcal{F}_{kl}^{ij}. \tag{75}$$

Here f_q^p is a matrix element of the Fock operator. The quantities (68)–(74 have been defined in our earlier works [9, 12]. The $\tilde{\mathcal{B}}_{kl}^{ij}$ defined by (75) have not been used previously, only the different (though related) quantity $\tilde{\mathcal{B}}_{kl}^{ij}(m,n)$, given later by (102) [12].

For the evaluation of (67) we can then use the fact that

$$\frac{1}{4}\mathrm{Re}\bar{R}_{ij}^{\kappa\lambda} < \Phi^{(0)}|a_{\kappa\lambda}^{ij}\hat{g}|\Phi^{(0)} >= \frac{1}{2}\mathrm{Re}\bar{R}_{ij}^{\kappa\lambda}\bar{g}_{\kappa\lambda}^{ij} = V_{ij}^{ij} \tag{76}$$

$$\begin{aligned}
&< \Phi^{(0)}|a_{\alpha\beta}^{ij}(H_0 - E_0)a_{kl}^{\gamma\delta}|\Phi^{(0)} >= f_\lambda^\kappa < \Phi^{(0)}|a_{\alpha\beta}^{ij}[a_\kappa^\lambda, a_{kl}^{\gamma\delta}]|\Phi^{(0)} > \\
&= f_\lambda^\kappa < \Phi^{(0)}|a_{\alpha\beta}^{ij}\left\{-\delta_k^\lambda a_{\kappa l}^{\gamma\delta} - \delta_l^\lambda a_{k\kappa}^{\gamma\delta} + \delta_\kappa^\gamma a_{kl}^{\lambda\delta} + \delta_\kappa^\delta a_{kl}^{\gamma\lambda}\right\}|\Phi^{(0)} > \\
&= \delta_{kl}^{ij}\left\{f_\varepsilon^\gamma \delta_{\alpha\beta}^{\varepsilon\delta} + f_\varepsilon^\delta \delta_{\alpha\beta}^{\gamma\varepsilon}\right\} - \delta_{\alpha\beta}^{\gamma\delta}\left\{\delta_{ml}^{ij} f_k^m + \delta_{km}^{ij} f_l^m\right\} \\
&= \delta_{kl}^{ij}\left\{f_\alpha^\gamma \delta_\beta^\delta - f_\beta^\gamma \delta_\alpha^\delta + f_\beta^\delta \delta_\alpha^\gamma - f_\alpha^\delta \delta_\beta^\gamma\right\} \\
&\quad -\delta_{\alpha\beta}^{\gamma\delta}\left\{f_k^i \delta_l^j - f_k^j \delta_l^i + f_l^j \delta_k^i - f_l^i \delta_k^j\right\}
\end{aligned} \tag{77}$$

$$\begin{aligned}
&\frac{1}{16}\bar{R}_{ij}^{\alpha\beta}\bar{R}_{\gamma\delta}^{kl} < \Phi^{(0)}|a_{\alpha\beta}^{ij}(H_0 - E_0)a_{kl}^{\gamma\delta}|\Phi^{(0)} > \\
&= \frac{1}{8}\bar{R}_{ij}^{\alpha\beta}\bar{R}_{\gamma\delta}^{ij}\left\{f_\alpha^\gamma \delta_\beta^\delta - f_\beta^\gamma \delta_\alpha^\delta + f_\beta^\delta \delta_\alpha^\gamma - f_\alpha^\delta \delta_\beta^\gamma\right\} \\
&\quad -\frac{1}{8}\bar{R}_{ij}^{\alpha\beta}\bar{R}_{\alpha\beta}^{kl}\left\{f_k^i \delta_l^j - f_k^j \delta_l^i + f_l^j \delta_k^i - f_l^i \delta_k^j\right\} \\
&= \frac{1}{8}\bar{R}_{ij}^{\alpha\beta}\left\{f_\alpha^\gamma \bar{R}_{\gamma\beta}^{ij} - f_\beta^\gamma \bar{R}_{\gamma\alpha}^{ij} + f_\beta^\delta \bar{R}_{\alpha\delta}^{ij} - f_\alpha^\delta \bar{R}_{\beta\delta}^{ij}\right\} \\
&\quad -\frac{1}{8}\bar{R}_{ij}^{\alpha\beta}\left\{f_k^i \bar{R}_{\alpha\beta}^{kj} - f_k^j \bar{R}_{\alpha\beta}^{ki} + f_l^j \bar{R}_{\alpha\beta}^{il} - f_l^i \bar{R}_{\alpha\beta}^{jl}\right\} \\
&= \frac{1}{2}\bar{R}_{ij}^{\alpha\beta} f_\alpha^\gamma \bar{R}_{\gamma\beta}^{ij} - \frac{1}{2}\bar{R}_{ij}^{\alpha\beta}\bar{R}_{\alpha\beta}^{kj} f_k^i \\
&= \frac{1}{2}\mathcal{B}_{ij}^{ij} - \frac{1}{2}X_{ij}^{kj} f_k^i - \frac{1}{2}X_{ij}^{il} f_l^j = \frac{1}{2}(\mathcal{B}_{ij}^{ij} - \mathcal{F}_{ij}^{ij}) = \frac{1}{2}\tilde{\mathcal{B}}_{ij}^{ij}.
\end{aligned} \tag{78}$$

As to the evaluation of $\mathcal{B}_{kl}^{ij}$ [as defined by (74)], there is a difference between the two variants A and B of the *standard approximation*. Approximation A is only used in the context of MP2-R12 as an alternative to approximation B, while in treatments on a higher level of approximation only approximation B is applied. In approximation B all those terms — that arise in the summation over the complete basis — are neglected that in atomic theory decrease as $(L+1)^{-7}$ or faster, while in approximation A even terms that decrease as $(L+1)^{-5}$ are neglected, where L is the highes orbital angular momentum contained in the CI expansion. Note that the error of a conventional CI without r_{12} terms goes as $(L+1)^{-3}$.

In the standard approximation B the result for the matrix elements of $\mathcal{B}$ [9, 11] is

$$\mathcal{B}_{kl}^{ij} = (-V - V^{\dagger} + \mathcal{U} + \mathcal{U}^{\dagger})_{kl}^{ij} + \frac{1}{2}(Q + Q^{\dagger} + \mathcal{F} + \mathcal{F}^{\dagger})_{kl}^{ij} \tag{79}$$

$$\tilde{\mathcal{B}}_{kl}^{ij} = \mathcal{B}_{kl}^{ij} - \mathcal{F}_{kl}^{ij}\,. \tag{80}$$

In approximation A the contribution of Q is ignored. Finally (80) becomes

$$\frac{1}{2}\tilde{\mathcal{B}}_{ij}^{ij} = \begin{cases} U_{ij}^{ij} - V_{ij}^{ij} + \frac{1}{2}Q_{ij}^{ij} & \text{in approximation B} \\ U_{ij}^{ij} - V_{ij}^{ij} & \text{in approximation A}\,. \end{cases} \tag{81}$$

Obviously $F_{2,R12}$ is a sum of diagonal elements

$$F_{2,R12} = \sum_{i<j} f_{ij} \tag{82}$$

$$f_{ij} = cV_{ij}^{ij} + \frac{1}{2}c^2\tilde{\mathcal{B}}_{ij}^{ij}\,. \tag{83}$$

Minimization with respect to c leads to

$$c = -\sum_{i<j} V_{ij}^{ij} / \sum_{k<l} \tilde{\mathcal{B}}_{kl}^{kl} = -Tr(V)/Tr(\tilde{\mathcal{B}}) \tag{84}$$

$$F_{2,R12}(opt) = -\frac{1}{2}(TrV)^2/Tr(\tilde{\mathcal{B}})\,. \tag{85}$$

This is the improvement of the MP2-energy that one would get by including the R12 correction with *a single common factor c* for all pairs. This can be a decent result only if the basis has already been large enough, that all pairs 'want' a factor $c = \frac{1}{2}$. Therefore the value (84) can be used as a test of the quality of the basis. It is currently used under the name 'diagnostic' and implemented in the program.

Actually one has to treat singlet and triplet pairs separately and one gets a different c for the singlet and the triplet part (for a good basis c approaches 1/2 for the singlet part and 1/4 for the triplet part) and independent expressions of type (83) for the singlet and the triplet parts.

7
Arbitrary and Extremal Pairs in MP2-R12 Theory

We postpone our aim to choose different c-factors for different pairs, and consider first the possibility of using other than canonical pairs. Then (64,66) will be replaced by

$$\Psi^{(1)} = \Phi^{(1)} + \frac{c}{2}\bar{R}^{\mu}_{\alpha\beta}a^{\alpha\beta}_{\mu}\,. \tag{86}$$

Note that μ now counts pair functions (rather than basis functions belonging to the complete basis). For (67), in terms of pair functions ψ_μ as defined by (5), we get

$$\begin{aligned} F_{2,R12} &= \frac{c}{2}\mathrm{Re}\ \bar{R}^{\alpha\beta}_{\mu} < \Phi^{(0)}|a^{\mu}_{\alpha\beta}V|\Phi^{(0)} > \\ &\quad + \frac{c^2}{4}\bar{R}^{\alpha\beta}_{\mu}\bar{R}^{\nu}_{\gamma\delta} < \Phi^{(0)}|a^{\mu}_{\alpha\beta}(H_0 - E_0)a^{\gamma\delta}_{\nu}|\Phi^{(0)} > \\ &= \frac{c}{2}\mathrm{Re}\ b^{ij}_{\mu}R^{\alpha\beta}_{ij}d^{\mu}_{kl} < \Phi^{(0)}|a^{kl}_{\alpha\beta}V|\Phi^{(0)} > \\ &\quad + \frac{c^2}{4}b^{ij}_{\mu}b^{\nu}_{kl}\bar{R}^{\alpha\beta}_{ij}\bar{R}^{kl}_{\gamma\delta}b^{\mu}_{mn}b^{op}_{\nu} < \Phi^{(0)}|a^{mn}_{\alpha\beta}(H_0 - E_0)a^{\gamma\delta}_{op}|\Phi^{(0)} > \,. \end{aligned} \tag{87}$$

In view of the orthogonality relations (10) of the b-coefficients, (87) is identical with (67). The counterpart of (82,83) is

$$F_{2,R12} = \sum_{\mu} f_{\mu} \tag{88}$$

$$f_{\mu} = cV^{\mu}_{\mu} + \frac{1}{2}c^2\tilde{\mathcal{B}}^{\mu}_{\mu} \tag{89}$$

with

$$V^{\mu}_{\mu} = \frac{1}{4}b^{kl}_{\mu}V^{ij}_{kl}b^{\mu}_{ij} = 1 - \frac{1}{2}\bar{r}^{pq}_{\mu}\bar{g}^{\mu}_{pq} \tag{90}$$

$$\tilde{\mathcal{B}}^{\mu}_{\mu} = \mathcal{B}^{\mu}_{\mu} - X^{\nu}_{\mu}\varepsilon^{\mu}_{\nu} \tag{91}$$

$$\mathcal{B}^{\mu}_{\mu} = \frac{1}{4}b^{\mu}_{kl}\mathcal{B}^{kl}_{ij}b^{ij}_{\mu} \tag{92}$$

$$\varepsilon^{\nu}_{\mu} = b^{kn}_{\nu}f^{m}_{k}b^{\mu}_{mn} \tag{93}$$

$$X^{\nu}_{\mu} = b^{kl}_{\mu}X^{ij}_{kl}b^{\nu}_{ij}\,. \tag{94}$$

MP2-R12 with a single coefficient c is invariant with respect to a unitary transformation of the pair functions.

Let us now consider the possibility of choosing different c-coefficients for different pairs.

When we first presented the MP2-R12 method [11], we did in fact choose a generalization of the ansatz (64,66)

$$\Psi^{(1)} = \Phi^{(1)} + \frac{1}{4}c_{ij}\bar{R}^{ij}_{\kappa\lambda}a^{\kappa\lambda}_{ij}\Phi^{(0)} \tag{95}$$

with a different c_{ij} for each pair (actually treating singlet and triplet pairs differently). This led to the following generalization of (82,83)

$$F_{2,R12} = c_{ij} V_{ij}^{ij} + \frac{1}{2}(c_{ij})^2 \tilde{\mathcal{B}}_{ij}^{ij} . \tag{96}$$

It turned out that this ansatz is not invariant with respect to a switch from canonical to localized orbitals, and that localized orbitals are actually the better choice [13]. This lack of unitary invariance is, of course, unsatisfactory, and one of us [14] has proposed a generalization in which orbital invariance is established. We come to this later. Let us first apply a pair transformation to (96).

$$\begin{aligned} F_{2,R12} &= c_{ij} b_{ij}^{\nu} V_{\nu}^{\mu} b_{\mu}^{ij} + \frac{1}{2}(c_{ij})^2 b_{ij}^{\nu} \tilde{B}_{\nu}^{\mu} b_{\mu}^{ij} \\ &= c_{\mu}^{\nu} V_{\nu}^{\mu} + \frac{1}{2} c_{\varrho}^{\nu} B_{\nu}^{\mu} c_{\mu}^{\varrho} \end{aligned} \tag{97}$$

with

$$c_{\mu}^{\nu} = \frac{1}{2} b_{\mu}^{ij} c_{ij} b_{ij}^{\nu} . \tag{98}$$

If for one choice of pair functions, $c_{ij}^{ij} = c_{ij}$ is a diagonal matrix, then — in order to get an invariant result — for any other choice of pairs c_{μ}^{ν} will be an arbitrary hermitean matrix. This implies, vice versa, that if we choose a hermitean matrix c_{kl}^{ij} for canonical pairs, there should be one set of pairs for which c_{μ}^{ν} is diagonal. Diagonality of c_{μ}^{ν} is obviously another criterion for extremal pairs. We note that there is a set of pairs, for which the choice of a diagonal c-matrix would mean no loss of generality.

Since there is no indication that the canonical pairs are extremal in this sense, it is appropriate to use a non-diagonal matrix for canonical pairs, i.e. to choose the following ansatz [14] instead of (64,66)

$$\Psi^{(1)} = \Phi^{(1)} + \frac{1}{8} \bar{R}_{\alpha\beta}^{kl} c_{kl}^{ij} a_{ij}^{\alpha\beta} \Phi^{(0)} . \tag{99}$$

With this ansatz the counterpart of (67) is (for an arbitrary choice of the unitary equivalent sets of orbitals)

$$\begin{aligned} F_{2,R12} &= \frac{1}{8} \mathrm{Re}\ c_{ij}^{kl} \bar{R}_{kl}^{\alpha\beta} < \Phi^{(0)} | a_{\alpha\beta}^{ij} \hat{g} | \Phi^{(0)} > \\ &\quad + \frac{1}{64} c_{ij}^{mn} \bar{R}_{mn}^{\alpha\beta} < \Phi^{(0)} | a_{\alpha\beta}^{ij} (H_0 - E_0) a_{kl}^{\gamma\delta} | \Phi^{(0)} > \bar{R}_{\gamma\delta}^{op} c_{op}^{kl} \\ &= \frac{1}{4} c_{kl}^{ij} V_{ij}^{kl} + \frac{1}{16} c_{ij}^{mn} \tilde{\mathcal{B}}_{mn}^{op}(k,l,i,j) c_{op}^{kl} \end{aligned} \tag{100}$$

$$\begin{aligned} \tilde{\mathcal{B}}_{mn}^{op}(i,j,k,l) &= \frac{1}{4} \bar{R}_{mn}^{\alpha\beta} \delta_{kl}^{ij} \left\{ f_{\alpha}^{\gamma} \delta_{\beta}^{\delta} - f_{\beta}^{\gamma} \delta_{\alpha}^{\delta} + f_{\beta}^{\delta} \delta_{\alpha}^{\gamma} - f_{\alpha}^{\delta} \delta_{\beta}^{\gamma} \right\} \bar{R}_{\gamma\delta}^{op} \\ &\quad - \frac{1}{4} R_{mn}^{\alpha\beta} \delta_{\alpha\beta}^{\gamma\delta} \left\{ f_k^i \delta_l^j - f_k^j \delta_l^i + f_l^j \delta_k^i - f_l^i \delta_k^j \right\} \bar{R}_{\gamma\delta}^{op} \end{aligned}$$

$$= \delta^{ij}_{kl}\bar{R}^{\alpha\beta}_{mn}f^{\gamma}_{\alpha}\bar{R}^{op}_{\gamma\beta} - \frac{1}{2}\bar{R}^{\alpha\beta}_{mn}\bar{R}^{op}_{\alpha\beta}\left\{f^i_k\delta^j_l - f^j_k\delta^i_l + f^j_l\delta^i_k - f^i_l\delta^j_k\right\}$$

$$= \delta^{ij}_{kl}\mathcal{B}^{op}_{mn} - X^{op}_{mn}\left\{f^i_k\delta^j_l - f^j_k\delta^i_l + f^j_l\delta^i_k - f^i_l\delta^j_k\right\}. \tag{101}$$

A simplification arises for canonical orbitals, where $f^i_k = \varepsilon_i\delta^i_k$

$$\tilde{\mathcal{B}}^{op}_{mn}(ij) = \delta^{ij}_{kl}\tilde{\mathcal{B}}^{op}_{mn}(i,j,k,l) = \delta^{ij}_{kl}\left\{\bar{R}^{\alpha\beta}_{mn}f^{\gamma}_{\delta}\bar{R}^{op}_{\gamma\delta} - \frac{1}{2}\bar{R}^{\alpha\beta}_{mn}\bar{R}^{op}_{\alpha p}(\varepsilon_i+\varepsilon_j)\right\}$$

$$= \delta^{ij}_{kl}\left\{\mathcal{B}^{op}_{mn} - X^{op}_{mn}(\varepsilon_i+\varepsilon_j)\right\} = \delta^{ij}_{kl}\left\{\tilde{\mathcal{B}}^{op}_{mn} + \mathcal{F}^{op}_{mn} - X^{op}_{mn}(\varepsilon_i+\varepsilon_j)\right\}$$

$$= \delta^{ij}_{kl}\left\{\tilde{\mathcal{B}}^{op}_{mn} + X^{op}_{mn}(\varepsilon_o+\varepsilon_p-\varepsilon_i-\varepsilon_j)\right\} \tag{102}$$

$$F_{2,R12} = \frac{1}{4}c^{ij}_{kl}V^{kl}_{ij} + c^{mn}_{ij}\tilde{\mathcal{B}}^{op}_{mn}(ij)c^{ij}_{op}\,. \tag{103}$$

When this ansatz was first proposed [14] in order to make the theory invariant with respect to pair transformations, the relation to extremal pairs was not yet seen. Rather the ansatz (99) has been justified in terms of the variational principle and it was not restricted to a hermitean c^{kl}_{ij} (which is sufficient to guarantee invariance with respect to pair transformations).

A comment is in order at this point. If we only consider the contributions of the diagonal coefficients c^{ij}_{ij} to (103) we get

$$\frac{1}{4}c^{ij}_{ij}V^{ij}_{ij} + c^{ij}_{ij}\tilde{\mathcal{B}}^{ij}_{ij}(ij)c^{ij}_{ij};\ \tilde{\mathcal{B}}^{ij}_{ij}(ij) = \tilde{\mathcal{B}}^{ij}_{ij} \tag{104}$$

i.e. if we only have a diagonal c-matrix, $\tilde{\mathcal{B}}(ij)$ can be replaced by $\tilde{\mathcal{B}}$.

This suggests that we should consider a variant with $\tilde{\mathcal{B}}^{op}_{mn}(ij)$ replaced by $\tilde{\mathcal{B}}^{op}_{mn}$ but *c not* replaced by its diagonal part. We consider this simplification only for approximation *A*, which is then characterized by

$$\frac{1}{2}\tilde{\mathcal{B}}^{op}_{mn}(ij) = \frac{1}{2}\tilde{\mathcal{B}}^{op}_{mn} = U^{op}_{mn} - V^{op}_{mn} \tag{105}$$

whereas in approximation B we use the full $\tilde{\mathcal{B}}(ij)$

$$\frac{1}{2}\tilde{\mathcal{B}}^{op}_{mn}(ij) = U^{op}_{mn} - V^{op}_{mn} + \frac{1}{2}Q^{op}_{mn}(\varepsilon_o+\varepsilon_p-\varepsilon_i-\varepsilon_j)\,. \tag{106}$$

For internal use we have also defined an approximation A', in which $\tilde{\mathcal{B}}^{op}_{mn}(ij)$ is not replaced by $\tilde{\mathcal{B}}^{op}_{mn}$, but where otherwise approximation A is used. So far we have not yet published results with the variant A'.

If one writes the ansatz (99) just formally, the physical meaning of the matrix elements c^{ij}_{kl} (especially the nondiagonal ones) is not obvious. We only know that within the limit of a large basis [measured e.g. in terms of the diagnostic (84) the diagonal elements c^{ij}_{ij} should reach the value 1/2 (for singlet pairs)], the off-diagonal elements should approach 0. Since in this limit all elements of the V and $\tilde{\mathcal{B}}$ matrices become very small (and hence not very accurate), the

c-matrix will have to be evaluated from an *ill-conditioned* system of equations. It appears more stable from a numerical point of view to find first the extremal pairs (at least approximatively) and to work with a diagonal c matrix. The additional flexibility and the gain of orbital invariance in choosing a full CI matrix rather than a diagonal one has inevitably to be paid for by a loss of numerical stability.

We have hardly observed numerical instabilities on the MP2-R12/A level, but rather often on the higher level of approximations. These numerical instabilities can be avoided if one no longer cares to determine the full matrix c_{kl}^{ij} from the stationarity condition (but rather only its eigenvalues) for eigenvectors (extremal pairs) obtained from other information.

Let us now start from the functional (100) and express this in terms of arbitrary pairs, related to the canonical pairs by the transformation b_{ij}^{μ}

$$\begin{aligned} F_{2,R12} &= \frac{1}{4} c_{kl}^{ij} b_{ij}^{\nu} V_{\nu}^{\mu} b_{\mu}^{kl} + c_{ij}^{mn} \left\{ \mathcal{B}_{mn}^{op} - X_{mn}^{op} (\varepsilon_i + \varepsilon_j) \right\} c_{op}^{ij} \\ &= c_{\mu}^{\nu} V_{\nu}^{\mu} + c_{ij}^{mn} b_{mn}^{\nu} \left\{ \mathcal{B}_{\nu}^{\mu} + X_{\nu}^{\mu} (\varepsilon_i + \varepsilon_j) \right\} b_{\mu}^{op} c_{op}^{ij} \\ &= c_{\mu}^{\nu} V_{\nu}^{\mu} + \frac{1}{2} b_{\varrho}^{ij} c_{ij}^{mn} b_{mn}^{\nu} \mathcal{B}_{\nu}^{\mu} b_{\mu}^{op} c_{op}^{ij} b_{ij}^{\varrho} \\ &\quad - \frac{1}{4} b_{ij}^{\varrho} b_{\varrho}^{kl} c_{kl}^{mn} b_{mn}^{\nu} X_{\nu}^{\mu} (\varepsilon_i + \varepsilon_i) b_{\mu}^{op} c_{op}^{rs} b_{rs}^{\sigma} b_{\sigma}^{ij} \\ &= c_{\mu}^{\nu} V_{\nu}^{\mu} + \frac{1}{2} c_{\varrho}^{\nu} \mathcal{B}_{\nu}^{\mu} c_{\mu}^{\varrho} - \frac{1}{2} c_{\varrho}^{\nu} X_{\nu}^{\mu} c_{\mu}^{\sigma} \varepsilon_{\sigma}^{\varrho} \end{aligned} \tag{107}$$

with

$$c_{\mu}^{\nu} = \frac{1}{4} b_{\mu}^{kl} c_{kl}^{ij} b_{ij}^{\nu}; \; \varepsilon_{\sigma}^{\varrho} = \frac{1}{2} b_{\sigma}^{ij} (\varepsilon_i + \varepsilon_j) b_{ij}^{\varrho} . \tag{108}$$

If we consider especially the approximation A, (107) reduces to

$$F_{2,R12} = c_{\mu}^{\nu} V_{\nu}^{\mu} + \frac{1}{2} c_{\varrho}^{\nu} \tilde{\mathcal{B}}_{\nu}^{\mu} c_{\mu}^{\varrho} . \tag{109}$$

Extremal pairs, which are optimal for the calculation of the R12-correction, are characterized by diagonality of c_{μ}^{ν}, i.e.

$$c_{\mu}^{\nu} = c_{\nu} \delta_{\mu}^{\nu} . \tag{110}$$

For this choice (107) becomes

$$F_{2,R12} = c_{\mu} V_{\mu}^{\mu} + \frac{1}{2} c_{\mu}^{2} \tilde{\mathcal{B}}_{\mu}^{\mu} \tag{111}$$

with

$$\tilde{\mathcal{B}}_{\mu}^{\mu} = 2(U_{\mu}^{\mu} - V_{\mu}^{\mu}) \text{ in Approximation A.} \tag{112}$$

We can now proceed essentially in two ways. The first would be to construct extremal pairs by some given criterion, e.g. by diagonalizing $1/r_{12}$. Then we

construct V_μ^μ and $\tilde{\mathcal{B}}_\mu^\mu$ for these pairs (possibly checking that the off-diagonal elements are actually small). We then minimize with respect to c_μ and get

$$c_\mu = V_\mu^\mu / \tilde{\mathcal{B}}_\mu^\mu \tag{113}$$

The other possibility is to regard diagonality of $F_{2,R12}$ as the criterion for extremal pairs, i.e. to determine the extremal pairs such that

$$\underline{\underline{V}}\, \underline{\underline{\tilde{\mathcal{B}}}}^{-1} \underline{\underline{V}} = \text{diagonal} \tag{114}$$

which is, so to speak, the most general condition for extremal pairs to be used in MP2-R12/A. This is essentially equivalent to choosing $\underline{\underline{c}}$ hermitean and to diagonalize it afterwards.

In MP2-R12/B (and also MP2-R12/A′, which we do not consider now) the situation is a little more complicated because we have to start from (107) and not the simplified form (109).

Then, instead of (111) we get

$$F_{2,R12} = c_\mu V_\mu^\mu + \frac{1}{2} c_\mu^2 \mathcal{B}_\mu^\mu - \frac{1}{2} c_\nu X_\nu^\mu \varepsilon_\mu^\nu c_\mu \,. \tag{115}$$

It is plausible that for this choice also X_ν^μ is close to diagonal. Assuming that X_ν^μ is diagonal if c_ν^μ is diagonal, we get the simpler expression:

$$F_{2,R12} = c_\mu V_\mu^\mu + \frac{1}{2} c_\mu^2 (\mathcal{B}_\mu^\mu - X_\mu^\mu \varepsilon_\mu^\mu) \tag{116}$$

which is a sum of pair contributions.

We finally observe that under the same conditions

$$\begin{aligned} \mathcal{B}_\mu^\mu &= b_\mu^{ij} \mathcal{B}_{ij}^{kl} b_{kl}^\mu = b_\mu^{ij} \left\{ \tilde{\mathcal{B}}_{ij}^{kl} + X_{ij}^{kl} (\varepsilon_i + \varepsilon_j) \right\} b_{kl}^\mu \\ &= \tilde{\mathcal{B}}_\mu^\mu + b_\mu^{ij} (\varepsilon_i + \varepsilon_j) b_{ij}^\varrho b_\varrho^{mn} X_{mn}^{kl} b_{kl}^\mu \\ &= \tilde{\mathcal{B}}_\mu^\mu + \varepsilon_\mu^\varrho X_\varrho^\mu = \tilde{\mathcal{B}}_\mu^\mu + \varepsilon_\mu^\mu X_\mu^\mu \,. \end{aligned} \tag{117}$$

Hence we get, at least, approximatively, (111) again but now with

$$\tilde{\mathcal{B}}_\mu^\mu = 2(U_\mu^\mu - V_\mu^\mu) + Q_\mu^\mu \text{ in Approximation B.} \tag{118}$$

Our experience has been that it is usually sufficient to determine first the extremal pairs that are optimal for MP2-R12/A (standard choice of extremal pairs) and to use them afterwards for MP2-R12/B as well as for all CC-R12 calculations. Some numerical results are given in the next section.

8
Numerical Examples of the Use of Extremal Pairs in R12-Theory

Our early calculations with the R12 method sometimes suffered from numerical instability problems. These arise naturally if one uses very large basis sets, because in the limit of a complete basis the R12-correction becomes undetermined. However, numerical instablities may also arise for relatively small basis sets, namely if the $\mathcal{B}$-matrices that are negative definite when evaluated

Table 1. N2, valence-only correlation energy, in millihartrees, basis 1 (11s6p3d2f)

Method	Diagonal	r_{12}^2	r_{12}^{-1}	Standard	Invariant
MP2-R12-A	−416.389	−416.278	−416.593	−416.902	−417.298
MP2-R12-B	−410.834	−410.743	−410.985	−411.268	−411.856
CCSD-R12	−404.095	−403.899	−404.109	−404.285	−404.659
CCSD[T]-R12	−424.594	−424.492	−424.596	−424.774	−425.148
CCSD(T)-R12	−423.739	−423.537	−423.743	−423.921	−424.294

Diagnostic A (B). S: 0.35 (0.27); T: 0.12 (0.08). Without R12 one gets MP2: −391.705

exactly, lose this property due the application of the standard approximation. A general experience has been that numerical instabilities hardly affect MP2 calculations. They are more serious in CC-methods, where one proceeds iteratively and where a feedback of numerical instability effects is possible. The orbital invariant method is more sensitive to numerical instability effects than the older non-invariant approach.

We have tried several tricks to avoid the numerical instabilities, such as setting some quantities beyond a threshold equal to zero or subjecting the c-coefficients to an upper bound condition. None of these tricks was fully satisfactory, but the use of extremal pairs worked in all cases.

Since it is difficult to construct the extremal pairs for the CC-R12 calculations in question, we have rather constructed the extremal pairs approximatively by some simple criterion, and compared the results of several such choices with the full 'orbital invariant' approach in those cases where it is stable. The fact that one has not the exact extremal pairs (which diagonalize the $\underline{\underline{c}}$-matrix) leads to some loss in the correlation energy, but this is in most cases negligibly small.

In Tables 1 to 4 we display the results of MP2-R12, CCSD-R12 and CCSD(T)-R12 calculations of F_2 and N_2 for various options of the pairs:

(a) Diagonal: a diagonal c-matrix in terms of canonical orbitals;
(b) r_{12}^2: extremal pairs in terms of the criterion $\Omega(1,2) = r_{12}^2 - r_{12}Pr_{12}$, with P the projector to the given basis;
(c) r_{12}^{-1}: extremal pairs in terms of the criterion $\Omega(1,2) = r_{12}^{-1}$;

Table 2. F2, valence-only correlation energy in millihartrees, basis 1 (11s6p3d2f)

Method	Diagonal	r_{12}^2	r_{12}^{-1}	Standard	Invariant
MP2-R12-A	−610.600	−612.260	−612.722	−612.876	−613.002
MP2-R12-B	−595.874	−598.029	−598.762	−598.946	−599.571
CCSD-R12	−599.056	−600.125	−600.310	−600.382	−600.868
CCSD[T]-R12	−620.696	−621.742	−621.923	−621.996	−622.478
CCSD(T)-R12	−619.869	−620.920	−621.101	−621.174	−621.658

Diagnostic A (B). S: 0.40 (0.29); T: 0.16 (0.12). Without R12 one gets MP2: −560.083

Table 3. N2, all electrons correlated, correlation energy in millihartrees

Method	Basis	Diagional	r_{12}^2	r_{12}^{-1}	Standard	Invariant
MP2-R12-A	1	−527.430	−528.307	−529.742	−530.240	−530.926
	2	−535.445	−535.845	−535.985	−536.049	−536.165
	3	−537.441	−537.497	−537.524	−537.542	−537.572
MP2-R12-B	1	−519.777	−522.385	−520.645	−521.552	−521.102
	2	−529.708	−530.116	−530.277	−530.335	−530.295
	3	−533.516	−533.563	−533.619	−533.621	−533.877
CCSD-R12	1	−516.831	−517.507	−517.940	−518.286	—
	2	−521.910	−522.268	−522.328	−522.364	—
	3	−523.659	−523.669	−523.694	−523.680	—
CCSD(T)-R12	1	−537.322	−537.989	−538.415	−538.765	—
	2	−542.796	−543.149	−543.208	−543.245	—
	3	−544.725	−544.730	−544.755	−544.740	—

Diagnostic A (B), basis

1, S: 0.30 (0.24) T: 0.09 (0.06)
2, S: 0.40 (0.34) T: 0.14 (0.09)
3, S: 0.43 (0.38) T: 0.18 (0.14)

Basis

1: 11s6p4d3f (aug-cc-pVTZ +S(1d1f))
2: 13s7p5d4f (aug-cc-pVQZ +S(1d1f))
3: 15s9p6d5f (aug-cc-pV5Z +S(1d1f))

Table 4. F2, all electrons correlated, correlation energy in millihartrees

Method	Basis	Diagonal	r_{12}^2	r_{12}^{-1}	Standard	Invariant
MP2-R12-A	1	−734.589	−735.937	−736.774	−737.597	−738.242
	2	−738.749	−739.552	−740.353	−740.588	−740.953
MP2-R12-B	1	−716.560	−718.265	−720.011	−721.030	−722.221
	2	−728.122	−729.222	−730.427	−730.690	−731.053
CCSD-R12	1	−725.496	−726.663	−727.087	−727.569	−727.644
	2	−726.277	−726.698	−727.139	−727.267	−727.301
CCSD(T)-R12	1	−747.054	−748.208	−748.628	−747.054	−749.227
	2	−748.324	−748.740	−749.178	−749.305	−749.346

Diagnostic A (B), basis

1, S: 0.35 (0.24) T: 0.12 (0.06)
2, S: 0.40 (0.32) T: 0.15 (0.09)

Basis

1: 11s6p4d3f (aug-cc-pVTZ +S(1d1f)
2: 13s7p5d4f (aug-cc-pVQZ +S(1d1f)

(d) Standard: canonical pairs which diagonalize the R12 contributions to the MP2-R12 energy;
(e) Invariant: the orbial invariant method with a general form of the *c*-matrix.

We also document the diagnostic, as defined by (84), at the MP2-R12/A and MP2-R12/B levels. For a saturated basis the diagnostic should reach 0.5

for a singlet pair and 0.25 for a triplet pair. The convergence to these limiting values is faster for MP2-R12/A than for MP2-R12/B. One makes the following observations:

(1) With increase of the basis size the different pair variants appear to converge to the same results.

(2) Among the different choices of extremal pairs, the choice 'standard', which diagonalizes the R12-contribution to MP2-R12/A is closest to the orbital invariant approach 'invariant', provided that the latter converges . The choice 'r_{12}^{-1}' is only slightly poorer, while r_{12}^{2}' is slightly more off, but definitely better than 'diag' with a diagonal *c*-matrix for canonical orbitals.

(3) The only case documented here, where the orbital-invariant approach diverges is that for the all-electron calculations of N_2 at CCSD-R12 and CCSD(T)-R12 levels, for all three basis sets. Even for this example the MP2-R12/A and MP2-R12/B results are numerically stable, and in valence-only calculations even CCSD-R12 and CCSD(T)-R12 converge. Our general experience is that the approaches based on extremal pairs don't suffer from numerical instabilities, at variance with the orbital-invariant method. To avoid divergencies we use 'standard' pairs as default option.

Acknowledgements. Support by the Alexander-von-Humboldt foundation, Fonds der Chemie, the EC programs 'Human Capital and Mobility' and COST, the Slovak agency VEGA (4227) and a fellowship to WKl by the Royal Netherlands Academy of Arts and Sciences are gratefully acknoweledged.

9 References

1. Edmiston C, Ruedenberg K (1963) Rev Mod Phys 33:457; (1965) J Chem Phys 43, 597
2. Foster JM, Boys SF (1960) Rev Mod Phys 32:296
3. What is now generally referred as due to Foster and Boys was actually first formulated by Edmiston and Ruedenberg with whose names the other criterion is associated correctly.
4. Kutzelnigg W, (1977) in: Schaefer HF (ed) Modern Theoretical Chemistry, Vol. IIIa. Plenum, New York
5. Vogtner St, Kutzelnigg W. (1996) Int J Quant Chem 60:235
6. Coleman AJ (1963) Rev Mod Phys 35:668
7. Löwdin PO (1955) Phys Rev 97:1474, 1490, 1589
8. Kutzelnigg W (1991) Theoret Chim Acta 80:349
9. Noga J, Klopper W, Kutzelnigg W (1997) in: Bartlett RJ (ed.) Recent Advances in Coupled Cluster Methods. World Scientific, Singapore, p 1
10. Kutzelnigg W (1985) Theoret Chimica Acta 68:445
11. Kutzelnigg W, Klopper W (1991) J Chem Phys 94:1985
12. Noga J, Kutzelnigg W (1994) J Chem Phys 101:7738
13. Klopper W, Kutzelnigg W (1991) J Chem Phys 94:2020
14. Klopper W (1991) Chem Phys Lett 186:583

Many-Body Perturbation Theory with Localized Orbitals — Kapuy's Approach

János Pipek[1] · Ferenc Bogár[2]

[1]Department of Theoretical Physics, Institute of Physics,
Technical University of Budapest, H–1521 Budapest, Hungary
[2]Department of Theoretical Physics,
József Attila University, H–1525 Szeged, Hungary

In this contribution the application of localized molecular orbitals for the separation of local and long-range correlation effects in extended systems is studied in the framework of the many-body perturbation theory. We first summarize the basic ideas developed by Professor Kapuy for extending diagrammatic methods based on localized one-electron states in correlation energy calculations. After describing some possible ways for characterizing the extension and separation of localized MOs we give a flexible procedure for the truncation of long-range correlation effects with the remarkable property that the range of the Coulomb interaction is still kept infinite. Analyzing numerical results the convergence of localization corrections is discussed and the separation of local correlation terms show that only the immediate neighborhood of a localized MO plays a considerable role in excitation processes.

Keywords: localized orbitals, many-body perturbation, local correlation, long-range correlation, extended systems

Topics in Current Chemistry, Vol. 203

1 Introduction

Independent particle (Hartree–Fock) theory has played a historical role in understanding the electronic structure of well known chemical species, small and medium sized molecules and the basic features of chemical bonding. Although the computational complexity of the Hartree–Fock calculation increases as N^4 with the number of electrons N, the unbelievable development of computational hardware devices as well as the refinement of theoretical methods triggered a continuous extension of the manageable size of the systems available for the theoretical description. Such a development, however, required tremendous efforts and already in the early sixties ideas had emerged about introducing useful chemical concepts like locality into the theory in order to ease and speed up calculations and at the same time make them more transparent. The need for this kind of simplification was given even larger emphasis as the collected independent particle results indicated that the quality of the Hartree–Fock calculations becomes poorer with increasing system size or if molecular interactions are targeted. Indeed, a naive physical picture suggests that the long-range part of the electron-electron interaction should be satisfactorily described by the effective averaged Coulomb and exchange interaction terms of the Hartree–Fock theory, as at large separation electrons hopefully do not "see" the details of the motion of others. This consideration indicates a natural connection between the remaining interaction not taken into account in independent particle models (commonly known as *electron correlation*), and localizability. Application of *localized orbitals* to the theoretical description of extended systems has always played a central role in the work of Professor Kapuy, who first studied transferability of local energy contributions and the localized orbitals themselves within the framework of the Hartree–Fock model [1–5] and later extended the concept of locality to electron correlation calculations mainly in the field of many-body perturbation theory (MBPT) [6, 7]. The main goal of this contribution is to summarize the basic ideas developed by Professor Kapuy for a possible extension of the diagrammatic MBPT using localized molecular orbitals and to show numerical experiences for convergence features of the perturbation expansion with localized orbitals, as well as to present some evidence for the expected negligibility of long-range correlation effects.

2 Concept of Locality

The existence of local properties in extended molecules has been known by chemists for a long time; especially in organic chemistry the use of relatively independent functional groups like $-CH_3$ or $-OH$, etc., as building blocks of the system is a well-working everyday practice. Also, the success of thermodyna-

mics and statistical physics by dividing large systems to (almost) independent non-interacting subsystems underlines the dominating role of local effects and the negligible amount of long-range energy contributions, at least far from phase transitions.

Quantum chemistry, however, seems to contradict to the above well known facts, as both the N-electron wavefunction $\Psi(x_1, \ldots, x_N)$ and the canonical Hartree–Fock one-particle orbitals $\varphi_i(x)$, $(i = 1, \ldots, N)$ are usually extended to the whole volume of the atomic system. Here, variable $x_j = (\mathbf{r}_j, \sigma_j)$ is used as a combined notation for the space and spin variables of electron j, and the independent particle wavefunction is constructed by the usual determinant formula [8] from one-particle functions

$$\Psi(x_1, \ldots, x_N) = (N!)^{-1/2} \det[\varphi_i(x_j)]. \tag{1}$$

The optimum set of one-particle orbitals $\{\varphi_i(x),\ (i = 1, \ldots, N)\}$ is determined from the minimum condition of the energy expectation value [9, 10]

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{2}$$

with the additional constraint of keeping the molecular orbitals (MOs) orthonormal $\langle \varphi_i | \varphi_j \rangle = \delta_{ij}$, $(i, j = 1, \ldots, N)$. By introducing Lagrangian multipliers f_{ij} for each of the above constraints one arrives at the complicated integro-differential coupled system of equations

$$F\,\varphi_i = \sum_{j=1}^{N} f_{ji}\,\varphi_j. \tag{3}$$

A formal decoupling of these equations is possible, however, as the Fock-operator

$$F = -\frac{1}{2}\Delta + V_n + V_C - V_X, \tag{4}$$

has the following remarkable invariance property. Although (besides the constant differential operator Δ and the multiplication operator of the nuclear-electron interaction V_n) the Coulomb and exchange operators V_C and V_X contain all unknown φ_i orbitals in an integral form, all terms of the Fock-operator, as well as the N-electron wavefunction $\Psi(x_1, \ldots, x_N)$, are invariant against a unitary transformation of the MOs

$$\varphi_i' = \sum_{j=1}^{N} U_{ji}\,\varphi_j, \qquad UU^{\dagger} = 1. \tag{5}$$

This fact can be used for choosing a specific orbital set $\{\varphi_i^{\mathrm{can}}(x),\ (i = 1, \ldots, N)\}$ for which the Lagrangian multiplier matrix f_{ij} becomes diagonal, leading to the canonical form of the Hartree–Fock equations

$$F\,\varphi_i^{\mathrm{can}} = \varepsilon_i\,\varphi_i^{\mathrm{can}}. \tag{6}$$

Canonical MOs have the definite advantage that the diagonal Fock-matrix elements (or Lagrangian multipliers) $\langle \varphi_i | F | \varphi_i \rangle = \varepsilon_i$, $(i = 1, \ldots, N)$ can be inter-

preted as approximate ionization energies of the orbitals φ_i [11]. We wish to emphasize, however, that no additional reason forces us to choose the canonical orbital set; all other equivalent sets of the form (5) can be used to describe molecular properties in the most convenient way.

Due to their relative mathematical simplicity canonical Hartree–Fock equations are used to solve the independent particle problem in an iterative, self-consistent-field (SCF) way. Since in each iterative step Eq. (6) behaves as an eigenvalue problem, solutions are not restricted to a finite set of so-called *occupied* orbitals, $\texttt{occ} = \{\varphi_i^{\text{can}},\ (i = 1, \ldots, N)\}$, but an infinite set, $\texttt{virt} = \{\varphi_a^{\text{can}},\ (a = N+1, \ldots, \infty)\}$, of *virtual* orbitals appears which does not contribute to the N-electron wavefunction Ψ. In practical calculations the MOs are expanded using a finite set of atomic orbitals (LCAO), and the set virt of MOs is also restricted to a finite number of orbitals. After convergence a unitary transformation of the form (5) with a properly chosen matrix U is used in order to achieve chemically well-interpretable molecular orbitals. This process is often called *localization*; however, in principle, any other physical or chemical criterion can be prescribed for the resulting MOs, e.g. finding the most delocalized orbitals [12], etc. Localization in the above sense has to be strictly distinguished from the Anderson-type localization phenomenon in disordered systems [13], where due to irregularities of the one-electron Hamiltonian F the canonical MOs themselves are localized in an exponential way to a given part of the whole system. In chemical species the canonical orbitals are usually distributed over a substantial part of the molecule and there exist unitary transformations which result in considerably more compact MOs.

As these compact orbitals are usually concentrated in the regions where core electrons, bonds, non-bonding electron pairs are expected, the introduction of localized MOs has played an important role in understanding the above classical chemical concepts and valence. Most of the localized orbitals, on the other hand, are also well displaced in space from each other. This property makes them an excellent tool for separating local and long-range electron interaction effects.

Finding a localization transformation is usually done by satisfying a given optimum criterion. Some well-known optimum criteria can be summarized as searching for the maximum or minimum value of the functional

$$G\{\varphi_i\} = \sum_{i=1}^{N} \langle ii|\Omega|ii\rangle \tag{7}$$

where the matrix elements of the two-electron operator $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ are defined by the expression

$$\langle ij|\Omega|kl\rangle = \int \varphi_i^*(\mathbf{r}_1)\, \varphi_j(\mathbf{r}_1)\, \Omega(\mathbf{r}_1, \mathbf{r}_2)\, \varphi_k^*(\mathbf{r}_2)\, \varphi_l(\mathbf{r}_2) \mathbf{d}r_1\, \mathbf{d}r_2\,. \tag{8}$$

The earliest approach to localized orbitals developed by Foster and Boys [14–16] applies the two-electron operator

$$\Omega^B(\mathbf{r}_1, \mathbf{r}_2) = (\mathbf{r}_1 - \mathbf{r}_2)^2 , \tag{9}$$

and functional G of (7) (i.e. the spatial extension of the MOs) is to be minimized, while the method of Edmiston and Ruedenberg [17–19] uses

$$\Omega^{ER}(\mathbf{r}_1, \mathbf{r}_2) = |\mathbf{r}_1 - \mathbf{r}_2|^{-1} , \tag{10}$$

and seeks the maximum of the corresponding functional G, i.e. the total self-repulsion energy of the electrons. A similar, but not so widely applied localization procedure of von Niessen [20] maximizes the charge density overlap functional G with

$$\Omega^N(\mathbf{r}_1, \mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_2) . \tag{11}$$

The main disadvantage in all of the above cases is represented by the fact that the computational complexity of the methods grows in proportion to N^5. In the specific case, however, of Boys' procedure a fortunate transformation of criterion (7) with (9) leads to the maximization of the equivalent optimum criterion

$$G^{B2}\{\varphi_i\} = \sum_{i=1}^{N} [\langle i|\mathbf{r}|i\rangle]^2 , \tag{12}$$

where only matrix elements of the one-electron position operator $\mathbf{r}$ appear, resulting in a much more favorable algorithmic behavior of N^3.

The same advantageous algorithmic complexity characterizes the so-called population localization method of Pipek and Mezey [21], where the functional of the form

$$G\{\varphi_i\} = \sum_{i=1}^{N} \sum_{A=1}^{n} [\langle i|P_A|i\rangle]^2 \tag{13}$$

is to be maximized. Here, the second summation runs over all atoms A of the molecule and the expectation value of the one-electron operator P_A gives the population of orbital φ_i on atom A. By an appropriate definition of the atomic population operator P_A the quantities

$$Q_A^i = \langle i|P_A|i\rangle \tag{14}$$

can be identified either as Mulliken's gross atomic populations [22, 23], or populations of Löwdin's symmetrically orthogonalized basis sets [24, 25], or even Bader's charges [26, 27], etc. Using Mulliken's populations, however, leads to easily evaluable, handy formulas.

Interestingly, and probably due to a very exciting connection between the Fermi-hole and the localized orbitals [28], various localization methods result in rather similar localized orbitals, except for the description of double bonds by a σ- and π-orbital-pair or two equivalent τ (banana) bonds. Boys' localization gives τ orbitals, while the Edmiston–Ruedenberg and the popula-

tion localization method leads to $\sigma - \pi$ separation, which is considered to be chemically preferable.

Some other alternatives (e.g. the procedure of Magnasco and Perico [29]) for solving the localization problem are not discussed here, as these unfortunately have the theoretical disadvantage that external requirements like *a priori* definitions of bonds, lone pairs, and core orbitals have to be introduced. A more detailed analysis of various localization methods can be found in Ref. [21].

2.1 Characterization of Localized Orbitals

The most popular way for visualizing MOs is the density or wave function contour plot. We can also introduce other quantities that can measure the position and the spatial extension of an LMO. The position of an LMO can be characterized by the so-called orbital centroid, the expectation value of the position of an electron on the given LMO, $\bar{\mathbf{r}}_i = \langle \varphi_i | \mathbf{r} | \varphi_i \rangle$. The spatial extension, the size of an LMO can be measured by the dispersion of electron coordinates placed on a given LMO

$$(\Delta r_{\alpha i})^2 = \langle \varphi_i | r_\alpha^2 | \varphi_i \rangle - \langle \varphi_i | r_\alpha | \varphi_i \rangle^2 \qquad \alpha = x, y, z$$

that is coordinate-system-dependent. If we want to choose coordinate-system-independent quantities, we can use the second moment matrix

$$[S_i]_{\alpha\beta} = \langle \varphi_i | (r_\alpha - \bar{r}_\alpha)(r_\beta - \bar{r}_\beta) | \varphi_i \rangle .$$

The ellipsoid determined by the $[S_i]^{\frac{1}{2}}$ and the volume of it proportional to $\det\left([S_i]^{\frac{1}{2}}\right)$ characterize the position and the spatial extension of the LMOs. The latest quantity is called orbital size [54, 59].

In our calculations Boys' localization procedure was applied to localize the occupied and virtual orbitals in two separate blocks. The properties of the localized virtual orbitals of normal saturated hydrocarbons are investigated in detail in [65] and for all-trans conjugated polyenes in [68]. In Fig. 1 the schematic plots of ellipsoids of selected occupied and virtual orbitals of C_5H_{12} in the 6-31G* basis set are presented in the plane of the CC bonds. Only two characteristic sets of LMOs are selected, a terminal CH bond and a CC bond. The left panel shows the occupied and the right the virtual LMOs. The plots of occupied and virtual LMOs in minimal basis are very similar to case (a) and they are not presented here. We can see that the localized orbitals are well separated and their spatial extent is restricted to a small part of the molecule even for the diffuse ϕ_{v2} and ϕ_{v3} LMOs. The virtual LMOs are spread along the axis of chemical bonds and they are not restricted to the region between the atoms.

In Table 1 the sizes of LMOs presented in Fig. 1 are shown. Using minimal basis the sizes of the virtual LMOs are approximately the same as the size of occupied LMOs. For the 6-31G* basis set we can also find virtual orbitals that

Table 1. The effective volume of LMOs presented in Fig. 1

	CH		CC	
	STO-3G	6-31G*	STO-3G	6-31G*
ϕ_{o1}	0.63	0.69	0.66	0.69
ϕ_{v1}	0.65	0.33	0.69	0.39
ϕ_{v2}		2.03		2.12
ϕ_{v3}		2.12		2.12
ϕ_{v4}		0.52		0.46
ϕ_{v4}				0.50

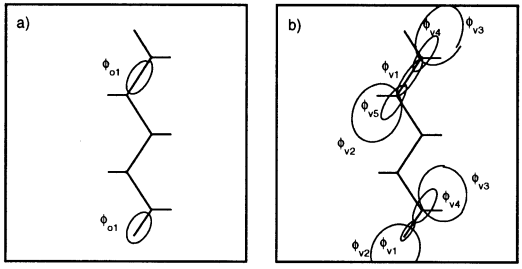

Fig. 1. Schematic plots of ellipsoids of a) occupied and b) virtual LMOs belonging to the CH and CC bonds of a C_5H_{12} molecule in the 6-31G* basis set

have sizes similar to the occupied orbitals, but orbitals that are tree times larger also appear.

3
Localized Many-Body Perturbation Theory

For treating the electron correlation problem many alternative approaches have appeared in the literature, but basically three main streams (and variants of them) can be distinguished.

The *full configuration interaction* method [34–36] is exact in the sense that after choosing appropriate atomic basis functions (defining the model in this way), the resulting many-electron wavefunction is an exact eigenfunction of the model Hamiltonian, the computational effort, nevertheless, increases in an exponential manner. Truncation of the full CI expansion (especially after single and double excitations, CI-SD) considerably reduces the necessary computational resources, but leads unfortunately to the serious problem of *non-size-consistency* [37, 38] which makes the results even for medium systems unrealistic. The *coupled-cluster* method [39, 40] theoretically properly describes extended systems as well, but numerous experiences show the enormous increase of computational work with the size of the system.

Perturbative approaches to the electron correlation problem have proved to be successful even when calculating second-order corrections only; but finer results require fourth order energy corrections, as we will see later in this text. The reliability of a perturbation expansion greatly depends on the partitioning of the exact Hamiltonian $H = H_0 + W$ to an unperturbed part H_0 and a perturbation W. Good quality approximations are to be expected if the N-particle operator H_0 is chosen as the sum of equivalent one-particle Fock operators

$$H_0(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{i=1}^{N} F(\mathbf{r}_i)\,, \qquad W = H - H_0\,. \tag{15}$$

The above Mœller–Plesset (MP) [50] partitioning scheme is based on the assumption that Hartree–Fock theory is adequate for describing the most important electron interaction effects. The mathematical form (15) of the unperturbed Hamiltonian corresponds to a non-interacting system which has the important consequence that all eigenfunctions of H_0 are exactly known, i.e. Slater-determinants built from canonical MOs (eigenfunctions of F). As we will see later this is a prerequisite of a perturbation theory based on the unperturbed Hamiltonian H_0. Rayleigh–Schrödinger perturbation theory formulas worked out for N-electron operators of the form (15) are, however, extremely complicated, hard to systemize and contain at each perturbative order terms which are exactly cancelled by higher-order perturbative corrections. A systematic and economic solution to these problems is presented by the so-called *many-body perturbation theory* (MBPT) where the Hamiltonian is rewritten using creation and annihilation operators of the one-particle MOs and elegant algorithmic methods, using graphical representations for the various perturbative corrections, have been worked out which automatically take care of cancellations in the perturbative expansion. For details we refer to the literature [41–43].

Introducing localized one-electron MOs into the perturbation theory does not seem to present any danger at the first sight, as the Fock-operator and consequently H_0 is invariant to canonical $\rightarrow$ localized transformations. A serious complication arises, however, because of the fact that the eigenfunctions of H_0 can *not* be written as simple Slater-determinants of the localized molecular orbitals any more. We have now the options either expressing the eigenfunctions of H_0 (i.e. Slater-determinants made of canonical MOs) as complicated linear combinations of determinants of LMOs or changing the unperturbed Hamiltonian to an operator with simple LMO-determinant eigenfunctions.

The first choice seems to be more natural since, H_0 being invariant, the partitioning scheme remains untouched of Mœller–Plesset type. The price to be paid for this principal simplicity, however, is high in calculational details, as the well-developed, systematic many-body graphical algorithms are *not* applicable if the unperturbed eigenfunctions bear a complicated structure. In a series of papers [44-48], Pulay and Sæbø developed formulas for the second- and third- and fourth-order perturbative corrections with localized orbitals using a CEPA-

based formalism. They also describe theoretical results for CI and coupled-cluster treatments. Their basic idea is that after localizing the occupied orbital space, the virtual orbitals are represented by atomic orbitals and excitations are allowed only from a localized MO to the AOs of its immediate neighborhood (domain). Computational results are fine, but the development of formulas to any higher order of perturbation theory requires extra efforts that are not algorithmizable.

3.1 Kapuy's Approach

Following the early footsteps of Amosh and Musher [51], and Davidson [52], Professor Kapuy suggested and developed [53] a many-body procedure for calculating ground-state energy, Green's function and cluster operator perturbative corrections. Formulas for corrections at any order of the perturbation expansion can be algorithmically generated and computerized. A summary of the various diagram generation methods can be found in Ref. [49]. The formalism of Professor Kapuy's approach is summarized briefly in this section.

The method uses the idea of redefining the unperturbed Hamiltonian outlined in the previous section. In the first step, occupied and virtual canonical orbitals are localized separately by the unitary matrices U and Q as

$$\varphi_i^{\mathrm{loc}} = \sum_j^{\mathtt{occ}} U_{ji}\, \varphi_j^{\mathrm{can}}, \qquad \varphi_a^{\mathrm{loc}} = \sum_b^{\mathtt{virt}} Q_{ba}\, \varphi_b^{\mathrm{can}} \tag{16}$$

where the first summation runs over the occupied canonical orbital set `occ` and the second one on the virtual set `virt`. The resulting orbitals belong to the localized occupied set, $\mathtt{locc} = \{\varphi_i^{\mathrm{loc}},\ (i = 1, \ldots, N)\}$, or to the localized virtual set, $\mathtt{lvirt} = \{\varphi_a^{\mathrm{loc}},\ (a = N+1, \ldots, M)\}$, where M is the number of atomic basis functions.

The Fock operator can be represented both by canonical and localized MOs as

$$\begin{aligned} F &= \sum_i^{\mathtt{occ}} |\varphi_i^{\mathrm{occ}}\rangle\, \varepsilon_i\, \langle\varphi_i^{\mathrm{occ}}| + \sum_a^{\mathtt{virt}} |\varphi_a^{\mathrm{occ}}\rangle\, \varepsilon_a\, \langle\varphi_a^{\mathrm{occ}}| \\ &= \sum_{i,j}^{\mathtt{locc}} |\varphi_i^{\mathrm{loc}}\rangle\, f_{ij}\, \langle\varphi_j^{\mathrm{loc}}| + \sum_{a,b}^{\mathtt{lvirt}} |\varphi_a^{\mathrm{loc}}\rangle\, f_{ab}\, \langle\varphi_b^{\mathrm{loc}}| \end{aligned} \tag{17}$$

with the Lagrangian multipliers $\varepsilon_i, \varepsilon_a, f_{ij}, f_{ab}$ defined by the expressions (6) and (3). Note, that off-diagonal Fock matrix elements between occupied and virtual orbitals do not appear in (17), as the sets `occ` and `virt` are localized separately. The diagonal part of the localized representation of F can be separated as

$$F^d = \sum_i^{\mathtt{locc}} |\varphi_i^{\mathrm{loc}}\rangle\, f_{ii}\, \langle\varphi_i^{\mathrm{loc}}| + \sum_a^{\mathtt{lvirt}} |\varphi_a^{\mathrm{loc}}\rangle\, f_{aa}\, \langle\varphi_a^{\mathrm{loc}}| \tag{18}$$

which we will use for an alternative definition of the unperturbed Hamiltonian

$$H_0^d(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{i=1}^{N} F^d(\mathbf{r}_i) \tag{19}$$

that leads to the partitioning $H = H_0^d + V + W$ with the off-diagonal one-electron perturbation $V = H_0 - H_0^d$ and the electron-electron interaction perturbation W is taken from (15).

As the ratio of the two perturbations is not known, the formalism of the double Rayleigh–Schrödinger (RS) perturbation theory can be used, which looks for the ground state of the total Hamiltonian

$$H(\mu, \lambda) = H_0^d + \mu V + \lambda W\,, \tag{20}$$

with the formal expansion coefficients μ and λ, taking finally the values $\mu = \lambda = 1$. Standard RS considerations lead to coupled non-linear recursive equations for the exact ground state $|\Phi\rangle$ of $H(\mu, \lambda)$ and the energy difference $\triangle\mathcal{E} = \mathcal{E} - E_0$ of the exact and unperturbed ground states in the intermediate normalization ($\langle\Psi_0|\Phi\rangle = 1$)

$$|\Phi\rangle \;=\; |\Psi_0\rangle + R_0 \left(\triangle\mathcal{E} - \lambda W - \mu V\right) |\Phi\rangle \tag{21}$$

$$\triangle\mathcal{E} \;=\; -\langle\Psi_0|\, \lambda W + \mu V\, |\Phi\rangle\,. \tag{22}$$

The resolvent operator

$$R_0 = \sum_{i>0} \frac{|\Psi_i\rangle\,\langle\Psi_i|}{E_i - E_0} \tag{23}$$

contains all eigenfunctions $|\Psi_i\rangle$ and eigenvalues E_i for $i \neq 0$ of the unperturbed Hamiltonian H_0^d. One of the main advantages of Kapuy's partitioning that all these eigenfunctions are known by simple and explicit formulas as $|\Psi_i\rangle$ are Slater-determinants of the localized one-particle MOs of the form (1) and the unperturbed energy values E_i are calculated by an appropriate sum of the diagonal Fock-matrix elements f_{jj}. Moreover, the unperturbed wavefunction $|\Psi_0\rangle$ is the Hartree–Fock ground state of the system.

Solving equations (21) and (22) iteratively one arrives at the formal double power series expansions

$$|\Phi\rangle = \sum_{m,n=0}^{\infty} \left|\Phi_{\mathrm{loc}}^{(m,n)}\right\rangle, \qquad\qquad \triangle\mathcal{E} = \sum_{\substack{m,n=0 \\ m+n>0}}^{\infty} \mathcal{E}_{\mathrm{loc}}^{(m,n)},$$

where

$$\left|\Phi_{\mathrm{loc}}^{(m,n)}\right\rangle = \lambda^m \mu^n \left|\Psi_{\mathrm{loc}}^{(m,n)}\right\rangle, \qquad\qquad \mathcal{E}_{\mathrm{loc}}^{(m,n)} = \lambda^m \mu^n E_{\mathrm{loc}}^{(m,n)}. \tag{24}$$

Convergence of these kind of formal expansions is questionable, however, it is believed today that at least asymptotic convergence behavior is to expected. Collecting terms in equations (21) and (22) that are proportional to the same power of λ and μ on the left and right hand sides we obtain recursion formulas

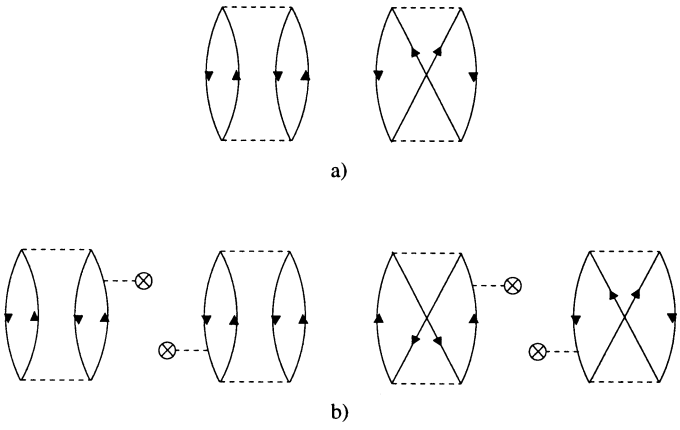

Fig. 2. Unlabelled Goldstone diagrams representing a) $\mathcal{E}_{\rm loc}^{(2,0)}$, b) $\mathcal{E}_{\rm loc}^{(2,1)}$ terms.

for the perturbation corrections:

$$\begin{aligned}\left|\Phi_{\rm loc}^{(m,n)}\right\rangle &= \sum_{i=0}^{m}\sum_{\substack{j=0\\ i+j>0}}^{n}\mathcal{E}_{\rm loc}^{i,j}R_0\left|\Phi_{\rm loc}^{(m-i,n-j)}\right\rangle \\ &\quad -R_0\lambda W\left|\Phi_{\rm loc}^{(m-1,n)}\right\rangle - R_0\mu V\left|\Phi_{\rm loc}^{(m,n-1)}\right\rangle \\ \mathcal{E}_{\rm loc}^{(m,n)} &= \left\langle\Psi_0\left|\lambda W\right|\Phi_{\rm loc}^{(m-1,n)}\right\rangle + \left\langle\Psi_0\left|\mu V\right|\Phi_{\rm loc}^{(m,n-1)}\right\rangle. \end{aligned} \tag{25}$$

Starting from the ground state of the unperturbed Hamiltonian we can obtain explicit expressions for these corrections, that are very complicated especially in higher orders. With the help of diagrammatic many-body perturbation theory these terms can be represented graphically (the detailed description of the diagrammatic method of time dependent and time independent MBPT can be found in Refs. [140] and [57], respectively). The contributions of certain diagrams will cancel and only those terms remain that are connected to linked diagrams [56, 57]. In Fig. 2 the unlabelled Goldstone diagrams [58], representing the terms $\mathcal{E}_{\rm loc}^{(2,0)}$ and $\mathcal{E}_{\rm loc}^{(2,1)}$, are presented.

The $\mathcal{E}_{\rm loc}^{(m,0)}$ terms are given by the usual diagrams and formulas but instead of canonical HF orbitals localized ones are used. The remaining $\mathcal{E}_{\rm loc}^{(m,n)}$ terms with $n \neq 0$ are called localization corrections. It is clear from the definitions above that summing up the localization corrections in any order of two-particle perturbation W we obtain the same order correction in MP partitioning. These terms are distinguished by the index "can":

$$|\Phi_{\rm occ}^{(m)}\rangle = \sum_{\substack{i=0\\ m+i>0}}^{\infty}\left|\Phi_{\rm loc}^{(m,i)}\right\rangle,$$

$$E_{\mathrm{occ}}^{(m)} = \lim_{n\to\infty} e_{\mathrm{loc}}^{(m,n)} \qquad \text{where} \qquad e_{\mathrm{loc}}^{(m,n)} = \sum_{i=0}^{n} \mathcal{E}_{\mathrm{loc}}^{(m,i)} .$$

The HF energy is given by

$$E_{HF} = E_0 + \mathcal{E}_{\mathrm{loc}}^{(1,0)} .$$

The localization corrections to the terms that are 0th or 1st order in the two particle perturbation are all zero because the diagonal part is omitted from the one-particle perturbation just like the matrix element between the occupied and virtual orbitals. The total electron correlation energy can be expressed as

$$E_{corr} = \sum_{m=2}^{\infty} \lim_{n\to\infty} e_{\mathrm{loc}}^{(m,n)} .$$

4 Application

In this section representative results are presented for a normal saturated hydrocarbon (C_5H_{12}) molecule as a model system. The CC bond length is 1.526 Å the CH distance is 1.094 Å and the bond angles are 120°. The LMOs were created using the Boys' localization procedure.

4.1 Convergence of the Localization Corrections

Comparing the canonical and localized MBPT treatment of the correlation energy we have seen that the introduction of localized orbitals results in additional terms in the correlation energy expression. Calculation of these terms requires extra work which depends on the speed of convergence of the additional localization corrections. This problem was investigated for our model system and is presented in Fig. 3. Convergence of summed up localization corrections $e_{\mathrm{loc}}^{(2,n)}$, $e_{\mathrm{loc}}^{(3,n)}$ as a function of the order n of the one-particle perturbation operator V is shown for the C_5H_{12} molecule in STO-3G and 6-31G* basis sets. Both for the second and third canonical order the convergence is faster in the minimal basis set than in 6-31G* one. Using the minimal basis set, the 0th order already gives approximately 95% of the canonical values in both cases. For the 6-31G* basis set in case (a) the 0th order of the second perturbation gives ~74% and in case (b) this value is only ~67%. In the third canonical order ($m = 3$) the convergence is slower than in second order ($m = 2$). To recover more than 99% of the canonical corrections we have to sum up the localization corrections in the second canonical order at least up to $n = 6$ and in the third canonical order up to $n = 8$.

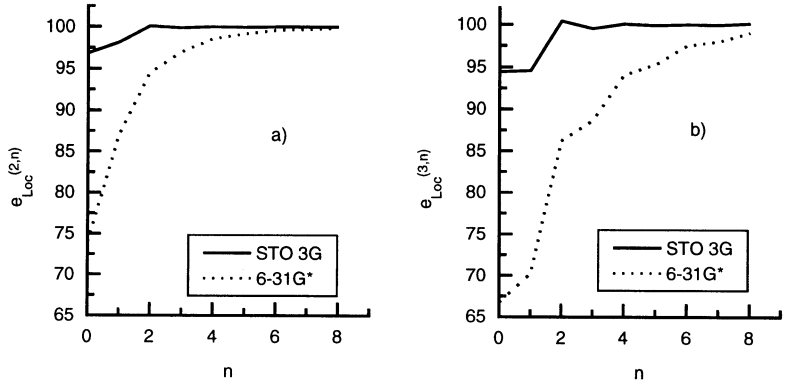

Fig. 3. Convergence of summed up localization corrections, a) $e_{\text{loc}}^{(2,n)}$ and b) $e_{\text{loc}}^{(3,n)}$, as a function of order n of one-particle perturbation for the C_5H_{12} molecule in STO-3G and 6-31G* basis sets. The energy values are given as percentages of $\mathcal{E}_{\text{occ}}^{(2)}$ and $\mathcal{E}_{\text{occ}}^{(3)}$, respectively.

4.2
Separation of Local Terms

Let us return now to the original aim of localizing molecular orbitals. The basic expectation is that interaction between distant localized orbitals turns out to be negligible. Before specifying more precisely what "interaction" in the above sense means, it is necessary to define the concept of "distance" between localized MOs. We will discuss here some alternatives that have all been tested in our numerical calculations.

Vicinity index. One such measure of separation is based on an exact upper bound for two-electron integrals of model wavefunctions which remains almost always valid for realistic LCAO orbitals, as well. The *vicinity index* [30, 31] of two MOs φ_i and φ_j fits well into framework of population localization, as it uses the atomic population definitions Q_A^i and Q_A^j of the orbitals defined by (14). The quantity

$$v_{ij} = \left| \sum_{A=1}^{n} Q_A^i \, Q_A^j \right|^{1/2} \tag{26}$$

is approximately 1 if both MOs are localized on the same atom (core orbitals or lone pairs), whereas $v_{ij} = 2^{-1/2}$ if one of the orbitals is localized on atom A, and the other orbital is in a bonding situation between atoms A and B, or if both orbitals lie in the same bonding region between A and B. The vicinity index decreases if the orbitals are more separated. In the case where φ_i and φ_j occupy different bonding regions, A–B and A–C, joining at atom A, $v_{ij} \approx 2^{-1}$, and for distant localized MOs that have no atoms in common the vicinity index

is practically zero. Even diagonal elements of Eq. (26) carry a physical meaning as v_{ii}^{-2} gives the approximate number of atoms where orbital φ_i is localized [32]. How the introduced quantities can be used for separating local and long-range interaction effects will be shown later in the discussion of numerical results.

An approach for separating local and long-range electron-electron interaction is based on the following idea. The two principal types of interaction quantities in localized MBPT are the two-electron integrals

$$\langle ij|kl\rangle = \int \varphi_i^*(\mathbf{r}_1)\,\varphi_j(\mathbf{r}_1)\,|\mathbf{r}_1 - \mathbf{r}_2|^{-1}\,\varphi_k^*(\mathbf{r}_2)\,\varphi_l(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \tag{27}$$

and the off-diagonal Fock-matrix elements

$$f_{ij} = \langle \varphi_i | F | \varphi_j \rangle. \tag{28}$$

Physical intuition suggests that the magnitude of $\langle ij|kl\rangle$ should decrease with increasing separation of orbitals φ_i and φ_j, and with increasing separation of φ_k and φ_l. In fact, the inequality

$$|\langle ij|kl\rangle| \leq L\, v_{ij}\, v_{kl} \tag{29}$$

holds exactly for a model system and is approximately true in realistic cases, as well [30]. Here L is an appropriate constant, v_{ij} and v_{kl} are the vicinity indices introduced previously. Considering (29) the two-electron integrals are replaced by a truncated value of

$$\langle ij|kl\rangle_t = \begin{cases} 0 & \text{if } v_{ij} \leq \mathrm{v_{lim}} \text{ or } v_{kl} \leq \mathrm{v_{lim}} \\ \langle ij|kl\rangle & \text{otherwise}. \end{cases} \tag{30}$$

The value of v_{lim}, a free control parameter of the calculations, is tunable in the range $(0, 1)$. According to the physical interpretation of v_{ij}, if $v_{\mathrm{lim}} = 0$ a full calculation without truncation is done, if v_{lim} is tuned in the range $0 \leq v_{\mathrm{lim}} < 0.5$ the long-range effects are switched off, while at $0.5 \leq v_{\mathrm{lim}}$ various local correlations are also neglected. Plotting the calculated correlation energy as a function of the control parameter v_{lim} the relative importance of local and long-range correlations can be traced [31].

In order to understand the physical meaning of the above truncation let us consider that integral $\langle ij|kl\rangle$ corresponds to a scattering process of the incoming electrons j and l to the states i and k, respectively. Truncation (30) restricts the scattering process but not the length of interaction among the electrons, i.e. the hopping of localized electrons j and l is restricted only to a given neighborhood but at the same time the interaction length of the Coulomb forces is still considered to be infinite.

A similar consideration for the off-diagonal Fock-matrix elements seems to be straightforward. The underlying inequality found by numerical calculations [33], however, has a quite different structure compared to (29)

$$|f_{ij}| \leq A\, \exp(\alpha\, v_{ij}). \tag{31}$$

The constant $A(\approx 0.05)$ is independent of the molecular system and of the applied atomic basis functions, whereas the value of α changes between 2.7

and 13.2 depending on the basis set and on the system studied. This result has to be interpreted as a reminder for being extremely cautious in neglecting long-range Fock-matrix elements as, even at large separation of MOs ($v_{ij} \approx 0$), there appear off-diagonal elements in the order of $f_{ij} \approx 0.05$. This is probably due to the well-known fact that localized orbitals have long oscillatory tails in order to achieve orthogonality, which are magnified by the Laplacian operator of F even far away from the center of the localized MO. Another possible reason for the infinite-distance one-particle interaction may be an artifact of the Hartree–Fock approximation, since it cannot properly describe dissociation, leading to non-physical interaction at infinite separation.

Neighborhood order. In minimal basis-set calculations both the localized and the virtual LMOs can be assigned to one or at most to two atoms. Using this assignment (as a practical realization of the vinicity index) we can define the neighborhood order O_{ij} of the LMOs φ_i and φ_j. Two orbitals are considered 0th neighbors if they are assigned to the same atom(s), first neighbors if they have one common atom and they are not 0th neighbors, 2nd neighbors if they are connected by one LMO, etc. Using this definition we can also assign a neighborhood order $O\left(ij\right) = O_{ij}$ to the one-particle matrix elements f_{ij} and $O\left(ijkl\right) = \max\left(O_{ij}, O_{kl}\right)$ to the two electron integrals $\left\langle ij|kl\right\rangle$.

Using the neighborhood order assigned to the two electron integrals we can investigate the behavior of $e_{\mathrm{loc}}^{(2,4)}$ by changing the cut-off neighborhood order O_{cut} and consider only those integrals where $O(ijkl) \leq O_{cut}$ (see Fig. 4) [61–64]. The corresponding correlation energy contribution is denoted by $e_{\mathrm{loc}}^{(2,4)}(O_{cut})$.

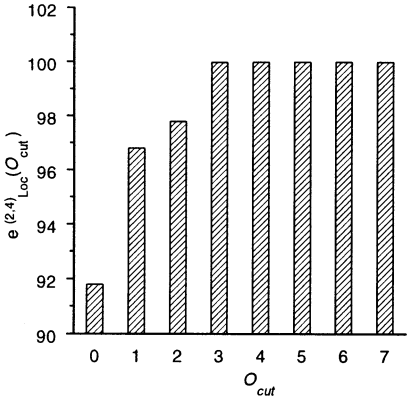

Fig. 4. The influence of the selection of O_{cut}, cut-off neighborhood order, on the value of the $e_{\mathrm{loc}}^{(2,4)}(O_{cut})$. Values are in %; the $e_{\mathrm{loc}}^{(2,4)}(O_{cut} = \infty)$ was taken as 100%.

We can see that up to 3rd neighbors we obtain the dominant part of this contribution.

The concept of neighborhood order and the truncation method based on it is similar in spirit to the method (30) based on vicinity indices. Neighborhood order, however, is more sensitive to long-range orbital separations as the vicinity index $v_{ij} \approx 0$ to a good approximation if $O_{ij} \geq 2$. In the light of the results shown in Fig. 4 this distinction does not play any significant role since long-range effects seem to be, according to our expectations, negligible and this finding is also confirmed by truncated correlation energy calculations using vicinity indices [31]. For a small molecule and minimal basis set the simple method described works really well but for more complicated systems there is no way to extend the definition of neighborhood orders.

Distance of orbitals. In larger basis sets the LMOs cannot be assigned to the atoms or atom pairs any more. A natural extension of the neighborhood order is the distance of LMOs φ_i and φ_j defined by $d_{ij} := \left|\bar{\mathbf{r}}_i - \bar{\mathbf{r}}_j\right|$, where $\bar{\mathbf{r}}_i$ and $\bar{\mathbf{r}}_j$ are the centroids of φ_i and φ_j, introduced previously. The definition is meaningful only if the distance of the orbital centroids is larger than the extension of the orbitals. Using d_{ij} we can assign a distance of $d\left(ij\right)$ to the one-electron integrals f_{ij} and a distance of $d\left(ijkl\right)$ to the two-electron integrals $\langle ij|kl\rangle$ in the following way

$$d\left(ij\right) = d_{ij}\,, \qquad\qquad d\left(ijkl\right) = \max\left\{d_{ij}, d_{kl}\right\}.$$

Similar to the previous case we can calculate the $e_{\mathrm{loc}}^{(2,n)}\left(d_0\right)$ and $e_{\mathrm{loc}}^{(3,n)}\left(d_0\right)$ contributions that include two-electron integrals with $d\left(ijkl\right) < d_0$ [66–68]. In Fig. 5 the influence of three different cut-off distances (d_0=2.0, 4.0 and 6.0

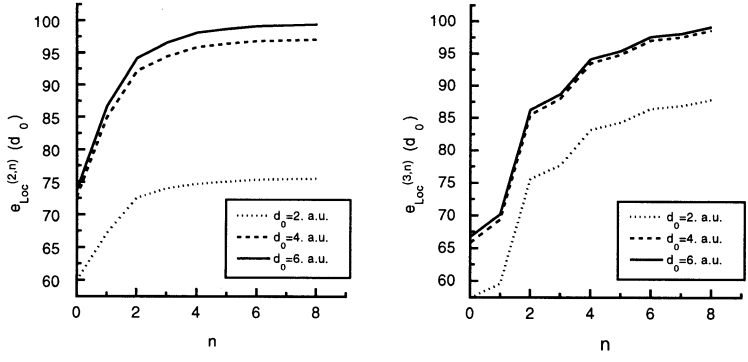

Fig. 5. The influence of the selection of d_0 cut-off distance on the value of the $e_{\mathrm{loc}}^{(2,n)}(d_0)$ (*left*) and $e_{\mathrm{loc}}^{(3,n)}(d_0)$ (*right*) as a function of n. Values are in %, the corresponding canonical values were taken as 100%.

a.u.) on the calculated second order (left panel) and third order (right panel) electron correlation energy is presented in the 6-31G* basis set. The cut-off does not change the convergence behavior that was presented in Fig. 3. For the second-order energy the neglected part of the complete contribution is ∼25% if $d_0 = 2.0$ a.u. and less than 1% if $d_0 = 6.0$ a.u. In the third order case the corresponding neglected parts are ∼15% and ∼0.1%, respectively.

5 Conclusions

Numerical calculations using Kapuy's partitioning scheme have shown that for covalent systems the role of one-particle localization corrections in many-body perturbation theory is extremely important. For good quality results several orders of one-particle perturbations have to be taken into account, although the additional computational power requirement is much less in these cases than for the two-electron perturbative corrections. Another alternative for increasing the precision of the calculations is to estimate of the asymptotic behavior of the double power series expansion (24) from the first few terms by applying Canterbury approximants [31], which is a two-variable generalization of the well-known Padé approximation method. It has also been found [6, 7] that in more metallic-like systems the relative importance of the localization corrections decreases, at least in PPP approximation.

The main advantage of Kapuy's method, however, is the excellent separation of local and long-range correlation effects, both of which are of theoretical and practical significance. Numerical results show that only the immediate neighborhood of a given localized MO plays a considerable role in excitation processes. A remarkable property of the method is that, even if long-range excitations are truncated, the range of the Coulomb interaction is still kept infinite. This possibility leads to a tremendous saving of the necessary computational resources even for medium-sized molecules. The advantages of applying localized MBPT depend on the balance between the increased efforts for calculating extra localization diagrams and the falling computational demands when long-range effects are omitted, making the method most probably more and more beneficial as the numerical treatment of extended molecular systems become available.

Acknowledgements. This work was supported by the Országos Tudományos Kutatási Alap (OTKA), Grant Nos. T024136, T021179 and F020665.

6 References

1. Kapuy E, Kozmutza C, Stephens ME (1976) Theoret Chim Acta 43:175
2. Kapuy E, Ozoróczy Z, Kozmutza C (1976) Acta Phys Hung 41:125
3. Kapuy E, Kozmutza C, Daudel R, Stephens ME (1978) Theoret Chim Acta 50:31

4. Kapuy E, Kozmutza C, Ozoróczy Z, Pipek J (1979) Acta Phys Hung 46:333
5. Kapuy E, Kozmutza C, Daudel R, Stephens ME (1979) Theoret Chim Acta 53:147
6. Kapuy E, Csépes Z, Kozmutza C (1983) Intern J Quantum Chem 23:981
7. Kapuy E, Csépes Z, Pipek J (1984) Acta Phys Hung 55:365
8. Slater JC (1930) Phys Rev 35:210
9. Hartree DR (1928) Proc Cambridge Phil Soc 24:89
10. Fock V (1930) Z Physik 61:126
11. Koopmans T (1933) Physica 1:104
12. Pipek J, Mezey PG (1988) Intern J Quantum Chem, Quantum Chem Symp 22:1
13. Anderson PW (1958) Phys Rev 109:1492
14. Boys SF (1960) Rev Mod Phys 32:296
15. Foster JM, Boys SF (1960) Rev Mod Phys 32:300
16. Boys SF (1966) in: Löwdin P-O (ed) *Quantum Theory of Atoms, Molecules and the Solid State*. Academic, New York p 253
17. Edmiston C, Ruedenberg K (1963) Rev Mod Phys 35:457
18. Edmiston C, Ruedenberg K (1965) J Chem Phys 43:S97
19. Edmiston C, Ruedenberg K (1966) in: Löwdin P-O (ed) *Quantum Theory of Atoms, Molecules and the Solid State*. Academic, New York p 263
20. von Niessen W (1972) J Chem Phys 56:4290
21. Pipek J, Mezey PG (1989) J Chem Phys 90:4916
22. Mulliken RS (1955) J Chem Phys 23:1833
23. Mulliken RS (1955) J Chem Phys 23:1841
24. Löwdin P-O (1950) J Chem Phys 18:365
25. Löwdin P-O (1970) Advan Quantum Chem 5:185
26. Bader RFW, Nguyen-Dang TT (1981) Advan Quantum Chem 14:63
27. Bader RFW, Nguyen-Dang TT, Tal Y (1981) Rept Progr Phys 44:893
28. Luken WL, Beratan DN (1982) Theoret Chim Acta 61:265
29. Magnasco V, Perico A (1967) J Chem Phys 47:971
30. Pipek J (1984) Chem Phys Lett 111:430
31. Pipek J, Ladik J (1986) Chem Phys 102:445
32. Pipek J (1989) Intern J Quantum Chem 36:487
33. Pipek J (1988) Chem Phys Lett 143:293
34. Roos B (1975) in: Dierksen GHF et al. (ed) *Computational Techniques in Quantum Chemistry and Molecular Physics*. Reidel, Dordrecht p 251
35. Shavitt I (1977) in: Schaefer HF (ed) *Modern Theoretical Chemistry*. Plenum, New York Vol. 3, p 189
36. Bartlett RJ, Purvis GD (1978) Intern J Quantum Chem 14:561
37. Pople JA, Seeger R, Krishnan R (1977) Intern J Quantum Chem, Quantum Chem Symp 11:149
38. Sasaki F (1977) Intern J Quantum Chem, Quantum Chem Symp 11:125
39. Čižek J (1966) J Chem Phys 45:4256
40. Čižek J (1969) Advan Chem Phys 14:35
41. Fetter AL, Walecka JD (1971) *Quantum Theory of Many–Particle Systems*. McGraw–Hill, New York
42. Paldus J, Čižek J (1975) Advan Quantum Chem 9:105
43. Wilson S (1981) in: *Theoretical Chemistry*. The Royal Society of Chemistry, London Vol. 4, p 1
44. Pulay P (1983) Chem Phys Lett 100:151
45. Pulay P (1986) J Chem Phys 85:1703
46. Pulay P, Sæbø S (1986) Theoret Chim Acta 69:357
47. Sæbø S, Pulay P (1987) J Chem Phys 86:914

48. Sæbø S, Pulay P (1988) J Chem Phys 88:1884
49. Csépes Z, Pipek J (1988) J Comput Phys 77:1
50. Moeller C, Plesset MS (1934) Phys Rev 46:618
51. Amosh AT, Musher JI (1971) J Chem Phys 54:2380
52. Davidson ER (1972) J Chem Phys 57:1999
Davidson ER, Bender CF (1972) J Chem Phys 56:4334
53. Kapuy E (1981) in: Laider KJ (ed) *International Union of Pure and Applied Chemistry*. 28th Congress, Vancouver, Canada, Abstracts p cc013
54. Csizmadia IG (1975) in: Chavlet O et al. (ed) *Localization and delocalization in quantum chemistry*. D. Reidel PC. p 349
55. March NH, Young WH, Sampathar S (1967) *The Many-Body Problem in Quantum Mechanics*, Cambridge University Press
56. Frantz LM, Mills RL (1960) Nucl Phys 15:16
57. Paldus J, Čižek J (1975) Adv Quantum Chem 9:105
58. Goldstone J (1957) Proc Roy Soc, Ser A 239:267
59. Daudel R, Stephens ME, Csizmadia IG, Kozmutza C, Kapuy E, Goddard JD (1977) Intern J Quantum Chem 11:665
60. Boys SF (1966) in: Löwdin P-O (ed) *Quantum Theory of Atoms Molecules and the Solid State*. Academic Press, New York-London, p 253
61. Kapuy E, Bartha F, Bogár F, Csépes Z, Kozmutza C (1987) Theoret Chim Acta, 72:337
62. Kapuy E, Bartha F, Kozmutza C, Bogár F (1988) J Mol Structure (THEOCHEM) 170:59
63. Kapuy E, Csépes Z, Bartha F, Bogár F, Kozmutza C (1988) Coll. Czechoslovak Chem Commun 53:2073
64. Kapuy E, Bartha F, Bogár F, Csépes Z, Kozmutza C (1990) Intern J Quantum Chem 38:139
65. Bartha F, Bogár F, Kapuy E (1990) Intern J Quantum Chem 38:215
66. Kapuy E, Bogár F, Bartha F, Kozmutza C (1991) J Mol Structure (THEOCHEM) 233:61
67. Kapuy E, Bogár F, Kozmutza C (1993) J Mol Structure, 297:365
68. Kapuy E, Bogár F, Tfirst E (1994) Intern J Quantum Chem 52:127

An Introduction to the Theory of Geminals

Péter R. Surján

Department of Theoretical Chemistry, Eötvös University, H-1518 Budapest P.O.B. 32, Hungary. *E-mail: surjan@para.chem.elte.hu*

Two-electron functions, called also geminals, have been around in quantum chemistry for some time. They represent a generalization of one-electron orbitals accounting for intra-orbital correlation effects. Geminal-based methods can be tailored to be variational as well as size-consistent and size-extensive. In spite of these appealing features, geminals became somewhat eclipsed in modern quantum chemistry because of their relative complexity and because the associated energies do not always cover a sufficient fraction of the correlation energy. However, several recent investigations revisit geminals and advocate the use of some extended geminal models which may turn out to offer useful alternatives to conventional approaches. In this paper, the formalism of two-electron functions will be reviewed in a simple fashion, focusing mainly on qualitative and conceptual points rather than technical details. After a short historical survey, the basic notions of geminals will be reviewed both in first- and second-quantized notations, the latter being especially advantageous when dealing with geminals. A few important points about the optimization of geminal-based wave functions will then be discussed, followed by a discussion about the inherent connection between geminals and the localization problem. We shall close with a few remarks on the prospect of geminal theories.

Keywords: Geminals Localization Arai theorem Brillouin theorem

Topics in Current Chemistry, Vol. 203

1 Introduction

Quantum chemical approaches rely on approximate solutions to the many-electron Schrödinger equation, which emerge either by introducing approximations of a mathematical nature (i.e., neglecting small terms in the equation to be solved), or by simplifying the Hamiltonian through the elaboration of some appropriate physical models. One of the most fruitful models in the history of quantum chemistry has been the *one-electron model* in which the interaction between electrons is either neglected (cf. Hückel scheme) or averaged (Hartree–Fock, HF). In both cases the appropriate form of the many-electron wave function satisfying the Pauli principle is a single Slater-determinant

$$\Phi(1,2,\ldots,N) = \hat{\mathcal{A}}\left[\varphi_1(x_1)\ \varphi_2(x_2)\ldots\varphi_N(x_N)\right], \tag{1}$$

with $\hat{\mathcal{A}}$ being the antisymmetrizer for N electrons and φ_i denoting a one-electron function (molecular spinorbital); the latter depends on the foursome of spatial- and spin-coordinates x_i. The vast majority of quantum chemical calculations starts with determining the one-electron functions φ_i which are of fundamental importance not only in traditional wave-function-based approaches but also in density-functional theories.

The importance of the one-electron model relies upon two facts: (i) the approximation represented by Eq. (1) is usually a good one, and (ii) assuming a *complete* set of one-electron orbitals $\{\varphi_i\}$ one may develop the *exact* wave function in terms of (an infinite number) of determinants of the form of Eq. (1).

For closed-shell systems, the HF wave function [Eq. (1)] is commonly replaced by the simpler form

$$\Phi(1,2,\ldots,N) = \hat{\mathcal{A}}\left[\psi_1(r_1)\alpha(1)\ \psi_1(r_2)\beta(2)\ldots\psi_{\frac{N}{2}}(r_N-1)\alpha(N-1)\psi_{\frac{N}{2}}(r_N)\beta(N)\right], \tag{2}$$

expressing that two electrons with opposite spins (α and β) occupy the same spatial orbital ψ_i which depends on the spatial coordinates r_i. Equation (2) has the important advantage over Eq. (1) that the former is always an eigenfunction of the $\hat{S}^2$ operator, in fact it is a singlet state.

In spite of the great success of the one-electron and the HF models, it soon turned out that Eq. (1) and especially Eq. (2), had serious drawbacks. These have been widely discussed in the literature. We note here merely that Eq. (2)

yields electronic energies that are too high; it completely neglects the Coulomb-correlation among electrons; it cannot describe bond dissociation processes; and it does not automatically reflect localized electron pairs, a fundamental concept of empirical chemistry.

A straightforward generalization of Eq. (2) emerges if, instead of the one-electron spin-orbitals $\varphi_1(1) = \psi_1(1)\alpha(1)$ and $\varphi_2(2) = \psi_1(2)\beta(2)$, one writes the wave function in terms of *two-electron orbitals* $\psi(1,2)$:

$$\Phi(1,2,\ldots,N) = \hat{\mathcal{A}}\left[\psi_1(x_1,x_2)\ \psi_2(x_3,x_4)\ldots\psi_{\frac{N}{2}}(x_N-1,x_N)\right], \tag{3}$$

where, if $\psi_i(x_1,x_2)$ is antisymmetric with respect to the electronic variables (spin-coordinates) x_1, x_2, the antisymmetrizer $\hat{\mathcal{A}}$ only takes care of inter-orbital permutations of the electron coordinates. Following perhaps Shull [1], the two-electron orbitals ψ_i are often called *geminals* while, being suggested by the form of Eq. (3), the term *separated pairs* is also used.

The analogy between the wave function of Eqs. (1–2) and Eq. (3) is obvious. However, the physical/chemical significance of these equations is quite different, which can be made evident by comparing the basis set expansion of the one-electron spin-orbitals $\varphi_i(x)$

$$\varphi_i(x) = \sum_{\mu} c_{\mu}^{i}\chi_{\mu}(x) \tag{4}$$

to that of the geminals $\psi_i(x_1,x_2)$:

$$\psi_i(x_1,x_2) = \sum_{\mu<\nu} C_{\mu\nu}^{i}\ \hat{\mathcal{A}}\ \left[\chi_{\mu}(x_1)\chi_{\nu}(x_2)\right]. \tag{5}$$

Clearly, the two-electron orbitals require a double-expansion; the $\mu < \nu$ restriction is introduced to avoid double-counting of the configurations.

In the HF model one optimizes the one-electron orbitals $\varphi_i(x)$ or their expansion coefficients c_{μ}^{i}, while in geminal theories one looks for energy-optimized two-electron functions $\psi_i(x_1,x_2)$ or the coefficients $C_{\mu\nu}^{i}$. This latter optimization is much more complicated than the HF procedure, but the advantageous properties of geminal-based functions may compensate us for the extra effort. The most appealing feature of a wave function built up from geminals is that it takes into account intra-pair electron correlation, which, in many cases, is a dominant fraction of the total correlation. More detailed properties of geminals, as well as those of the associated wave functions and optimization strategies, will be discussed after a short summary of the history of applying two-electron functions.

2
History of Geminals

2.1
Beginnings

As early as 1950, Fock [2] had already suggested the use of two-electron orbitals. Not too much later, Hurley, Lennard-Jones and Pople published their fundamental paper [3] elaborating the basic theory of geminals. Their work was followed in 1958 by that of Parks and Parr [4], who discussed also the minimalization of the energy of a system of separated electron pairs. In the same year, Ede Kapuy started a long series of papers that covered a detailed theory of orthogonal [5, 6, 7] and non-orthogonal [8, 9] geminals, methods for their construction [10, 11], possible improvements of the conventional formalism by means of configuration interaction [12, 13] or perturbation theory [14, 15], elaboration of the connection between the separated pair theory and the alternant molecular orbital (AMO) method [16, 17], as well as some applications [18, 15, 19]. Kapuy finished this series with a short review [20] summarizing the developments in the separated pair theory up to 1969.

In the meantime, the field grew and many authors joined in. There is no room here to list all important works; merely a few will be dealt with, more or less in chronological order.

A fundamental theorem was proven in 1960 by Arai [21] stating that the so-called strong orthogonality *(vide infra)* property of geminals is equivalent to expanding them on mutually exclusive orthogonal subspaces of one-electron functions. A proof for this theorem was also given by Löwdin [22]. Allen and Shull [23] emphasized the inherent connections between geminals and the concept of the chemical bond. An important next step was made in 1964 by Kutzelnigg [24], who also gave a simple proof for Arai's theorem, introduced the natural expansion of geminals, formulated and proved a series of new theorems, and proposed approximate procedures for the optimization of the wave function of Eq. (3), which he named APSG (*antisymmetric product of strongly orthogonal geminals*), provided that the two-electron functions satisfy the condition of strong orthogonality.

In 1965 Ebbing and Henderson [25] analyzed geminals in the LiH molecule and optimized them by inspecting total energy curves as functions of orbital rotation parameters.

Geminal-based methods were shown to be special cases of the more general group-function formalism [26, 27, 28]. Csizmadia et al. [29] performed calculations on the LiH molecule. Though their geminals, being composed from a single determinant, were not correlated, these calculations indicated that the separation of electron pairs can be a good approximation in well localized systems. Klessinger and McWeeny [30] performed group function (basically un-optimized geminal) calculations on methane, addressing even excited and

ionized states. McWeeny and Steiner [31] described a "pair-see" formulation where two fully correlated electrons move in the field of the others. Van der Hart presented a number of applications of the group function method, mostly in the geminal approximation [32, 33, 34, 35].

Blatt [36], Coleman [37, 38], then Bratož and Durand [39] investigated a special N-electron wave function in which all geminals were constrained to have the same form, and established the relationship between this function and that used by Bardeen, Cooper and Shieffer (BCS) [40] to describe superconducting systems with electron pairs. The underlying wave function was termed as the *antisymmetrized geminal power* (AGP) function.

In 1968, a group of keynote papers was published by Miller and Ruedenberg [41, 42, 43]. They introduced an optimization procedure for geminals, presented a proof for Arai's theorem, analyzed the advantages and the limitations of the simple APSG wave function, introduced the *augmented separated-pair expansion* (which develops the wave function in terms of APSG-type terms), and tested the theory on beryllium-like atoms.

Extensive numerical studies were performed in 1970 by Silver et al. [44, 45, 46]. These gave rise to the conclusion that, in contrast to earlier expectations, the fraction of the correlation energy covered by APSG is not too high; it can be sometimes as low as 50%. This negative experience justified the development of the antisymmetrized product of geminals (APG) model [47, 48] in which the strong orthogonality constraint was relaxed. Nicely and Harrison [49] constructed APG wave functions for LiH and BH, bringing substantial improvement with respect to APSG. For the BH molecule, for which APSG was known to fail [45], they found that the fraction of the correlation energy increased from 80% (APSG) to 98% (APG). Nicely and Harrison [49] have also considered odd-electron systems by augmenting the singlet coupled geminals with appropriate doublet one-electron functions.

An interesting idea was used by Franchini and Vergani [50], who avoided the expensive optimization of geminals, rather they expanded them in a localized MO basis. A notable study was published in early seventies by Robb and Csizmadia [51, 52, 53, 54], entitled *'The Generalized Separated Electron Pair Model'*. They have also determined geminals on a localized basis, introduced a multi-configuration APSG ansatz, and considered one- and two-electron transfer terms as well. They performed calculations on NH_3 and its isoelectronic series [51, 52, 53] as well as for the CO molecule [54].

The generalized valence bond method [55] can be considered as a special APSG technique, where each geminal consists only of two spatial orbitals.

Wilson and Geratt [56] discussed a pair-function model constructing geminals from non-orthogonal one-electron orbitals. Their calculations, performed on the water molecule, supported qualitative valence-shell electron-pair (VSEPR) models [57] of directed valence.

Further theoretical development was achieved in the late seventies, some has proved to be useful for future applications, others represented some *cul de*

sacs. Based on an idea by Silver [58], Náray-Szabó [59] investigated all-pair N-electron wave functions constructed from $N(N-1)/2$ geminals. Fantucci and Polezzo [60, 61, 62, 63, 64] elaborated optimization techniques for multiconfigurational SCF wave functions which were used subsequently for optimizing geminals [65]. Besides strong orthogonality, a set of general orthogonality conditions was investigated by Wilson [66]. Interest in non-orthogonal geminals was also increased in this period [67, 68].

There is no room here to collect together all important achievements in quantum chemistry which are strongly or loosely connected to geminals; or at least, their development was perhaps motivated by the successes or failures of geminal theories (cf. e.g. the self-consistent electron pair method by Meyer and its reformulations [69, 70, 71], or the connections to valence bond theory [72]). We have to mention an important book by Hurley [73] in which significant attention was paid to geminals and their localization.

In summarizing, the fundamental knowledge about constructing and using simple as well as generalized geminal-based wave functions had been collected by the 1980s. In course of this work, the pioneering studies of Ede Kapuy played an important role. The calculations performed with such type of wave functions were mostly limited at that time, however, to small molecules and to small basis sets.

2.2 Recent Developments

Since 1980, developments in the area proceeded along several lines, a few examples of which are listed below.

2.2.1 *Use of Relative Electron Coordinates*

Already in 1970, Robb and Csizmadia discussed the possibility of incorporating *relative electron-coordinates* in geminal-type wave functions [52]. This idea comes up naturally as two-electron functions may, in principle, depend on the coordinate differences r_{ij}, though this is never used in the so-called *algebraic approximation* that expresses any N-electron function in terms of products of one-electron functions. The result is a better account of electron correlation, e.g. the correlation cusp [74] of the wave function. The correlation energy will, therefore, converge better with increasing basis set as compared to the algebraic approximation. Geminals that contain relative electron coordinates are often termed explicitly correlated.

A scientific research group, concentrated mainly at Gainesville, Florida, carried out a detailed study on various quantum chemical methods, incorporating explicitly correlated Gaussian geminals into the basis set [75, 76, 77, 78, 79]. The research in this direction is still in progress [80, 81, 82, 83, 84].

2.2.2
Treatment of Extended (Large) Systems

Though recent computational facilities make the use of larger basis sets possible, standard geminal wave functions can only be determined accurately for small molecules. For extended systems, including large molecules, polymers and solids, severe approximations have to be made.

Valdemoro and Rubio proposed a geminal approach to treat covalent crystals [85]. They worked in a minimal basis set of Bloch orbitals and constructed k-vector dependent geminals.

Kirtman formulated a density matrix treatment using a separated pair ansatz [86]. He derived variational equations in the local space approximation to determine fragment wave functions and discussed the role of strong orthogonality in the localization of the wave function.

Following the work by Ukrainskii [87], several other authors applied geminals to describe infinite systems, mostly in the π-electron approximation [88, 89, 90, 91]. Kuprievich treated ground and excited states of a quasi-metallic ring of 50 atoms with strongly orthogonal geminals [89]. Karadakov et al. discussed simple geminals both in delocalized [90] and localized [91] representations, and found the latter superior. This confirms a general feature of geminals that will be discussed in Sect. 5 in more detail.

2.2.3
Uniform Geminal Models

Several authors revisited the AGP wave function and its extensions. In the original form, this function is simply the product (the N-th power) of N identical geminals. In spite of this simple form, it can offer an approximation better than HF if the geminal is properly optimized. Fukutome has dealt with the projected BCS wave function [92] which is essentially equivalent APG. Some key publications in this field are due to Ortiz et al. [93] and Sangfelt et al. [94]. Further, Öhrn and Linderberg have shown that the AGP wave function may serve as an appropriate (i.e., consistent) reference state for excited state calculations in the random phase approximation (RPA) [95, 96]. (We note only in passing that, subject to some restriction in the excitation ansatz, the full APSG wave function can also be used as a consistent ground state in equation-of-motion (EOM) techniques [97, 98].) Liu et al. has extended the AGP model to allow for two kinds of geminals [99, 100].

2.2.4
Second Quantized Theory, Composite Particles

Second quantization offers a great help in dealing with elementary particles (e.g., electrons). Geminals are entities composed of two electrons, thus their second quantized description requires special features.

Second quantization for composite particles, in the context of quantum chemistry, was elaborated by several authors, e.g. by Girardeau [101], Kvasnička [102], Fukutome [103, 92] , and Valdemoro [104, 105], to name a few. The present author used creation operators composed of two fermion operators to describe geminals in orthogonal [106, 107, 108, 98] and non-orthogonal [109, 110, 111, 112, 113] basis sets. Second quantization for geminals will be reviewed in Sect. 3.2.

2.2.5
Extended Geminal Models

In a series of papers, Røggen has formulated a general procedure to approach the exact (full-CI) solution in terms of geminals. His approach is called the *extended geminal model*. First [114, 115], he introduced an extra subset of orbitals, common to all geminals, in order to describe inter-geminal correlation effects, and avoided the violation of strong orthogonality by the method of moments [116, 117, 118]. This theory later evolved into a general formulation of the many-electron problem, specifying a new route towards the exact solution of the many-electron problem (see [119] and references therein, as well as the contribution by Røggen in the present book [120]).

Extending the structure of the wave function is not the only way of improving the APSG approximation. In our laboratory [109, 107, 108], we proposed the biorthogonal formulation to take care of intergeminal overlap effects, and derived simple formulae to account for delocalization and dispersion interactions using either perturbation theory or a linearized coupled-cluster-type ansatz with the APSG reference state.

2.2.6
Applications, Qualitative Explanations, Localization

Among the links to qualitative theory, the connection to the VSEPR theory has already been mentioned above. Another conceptually important field of application offered by geminal-based theories is the description of two-electronic fragments (inner shells, valence-shell two-center bonds, lone pairs, etc.) in a polyatomic system [113]. The inherent relation between the theory of geminals and the localization problem has been emphasized for a long time. Due to its importance this issue will be the focus of Sect. 5.

An important qualitative feature of geminal-based wave functions is that they, unlike closed-shell HF, may properly describe the bond dissociation process. This advantage was evident from the very beginning and it was analyzed later in great detail [121, 122] including the extension to describe dissociation of multiple bonds [123]. Qualitatively correct potential curves have even been reported as obtained from drastically simple approximate geminals [106, 110].

3
The Form of Geminal Wave Functions

3.1
First-Quantized Formulation

Since the pioneering work of Heitler and London [124], that explained the classically non-existent bond in the H_2 molecule, two-electron two-center bonds have been the targets of many investigations. The most essential features can be understood on a minimal basis set model, where each of the two constituents contribute one basis function, say χ_A and χ_B. Then, the normalized combinations of these,

$$\phi_1 = \frac{1}{\sqrt{1+\lambda^2+2\lambda s}}\left(\chi_A + \lambda\chi_B\right) \tag{6}$$

$$\phi_2 = \frac{1}{\sqrt{1+\mu^2-2\mu s}}\left(\mu\chi_A - \chi_B\right), \tag{7}$$

with s being the $\langle\chi_A|\chi_B\rangle$ overlap, can be considered as bonding and antibonding molecular orbitals, respectively. The λ and μ polarity parameters, from the orthogonality of ϕ_1 to ϕ_2, are related as $\mu = \frac{\lambda+s}{1+\lambda s}$. They are equal if $s = 0$ (cf. semiempirical theories or the case of orthogonalized basis functions), and both are 1 for a homopolar bond. In the general case, λ has to be optimized variationally.

The singlet HF wave function in terms of MOs is clearly written as

$${}^1\Psi_{HF} = \hat{\mathcal{A}}\phi_1(r_1)\alpha(1)\phi_1(r_2)\beta(2) = \phi_1(r_1)\phi_1(r_2)\,\frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2)-\alpha(2)\beta(1)\right]. \tag{8}$$

That is, ${}^1\Psi_{HF}$ is described as a product of a symmetric spatial and an antisymmetric spin function ${}^1\vartheta(1,2) = \frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2)-\alpha(2)\beta(1)\right]$. The same function can be transformed to the AO basis by substituting (6) and (7) into (8):

$${}^1\Psi_{HF} = \frac{\chi_A(r_1)\chi_A(r_2)+\lambda^2\chi_B(r_1)\chi_B(r_2)+\lambda(\chi_A(r_1)\chi_B(r_2)+\chi_B(r_1)\chi_A(r_2))}{1+\lambda^2+2\lambda s}\;{}^1\vartheta(1,2). \tag{9}$$

The first two terms in the numerator describe configurations where both electrons are on A and B, respectively. These are called ionic terms. The last two ones describe the sharing of electrons in a symmetrized way, these are the covalent terms. For $\lambda = 1$ (homopolar bond), the coefficients of the ionic and

covalent terms are equal, though physical intuition predicts the ionic terms to have smaller weights due to correlation effects (Coulomb hole). The neglect of electron correlation introduces a quantitative error in ${}^1\Psi_{HF}$, and, even worse, it results in a qualitative failure in the dissociation limit when χ_A and χ_B are very far from each other. Here the weight of ionic terms should go to zero while it is kept constant in ${}^1\Psi_{HF}$ for $\lambda = 1$ ("dissociation catastrophe"). The famous Heitler–London (HL) wave function is obtained by keeping the covalent terms only; it dissociates properly but, since it neglects ionic terms completely, it is strongly over-correlated. Clearly, the best possible two-electron wave function in this basis set emerges if one introduces an extra variational parameter in front of the ionic terms, e.g.

$$\begin{aligned} {}^1\Psi_{Weinb} \;=\; & \big(c_1\chi_A(r_1)\chi_A(r_2) + c_2\chi_B(r_1)\chi_B(r_2) \\ & +\kappa\left[\chi_A(r_1)\chi_B(r_2) + \chi_B(r_1)\chi_A(r_2)\right]\big) \;\; {}^1\vartheta(1,2)\,, \end{aligned} \tag{10}$$

with the normalization condition $c_1^2 + c_2^2 + 2\kappa^2 + 2c_1c_2s^2 + 4\kappa s(c_1 + c_2) + \kappa^2 s^2 = 1$. Such a wave function was used (with optimized orbital exponents) by Weinbaum [125]. It is important to realize that the same form emerges if one considers the two-configuration wave function

$${}^1\Psi = \left[\phi_1(r_1)\phi_1(r_2) + \eta\phi_2(r_1)\phi_2(r_2)\right] \;\; {}^1\vartheta(1,2)\,, \tag{11}$$

with an appropriate mixing coefficient η ("pair-excited CI"). The equivalence of (10) and (11) can be seen by inserting the forms of ϕ_1 and ϕ_2 from Eqs. (6) and (7). This equivalence means that, for a two-electron system, no open-shell configurations (i.e., $\phi_1(r_1)\phi_2(r_2)$ type terms) occur in the MO basis. In other words, the analogs of the covalent configurations are missing if the MOs are used. We emphasize that within this limited two-orbital basis set the wave functions (10) or (11) are exact, they correspond to the full-CI solution of the two-electron problem. The HF or HL functions are obviously special cases of Eq. (10) which can be called a geminal wave function for the two electrons.

More generally, a singlet geminal in a larger AO basis can be put down as

$${}^1\Psi_{geminal} = \sum_{mn} C_{mn}\;\chi_m(r_1)\chi_n(r_2) \;\; {}^1\vartheta(1,2)\,, \tag{12}$$

where the diagonal terms ($m = n$) constitute the ionic terms while $m \neq n$ contribute to the covalent configurations. The same wave function was quoted in Eq. (5) in terms of spin-orbitals; now it is written in terms of spatial orbitals with the spin function separated. Equation (12) is a generalization of Eq. (10). The singlet spin function ${}^1\vartheta(1,2)$ being antisymmetric, the coefficients C_{mn} must be symmetric to preserve the antisymmetry of the entire wave function. One can therefore turn to the basis set where matrix C_{mn} is diagonal:

$${}^1\Psi_{natgem} = \sum_{m} C_m\;\varphi_m(r_1)\varphi_m(r_2) \;\; {}^1\vartheta(1,2)\,, \tag{13}$$

which is a generalization of Eq. (11). For reasons given below, orbitals φ are called natural orbitals and Eq. (13) is the natural representation of the geminal.

The total wave function of a many-electron system can be constructed as an antisymmetrized product of individual geminals [see Eq. (3)]. Dealing with this product is substantially simpler if the geminals are kept orthogonal to each other in the strong sense, i.e.

$$\int \psi_i(r_1, r_2)\psi_k(r_1, r_2)dr_1 \;=\; 0\,. \tag{14}$$

In terms of the expansion coefficients C_{mn} this means

$$\sum_n C^i_{mn}\, C^k_{rn} \;=\; 0 \quad \text{for all } m, r\ (i \neq k)\,. \tag{15}$$

The antisymmetrized product of strongly orthogonal geminals is denoted by the acronym APSG. Properties of this wave function, its construction and use, will be discussed in the forthcoming sections.

3.2
Second-Quantized Formulation

Application of the second quantized formalism to the theory of geminals does not introduce any essentially new features but it does makes many derivations much easier and more transparent. If the reader is not familiar with the formalism of second quantization [126], she or he may skip the following formulae and focus merely on the results of this section.

The second quantized analogue of Eq. (5) is

$$\psi_i^+ \;= \sum_{\mu<\nu} C^i_{\mu\nu}\, a^+_\mu a^+_\nu \tag{16}$$

with antisymmetric coefficient matrices $C^i_{\mu\nu}$. Here a^+_μ is an electron creation operator on spinorbital χ_μ. One can interpret the symbol ψ_i^+ as a *composite particle creation operator*, since it creates the two-electron wave function for the geminal i.

For the sake of simplicity, we shall assume in this review that the basis orbitals χ_μ form an orthonormal set. Generalization to the nonorthogonal case can be done in a straightforward manner using the biorthogonal technique [127, 128, 129, 109, 113, 126].

In an orthogonal basis, the adjoints of creation operators a^+_μ are the annihilation operators a_μ. The electron creation and annihilation operators obey simple anticommutation rules

$$\left[a^+_\mu,\ a^+_\nu\right]_+ \;=\; \left[a_\mu,\ a_\nu\right]_+ \;= 0\, \left[a^+_\mu,\ a_\nu\right]_+ \;=\; \delta_{\mu\nu}\,. \tag{17}$$

The algebra of the composite particle creation operators ψ_i^+ is more complicated. Denoting the adjoint operators by ψ_i^-, the commutation rules, in the general case, read:

$$\begin{aligned} \left[\psi_i^+,\ \psi_k^+\right]_- \;&=\; \left[\psi_i^-,\ \psi_k^-\right]_- \;= 0 \\ &\left[\psi_i^-,\ \psi_k^+\right]_- \;=\; \hat{Q}_{ik}\,, \end{aligned} \tag{18}$$

which can be seen by inserting the expansion (16) of the composite operators and using Eq. (17). Here $\hat{Q}_{ik}$ is a non-diagonal matrix of operators [104] that causes serious difficulties when handling operators ψ in matrix elements, so we shall get rid of it.

An important difference between Eqs. (17) and (18) is that the latter contains commutators instead of anticommutators, expressing that the quasi-particles composed of two electrons are bosons.

To avoid the difficulties represented by the non-diagonal operator matrix $\hat{Q}$, we shall introduce the strong orthogonality (SO) condition (15) under which its form becomes substantially simpler:

$$\hat{Q}_{ik} = \delta_{ik}\, \hat{Q}_i\,, \tag{19}$$

with

$$\hat{Q}_i = 1 - \sum_{\mu\nu} P^i_{\mu\nu} a^+_\mu a_\nu\,, \tag{20}$$

where $P^i_{\mu\nu}$ is a matrix element of the first-order density matrix for geminal i:

$$P^i_{\mu\nu} = \langle \psi_i | a^+_\nu a_\mu | \psi_i \rangle\,. \tag{21}$$

As a result of this simplification, the rules for the evaluation of matrix elements between APSG-type wave functions become similar to those between single determinants [126]. In particular, full density matrices as well as the energy formula can easily be evaluated.

An easy way of ensuring SO is to introduce a partition of the set of one-electron basis functions and to expand each geminal only within a single subspace:

$$\psi^+_i = \sum_{\mu<\nu\in i} C^i_{\mu\nu}\; a^+_\mu a^+_\nu\,, \tag{22}$$

Then, the full first order density matrix becomes block-diagonal in geminal indices:

$$P_{\mu\nu} = \langle \Psi | a^+_\nu a_\mu | \Psi \rangle = \delta_{ik}\, P^i_{\mu\nu} \quad (\mu \in i,\ \nu \in k)\,, \tag{23}$$

where the first-order density matrix for geminal i takes the simple form

$$P^i_{\mu\nu} = \sum_{\lambda\in i} C^i_{\mu\lambda} C^i_{\nu\lambda} \quad (\mu,\nu \in i)\,. \tag{24}$$

The intra-geminal elements of the second order density matrix read:

$$\Gamma^i_{\sigma,\lambda,\mu,\nu} = \langle a^+_\mu a^+_\nu a_\lambda a_\sigma \rangle = C^i_{\mu\nu} C^i_{\sigma\lambda} \quad (\mu,\nu,\lambda,\sigma \in i)\,, \tag{25}$$

while the intergeminal elements have the same structure as in HF theory:

$$\Gamma_{\sigma,\lambda,\mu,\nu} = \langle a^+_\mu a^+_\nu a_\lambda a_\sigma \rangle = P_{\mu\sigma} P_{\nu\lambda} - P_{\nu\sigma} P_{\mu\lambda} \tag{26}$$

[Many of these matrix elements are zero due to Eq. (23)]. The energy expression for the APSG wave function is, therefore, particularly simple:

$$E = \langle \hat{H} \rangle = \sum_i \sum_{\mu,\nu \in i} h_{\mu\nu}^{eff} \, P_{\nu\mu}^i + \frac{1}{2} \sum_i \sum_{\mu,\nu,\lambda,\sigma \in i} [\mu\nu|\lambda\sigma] \, C_{\mu\nu}^i C_{\lambda\sigma}^i - \frac{1}{2} \sum_{i \neq K} \sum_{\mu,\nu \in i} \sum_{\lambda,\sigma \in K} [\mu\lambda||\nu\sigma] \, P_{\mu\nu}^i P_{\lambda\sigma}^K , \tag{27}$$

with the effective one-electron Hamiltonian for geminal i

$$h_{\mu\nu}^{eff} = h_{\mu\nu} + \sum_{k(\neq i)} \sum_{\lambda,\sigma \in k} P_{\lambda\sigma}^k \, [\mu\lambda||\nu\sigma] . \tag{28}$$

In the above equations, $h_{\mu\nu}$ are the usual one-electron integrals while $[\mu\nu|\lambda\sigma]$ and $[\mu\lambda||\nu\sigma]$ are the standard bare and antisymmetrized two-electron integrals, respectively. To derive these formulae, one has merely to substitute the second quantized form of the total Hamiltonian and apply the above rules for the density matrix elements. The analogy of Eq. (27) to the corresponding HF formula is obvious.

Determination of the appropriate expansion coefficients $C_{\mu\nu}^i$ and the optimal choice of the basis functions μ by which the geminals are expanded will be discussed in the following section.

4
Optimization of Geminals

4.1
Arai's Theorem

Because of the strong orthogonality condition, two-electron functions are easier to construct, and to deal with. As mentioned above, a straightforward way of ensuring SO is to expand the geminals in mutually exclusive and orthogonal subspaces. This seems to be a very severe restriction, but, because of a famous statement that we call Arai's theorem, the fact is that the existence of an expansion of this kind and SO are completely equivalent. The theorem can be formulated as follows.

Theorem (Arai):

> Given a set of geminals $\{\psi_i\}$ satisfying the SO condition [Eq. (15)] and having a basis set expansion in an orthogonal set of one-electron orbitals [Eq. (5)], it is always possible to find a unitary transformation in the one-electron space so that in the new basis each geminal is expanded in a subset of the one-electron functions the subsets having no common elements.

The importance of this statement cannot be over-emphasized. Strongly orthogonal geminal expansions in mutually exclusive subspaces can be done without any loss of generality provided one determines these subspaces which, for brevity, will be called *Arai-subspaces* hereinafter.

As listed in Sect. 2., several authors have proved this theorem [21, 22, 42, 24]. A simple proof will be presented below.

Consider two geminals ψ_i and ψ_k having expansions (16). The associated density matrices are:

$$P^i_{\rho\tau} = \langle \psi_i | a^+_\tau a_\rho | \psi_i \rangle = \sum_\gamma C^i_{\rho\gamma} C^i_{\tau\gamma} \tag{29}$$

and similarly for P^k. The product of these two density matrices will be:

$$\left(P^i P^k\right)_{\rho\tau} = \sum_{\lambda\gamma\eta} C^i_{\rho\gamma} C^i_{\lambda\gamma} C^k_{\lambda\eta} C^k_{\tau\eta} \,. \tag{30}$$

One may observe that under the strong orthogonality condition (15), due to the summation over λ, this product matrix vanishes:

$$\sum_\lambda P^i_{\rho\lambda} P^k_{\lambda\tau} = 0 \,. \tag{31}$$

Consequently, density matrices of different SO geminals will commute, thus they may be brought to the diagonal form simultaneously:

$$P^i_{\rho\tau} = n^i_\rho \delta_{\rho\tau} \tag{32}$$

and

$$P^k_{\rho\tau} = n^k_\rho \delta_{\rho\tau} \,. \tag{33}$$

Substitution of this diagonal form into (31) gives, for $i \neq k$,

$$n^k_\rho n^i_\rho = 0 \,, \tag{34}$$

which means that for any orbital ρ either of the occupation numbers n^i_ρ or n^k_ρ must be zero, thus in the corresponding basis sets the two subsets cannot have a common element, *Q.e.d.*

In terms of spatial orbitals, the geminal coefficient matrix C^i_{mn} is symmetric for singlets, and thus may be diagonalized by an appropriate unitary transformation [Eq. (13)]. Then, the associated spatial density matrices become also diagonal:

$$P^i_{mn} = 2 \sum_r C^i_{mr} C^i_{nr} \;=\; 2\, \delta_{mn} \, \left(C^i_m\right)^2 \,. \tag{35}$$

That is, diagonalizing all C^i matrices in their respective subspaces, one arrives at a basis set in which the entire density matrix is diagonal with occupation numbers $n^i_m = 2\left(C^i_m\right)^2$. The orbitals spanning this basis set are, therefore, the natural orbitals of the APSG wave function, justifying the name 'natural geminals' used in Sect. 3.1.

Arai's theorem suggests that the optimization of geminals concerns two issues: (i) finding the optimal Arai subspaces and (ii) solving the local two-electronic Schrödinger equations within each subspace. These points will be addressed below.

4.2 Local Schrödinger Equations

In this section, assuming that optimal or approximate subspaces for all geminals are available, we discuss the optimization of the expansion parameters $C^i_{\mu\nu}$ [Eq. (22)]. These are linear variational parameters, thus their optimum values result from a local Schrödinger equation

$$\hat{H}^i \psi_{iq} = E_{iq} \psi_{iq} \,, \tag{36}$$

one for each geminal. Here q labels the various states of geminal i, which are eigenfunctions of the effective Hamiltonian $\hat{H}^i$. Substituting the basis set expansion for the geminals, we arrive at the matrix equation

$$\sum_{\mu<\nu\in i} H^i_{\lambda\sigma,\mu\nu} C^{iq}_{\mu\nu} = E^{iq} C^{iq}_{\lambda\sigma} \quad (\lambda < \sigma \in i) \,, \tag{37}$$

where the matrix elements of the effective Hamiltonian are defined by

$$H^i_{\lambda\sigma,\mu\nu} = \langle vac | a_\sigma a_\lambda \; \hat{H}^i \; a^+_\mu a^+_\nu | vac \rangle \,. \tag{38}$$

The local effective Hamiltonians compress all integrals whose indices belong to the associated subspaces:

$$\hat{H}^i \;=\; \sum_{\mu\nu\in i} h^{eff}_{\mu\nu} \; a^+_\mu a_\nu \;+\; \frac{1}{2} \sum_{\mu\nu\lambda\sigma\in i} [\mu\nu|\lambda\sigma] \; a^+_\mu a^+_\nu a_\sigma a_\lambda \,, \tag{39}$$

where, in the spirit of group function theory [26, 27], the effective one-electron integrals are the ones defined in Eq. (28). Substituting Eq. (39) into Eq. (38), the matrix elements of the local Hamiltonian are obtained in a particularly simple form:

$$H^i_{\lambda\sigma,\mu\nu} \;=\; \delta_{\sigma\nu} h^{eff}_{\lambda\mu} \;-\; \delta_{\sigma\mu} h^{eff}_{\lambda\nu} \;+\; \delta_{\lambda\mu} h^{eff}_{\sigma\nu} \;-\; \delta_{\lambda\nu} h^{eff}_{\sigma\mu} \;+\; [\mu\nu||\lambda\sigma] \,. \tag{40}$$

Variational determination of the expansion coefficients $C^i_{\mu\nu}$, requires the construction and diagonalization of Hamiltonian matrices, one for each geminal. Diagonalizations have to be done iteratively, because, due to the presence of the effective core h^{eff}, the Schrödinger equations of the geminals are coupled. These iterations, however, usually converge quite fast. The Hamiltonian matrices are of limited size, as $\hat{H}^i$ is represented only in the subspace of basis functions assigned to the geminal in question. Optimization of the subspaces themselves is a more complicated issue.

4.3 Brillouin Theorem for Geminals

In HF theory, one has the Brillouin theorem (BT) stating that singly excited configurations do not interact with the ground state determinant [130]. The proof commonly proceeds by utilizing the properties of the HF wave function. An alternative route was followed by Mayer [131, 132, 133] who derived the BT directly from the variational principle, permitting one to obtain the HF

equations from the Brillouin condition. We shall follow here the same idea in studying whether a generalization of the BT exists for the APSG wave function.

The conjunction that such a generalization must exist, is based on an analogy of the BT for group functions discussed by McWeeny [26].

We start from the general variational condition $\delta\left[\langle\Psi|\hat{H}|\Psi\rangle - E\langle\Psi|\Psi\rangle\right] = 0$, where E is a Lagrangian multiplier, to ensure the normalization of Ψ. Carrying out the variation of the brackets, one finds

$$\langle\delta\Psi|\hat{H} - E|\Psi\rangle = 0\,, \tag{41}$$

where $\delta\Psi$ is an infinitesimal variation of the wave function within the permitted class of functions specified by the type of Ψ. If the variation is kept orthogonal to the wave function, i.e., $\langle\delta\Psi|\Psi\rangle = 0$, Eq. (41) simplifies to

$$\langle\delta\Psi|\hat{H}|\Psi\rangle = 0\,. \tag{42}$$

Let us now study the consequences of this general result when Ψ is the APSG wave function for N geminals

$$\Psi_0 = \psi_{10}^+\psi_{20}^+\ldots\psi_{N0}^+\,|vac\rangle\,, \tag{43}$$

where, with the subscript 0, we indicate that each geminal is in its ground state, yielding the total ground state APSG wave function Ψ_0. Varying the individual geminals as $\psi_{i0} \to \psi_{i0} + \delta\psi_i$, we can write the varied APSG wave function, up to first order, as

$$\Psi_0 \to \Psi_0 + \delta\Psi = \Psi_0 + \sum_{i=1}^{N} \psi_{10}^+\ldots\delta\psi_i^+\ldots\psi_{N0}^+\,|vac\rangle + \mathcal{O}(2)\,. \tag{44}$$

Accordingly, the first order variation of the wave function takes the form

$$\delta\Psi = \sum_{i=1}^{N} \psi_{10}^+\ldots\delta\psi_i^+\ldots\psi_{N0}^+\,|vac\rangle\,. \tag{45}$$

The next step is to study the individual variations $\delta\psi_i$. As any two-electron function, these variations can be expanded according to products of pairs of one-electron functions. In the second quantized notation we may therefore, in general, write

$$\delta\psi_i^+ = \sum_{\mu\nu} \eta_{\mu\nu}\, a_\mu^+ a_\nu^+\,, \tag{46}$$

where η-s are first-order infinitesimals. However, any choice of η-s does not represent a permitted variation (that keeps the APSG form). To remain within the permitted function class, let us separate three possible types of variations:

$$\delta\psi_i^+ = \sum_{\mu\nu\in i} \eta_{\mu\nu}^i\, a_\mu^+ a_\nu^+ + \sum_{\mu\in i}\sum_{k(\neq i)}\sum_{\nu\in k} \eta_{\mu\nu}^{ik}\, a_\mu^+ a_\nu^+ + \sum_{k,l(\neq i)}\sum_{\mu\in k}\sum_{\nu\in l} \eta_{\mu\nu}^{kl}\, a_\mu^+ a_\nu^+\,. \tag{47}$$

It is not difficult to see that the third term completely destroys geminal i, putting both electrons into other subspaces. Since such a variation will not be permitted, we set $\eta_{\mu\nu}^{kl} = 0$. The first term is also easy to manage, as it preserves geminal i in the sense that it puts both electrons onto subspace i.

Such a function can be expanded using the locally complete set of two-electron functions ψ_{iq}:

$$\sum_{\mu\nu\in i} \eta^{i}_{\mu\nu}\; a^{+}_{\mu} a^{+}_{\nu} \;=\; \sum_{q\neq 0} \delta t_{iq}\; \psi^{+}_{iq}\,, \tag{48}$$

where q runs over all excited states of geminal i. The ground state ($q = 0$) is left out as it would not be a variation. The varied wave function is thus:

$$\begin{aligned} \delta\Psi \;&=\; \sum_{i=1}^{N} \sum_{q\neq 0} \delta t_{iq}\; \psi^{+}_{10}\ldots\psi^{+}_{iq}\ldots\psi^{+}_{N0}\;|vac\rangle\,. \\ &=\; \sum_{i=1}^{N} \sum_{q\neq 0} \delta t_{iq}\; \psi^{+}_{iq}\psi^{-}_{i0}\;|\Psi_0\rangle \end{aligned} \tag{49}$$

It is easy to see by substitution that such a variation remains orthogonal to Ψ_0, thus the variational condition becomes:

$$\sum_{i=1}^{N} \sum_{q\neq 0} \delta t_{iq}\; \langle \Psi_{i_0\to i_q}\;|\hat{H}|\;\Psi_0\rangle \;=\; 0 \tag{50}$$

which, for arbitrary δt_{iq}, can hold only if

$$\langle \Psi_{i_0\to i_q}\;|\hat{H}|\;\Psi_0\rangle \;=\; 0\,. \tag{51}$$

Verbally, excitations that remain within a geminal do not interact with the ground state. This result can be referred to as the *local Brillouin theorem*. It is equivalent to solving the local Schrödinger equations as described in the previous section.

It remains to consider the third type of variation, the middle term in Eq. (47). It has the effect of moving a single electron from geminal i to k. It can be transformed to the form

$$\sum_{\mu\in i} \sum_{k(\neq i)} \sum_{\nu\in k} \eta^{ik}_{\mu\nu}\; a^{+}_{\mu} a^{+}_{\nu} \;=\; \sum_{k(\neq i)} \sum_{\mu\in i} \sum_{\nu\in k} \xi^{ik}_{\mu\nu}\; a^{+}_{\nu} a_{\mu} \psi^{+}_{i}\,, \tag{52}$$

where the infinitesimals ξ-s are in a one-to-one correspondence with η-s:

$$\eta^{ik}_{\mu\nu} \;=\; \sum_{\lambda\in i} C^{i}_{\lambda\mu}\; \xi^{ik}_{\lambda\nu}\,,$$

which can easily be verified by substituting the expansion of geminal ψ^{+}_{i}, letting Eq. (52) act on the vacuum state, and comparing the coefficients. Since such a variation remains also orthogonal to Ψ_0, the variational condition yields

$$\sum_{i\neq k} \sum_{\mu\in i} \sum_{\nu\in k} \xi^{ik}_{\mu\nu}\; \langle \Psi_0 \mid\; a^{+}_{\nu} a_{\mu} \hat{H} \mid \Psi_0 \rangle \;=\; 0\,, \tag{53}$$

which, for arbitrary variations implies

$$\langle \Psi_0 \mid\; a^{+}_{\nu} a_{\mu} \hat{H} \mid \Psi_0 \rangle \;=\; 0 \quad (\mu \in i,\; \nu \in k,\; i \neq k)\,. \tag{54}$$

These equations constitute the *nonlocal* part of the *Brillouin theorem* for geminals. They can be formulated as the APSG wave function is stationary with

respect to variations that move an electron from one geminal to another. Such variations are known to be described by an appropriate unitary transformation within the orbital space [134], thus satisfaction of the Brillouin theorem is equivalent to finding the optimal Arai-subspaces.

4.4 MCSCF Philosophy

Optimization of the APSG wave function requires the fulfillment of both the local and nonlocal Brillouin conditions [Eqs. (51) and (54)]. The former can be achieved by solving the local Scrödinger equations [Sect. 4.2] while the latter requires a laborious orbital optimization.

Efficient techniques for optimizing orbitals have been elaborated in multiconfiguration SCF (MCSCF) theory (see e.g. the book by McWeeny [134] and refs. therein). Since the APSG wave function represents a special class among MCSCF functions, these procedures can be applied to determine the optimal Arai-subspaces [65].

The basic idea of the optimization techniques mentioned is to express the energy as a function of orbital rotation parameters, and to make the energy stationary with respect to the variation of these parameters. The optimization is most simply done by the gradient technique. For the rotation of two orbitals, m and n, one has the gradient

$$g_{mn} = 2(F_{mn} - F_{nm}) , \tag{55}$$

where F is a generalized Fockian, which for an APSG wave function reads

$$F_{mn} = \sum_{l\in i} h_{ml} P^i_{ln} + \sum_{spl\in i} \left[sl|pm\right] \Gamma^i_{pnsl} + \sum_{l\in k} \sum_{j(\neq i)} \sum_{ps\in j} \left[sl|pm\right] \Gamma_{pnsl} \tag{56}$$
$$(m \in k, \quad n \in i) .$$

To reach the minimum, one has to set up an iterative procedure. Starting with an initial guess for the orbitals, the local Schrödinger equations are solved to get the geminals, from which the generalized Fockian and the gradients are evaluated. Then the orbitals are transformed in the direction of gradients; this transformation can be made more effective (but more laborious) with the aid of a Hessian. In the resulting new orbital basis new geminals have to be constructed and the procedure is repeated till convergence. Invoking natural geminals further increases the effectivity of the procedure.

As a result of optimization, unique Arai-subspaces emerge, but the form of the orbitals spanning the subspaces is arbitrary. The orbitals can be made unique by fixing them as the natural orbitals in each subspace.

An interesting feature of this optimization procedure is that the *dimensions* of Arai-subspaces have to be given as *input parameters*. They specify the overall structure of the wave function and can be guessed by chemical intuition or optimized in a "trial-and-error" scheme.

Table 1. Test calculations for LiH in STO-6G basis for comparison to another optimized APSG method

Method	Energy [atomic units]
HF	−7.96663
Optimized in Ref. [65]	−7.97981
SLG-Boys	−7.98085
Optimized APSG	−7.98090

One has to keep in mind that orbital rotation is a highly nonlinear procedure, and, unless checked, one can never be sure whether it converges to the absolute minimum in the parameter space: local minima as well as saddle points may also be reached. An example of this will be shown in Table 1.

4.5 Using Localized Orbitals

To speed up the convergence of orbital optimizations, the selection of a proper initial guess has to be done with care. For reasons given in the following paragraph, localized MOs obtained at the HF level often provide a good starting point.

An example is given in Table 1 where the total energy of the LiH molecule is shown at different levels. Although the basis is very small (minimal STO-6G), the subspace optimization is not trivial as shown by the second row of the table: the corresponding energy was supposed to be optimized if Ref. [65], but, using Boys' localized MOs (LMOs) as an initial guess, one gets a better energy without any optimization. Varying the Boys orbitals one may still achieve an energy lowering of 0.05 mH. (The acronym 'SLG-Boys' in the Table means 'strictly localized geminals', expressing that the geminals are not fully optimized but are expanded in orthogonal subspaces of Boys' localized MOs.)

Localized HF MOs can be used not only for accelerating convergence, but even to bypass the whole expensive optimization process simply by approximating the optimum orbitals with the localized ones. Clearly, besides occupied HF orbitals, the virtual ones must also be localized with some criterion. If the Boys localization is applied, the virtuals can be assigned to the occupied ones by a simple distance criterion. It is not necessarily true, however, that Boys orbitals represent the best choice. It may be worthwhile checking energetically localized Edmiston–Ruedenberg LMOs [135, 136], or any other set of LMOs. The idea by Pulay [137] to substitute virtual MOs with atomic orbitals, projected out from the occupied space, is also promising.

The fact that Boys LMOs represent an appropriate initial guess can also be inferred from Fig. 1. We plot there the variation of the total energy of some molecules as a function of a single selected orbital rotation parameter. The scale is chosen so that the Boys LMOs correspond to 0 degree. It is apparent

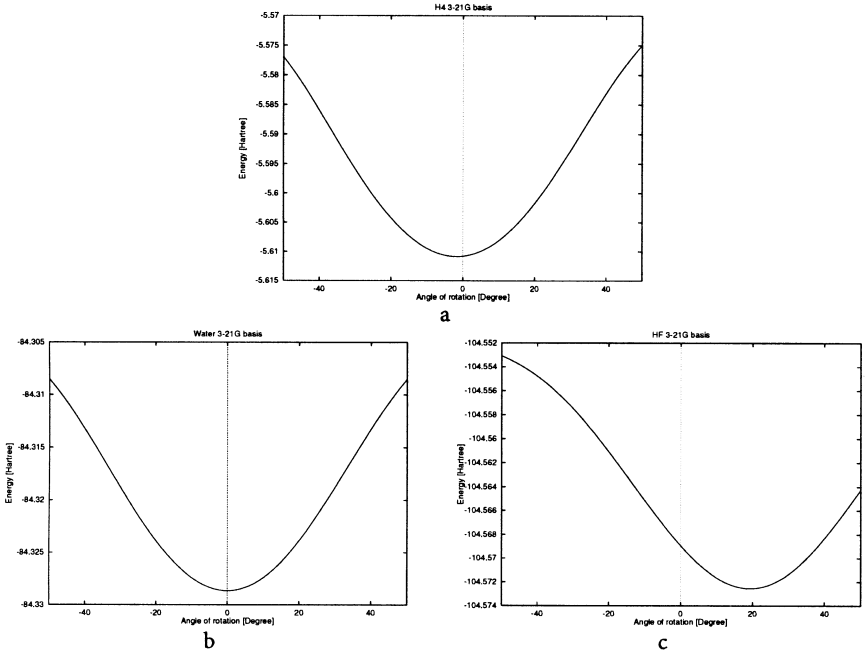

Fig. 1. Dependence of the total energy of the H_4 cluster, the water and HF molecules, as a function of an orbital rotation parameter. **a** H_4 cluster, mixing the occupied MOs of the two H_2 molecules. **b** H_2O, mixing the occupied MOs of the two O–H bonds **c** HF, mixing the bonding LMO of the F–H bond to the lowest nonbonding MO on F

that in two of the cases the variational minimum is almost at 0°, while for the hydrogen fluoride it is at around 20° but it is still closer to the Boys limit than, let us say, to the canonical MOs.

5
Geminals and Electron Localization

5.1
Uniqueness of Geminals

While a single-determinant wave function is known to be invariant — up to an irrelevant phase factor — against any unitary transformation among the orbitals it is constituted from, no such degree of freedom exists for an APG wave function.This is easily seen by mixing geminals ψ_1 and ψ_2 in $\Psi = \psi_1^+ \psi_2^+ |vac\rangle$:

$$\begin{aligned} \psi_1' &= \cos\alpha\ \psi_1 - \sin\alpha\ \psi_2 \\ \psi_2' &= \sin\alpha\ \psi_1 + \cos\alpha\ \psi_2 \end{aligned} \tag{57}$$

by which the transformed wave functions becomes

$$\begin{aligned}\Psi' &= \psi'^{+}_1\ \psi'^{+}_2\ |vac\rangle \qquad (58)\\ &= \left[\cos\alpha\sin\alpha\left(\psi_1^+\psi_1^+ - \psi_2^+\psi_2^+\right) + \cos^2\alpha\ \psi_1^+\psi_2^+ - \sin^2\alpha\ \psi_2^+\psi_1^+\right]|vac\rangle\,.\end{aligned}$$

For a general α, this wave function is *not* equal to Ψ for two reasons: (i) diagonal terms like $\psi_1^+\psi_1^+$, which vanish for single determinants, now survive; (ii) $\psi_2^+\psi_1^+ = \psi_1^+\psi_2^+$ without sign change. This latter point affects only the normalization of Ψ, but the former brings in configurations absent in Ψ. Both are consequences of the Bose-like commutation rules [Eq. (18)].

There are only special values of the rotation angle α for which $\Psi' = \Psi$: if either $\cos\alpha$ or $\sin\alpha$ is zero. These values correspond to leaving the geminals unchanged ($\sin\alpha = 0$) or to interchanging ψ_1 and ψ_2 ($\cos\alpha = 0$). Accordingly, apart from trivial permutations, geminals cannot be subjected to unitary transformations without changing the many-electron wave function.

This "no-transformation-theorem" for geminals has important consequences. In HF theory, one can freely localize the MOs, or keep them in canonical form, or even make them maximally delocalized [138]. This whole stuff is, however, completely arbitrary. In contrast, the uniqueness of individual geminals in an APG wave function offers important information on localization and delocalization: if a geminal happens to be localized in a molecule, it is not a matter of an ambiguous transformation but it is an inherent property of the geminal. We must note here that geminal-based wave functions can be, and usually are, superior to the HF wave function.

A last remark needs to be made about the connection of unitary transformations and strong orthogonality. The reader may easily verify that the commutation rule (18), expressing the SO condition algebraically, is *not* invariant against a unitary transformation of geminals. Such transformations, therefore, may preserve only the weak orthogonality but not SO. This fact underlines the irrelevance of global geminal transformations in APSG theory, emphasizing the uniqueness of individual geminals in a many-electron wave function.

5.2
Qualitative Considerations

The conclusion of the previous paragraph is that a geminal is either localized or delocalized, independently of any auxiliary conditions. Qualitative arguments have led to the further conclusion that geminals are usually *localized* in space. To see this, one has to recall that correlation always stabilizes the energy, and that a geminal-type wave function accounts for nothing but intra-geminal correlation. Consequently, letting the shape of a geminal become optimized in space, it will be energetically favorable to end with a compact function for which the dynamic correlation energy of the two electrons sitting on it is larger. This argument does not hold for situations where the static correlation is

dominant, for in this case a spatially extended geminal may emerge to account for the long-range static correlation effect (cf. bond dissociation process).

The fact that energy-optimized geminals are usually localized in space was predicted a long time ago by Hurley [73], and was investigated in more detail by Luken [139]. The essence of these investigations is that in the natural orbital representation [Eq. (13)] the geminals are dominated by a *principal orbital* having the highest (close to 2) occupation number and a *secondary orbital* being also occupied appreciably (the rest of the orbitals usually have quite small occupation numbers). An approximate minimization of the energy leads to conditions for the forms of the principal and secondary orbitals that are similar to those of a localization criterion.

The spontaneous localization of geminals may also be important in extended systems, where long-range correlation effects may appear in the form of localization. For example, there is no way to describe the so called Wigner-crystal in a free electron gas (see e.g. [140]) at the HF level, while in principle it should be possible with a geminal wave function.

These conceptually important results may have a practical impact, too. They constitute the basis of the approximate procedures discussed in Sect. 4.5, which avoid the time-consuming optimization of the orbitals by constructing the geminals from localized MOs.

6 Outlook

When they were introduced and first investigated, geminals were of interest to quantum chemists because they promised to give the correlation energy to chemical accuracy. Though it soon turned out that it was not the case, geminals have still retained their significance because they exhibit certain properties that serve as some kind of standards for most sophisticated wave functions. A proper many-electron method has to obey certain conditions: (i) it should give an upper limit to the energy (variational character); (ii) it should be exact for a two-electron unit; (iii) should be it size-consistent/extensive. Most of the widely accepted modern quantum chemical approaches violate one or more of these conditions: perturbational and coupled-cluster approaches are not variational, truncated CI schemes are not size-extensive, and neither of them takes care properly of two-electron units in their simplest (single-reference) forms; thus they are unable to describe the bond-fission process. Geminal-based methods belong to that family of multi-configurational schemes which simultaneously satisfy the above requirements. Therefore, if their construction and optimization can be done more effectively and their deficiencies solved at a reasonable cost, geminals may show a great promise.

Among multi-configurational approaches, the geminal approach, in particular the APSG method, is distinguished for several reasons. First, it correlates all electron pairs; thus APSG embodies highly excited determinants relative to

HF, while keeping the structure of the HF wave function as far as possible. In its fully variational form, it may yield a smooth potential surface with any kind of geometry, including extremely large bond distances, and it is free from the intruder-state problem (unfortunately, this does not hold for approximately optimized APSG).

The *a priori* localized nature of geminals is quite important and promising. Computationally, localization permits one to develop efficient algorithms that scale, in the limit of large N, quadratically or even linearly with the number of geminals (N). Conceptually, localization means that a geminal-type wave function may represent classical chemical concepts (such as the two-electron bond, lone pairs, etc.) in an unambiguous manner. In fact, we believe that the best possible quantum chemical representation of local bonds in a molecule is an APG-type wave function. Thus the limitations of this model reflect the limitations of these concepts themselves.

The above-mentioned items are mainly valid for the ground state. Calculation of excited and ionized states with geminals has only been marginally dealt with in the literature, though it is not *ab ovo* evident that such wave functions were less appropriate for that purpose. Formally, the APSG wave function may serve as a correlated reference state for various excited-state methods, though the localized features of the wave function will be almost certainly lost. This, again, reflects the objective fact that excited (or ionized) states of molecules have, in most cases, delocalized characters.

Much work has still to be done before the above expectations and hopes can be verified or disproved, allowing geminal-based approaches to occupy their due place in the arsenal of modern quantum chemistry. The recent increasing interest in these schemes indicates that the substantial efforts of Ede Kapuy in the fifties and sixties will not have been wasted.

Acknowledgements. The author is indebted to M. Kállay and Á. Szabados for many useful discussions and for reading the present manuscript. This work was supported in part by the grants of the Hungarian research fund SOTKA T021179–T023052 and FKFP 0165.

7 References

1. Shull H (1959) J Chem Phys 30:1405
2. Fock FA (1950) Dokl Akad Nauk USSR 73:735
3. Hurley AC, Lennard-Jones J, Pople JA (1953) Proc Roy Soc (London) A220:446
4. Parks JM, Parr RG (1957) J Chem Phys 28:335
5. Kapuy E (1958) Acta Phys Hung 9:237
6. Kapuy E (1960) Acta Phys Hung 11:97
7. Kapuy E (1960) Acta Phys Hung 12:185
8. Kapuy E (1959) Acta Phys Hung. 10:125
9. Kapuy E (1966) J Chem Phys 44:956
10. Kapuy E (1960) Acta Phys Hung 11:409
11. Kapuy E (1961) Acta Phys Hung 13:461

12. Kapuy E (1961) Acta Phys Hung 12:351
13. Kapuy E (1961) Acta Phys Hung 13:345
14. Kapuy E (1966) Theor Chim Acta 6:281
15. Kapuy E (1968) Theor Chim Acta 12:397
16. Kapuy E (1965) Theor Chim Acta 3:379
17. Kapuy E (1968) Acta Phys Hung 24:307
18. Kapuy E, March N (1967) J Math Phys 8:1915
19. Kapuy E (1968) Chem Phys Letters 3:43
20. Kapuy E (1969) Acta Phys Hung 27:179
21. Arai T (1960) J Chem Phys 33:95
22. Löwdin P-O (1961) J Chem Phys 35:78
23. Allen T, Shull H (1961) J Chem Phys 35:1644
24. Kutzelnigg W (1964) J Chem Phys 40:3640
25. Ebbing D, Henderson R (1965) J Chem Phys 42:2225
26. McWeeny R (1959) Proc Roy Soc (London) A253:242
27. McWeeny R (1960) Rev Mod Phys 32:335
28. Wilson S, Gerratt J (1975) Mol Phys 30:765
29. Csizmadia IG, Sutcliff BT, Barnett MP (1964) Canadian J Chem 42:1645
30. Klessinger M, McWeeny R (1965) J Chem Phys 42:3343
31. Steiner RME (1965) Adv Quant Chemistry 2:93
32. van der Hart WJ (1971) Mol Phys 20:385
33. van der Hart WJ (1971) Mol Phys 20:399
34. van der Hart WJ (1971) Mol Phys 20:407
35. Meerman CM, van der Hart WJ (1975), Theor Chim Acta 37:319
36. Blatt JM (1960) Prog Theor Phys 23:447
37. Coleman A (1963) Rev Mod Phys 36:668
38. Coleman A (1965) J Math Phys 6:1425
39. Bartož S, Durand P (1963) J Chem Phys 43:2670
40. Baarden L, Cooper L, Schrieffer J (1957) Phys Rev 108:1175
41. Miller KJ, Ruedenberg K (1968) J Chem Phys 48:3413
42. Miller KJ. Ruedenberg K (1968) J Chem Phys 48:3444
43. Miller KJ, Ruedenberg K (1968) J Chem Phys 48:3450
44. Silver D, Mehler E, Ruedenberg K (1970) J Chem Phys 52:1174
45. Mehler E, Ruedenberg K, Silver D (1970) J Chem Phys 52:1181
46. Silver D, Ruedenberg K, Mehler E (1970) J Chem Phys 52:1206
47. Silver D (1969) J Chem Phys 50:5108
48. Silver D (1970) J Chem Phys 52:299
49. Nicely V, Harrison JF (1971) J Chem Phys 54:4363
50. Franchini PF, Vergani C (1968) Theor Chim Acta 13:46
51. Robb MA, Csizmadia IG (191) J Chem Phys 54:3646
52. Robb MA and Csizmadia IG (1970) Int J Quantum Chem 4:365
53. Robb MA, Csizmadia IG (1971) Int J Quantum Chem 5:605
54. Robb MA, Csizmadia IG (1972) Int J Quantum Chem 6:367
55. Hunt WJ, Hay PJ, Goddard W (1972) J Chem Phys 57:738
56. Wilson S, Gerratt J (1975) Mol Phys 30:89
57. Gillespie RJ (1973), *Molecular Geometry* (Van Nostrand Reinhold, London)
58. Silver D (1971) J Chem Phys 55:1461
59. Náray-Szabó G (1975) Int J Quantum Chem 9:9
60. Polezzo S (1975) Theor Chim Acta 38:211
61. Fantucci P, Polezzo S, Stabilini MP (1976) Theor Chim Acta 41:311
62. Fantucci P, Polezzo S (1977) Theor Chim Acta 45:317

63. Polezzo S, Fantucci P (1980) Mol Phys 39:1527
64. Polezzo S, Fantucci P (1980) Mol Phys 40:759
65. Polezzo S, Fantucci P (1978) Mol Phys 36:1835
66. Wilson S (1976) J Chem Phys 64:1692
67. Lyast IT (1978) Int J Quantum Chem 13:83
68. Wilson S (1978) Mol Phys 35:1381
69. Meyer W (1976) J Chem Phys 64:2901
70. Pulay P, Saebø S, Meyer W (1984) J Chem Phys 81:1901
71. Polezzo S, Fantucci P, Trombetta L (1981) Int J Quantum Chem 19:493
72. Raimondi M (1999) in: Surján PR (ed) Correlation and Localization. Top Curr Chem, Vol. 203, Springer, Berlin Heidelberg New York, p. 105
73. Hurley AC (1976) Electron correlation in small molecules. Academic, New York
74. Kato T (1957) Commun Pure Appl Math 10:15
75. Szalewitz K, Jeziorski B, Monkhorst H, Zabolitzki J (1983) J Chem Phys 78:1420
76. Szalewitz K, Jeziorski B, Monkhorst H, Zabolitzki J (1983) J Chem Phys 79:5543
77. Jeziorski B, Szalewitz HMK, Zabolitzki J (1984) J Chem Phys 81:368
78. Szalewitz K, Zabolitzki J, Jeziorski B, Monkhorst H (1984) J Chem Phys 81:2723
79. Alexander S, Monkhorst H, Roeland R (1990) J Chem Phys 93:4230
80. Bukowski R, Szalewitz K, Jeziorski B (1995) J Chem Phys 224:155
81. Persson B, Taylor P (1996) J Chem Phys 105:5915
82. Korona T, Willians H, Szalewicz K (1998) J Chem Phys 106:5109
83. Bukowski R, Jeziorski B, Szalewicz K (1998) J Chem Phys 108:7946
84. Gilmore D, Kozlowski PM, Adamowitz L (1997) Int J Quantum Chem 63:991
85. Valdemoro C, Rubio J (1980) Annales de Fisica 76:26
86. Kirtman B (1983) J Chem Phys 79:835
87. Ukrainskii II (1978) Theor Math Phys 32:816
88. Tchugreev A, Misurkin I (1987) Teor Eksp Khim 6:665
89. Kuprievich V (1989) Phys Rev B 40:3882
90. Karadakov P, Castaño O, Calais J-L (1990) J Chem Phys 92:3021
91. Karadakov P, Castaño O, Calais J-L (1990) J Chem Phys 92:3027
92. Takahashi M, Fukutome H (1983) Int J Quantum Chem 24:603
93. Ortiz J, Weiner B, Öhrn Y (1981) Int J Quantum Chem S15:113
94. Sangfelt E, Goscinski O, Elander N, Kurtz H (1981) Int J Quantum Chem S15:133
95. Öhrn Y, Linderberg J (1977) Int J Quantum Chem 12:161
96. Öhrn Y, Linderberg J (1979) Int J Quantum Chem 15:343
97. Surján PR (1998) Croatica Chemica Acta 71:489
98. Surján PR, Kállay M, Szabados Á (1998) Int J Quantum Chem 70:571
99. Liu C, Deng C, Hu H, Jin B (1992) Int J Quantum Chem 42:339S
100. Liu C, Deng C, Jin B, Hu H (1992) Int J Quantum Chem 43:301
101. Girardeau M (1963) J Math Phys 4:1096
102. Kvasnička V (1982) Czech J Phys B32:947
103. Fukutome H (1978) Prog Theor Phys 60:1624
104. Valdemoro C (1985) Phys Rev A 31:2114
105. Valdemoro C (1985) Phys Rev A 31:2123
106. Surján PR (1984) Phys Rev A 30:43
107. Surján PR (1994) Int J Quantum Chem 52:563
108. Surján PR (1995) Int J Quantum Chem 55:109
109. Surján PR, Mayer I, Lukovits I (1985) Phys Rev A 32:748
110. Poirier RA, Surján PR (1987) J Comput Chem 8:436
111. Surján PR (1984) Croatica Chemica Acta 57:833
112. Surján PR (1989) Croatica Chemica Acta 62:579

113. Surján PR (1990) The Two-electron Bond as a Molecular Building Block. In: Maksić ZB (ed) Theoretical Models of Chemical Bonding, Vol 2 Springer, Berlin Heidelberg New York, p. 205
114. Røeggen I (1981) Int J Quantum Chem 19:319
115. Røeggen I (1981) Int J Quantum Chem 20:817
116. Szondy E, Szondy T (1966) Acta Phys Hung 20:253
117. Hegyi M, Mezei M, Szondy T (1971) Theor Chim Acta 21:168
118. Ladányi K, Lengyel V, Szondy T (1971) Theor Chim Acta 21:176
119. Røggen I, Wind PA (1996) J Chem Phys 105:2751
120. Røggen I (1999), in: Surján PR (ed) Correlation and Localization. Springer, Berlin Heidelberg New York, p. 000
121. England WB (1982) J Phys Chem 86:1204
122. England WB (1983) Int J Quantum Chem 17:357
123. England WB, Silver DM, Steinborn EO (1984) J Chem Phys 17:81
124. Heitler W, London F (1927) Z Physik 44:455
125. Weinbaum S (1933) J Chem Phys 1:593
126. Surján PR (1989) Second Quantized Approach to Quantum Chemistry. Springer, Berlin Heidelberg New York
127. Longuet-Higgins H (1966) in: Löwdin P-O (ed) Quantum Theory of Atoms, Molecules, and the Solid State. Academic, New York, p. 105
128. Moshinsky M, Seligman TH (1971) Annals of Phys 66:311
129. Gouyet J (1973) J Chem Phys 59:4637
130. Brillouin M (1933) Actualites Sci et Ind 71
131. Mayer I (1971) Acta Physica Hung 30:373 (1971)
132. Mayer I (1973) Acta Physica Hung 34:83
133. Mayer I (1974) Acta Physica Hung 36:11
134. McWeeny R (1989) Methods of Molecular Quantum Mechanics. Academic, London
135. Edmiston C, Ruedenberg K (1963) Rev Mod Phys 35:457
136. Edmiston C, Ruedenberg K (1965) J Chem Phys 43:S97
137. Pulay P (1983) Chem Phys Letters 100:151
138. Pipek J (1989) Int J Quantum Chem 36:487
139. Luken WL (1983) J Chem Phys 78:5729
140. March NM (1999) in: Surján PR (ed) Correlation and Localization. Top Curr Chem, Vol. 203, Springer, Berlin Heidelberg New York, p. 201

Extended Geminal Models

I. Røeggen

Department of Physics, Faculty of Science, University of Tromsø,
N-9037 Tromsø, Norway. *E-mail: inge.roeggen@phys.uit.no*

The extended geminal models are reviewed with emphasis both on their conceptual structure and computational feasibility. A new numerical model which drastically reduced the computation time at the cost of a very small reduction in accuracy, is introduced. A review of the applications of the extended geminal models to studies of intermolecular interactions is given, and the neon dimer is considered in more detail to illustrate the properties of these models.

Keywords: Localization, interpretability, electron correlation, intermolecular interactions

Topics in Current Chemistry, Vol. 203

1 Introduction

Localization is a key element in order to correlate quantum mechanical calculations and central chemical concepts such as the chemical bond, lone pairs, atoms in a molecule and valence. In a conventional approach one usually applies a localization procedure to the orbitals of a Slater determinant which defines a many-electron function. Unfortunately, these localization procedures are not a proper part of the basic theory. Even though these procedures might be based on reasonable physical arguments, they are subjected to an element of arbitrariness. However, there is a class of self-consistent field models where the occupied orbitals are uniquely determined, and where the localized character of the orbitals is a result of the optimization of the many-electron function. Within this class we notice the APSG (anti-symmetric product of strongly orthogonal geminals) model [1] and modern valence bond theory [2–5]. The extended geminal (EXGEM) models introduced by Røeggen [6–14] can be based on an APSG function as the zero-level approximation. Hence, the interpretative features which can be attributed to the APSG model can be transferred to this class of EXGEM models. In addition the EXGEM models have an additive structure which facilitates the analysis of electron correlation contributions.

In this work we shall review the essential features of the extended geminal models, sketch some numerical refinements of the models, consider some important applications and present some new numerical results on the neon dimer.

2 Extended Geminal Models (EXGEM)

In this section we present the extended geminal (EXGEM) ansatz, discuss the choice of root function, and comment on numerical models, i.e. different approximations to the full configuration interaction (FCI) equations defining the terms in the general EXGEM model.

2.1 The General EXGEM Ansatz

The general extended geminal model is a particular way of constructing a FCI wave function [6]. For a closed shell $2N$-electron system we have the following ansatz for the wave function:

$$\Phi^{\mathrm{EXG}} = \Phi^{\mathrm{APSG}} + \sum_{K=1}^{N} M^{[N-1,1]} \mathcal{A}^{[2N]} \left\{ \Phi_{[K]}^{[2N-2]} \Omega_K^{[2]} \right\}$$

$$
\begin{aligned}
&+ \sum_{K_1<K_2}^{N} M^{[N-2,2]} \mathcal{A}^{[2N]} \left\{ \Phi^{[2N-4,4]}_{[K_1K_2]} \Omega^{[4]}_{K_1K_2} \right\} + \cdots \\
&+ \sum_{K_1<K_2<\cdots<K_p}^{N} M^{[N-p,p]} \mathcal{A}^{[2N]} \left\{ \Phi^{[2N-2p,2p]}_{[K_1K_2\cdots K_p]} \Omega^{[2p]}_{K_1K_2\cdots K_p} \right\} + \cdots + \Omega^{[2N]}_{K_1K_2\cdots K_N} \\
&= \Phi^{\mathrm{APSG}} + \sum_{K=1}^{N} \Psi_K + \sum_{K_1<K_2}^{N} \Psi_{K_1K_2} + \cdots + \sum_{K_1<K_2<\cdots<K_p}^{N} \Psi_{K_1K_2\cdots K_p} + \cdots + \Psi_{1\,2\cdots N} \\
&= \Phi^{\mathrm{FCI}} . \qquad (1)
\end{aligned}
$$

In Eq. (1) Φ^{APSG} denotes the APSG function, $\{M^{[N-p,p]}\}$ are normalization constants, $\mathcal{A}^{[2N]}$ is the antisymmetrizer, $\Phi^{[2N-2,2]}_{[K]}$ is the APSG approximation of the $(2N-2)$ electron cluster generated by deleting the geminal of electron pair K in Φ^{APSG}, $\Omega^{[2]}_{k}$ is a two-electron correction function. There are analogous definitions of the higher order terms in Eq. (1). In the more compact notation, Ψ_K represents a single pair correction term, $\Psi_{K_1K_2}$ a double pair correction term, and so on. The expansion Φ^{EXG} is identical to a FCI expansion.

The energy can formally be evaluated within the framework of the method of moments. Since by construction we have

$$
\left\langle \Phi^{\mathrm{APSG}} \mid \Phi^{\mathrm{EXG}} \right\rangle = 1 , \qquad (2)
$$

it follows

$$
\begin{aligned}
E^{\mathrm{EXG}} &= \left\langle \Phi^{\mathrm{APSG}} \mid H\Phi^{\mathrm{EXG}} \right\rangle \\
&= E^{\mathrm{APSG}} + \sum_{K=1}^{N} \varepsilon_K + \sum_{K_1<K_2}^{N} \varepsilon_{K_1K_2} + \cdots \\
&\quad + \sum_{K_1<K_2<\cdots<K_p}^{N} \varepsilon_{K_1K_2\cdots K_p} + \cdots + \varepsilon_{1\,2\cdots N} , \qquad (3)
\end{aligned}
$$

where

$$
\begin{aligned}
E^{\mathrm{APSG}} &= \left\langle \Phi^{\mathrm{APSG}} \mid H\Phi^{\mathrm{APSG}} \right\rangle && (4) \\
\varepsilon_K &= \left\langle \Phi^{\mathrm{APSG}} \mid H\Psi_K \right\rangle && (5) \\
\varepsilon_{K_1K_2} &= \left\langle \Phi^{\mathrm{APSG}} \mid H\Psi_{K_1K_2} \right\rangle && (6) \\
\varepsilon_{K_1K_2\cdots K_p} &= \left\langle \Phi^{\mathrm{APSG}} \mid H\Psi_{K_1K_2\cdots K_p} \right\rangle && (7) \\
\varepsilon_{1\,2\cdots N} &= \left\langle \Phi^{\mathrm{APSG}} \mid H\Psi_{1\,2\cdots N} \right\rangle && (8)
\end{aligned}
$$

In Eqs. (3)–(8), H is the conventional spin-independent electronic Hamiltonian, i.e.

$$
H = \sum_{i=1}^{2N} h(i) + \sum_{i<j}^{2N} g(i,j) . \qquad (9)
$$

The interpretation of the terms in Eq. (1) is very simple. The function

$$\Psi^{[K]} = \Phi^{\mathrm{APSG}} + \Psi_K \tag{10}$$

describes the correlated state of electron pair K when the rest of the system is described by an APSG approximation. Similarly,

$$\Psi^{[K_1K_2]} = \Phi^{\mathrm{APSG}} + \Psi_{K_1} + \Psi_{K_2} + \Psi_{K_1K_2} \tag{11}$$

describes the correlated state of the electron pairs K_1 and K_2 when the rest of the system is described by an APSG function. The interpretation of the higher order terms has an equally simple structure.

The derivation of the general extended geminal model [6] is based on infinite order Rayleigh–Schrödinger perturbation theory. A reordering of the terms in the perturbation expansion leads to a set of FCI equations defining the terms $\{\Psi_K\}, \{\Psi_{K_1K_2}\}, \ldots$ and the corresponding energy corrections. These FCI equations are in accordance with the interpretation of the terms $\{\Psi_K\}, \{\Psi_{K_1K_2}\}, \ldots$ given in the preceding paragraph.

2.2 The Root Functions in EXGEM Models

The APSG functions in Eq. (1) has the form

$$\Phi^{\mathrm{APSG}} = M^{[N]}\mathcal{A}^{[2N]}\left\{\prod_{K=1}^{N}\tilde{\Lambda}_K(x_{2K-1}, x_{2K})\right\}, \tag{12}$$

where $M^{[N]}$ is a normalization constant determined by the equation

$$\left\langle \Phi^{\mathrm{APSG}} \mid \Phi^{\mathrm{APSG}} \right\rangle = 1\,. \tag{13}$$

The geminals $\{\tilde{\Lambda}_K\}$ are products of an orbital part and a spin function which is singlet coupled, i.e.

$$\tilde{\Lambda}_K(x_1, x_2) = \Lambda_K(\mathbf{r}_1, \mathbf{r}_2)\vartheta_0(\sigma_1, \sigma_2)\,. \tag{14}$$

The orbital space is partitioned into $N + 1$ subspaces, one subspace for each electron pair and a virtual space. The geminals are defined in terms of the orbitals of the associated orbital space:

$$\Lambda_K(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k=1}^{n_K} a_k^K \varphi_k^K(\mathbf{r}_1)\varphi_k^K(\mathbf{r}_2)\,. \tag{15}$$

In Eq. (15), $\{\varphi_k^K\}$ are natural orbitals (NOs). The ordering of the orbitals is according to the relation

$$|a_k^K| \geq |a_{k+1}^K|; \qquad k = 1, ..., n_K\,. \tag{16}$$

Since the geminals are normalized we have

$$\sum_{k=1}^{n_K} |a_k^K|^2 = 1\,. \tag{17}$$

Let $\{\varphi_i^V;\ \ i = 1, ..., n_V\}$ denote an arbitrary orthonormal basis of the virtual space. The following orthonormality conditions are then valid:

$$\left\langle \varphi_k^K \mid \varphi_l^L \right\rangle = \delta_{KL}\delta_{kl}\,, \tag{18}$$

$$\left\langle \varphi_k^K \mid \varphi_i^V \right\rangle = 0\,, \tag{19}$$

$$\left\langle \varphi_i^V \mid \varphi_j^V \right\rangle = \delta_{ij}\,. \tag{20}$$

Equation (18) expresses the strong orthogonality condition.

A proper APSG function is characterized by $n_K > 1$ for all geminals. In the case $n_K = 1$, $K = 1, ..., N$; the APSG function is simply a restricted Hartree–Fock (RHF) wave function. By using $n_K = 1$ for the core geminals and $n_K = 2$ for the valence geminals, the APSG approximation is identical to a generalized valence bond perfect pairing (GVB/PP) scheme [15–17]. It should be emphasized at this point that it is only the dimensions of the orbital subspaces which are fixed a priori. The form of the orbitals and the expansion coefficients $\{a_k^K\}$ are determined by the optimization procedure [18–20].

By assuming $n_K > 1$ for the valence geminals, one of the advantages of using an APSG function, is the unique character of the geminals. There can be no mixing of the geminals without changing the wave function Φ^{APSG}. For a large group of molecules, the geminals have a localized character and thereby allowing a simple identification with lone pairs and electron-pair bonds. An excellent exposition of this aspect of the APSG model is given by McWeeny [21] in his monograph "Coulson's Valence".

The localized character of the geminals can be displayed by introducing the concepts of charge centroids [22] and charge ellipsoids [23, 24] of the geminal one-electron densities. By using these localization measures, we can classify the geminals as core electron pairs, lone pairs and bond pairs [25].

There is an arbitrariness in the choice of dimensions $\{n_K\}$ of the geminal orbital subspaces. However, one can argue that for the valence geminals the proper choice is $n_K = 2$ as in the GVB/PP scheme. In this case there is one orbital for each electron. One is then guaranteed a qualitatively correct picture of the disruption of an electron-pair bond.

There is a subtle weakness in adopting a proper APSG function as the root function in the EXGEM models. It has its origin in the basis sets used for calculating the energy terms $\{\varepsilon_K\}, \{\varepsilon_{KL}\}, \ldots$, Eqs. (4)–(7). The correlation term $\Omega_K^{[2]}$ is expressed in terms of the orbital basis set

$$\{\varphi_k^K;\ \ k = 1, ..., n_K\} \cup \{\varphi_i^V;\ \ i = 1, ..., m_V\}\,,$$

the term $\Omega_{K_1K_2}^{[4]}$ is expressed in terms of

$$\{\varphi_{k_1}^{K_1};\ \ k_1 = 1, ..., n_{K_1}\} \cup \{\varphi_{k_2}^{K_2};\ \ k_2 = 1, ..., n_{K_2}\} \cup \{\varphi_i^V;\ \ i = 1, ..., m_V\}\,,$$

and with similar constructions for the higher order terms. As a consquence the weakly occupied natural orbitals (NOs) of the partner geminals to the geminal K, i.e.

$$\bigcup_{L\neq K}\{\varphi_l^L;\ \ l=2,...,n_L\},$$

are excluded in describing the correlation term $\Omega_K^{[2]}$. Similarly, the weakly occupied NOs

$$\bigcup_{L\neq K_1,K_2}\{\varphi_l^L;\ \ l=2,...,n_L\}$$

are excluded in describing the correlation term $\Omega_{K_1K_2}^{[4]}$. For triple pair terms and higher order terms we have an effect of the same kind. This orbital deficiency at a certain level is corrected at the next level. However, from a computational point of view, it is very inconvenient since it implies that to obtain a high accuracy in a calculation, the truncation of the EXGEM expansion, Eq. (1), must be at a high order. This particular weakness of the EXGEM models was discovered when the models were applied to weakly bonded complexes. In a calculation on the HF dimer, using an APSG function with $n_K = 2$ for the valence geminals, and an EXGEM expansion truncated at the double pair level, the EXGEM value for the binding energy of the dimer was only half the value of the correct one. A careful analysis showed that this disastrous result could be traced back to the mentioned orbital deficiency when the intersystem correlation terms $\{\varepsilon_{K_1K_2}\}$ were calculated. Hence, when EXGEM models are used for a study of intermolecular interactions, it is most convenient to use a RHF function as the root function or basic approximation.

By adopting a RHF function as the root function, the occupied orbitals $\{\varphi_1^K\}$ are no longer unique. The RHF function is invariant by any unitary transformation of the orbitals. To obtain localized orbitals we have to resort to a localization scheme. In most applications of the EXGEM models based on a RHF function as the basic approximation, localized orbitals are determined by minimizing the Coulomb repulsion energy between the corresponding electron pairs [26]. Recently, Ahmadi and Røeggen [27] introduced a new localization scheme for the occupied orbitals in supermolecule calculations of intermolecular interactions. The localized supersystem orbitals are obtained by minimizing a least-square deviation from the isolated subsystem orbitals. This procedure eliminates the arbitrariness inherent in conventional localization schemes.

2.3
Energy Partitioning as an Interpretative Tool

One important feature of extended geminal models is an energy partitioning which leads to a simple interpretation of the total energy. When the geminals are localized in space, it is possible to identify molecular fragments. These

fragments can be of three types: (1) the fragment consists both of nuclei and electron pairs, (2) the fragment consists of only electron pairs, and (3) the fragment consists of only nuclei. Røeggen and Wisløff-Nilssen [28] demonstrated that the total electronic energy, defined as the total energy in the absence of nuclear motion, i.e.

$$E_{\text{tot}}^{\text{EXG}} = E^{\text{EXG}} + E_{\text{nuc}}\,, \tag{21}$$

where E_{nuc} is the nuclear electrostatic energy, can be written as a sum of intra- and interfragment energies. If $\gamma, \delta, \tau, \ldots$ denote the different fragments, we have

$$E_{\text{tot}}^{\text{EXG}} = \sum_{\gamma} E_{\gamma} + \sum_{\gamma<\delta} E_{\gamma\delta} + \sum_{\gamma<\delta<\tau} E_{\gamma\delta\tau} + \cdots\,. \tag{22}$$

In Eq. (22), E_{γ} denotes the internal energy of a fragment, $E_{\gamma\delta}$ denotes the interaction between the fragments γ and δ, and so on. The two-fragment term $E_{\gamma\delta}$ includes the Coulomb interaction between the relevant charge and charge distributions, exchange interaction (with its origin in the antisymmetrization of the wave function) and correlation terms. The three-fragment terms, i.e. $\{E_{\gamma\delta\tau}\}$, and higher-order terms are interfragment correlation contributions.

There is of course no unique way of partitioning a molecular complex. However, there might exist natural fragments. A natural fragment can be defined with respect to a particular distortion of the molecular system. If due to the distortion, the fragment γ exhibits a change in the internal energy E_{γ} which is of the same order of magnitude or smaller than the corresponding change in the total electronic energy, one might denote γ as a natural fragment with respect to this distortion. Røeggen and Wisløff-Nilssen used this approach in describing the difference in equilibrium bond angles in H_2O and H_2S, NH_3 and PH_3. For weakly bonded complexes there are obvious natural fragments or subsystems. A particular energy decomposition scheme for these type of complexes has been advocated by Røeggen [29] and Røeggen and Dahl [30].

2.4 Numerical Models

The energy correction $\varepsilon_{K_1K_2\cdots K_p}$ is in principle defined by the lowest eigenvalue λ_{min} of the FCI equation

$$H_{\text{eff}}(1,2,...,2p)\Phi_{\lambda}(1,2,...,2p) = \lambda\Phi_{\lambda}(1,2,...,2p)\,, \tag{23}$$

and corrections obtained at the lower levels [6]. In Eq. (23), H_{eff} is an effective Hamiltonian for the $2p$-electron cluster in question. To solve FCI equations for higher order clusters, i.e. $2p \geq 4$, with large basis sets, is out of question from a computational point of view. Approximations have to be introduced. In the development of EXGEM models, several approximations have been constructed [7–14]. The very recent model advocated by Røeggen and Wind [14] is the most elaborate one. This latter model is based on an RHF function as the root function. The FCI equations for the four-electron clusters are solved by

using a restricted set of NOs. The six-electron FCI equations are approximated by coupled cluster (CC) approximations: CCSD, CCSD(T) and CCSD(TQ). The acronyms are the conventional ones; SD means single and double excitations, CCSD(T) is a model correct up to fourth order, and CCSD(TQ) includes all corrections necessary to make the model correct up to fifth order energy.

To improve the computational feasibility of the EXGEM models, we now introduce the following simplifications. For the single pair corrections $\{\varepsilon_K\}$ we first calculate an MP2-approximation, denoted by $\{\varepsilon_K^{\mathrm{MP2}}\}$, using the proper subsystem basis [14]. The MP2-approximation requires only a small subset of transformed two-electron integrals. On the basis of the MP2-approximations, the geminals are ordered in groups. Within a fixed numerical accuracy, all the geminals in a group have the same value for the MP2 correction. This procedure defines which geminals are symmetry related. By following the assumption that if the MP2 corrections are equal for a set of geminals, then the FCI corrections must also be equal, we calculate only *one* FCI correction for each group. Furthermore, this FCI correction is calculated using a set of NOs constructed using the first order wave function correction. The dimension of the orbital space defined by this set of NOs, is considerably smaller than the full virtual orbital space for the subsystem. Let $\tilde{\varepsilon}_K^{\mathrm{FCI}}$ and $\tilde{\varepsilon}_K^{\mathrm{MP2}}$ denote respectively the FCI and MP2 corrections when this truncated basis set is used for the calculations. We then introduce the following approximation for the single pair correction ε_K:

$$\begin{aligned} \varepsilon_K &= \tilde{\varepsilon}_K^{\mathrm{FCI}} - \tilde{\varepsilon}_K^{\mathrm{MP2}} + \varepsilon_K^{\mathrm{MP2}} \\ &= \tilde{\varepsilon}_K^{\mathrm{FCI}} + \left(\varepsilon_K^{\mathrm{MP2}} - \tilde{\varepsilon}_K^{\mathrm{MP2}}\right) \\ &= \tilde{\varepsilon}_K^{\mathrm{FCI}} + \varepsilon_K^{\mathrm{BSE}}. \end{aligned} \tag{24}$$

The term $\varepsilon_K^{\mathrm{BSE}}$ represents a basis set extension (BSE) effect calculated at the MP2-level. As the dimension of the truncated orbital space is increased, the magnitude of the term $\varepsilon_K^{\mathrm{BSE}}$ is reduced.

For the double pair corrections we introduce a similar procedure. Both for intra- and intersubsystem corrections we calculate MP2-like corrections $\{\varepsilon_{KL}^{\mathrm{MP2}}\}$. Based on these corrections, the four-electron clusters are ordered in groups, all elements in a group having the same value for the MP2-like correction. As in the single pair case, all elements or four-electron clusters in a group are symmetry related. For each group we calculate a four-electron FCI correction based on a truncated set of NOs [14]. Hence, we use the following approximation for the four-electron FCI corrections:

$$\begin{aligned} \varepsilon_{KL} &= \tilde{\varepsilon}_{KL}^{\mathrm{FCI}} - \tilde{\varepsilon}_{KL}^{\mathrm{MP2}} + \varepsilon_{KL}^{\mathrm{MP2}} \\ &= \tilde{\varepsilon}_{KL}^{\mathrm{FCI}} + \left(\varepsilon_{KL}^{\mathrm{MP2}} - \tilde{\varepsilon}_{KL}^{\mathrm{MP2}}\right) \\ &= \tilde{\varepsilon}_{KL}^{\mathrm{FCI}} + \varepsilon_{KL}^{\mathrm{BSE}}. \end{aligned} \tag{25}$$

As in the two-electron case, the superscript ˜ denotes a quantity calculated in terms of the truncated set of NOs. The basis set extension effect (BSE) is calculated at the MP2-level.

Since the FCI corrections are based on different truncated virtual orbital spaces, it is very important to have an efficient algorithm for the transformation of two-electron integrals. A Cholesky decomposition of the two-electron matrix is then very convenient [31, 32]. A two-electron integral $[\mu\nu \mid \lambda\sigma]$ is related to the integral tables (obtained by the Cholesky decomposition) $\{L_{\mu\nu;t};\quad t = 1, ..., r_\delta\}$ by the relation

$$[\mu\nu \mid \lambda\sigma] = \sum_{t=1}^{r_\delta} L_{\mu\nu;t} L_{\lambda\sigma;t}\,, \tag{26}$$

where r_δ is the effective numerical rank of the two-electron matrix. The transformation of the two-electron integrals is performed in accordance with the formulas

$$\psi_i = \sum_{\mu=1}^{m} X_\mu U_{\mu i} \tag{27}$$

$$L_{ij;t} = \sum_{\mu,\nu=1}^{m} U_{\mu i} L_{\mu\nu;t} U_{\nu j} \tag{28}$$

$$[ij \mid kl] = \sum_{t=1}^{r_\delta} L_{ij;t} L_{kl;t} \tag{29}$$

The algorithm defined by Eqs. (27)–(29) can be effectively coded for both vector and parallel processor computers.

3 Applications

In this section we shall briefly comment on previous works based on extended geminal models (Sect. 3.1) and present a case study of the neon dimer (Sect. 3.2).

3.1 Previous Works

The extended geminal models have been used to calculate the interatomic potential for the ground state of diatomic complexes comprising an alkali ion and a noble gas atom: $NeLi^+$ [33]; $ArLi^+$ [34]; $ArNa^+$, $NeNa^+$, $HeNa^+$ [35]; $HeKa^+$ [36]. On the basis of the potentials for $NeLi^+$ and $HeKa^+$, mobility coefficients were calculated [37, 38]. There was a very good agreement between the calculated and measured mobility coefficients. The deviation being of the order of 1% or lower.

Røeggen and Ahmadi [39] have studied the F- and H-bonded isomers of HF/ClF. The F-bonded isomer was found to have the largest binding energy. The energy partitioning scheme was utilized in order to obtain a simple phy-

sical explanation concerning the prediction of the shapes of van der Waals complexes by pure electrostatic interactions. An analysis of the electron donor-acceptor complexes H_2OF_2, H_2OCl_2 and H_2OClF, was performed by Dahl and Røeggen [40]. The partitioning scheme was utilized also in this work to illuminate the bonding of the complexes.

The dimer $(H_2)_2$ and the trimer $(H_2)_3$ have been extensively studied using extended geminal models. Røeggen and Wind [41] presented a detailed analysis of the bonding of the $(H_2)_2$ dimer. Two extreme geometrical structures were considered: the T structure and the linear structure. A full potential energy surface of $(H_2)_2$ within the rigid rotor approximation was calculated by Wind and Røeggen [42]. Wind and Røeggen [43] demonstrated that the isotropic potential of the dimer, could be calculated directly to a very good approximation by considering a special orientation of the two molecules in the complex. The relative error of the isotropic potential calculated in this way, was less than 1%. Wind and Røeggen [44] have also studied the non-additivity in the $(H_2)_3$ trimer. The isotropic part of the three-body non-additive energy was derived. Near the equilibrium geometry the global error of the three-body potential was estimated to be 7% or less. At large distances the results were found to be in accordance with the Axilrod–Teller–Muto triple-dipole approximation.

Røeggen and Almlöf [45, 46] have studied the three-body potential for the ground state of He_3 using extended geminal models. For the triangular structure and an interatomic distance $R = 5.6$ au (i.e. a distance close to the equilibrium distance for the dimer) the following value was found for the three-body potential: -0.20 ± 0.02 μhartree. In the long range the calculated three-body potential was very close to the Axilrod–Teller–Muto triple dipole energy.

Røeggen and Almlöf [47] used an extended geminal to calculate the interatomic potential for the ground state of Be_2. The calculated binding energy was -3.79 mhartree, to be compared with the experimental result of -3.82 ± 0.05 mhartree.

3.2
The Neon Dimer as a Case Study

The basis set adopted in the study of the neon dimer is a contracted $[8s,6p,5d,4f,3g]$ set of Gaussian type functions (GTFs). The set is constructed in the following way. We start by an uncontracted $(12s,7p)$ set of GTFs [49]. This set is contracted to $[6s,4p]$ using contraction coefficients from an atomic SCF calculation and keeping the most diffuse functions uncontracted. To this set of contracted s- and p-type functions we add a set of $(3d,2f,1g)$ polarization functions [50]. The resulting set of $[6s,4p,3d,2f,1g]$ GTFs is augmented by two sets of diffuse functions in each symmetry. The exponents of the diffuse functions are determined as an even-tempered extension of the smaller set, and for the g-type functions we use the same ratio as for the f-type functions.

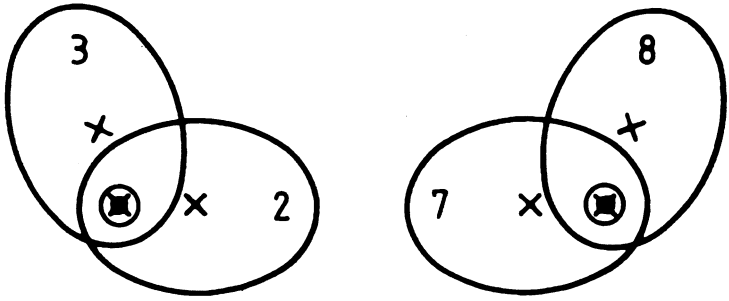

Fig. 1. Intersection between the xy-plane and the charge ellipsoids of the localized orbitals of the neon dimer. Relative interatomic distance twice as large as on figure. Charge centroids are marked with a cross and nuclear positions with a dot. The numbering of the ellipsoids is utilized in Table 3.

In this work we use an integral threshold of 10^{-8} au for the Cholesky decomposition of the two-electron matrix.

For the dimer we choose an interatomic distance equal to 5.84 au. This distance is close to the assumed minimum of the potential [51].

The construction of the localized orbitals are as follows. For the isolated neon atom the 1*s*-orbital is chosen to be identical to the canonical 1*s*-orbital. Four equivalent valence orbitals are determined by the Boys localization procedure [52]. These orbitals have a tetrahedral-like structure, i.e. the charge centroids of the orbitals coincide with the corners of a tetrahedron. As a starting point for the minimal distortion localization (MDL) of the orbitals of the dimer, for each isolated atom a valence charge centroid is positioned along the interatomic axis and between the nuclei. Hence, for each atom three charge centroids are pointing away from the neighbouring atom. Furthermore, the charge centroids outside the interatomic region are rotated to an eclipsed position (see Fig. 1). The MDL procedure yields localized orbitals which are very similar to the orbitals of the isolated atoms [27].

The first question to be addressed is the question of the accuracy of the advocated numerical model. For the intrasystem terms, the errors in the calculated correlation terms compared with the FCI definition of the corresponding terms, should be smaller than the truncation error of the extended geminal expansion. As demonstrated by Røeggen and Wind [14], by truncating the extended geminal expansion at the triple pair level, the error in the calculated correlation energy compared with the FCI result, is less than 1%. With the adopted basis set, the calculated correlation energy for the isolated neon atom is -0.33070071 hartree. By using a truncated virtual space in approximating ε_K and ε_{KL}, Eq. (24) and (25), the sum of these errors should be less than say 1% of the calculated correlation energy. Since there are five single pair terms and ten double pair terms involved, this amounts to a maximum of error for the individual terms of the order of 0.0002 hartree. In Table 1 we have displayed the value

Table 1. A valence shell single pair correction term for the neon atom as a function of the dimension of the truncated virtual space[a,b]

m	$\tilde{\varepsilon}_K^{\mathrm{FCI}}$	$\tilde{\varepsilon}_K^{\mathrm{BSE}}$	ε_K
20	−0.027 806 32	−0.000 694 08	−0.028 500 41
30	−0.028 143 72	−0.000 243 08	−0.028 386 80
40	−0.028 254 47	−0.000 081 13	−0.028 335 60
50	−0.028 285 58	−0.000 032 53	−0.028 318 11
60	−0.028 297 31	−0.000 011 86	−0.028 307 17

[a] Energies in hartree.

[b] All valence geminals are equivalent.

of a valence shell single pair correction term as a function of the dimension of the truncated virtual space. By considering the basis set extension effect, $\tilde{\varepsilon}_K^{\mathrm{BSE}}$, we can see that this criterion is satisfied for a dimension of the truncated space between 30 and 40. The double pair correction terms require a somewhat larger truncated space. According to Table 2 we must use a dimension which is between 50 and 60. For the intersystem correlation terms the criterion must be formulated in absolute terms. Since the error in the calculated potential is assumed to be less than 5 μhartree, the error in the calculated intersystem double pair correction terms should be smaller than say 0.2 μhartreee. According to Table 3, this criterion is satisfied when the dimension of the truncated virtual space is equal to 60. In the calculation on the neon dimer we therefore use 60, 60 and 70 as the dimensions of the truncated virtual spaces for the single, double and triple pair corrections, respectively.

In Table 4 we present results from the calculation on the neon dimer. The equilibrium distance of the potential is not determined in this work. But based on previous experience with extended geminal models, say on the He_2 dimer [45] and the Be_2 dimer [47], we can safely assume that the equilibrium distance obtained by a geometry optimization, is close to the experimental value. Hence, the quantity $-U$ ($R = 5.84$ au), where $U(R)$ is the interatomic

Table 2. A valence shell double pair correction term for the neon atom as a function of the dimension of the truncated virtual space[a,b]

m	$\tilde{\varepsilon}_{KL}^{\mathrm{FCI}}$	$\varepsilon_{KL}^{\mathrm{BSE}}$	ε_{KL}
20	−0.030 217 10	−0.002 960 06	−0.033 177 16
30	−0.031 846 65	−0.001 351 78	−0.033 198 43
40	−0.032 634 74	−0.000 577 03	−0.033 211 77
50	−0.032 948 94	−0.000 238 85	−0.033 187 79
60	−0.033 104 46	−0.000 064 96	−0.033 169 42

[a] Energies in hartree.

[b] All pairs of valence geminal are equivalent.

Table 3. Intersystem valence double pair correction terms for the neon dimer as a function of the dimension of the truncated virtual space[a,b,c]

m	(K,L)	$\tilde{\varepsilon}_{KL}^{\mathrm{FCI}}$	$\varepsilon_{KL}^{\mathrm{BSE}}$	ε_{KL}
20	(2,7)	–0.000 053 76	–0.000 005 50	–0.000 059 26
	(2,8)	–0.000 016 96	–0.000 001 79	–0.000 018 75
30	(2,7)	–0.000 058 21	–0.000 002 07	–0.000 060 29
	(2,8)	–0.000 018 19	–0.000 000 81	–0.000 019 00
40	(2,7)	–0.000 059 86	–0.000 001 04	–0.000 060 90
	(2,8)	–0.000 018 63	–0.000 000 50	–0.000 019 13
50	(2,7)	–0.000 060 04	–0.000 000 90	–0.000 060 94
	(2,8)	–0.000 018 81	–0.000 000 37	–0.000 019 18
60	(2,7)	–0.000 060 39	–0.000 000 67	–0.000 061 06
	(2,8)	–0.000 018 90	–0.000 000 28	–0.000 019 18

[a] Energies in hartree.
[b] See Fig. 1 for the numbering of geminals.
[c] Interatomic distance $R = 5.84$ au.

potential, should be close to the binding energy. In this work then we estimate the binding energy to be 135.80 μhartree. According to Asiz and Slaman, the best experimental value is 133.80 μhartree. Pertaining to the results in Table 4, we would like to emphasize the following points. First, there is a small but significant change in the intrasystem correlation energy. For each atom this

Table 4. Partitioning of the interatomic potential for the neon dimer[a,b]

Component	Isolated monomer	Dimer	ΔU
$E_{\gamma}^{\mathrm{RHF}}$	–128 542 001.42	–128 541 701.94	299.48
$\sum_{K=1}^{5} \varepsilon_K$	–127 606.59	–127 601.41	5.18
$\sum_{K<L}^{5} \varepsilon_{KL}$	–213 140.11	–213 125.62	14.49
$\sum_{K<L<M}^{5} \varepsilon_{KLM}$	10 045.99	10 043.44	–2.55
$E_{1,2}^{\mathrm{coul}}$		–404.10	–404.10
$E_{1,2}^{\mathrm{exch}}$		–112.73	–112.73
$\sum \varepsilon_{KL}$ (intersystem)		–236.38	–236.38
$\sum \varepsilon_{KLM}$ (intersystem)		–25.81	–25.81
$\sum \varepsilon_{K_1 \cdots K_p}$ (intersystem – higher order)[c]		10.02	10.02
U			–135.80

[a] Energies in μhartree.
[b] Interatomic distance $R = 5.84$ au.
[c] From Wind and Røeggen [48].

change is 17.12 μhartree, compared with $-330\ 700.71$ μhartree for the total correlation energy for the isolated atom. Second, if for intersystem correlation we include only double pair correction terms, the calculated value for the potential is -120.01 μhartree. By including the triple pair correction terms, this value changes to -145.82 μhartree. Hence, an accurate value for the potential requires that even higher order terms are included. In this work we have added a coupled-cluster correction for this higher order term [44]. This leads finally to the accurate result displayed in Table 4.

4 Concluding Remarks

The extended geminal models have two main advantages. First, the conceptual structure which facilitates interpretation. This property is utilized in several studies on intermolecular interactions where energy decomposition schemes illuminate the character of the bonding. Second, the models are highly accurate. This feature is related to the FCI corrections on which the models are based. The reported calculations on few-electron systems illustrate this point. However, as demonstrated by the calculation on the neon dimer reported in this work, a high accuracy of a calculation on larger systems, require that at least triple pair corrections are included.

The extended geminal models take care of the electron correlation problem for closed shell systems. The basis set problem becomes the bottleneck. By adopting the Cholesky decomposition of the two-electron matrix, and storing the integral tables, one can handle basis sets comprising up to say, 500 functions. For systems requiring more than 500 functions one possibility is to recalculate integrals. However, we are presently working on an alternative approach. This approach is based on defining atoms in a molecule or solid. For intraatomic wave-function terms a local basis set is adopted. When considering the diatomic terms, a two-center basis is adopted, the triatomic terms require a three-center basis and so on. In this way one might deal with large systems without the necessity of adopting extremely large basis sets. Progress in this particular research program will be reported elsewhere.

5 References

1. Hurley AC, Lennard-Jones J, Pople JA (1953) Proc Roy Soc Ser A220:446
2. Gerratt J (1971) Adv Atom Mol Phys 7:141
3. Gerratt J (1976) Proc Roy Soc Lond A350:363
4. Gerratt J, Raimondi M (1976) Proc Roy Soc Lond A371:525
5. Cooper DL, Gerratt J, Raimondi M (1991) Chem Rev 91:929
6. Røeggen I (1983) J Chem Phys 79:5220
7. Røeggen I (1983) J Chem Phys 78:2485
8. Røeggen I (1986) J Chem Phys 85:262

9. Røeggen I (1987) Int J Quantum Chem 31:951
10. Røeggen I (1988) J Chem Phys 89:441
11. Røeggen I (1990) Int J Quantum Chem 37:585
12. Røeggen I, Ahmadi GR, Wind PA (1993) J Chem Phys 99:277
13. Røeggen I, Almlöf J, Ahmadi GR, Wind PA (1995) J Chem Phys 102:7088
14. Røeggen I, Wind PA (1996) J Chem Phys 105:2751
15. Ladner RC, Goddard III WA (1969) J Chem Phys 51:1073
16. Goddard III WA, Ladner RC (1971) J Am Chem Soc 93:6750
17. Hunt WJ, Hay PJ, Goddard III WA (1972) J Chem Phys 57:738
18. Saunders VR, Guest MF (1974) In: Proceedings of the SRC Atlas Symposium, No. 4: Quantum
    Chemistry – The State of the Art
19. Bobrowicz FW, Goddard III WA (1979) In: Schaefer III HF (ed) Modern Theoretical Chemistry.
    Plenum Press, New York
20. Røeggen I (1981) Int J Quantum Chem 19:319
21. McWeeny R (1979) Coulson's Valence, Oxford University Press, Oxford
22. Boys SF (1966) In: Löwdin PO (ed) Quantum Theory of Atoms, Molecules and the Solid State.
    Academic Press, New York
23. Robb MA, Haines WJ, Csizmadia IG (1973) J Am Chem Soc 95:42
24. Csizmadia IG (1985) In: Chalvet O, Daudel R, Diner S, Malrieu JP (eds) Localization and Delocalization in Quantum Chemistry. Reidel, Dordrecht
25. Røeggen I (1991) Int J Quantum Chem 40:951
26. Edmiston C, Ruedenberg K (1963) Rev Mod Physics 35:457
27. Ahmadi GR, Røeggen I (1977) Theor Chem Acc 97:41
28. Røeggen I, Wisløff–Nilssen E (1987) J Chem Phys 86:2869
29. Røeggen I (1986) J Chem Phys 85:262
30. Røeggen I, Dahl T (1992) J Am Chem Soc 114:511
31. Beebe NHF, Linderberg J (1977) Int J Quantum Chem 12:683
32. Røeggen I, Wisløff–Nilssen E (1986) Chem Phys Lett 132:154
33. Røeggen I, Skullerud HR (1992) J Phys B At Mol Opt Phys 25:1795
34. Ahmadi GR, Røeggen I (1994) J Phys B At Mol Opt Phys 27:5603
35. Ahmadi GR, Almlöf J, Røeggen I (1995) Chem Phys 199:33
36. Røeggen I, Skullerud HR, Elford MT (1996) J Phys B At Mol Opt Phys 29:1913
37. Skullerud HR, Røeggen I, Løvaas TH (1992) J Phys B At Mol Opt Phys 25:1811
38. Skullerud HR, Elford MT, Røeggen I, (1996) J Phys B At Mol Opt Phys 29:1925
39. Røeggen I, Ahmadi GR (1994) J Mol Struct (Theochem) 307:9
40. Dahl T, Røeggen I (1996) J Am Chem Soc 118:4152
41. Røeggen I, Wind P (1992) Chem Phys 167:247
42. Wind P, Røeggen I (1992) Chem Phys 167:263
43. Wind P, Røeggen I (1993) Chem Phys 174:345
44. Wind P, Røeggen I (1996) Chem Phys 211:179
45. Røeggen I, Almlöf J (1995) J Chem Phys 102:7095
46. Røeggen I, Almlöf J (1996) J Mol Struct (Theochem) 388:331
47. Røeggen I, Almlöf J (1996) Int J Quantum Chem 60:453
48. Wind P, Røeggen I (1996) Chem Phys 206:307
49. Van Duijneveldt FB (1971) IBM Research Report RJ945
50. Dunning Jr TH (1989) J Chem Phys 90:1007
51. Aziz RA, Slaman MJ (1987) Chem Phys 130:187
52. Boys SF (1960) Rev Mod Phys 32:296

Ab Initio Modern Valence Bond Theory

Mario Raimondi[1] · David L. Cooper[2]

[1]Dipartimento di Chimica Fisica ed Elettrochimica and Centro CNR CSRSRC,
Università di Milano, Via Golgi 19, I-20133 Milano, Italy. *E-mail: raim@rs6.csrsrc.mi.cnr.it*
[2]Department of Chemistry, University of Liverpool, P.O. Box 147,
Liverpool, L69 7ZD UK. *E-mail: dlc@liv.ac.uk*

We concentrate in the present account on certain recent developments linked to extensions of spin-coupled theory. In particular, we describe the so-called SCVB* strategy, which employs optimised virtual orbitals to reduce still further the number of nonorthogonal configurations required in accurate calculations of ground and excited state potential energy surfaces. We also outline the CASVB approach, which may be used to generate modern valence bond representations of complete active space self-consistent field wavefunctions or, alternatively, to perform the fully-variational optimisation of quite general types of single- and multiconfiguration modern VB wavefunctions for ground and excited states.

Keywords: Valence bond, Spin-coupled theory, CASVB, Excited states, Intermolecular forces

Topics in Current Chemistry, Vol. 203

1 Introduction

Valence bond theory will always have a special place in chemistry, not least because it provides such appealing visual representations and interpretations of molecular electronic structure. Unfortunately, early computational implementations were hampered by difficulties linked to the nonorthogonality of the orbitals and/or of the structures. This was one reason why valence bond (VB) approaches were largely eclipsed by molecular orbital (MO) theory. These technical problems have now largely been overcome, but the fact remains that calculations that include only the traditional covalent structures tend to deliver disappointing accuracy. In order even to compete with self-consistent field (SCF) calculations, it proves necessary to include significant numbers of additional structures, particularly ionic ones. All of this detracts, of course, from the original appealing simplicity of the basic VB picture. Nonetheless, the language of VB theory continues to permeate much of chemistry.

Coulson and Fischer [1] realised that the admixture of ionic structures into the Heitler and London [2] covalent description of H_2 was entirely equivalent to allowing a (relatively small) degree of delocalisation of the H(1*s*) functions, and it is now widely accepted that the quality of compact VB wavefunctions is strongly related to the freedom of the orbitals to deform and to overlap in the bond-forming directions. Part of the philosophy of so-called modern VB approaches is to allow unprejudiced optimisation of the orbitals. Specifically, we use the label 'classical VB' for wavefunctions based on strictly localised orbitals, whereas the term 'modern valence bond' signifies one or more spatial configurations constructed from completely general, nonorthogonal orbitals, and combined with all allowed ways of coupling together the electron spins so as to obtain the required resultant. Although the orbitals often turn out to be fairly localised, the generation of localised (e.g. atomic-orbital-like) orbitals is not a goal in itself.

Recent years have certainly seen the re-emergence of ab initio VB theory, particularly in 'modern' form, as a serious tool for computational chemistry, with a significant role being played by spin-coupled theory [3–8] and its various extensions. The spin-coupled approach was probably first used by Kaldor [9], who demonstrated for atoms the power of using different spin functions to describe correctly, in a compact way, delicate properties such as electron spin density at the nucleus. Spin-coupled (SC) theory was developed for molecules by Gerratt [3] and his subsequent collaborators. Early work on molecules was also carried out by Goddard and coworkers [10], who initially used an orbital equation method to optimise the orbitals, and named their approach generalised valence bond (GVB). The SC formalism corresponds to *full* GVB. However, most GVB calculation involve strong orthogonality (SO) constraints, according to which only two orbitals at a time are allowed to overlap and, in general, these two electrons are constrained to have opposite spin [11]. An approach to

the fully nonorthogonal case was made by Chipman and Palke [12], in which orbitals were approximately optimised in a cyclic procedure. Their method employed an approximate expression for the energy and relied on using only the perfect-pairing (PP) spin function.

The combination of the SC (*full* GVB) approach and a traditional VB code gave rise to spin-coupled valence bond (SCVB) calculations [4]. Many other groups have also been active in the general area of ab initio modern VB calculations and we mention important contributions such as resonating VB [13], the TURTLE program [14], the breathing orbital method [15], the general multistructural method [16], the tableau function approach of Gallup [17], and the bonded tableau valence bond formalism [18, 19]. Various spin-coupled-like approaches have also been developed by others, including Penotti [20] and Manby and Doggett [21]. In the present article, however, we concentrate on a brief account of some aspects of our own work. Given the availability of a number of reviews of the SC and SCVB approaches [5–8], we emphasise instead the recent development of the SCVB* and CASVB strategies.

2
Spin-Coupled Theory

The spin-coupled wavefunction is based on a single configuration of N singly occupied nonorthogonal orbitals:

$$\Psi = \hat{A}\left(\Phi^{\text{core}}\phi_1\phi_2\ldots\phi_N\Theta^N_{SM}\right) . \tag{1}$$

The spin-coupled orbitals, ϕ_μ, are expanded in an underlying atomic orbital (AO) basis set $\{\chi_i\}$:

$$\phi_\mu = \sum_{i=1}^{M} c_{i\mu}\chi_i \tag{2}$$

without preconceptions as to their degree of localisation. Using an efficient second-order procedure, the SC wavefunction is optimised variationally with respect to the ϕ_μ and to the spin-coupling coefficients, c_{Sk}, which define the optimal N-electron spin eigenfunction Θ^N_{SM} (with total spin S and projection M):

$$\Theta^N_{SM} = \sum_{k=1}^{f^N_S} c_{Sk}\Theta^N_{SM;k} . \tag{3}$$

For the construction of spin eigenfunctions see, for example, Ref. [22]. There are obviously many parallels to the multiconfiguration self-consistent field (MCSCF) methods of MO theory, such as the restriction to a relatively small 'active space' describing the chemically most interesting features of the electronic structure. The 'core' wavefunction for the inactive electrons, Φ^{core}, may be taken from prior SCF or complete active space self-consistent field (CASSCF) calculations, or may be optimised simultaneously with the ϕ_μ and c_{Sk}.

In general, SC wavefunctions are not invariant to linear transformations of the SC orbitals, and so the ϕ_μ and Θ^N_{SM} are a unique outcome of the optimization procedure. The optimal SC orbitals often, but by no means always, turn out to be well-localised functions, similar in shape to their classical VB counterparts. A key difference, however, is the existence of small, but crucial distortions of the orbitals towards other centres. The total spin function, characterised by the spin-coupling coefficients, c_{Sk}, provides useful information about the recoupling of electronic spins as bonds are broken and new ones are formed during chemical reactions. The SC wavefunction may be extended by adding further configurations, in which case we may speak of a multiconfiguration spin-coupled (MCSC) description.

Some of the most efficient algorithms currently available for spin-coupled calculations involve the expansion of the SC wavefunction in terms of Slater determinants, leading to a summation over $(N-p)$-dimensional cofactors for the p-particle density matrix [23]. For calculations involving the full spin space, a further saving can be achieved by the use of projected spin functions [24]. An alternative strategy is provided by CASVB (see Sect. 5).

The SC approach has now been applied to a very wide range of problems, spanning all of the main branches of chemistry, often providing important new insights into the nature of chemical bonding. A number of recent reviews are available [6–8].

3 Spin-Coupled Valence Bond Calculations

At convergence of the spin-coupled minimisation procedure, the orbitals satisfy equations which can be recast in the form [4]

$$\hat{F}^{\text{eff}}_\mu \phi^j_\mu = \varepsilon^j_\mu \phi^j_\mu \,. \tag{4}$$

The effective operators $\hat{F}^{\text{eff}}_\mu$ are each of dimension M, the total number of basis functions. Diagonalisation of each operator gives rise to M orbitals, one of which is the corresponding occupied orbital. The virtual orbitals are denoted ϕ^j_μ, where j labels the position in the list and ε^j_μ plays the role of an orbital energy. In this fashion, a 'stack' of virtual orbitals is obtained for each electronic coordinate. The orbitals are mutually orthogonal within a given stack but they are not in general orthogonal to orbitals in the other stacks.

Excited configurations may be constructed by replacing one or more occupied SC orbitals with a virtual orbital, frequently taken from the same stack as the occupied orbital, in which case it is referred to as a 'vertical' excitation. If the occupied orbital is replaced with a virtual from a different stack, then this is referred to as a 'cross' excitation. A linear combination of the reference SC configuration and the excited configurations described above constitutes a so-called spin-coupled valence bond (SCVB) wavefunction. We use the term *configuration* to denote a particular orbital product, with all possible modes of coupling

together the electron spins. The term *structure* is used to denote instead a particular orbital product together with a particular mode of spin coupling. By allowing double occupancy of the occupied and virtual SC orbitals it is possible to generate additional configurations, which are called spin-coupled ionic configurations, where this nomenclature comes from classical VB terminology. Obtaining the expansion coefficients via resolution of the corresponding secular problem, which requires the evaluation of the hamiltonian and overlap matrices between nonorthogonal VB structures, gives the total energy of the system. This 'nonorthogonal configuration interaction' SCVB procedure can be used for further refinement of the spin-coupled wavefunction, taking account of those contributions to the correlation energy which are not recovered by the single-configuration spin-coupled wavefunction. This same procedure also produces excited states of the N-electron system.

Applications of SCVB up to ca. 1991 were reviewed in Ref. [6]. More recent studies have concentrated on excited states of organic 'π-electron' systems, on intermolecular forces and on charge transfer processes. The SCVB approach turns out to be well suited to the study of charge transfer collisions involving highly stripped atomic ions with helium or atomic hydrogen. Accurate state-dependent cross sections and rate constants are required not only for the modelling of various astrophysical environments but also for fusion plasma density impurity diagnostics. In general, it proves necessary to investigate the potential energy curves, avoided crossings and nonadiabatic radial couplings, $\langle \Psi_i | \partial / \partial R | \Psi_j \rangle$, for *several* states of the quasimolecule over an extended range of nuclear separation, R. Accurate asymptotic energy separations are important, but it is also crucial to maintain a consistently high level of accuracy for all the relevant states over the entire range of R. It has now become almost routine to calculate several states of the same symmetry, as well as other symmetries, with asymptotic energies accurate to very few tenths of an eV, and to compute the nonadiabatic couplings. The state-dependent cross sections obtained via subsequent fully-quantum-mechanical close-coupling calculations [25, 26] are often in close harmony with reliable experimental measurements, when available.

The SCVB method has been used to study all the singlet and triplet valence excited states, as well as the $n = 3{,}4$ singlet and triplet Rydberg states of benzene below the first ionisation potential at 9.25 eV. The 'covalent' valence excited states may be well described using σ/π separation and a frozen σ core, whereas the error in the computed transition energies to the 'ionic' states was somewhat larger, indicating that these states require additional σ/π correlation for their proper description. The Rydberg states are very well described, provided that a suitable σ core is used, derived from calculations on the $C_6H_6^+$ ion. The numerical accuracy for the transition energies [27] was comparable to that from the largest MO-CI or CASSCF-CI calculations reported at the time. The SCVB method is clearly a powerful tool for describing the excited states of medium-sized molecules such as benzene, affording at the same time a clear view of the wavefunctions for the various states. While it turns out that there

is room for improvement, particularly for the excited states of ionic character, it is clear that the SCVB approach is able to identify the physical reasons why certain states are more difficult to describe and thus require a more advanced treatment.

The SCVB method has also been applied to the study of weakly interacting systems. The resulting approach is fully variational and carefully avoids one of the main difficulties that plagues conventional MO supermolecule calculations, namely the basis set superposition error (BSSE). This is achieved in an a priori fashion by expanding the occupied SC orbitals (and the corresponding virtuals) in the basis functions located only on the sub-system with which the orbital is associated. An early application to the LiH … He system gave encouraging results, which highlighted the accuracy of the method relative to previous supermolecule studies. The absolute minimum was found for the collinear approach of the helium atom to the lithium end of LiH. A small shallow secondary minimum, which might influence the dynamics at very low collision energies, was found for the collinear approach, i.e. of helium to the $H^{\delta-}$ end of LiH [28].

The SCVB method is very successful for including dynamical correlation into the description of ground and excited states of systems with a small number of electrons. Higher and higher accuracy requires a significant number of SC virtual orbitals for each active electron and thus a wavefunction formed from many structures. It is obvious that calculations on systems with larger numbers of electrons will rapidly become untenable. To overcome this severe size restriction, it is necessary to find ways to achieve equivalent, or almost equivalent, results with much smaller numbers of virtual orbitals. Considering that the standard SC virtuals already come from diagonalisation of a physically reasonable operator, the next stage in improving virtuals must require a more specific optimisation of the orbitals with respect to the correlation energy contributions to be taken into account by the wavefunction. This is the topic of the next section.

4 SCVB*

We consider first an N-electron all-double-excitation SCVB expansion with one optimal virtual orbital ϕ_i^+ for each stack i:

$$\Psi = C_0\Psi_0 + \sum_{i=1}^{N}\sum_{j>i}^{N} C_{ij}\Psi_{ij} \tag{5}$$

where

$$\Psi_{ij} = \hat{A}\left(\phi_1^0\phi_2^0\ldots\phi_i^+\ldots\phi_j^+\ldots\phi_N^0\sum_{k=1}^{f_S^N} c_{Sk}^{ij}\Theta_{SM;k}^N\right) . \tag{6}$$

In order to reduce the computational effort, we adopt the following simplifying strategy [29, 30]:

(1) We first perform a standard SC calculation to obtain orbitals ϕ_i^0 and the spin-coupling coefficients c_{Sk}^0 (cf. Eqs. (1)–(3)).
(2) The corresponding spin-coupling coefficients c_{Sk}^{ij} for the excited configurations are taken to be the same as c_{Sk}^0.
(3) The virtual orbitals ϕ_i^+ are optimised using a second-order perturbation theory approximation to the energy [31] so that we need only evaluate the diagonal and first row elements of the hamiltonian and overlap matrices:

$$E^{(2)} = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{\left[\langle \Psi_0 | \hat{H} | \Psi_{ij} \rangle - \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \langle \Psi_0 | \Psi_{ij} \rangle \right]^2}{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle \langle \Psi_{ij} | \Psi_{ij} \rangle - \langle \Psi_{ij} | \hat{H} | \Psi_{ij} \rangle} \quad (7)$$

(4) An approximate expression is adopted for the hessian, requiring no additional terms not already computed for the gradient of the energy with respect to the free parameters.
(5) To overcome the possible onset of linear dependence, we project out the occupied SC orbitals from the full basis set, and expand the virtual orbitals in this projected basis.
(6) The resulting orbitals are then used in a standard nonorthogonal CI expansion (single and double vertical excitations) in order to relax the spin-coupling coefficients and to find a variational bound for the energy. We refer to such an expansion as an SCVB* wavefunction. In order to improve the description of one-electron properties, we may choose to include, at very little additional cost, configurations with double occupancy of the occupied and/or virtual orbitals.

The combination of an approximate form for the hessian with a second-order perturbation expression for the energy results in an overall orbital optimisation strategy that scales extremely favourably both with the number of 'active' electrons and with the number of basis functions. In this way, the current upper limit to the applicability of the method is fixed by the determination of the SC occupied orbitals, ϕ_i^0; the number of active electrons that can be treated is therefore currently about 12–14. Obviously, the use of an alternative formalism for the reference configuration, e.g. SCF or GVB–SOPP, could allow the basic method to be extended to larger systems.

Calculations for LiH and Li_2 showed that four-configuration SCVB* expansions represent a significant energy improvement over the standard single-configuration SC wavefunction [30]. For the $^1\sum^+$ electronic ground state of LiH, configurations based on σ^2 and π_x^2 (or π_y^2) double excitations make a contribution, providing radial and angular correlation, respectively. The σ and π_x virtual orbitals were optimised separately, and the π_y components were obtained by symmetry. The final variational wavefunction, denoted SCVB*–4, contained just the SC configuration and double excitations for the valence el-

ectrons into σ, π_x and π_y virtuals. While the SC wavefunction overestimates the position of the minimum, R_e, by 2.4%, the SCVB*–4 wavefunction underestimates it by only 0.95%. A much greater difference is seen, however, in the values of the dissociation energy, D_e. The SC wavefunction recovers 77% of the experimental value while SCVB*–4 obtains 95%. Full-valence complete active space (FVCAS) calculations in the same basis set give essentially the same value of R_e and a slightly superior value of D_e (by 0.03 eV), recovering 96% of the experimental value. A 1269–configuration 2-electron full CI (FCI) in the same basis gives a very slight improvement in R_e and recovers 98.7% of D_e, i.e. SCVB*–4 is less than 4% (0.1 eV) short of the best possible valence-only result for this basis. Furthermore, comparisons with conventional SCVB calculations suggest that the SCVB*–4 optimised virtuals span a significant proportion of the energetically-useful available space.

Analogous calculations were performed for Li_2, for which the SC wavefunction overestimates the value of R_e by 9.8% and recovers only 42% of the experimental value of D_e. The SCVB*–4 wavefunction, however, represents a significant further improvement, overestimating R_e by only 0.7% and recovering 91.5% of D_e. FVCAS calculations in the same basis set recover 94% of D_e (a 0.03 eV improvement), while 2-electron FCI calculations produce a further energy lowering of less than 0.004 eV.

The various approximations implicit in the SCVB* approach make the scaling of computational effort with numbers of electrons and of basis functions somewhat less severe than is the case for the multiconfiguration spin-coupled method of Penotti [20], making it possible to perform calculations on larger systems, and SCVB* may compare favourable with CASVB, which is described in a later section.

Our interest in LiH and LiH^+ arises mainly from the role that such molecules are expected to play in the chemistry of the early universe. It is important to be able to analyse in detail the efficiency of various processes that could lead to the formation of rotationally and/or vibrationally hot LiH molecules by collision with the most abundant partners, such as He, H, and H^+. We have applied the SCVB* approach to the computation of accurate potential energy surfaces involving these weakly interacting systems. We have already commented on the possibility of performing BSSE-free ab initio variational calculations in the context of SC theory. The agreement for the LiH … He system between the values obtained with the SCVB* method and the previous standard SCVB results [28] was very good everywhere. The SCVB* results turned out to be of better quality and comparable to standard SCVB calculations including up to 15 virtual orbitals and corresponding to a set of $\sim$ 1000 doubly excited configurations. Even the small secondary minimum of about 0.01 mHartree at long distance was well reproduced.

Our ability to describe accurately the interaction of LiH with He offered the opportunity to consider the formation of LiH in highly rotationally and/or vibrationally excited states during collisions with He in the primordial clouds.

For this purpose, we employed the SCVB* method to study the general features of the full potential surface. The target molecule LiH was initially treated as a rigid rotor at its equilibrium geometry and the behaviour of the interaction was examined over a wide range of orientations and relative distances for the impinging He projectile. The general features of the anisotropy of the potential were analysed in relation to earlier MBPT calculations [32] on the same system and were found to be rather different, especially in the short-range repulsive region for both directions of He approach along the diatomic bond. The full vibrational potential energy surface was also presented by performing the calculations for five different values of the molecular bond length [33]. The ab initio surface was then employed in the calculation of rotationally inelastic state-to-state cross sections. The comparison with the available experiments and with earlier calculations allowed us to relate in detail the features of the computed anisotropic, short-range interaction potential with the outcomes of the collisional excitation processes. The corresponding excitation rates were also evaluated between the lower-lying rotational levels and for a range of temperatures of astrophysical interest. The detailed comparisons with experiment indicate that the SCVB* potential energy surface provides better agreement with the measurements than does the earlier MBPT surface of Silver [32]. The new potential energy surface is less anisotropic in the repulsive region and also exhibits a weaker long-range coupling with rotations, but it produces inelastic cross sections which are in better accord with the experiments, justifying the expectation that the SCVB* surface provides a rather realistic description of the intermolecular interactions.

The vibrational heating efficiency of LiH molecules in collisions with He atoms was the subject of further study [34]. The excitation and relaxation rates over a broad range of temperatures were reported, together with the average energy transfer indices. It was found that in spite of the weak nature of the van der Waals interaction, the strong anisotropy of the surface leads to rovibrational excitation rates which are larger, for example, than those exhibited by the He–CO [35] or He–N_2 [36] systems.

We have also analysed the reaction of Li with H_2, with the aim of performing classical and quantum dynamical studies on the potential energy surface. To maintain absolute consistency in the calculations as the spatial arrangement of the three atoms changed symmetry, we have evaluated separately four discrete sets of virtual orbitals: each was expanded only in a subset of the AO basis relating to certain classes of functions. The four sets were created by dividing the full AO basis into one group containing just *s* functions and three groups each containing just one of the components of the *p* functions and the appropriate components of the *d* functions. Each of the four sets of virtual orbitals was optimised separately. From these orbitals, it was possible to create an SCVB* wavefunction which was equally flexible over the entire potential surface and did not change in character when the spatial arrangement of atoms changed symmetry. Without this technique, problems could have been experienced, par-

ticularly when the system passes through the collinear geometries. The final wavefunction consisted in total of just 73 configurations (293 structures); this extremely compact set is composed of all possible doubly excited configurations which involve only vertical excitations of SC-occupied orbitals into one of their four associated virtual orbitals, supplemented with all the singly ionic configurations in which a single virtual orbital is doubly occupied.

Rather extensive calculations were performed for the most important features of the lowest reactive surface of the process $H + LiH \rightarrow Li + H_2$. The five occupied orbitals from the single–configuration SC calculation included a clearly defined lithium $1s^2$ pair, which remains overwhelmingly singlet coupled at all points on the surface. At some positions on the potential energy surface, the three remaining orbitals are heavily localised on a particular atomic centre (for instance, when there is one isolated hydrogen atom) whereas at others they are delocalised over more than one centre (for instance, when all three atoms are in close proximity). It is clearly a very important feature of the SC method that there is no external predetermination of the degree of localisation of the orbitals. The spin coupling of the electrons in the three occupied valence orbitals also varies with geometry, as the nature of the orbitals changes. A perfect pairing scheme dominates at some points, whilst at others a more complicated situation occurs. The SCVB*–73 wavefunction recovers 93% of the binding energy, overestimating R_e by only 0.015 Å. The same kind of accuracy is obtained for the H_2 asymptotic potential curve: the SC and SCVB* wavefunctions recover 87% and 97% of the binding energy, respectively, and both reproduce the position of the minimum almost exactly. In the non-asymptotic regions of the potential, it is interesting to note that the SCVB* approach lowers to 2 mH (0.05 eV) the height of the barrier that blocks the path to a potential well in the linear Li–H_1–H_2 arrangement, as $R(H_1–H_2)$ is reduced while R(Li–H_1) is kept fixed at $3\,a_0$, a geometry at which the height of the SC barrier is about 10 mH (0.27 eV). The SCVB* ab initio points were interpolated using a two-dimensional spline function in order to be able to treat the two variables within the dynamical codes employed for an exploratory analysis of its reactive behaviour in the special two-dimensional arrangements, involving the Li–H–H and H–Li–H collinear structures. Both the highly exothermic depletion reaction $H + LiH(v) \rightarrow Li + H_2$ and the endothermic $Li + H_2(v) \rightarrow LiH + H$ formation reaction where analysed in terms of classical reactive trajectories and of a quantal treatment based on the time-dependent wave packet method. These preliminary collinear results based on the new SCVB* surface already appeared to be able to say a great deal about specific features of the reaction and the likely success of classical calculations for a full treatment [37].

The description of the interaction of LiH colliding with H^+ presents new problems, mostly because the surface describing $LiH + H^+$ corresponds to the asymptotic region of the excited state of LiH_2^+, whose ground state dissociates to $LiH^+ + H$. The separation between these asymptotic states is 5.5–6 eV. The states remain well separated also in the other asymptotic region corresponding

to $Li^+ + H_2$ and $Li + H_2{}^+$, respectively. We describe briefly a new approach for treating excited states belonging to the same symmetry as the ground state. We first identified the ground state SC orbital with the highest orbital energy, and projected it out of the original atomic basis set. Subsequent optimisations were then carried out in this projected basis set, so that all of the occupied and virtual orbitals for the excited state are necessarily orthogonal to the chosen ground state orbital. Finally, by constructing a wavefunction represented by the ground and excited state SC configurations and by single and double excitations to the optimal virtual orbitals of each state, and including also all of the singly ionic structures, a very compact set of 125 structures is generated, describing both the $LiH + H^+$ and the $LiH^+ + H$ systems. Solution of the usual nonorthogonal secular problem provides the description of both the ground and excited states. We denote the final variational wavefunctions SCVB**–125. Comparison with FCI calculations in the same basis set ($\sim 2.5 \times 10^6$ configurations) suggests that the performance of the SCVB** wavefunction is astonishingly good for both the ground and excited states [38]. The accuracy of the results not only shows that expanding SC-occupied and virtual orbitals for the excited state in the orthogonal complement of just a single ground state SC-occupied orbital is viable for this system, but also that this is a promising method for future work. An alternative approach for the direct optimisation of orbitals for excited states is provided by CASVB.

5
CASVB

The CASVB strategy [39–48] exploits the invariance of wavefunctions of full-CI form to arbitrary nonsingular linear transformations of the active orbitals. The method may be used to carry out a (typically) nonunitary orbital transformation such that the representation of the total CASSCF wavefunction is dominated by a compact, easy-to-interpret component of modern VB form. In order to be able to realise this goal, we require efficient procedures for carrying out exactly the transformation of the structure space induced by a given general orbital transformation, as well as schemes for determining orbital transformations that lead to 'useful' representations of the CASSCF wavefunction. The CASVB algorithms may be adapted so as to enable fully-variational optimisation of fairly general types of (multiconfiguration) modern VB wavefunctions, so that this strategy also represents a significant advance in the development of VB methodology. In all modes of using the CASVB approach, important features for the quality of the final description include the unbiased optimisation of both the VB orbitals and the mode of spin coupling, and also flexibility in the choice of the form of wavefunction. In order to ensure reliable convergence, we prefer an exact second-order optimisation scheme involving first and second derivatives with respect to all of the variational parameters.

For a many-electron space that is closed under linear transformations **O** of the defining orbitals, we decompose such a transformation into simpler 'orbital updates':

$$\mathbf{O} = \mathbf{O}_{11}(\lambda_1)\mathbf{O}_{12}(\lambda_2)\mathbf{O}_{13}(\lambda_3)\ldots\mathbf{O}_{mm}(\lambda_{m^2}) \tag{8}$$

with

$$\mathbf{O}_{\mu\nu}(\lambda) : \phi_\nu \rightarrow \phi_\nu + \lambda\phi_\mu \,. \tag{9}$$

The effect of the corresponding linear transformation of the many-electron structure space, $\mathbf{T}(\mathbf{O})$, on a given vector of CI coefficients, may then be realised by m^2 applications of the replacement operators:

$$\mathbf{T}(\mathbf{O}_{\mu\nu}(\lambda)) = (\mathbf{1} + \lambda\mathbf{E}^{(\alpha)}_{\mu\nu})(\mathbf{1} + \lambda\mathbf{E}^{(\beta)}_{\mu\nu}) = \mathbf{1} + \lambda\mathbf{E}^{(1)}_{\mu\nu} + \lambda^2\mathbf{E}^{(2)}_{\mu\nu} \,, \tag{10}$$

in terms of the spin-orbital single-replacement operators $\mathbf{E}^{(\alpha)}_{\mu\nu}$ and $\mathbf{E}^{(\beta)}_{\mu\nu}$, or, equivalently, in terms of the spin-averaged single- and double-replacement operators $\mathbf{E}^{(1)}_{\mu\nu}$ and $\mathbf{E}^{(2)}_{\mu\nu}$.

We may partition a CASSCF wavefunction, Ψ_{CAS}, according to

$$\Psi_{\mathrm{CAS}} = c_{\mathrm{VB}}\Psi_{\mathrm{VB}} + c^{\perp}_{\mathrm{RES}}\Psi^{\perp}_{\mathrm{RES}} \,, \tag{11}$$

in which Ψ_{VB} is of VB type, and the 'residual' $\Psi^{\perp}_{\mathrm{RES}}$ is the orthogonal complement. To this end, we adopt the so-called 'overlap criterion', in which we maximise

$$S_{\mathrm{VB}} = \frac{\langle\Psi_{\mathrm{CAS}}|\Psi_{\mathrm{VB}}\rangle}{\langle\Psi_{\mathrm{VB}}|\Psi_{\mathrm{VB}}\rangle^{\frac{1}{2}}} \,. \tag{12}$$

With a suitably chosen Ψ_{VB}, more than 99% of a ground-state CASSCF wavefunction may typically be brought to VB-like form. Provided CASSCF solutions are available, it is straightforward to extend this strategy to obtain modern valence bond wavefunctions for states that are second or higher within a particular symmetry [46, 47].

The variational description of a ground state is obtained of course by minimising the energy expectation value

$$E_{\mathrm{VB}} = \frac{\langle\Psi_{\mathrm{VB}}|\hat{H}|\Psi_{\mathrm{VB}}\rangle}{\langle\Psi_{\mathrm{VB}}|\Psi_{\mathrm{VB}}\rangle} \tag{13}$$

with respect to all the free parameters. In the CASVB strategy, we employ an iterative two-step procedure:

(1) optimise the active-space parameters, i.e. the orbital transformation and the vector of VB structure coefficients,
(2) with the set of CASSCF CI coefficients defined by step (1), optimise the orbitals (in the MO basis) with respect to core-active, core-virtual, and active-virtual orbital rotations (using a CASSCF procedure).

Because the final wavefunction is essentially a constrained form of the CASSCF function, much of the impressive technology of CASSCF procedures immediately becomes available, including the efficient evaluation of analytic energy

gradients (e.g. for geometry optimization). For excited states, optimisation of the i^{th} root is achieved by seeking a saddle point of order $i-1$ [46, 47]. Multiconfiguration wavefunctions [43] for ground and excited states, obtained with the CASVB method, provide useful benchmarks for the (approximately) optimised orbitals used in the SCVB* and SCVB** approaches.

As an alternative to fully-variational calculations, carrying out step (1) on its own simply gives an 'energy-based' representation of Ψ_{CAS}. We have generally found very close agreement for ground and excited states between the descriptions generated by the energy-based and (cheaper) overlap-based criteria.

The excited states investigated so far with CASVB can be classified into three main categories:

(1) recoupling of electron spins;
(2) 'valence' → 'valence' orbital excitations, so that the excited state has doubly occupied orbitals (so-called 'ionic' configurations);
(3) 'valence' → 'virtual' orbital excitations, including those leading to Rydberg states.

The optimisation of excited states of types (1) or (3) is usually straightforward, but excited states of type (2) may require particular care in the choice of reference function [46, 47]. In addition to excited states, recent applications have been concerned, inter alia, with the mechanisms of gas phase pericyclic reactions [48] and with the modes of spin coupling for rival spin multiplicities in organic π-electron systems [47, 49].

Our CASVB code includes sophisticated techniques for exploiting point group symmetry [41, 47] and for analysing the resulting wavefunctions [44, 47]. The practical limit on the number of active electrons ($\sim$14) for calculations with the full spin space is determined by the current speed of CASSCF procedures. CASVB is incorporated as a standard feature in MOLPRO [50] and we are actively seeking ways to make all of the methodology more widely available via other packages.

6
Final Remarks

As has been described in numerous review articles, approaches based on spin-coupled theory can provide highly visual representations of correlated electronic structure and have made important contributions to the development of ab initio modern valence bond theory. We have chosen in the present account to concentrate on certain recent developments.

We have indicated the ability of spin-coupled valence bond (SCVB) calculations to provide compact, though accurate, nonorthogonal configuration interaction expansions for ground and excited states. These wavefunctions can be made even more compact using the SCVB* method, which starts from direct

physical ideas and a perturbation theory approach to generate virtual orbitals that are particularly suitable for the inclusion of electron correlation effects for the problem at hand. The overall strategy scales favourably with the numbers of active electrons and of basis functions. An important aspect is the ability to describe correctly intermolecular interactions in a framework that avoids basis set superposition error. Applications have been described here for reactions that are important for understanding the lithium chemistry of the early universe. We have also illustrated the progress made so far for optimising occupied and virtual orbitals for excited states within an SCVB*-like formalism.

We have outlined the basic features of the CASVB approach, which may be used to generate modern VB representations of complete active space self-consistent field wavefunctions for ground and excited states. The same code may be used for the fully-variational simultaneous optimisation of the inactive and active orbitals and structure coefficients for fairly general types of ground and excited state modern VB wavefunctions, including single- and multiconfiguration spin-coupled descriptions. The methodology also provides analytic gradients as required, for example, in geometry optimisations.

It should be clear that the use of nonorthogonal orbitals (and structures) does not lead to formalisms of hopeless complexity, as might have been supposed, but instead opens up new avenues and novel strategies for incorporating physical ideas into algorithms that reduce the complexity of accurate descriptions of molecular electronic structure.

7 References

1. Coulson CA, Fischer I (1949) Phil Mag 40:386
2. Heitler W, London F (1927) Z Phys 44:455
3. Gerratt J (1971) Adv Atom Mol Phys 7:141
4. Gerratt J, Raimondi M (1980) Proc Roy Soc (London) A 371:525
5. (a) Cooper DL, Gerratt J, Raimondi M (1987) Adv in Chem Phys 69:319.
   (b) Cooper DL, Gerratt J, Raimondi M (1988) Int Rev Phys Chem 7:59.
   (c) Gerratt J, Cooper DL, Raimondi M (1990) The spin-coupled valence bond theory of molecular electronic structure. In: D.J. Klein DJ, TrinajsticĐ N (eds) Valence bond theory and chemical structure, Elsevier, Amsterdan, p 287
6. Cooper DL, Gerratt J, Raimondi M (1991) Chem Rev 91:929
7. Gerratt J, Cooper DL, Karadakov PB, Raimondi M (1997) Chem Soc Rev 26:87
8. (a) Cooper DL, Gerratt J, Raimondi M (1999) The spin-coupled description of aromatic, antiaromatic and nonaromatic systems. In: Maksic ZB, Orville-Thomas WJ (eds) Pauling's Legacy — Modern Modelling of the Chemical Bond, Elsevier, Amsterdam.
   (b) Cooper DL, Gerratt J, Raimondi M (1999) Hypercoordinate bonding to main group elements: the spin-coupled point of view. In: Maksic ZB, Orville-Thomas WJ (eds) Pauling's Legacy — Modern Modelling of the Chemical Bond, Elsevier, Amsterdam
9. (a) Kaldor U, Harris FE (1969) Phys Rev 183:1.
   (b) Kaldor U (1970) Phys Rev A 1:1586.
   (c) Kaldor U (1970) Phys Rev A 2:1267

10. (a) Goddard WA (1967) Phys Rev 157:81.
(b) Palke WE, Goddard WA (1969) J Chem Phys 50:4524.
(c) Ladner RC, Goddard WA (1969) J Chem Phys 51:1073.
(d) Melius CF, Goddard WA (1972) J Chem Phys 56:3348.
(e) Blint RJ, Goddard WA (1972) J Chem Phys 57:5296
11. (a) Hay PJ, Goddard WA (1972) J Chem Phys 57:738.
(b) Voter, Goddard WA (1981) J Chem Phys 75:3638.
(c) Bobrowicz FW, Goddard WA (1977). In: Schaefer HF (ed) Methods of Electronic Structure Theory, Plenum, New York
12. Chipman DM, Kirtman B, Palke WE (1976) J Chem Phys 65:2556
13. Schultz P, Messmer RP (1987) Phys Rev Lett 58:2416
14. Verbeek J, Langenberg JH, Byrman CP, van Lenthe JH (1993) TURTLE - ab initio VB/VBSCF/VBCI program, Theoretical Chemistry Group, Debye Institute, University of Utrecht
15. Hiberty PC, Flament JP, Noizet E (1992) Chem Phys Lett 189:259
16. (a) Hollauer, Nascimento (1991) Chem Phys Lett 184:470.
(b) (1993) J Chem Phys 99:1207
17. Gallup GA, Vance RL, Collins JR, Norbeck JM (1982) Adv in Quant Chem 16:229
18. McWeeny R (1988) Int J Quant Chem 34:25
19. Li X, Zhang Q (1989) Int J Quant Chem 36:599
20. Penotti F (1996) Int J Quant Chem 59:349
21. Manby FR and Doggett G (1998) J Phys B 31:949
22. Pauncz R (1979) Spin Eigenfunctions. Plenum, New York
23. Cooper DL, Gerratt J, Raimondi M, Sironi M, Thorsteinsson T (1993) Theor Chim Acta 85:261
24. Friis-Jensen B, Cooper DL, Rettrup S (1998) Theor Chem Acc 99:64
25. (a) Zygelman B, Cooper DL, Ford MJ, Dalgarno A, Gerratt J, Raimondi M (1992) Phys Rev A 46:3846.
(b) Jamieson MJ, Cooper DL, Zygelman B (1994) J Phys B 27:L73.
(c) Stancil PC, Zygelman B, Clarke NJ, Cooper DL (1997) J Phys B 30:1013.
(d) Zygelman B, Stancil PC, Clarke NJ, Cooper DL (1997) Phys Rev A 56:457
26. (a) Stancil PC, Zygelman B, Clarke NJ, Cooper DL (1997) Phys Rev A 55:1064.
(b) Clarke NJ, Stancil PC, Zygelman B, Cooper DL (1998) J Phys B 31:533.
(c) Clarke NJ, Cooper DL (1998) J Chem Soc Faraday Trans 94:3295.
(d) Stancil PC, Clarke NJ, Zygelman B, Cooper DL (1999) J Phys B, to be published
27. da Silva EC, Gerratt J, Cooper DL, Raimondi M (1994) J Chem Phys 101:3866
28. Matías MA, Raimondi M, Tornaghi E, Cooper DL, Gerratt J (1994) Mol Phys 83:89
29. Raimondi M, Sironi M, Gerratt J, Cooper DL (1996) Int J Quant Chem 60:225
30. Clarke NJ, Raimondi M, Sironi M, Gerratt J, Cooper DL (1998) Theor Chem Acc 99:8
31. Löwdin P-O (1951) J Chem Phys 19:1396
32. Silver DM (1980) J Chem Phys 72:6445
33. Gianturco FA, Kumar S, Pathak SK, Raimondi M, Sironi M, Gerratt J, Cooper DL (1997) Chem Phys 215:227
34. (a) Bodo E, Kumar S, Gianturco FA, Famulari A, Raimondi M, Sironi M (1998) J Phys Chem A 102:9390.
(b) Bodo E, Buonomo E, Gianturco FA, Kumar S, Famulari A, Raimondi M, Sironi M (1998) Chem Phys 237:315
35. Reid JP, Simpson CJSM, Quiney HM (1997) J Chem Phys 107:9929
36. Reid JP, Thakkar AJ, Barnes PW, Archibong EF, Quiney HM, Simpson CJSM (1997) J Chem Phys 107:2329

37. Clarke NJ, Sironi M, Raimondi M, Kumar S, Gianturco FA, Buonomo E, Cooper DL (1998) Chem Phys 233:9
38. Raimondi M, Martinazzo R, Famulari A, Sironi M (1999) J Chem Phys
39. Thorsteinsson T, Cooper DL, Gerratt J, Karadakov PB, Raimondi M (1996) Theor Chim Acta 93:343
40. Thorsteinsson T, Cooper DL (1996) Theor Chim Acta 94:233
41. Thorsteinsson T, Cooper DL, Gerratt J, Raimondi M (1997) Theor Chim Acta 95:131
42. Thorsteinsson T, Cooper DL, Gerratt J, Raimondi M (1997) A new approach to valence bond calculations: CASVB. In: McWeeny R, Maruani J, Smeyers YG, Wilson S(eds) Quantum Systems in Chemistry and Physics: Trends in Methods and Applications, Kluwer, Dordrecht
43. Cooper DL, Thorsteinsson T, Gerratt J (1997) Int J Quant Chem 65:439
44. Thorsteinsson T, Cooper DL (1998) J Math Chem 23:105
45. Cooper DL, Thorsteinsson T, Gerratt J (1998) Adv Quant Chem 32:51
46. Thorsteinsson T, Cooper DL (1998) Int J Quant Chem 70:637
47. Thorsteinsson T, Cooper DL (1999) Progress in Theoretical Chemistry and Physics
48. (a) Karadakov PB, Cooper DL, Gerratt J (1998) J Am Chem Soc 120:3975.
(b) Karadakov PB, Cooper DL, Gerratt J (1998) Theor Chem Acc.
(c) Karadakov PB, Cooper DL, Thorsteinsson T, Gerratt J (1999) Progress in Theoretical Chemistry and Physics
49. Friis-Jensen B, Cooper DL, Rettrup S, Karadakov PB (1998) J Chem Soc Faraday Trans 94:3301
50. MOLPRO is a package of *ab initio* programs written by Werner H-J, Knowles PJ, with contributions from Almlöf J, Amos RD, Berning A, Cooper DL, Deegan MJO, Eckert F, Elbert ST, Hampel C, Lindh R, Meyer W, Nicklaß A, Peterson K, Pitzer R, Stone AJ, Taylor PR, Mura ME, Pulay P, Schütz M, Stoll H, Thorsteinsson T

Modern Correlation Theories for Extended, Periodic Systems

Jun-Qiang Sun · Rodney J. Bartlett

Quantum Theory Project, Departments of Chemistry and Physics,
University of Florida, Gainesville, FL 32611, USA.
E-mail: sun@qtp.ufl.edu, bartlett@qtp.ufl.edu

Recent progress on correlated methods for extended, periodic systems is briefly summarized. Special difficulties compared to finite systems such as convergence with infinite summations over lattices and integrations over reciprocal vectors are addressed. We also consider the comparative advantages of Bloch orbitals and Wannier orbitals. The applications of MBPT(2) to polymers for determining their equilibrium structures, vibrational frequences, XPS and UPS spectra, and band gaps are examined. It is shown that MBPT(2) corrections have significant contributions to all of the above quantities and make significant improvements to the HF results, especially for the XPS and UPS spectra.

Keywords: Electron correlation, Polymers, XPS and UPS, Vibrational frequencies, Band gaps and Excitations

Topics in Current Chemistry, Vol. 203

1 Introduction

It is well known that electron correlation plays a key role in understanding the most interesting phenomena in molecules. It has been the focal point of atomic and molecular theory for many years [1] and various correlated methods have been developed [2]. Among them are many-body perturbation theory [3] (MBPT) and its infinite-order generalization, coupled cluster (CC) theory [4, 5], which provides a systematic way to obtain the essential effects of correlation. Propagator [6–9] or Green's function methods (GFM) [10–14] provide another correlated tool to calculate the electron correlation corrections to ionization potentials (IPs), electron affinites (EAs), and electronic excitations.

As in finite systems, electron correlation is critical to a description of many of the most interesting problems in extended systems, polymers, surfaces, and crystals [15, 16]. These include band structures and band gaps, phonon spectra, mechanic properties, cohesive energies, optical and magnetic properties, superconductivity, etc. Although it remains one of the foremost problems in the field of electronic structure, some significant progress has recently been made in the theory and their applications.

Compared to finite systems, extended systems are much more difficult to treat. One approach is to treat increasely large clusters, perhaps assisted by some kind of localization approach [17], and extrapolate to infinity [18–19]. However, such extrapolations are difficult, slowly convergent, and not very satisfying formally. We believe a superior approach is to properly treat the periodic symmetry at the correlated level. Since this introduces lattice summations which go to infinity in both Hartree–Fock (HF) and correlated methods, it is essential to be sure the calculated results are converged with the summations [12, 20–25]. The band energies, in particular, are more sensitive to the cutoff of the lattice summations than the total energy per unit cell [26, 27]. Multipole expansion techniques have been applied in HF calculations to obtain converged results and also to accelerate the calculations [22]. After thirty years of development, converged and reliable HF packages for polymers and crystals have become available [28–36], although further improvement is certainly possible.

Since the size of a periodic polymer or crystal is infinite, correct scaling with size (i.e. size extensivity [37, 38]) is an essential prerequisite for a correlated theory to be applied to extended systems. It is well known that CC/MBPT has this property [37, 38], and so they are appropriate tools for extended systems while the traditional configuration interaction (CI) methods are not. This raises the prospect that a converging series of approximations somewhat analogous to the molecular paradigm, $SCF < MBPT(2) \sim MBPT(3) < CCD < CCSD \sim MBPT(4) < CCSD(T) < CCSDT < Full\ CI$, can be envisioned, though, of course, full CI is never possible for an infinite system. The availability of this paradigm would complement improving DFT methods and GW approximations in resolving the puzzles of electron correlation in polymers, surfaces,

and crystals [39–42]. In brief, we would like to be able to make the same kind of applications for infinite systems as we do today for finite ones, and these include higher-order correlated CC methods, analytical gradients and hessians, plus methods for excited and ionized states.

In particular, orbital energies in extended systems are important quantities. They form the so-called band energies or band structure. The HF band energies contribute the first approximations for IPs and EAs [43]. Beyond the HF approximation, one defines correlated quasi-particle band energies as IPs and EAs for occuiped and unoccpied orbitals, respectively [44], via MBPT [44–48], CC [49–51], and propagator [6–9] or Green's function methods (GFM) [10–14]. Explicit expressions for any given-order correction have been recently derived [52].

The simplest correlated method is second-order MBPT [MBPT(2)], and the first element in the above paradigm. MBPT(2) has been shown in molecules to typically account for > 90% of the correlation energy and to offer a factor of two improvement in other properties [5]. MBPT(2) was first applied to extended systems by Kunz and coworkers [44], They derived the formulas and discussed the effect of the MBPT(2) correlation correction to band energies for insulators. The method was implemented by Suhai for one-dimensional, periodic systems and applied to determine the structures and band gaps in polyacetylene and some other systems [53–56]. He showed that MBPT(2) corrections have a strong effect on both equilibrium structures and band gaps [53, 55]. From the inverse Dyson equation with the irreducible self-energy part in the diagonal approximation (e.g. GFM), Liegener [57] calculated the second- and third-order corrections to the band gap for alternating trans-polyacetylene. However, there was a large inconsistency between the MBPT(2) and GFM corrections for band energies.

Recently, we implemented both MBPT(2) and second-order GFM for polymers, and studied the convergence of the corrections with lattice summation cutoff (N) and number of **k**-points (K) taken in the first Brillouin zone in the integrations over the reciprocal lattice [26]. We found the corrections to band energies to converge much slower with N and to demand much larger K values than that of the total energy per unit cell. We calculated the total energy and band gap for alternating trans-polyacetylene at several geometries and basis sets and showed that the previous results had been obtained before convergence with the lattice summations had been properly reached [26]. We also showed both theoretically and numerically that the MBPT(2) correction to the total energy per unit cell converges with lattice summation cutoff N as $1/N^3$ while the MBPT(2) correction to the band energy converges one power slower [27]. It has also been proven recently that any-order corrections of MBPT, CC, and MBGF will converge with lattice summations, provided that the evaluation is correctly treated [24, 25].

X-ray (XPS) [58–61] and ultraviolet (UPS) [62–65] photoelectron spectroscopy provide rich information about the valence bands and useful tools to

check the accuracies of developing theories. We applied MBPT(2) to polyethylene and showed that electron correlation effects for the valence bands in polyethylene vary from 1.5 eV to 5.4 eV [66]. The correlated quasi-particle band energies given by MBPT(2) with a polarized 6-31G** basis accurately explain the measured photoelectron spectra (XPS, UPS) of polyethylene and resolve long-standing disagreements among the experiments.

MBPT(2) has also been applied to calculate vibrational frequencies of polymers. With the translational symmetry, one can only calculate the vibrational modes with the reciprocal vector $\mathbf{k} = 0$. These modes are of particular importance since they give rise to infrared and Raman spectra [67]. We applied MBPT(2) to polymethineimine and calculated its equilibrium structure, band gap, and vibrational frequencies with basis sets STO-3G, 6-31G and 6-31G** [68]. Both basis set and electron correlation have a strong influence on its vibrational frequencies as well as its optimized geometry and band gap. With respect to in-phase (k=0) nuclear displacements, Hirata and Iwata very recently calculated the MBPT(2) vibrational frequencies of polyacetylene for basis sets STO-3G and 6-31G with analytical gradients [69]. They showed that MBPT(2) greatly improves the HF vibrational frequencies for polyacetylene.

Recently, CC methods have been pursued by Förner, Ladik and coworkers [70]. They derived the equations with localized orbitals and applied the L-CCD and CCD to polyacetylene. Since it is much more difficult to implement and requires more computational effort than MBPT, the method was executed with further approximations which prohibit an accurate assessment of the method. MBPT(2) has also been applied to cases where conduction and valence bands are quasidegenerate [71, 72] and to study high T_c superconductivity [72].

In this short review, we will concentrate on MBPT and GFM methods and their application to polymers. In Sect. 2, we will give a brief summary for HF theory and establish notations. Then in Sects. 3 and 4, we will give the basic formulas for MBPT(2) corrections to the total energy per unit cell and band energies. The second-order GFM equation will also be given. In Sect. 5, we will discuss the convergence of MBPT(2) with the cutoffs for lattice summations and integrations over the reciprocal space. In Sect. 6, we will address higher-order MBPT corrections and their convergence with lattice summations. In Sect. 7, we will compare Bloch and Wannier orbitals. In Sects. 8, 9, and 10, we will address XPS and UPS, geometry and vibrational frequencies, band gaps and excitations, and the applications of MBPT(2) to these properties. Conclusions will follow in Sect. 11.

2
Hartree-Fock Method

An infinite system which is periodic in ρ directions ($\rho = 1, 2, 3$ for polymers, surfaces, and crystals, respectively) is described by a reference unit cell and ρ basis vectors, $\mathbf{a}_1$, ..., $\mathbf{a}_\rho$. These vectors point to the directions of the periodicity

and their sizes give the repeat lengths in the directions. Then the nonrelativistic Hamiltonian of the system is

$$H = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}{\sum_{l,l',A,B}}' \frac{Z_A Z_B}{|\mathbf{R}_{l'} + \mathbf{R}_B - \mathbf{R}_l - \mathbf{R}_A|} - \sum_{i,l,A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_l - \mathbf{R}_A|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (1)$$

where $\mathbf{R}_l = l_1\mathbf{a}_1 + ... + l_\rho \mathbf{a}_\rho$ and $\mathbf{R}_{l'} = l'_1\mathbf{a}_1 + ... + l'_\rho \mathbf{a}_\rho$ are the lattice vectors, Z_A and Z_B are the charges of the nuclei A and B at position R_A and R_B in the reference unit cell, respectively, and the prime excludes $\mathbf{R}_l = \mathbf{R}_{l'}$ and $A = B$.

The periodicity of the system means that it has translational symmetry, whose irreducible representations are all one-dimensional. They can be classified by the so-called reciprocal lattice vector $\mathbf{k} = k_1\mathbf{b}_1 + ... + k_\rho \mathbf{b}_\rho$ where $\mathbf{b}_j$ $(j = 1, ..., \rho)$ are determined by $\mathbf{a}_i \cdot \mathbf{b}_j = \delta_{ij}$, $i = 1, ..., \rho$. When k_i $(i = 1, ..., \rho)$ are limited to the region $[-\pi, \pi]$, the $\mathbf{k}$ space is called the first Brillouin zone.

Let $\{\chi_\alpha(\mathbf{r})\}$ be atomic orbitals in the reference unit cell. Then according to the periodicty of the system, the atomic orbitals in the ith unit cell are $\{\chi^l_\alpha(\mathbf{r}) = \chi_\alpha(\mathbf{r} - \mathbf{R}_l)\}$. With the atomic orbitals, one can construct symmetrized orbitals by

$$\phi_{\mathbf{k}\alpha}(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{N}}} \sum_l e^{i\mathbf{k}\cdot\mathbf{R}_l} \chi^l_\alpha(\mathbf{r}), \quad (2)$$

where $\mathcal{N} \to \infty$ and $\mathbf{k}$ is limited to the first Brillouin zone. The overlap matrix of the symmetrized orbitals is block diagonal, e.g.

$$\langle \phi_{\mathbf{k}\alpha}(\mathbf{r}) | \phi_{\mathbf{k}'\beta}(\mathbf{r}) \rangle = \delta_{\mathbf{k},\mathbf{k}'} S^{\mathbf{k}}_{\alpha\beta}, \quad (3)$$

where

$$S^{\mathbf{k}}_{\alpha\beta} = \sum_l e^{i\mathbf{k}\cdot\mathbf{R}_l} \langle \chi_\alpha(\mathbf{r}) | \chi^l_\beta(\mathbf{r}) \rangle. \quad (4)$$

With symmetrized orbitals and closed shell assumption, the Fock matrix of the system can be expressed as

$$F_{\mathbf{k}\alpha,\mathbf{k}'\beta} = \delta_{\mathbf{k},\mathbf{k}'} F^{\mathbf{k}}_{\alpha\beta}, \quad (5)$$

which is also block diagonal and whose diagonal blocks are given by

$$F^{\mathbf{k}}_{\alpha\beta} = \sum_l e^{i\mathbf{k}\cdot\mathbf{R}_l} \left[T^l_{\alpha\beta} + V^l_{\alpha\beta} + 2J^l_{\alpha\beta} - K^l_{\alpha\beta} \right]. \quad (6)$$

$T^l_{\alpha\beta}$, $V^l_{\alpha\beta}$, $J^l_{\alpha\beta}$, and $K^l_{\alpha\beta}$ in Eq. (6) are the kinetic, electron-nucleus, Coulomb and exchange energy matrices, respectively. Then HF orbitals can also be classified according to the reciprocal lattice $\mathbf{k}$ and be written as

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_\alpha C^{n\mathbf{k}}_\alpha \phi_{\alpha\mathbf{k}}(\mathbf{r}), \quad (7)$$

where $C_\alpha^{n\mathbf{k}}$ are required to satisfy

$$\sum_{\alpha\beta} \left(C_\alpha^{n\mathbf{k}}\right)^* C_\beta^{n'\mathbf{k}} S_{\alpha\beta}^{\mathbf{k}} = \delta_{nn'} \tag{8}$$

by orthonormalization. The Hartree–Fock equation for the system is [26, 28–36]

$$\mathbf{F}^{\mathbf{k}}\mathbf{C}^{n\mathbf{k}} = \varepsilon_{n\mathbf{k}}^{HF}\mathbf{S}^{\mathbf{k}}\mathbf{C}^{n\mathbf{k}}. \tag{9}$$

The orbital energies $\varepsilon_{n\mathbf{k}}^{HF}$ form the well-known HF energy bands, where n is the band index. According to Koopmans' theorem [43], the band energies for the occupied bands i and unoccupied bands a, are equal to the corresponding ionization potentials and electron affinities, respectively. When $(a\mathbf{k}_a)$ and $(i\mathbf{k}_i)$ are the lowest unoccupied and the highest occupied orbitals, respectively, $\Delta\varepsilon^{HF}(i\mathbf{k}_i \rightarrow a\mathbf{k}_a)$ reaches its minimum. This minimum value is called the HF band gap. Let us denote it as E_g^{HF}.

From the HF block orbitals $\phi_{n\mathbf{k}}(\mathbf{r})$, the HF ground state of the system can be constructed as

$$\Phi_0^{HF} = \mathcal{A}\prod_{n\mathbf{k}\omega}(n\mathbf{k}\omega), \tag{10}$$

where $\mathcal{A}$ is the antisymmetrizing operator and ω is the spin quantum number of the spin orbitals. The total HF energy of the system is infinite while the total energy per unit cell, e.g. the total energy divided by the number of unit cells, has a finite value

$$\begin{aligned} E_{uc}^{HF} = &\sum_{l\alpha\beta}\left[2T_{\alpha\beta}^l + 2V_{\alpha\beta}^l + 2J_{\alpha\beta}^l - K_{\alpha\beta}^l\right]D_{\alpha\beta}^{0l} \\ &+\frac{1}{2}{\sum_{l,A,B}}' \frac{Z_A Z_B}{|\mathbf{R}_l - \mathbf{R}_B + \mathbf{R}_A|}, \end{aligned} \tag{11}$$

where $D_{\alpha\beta}^{0l}$ is the electron density matrix [26, 36] and the prime excludes $\mathbf{R}_j - \mathbf{R}_A + \mathbf{R}_B = 0$.

In Eqs. (4), (6) and (11), there are infinite summations in the overlap matrix $\mathbf{S}^{\mathbf{k}}$, the Fock matrix $\mathbf{F}^{\mathbf{k}}$, and the total energy per unit cell E_{uc}^{HF}. There are also infinite summations in T_{pq}^l, V_{pq}^l, J_{pq}^l, and K_{pq}^l. The infinite lattice summations in $\mathbf{S}^{\mathbf{k}}$, $\mathbf{T}^l$ and $\mathbf{K}^l$ converge themselves while $\mathbf{V}^l$, $\mathbf{J}^l$ and the internuclear interactions have to be summed together to get converged results [20–22]. In real calculations, cutoffs for the lattice summations have been imposed and multipole expansion techniques have been applied to hasten the convergence [22].

3 Correlated Method for Total Energies

From perturbation theory, we know that the exact total energy of the system can be expressed as

$$E = E^{HF} + E^{(2)} + E^{(3)} + \ldots . \tag{12}$$

Since the total energy is infinite in an extended system, we are more interested in the total energy per unit cell, which is defined as

$$\begin{aligned} E_{uc} &= E/\mathcal{N} \\ &= E_{uc}^{HF} + E_{uc}^{(2)} + E_{uc}^{(3)} + \ldots , \end{aligned} \tag{13}$$

for which the analytical expression of E_{uc}^{HF} has already been given in Eq. (11).

From Raleigh–Schrödinger perturbation theory [3], one can get analytical expressions for any given-order MBPT correction to the total energy per unit cell. Here we focus on the second-order MBPT correction which is given by [26]

$$\begin{aligned} E^{(2)} = \frac{1}{\mathcal{N}} \sum_{i\mathbf{k}_i j\mathbf{k}_j} \sum_{a\mathbf{k}_a b\mathbf{k}_b} \Big\{ & 2|\langle i\mathbf{k}_i j\mathbf{k}_j|r_{12}^{-1}|a\mathbf{k}_a b\mathbf{k}_b\rangle|^2 - Re[\langle i\mathbf{k}_i j\mathbf{k}_j|r_{12}^{-1}|a\mathbf{k}_a b\mathbf{k}_b\rangle \\ & \times \langle b\mathbf{k}_b a\mathbf{k}_a|r_{12}^{-1}|i\mathbf{k}_i j\mathbf{k}_j\rangle] \Big\} / \left[\varepsilon_{i\mathbf{k}_i}^{HF} + \varepsilon_{j\mathbf{k}_j}^{HF} - \varepsilon_{a\mathbf{k}_a}^{HF} - \varepsilon_{b\mathbf{k}_b}^{HF} \right] , \end{aligned} \tag{14}$$

where i, j and a, b denote occupied and unoccupied bands, respectively, and $Re[x]$ is a function which takes the real part of x. The two-electron integrals in Eq. (14) are infinitesimal numbers [15]. They approach zero when $\mathcal{N} \to \infty$. To separate the infinitesimal factor, we can write [26]

$$\langle (p\mathbf{k}_p)(q\mathbf{k}_q)|y_{12}^{-1}|(r\mathbf{k}_r)(s\mathbf{k}_s)\rangle = \delta_{\mathbf{k}_q,\mathbf{T}(\mathbf{k}_r+\mathbf{k}_s-\mathbf{k}_p)} Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)/\mathcal{N} , \tag{15}$$

where

$$\begin{aligned} Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s) = & \sum_{\mathbf{R}_1,\mathbf{R}_2,\mathbf{R}_3} exp\left[i\left(\mathbf{k}_r \cdot \mathbf{R}_1 - \left(\mathbf{k}_r - \mathbf{k}_p\right) \cdot \mathbf{R}_2 + \mathbf{k}_s \cdot \mathbf{R}_3\right)\right] \\ & \times \sum_{pqrs} \left(C_\alpha^{i\mathbf{k}_p}\right)^* \left(C_\gamma^{qT(\mathbf{k}_r+\mathbf{k}_s-\mathbf{k}_p)}\right)^* C_\beta^{r\mathbf{k}_r} C_\vartheta^{s\mathbf{k}_s} \\ & \left(\alpha\beta^{\mathbf{R}_1}|\gamma^{\mathbf{R}_2}\vartheta^{\mathbf{R}_2+\mathbf{R}_3}\right) \end{aligned} \tag{16}$$

has a finite value. Substituting Eq. (15) in Eq. (14) and replacing the summations over the reciprocal vector $\mathbf{k}$ in the first Brillouin zone, we get [26]

$$\begin{aligned} E^{(2)} = \frac{1}{\mathcal{W}^3} \sum_{ijab} \int_{BZ} d\mathbf{k}_i \int_{BZ} d\mathbf{k}_a \int_{BZ} d\mathbf{k}_b \Big\{ & 2|Q(ijab\mathbf{k}_i\mathbf{k}_a\mathbf{k}_b)|^2 \\ & - Re\left[Q(ijab\mathbf{k}_i\mathbf{k}_a\mathbf{k}_b)Q^*(ijba\mathbf{k}_i\mathbf{k}_b\mathbf{k}_a)\right] \Big\} \\ & \Big/ \left(\varepsilon_{i\mathbf{k}_i}^{HF} + \varepsilon_{jT(\mathbf{k}_a+\mathbf{k}_b-\mathbf{k}_i)}^{HF} - \varepsilon_{a\mathbf{k}_a}^{HF} - \varepsilon_{b\mathbf{k}_b}^{HF}\right) , \end{aligned} \tag{17}$$

where BZ means the integral range is the first Brillouin zone and $\mathcal{W}$ is the volume of the first Brillouin zone.

4 MBPT and GFM for Band Structure

Beyond the HF approximation, the band energies are defined as ionization potentials for occupied or valence bands, and as electron affinites for unoccupied or conduction bands, respectively [44]. They are no longer one-electron energies, but are quasi-particle energies, which can be calculated via MBPT [26, 44–48], CC [4(d), 5, 49–51], and propagator [6–9] or Green's function methods (GFM) [10–14].

From Eqs. (6) and (9), one can see that the orbital relaxation due to the removal or addition of an electron to the neutral system is infinitesimal. Then it is reasonable to use HF orbitals of the neutral system as the orbitals of its ion and anion systems and use the orbitals to construct their zeroth-order wavefunctions. Let us use $E^+(i\mathbf{k}_i)$ to denote the total energy of the ion system whose zeroth-order wavefunction can be obtained by removing an electron in the $\psi_{i\mathbf{k}_i}$ orbital from the ground state HF wavefunction of the neutral system. Then the band energy $\varepsilon_{i\mathbf{k}_i}$ is given by

$$\varepsilon_{i\mathbf{k}_i} = E - E^+(i\mathbf{k}_i)\,. \tag{18}$$

Similarly, let us use $E^-(a\mathbf{k}_a)$ to denote the total energy of the anion system whose zeroth-order wavefunction can be obtained by adding an electron to the $\psi_{a\mathbf{k}_a}$ orbital to the HF wavefunction of the neutral system and then the band energy $\varepsilon_{a\mathbf{k}_a}$ is determined by

$$\varepsilon_{a\mathbf{k}_a} = E^-(a\mathbf{k}_a) - E\,. \tag{19}$$

It should be mentioned that although the orbital relaxation in each orbital is infinitesimal, the collective effect on the total energy has a finite contribution since the number of orbitals is infinite [52]. The orbital relaxation can be accounted for by an operator when the total energy of an ion system is calculated. Since the total energies on the right-hand sides of Eqs. (18) and (19) are all infinite, we need to directly calculate the difference of the two energies on the right-hand side of each equation, which has a finite value. The orbital relaxation starts to have finite contributions in the third-order correction, or even in a first-order correction if non-canonical orbitals are used [52].

Just as for the total energy, we can use perturbation theory to express $\varepsilon_{n\mathbf{k}}$ as

$$\varepsilon_{n\mathbf{k}} = \varepsilon_{n\mathbf{k}}^{HF} + \varepsilon_{n\mathbf{k}}^{(2)} + \varepsilon_{n\mathbf{k}}^{(3)} + \ldots\,, \tag{20}$$

where n indicates either an occupied or unoccupied band. Using MBPT for E, $E^+(i\mathbf{k}_i)$, and $E^-(a\mathbf{k}_a)$, respectively, and then taking differences of the energies in Eqs. (18) and (19), one can get an analytical expression for a *direct* evaluation of any given-order correction to the band energies [52]. For the second-order correction, the analytical expression can be written as [26]

$$\varepsilon_{n\mathbf{k}}^{(2)} = U(n\mathbf{k}, 0) + V(n\mathbf{k}, 0)\,, \tag{21}$$

where

$$U(n\mathbf{k},\Delta\varepsilon) = \left(\frac{1}{\mathcal{W}}\right)^2 \sum_{iab} \int_{BZ} d\mathbf{k}_a \int_{BZ} d\mathbf{k}_b \Big\{ 2|Q(niab\mathbf{k}\mathbf{k}_a\mathbf{k}_b)|^2 - Re\big[Q(niab\mathbf{k}\mathbf{k}_a\mathbf{k}_b)Q^*(niba\mathbf{k}\mathbf{k}_b\mathbf{k}_a)\big]\Big\} \Big/ \left(\Delta\varepsilon + \varepsilon^{HF}_{n\mathbf{k}} + \varepsilon^{HF}_{jT(\mathbf{k}_a+\mathbf{k}_b-\mathbf{k}_i)} - \varepsilon^{HF}_{a\mathbf{k}_a} - \varepsilon^{HF}_{b\mathbf{k}_b}\right), \tag{22}$$

$$V(n\mathbf{k},\Delta\varepsilon) = \left(\frac{1}{\mathcal{W}}\right)^2 \sum_{aij} \int_{BZ} d\mathbf{k}_i \int_{BZ} d\mathbf{k}_j \Big\{ 2|Q(naij\mathbf{k}\mathbf{k}_i\mathbf{k}_j)|^2 - Re\big[Q(naij\mathbf{k}\mathbf{k}_i\mathbf{k}_j)Q^*(naji\mathbf{k}\mathbf{k}_j\mathbf{k}_i)\big]\Big\} \Big/ \left(\Delta\varepsilon + \varepsilon^{HF}_{n\mathbf{k}} + \varepsilon^{HF}_{aT(\mathbf{k}_i+\mathbf{k}_j-\mathbf{k})} - \varepsilon^{HF}_{i\mathbf{k}_i} - \varepsilon^{HF}_{j\mathbf{k}_j}\right). \tag{23}$$

There are no singularies in Eqs. (22) and (23), e.g. the denominators do not vanish, provided $\varepsilon^{HF}_{n\mathbf{k}}$ satisfies the condition that

$$\varepsilon^{HF}_{lu} + E^{HF}_g > \varepsilon^{HF}_{n\mathbf{k}} > \varepsilon^{HF}_{ho} - E^{HF}_g, \tag{24}$$

where ε^{HF}_{lu} and ε^{HF}_{ho} denote the lowest unoccupied and highest occupied HF orbital energies, respectively. In the range given in Eq. (24), the single determinant approximation for ion and anion systems should be reasonable. Beyond that range, the approximation may break down since degeneracy among zeroth-order determinants could occur. It is also reasonable to argue according to Eqs. (22) and (23) that the second-order correction $\varepsilon^{(2)}_{i\mathbf{k}_i}$ for the valence bands are positive while the second-order correction $\varepsilon^{(2)}_{a\mathbf{k}_a}$ for the conduction bands are negative. (The U(P) and V(P) should be exchanged in Eq. (55) and the associated paragraph in reference [26].) Then the second-order MBPT energy band gap is smaller than the HF energy band gap.

The propagator [6–9] or Green's function method (GFM) [10–14] provides another approach to calculate the quasi-particle energy bands. The Dyson equation provides the exact $E(N \pm 1)$ energies in a formally one-particle picture, but the equation can only be solved approximately in real applications [57]. With the irreducible self-energy part in the diagonal approximation being correct to second-order, the inverse Dyson equation can be written as [26]

$$\varepsilon^{(d)}_{n\mathbf{k}} = U(n\mathbf{k},\varepsilon^{(d)}_{n\mathbf{k}}) + V(n\mathbf{k},\varepsilon^{(d)}_{n\mathbf{k}}) \tag{25}$$

and solved iteratively. The first iteration corresponds to Eq. (21). When one goes to a third-order approximation, one needs to consider the off-diagonal elements of the second-order correction since they are likely to be more important than the third-order corrections. However, it is much more expensive to calculate the off-diagonal elements.

5 Convergence of MBPT(2) Corrections with Cutoffs

From Eq. (16), we know that there are three lattice summations which go to infinity, in the expression of $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$. In the real calculations, we have to cutoff the summations. N denotes the cutoff for the lattice summations. In Eqs. (17), (22), and (23), there are integrations over $\mathbf{k}$ in the first Brillouin zone. In real calculations, the integrations are done with a numerical method, e.g. the integrations become summations of the integrand at a set of points chosen in the first Brillouin zone. For crystals, one may need to take the so-called "special" sets of points which have been shown to be efficient [73, 74] while for polymers, we can simply let the chosen points be evenly distributed in the first Brillouin zone. Let us use K to denote the number of the points. Then the MBPT(2) corrections to both the total energy per unit cell and the band energy are functions of the cutoff, N and K. We have shown that the correction to the band energy converges much slower than that to the total energy per unit cell [26]. To get reliable and meaningful results, one must check the convergence of the corrections with the cutoffs, especially for the corrections to the band energies.

For polymers, we can get analytical expressions for the convergence of $E_{uc}^{(2)}$ and $\varepsilon_{n\mathbf{k}}^{(2)}$ with the lattice summation cutoff N. Let us use $E_{uc}^{(2)}(N)$ and $\varepsilon_{n\mathbf{k}}^{(2)}(N)$ to denote the approximate MBPT(2) corrections to the total energy per unit cell and energy of the nth band with the reciprocal vector $\mathbf{k}$, respectively. With multipole expansion techniques, we have shown that when N is large enough, $E_{uc}^{(2)}(N)$ is related to the exact value $E_{uc}^{(2)}$ by [27]

$$E_{uc}^{(2)}(N) = E_{uc}^{(2)} - A/N^3 + \ldots, \tag{26}$$

where A is a constant. This indicates that $E_{uc}^{(2)}$ in polymers converges with the lattice summation at least as fast as $1/N^3$. For the band energy, one of the two terms, U and V, converges with N as fast as $E_{uc}^{(2)}$ while the other is one power slower. Then $\varepsilon_{n\mathbf{k}}^{(2)}(N)$ can be expressed as [27]

$$\varepsilon_{n\mathbf{k}}^{(2)}(N) = \varepsilon_{n\mathbf{k}}^{(2)} - B_{n\mathbf{k}}/N^2 - C_{n\mathbf{k}}/N^3 + \ldots, \tag{27}$$

where $B_{n\mathbf{k}}$ and $C_{n\mathbf{k}}$ are constants for a given $n\mathbf{k}$. From Eqs. (26) and (27), we know that the MBPT(2) correction to a band energy converges one power slower than that to the total energy per unit cell.

6 Higher-Order MBPT Corrections and Their Convergence with Lattice Summations

The formal expressions for higher-order MBPT corrections for both the total energy and band energy are the same as those for the total energy and quasi-particle energy, respectively, in finite systems. By substituting Eq. (15) and

replacing the summations over the reciprocal vector $\mathbf{k}$ by $(\mathcal{N}/\mathcal{W})\int_{BZ} d\mathbf{k}$ in the first Brillouin zone, all the infinite and infinitesimal factors in the MBPT expressions for the total energy per unit cell and band energy cancel each other. Then in the expressions, there are only summations over band indices and integrations over the reciprocal vector in the first Brillouin zone [24]. These arithmetic operations do not introduce divergence but $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$ can, and then the integrands may have singularities [12, 75].

By using the multipole expansion for $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$, one can show that it becomes divergent or singular when and only when [24]

$$p = r\,, \qquad \mathbf{k}_p = \mathbf{k}_r\,, \qquad q = s\,, \qquad \mathbf{k}_q = \mathbf{k}_s\,. \tag{28}$$

It is easy to see that some of the $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$ integrals in the third-order MBPT become singular at certain points in the the space formed by the $\mathbf{k}$ vectors [24, 75].

The MBPT diagrams can always be classified into three types. In the first type of the diagrams, the two-electron integral $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$ never satisfies the above condition, such as the diagrams of $E_{uc}^{(2)}$ and $\varepsilon_{n\mathbf{k}}^{(2)}$. For these types of diagrams, there is no divergent problem at all since each $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$ converges with lattice summations at any point of $\mathbf{k}$ space. The second type of diagrams are those, in which some of the two-electron integrals $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$ diverge at discrete points in $\mathbf{k}$ space. We have shown that in this case, the value of such a diagram is still uniformly convergent with lattice summations [24]. In the third type of diagrams, some of the $Q(pqrs\mathbf{k}_p\mathbf{k}_r\mathbf{k}_s)$ diverge in a region. Then the value of each diagram becomes infinite or diverges with the lattice summations. However, these diagrams always occur in pairs [12, 75]. In each pair, the divergent parts of the two diagrams cancel each other [12, 24, 25]. Therefore, any order of MBPT correction either to the total energy per unit cell or to the band energies converges with lattice summations [24, 25].

7
Wannier Orbitals Versus Bloch Orbitals

Besides Bloch orbitals, Wannier orbitals [76] are also widely used in solid state physics. They are defined as Fourier transformations of the Bloch orbitals, e.g.

$$\begin{aligned} W_{nl}(\mathbf{r}) &= W_n(\mathbf{r} - \mathbf{R}_l) \\ &= \frac{1}{\sqrt{\mathcal{N}}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_l} \psi_{n\mathbf{k}}(\mathbf{r})\,. \end{aligned} \tag{29}$$

They can also be expressed as summations over atomic orbitals [15]

$$W_{nl}(\mathbf{r}) = \sum_j \sum_\alpha d_{\alpha j}^n \chi_\alpha(\mathbf{r} - \mathbf{R}_\alpha - \mathbf{R}_l - \mathbf{R}_j)\,, \tag{30}$$

where $d_{\alpha j}^n$ are Fourier transformations of the HF orbital coefficients $C_\alpha^{n\mathbf{k}}$. For insulators and semiconductors, the coefficients $d_{\alpha j}^n$ decrease rapidly. Then with

a given cutoff, each Wannier orbital spans a finite number of unit cells, i.e. it is localized. Since Bloch orbitals have freedom to add phase factors, the Wannier orbitals are not unique. Thus, one can obtain better localized Wannier orbitals by adjusting the phase factors of the Bloch orbitals. Several approaches have been developed to obtain optimized Wannier orbitals [54, 77]. However, the *intrinsic* sizes of the Wannier orbitals, e.g. the best localized Wannier orbitals which one can obtain by any optimization approach, are determined by the interactions of the particles in the system and the basis set [27]. Compared to Bloch orbitals, Wannier orbitals are localized orbitals while compared to atomic orbitals, they are delocalized. The optimized Wannier orbitals in a conjugate polymer with an STO-3G basis set typically span five unit cells on each side of the reference cell [15]. Their sizes may become larger when a larger basis set is used, especially the Wannier orbitals of the conduction bands. Of course, the size of a Wannier orbital is also dependent on the cutoff of the coefficients $d^n_{\alpha j}$.

One can transform the two-electron integrals over atomic orbitals to those over the Wannier orbitals and then transform the integrals to those over Bloch orbitals [15, 53]. As in finite systems, the number of the nonzero two-electron integrals over localized Wannier orbitals is much less than that of those over delocalized Bloch orbitals. In fact, the latter has an infinite value considering $\mathbf{k}$ as a continuous variable. However, the two-electron integrals over Bloch orbitals are continuous functions of the reciprocal vectors. The continuity enables us to calculate the integrals only at a set of discrete points in the space of the $\mathbf{k}$ vectors. Then the number of the two-electron integrals which need to be calculated over Bloch orbitals is not necessarily larger than that of those over the Wannier orbitals for a given accuracy, yet the two-step procedure brings an extra approximation, to which the MBPT corrections to band energies are more sensitive than that to the total energy per uint cell, since the former converges one power slower with the cutoff of the lattice. Furthermore, a seven-step procedure [78] can be developed for the direct transformation described in Eq. (16) while it cannot when Wannier orbitals are used. The seven-step procedure reduces the computational cost approximately by K, which is about twenty for polymers, compared to the conventional four-step procedure. So far, all our MBPT(2) calulations for polymers [26, 27, 66, 68] have been done on IBM/R6000 workstations.

8
Structure and Phonon Spectra

8.1
Geometry Optimization

For an infinite system which is periodic in ρ directions, the independent variables required to describe the nuclear structure are the internal coordinates in one unit cell and the ρ basis vectors, $\mathbf{a}_1$, ..., $\mathbf{a}_\rho$. Besides the translational

symmetry, the system may have point symmetry. In this case, the number of independent coordinates can be further reduced. Let us use $\mathbf{r} = r_1, r_2, ..., r_m$ to denote the independent internal coordinates of the system. Then the Hartree–Fock (HF) total energy per unit cell E_{uc}^{HF}, and the MBPT(2) total energy per unit cell $E_{uc}^{MBPT(2)}$, where

$$E_{uc}^{MBPT(2)} = E_{uc}^{HF} + E_{uc}^{(2)}\,, \tag{31}$$

(or total energies per unit cell of other correlated methods for a given basis set) are functions of r_1, r_2, ..., r_m. They are also the functions of $\mathbf{a}_1$, ..., $\mathbf{a}_\rho$. The optimized geometries of a method within the basis set are the geometries at which the energy of the method has minimum values. There can be more than one optimized geometry for a system. The geometry at which the system has the lowest energy is the most stable structure of the system. The other minima also correspond to other equilibrium structures. In a crystal, different equilibrium structures correspond to different phases. In a polymer, they correspond to different forms, e.g. polyacetylene has all-trans, trans-cisoid, and cis-transoid forms. The the transitions and the transition paths among phases or forms of a system can also be determined by locating saddle points [79] and following steepest descent lines [80], which connect the saddle points and the minima, on the energy surface. The energy of the system at an equilibrium structure also provides the information for determining the cohesive energy of the system at the structure.

8.2 Vibrational Frequencies

Just as in molecules, the nuclei in an infinite, periodic system vibrate around their equilibrium positions according to their vibrational modes which are called phonons in solid state physics. They determine the system's specific heat, thermal conductivity, thermal expansion, infrared and Raman spectroscopy, and other mechanical properties [81–83].

When the nuclei vibrate, they may break the system's translational symmetry. However, one can still classify the vibrational modes by the reciprocal vector $\mathbf{k}$. The vibrational frequencies $\omega(\mathbf{k})$ which belong to the reciprocal vector $\mathbf{k}$ are determined by [81, 84]

$$Det\left|\mathbf{M}^{-1/2}\mathbf{F}_X(\mathbf{k})\mathbf{M}^{-1/2} - \omega(\mathbf{k})^2\mathbf{I}\right| = 0\,, \tag{32}$$

where $\mathbf{M}$ is the mass matrix which is diagonal, $\mathbf{I}$ is an identity matrix, and $\mathbf{F}$ is the matrix of the force constants given by [85]

$$\{\mathbf{F}_X(\mathbf{k})\}_{A'\tau',A\tau} = \frac{\partial^2 E_{uc}}{\partial Y^{\dagger}_{\mathbf{k}A'\tau'}\partial Y_{\mathbf{k}A\tau}}\,. \tag{33}$$

The energy E_{uc} in Eq. (33) can be E_{uc}^{HF}, $E_{uc}^{MBPT(2)}$ or other correlated total energy per unit cell. The coordinates $Y_{\mathbf{k}A}$ are Fourier transforms of the nuclei's Carte-

sian coordinates. They break the system's translational symmetry except those where $\mathbf{k} = 0$.

When $\mathbf{k} = 0$, the nuclei in different units vibrate in phase, e.g., the translational symmetry is always kept. The vibrational modes for $\mathbf{k} = 0$ can be observed in an infrared (IR) spectra and then are called fundamental frequencies [67]. For a polymer with screw symmetry such as polyethylene, the vibrational modes for $k = \pi/a$ are also observable in an IR spectra [67]. For the vibrational modes for $\mathbf{k} = 0$, the matrix of force constants becomes [68]

$$\{\mathbf{F}_X\}_{ij} = \frac{\partial^2 E_{uc}}{\partial x_i \partial x_j}, \tag{34}$$

where x_i and x_j are the Cartesian coordinates of the nuclei in the reference unit cell. Since the translational symmetry is always kept, the energy derivatives $\partial^2 E_{uc}/\partial x_i \partial x_j$ can be calculated by either finite difference or analytical methods.

When $\mathbf{k} \neq 0$, the translational symmetry is broken in the vibrational motions. Then the finite difference method can no longer provide the energy derivatives required in Eq. (33) since the translational symmetry is required in all *ab initio* methods applied to calculate the energy of an extended, periodic system. One may cause the reciprocal vector of a certain vibrational mode to vanish by enlarging the unit cell of the system (the so called "super cell") [81], but this is very expensive.

For $\mathbf{k} = 0$, the Fock matrix and its derivatives with respect to the displacements of the nuclei are always block diagonal. Then one can directly apply the analytical derivative methods developed for finite systems to extended systems [69, 86, 87, 88]. But when the displacements break the translational symmetry, the Fock matrix and its derivatives are no longer block diagonal. To solve the CPHF equations, one needs to use the symmetrized (normal mode) coordinates instead of the Cartesian coordinates of the nuclei. Efficient analytical methods have been developed to calculate the energy derivatives for $\mathbf{k} \neq 0$ with both plane wave [89–90] and general basis functions [85]. The latter can be functions of nuclear coordinates and have linear dependence. These methods reduce the computational cost required to calculate the phonon spectrum with $\mathbf{k} \neq 0$ to the same as that needed for the spectrum at $\mathbf{k} = 0$.

8.3 Application to Polymethineimine

Figure 1 shows the structure of all-trans-polymethineimine and Table 1 lists the optimized geometries for both HF and MBPT(2) methods with three basis sets, STO-3G, 6-31G, and 6-31G** [68]. We can see that both the size of the basis sets and electron correlation have a strong influence on the stable structure, especially on the difference between the two bond lengths, $r_{N=C}$ and r_{C-N}.

Table 2 presents the HF and MBPT(2) fundamental vibrational frequencies for all-trans-polymethineimine [68]. The vibrational frequencies listed in each

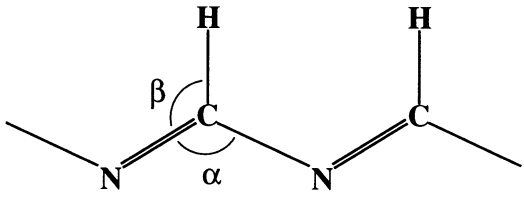

Fig. 1. The structure of polymethineimine and the five independent internal coordinates: $r_{N=C}$, r_{C-N}, r_{C-H}, α, and β

Table 1. The optimized geometries for polymethineimine using both HF and MBPT(2) with three basis sets, STO-3G, 6-31G, and 6-31G** (units: Å and degree) [68]

	$r_{N=C}$	r_{C-N}	r_{C-H}	α	β
HF/STO-3G	1.2814	1.4515	1.0973	116.63	125.03
MBPT(2)/STO-3G	1.3367	1.4836	1.1190	113.76	126.33
HF/6-31G	1.2657	1.3839	1.0876	120.30	121.82
MBPT(2)/6-31G	1.3039	1.4098	1.1088	118.31	122.96
HF/6-31G**	1.2616	1.3617	1.0939	119.61	121.55
MBPT(2)/6-31G**	1.2859	1.3717	1.1023	118.19	122.16

Table 2. Fundamental vibrational frequencies with symmetry A' calculated with different methods and basis sets (unit:1/cm). Unscaled [68]

	C-N str.	C-H def.	C=N str.	C-H str.
HF/STO-3G	1305.94	1517.19	1866.66	3576.59
MBPT(2)/STO-3G	1014.92	1352.07	1505.99	3340.30
HF/6-31G	1257.19	1527.82	1696.89	3176.84
MBPT(2)/6-31G	1173.12	1396.08	1596.59	2957.92
HF/6-31G**	1140.57	1379.49	1607.04	3135.55
MBPT(2)/6-31G**	1009.77	1265.55	1582.76	2996.73
Experiment [91]		1410	1620	3170

row are calculated by the combination of method and basis given in the first column at the corresponding optimized structure listed in Table 1. It is easy to see that both basis set and electron correlation have a large influence on the vibrational frequencies.

For STO-3G, the HF vibrational frequencies are larger than the experimental values. MBPT(2) [91] improves the numerical results by a few hundred wave numbers. When the 6-31G basis is used, both HF and MBPT(2) provide good agreement with experiment. When the larger basis set 6-31G** is used, MBPT(2) frequencies are even smaller than those obtained with 6-31G, although MBPT(2)/6-31G** frequencies still match the measured values reasonably well.

It is not clear that the experimental results are for pure all-trans-polymethineimine, as other cis-transoid and trans-cisoid structures are possible. In fact, Hirao and Iwata recently stated that the experiment results are for cis-transoid-polymethineimine, and reported DFT results with the L3YP functional [87]. To clarify this issue, we are calculating the MBPT(2) structures and vibrational frequencies for other isomers of polymethineimine. Further experiments are also needed to compare with theory.

9
X-Ray and Ultraviolet Photoelectron Spectroscopy

9.1
Formalism

X-ray (XPS) [58–61] and ultraviolet photoelectron spectroscopy (UPS) [62–65] provide rich information about the valence bands of extended systems. The further development of angle-resolved UPS (ARUPS) can even be used to directly observe the band structures. Then they can provide a tool to directly check theory.

The line intensities of the measured XPS and UPS spectra depend on both the photoionization cross sections and the frequency distribution of the incident radiation, e.g. [66]

$$I(E) = \sum_i \int t_i(E') d_i(E') F(E - E', \Gamma) dE' , \tag{35}$$

where $t_i(E')$ is the photoelectron cross-section of the orbital on the ith band with orbital energy E', $d_i(E')$ is the band's density of states (DOS), and $F(E - E', \Gamma)$ is the energy distribution of the incident radiation around the frequency of the maximum intensity. The photoionization cross-sections are functions of the angle of incident radiation, the energy of the radiation, and the angle of the emitted electrons. Hence, the ARUPS spectra varies with these three parameters [65]. The energy distribution of the incident radiation can be described by a linear combination of Lorentzian and Gaussian curves with a half-width at half-maximum Γ, e.g.,

$$F(x, \Gamma) = \frac{w_l}{\pi} \frac{\Gamma}{x^2 + \Gamma^2} + \frac{w_g}{\Gamma} \sqrt{\frac{ln2}{\pi}} exp \left[-\frac{ln2}{\Gamma^2} x^2 \right] , \tag{36}$$

where w_l and w_g are the weight of the Lorentzian and Gaussian curves, respectively.

9.2
Application to Polyethylene

Polyethylene has been well studied both experimentally and theoretically, although the purity of a polymer sample is always an issue. Several experimental

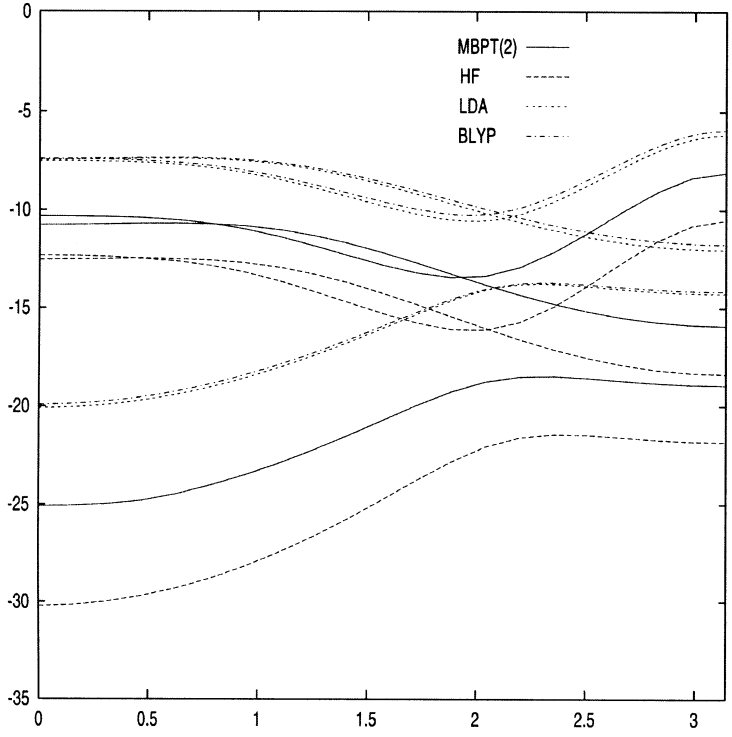

Fig. 2. The HF, LDA, BLYP, and MBPT(2) valence bands of all-*trans* polyethylene with basis set 6-31G as a function of k [66]

XPS and UPS for polyethylene have been reported while there are some discrepancies among them [60, 61, 65]. Semi-empirical methods provided some agreement with experiment in selected energy ranges, but failed in others [65]. HF offered a better description of the general features [92–94], but the difference between theory and experiment varied from 2 eV to 5 eV in the energy range of the valence bands. Recently, we applied MBPT(2) to interpret the measured XPS and UPS data and help resolve the discrepancies [66].

Figure 2 shows the HF, two DFT variants (LDA and gradient corrected BLYP), and MBPT(2) band structures of polyethylene calculated with the 6-31G basis set [66]. The MBPT(2) bands are above the HF and below the DFT ones. The correlation shift is different at different points in the bands, being around 2 eV for the first two bands. For the third band, the shift is about 5 eV at 0 and 3 eV at π/a. Table 3 list the calculated and measured peaks, the positions of which are not sensitive to the distribution of the incident radiation in the spectra [66]. The HF values are much larger than the measured data while DFT results are too small. As expected, MBPT(2) greatly reduces the

Table 3. Comparison among the peaks in the density of states calculated using HF, BLYP, and MBPT(2) with basis sets 6-31G and 6-31G**, respectively, and those measured by XPS and ARUPS [66] unit: eV

		I.P.	I_s	I_1	I_2	I_x	I_3	I_4	II_1	II_2
HF	6-31G	10.51		12.31	12.52		16.08	18.37	21.43	30.21
HF	6-31G**	10.55		12.37	12.64		15.99	18.27	21.50	30.03
BLYP	6-31G	5.93		7.39	7.40		10.24	11.73	13.70	19.92
MBPT(2)	6-31G	8.08	9.50	10.32	10.75		13.44	15.91	18.48	25.09
MBPT(2)	6-31G**	8.40[a]		10.59	11.06		13.54	15.99	18.39	24.66
XPS [60]	Polyethylene	8.6	9.6	11.2		12.6	13.8		18.0	23.6
XPS [61]	$C_{36}H_{74}$		9.8	11.1			13.7	15.4	18.0	23.8
ARUPS [65]	$C_{36}H_{74}$			10.5-12.0			14.0	15.5	18.3	24.6

[a] MBPT(2)/6-31G** calculations have been done only at the peaks.

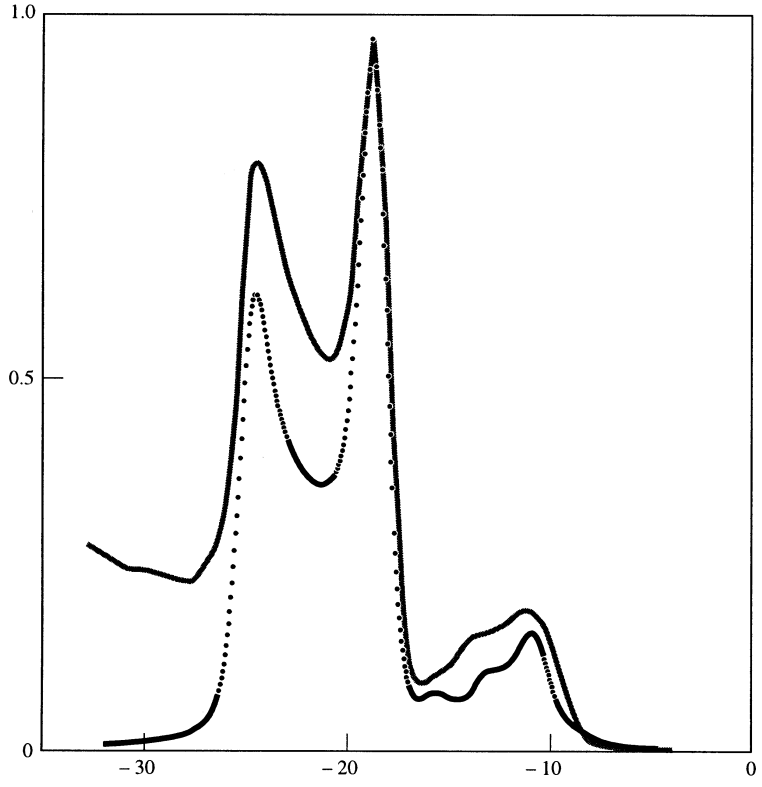

Fig. 3. Comparison between the experimental and the MBPT(2) XPS for polyethylene [66]. Solid line: experimental XPS measured by Pireaux et al. [16b]; circles: the MBPT(2) photoelectron spetra shown in Fig. 2c.

differences between theory and experiment. By considering the width of the incident radiation, the I_s has been shown as the result of the overlap among I_1, I_2 and the IP peaks while I_x is an artifact in the earlier experiment.

Figure 3 is a direct comparison between the MBPT(2) photoelectron spectra and the experimental XPS spectra measured by Pireaux et al. [61b] where the II_1 peak of the two spectra have been superimposed. In our calculation, the Γ in Eq. (36) was taken as 0.75 eV, which is the FWHM of the radiation used in the experiment [95]. Figure 3 shows that the two spectra match very well, even including the three small peaks.

Shake-up effects may start to contribute from the II_2 peak. That may be the reason why the measured peak is higher than calculated and also the reason the intensity beyond the II_2 peak is strong.

10
Band Gaps and Excitation Energies

The band gap in an extended system is defined as the difference between the lowest band energy in the conduction band and highest band energy in the valence band [15]. In the zeroth-order approximation, it equals the lowest excitation energy. But it is no longer correct beyond the zeroth-order approximation. Thus one should be careful when comparing calculated energy band gaps with experiment. Most of the measured band gaps in the literature are obtained from absorption spectra which are determined by excited states (excitons) of the systems.

The zeroth-order wavefunction of an excited state in an extended system cannot be described by a single determinant since the energies of the excited determinants form continuous bands. The states associated with the same band are degenerate since the energy difference between two adjacent states is an infinitesimal number. Thus, to calculate the energy spectrum of excited states, one needs to use degenerate MBPT, two-particle green function theory or a method like EOM-CC [5].

10.1
Polyacetylene

The optimized MBPT(2) geometries of polyacetylene using MBPT(2) with several basis sets were given by Suhai [53, 55]. At these geometries, he calculated the MBPT(2) band gaps of the system with the basis sets. Using GFM described in Eq. (25), Liegener [57] also calculated the band gaps of polyacetylene with three different basis sets at one geometry given by Suhai. There was a large inconsistency among these calculated band gaps. Recently, we showed that the inconsistency should come from the nonconvergence of MBPT(2) corrections to the band energies with lattice summation cutoff N [26]. As mentioned before, the MBPT(2) correction to a band energy converges one power slower with

Table 4. E_g^{HF}, $E_g^{(2)}$ and $E_g^{HF}+E_g^{(2)}$ for alternating trans-polyacetylene (unit: eV) [26]

Geometry	G1 and 6-31G**	G2 and DZP	G3 and DZP
E_g^{HF}	6.394	6.284	5.574
$E_g^{(2)}$	2.361	2.540	2.352
$E_g^{HF} + E_g^{(2)}$	4.033	3.744	3.222

N and is more sensitive to K than that of the total energy per unit cell [27]. To get reliable results, one has to check the convergence of the corrections with the cutoffs carefully.

Table 4 lists the MBPT(2) band gaps of polyacetylene calculated with basis set 6-31G** and DZP at three different geometries by us [36]. The cutoffs N and K are both 21. The geometries used in the calculations are listed in Table 5. The first two were given by Suhai [53, 55] and the last one was an experimentally estimated geometry [97]. The band gaps obtained are 4.033, 3.744, and 3.222 eV, respectively. There is no direct measurement of the band gap, defined as a quasi-particle energy difference of the lowest unoccupied and highest occupied orbitals. Instead, the absorption spectrum of polyacetylene crystalline films rises sharply at 1.4 eV and has a peak around 2.0 eV [97]. To explain this measured spectrum, one needs to calculate the density of the system's excited states and the absorption coefficients of the states.

Let us use Δ to denote the bond alternation of polyacetylene, e.g. $\Delta = r_2 - r_1$. Then the data in Table 4 show that the smaller the value of Δ, the smaller the band gap. However, the HF and MBPT(2) band gaps do not approach zero when Δ approaches zero [26]. They even do not become zero when Δ vanishes, except for the screw symmetry enforced in the HF calculation. The discontinuity of the HF band gaps around $\Delta = 0$ is due to fact that a single determinant cannot describe the system when $\Delta = 0$. The discontinuity is also known in finite systems. This phenomenon occurs when a system changes symmetry from a geometry which has a higher symmetry than those around it and at which there are at least two orbitals of the same representation, having exactly the same orbital energy, but when one is occupied while the other is not. This is the situation of polyacetylene at $\Delta = 0$.

10.2 Polymethineimine

Table 6 lists the HF and MBPT(2) band gaps with three basis sets for polymethineimine. The structure of the system used in each calculation is the optimized geometry obtained with the same method and basis. The number of unit cells in the lattice summation is the same as that used in geometry optimization, namely 21. From the table, we can see that electron correlation

Table 5. The geometries of polyacetylene used in Table 4 (units: Å and degree) [26]

	$r_{C=C}$	r_{C-C}	r_{C-H}	α	γ
G1 [53]	1.3660	1.4500	1.0900	123.90	118.05
G2 [55]	1.3634	1.4450	1.0872	123.75	118.13
G3 [96]	1.3800	1.4300	1.0900	122.00	118.50

Table 6. Band gaps calculated using HF and MBPT(2) with three basis sets, STO-3G, 6-31G, and 6-31G** (unit: eV) [68]

	$\varepsilon_{v,max}$	$\varepsilon_{c,min}$	E_g
HF/STO-3G	−7.3126	2.7502	10.0628
MBPT(2)/STO-3G	−6.1099	2.1524	8.2623
HF/6-31G	−9.0705	−0.8250	8.2454
MBPT(2)/6-31G	−7.5709	−2.5668	5.0041
HF/6-31G**	−8.9331	−0.3900	8.5431
MBPT(2)/6-31G**	−7.6166	−2.8340	4.7826

has a dramatic effect on the band gap. The MBPT(2) band gap obtained with 6-31G** is 4.7826 eV. Considering that the band gap of polyacetylene computed at the same level is 4.033 eV [26], we can estimate that the first peak in the polymethineimine's absorption spectrum would occur at about 2.75 eV if we can assume the same difference between the MBPT(2)/6-31G** band gaps and the first peak in the absorption spectrum for the two systems.

11 Conclusions

Compared to finite systems, it is of critical importance for extended systems that the calculated electron correlation corrections properly converge with lattice summations and the integrations over the reciprocal vectors. It has been shown that any-order MBPT corrections to both the total energy per unit cell and band energies converge with the summations and the integrations. For the MBPT(2) corrections, the analytical expressions for their convergence behavior with the lattice summation cutoff have been derived, which show that the MBPT(2) correction to a band energy converges one power slower than that to the total energy per unit cell.

MBPT(2) has been applied to several polymers to determine their equilibrium structures, vibrational frequences, XPS and UPS spectra, and band gaps. It has been shown that even the MBPT(2) correction has a strong effect on their structures and remarkable improvements to the HF results for the other three quantities, especially the XPS and UPS spectra, have been made.

MBPT(2) is just the first step in the construction of an electron correlation paradigm for polymers, surfaces, and crystals. We certainly believe that higher-

order correlations correction are necessary and are working on the next step in the paradigm. We are also developing similar MBPT(2) level approximations for electronic excitations. As in all systems, the interplay between band structure (a one-particle propagator quantity) and excited states (a two-particle propagator quantity) will be crucial in the reliable treatment of optical properties.

Acknowledgements. We would like to thank Prof. H.J. Monkhorst for valuable comments. This work is supported by the U. S. Office of Naval Research under grant No. N00014-92-J-1100.

12 References

1. (a) Löwdin PO (1967) Adv Chem Phys 14:283.
   (b) ibid (1959) ibid 2:207
2. Yarkony DR (1995) Modern Electronic Structure Theory, Vols. I and II. World Scientific, Singapore (and references therein)
3. (a) Kelly HP (1969) Adv Chem Phys 14:129.
   (b) Bartlett RJ (1981) Ann Rev Phys Chem 32:359 (and references therein)
4. (a) Čìžek J (1969) Adv Chem Phys 14:35.
   (b) Čìžek J, Paldus J (1971) Int J Quantum Chem 5:359.
   (c) Bartlett RJ (1989) J Phys Chem 93:1697.
   (d) Bartlett RJ (1995) in: Yarkony DR (ed) Coupled-Cluster Theory: An Overview of Recent Developments in Modern Electronic Structure Theory. World Scientific, Singapore, p 1047 (and references therein)
5. Bartlett RJ, Stanton JF (1994) in: Lipkowitz KB, Boyd DB (eds) Applications of Post-Hartree-Fock Methods: A Tutorial, in Reviews in Computational Chemistry 5. VCH, New York, p 65
6. (a) Linderberg J, Öhrn Y (1973) Propagators in Quantum Chemistry. Academic, London.
   (b) Purvis GD, Öhrn Y, (1974) J Chem Phys 60:4063.
   (c) ibid (1975) ibid 62:2045.
   (d) Öhrn Y (1976) Propagator Theory of Atomic and Moleculer Structure. In: Pullman B, Parr R (eds) The New World of Quantum Chemistry. Reidel, Boston, p 57
7. Nichols JA, Yeager DL, and Jørgensen P (1984) J Chem Phys 80:293
8. Golab JT, Yeager DL (1987) J Chem Phys 87:2925
9. Oddershede J (1987) Adv Chem Phys 69:201
10. Pickup PT, Goscinski O (1973) Mol Phys 26:1013
11. Csanak G, Taylor HS, Yaris R (1971) Advan Atomic Mol Phys 7:287
12. Cederbaum LS, Domcke W (1977) Adv Chem Phys 36:205
13. Paldus J, Čìžek J (1974) J Chem Phys 60:149
14. (a) Nooijen M, Snijders JG (1992) Int J Quantum Chem Symp 26:55.
   (b) ibid (1995) J Chem Phys 102:22
15. Ladik JJ (1988) Quantum Theory of Polymers as Solids. Plenum, New York
16. Mahan GD (1990) Many-Particle Physics. Plenum, New York
17. Kapuy E, Bogar F, Tfirst F (1994) Int J Quantum Chem 52:127
18. (a) Deleuze M, Delhalle J, Anderé JM (1992) Int J Quantum Chem 41:243.
   (b) Deleuze M, Cederbaum LS (1996) Phys Rev B 53:13326
19. (a) Grafenstein J, Stroll H, Fulde P (1997) Phys Rev B 55:13588.
   (b) Choi CH, Kertesz M, Karpfen A (1997) J Chem Phys 107:6712
20. Berggreen KF, Martino F (1969) Phys Rev 184:484.
21. Harris FE, Monkhorst HJ (1969) Chem Phys Lett 23:1026

22. (a) Delhalle J, Piela L, Brédas JL, André JM (1980) Phys Rev B 22:6254.
(b) Piela L, Delhalle J (1978) Int J Quantum Chem 13:605.
(c) Brédas JL, André JM, Delhalle J (1980) J Chem Phys 45:109
23. Deleuze M, Delhalle J, Pickup BT, Calais JL (1992) Phys Rev B 46:15668
24. Sun JQ, Bartlett RJ (1997) J Chem Phys 106:5554
25. Nooijen M, Bartlett RJ (1997) Int J Quantum Chem 63:603
26. Sun JQ, Bartlett RJ (1996) J Chem Phys 104:8553
27. Sun JQ, Bartlett RJ (1998) Phys Rev Lett 80:3669
28. Del Re G, Ladik JJ, Biczó G (1967) Phys. Rev. 155:997
29. (a) André JM, Gouverneur L, Leroy G (1967) Int J Quantum Chem 1:427.
(b) ibid 1:451
30. Kertész M (1976) Acta Phys Acad Sci Hung 41:127
31. (a) Suhai S, Ladik JJ (1977) Solid State Commun 22:227.
(b) Otto P, Clementi E, Ladik JJ (1982) IBM Technical Report, POK-13
32. Karpfen A (1981) Int J Quantum Chem 19:1297
33. (a) Monkhorst HJ (1979) Phys Rev B 20:1504.
(b) Monkhorst HJ, Kertész M (1981) Phys Rev B 24:3015.
(c) Stolarczyk LZ, Jeziorska M, Monkhorst HJ (1988) Phys Rev B 37:10646.
(d) Jeziorska M, Stolarczyk LZ, Paldus J, Monkhorst HJ (1990) Phys Rev B 41:12473.
(e) Aissing G, Monkhorst HJ (1993) Int J Quantum Chem Quantum Chem Symp 27:81
34. André JM, Delhalle J, Brédas JL (1991) Quantum Chemistry Aided design of Organic Polymers. World Scientific, Singapore
35. (a) Pisani C, Dovesi R, Roetti C (1988) Hartree-Fock ab initio Treatment of Crystalline Systems. Springer, Berlin, Heidelberg, New York.
(b) Dovesi R, Saunders VR, Roetti C (1992), An ab-initio Hartree-Fock LCAO Program for Periodic Systems (CRYSTAL92)
36. (a) André JM, Vercauteren DP, Bodart VP, Brédas JL, Delhalle J, Fripiat JG (1983) Documentation for an ab initio polymer program (PLH), IBM IS&TG, POK-28.
(b) André JM, Mosley DH, Champagne B, Delhalle J, Fripiat JG, Brédas JL, Vanderveken DJ, Vercauteren DP (1993) in: Clementi E (ed) METEXX-94, Methods and Techniques in Computational Chemistry, Vol. B. Stef, Cagliari, Chap. 10, p 423.
(c) André JM, Mosley DH, Champagne B, Delhalle J, Fripiat JG, Brédas JL (1993) LCAO Ab Initio Band Structure Calculations for Polymers (PLH93)
37. (a) Brueckner KA (1955) Phys Rev 97:1353.
(b) ibid 100:36.
(c) Goldstone J (1957) Proc Roy Soc London 239:267
38. (a) Bartlett RJ, Purvis GD III (1978) Int J Quantum Chem 14:561.
(b) ibid (1980) Physica Scripta 21:255.
(c) Bartlett RJ, Shavitt I, Purvis GD (1979) J Chem Phys 71:281
39. Hedin L (1965) Phys Rev 139A:796
40. (a) Hybertsen MS, Louie SG (1985) Phys Rev Lett 55:1418.
(b) ibid (1985) Phys Rev B 32:7005.
(c) ibid (1986) ibid 34:5390
41. (a) Godby RW, Schluter M, Sham LJ (1986) Phys Rev Lett 56:2415.
(b) ibid (1987) Phys Rev B 36:6497.
(c) ibid (1988) ibid 37:10159
42. a) Rohlfing M, Louie SG (1998) Phys Rev Lett 80:3320 b) ibid 81:2312
43. Koopmans TA (1933) Physica 1:104
44. (a) Kunz AB (1972) Phys Rev B 6:606.
(b) Pantelides ST, Mickish AJ, Kunz AB (1974) Phys Rev B 10:2602
45. Kvasnička V, Hubač I (1974) J Chem Phys 60:4483

46. Chong DP, Herring FG, McWilliams D (1974) J Chem Phys 61:78
47. Paldus J, Čìžek J (1975) Advan Quantum Chem 9:105
48. Born G, Kurtz HA, Öhrn Y (1978) J Chem Phys 68:74
49. (a) Monkhorst HJ (1977) Int J Quantum Chem Symp 11:421.
(b) Dalgaard E, Monkhorst HJ (1983) Phys Rev A 28:1217
50. (a) Sekino H, Bartlett RJ (1984) Int J Quantum Chem Symp 18:255.
(b) Geertsen J, Rittby M, Bartlett RJ (1989) Chem Phys Lett 164:57
51. Nooijen M, Bartlett RJ (1995) J Chem Phys 102:3629
52. Sun JQ, Bartlett RJ (1997) J Chem Phys 107:5058
53. Suhai S (1983) Int J Quantum Chem 23:1239
54. (a) Suhai S (1983) Chem Phys Lett 96:619.
(b) ibid Phys Rev B 27:3506
55. Suhai S (1992) Int J Quantum Chem 42:193
56. (a) Suhai S (1993) Int J Quantum Chem, Quantum Chem. Symp 27:147.
(b) ibid (1995) Phys Rev B 51:16553.
(c) ibid 52:1674
57. Liegener CM (1988) J Chem Phys 88:6999
58. (a) Clark DT, Kilcast D (1971) Nature 233:77.
(b) Ginnard CR, Riggs WM (1972) Anal Chem 44:1310.
(c) Clark DT et al. (1972) J Polym Sci A-1 10:1637
59. (a) Wood MH, et al. (1972) J Chem Phys 56:1788.
(b) André JM, Delhalle J (1972) Chem Phys Lett 17:145
60. (a) André JM, et al. (1973) Chem Phys Lett 23:206.
(b) Delhalle J, et al. (1974) J Chem Phys 60:595
61. (a) Pireaux JJ, Candano R, Verbist J (1974) J Electron Spectry 5:267.
(b) Pireaux JJ, et al. (1976) Phys Rev A 14:2133
62. Duke CB, et al. (1978) Chem Phys Lett 59:146
63. Fujihira M, Inokuchi H (1972) Chem Phys Lett 17:554
64. (a) Seki K, et al. (1977) J Chem Phys 66:3644.
(b) Ueno N, et al. (1990) Phys Rev B 41:1176
65. (a) Seki K, Inokuchi H (1982) Chem Phys Lett 89:268.
(b) Seki K, et al. (1986) Chem Phys 105:247 (and references therein)
66. Sun JQ, Bartlett RJ (1996) Phys Rev Lett 77:3669
67. Koenig JL (1992) Spectroscopy of Polymers. American Chem Society, Washington DC
68. Sun JQ, Bartlett RJ (1998) J Phys Chem 108:301
69. Hirata S, Iwata S (1998) J Chem Phys 109:4147
70. (a) Förner W (1992) Int J Quantum Chem 43:221.
(b) Ye J, Förner W, Ladik JJ (1993) Chem Phys 178:1.
(c) Knab R, Förner W, Čìžek J, Ladik JJ (1996) J Mol Struct (Theochem) 366:11.
(d) Knab R, Förner W, Ladik JJ (1997) J Phys: Condens Matter 9:2043.
(f) Förner W, Knab R, Čìžek J, Ladik JJ (1997) J Chem Phys 106:10248
71. Surján PR, Szabados Á, Bogár F, Ladik JJ (1997) Solid State Comm 103:639
72. (a) Li JB, Liu HL, Ladik JJ (1994) Chem Phys Lett 230:414.
(b) Wechsler D, Ladik JJ (1997) Phys Rev B 55:8544.
(c) Li JB, Ladik JJ (1997) Solid State Comm 102:769
73. Monkhorst HJ, Pack JD (1976) Phys Rev B 13:5189
74. Evarestov RA, Smirnov VP (1983) Phys Status Solidi B 119:9
75. Deleuze M, Pickup BT (1997) Int J Quantum Chem 63:483
76. Wannier GH (1934) Phys Rev 52:191
77. Blount ED (1962) Solid State Phys 13:305
78. Sun JQ, Bartlett RJ (unpublished)

79. (a) Schlegal HB (1987) Adv Chem Phys 67:249.
(b) Sun JQ, Ruedenberg K (1994) J Chem Phys 101:2157 (and references therein)
80. (a) Mckelvey JM, Hamilton JF (1984) J Chem Phys 80:579.
(b) Page M, McIver JW (1988) ibid 88:922.
(c) Sun JQ, Ruedenberg K (1993) ibid 99:5257
81. Srivastava GP (1990) The Physics of Phonons. Adam Hilger, Bristol
82. Callaway J (1976) Quantum Theory of the Solid State. Academic Press, New York
83. Leburton JP, Pascual J, Torres CS (1993) Phonons in Semiconductor Nanostructures. Kluwer Academic Publishers, Dordrecht
84. Piseri L, Zerbi G (1961) J Mol Spectrosc 26:254
85. Sun JQ, Bartlett RJ (1998) J Chem Phys 109:4209
86. (a) Teramae H, Yamabe T, Satoko C, Imamura A (1983) Chem Phys Lett 101:149.
(b) Teramae H, Yamabe T, Imamura A (1984) J Chem Phys 81:3564
87. Hirata S, Iwata S (1997) J Chem Phys 107:10075
88. (a) Hirata S, Iwata S (1998) J Chem Phys 108:7901.
(b) ibid J Mol Struc (THEOCHEM) 451:121.
(c) ibid J Phys Chem 102:8426
89. (a) Baroni S, Giannozzi P, Testa A (1987) Phys Rev Lett 58:1861.
(b) Giannozzi P, de Gironcoli S, Pavone P, Baroni S (1991) Phys Rev B 43:7231
90. Quong AA, Klein BM (1992) Phys Rev B 46:10734
91. (a) Wöhrle D (1971) Tetrahedron Lett 1969.
(b) ibid (1974) Makromol Chem 175:1751
92. André JM, Leroy G (1970) Chem Phys Lett 5:71
93. Karpfen A (1981) J Chem Phys 75:238
94. Teramae H, et al. (1983) Theoret Chim Acta 64:1
95. Deleuze M, Denis JP, Dlhalle J, Pickup BP (1993) J Phys Chem 97:5115
96. Chien JCW (1984) Polyacetylene Chemistry, Physics, and Material Science. Academic Press, London
97. Fincher Jr CR, Ozaki M, Tanaka M, Peebles D, Lauchlan L, Heeger AJ, MacDiarmid AG (1979) Phys Rev B 20:1589

Local Space Approximation Methods for Correlated Electronic Structure Calculations in Large Delocalized Systems that are Locally Perturbed

Bernard Kirtman

Department of Chemistry, University of California, Santa Barbara, CA 93106, USA.
E-mail: kirtman@chem.ucsb.edu

Rigorous electronic structure calculations including correlation at all common levels of theory are feasible for large delocalized systems that are locally perturbed using the generic local space approximation (LSA). The basic ideas involved in the LSA are reviewed in the context of a single determinant Hartree–Fock or Kohn–Sham treatment. It is, then, shown how this treatment can be extended so as to explicitly incorporate local many-body interactions. For a problem such as a catalytic reaction on a metal surface the translational symmetry of the bare metal can be readily exploited. Several 'proof of the method' studies are discussed which demonstrate that accuracy improves monotonically as the size of the local region is increased and that the initial rate of convergence is rapid.

Keywords: Electronic structure, Localization, Correlation, Chemisorption

Topics in Current Chemistry, Vol. 203

1 Introduction

It is a pleasure to contribute to this volume in tribute to Professor Ede Kapuy. The primary issues that will be addressed in the current chapter, namely localization and correlation, are those that were advanced by Professor Kapuy himself during the major part of his scientific career. Like him we have adopted a non-conventional approach that will, hopefully, serve as a strong stimulus for further activity in the field.

One of the most difficult problems for ab initio quantum chemistry is to determine the potential energy function for a chemical reaction on a metal surface. Why is this so? First of all, the metal substrate is strongly delocalized. This means that the system cannot be modeled [1] by considering just a small or medium-sized cluster of metal atoms. On the other hand, the band structure techniques that would simplify calculations for a bare metal surface cannot be directly applied because the translational symmetry is broken by the presence of the reactants. As a result one has the difficulty of dealing with extended interactions without the benefit of simplifications due to symmetry. Many problems involving surfaces, interfaces, impurities, or defects in solid state materials fall under this broad rubric along with various solution phenomena as well.

For a number of years we have been developing a very general approach to the type of situation described above. Our family of methods, based on the Local Space Approximation (LSA), may be viewed either as an embedding technique or as a procedure for combining various fragments of a system in order to generate the whole. As such it has elements in common with some of the numerous specific methods that have been presented for embedding or for combining fragments. However, the LSA possesses the following unique set of properties: (1) it can be applied to infinite systems; (2) it can be adapted for use with any level of electronic structure theory or with different levels for different regions; (3) a systematic scheme for monotonic improvement towards the 'exact' result is obtained; (4) long-range interactions between the local region of interest and the surroundings are automatically included; and (5) electronic charge may be freely transferred between the local region and its surroundings.

Although this chapter is intended as a review, albeit with primary emphasis on our own work, in the course of writing several new developments have occurred. Thus, in addition to presenting the basic ideas in terms of a single determinant model (Hartree–Fock, Kohn–Sham density functional theory), Sect. 2 also contains an improved procedure for including the orbital overlap between fragments. Then, in Sect. 3, we develop an entirely new hybrid LSA method where correlation in the local interaction region is treated by conventional ab initio quantum chemistry techniques (including triple excitations as well as singles, doubles and quadruples) while the surroundings are treated within Kohn–Sham theory. The simplifications that can be achieved due to translatio-

nal symmetry, when one of the fragments is infinitely periodic, are explored in Sect. 4. Finally, Sect. 5 contains a review of the studies carried out, thus far, to test the efficacy of the LSA methodology.

2
Basic Ideas/Hartree–Fock and Kohn–Sham Density Functional Theory

Since the LSA is a density-matrix-oriented approach, most of the basic ideas can be conveyed in terms of either the Hartree–Fock (HF) version or the version based on Kohn–Sham density functional theory (KS–DFT). We start by partitioning the entire system into two or more fragments. It is assumed that the fragments are amenable to an HF or KS–DFT treatment when separate but not, at least by ordinary means, when combined. An example is a catalytic reaction on a metal surface which may be described by two fragments — the bare surface on the one hand and the molecular reactant on the other. Depending upon the nature of the surface reaction it may be advantageous to further partition the molecular reactants. In general, the fragments may be either strongly, weakly, or non-bonded entities; the bonds may be either localized or delocalized; and the division may occur at a reactive site or far from such a site. Obviously, the possibilities are legion. The only strict requirement is that an HF or KS–DFT calculation be feasible for each of the fragments.

We wish to determine the HF or KS–DFT solution when the fragments are allowed to interact. As usual this problem may be solved iteratively starting with some initial guess for the density matrix. (Strictly speaking, in the KS–DFT case one only needs the density function which may, of course, be derived from the density matrix.) In order to generate an initial guess let us assume that an unrestricted (for convenience) HF or KS–DFT calculation has been carried out on each fragment and that the spins are correctly paired between entities where bonding interactions will take place. In that event, one may construct (for each spin) an approximate initial density matrix, $\mathbf{R}_0$, for the entire system as the direct sum of the density matrices of the individual fragments. However, $\mathbf{R}_0$ will not be idempotent because the basis functions on one fragment are not orthogonal to those on another.

The lack of idempotency indicates that the density matrix is not derivable from a single determinant wavefunction and the first step is to correct this deficiency. A couple of 'purification' schemes are available. One of them, originally proposed by McWeeny [2], has proven useful of late [3–5] in furtherance of the quest for linear scaling. In our original formulation [6] we suggested another procedure based on a sequence of symmetric orthogonalizations. Here we present a new treatment that is more in keeping with the overall LSA method. If the direct sum overlap matrix corresponding to $\mathbf{R}_0$ is denoted by $\mathbf{S}_0 (\mathbf{R}_0 \mathbf{S}_0 \mathbf{R}_0 = \mathbf{R}_0)$, then the idempotency condition for the interacting fragments may be written:

$$(\mathbf{R}_0 + \Delta\mathbf{R}^S)(\mathbf{S}_0 + \Delta\mathbf{S})(\mathbf{R}_0 + \Delta\mathbf{R}^S) = \mathbf{R}_0 + \Delta\mathbf{R}^S \tag{1}$$

where $\Delta\mathbf{S}$ contains the inter-fragment overlaps, and $\Delta\mathbf{R}^{\mathrm{S}}$ restores idempotency. It is easy to show that

$$\Delta\mathbf{R}^{\mathrm{S}} = -\mathbf{R}_0\Delta\mathbf{S}\,\mathbf{A}^{-1}\mathbf{R}_0 = -\mathbf{R}_0(\mathbf{A}^{-1})^{\dagger}\Delta\mathbf{S}\,\mathbf{R}_0 \tag{2}$$

with $\mathbf{A} = [\mathbf{1} + \mathbf{R}_0\,\Delta\mathbf{S}]$. Exactly analogous relations hold for the unoccupied counterpart of $\mathbf{R}$, which is denoted by $\mathbf{U}$ and satisfies

$$\mathbf{RSU} = \mathbf{0} = \mathbf{USR}\,, \qquad \mathbf{USU} = \mathbf{U}\,. \tag{3}$$

$\Delta\mathbf{U}^{\mathrm{S}}$ may be otained from Eq. (2) by replacing $\mathbf{R}_0$ with $\mathbf{U}_0$ and $\mathbf{A}$ with $\mathbf{1}+\mathbf{U}_0\,\Delta\mathbf{S}$. The matrix elements of $\Delta\mathbf{S}$ fall off exponentially with the distance between atomic centers. Thus, in practice, one can specify a region in space beyond which the overlap integrals between fragments may be neglected. If $\Delta\mathbf{S}_{\mathrm{C}}$ is the projection of $\Delta\mathbf{S}$ onto this cut-off region, then $\Delta\mathbf{R}^{\mathrm{S}}$ becomes

$$\Delta\mathbf{R}^{\mathrm{S}} = -\mathbf{R}_0\,\Delta\mathbf{S}_{\mathrm{C}}[\mathbf{1}_{\mathrm{C}} + (\mathbf{R}_0)_{\mathrm{C}}\,\Delta\mathbf{S}_{\mathrm{C}}]^{-1}\mathbf{R}_0 \equiv -\mathbf{R}_0\,\mathbf{Y}_{\mathrm{C}}^{\mathrm{R}}\,\mathbf{R}_0\,. \tag{4}$$

At this point we define the 'local space', L, as the subset of the total atomic orbital (AO) basis that is strongly engaged in the coulombic interactions between fragments. This assumes that these interactions are localized to some degree, which is the only assumption of our method. The choice of a particular L is, clearly, a matter of judgment. However, as L is systematically enlarged we know that an increasingly more accurate description of the interacting system will be obtained. At the same time the interaction energy monotonically approaches the correct value for the total AO basis. Thus, varying the size of L is exactly like varying the size of the basis set in an ordinary calculation. As a side benefit, the LSA method can be used [7] to determine the location of important interactions between fragments. Of course, the subspaces L and C may or may not be the same.

According to the assumption we have made the change in the density matrix, $\Delta\mathbf{R}^{\mathrm{X}}$, due to the coulombic interaction between fragments will be more or less localized. It is tempting to set $\Delta\mathbf{R}^{\mathrm{X}} = \mathbf{X}_{\mathrm{L}}$. By doing that, however, one is forced [8] to split off the local space from the remainder of the system to satisfy the idempotency condition. This results in an ordinary cluster model which does not allow electron transfer to or from the surroundings and, as we will see in Sect. 5, is unsuitable for our purposes. In order to properly embed the cluster we take advantage of the fact that the sum of the occupied and unoccupied molecular orbital (MO) spaces is identical to the total AO space. So, instead of $\Delta\mathbf{R}^{\mathrm{X}} = \mathbf{X}_{\mathrm{L}}$, we write

$$\Delta\mathbf{R}^{\mathrm{X}} = \mathbf{R}\sigma_{\mathrm{L}}^{\mathrm{R}}\mathbf{R} + \mathbf{R}\sigma_{\mathrm{L}}^{\mathrm{RU}}\mathbf{U} + \mathbf{U}\sigma_{\mathrm{L}}^{\mathrm{UR}}\mathbf{R} + \mathbf{U}\sigma_{\mathrm{L}}^{\mathrm{U}}\mathbf{U} \tag{5}$$

where $\mathbf{R}$ is initially $\mathbf{R}_0^{\mathrm{S}} = \mathbf{R}_0 - \mathbf{R}_0\,\mathbf{Y}_{\mathrm{C}}^{\mathrm{R}}\,\mathbf{R}_0$ and $\mathbf{U}$ is given by an exactly analogous expression.

The effect of the projectors $\mathbf{R}$ and $\mathbf{U}$ is to spread out the local interaction over the entire system. In general, then, only part of each occupied orbital of the interacting fragments will be included within L. As a result L will contain a non-integral number of electrons as desired for appropriate transfer of electronic

charge between the local space and the surroundings. As the size of the local space is increased more and more of the interaction is included. Thus, the only question is — how rapidly do the calculations converge to the 'exact' result?

The local space matrices σ_L^R, σ_L^{RU}, ... must be chosen in such a way as to maintain idempotency. We have shown elsewhere [9] that this can be done by means of the following prescription

$$\sigma_L^{RU} = \mathbf{X}_L\,[\mathbf{1}_L + \mathbf{U}_L\,\mathbf{X}_L\,\mathbf{R}_L\,\mathbf{X}_L]^{-1} = \left(\sigma_L^{UR}\right)^{\dagger}\,, \tag{6a}$$

$$\sigma_L^{R} = -\sigma_L^{RU}\,\mathbf{U}_L\,\mathbf{X}_L\,, \quad \sigma_L^{U} = \mathbf{X}_L\,\mathbf{R}_L\,\sigma_L^{RU}\,, \tag{6b}$$

where $\mathbf{X}_L$ is now a completely arbitrary local space matrix to be determined. The initial $\mathbf{R}_L$ is $(\mathbf{R}_0^S)_L$. On subsequent iterations $\mathbf{R}_L$ will be the local space projection of the updated density matrix. It is convenient to rewrite $\mathbf{R} = \mathbf{R}_0^S + \Delta\mathbf{R}^X$ in terms of the initial projectors (after correction for overlap), i.e.

$$\Delta\mathbf{R}^X = \mathbf{R}_0^S\,\tau_L^R\,\mathbf{R}_0^S + \mathbf{R}_0^S\,\tau_L^{RU}\,\mathbf{U}_0^S + \mathbf{U}_0^S\,\tau_L^{UR}\,\mathbf{R}_0^S + \mathbf{U}_0^S\,\tau_L^{U}\,\mathbf{U}_0^S \tag{7}$$

so that all the updating information is contained in the local space matrices τ_L^R, τ_L^{RU}, ...

After the first iteration $\tau_L^R = \sigma_L^R$, $\tau_L^{RU} = \sigma_L^{RU}$, ... On succeeding iterations we use $\Delta\mathbf{R}^X$ from Eq. (7) to obtain $\mathbf{R} = \mathbf{R}_0^S + \Delta\mathbf{R}^X$ and $\Delta\mathbf{U}^X = -\Delta\mathbf{R}^X$ to obtain $\mathbf{U} = \mathbf{U}_0^S + \Delta\mathbf{U}^X$. Then substitution into Eq. (5) determines the updated τ_L^R, τ_L^{RU}, ... in terms of the σ matrices. The resulting formulas involve straightforward matrix multiplications and additions entirely within the local space. In Eq. (7) $\Delta\mathbf{R}^X$ is projected in terms of $\mathbf{R}_0^S$ and $\mathbf{U}_0^S$. Later we will find it convenient to go back one step further to $\mathbf{R}_0$ and $\mathbf{U}_0$. This is readily done by taking advantage of Eq. (4).

In order to complete our description of the LSA method all that remains is to specify $\mathbf{X}_L$ and, then, to evaluate the interaction energy. An optimum choice for $\mathbf{X}_L$ is determined by the variation condition which yields [10] the local space analogue of a familiar result, namely $(\mathbf{RFU} + \mathbf{UFR})_L = \mathbf{0}$. Here $\mathbf{F}$ is either the Fock or the Kohn–Sham matrix. Since $\mathbf{R}$, $\mathbf{F}$ and $\mathbf{U}$ all depend upon $\mathbf{X}_L$, this is a nonlinear relation that must be solved iteratively. The simplest, but least efficient, method of solution is steepest descents which corresponds to the choice

$$\mathbf{X}_L = \xi(\mathbf{RFU} + \mathbf{UFR})_L \tag{8}$$

where the optimum scaling constant, ξ, is determined by minimizing either the total energy (i.e., maximizing the magnitude of the interaction energy) or the trace (tr) of $[(\mathbf{RFU} + \mathbf{UFR})_L(\mathbf{RFU})_L]$. More efficient conjugate gradient and quasi-Newton procedures are applicable as well. The simple relationship between $\mathbf{X}_L$ and $\mathbf{R}$ (or $\mathbf{U}$) through Eqs. (6a), (6b), and (7), indicates that analytical gradients should be straightforward to implement.

We turn now to construction of the Fock matrix and calculation of the energy. The quantities that are required are the projections $(\mathbf{RFR})_L$, $(\mathbf{RFU})_L = (\mathbf{UFR})_L^{\dagger}$, and $(\mathbf{UFU})_L$. Following exactly the same procedure used to obtain

the τ matrices in Eq. (7) from the σ matrices in Eq. (5) one can determine [11] $(\mathbf{RFR})_{\mathrm{L}}$, $(\mathbf{RFU})_{\mathrm{L}}, \ldots$ from $(\mathbf{R}_0^{\mathrm{S}}\mathbf{FR}_0^{\mathrm{S}})_{\mathrm{L}}$, $(\mathbf{R}_0^{\mathrm{S}}\mathbf{FU}_0^{\mathrm{S}})_{\mathrm{L}}, \ldots$ through simple matrix multiplications and additions on the local space. Furthermore, as indicated earlier, it is straightforward to express the latter in terms of the original $\mathbf{R}_0$ and $\mathbf{U}_0$ utilizing Eq. (4). Assuming that $C = L$, the pivotal quantities become $(\mathbf{R}_0\mathbf{FR}_0)_{\mathrm{L}}$, $(\mathbf{R}_0\mathbf{FU}_0)_{\mathrm{L}}, \ldots$. Otherwise, we would need to use whichever space (C or L) is the larger of the two. One convenient strategy is to follow convergence to the 'exact' result along the path on which $C = L$ as the size of the local space is increased.

The next step is to take into account the fact that $\mathbf{F}$ depends upon $\mathbf{R}$. Indeed, in the spin-polarized treatment $\mathbf{F}$ is a function of both spin density matrices (although the matrices $(\mathbf{R}_0\mathbf{FR}_0)_{\mathrm{L}}$, $(\mathbf{R}_0\mathbf{FU}_0)_{\mathrm{L}}, \ldots$ above are all associated with one spin or the other). The corrections to $\mathbf{F}$ due to interaction between fragments come from three sources. Since $\mathbf{F}$ depends upon $\mathbf{R}$ there will be overlap effects due to $\Delta\mathbf{R}^{\mathrm{S}}$ and electronic charge redistribution effects due to $\Delta\mathbf{R}^{\mathrm{X}}$. In addition, there will be new nuclear-electron attraction terms. In analogy with $\mathbf{R}$ (cf., Eq. (7)) we write

$$\mathbf{F} = \mathbf{F}_0 + \Delta\mathbf{F}^{\mathrm{S}} + \Delta\mathbf{F}^{\mathrm{X}} \tag{9}$$

where $\Delta\mathbf{F}^{\mathrm{S}}$ includes the new nuclear-electron interactions as well as the contribution due solely to $\Delta\mathbf{R}^{\mathrm{S}}$ (i.e. $\Delta\mathbf{R}^{\mathrm{X}} = 0$). We will focus on the more difficult $\Delta\mathbf{F}^{\mathrm{X}}$ term which arises from the electron-electron interactions.

In HF theory the change, $\Delta\mathbf{F}^{\mathrm{X}}$, due to a change in the density matrix of the same spin is

$$(\Delta\mathbf{F}^{\mathrm{X}})_{\mu\nu} = \sum_{\lambda,\sigma}(\Delta\mathbf{R}^{\mathrm{X}})_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\right] \tag{10}$$

where the Greek indices refer to AOs of the entire system. The contribution due to a change in the density matrix of opposite spin has exactly the same form as the first term on the right hand side (rhs) of Eq. (10). Since no new considerations are involved, we will ignore the latter in order to simplify the presentation. Using Eqs. (4) and (7) in (10) leads to a set of terms that are typified by

$$\sum_{ab\in\mathrm{L}} \mathbf{q}_{\mathrm{ab}}^{\mathrm{RU}} \sum_{\lambda,\sigma}(\mathbf{R}_0)_{\lambda a}(\mathbf{U}_0)_{\sigma b}\left[(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)\right] \tag{11}$$

where $\mathbf{q}_{\mathrm{L}}^{\mathrm{RU}}$ is a local space matrix which is a sum of various products of the τ_{L}, $\mathbf{Y}_{\mathrm{L}}$, $(\mathbf{R}_0)_{\mathrm{L}}$ and $(\mathbf{U}_0)_{\mathrm{L}}$ matrices. Note that the indices $a, b, \ldots$ denote local space AOs. In addition to the $\mathbf{R}_0 \ldots \mathbf{U}_0$ term displayed in (11) there are also $\mathbf{R}_0 \ldots \mathbf{R}_0$ and $\mathbf{U}_0 \ldots \mathbf{U}_0$ terms. The double sum over λ, σ represents a half-transformation of the two-electron integrals from the full space to the local space. It is exactly analogous to the first two steps of the usual 4-index transformation except that the MO coefficients are replaced here by matrix elements of $\mathbf{R}_0$ or $\mathbf{U}_0$. These calculations are time-consuming and it is fortunate, therefore, that they can be circumvented in the conventional HF procedure. In the LSA/HF treatment,

when applied to problems such as chemisorption on metals, there are major simplifications that follow from the translational symmetry of the substrate, as we will see in Sect. 4. Furthermore, we intend to go beyond the HF level (cf., Sect. 3), in which case such an integrals transformation (or its equivalent) cannot be avoided even in conventional calculations. Finally, an alternative to the usual method of evaluation, based on numerical integration, is discussed immediately below. The latter approach can be very efficient when used in standard quantum chemistry methods and has a number of advantages in the present context. In any event the double sum over the full space is done externally, i.e. prior to the SCF iterations, whereas the double sum over the local space is done internally.

A recent development in ab initio quantum theory has been the introduction of (partially) numerical schemes for dealing with the two-electron integrals in a way that reduces the scaling with the size of the system. One of these is the pseudospectral (PS) [12, 13] technique, which is closely related to another procedure known as resolution of the identity (RI) [14–16]. The use of these schemes in conjunction with the LSA has been discussed in detail elsewhere [17]. Here we present just the basic idea behind the PS approach.

Referring to expression (11) we look at the coulomb term given by $(\mu\nu|\lambda\sigma)$. First, an analytical integration is done to obtain the set of numerical one-electron potentials:

$$\mathbf{A}_{\lambda\sigma}(\mathbf{r}_g) = \int \mathrm{d}\mathbf{r} \frac{\phi_\lambda(\mathbf{r})\phi_\sigma(\mathbf{r})}{|\mathbf{r}_g - \mathbf{r}|} \tag{12}$$

on a preset grid of points $\mathbf{r}_g$. Then $\mathbf{A}_{\lambda\sigma}(\mathbf{r}_g)$ is transformed to the local space by constructing

$$\mathbf{A}^{\mathrm{RU}}_{ab}(\mathbf{r}_g) = \sum_{\lambda\sigma} (\mathbf{R}_0)_{\lambda a}\, \mathbf{A}_{\lambda\sigma}(\mathbf{r}_g)(\mathbf{U}_0)_{\sigma b}\,. \tag{13}$$

The advantage of Eq. (13) is that the double sum does not have to be repeated for each μ, ν pair. On the other hand, it must be evaluated at each grid point. We will see later (Sect. 4) that the calculation is considerably simplified when one of the fragments has translational symmetry. Additional simplifications due to the absence of inter-fragment matrix elements in $\mathbf{R}_0$ will also be explored at that time.

The interaction energy (ΔE) is given by a term, $\frac{1}{2}\mathrm{tr}\,[\Delta\mathbf{V}_{\mathrm{ne}}]\,\mathbf{R}_0$, due to switching on the nuclear-electron attraction between fragments plus the overlap and electron redistribution contributions [11]

$$\Delta \mathrm{E}^{\mathrm{S,X}} = \frac{1}{2}\mathrm{tr}\,[\mathbf{F}(\mathbf{R}_0) + \mathbf{F}(\mathbf{R})]\,\Delta\mathbf{R}\,. \tag{14}$$

In Eq. (14) a sum over both spin components is understood. The most difficult part of ΔE to calculate is the charge redistribution term that comes from $\mathbf{F}(\mathbf{R})\,\Delta\mathbf{R}$ in Eq. (14). In order to evaluate this contribution we must multiply

the expression in (11) by $\Delta\mathbf{R}_{\mu\nu}$, which (using the same analysis that led to (11)) may be written as a sum of terms typified by

$$\sum_{c,d\in \mathrm{L}} \mathbf{p}^{\mathrm{RU}}_{\mathrm{cd}}(\mathbf{R}_0)_{\mu c}\,(\mathbf{U}_0)_{\nu d}\,. \tag{15}$$

(Here $\mathbf{p}^{\mathrm{RU}}$ is similar to $\mathbf{q}^{\mathrm{RU}}$ except that contributions from $\Delta\mathbf{R}^{\mathrm{S}}$ are also included.) Multiplying (11) by the above expression and, then, determining the trace (i.e. summing over μ, ν) completes the remaining half of the 4-index integrals-transformation. In the numerical procedure this may be accomplished by means of two quarter-transformations which utilize

$$\mathbf{Z}^{\mathrm{U}}_a(\mathbf{r}_g) = \sum_{\mu}(\mathbf{U}_0)_{a\mu}\,\phi_\mu(\mathbf{r}_g) \quad \text{and} \quad \mathbf{Z}^{\mathrm{R}}_a(\mathbf{r}_g) = \sum_{\mu}(\mathbf{R}_0)_{a\mu}\,\phi_\mu(\mathbf{r}_g)\,. \tag{16}$$

Again, the sum over the full space indices is done externally leaving a sum over the four local space indices to be carried out on each iteration along with a weighted sum over the grid points. Note that the sum over the local space orbital indices of $\mathbf{A}^{\mathrm{RU}}(\mathbf{r}_g)$ may be carried out separately from the sum involving the orbital indices of the local space $\mathbf{Z}$ vectors. Although we have considered only the coulomb energy it should be clear that the same general approach is applicable to the exchange energy. Important modifications [13, 17] can be made to the numerical method to improve accuracy but they do not affect the general description given here. In the RI scheme a set of fitting functions plays a somewhat similar role to the points on a grid in the PS technique. The RI approach is even more similar to the numerical procedures usually employed in connection with DFT.

The construction of $\mathbf{F}^{\mathrm{KS}}$ and calculation of the energy in DFT is akin to the PS version of the HF treatment in that both combine numerical and analytical techniques. For the coulomb (Hartree) term, which is identical in either case, exactly the same procedure could be followed. However, in DFT it is usual to replace the matrix $\mathbf{A}_{\lambda\sigma}(\mathbf{r}_g)$ with a vector $\mathbf{A}_i(\mathbf{r}_g)$ obtained by fitting the density function ρ using an auxiliary set of basis functions $\mathrm{f_i}$. The contribution to ρ from the effects of inter-fragment overlap and electron redistribution may be found using (15). Multiplication of the typical term in $\Delta\mathbf{R}_{\mu\nu}$ — see expression (15) — by $\phi_\mu(\mathbf{r})\,\phi_\nu(\mathbf{r})$, followed by summation over μ and ν, yields the corresponding term in $\Delta\rho$:

$$\Delta\rho(\mathbf{r}) = \sum_{a,b}\mathbf{Z}^{\mathrm{R}}_a(\mathbf{r})\,\mathbf{Z}^{\mathrm{U}}_b(\mathbf{r})\mathbf{p}^{\mathrm{RU}}_{ab} + \ldots\,. \tag{17}$$

One can either fit the individual products $\mathbf{Z}_{\mathrm{a}}(\mathbf{r})\,\mathbf{Z}_{\mathrm{b}}(\mathbf{r})$, in complete analogy with the RI approach, or fit $\Delta\rho(\mathbf{r})$ after summing over a, b (as well as over all 4 components indicated by ... in Eq. (17)).

Next we consider the exchange–correlation potential (and the binding energy) which can also be evaluated by means of an intermediate fitting procedure. That is, in fact, the approach taken in our original paper [8] based on Sambe and Felton's [18] DFT treatment. (It is feasible as well, to determine

$\mathbf{F}^{KS}$ by integrating directly over the exchange–correlation potential $\mathbf{v}_{XC}[\rho]$). The corresponding contribution to the binding energy is obtained from a numerical integration of the expression

$$\Delta E_{XC} = \int \left(\varepsilon_{XC}[\rho_0] + \Delta\varepsilon_{XC}\right) \Delta\rho \mathbf{dr} + \int \Delta\varepsilon_{XC}\rho_0 \mathbf{dr}$$
$$\left(\Delta\varepsilon_{XC} = \varepsilon_{XC}[\rho_0 + \Delta\rho] - \varepsilon_{XC}[\rho_0]\right) \tag{18}$$

where ε_{XC} is the energy density. The remaining contributions to the binding energy are the same as in the HF model. In Eq. (18) ρ_0 (and $\Delta\rho$) is the total density, obtained as the sum of the two spin components.

A major appeal of DFT lies in its computational efficiency. Even if a conventional method is deemed necessary for the local space it will often be desirable to treat the surroundings at the DFT level. Thus, one is led to consider hybrid treatments within the LSA as discussed in the next section.

3 Hybrid LSA Methodology

Thus far we have focused on a single determinant treatment, either HF or KS–DFT, of the *entire* system. For various reasons — greater accuracy, transition states, excited states, etc. — it may be desirable to undertake a configuration interaction (CI) or a coupled cluster or many-body perturbation theory (CC or MBPT) calculation on just the local space. We present below the first formulation of a hybrid approach that will do this within the LSA methodology. Much of the following will be general; where specifics are required we particularize to the KS procedure for the surroundings. The HF method could be utilized as well, although for a metal it is not as appropriate.

It is assumed that a single determinant calculation has been carried out using the LSA and that the resulting (HF or KS) density matrix is available in the form of Eq. (7). As discussed on other occasions [6, 19], the first step is to generate a set of localized MOs (i.e., LMOs) that are associated with the local space. These LMOs define the occupied electron pairs (and triples, quadruples, etc.) that will be considered in the subsequent CI, CC, or MBPT treatment. An appropriate density-matrix-based localization procedure has been developed by Foster and Weinhold (FW) [20] following earlier ideas due to Roby [21]. Starting with $\mathbf{R}_L$ the FW scheme leads to a set of orthogonal local space natural hybrid orbitals (NHOs) associated with bi-centric bonds and mono-centric lone pairs. These NHOs are appropriate for systems where the bonds (and lone pairs) are well-localized. However, for chemisorption on a metal surface and other problems of interest this will often not be the case. In general, then, we prefer to obtain LMOs from the entire $\mathbf{R}_L$ without projecting onto atomic and di-atomic subspaces. (If it is possible to further localize any of the resulting orbitals, then that can be determined by a separate calculation on the appropriate subspace). The transformation from AOs to LMOs on the local

space will be denoted by $\mathbf{B}_L$. LMOs with small occupancy can be eliminated at this point. Unless they have exactly unit occupancy, even the most highly occupied LMOs will contain contributions from the unoccupied (virtual) space. However, we want to work with orbitals that are completely occupied despite the fact that this can only be accomplished through a partial delocalization. Thus, the unoccupied contributions are annihilated, as in our previous work [19], by projection with **R**:

$$\phi_L^{LMO} = \mathbf{B}_L \, \mathbf{R} \, \chi^{AO} . \tag{19}$$

Here ϕ_L^{LMO} is the column vector of LMOs that are associated with the local space but contain contributions from the entire set of AOs in the column vector χ^{AO}. In Ref. [19] **SR** was used instead of **R**. Either choice will suffice to eliminate all virtual state contributions but it is simplest to use Eq. (19). At this point the orbitals ϕ_L^{LMO} are non-orthogonal; they will be orthogonalized later on.

We never explicitly utilize the LMOs of the surroundings. They could be determined in a manner similar to the local space LMOs except that they must be orthogonal to the latter. In principle, the usual Gram–Schmidt procedure could be employed for that purpose. Finally, it is convenient to symmetrically orthogonalize the local space LMOs which involves only local space operations since, from Eq. (19), $\mathbf{S}_L^{LMO} = \mathbf{B}_L \, \mathbf{R}_L \, \mathbf{B}_L^{\dagger}$:

$$\tilde{\phi}_L^{LMO} = (\mathbf{S}_L^{LMO})^{-1/2} \, \phi_L^{LMO} . \tag{20}$$

The general approach used in LSA methods to introduce electron correlation is based on Meyer's [22] self-consistent electron pair (SCEP) theory, later extended by Ahlrichs [23] and Dykstra et al. [24–27]. Additional important improvements in SCEP theory were made by Saebo and Pulay [28–36] leading ultimately to their very successful 'local correlation' treatment. We have adopted many of the specific features devised by Saebo and Pulay. Indeed, at first glance, the similarity between their treatment and ours might be more evident than the differences. However, the LSA deals particularly with the situation where a calculation of the entire system is impractical at *any* level. Thus, in contrast with Saebo and Pulay's procedure, our method applies even for the HF model. Moreover, because we start with isolated fragments the LSA method can be utilized to treat fully delocalized problems like chemisorption on a metal surface which are not amenable to a 'local correlation' treatment.

At the heart of SCEP theory lie the matrices, $\mathbf{C}_P$, that give the coefficients for excitation of electrons from a given pair of occupied MOs ($P = ij$) to an arbitrary pair of AOs p, q. In the spin-unrestricted treatment i and j refer to spinorbitals with $i > j$. The use of a matrix formulation in an AO basis leads to a perfect fit between SCEP theory and either the KS/ or HF/ LSA method. In treating pair correlations the idempotency condition on **R** is replaced by a strong orthogonality requirement on the double excitations — namely, $\mathbf{C}_P\mathbf{SR} = \mathbf{0} = \mathbf{RSC}_P$ or, equivalently,

$$\mathbf{C}_P = \mathbf{U} \, \mathbf{X}_P \, \mathbf{U} \tag{21}$$

where $\mathbf{X}_P$ is arbitrary. Clearly, it is much easier to satisfy the strong orthogonality for $\mathbf{C}_P$ than the more complicated idempotency condition for $\mathbf{R}$.

The form of the SCEP treatment will vary in certain aspects depending upon whether it is employed to carry out a CI, CC or Moller-Plesset (MP) perturbation theory calculation. However, the differences are modest and the same quantities appear in one place or another. For convenience we utilize here the MP perturbation theory version of SCEP as formulated by Pulay and Saebo [30, 31] for their local correlation treatment. The (Hylleraas) variation condition on the first-order coefficient matrix, $\mathbf{C}_P^{(1)} = \mathbf{C}_P$, may be written in the form

$$\langle \mathbf{T}_{ij}^{(2)} \delta \mathbf{C}_{ij}^{+} \rangle = \mathbf{0} \tag{22}$$

where the angular brackets signify the trace. In Eq. (22) $\mathbf{T}_{ij}^{(2)}$ is the second order residuum matrix which, in an AO spinorbital representation, is given by

$$\mathbf{T}_{ij}^{(2)} = (\mathbf{K}_{ij} - \mathbf{K}_{ji}) + (\mathbf{F}\,\mathbf{C}_{ij}\,\mathbf{S} + \mathbf{S}\,\mathbf{C}_{ij}\,\mathbf{F}) - \mathbf{S}\Big[\sum_k F_{ik}\,\mathbf{C}_{kj} + F_{kj}\,\mathbf{C}_{ik}\Big]\mathbf{S}\,. \tag{23}$$

Here $(\mathbf{K}_{ij})_{pq} = (ip|jq)$, $\mathbf{F}$ is the AO matrix corresponding to the operator

$$\mathbf{F} = \mathbf{h} + \sum_i (\mathbf{J}_{ii} - \mathbf{K}_{ii})\,. \tag{24}$$

$(\mathbf{J}_{ii})_{pq} = (ii|pq)$, and F_{ik} is the matrix element $\langle i|\mathbf{F}|k\rangle$. As above the AOs are denoted by $p, q, r, \ldots$ while the LMOs are $i, j, k, \ldots$. Even for $k \neq i$, F_{ik} will generally be non-zero in an LMO basis. Furthermore, the KS (as opposed to HF) orbitals are only approximate eigenfunctions of $\mathbf{F}$ since the latter is an HF Hamiltonian. This will show up later in the fact that $\mathbf{RFU}, \mathbf{UFR} \neq \mathbf{0}$.

For a complete SCEP treatment of the entire system the sum in Eq. (24) would run over all occupied LMOs. In the hybrid LSA method we want to limit the sum to the LMOs associated with the local space and replace the remainder by an appropriate potential. Within the KS procedure we make the replacement

$$\sum_i (\mathbf{J}_{ii} - \mathbf{K}_{ii}) \rightarrow \sum_{i\in L} (\mathbf{J}_{ii} - \mathbf{K}_{ii}) + \big\{\mathbf{v}_H[\rho] + \mathbf{v}_{XC}[\rho]\big\} - \big\{\mathbf{v}_H[\rho_L] + \mathbf{v}_{XC}[\rho_L]\big\} \tag{25}$$

where ρ_L is the density calculated from the orthogonalized local space LMOs. Note that $\mathbf{v}_H[\rho] + \mathbf{v}_{XC}[\rho]$ accounts for electron correlation throughout the entire system. After subtracting $\mathbf{v}_H[\rho_L] + \mathbf{v}_{XC}[\rho_L]$ correlation effects within the surroundings, *as well as* those between the surroundings and the local space, will remain. In treating correlation within the local space, therefore, we must exclude terms that couple the local space with the surroundings so that such terms are not double-counted. Consequently, the sum over k in Eq. (23) is restricted to local space LMOs.

At this point we introduce the LSA by substituting $(\mathbf{X}_P)_L$ for $\mathbf{X}_P$ in Eq. (21). Then the variation condition in Eq. (22) becomes $\langle (\mathbf{U}\,\mathbf{T}_{ij}^{(2)}\mathbf{U})_L (\delta X_{ij}^{\dagger})_L \rangle = 0$, or

$$(\mathbf{U}\,\mathbf{T}_{ij}^{(2)}\mathbf{U})_L = \mathbf{0} \tag{26}$$

since all the elements of $(\delta \mathbf{X}_P^{\dagger})_{\mathrm{L}}$ are freely variable. In combination with the expression for the residuum $\mathbf{T}_{ij}^{(2)}$ (cf. Eq. (23)) this condition leads to a set of linear coupled equations that determine the matrices $(\mathbf{X}_{ij})_{\mathrm{L}}$ for $ij \in L$. It follows immediately that $(\mathbf{X}_{ij})_{\mathrm{L}} = -(\mathbf{X}_{ji})_{\mathrm{L}} = -(\mathbf{X}_{ij})_{\mathrm{L}}^{\dagger}$.

Now let us examine, in turn, each of the local space matrices arising from the rhs of Eq. (23). The first is $\left[\mathbf{U}\,\overline{\mathbf{K}}_{ij}\,\mathbf{U}\right]_{\mathrm{L}}$ with $\overline{\mathbf{K}}_{ij} = \mathbf{K}_{ij} - \mathbf{K}_{ji}$. This matrix can be evaluated by exactly the same methods that were discussed earlier in connection with the HF/LSA method. In fact, using Eqs. (19) for the (non-orthogonal) local space LMOs one is led to exactly the same integral transformation that was discussed in Sect. 2. If this is done by the conventional method, then the entire set (all $i > j$) of $\left[\mathbf{U}\,\overline{\mathbf{K}}_{ij}\,\mathbf{U}\right]_{\mathrm{L}}$ matrices may be determined prior to the self-consistent iterations. Alternatively, the numerical scheme of Sect. 2 can be employed. Then $\left[\mathbf{U}\,\overline{\mathbf{K}}_{ij}\,\mathbf{U}\right]_{ab}$ can be written as a sum of terms each having the form:

$$\sum_{\mathbf{r}_g} \mathbf{Z}_a(\mathbf{r}_g)\mathbf{Z}_i(\mathbf{r}_g)\mathbf{A}_{bj}(\mathbf{r}_g)\,. \tag{27}$$

Here we have omitted the superscripts R and/or U on **Z** and **A**, which will vary from term to term. Because of the factorization in the above expression it is straightforward to calculate all matrix elements (ab) for all pairs (ij) in parallel. Although one can delay the sum over grid points until solving for $(\mathbf{X}_{ij})_{\mathrm{L}}$ that does not appear to be profitable.

Thus far we have not utilized Saebo and Pulay's partitioning of the unoccupied space into domains [29, 31, 32] although that is the primary reason for the computational savings they achieve over conventional correlation treatments. This was done deliberately because we envision application to problems where there is substantial delocalization, such as chemisorption on metal surfaces. Saebo and Pulay's procedure, on the other hand, was designed specifically for the case of a system with localized bonds. Nonetheless, the general idea of domains (of arbitrary size) could be incorporated within the LSA method simply by allowing for a different local space to be associated with each $P = ij$ pair. For well-localized LMOs (small domains) this should improve efficiency; for delocalized LMOs, especially if there are large components near the periphery of L, one might improve accuracy by using a local space that is centered near the periphery.

Now let us proceed to the second term on the rhs of Eq. (23). Using Eq. (40) for $\mathbf{C}_{ij}$ and, then, taking the local space projection of the **U** ... **U** component as in Eq. (26), this term becomes

$$(\mathbf{UFU})_{\mathrm{L}}(\mathbf{X}_{ij})_{\mathrm{L}}\,\mathbf{U}_{\mathrm{L}} + \mathbf{U}_{\mathrm{L}}\,(\mathbf{X}_{ij})_{\mathrm{L}}(\mathbf{UFU})_{\mathrm{L}}\,. \tag{28}$$

As defined by Eqs. (24)–(25) **F** is a combination of terms from the HF and KS Hamiltonians. Thus, we have already described how to evaluate all contributions to $(\mathbf{UFU})_{\mathrm{L}}$ in Sect. 2. The set of matrices $(\mathbf{X}_{ij})_{\mathrm{L}}$ $(i > j)$ are the quantities to be determined. They may be obtained in the usual manner by transforming

to a temporary molecular orbital type basis on the local space which diagonalizes $(\mathbf{UFU})_{\mathrm{L}}$. If $\mathbf{U}_{\mathrm{L}}$ were diagonal in this basis, and the interpair coupling (last term on the rhs of Eq. (23)) could be ignored, this would yield an immediate solution. Since that is not the case $(\mathbf{X}_{ij})_{\mathrm{L}}$ must be determined iteratively as in the usual SCEP treatment.

The last term on the rhs of Eq. (23) describes the coupling between pairs. When substituted into Eq. (26) it takes the form

$$-\mathbf{U}_{\mathrm{L}} \sum_{k\in \mathrm{L}} \left[F_{ik}\,(\mathbf{X}_{kj})_{\mathrm{L}} + F_{jk}\,(\mathbf{X}_{ik})_{\mathrm{L}}\right] \mathbf{U}_{\mathrm{L}}\,. \tag{29}$$

We observe that the sum over k is restricted here to the local space as discussed in connection with Eq. (25). Finally, using the self-consistent set of $(\mathbf{X}_{ij})_{\mathrm{L}}$ the local MP second-order energy for a UHF reference function may be obtained as [31]:

$$\mathrm{E}_{\mathrm{L}}^{(2)} = \frac{1}{4} \sum_{i,j\in \mathrm{L}} \langle \left[\mathbf{U}\,\overline{\mathbf{K}}_{ij}\,\mathbf{U}\right]_{\mathrm{L}} (\mathbf{X}_{ji})_{\mathrm{L}} \rangle\,. \tag{30}$$

The last step in the hybrid (DFT + MP2)/LSA treatment is to calculate the total binding energy, $\Delta\mathrm{E}$. In analogy with Eqs. (24)–(25) the hybrid expression is given by

$$\Delta \mathrm{E} = \Delta \mathrm{E}^{\mathrm{KS}}[\rho] + \left\{\Delta \mathrm{E}_{\mathrm{X;L}}^{\mathrm{HF}} + \Delta \mathrm{E}_{\mathrm{L}}^{(2)} - \Delta \mathrm{E}_{\mathrm{XC}}^{\mathrm{KS}}[\rho_{\mathrm{L}}]\right\} \tag{31}$$

where the quantity in curly brackets is the difference between the HF + MP2 energy and the KS energy for the local space. Since the one-electron terms and the Hartree energy are the same in either case (both calculated with KS orbitals) only the exchange and correlation contributions remain. As before, $\Delta\mathrm{E}$ denotes the energy change arising from the interaction between fragments due to orbital overlap and nuclear-electron attraction as well as the resulting electron redistribution.

The SCEP formalism is readily extended to yield the MP3 and MP4 methods. Using the third-order energy expression of Saebo and Pulay [31] together with Eq. (21) one obtains

$$\mathrm{E}_{\mathrm{L}}^{(3)} = \frac{1}{8} \sum_{ij} \langle (\mathbf{U}\,\mathbf{T}_{ij}^{(3)}\mathbf{U})_{\mathrm{L}} (\mathbf{X}_{ji})_{\mathrm{L}} \rangle \tag{32}$$

with

$$\begin{aligned}\mathbf{T}_{ij}^{(3)} = \mathbf{T}_{ij}^{(2)} + \mathbf{K}(\mathbf{C}_{ij}) + 2\sum_{k} \big\{ &(\mathbf{K}_{ik} - \mathbf{J}_{ik})\mathbf{C}_{kj}\mathbf{S} - \mathbf{S}\mathbf{C}_{jk}(\mathbf{K}_{ki} - \mathbf{J}_{ki}) - (\mathbf{K}_{jk} - \mathbf{J}_{jk})\mathbf{C}_{ki}\mathbf{S} \\ &+ \mathbf{S}\mathbf{C}_{ik}(\mathbf{K}_{kj} - \mathbf{J}_{kj})\big\} + \sum_{kl} \big\{(ki|lj) - (kj|li)\big\}\,\mathbf{S}\mathbf{C}_{kl}\mathbf{S}\,.\end{aligned} \tag{33}$$

Except for

$$\left[\mathbf{K}(\mathbf{C}_{ij})\right]_{\mu\nu} = \sum_{\lambda\sigma} (\mu\lambda|\nu\sigma)(\mathbf{C}_{ij})_{\lambda\sigma} \tag{34}$$

all the terms on the rhs of Eq. (33) have appeared previously. When substituted into Eq. (32) the latter term gives the energy contribution:

$$\sum_{abcd}\sum_{ij}(\mathbf{X}_{ji})_{ab}(\mathbf{X}_{ij})_{cd}\sum_{\mu\nu\lambda\sigma}\mathbf{U}_{a\mu}\mathbf{U}_{b\nu}\mathbf{U}_{c\lambda}\mathbf{U}_{d\sigma}(\mu\lambda|\nu\sigma)\,. \tag{35}$$

It is easy to show that the summation over μ, ν, λ, σ leads to the same 4-index transformations that were discussed in Sect. 2 and, therefore, no new considerations arise in evaluating this expression.

The appropriate SCEP formulas for the fourth-order energy, excluding the triples contribution, have also been given by Saebo and Pulay [31]. Although there are some new aspects as far as the singles term is concerned, much of the treatment is very similar to what we have already presented. For example, the equation for the second-order pair coefficient matrix, $\mathbf{C}_{ij}^{(2)}$, is the same as Eq. (26) except that one must substitute

$$\mathbf{T}_{ij}^{(4)} = \mathbf{T}_{ij}^{(3)} - \overline{\mathbf{K}}_{ij} + \left[\mathbf{T}_{ij}^{(2)}\right]^{\mathbf{c}_{ij}\rightarrow\mathbf{c}_{ij}^{(2)}} \tag{36}$$

for $\mathbf{T}_{ij}^{(2)}$. Then the local energy due to the second-order pair excitations may be calculated from Eq. (32) simply by relacing $(\mathbf{X}_{ji})_{\mathrm{L}}$ with its second-order counterpart and multiplying by 2. The fourth-order energy due to quadruple excitations is given by

$$\mathrm{E}_{\mathrm{Q;L}}^{(4)} = \frac{1}{8}\sum_{i,j\in\mathrm{L}}\langle(\mathbf{u}_{ij} - \mathbf{u}_{ji})\mathbf{S}\mathbf{C}_{ij}\mathbf{S}\rangle \tag{37}$$

in which

$$\begin{aligned}\mathbf{u}_{ij} = \sum_{k,l\in\mathrm{L}}\big\{&\langle\overline{\mathbf{K}}_{kl}\,\mathbf{C}_{ji}\rangle\mathbf{C}_{kl} \;-\; 4\mathbf{C}_{ij}\,\overline{\mathbf{K}}_{kl}\,\mathbf{C}_{lk}\\ &-\; 4\langle\overline{\mathbf{K}}_{kl}\,\mathbf{C}_{lj}\rangle\mathbf{C}_{ik} + 8\mathbf{C}_{ik}\,\overline{\mathbf{K}}_{kl}\,\mathbf{C}_{lj}\big\}\,.\end{aligned} \tag{38}$$

Applying the LSA we find, for example, that the second term on the rhs of Eq. (38) generates

$$-\frac{1}{2}\sum_{ijkl\in\mathrm{L}}\langle\mathbf{U}_{\mathrm{L}}(\mathbf{X}_{ij})_{\mathrm{L}}(\mathbf{U}\,\overline{\mathbf{K}}_{kl}\,\mathbf{U})_{\mathrm{L}}(\mathbf{X}_{lk})_{\mathrm{L}}\mathbf{U}_{\mathrm{L}}(\mathbf{X}_{ji})_{\mathrm{L}}\rangle\,. \tag{39}$$

All of the other contributions are similar in the sense that they involve local space matrix products where $\mathbf{U}_{\mathrm{L}}$ occurs twice, $\mathbf{X}_{\mathrm{L}}$ occurs three times, and $(\mathbf{U}\,\overline{\mathbf{K}}\,\mathbf{U})$ occurs once.

In the SCEP formalism single excitations are included through an *AO* matrix $\mathbf{d}$ which has only an $\mathbf{R}$ ... $\mathbf{U}$ component or, within the LSA,

$$\mathbf{d} = \mathbf{R}\mathbf{y}_{\mathrm{L}}\mathbf{U}\,. \tag{40}$$

The analogue of Eq. (26) — combined with Eq. (23) — is then

$$(\mathbf{RFU})_{ab} + \left[\mathbf{R}_{\mathrm{L}}\mathbf{y}_{\mathrm{L}}(\mathbf{UFU})_{\mathrm{L}} \;-\; (\mathbf{RFR})_{\mathrm{L}}\mathbf{y}_{\mathrm{L}}\mathbf{U}_{\mathrm{L}} - \frac{1}{2}\sum_{i,j\in\mathrm{L}}(\mathbf{R}\,\overline{\mathbf{K}}_{ij}\,\mathbf{U})_{\mathrm{L}}(\mathbf{X}_{ji})_{\mathrm{L}}\mathbf{U}_{\mathrm{L}}\right]_{ab}$$

$$- \frac{1}{2} \sum_{i,j \in \mathrm{L}} (\mathbf{RS})_{ai} \left[\mathbf{RK}(\mathbf{C}_{ij}) \mathbf{U} \right]_{jb} = 0 \quad (a, b \in \mathrm{L}) \,. \tag{41}$$

This relation is similar to the one used to determine $(\mathbf{X}_{ij})_{\mathrm{L}}$, which was obtained by substituting the expressions in (28) and (29) into Eqs. (23) and (26). Thus, the solution for $\mathbf{y}_{\mathrm{L}}$ may be found iteratively in the manner described earlier for $(\mathbf{X}_{ij})_{\mathrm{L}}$. An $(\mathbf{RFU})_{ab}$ term appears here because the KS orbitals are not exact eigenfunctions of the local space HF Hamiltonian. By including $(\mathbf{RFU})_{\mathrm{L}}$ together with the usual single excitation terms the former is implicitly being treated as a second-order perturbation (although it could easily be changed to first-order if necessary). Given $\mathbf{y}_{\mathrm{L}}$ the singles contribution to the local fourth-order energy can be evaluated as

$$\mathrm{E}^{(4)}_{\mathrm{S;L}} = \langle (\mathbf{RFR})_{\mathrm{L}} \mathbf{y}_{\mathrm{L}} \mathbf{U}_{\mathrm{L}} \mathbf{y}_{\mathrm{L}}^{\dagger} - \mathbf{y}_{\mathrm{L}}^{\dagger} \mathbf{R}_{\mathrm{L}} \mathbf{y}_{\mathrm{L}} (\mathbf{UFU})_{\mathrm{L}} \rangle \,. \tag{42}$$

The extension of the SCEP formalism to account for the computationally demanding triples contribution to the fourth-order energy has not previously been addressed either on its own or in connection with the 'local correlation' treatment. However, it is not too difficult to do so since an expression (albeit messy) for the key tensor quantity, $(\mathbf{w}_{ijk})_{pqr}$, has already been given by Krishnan et al. [37]. In analogy with Eq. (21) we write the triples coefficients in the form

$$(\mathbf{C}_{ijk})_{pqr} = \mathbf{U}_{pp'} \mathbf{U}_{qq'} \mathbf{U}_{rr'} (\mathbf{Z}_{ijk})_{p'q'r'} \tag{43}$$

using the summation convention (both here and in the remainder of this section) for repeated indices. (The tensor $\mathbf{Z}_{ijk}$ in Eq. (43) and the following should not be confused with the column vector $\mathbf{Z}(\mathbf{r}_g)$ in Sect. 2.). From the analogue of Eq. (26) one obtains the condition that determines $\mathbf{Z}_{ijk}$. Comparing with Eq. (23) we now have $(\mathbf{w}_{ijk})_{abc}$ instead of $(\mathbf{U}\,\overline{\mathbf{K}}_{ij}\,\mathbf{U})_{ab}$; expression (28) is replaced by

$$[(\mathbf{UFU})_{aa'} \mathbf{U}_{bb'} \mathbf{U}_{cc'} + \mathbf{U}_{aa'} (\mathbf{UFU})_{bb'} \mathbf{U}_{cc'} + \mathbf{U}_{aa'} \mathbf{U}_{bb'} (\mathbf{UFU})_{cc'}] \, (\mathbf{Z}_{ijk})_{a'b'c'} \tag{44}$$

and the coupling terms (cf. Eq. (29)) become

$$-\mathbf{U}_{aa'} \mathbf{U}_{bb'} \mathbf{U}_{cc'} \left[F_{il} \mathbf{Z}_{ljk} + F_{jl} \mathbf{Z}_{ilk} + F_{kl} \mathbf{Z}_{ijl} \right]_{a'b'c'} \,. \tag{45}$$

Finally, the local fourth-order energy due to triple excitations is

$$\mathrm{E}^{(4)}_{\mathrm{T;L}} = (\mathbf{w}_{ijk})_{abc} (\mathbf{Z}_{ijk})_{abc} \tag{46}$$

All indices in Eqs. (44)–(46) refer to the local space.

This completes our treatment of MP perturbation theory through fourth-order. We have shown elsewhere [6] that the SCEP approach is readily adapted for CC and CI calculations within the LSA as well. As already noted, essentially the same quantities appear in all cases. Thus, our formalism is now sufficient to include the CCSD(T) procedure, which is the current method of choice where feasible.

4 Application of Translation Symmetry for Fragments with Infinite Periodicity

If one or more of the separated fragments is infinitely periodic in 1, 2, or 3 directions the translational symmetry can lead to considerable simplification. For sake of discussion let us examine the situation where a relatively small atomic or molecular fragment (I) interacts with an infinitely periodic system. The latter fragment (II) may be divided into geometrically identical unit cells along the direction(s) of translational symmetry. Prior to interaction the AOs and all the electronic properties of these unit cells will be the same. After interaction the electronic properties will change but the AOs remain unaltered.

We concentrate on the 4-index integrals transformation from the full space to the local space, which is a major computational bottleneck in the hybrid treatment (although only half-transformations are needed in second-order perturbation theory). Along the way the simplifications for the initial KS–DFT/LSA calculation due to symmetry will be evident. One procedure for handling the two-electron integrals is the numerical prescription given in Sect. 2 along with Eqs. (12)–(16). In that event we utilize the matrices $\mathbf{A}_{\lambda\sigma}(\mathbf{r}_g)$ defined in Eq. (12). The AOs ϕ_λ and ϕ_σ may be associated with either fragment. For fragment II each orbital can be specified by the unit cell, the atomic center within that cell, and the orbital type. The location of the unit cell, in turn, requires an index (possibly compound) for the position in the direction(s) of translational symmetry and another for the remaining directions. In the following discussion we suppress all indices except the one that refers to the position of the unit cell in the direction(s) of translational symmetry and that index will be denoted by α, β, etc. For fragment (I) we associate a different cellular index with each atomic center.

The grid points, $\mathbf{r}_g$, have yet to be specified. We adopt the viewpoint [38] that any chemical system may be divided into space-filling regions as follows. First, an inner sphere is constructed for each atomic core. Then, the interstitial region associated with a particular atom is obtained by surrounding the inner sphere with a Voronoi polyhedron generated by the orthogonal planes that bisect the lines connecting that atom to all of its neighbors. At the same time a boundary surface is introduced (except for an infinite crystal) so that atoms on the periphery will be completely bounded by a polyhedron as are atoms in the interior. Finally, the outer region consists of all space external to the boundary surface. In principle, this procedure could be applied at any given geometry to both the interacting and the non-interacting fragments. However, if each non-interacting fragment is treated as if the other fragment were not present, then the polyhedra for atoms near the reactive site would generally be different for the two cases (interacting vs. non-interacting). In order to avoid this situation, and maintain translational symmetry as much as possible, we use the polyhedra of the isolated fragment II even when the interaction is switched on. For exactly

the same reasons we modify the polyhedra for the atoms of isolated fragment I if necessary so that they do not overlap with those of fragment II. On the other hand, there is no way to retain complete translational symmetry in the outer region. For that region we use the space which is external to both fragments. This means that the following translational symmetry arguments apply to all grid points associated with the atoms of fragment II but only to that portion of the external region which is unmodified by the presence of fragment I. We label the grid points of fragment II, like the AOs, using just the unit cell index in the direction(s) of translational symmetry.

The quantities $\mathbf{A}_{\alpha\beta}(\mathbf{r}_\gamma)$ where $\alpha, \beta, \gamma, \in$ II are exactly the same for both the interacting and the non-interacting fragments. Because of translational symmetry we have

$$\mathbf{A}_{\alpha\beta}(\mathbf{r}_{g\gamma}) = \mathbf{A}_{o\beta'}(\mathbf{r}_{g\gamma'}) \quad (\beta' = \beta - \alpha; \gamma' = \gamma - \alpha) \tag{47}$$

where the rhs has been obtained by carrying out a uniform translation so that cell α is at the origin, denoted by o. Thus, in terms of unit cells, $\mathbf{A}_{\alpha\beta}(\mathbf{r}_\gamma)$ is effectively a two-dimensional array. Although the array dimensions are infinite, in practice they can be truncated to a finite range in keeping with the following discussion. Assuming that the AOs are Gaussian-type functions, then $\mathbf{A}_{\alpha\beta}(\mathbf{r}_\gamma)$ will fall-off as $\exp(-\zeta|\mathbf{R}_\alpha - \mathbf{R}_\beta|^2)$ for all $\mathbf{r}_\gamma$, where $|\mathbf{R}_\alpha - \mathbf{R}_\beta|$ is the distance between the unit cells containing the two orbitals (here $\mathbf{R}$ is a position vector not to be confused with the density matrix) and ζ is an appropriate orbital exponent. This exponential decay provides a basis for discarding all α, β pairs with $|\mathbf{R}_\alpha - \mathbf{R}_\beta|$ greater than some preset value. On the other hand, if α and β refer to the same unit cell then $\mathbf{A}_{\alpha\alpha}(\mathbf{r}_\gamma)$ will fall-off as $1/|\mathbf{R}_\gamma - \mathbf{R}_\alpha|$. For α in (or near) the local space a saving feature is that $\mathbf{A}_{\alpha\alpha}(\mathbf{r}_\gamma)$ is later multiplied by a pair of elements from the local space $\mathbf{Z}_\mathrm{L}(\mathbf{r}_\gamma)$ vector, which leads to an additional fall-off factor of $1/|\mathbf{R}_\gamma - \mathbf{R}_\alpha|^6$ in a 3-dimensional metal as we will see shortly. The same factor will also occur for γ in (or near) the local space and α in the surroundings. In this case it is due to the density matrix elements that appear in Eq. (13).

The two-electron integrals involve a second AO pair which leads to the introduction of local space $\mathbf{Z}_\mathrm{L}$ vectors (cf. Eq. (16)). Let the local space orbital a in Eq. (16) belong to unit cell α in fragment II and let the grid point lie in unit cell $\gamma(\gamma \in$ II). Then, by the same reasoning that led to Eq. (47), $\mathbf{Z}^{\mathrm{II}}_\alpha(\mathbf{r}_\gamma) = \mathbf{Z}^{\mathrm{II}}_o(\mathbf{r}_{\gamma'})$ with $\gamma' = \gamma - \alpha$. (Note that $\mathbf{R}_0$ and $\mathbf{U}_0$ have no matrix elements connecting different fragments.) In terms of unit cells, this means that $\mathbf{Z}^{\mathrm{II}}_\mathrm{L}(\mathbf{r}_\gamma)$ is, operationally, a one-dimensional vector quantity. For Gaussian-type AOs the value of $\mathbf{Z}^{\mathrm{II}}_\mathrm{L}(\mathbf{r}_\gamma)$ will be dominated by contributions from orbitals in γ or its vicinity. Hence, if γ is far from the local space the value of $\mathbf{Z}^{\mathrm{II}}_\alpha(\mathbf{r}_\gamma)$ will fall-off with distance in the same manner as the density matrix element $(\mathbf{R}_0)_{\alpha\gamma}$, or $(\mathbf{U}_0)_{\alpha\gamma}$, which is $1/|\mathbf{R}_\alpha - \mathbf{R}_\gamma|^3$ in a 3-dimensional metal. This explains the $1/|\mathbf{R}_\alpha - \mathbf{R}_\gamma|^6$ decay discussed above in connection with $\mathbf{A}_{\alpha\alpha}(\mathbf{r}_\gamma)$.

In the foregoing analysis the symmetry arguments were applied separately to $\mathbf{A}_{\alpha\beta}(\mathbf{r}_\gamma)$ and $\boldsymbol{Z}_\alpha(\mathbf{r}_\gamma)$. Thus, even if only one (or two or three) of the local space orbital indices is associated with fragment II, it is possible to take some advantage of translational symmetry. On the other hand, this does not pertain to the numerical integration over the unit cells that lie outside the periodic region. If the alternative procedure of Sect. 2 is followed — i.e., the two-electron integrals are evaluated analytically at the outset and subsequently transformed to the local space — then the application of symmetry is more limited. In that event there are two scenarios. When all four local space orbitals are associated with fragment II, the translational symmetry may be utilized in exactly the same manner as it is in a conventional band structure calculation of a perfectly periodic system. Otherwise, there is no simplification due to symmetry.

5
'Proof of the Method' Calculations

Although some *ab initio* calculations using the LSA have been carried out [4, 9] most of the treatments, thus far, have employed semi-empirical Hamiltonians with the aim of proving the efficacy of the method. One key question is the convergence behavior with respect to increasing the size of the local space, particularly in delocalized systems such as a chemisorbed atom or molecule on a metal surface. For the case of atomic hydrogen adsorbed on tungsten, a comparison with the finite cluster approach has been made using an Anderson–Newns Hamiltonian. Two important points have emerged from this study [1]. First, the chemisorption binding energy (BE) calculated by the LSA method approaches the correct value monotonically when the local space is systematically increased in size. This may be contrasted with the finite cluster result which shows large oscillations even when hundreds of metal atoms are included. These oscillations, which are now recognized as a common phenomenon, cannot generally be prevented [39] by any known scheme. The second point is that the initial convergence in the LSA treatment is rapid. When we include in the local space just those metal AOs that interact directly with the adsorbate the error in the BE is 13.9%. This is reduced to 9.7%, 6.8%, 5.5%, 4.6%, ... as successive shells of metal atom nearest neighbors are added. The monotonic behavior suggests that extrapolation procedures can be developed to improve accuracy. No hybrid LSA calculations were attempted since the theory was not available at the time.

In a related vein, it is of interest to apply the LSA to a case where the local interaction occurs *within* the delocalized system rather than at a surface. One example is an impurity embedded in a 3-dimensional metal. Another example, that we have examined [40], is the cleavage of a (formal) C–C single bond in a π-conjugated polyacetylene chain. Using an INDO Hamiltonian it was found that a local space containing only one or, at most, two carbon π AOs on either side of the broken bond is sufficient to completely restore the π-conjugation.

Along the carbon backbone the addition of σ AOs on the originally bonded atoms plus those on the nearest and next-nearest neighbors leads to a BE that is accurate to well within 1%.

The ability of the LSA to account for long-range polarization was tested [41] by means of a CNDO/2 calculation on the α-glycylglycine Zwitterion dimer. In this example the dimer is formed through an ion-pair hydrogen bond ('salt bridge') connecting the monomers. The large BE ($\sim$ 50 kcal/mol) of the salt bridge is strongly influenced by the equal and opposite ionic charges on the end groups. Indeed, all cluster models (which always omit the end groups) yield a BE roughly a factor of 2 too large. In comparison, using a local space consisting of just the AOs directly involved in the N–H...O hydrogen bond, the LSA treatment gives about the same 5% accuracy that is obtained for an ordinary weak (BE $\sim$ 5 kcal/mol) hydrogen bond in a species without charge separation. Thus, the long-range polarization is automatically taken into account when starting with the separated fragments.

The semiempirical results described above show considerable promise. However, the LSA methodology has yet to be fully implemented in an ab initio framework. Those ab initio calculations that have been done were carried out on small well-localized systems. For such cases the HF results [11] confirm our conclusions drawn from semiempirical treatments. In the one instance where electron correlation was considered [6], the error in the LSA pair correlation energy was less than 0.1 kcal/mol at the equilibrium geometry. It is clear that further tests are warranted, particularly for delocalized systems with correlation included either at the KS–DFT level throughout or by means of the hybrid treatment given in Sect. 3.

Acknowledgements. I would like to express my appreciation to several major participants in developing the LSA methodology. Celso de Melo was involved from the outset and, along with Maria Matos, was instrumental in the application to chemisorption and polymeric systems. Clifford Dykstra made the original suggestion that the Hartree–Fock version might be extended to include correlation by means of the SCEP approach. Kathleen Robins wrote the ab initio Hartree–Fock computer code for well-localized systems while Hideo Sekino stimulated the introduction of pseudospectral techniques as well as the new treatment of overlap between fragments. In the course of preparing this chapter, I had numerous fruitful discussions with William Palke, Benoit Champagne, and Stan van Gisbergen.

6 References

1. Matos M, Kirtman B, DeMelo CP (1988) J Chem Phys 88:1019
2. McWeeny R (1960) Rev Mod Phys 32:335
3. Li XP, Nunes RW, Vanderbilt D (1993) Phys Rev B 47:10891
4. Daw MS (1993) Phys Rev B 47:10895
5. Ochsenfeld C, Head-Gordon M (1997) Chem Phys Lett 270:399
6. Kirtman B, Dykstra CE (1986) J Chem Phys 85:2791
7. Kirtman B, DeMelo CP (1986) Int J Quantum Chem 29:1209
8. Kirtman B, DeMelo CP (1981) J Chem Phys 75:4592

9. Kirtman B (1982) J Chem Phys 86:1059
10. Dykstra CE, Kirtman B (1990) Annu Rev Phys Chem 41:155
11. Robins KA, Kirtman B (1993) J Chem Phys 99:6777
12. Friesner R (1985) Chem Phys Lett 116:39
13. Ringnalda MN, Belhadj M, Friesner RA (1990) J Chem Phys 93:3397
14. Feyereisen MW, Fitzgerald G, Korminicki A (1993) Chem Phys Lett 208:359
15. Vahtras O, Almlof J, Feyereisen MW (1993) Chem Phys Lett 213:514
16. Bernholdt DM, Harrison RJ (1996) Chem Phys Lett 250:477
17. Kirtman B, Sekino H (1996) Chem Phys Lett 263:313
18. Sambe H, Felton RH (1973) J Chem Phys 62:1122
19. Kirtman B (1995) Int J Quantum Chem 55:103
20. Foster JP, Weinhold F (1980) J Am Chem Soc 102:7211
21. Roby KR (1974) Mol Phys 28:1441
22. Meyer W (1976) J Chem Phys 64:2901
23. Ahlrichs R (1979) Comp Phys Commun 17:31
24. Dykstra CE, Schaefer HF, Meyer W (1976) J Chem Phys 65:2740
25. Dykstra CE, Chiles RA, Garrett MD (1981) J Comput Chem 2:266
26. Jasien PG, Dykstra CE (1982) J Chem Phys 76:4564
27. Chiles RA, Dykstra CE (1981) J Chem Phys 74:4544
28. Pulay P (1983) Chem Phys Lett 100:151
29. Saebo S, Pulay P (1985) Chem Phys Lett 113:13
30. Pulay P, Saebo S (1986) Theor Chim Acta 69:357
31. Saebo S, Pulay P (1987) J Chem Phys 86:914
32. Saebo S, Pulay P (1988) J Chem Phys 88:1884
33. Saebo S (1990) Int J Quantum Chem 38:641
34. Saebo S (1992) Int J Quantum Chem 42:217
35. Saebo S, Pulay P (1993) Annu Rev Phys Chem 44:213
36. Pulay P, Saebo S, Meyer W (1984) J Chem Phys 81:1901
37. Krishnan R, Frisch MJ, Pople JA (1980) J Chem Phys 72:4244
38. See, for example, te Velde G, Baerends EJ (1992) J Comput Phys 99:84
39. Matos M, Kirtman B (1995) Surf Sci 341:162
40. Kirtman B, DeMelo CP (1987) J Chem Phys 86:1624
41. Kirtman B, DeMelo CP (1986) Int J Quantum Chem 29:1209

Local Electron Densities and Functional Groups in Quantum Chemistry

Paul G. Mezey[1, 2]

[1] Institute for Advanced Study, Collegium Budapest, Szentháromság u. 2, H-1014 Budapest, Hungary
[2] Mathematical Chemistry Research Unit, Department of Chemistry and Department of Mathematics and Statistics, University of Saskatchewan, 110 Science Place, Saskatoon, Canada S7N 5C9 *E-mail: mezey@sask.usask.ca*

The connections between the quantum chemical concept of local electron densities and the traditional concept of functional groups of organic chemistry are explored. The Density Domain criterion of functional groups and the Additive Fuzzy Density Fragmentation (AFDF) method are the tools for the construction of representative quantum chemical models for both local electron densities and functional groups. The applications of the method of fuzzy electron density partitioning of molecular electron density clouds to the description of the local fuzzy density contributions of functional groups are explored. The constraints implied by the recently proven "Holographic Electron Density Theorem" are discussed, with special emphasis on the predictability of differences in the activity of functional groups of a given type located within different molecular environments. The special role of symmetry and symmetry deficiency are discussed, and the deviations from local symmetry, as a diagnostic tool for the assessment of the reactivity of functional groups are explored.

Keywords: Quantum chemical localization, Subsystems of molecules, Fuzzy density fragmentation, Holographic density theorem, Symmetry deficiency

Topics in Current Chemistry, Vol. 203

1
Introduction and Conceptual Framework

Molecular electron densities determine molecular properties. A formal proof of this intuitively evident principle was provided by the Hohenberg–Kohn theorem [1], establishing that a ground state, nondegenerate electron density uniquely determines the energy, as well as all other properties of the molecule [2–5]. However, local electron densities, their representation, interpretation, and chemical role within the general framework of quantum chemical localization [6–13], represent an important challenge for both theoretical and empirical chemistry, and in this contribution some of these problems and potential answers will be reviewed, with special emphasis on chemical functional groups.

The quantum chemical concept of local electron densities and the traditional organic chemistry concept of functional groups have much in common.

Functional groups are among the universally accepted and frequently used conceptual tools of organic chemists, invoked in most mechanistic models of organic reactions, synthesis planning, and in the interpretations and explanations of biochemical processes. By contrast, the concept of local, fuzzy electron density clouds, as a potential tool for interpreting static and dynamic molecular properties, has been only recently employed in a systematic way within the framework of quantum chemistry. In fact, functional groups and local electron densities have analogous roles in the description of molecular properties, and the increasing recognition of this connection and its actual level of utilization by chemists is a reflection of the accelerating pace at which rigorous theoretical chemistry tools, such as quantum chemical electron density clouds, are becoming a routine component of the conceptual tool-box of chemists. However, some contrast between the interpretations of traditional and quantum chemical functional group concepts remain, and the purpose of this chapter is to point out the essential connections between the empirical model and the modern quantum chemical approaches.

This contrast is in part due to the difficulty of invoking fuzzy three-dimensional electron density clouds in a mechanistic interpretation of reactions, which is also hindered by the perceived classical nature of the concept of localization. Classical objects, often used in analogies when modeling molecular structures, usually exhibit localized features, yet the very concept of localization apparently conflicts with the delocalized nature of molecular electron distribution and the Heisenberg uncertainty relation.

Within the quantum chemical description of molecular electron density clouds, a natural criterion, the Density Domain criterion, provides a quantum chemical definition for functional groups [14–18]. Furthermore, techniques that generate fuzzy electron density contributions for local molecular moieties that are analogous to the fuzzy electron density clouds of complete molecules, determined by the analytic Additive Fuzzy Density Fragmentation (AFDF) method [19–21], or the earlier numerical-grid MEDLA method [22, 23], are also

applicable for the computation of local electron density clouds suitable to represent functional groups within each molecule. Hence, functional groups can be described and analyzed using essentially the same tools which are employed in the electron density analysis and shape analysis of complete molecules.

1.1 The Quantum Chemical Concept of Localization

The chemical concept of localization is based on the classical assumption that molecular electron densities are confined to finite regions of space. Although various quantum chemical approaches have been proposed to implement this concept within a theoretical framework, this is essentially a classical idea that disregards some aspects of the Heisenberg uncertainty relation. Nevertheless, the interpretation of molecular properties and chemical reactions strongly relies on the inherent assumption that both complete molecules and various, chemically identifiable, molecular pieces can be assigned to various regions of space, at least in some approximate sense. The explanation of the apparent success of this essentially classical assumption within a quantum chemical framework is a nontrivial problem.

If the model is confined to nonrelativistic quantum mechanics, then the consequences of the Heisenberg relation are not incompatible with a simple, pictorial interpretation of quantum mechanical localization. A simple localization scheme involves a series of monotonically shrinking volumes, contracting to a single point of the space. As the position information becomes less uncertain, eventually leading to complete localization, any such process implies a gradual and eventually complete loss of momentum information. Nevertheless, the idea of localization by this scheme, however unlikely as a practical proposition, is still meaningful within nonrelativistic quantum mechanics.

More severe conceptual problems arise in relativistic quantum mechanics, where the analogous scheme leads to completely non-physical results. In the process leading to formal localization and to complete uncertainty of momentum information, the information on the reference frame itself is lost, consequently, one can no longer determine what events can be regarded as simultaneous. Consequently, a conventional, classically interpreted localization approach within a relativistic quantum mechanical framework leads to nonsensical results. For example, if one assumes that localization is possible [6] and, as an initial condition, a relativistic particle is forced to be fully localized at some time $t = 0$, then the complete uncertainty of the reference frame implies that the same particle is already spread over the whole space at any later time $t > 0$. Within such a model causality is lost, and the relativistic model itself becomes self-contradictory [6].

Nevertheless, in an approximate sense, localization is a useful concept for the interpretation of molecular properties and the dynamic aspects of molecular interactions. Since atomic nuclei within molecules dominate electron den-

sity distributions, and since rather reliable approximate models of nuclear behavior in molecules can be obtained using models involving some degree of nuclear localization, the influence of localized nuclear arrangements on the primarily delocalized molecular electron density cloud also provides a basis for approximate localization schemes of molecular electron distributions.

Within the approximate molecular orbital framework several useful localization schemes have been introduced for the determination of localized molecular orbitals. These localized orbitals have been assumed to model some of the earlier, semi-empirical concepts of bonding electron pairs, lone pairs, and other local molecular regions assumed to have a dominant role in some chemical reactions [7, 8, 9, 11, 12]. These approaches include an orbital localization technique that can be easily adapted to the reverse process of delocalization [11, 12]. This latter method has also been proposed for the determination of the "most delocalized" molecular orbitals [11, 12, 13], reflecting some of the long-range properties of electron distributions within molecules. Most of these methods are options within some of the current quantum chemistry program packages, such as the *GAMESS Quantum Chemistry Program Package* [13].

1.2 Subsystems of Molecular Systems

Subsystems of electron density distributions of molecules can be obtained by a variety of methods; here we shall be concerned with two, fundamentally different approaches:

(i) subdivision of the space into various compartments d_i, resulting in molecular fragments of sharp boundaries;
(ii) a fuzzy decomposition of the molecular electron density into fragments of no boundaries, yielding fuzzy fragment densities that are not only reminiscent of the electron densities of complete molecules but also can be superimposed on one another, resulting in an exact reconstruction of the original electron density of the complete molecule.

The former approach is subject to the constraints imposed by the recently proven "Holographic Electron Density Theorem" [24, 25]:

> *"Any nonzero volume piece $\rho_d(\mathbf{r})$ of the nondegenerate ground state electron density fully determines the ground state electron density $\rho(\mathbf{r})$ of the entire, boundaryless molecular system."*

Note that the holographic electron density theorem has been proven for boundaryless electron densities, where the boundaryless, fuzzy nature of the complete electron density is an essential feature if realistic molecular representations are to be considered [24, 25]. Note that an earlier result on subsystems used models where both the subsystem and the complete system were assumed

to be finite and bounded [10], that is, confined to finite, bounded domains of the space, a condition that is not valid for real molecular electron densities.

According to the fundamental Hohenberg–Kohn theorem of density functionals [1], the ground state energy E, and the ground state wavefunction Ψ (hence, essentially all ground state properties of the molecule) are uniquely determined by the nondegenerate ground state electron density $\rho(\mathbf{r})$ [2–5]. Since any nonzero volume piece $\rho_d(\mathbf{r})$ determines the complete density, which in turn determines all molecular properties, a combination of the holographic electron density theorem with the Hohenberg–Kohn theorem provides a statement stronger than the original Hohenberg–Kohn theorem. This stronger result can be stated as the "Local Density Information Completeness Theorem" [26]:

> *"Any nonzero volume piece $\rho_d(\mathbf{r})$ of the nondegenerate ground state electron density fully determines the ground state energy E, the ground state wavefunction Ψ (up to a phase factor), and the expectation values of all spin-free operators defined by the ground state wavefunction Ψ".*

Take now any local, nonzero volume domain c of arbitrary but fixed shape and size within the ordinary, three-dimensional space, and refer to it as the standard domain c. For example, take c as the unit cube, using a unit length convenient for molecular size. Consider all possible ground state, nondegenerate local electron densities $\rho_c(\mathbf{r})$ within this domain c. The Local Density Information Completeness Theorem [26] implies that the energy E of complete molecules is a unique functional of the local electron density $\rho_c(\mathbf{r})$ within the standard domain c:

$$E = E(\rho_c)\,. \tag{1}$$

This result, known as the *Local Density Functional Energy Theorem Over Standard Domains* [26], is a statement stronger than the Hohenberg–Kohn theorem, implying that the molecular energy is a unique functional not only of the complete ground state electron density, but also of the electron density within any local, nonzero volume standard domain of arbitrary but fixed shape and size. Of course, this theorem contains the original Hohenberg–Kohn theorem as a special case, if the standard domain c is chosen as the full space, $c = E^3$.

Energy is not the only property that is so determined by the electron density fragment ρ_c. Since the (non-degenerate, ground state) local electron density $\rho_c(\mathbf{r})$ in any standard domain c fully determines the complete density $\rho(\mathbf{r})$, which in turn fully determines the molecular wavefunction Ψ (up to a phase factor), all molecular properties P which can be expressed as expectation values of spin-free operators defined by the ground state wavefunction Ψ are also determined by the local electron density $\rho_c(\mathbf{r})$ in the standard domain c. Consequently, any such property P is also a unique functional of the local electron density $\rho_c(\mathbf{r})$ within the standard domain c:

$$P = P(\rho_c)\,. \tag{2}$$

This result, known as the *Local Density Functional Property Theorem Over Standard Domains* [26], is the fundamental statement on the role of local molecular domains in determining the properties of complete molecules.

These results concern the theoretical basis of chemistry. However, they are also of relevance to the application of local shape analysis and to the subsequent use of numerical local shape descriptors in correlations with various, not-well-understood experimental results, such as biochemical activities in complex biochemical systems, potency data in pharmaceutical drug design approaches, experimental toxicities in toxicological risk assessments, etc. [27–36].

If fuzzy decompositions of molecular electron densities are considered, then the fragments themselves are fuzzy objects with no boundaries; far away from the nearest nucleus, the density value for each fragment density converges to zero exponentially with distance. The principles of fuzzy electron density fragmentation schemes are reviewed in Sect. 2. Fuzzy fragment densities can be studied using level sets, that is, domains of the space enclosed by isodensity contours. Such domains, called density domains, provide a theoretical basis for the study of functional groups. Density domains have been discussed from the perspective of molecular similarity measures in an earlier study [19], and in Sect. 3 only a brief summary will be given, followed by a discussion of density domain properties relevant to functional group analysis.

2
Local Electron Densities, the Additive Fuzzy Density Fragmentation (AFDF) Principle, and the Adjustable Density Matrix Assembler (ADMA) Method

If one follows approach (ii) to the study of subsystems of molecules, as outlined in Sect. 1.2, then it is natural to consider subsystems having properties analogous to complete molecules. In fact, it is possible to define molecular subsystems with electron densities whose convergence properties are analogous to those of complete molecules.

2.1
Representation of Local Electron Densities: the Additive Fuzzy Density Fragmentation (AFDF) Principle

Consider an arbitrary classification of the nuclei of the molecule into various families

$$f_1, f_2, \ldots, f_k, \ldots, f_m . \tag{3}$$

Whereas this classification of nuclei into families is arbitrary, the fuzzy electron density fragments obtained are simpler to interpret and identify with chemically relevant local molecular regions if the nuclei within each family f_k are all those which fall within some local domain of the three-dimensional space.

A general, additive, fuzzy density fragmentation scheme (AFDF scheme) has been described in [19], as outlined below.

Consider a small molecule in some fixed nuclear configuration K and the corresponding *ab initio* LCAO molecular wavefunction Ψ expressed in terms of a basis set of n atomic orbitals $\varphi_i(\mathbf{r})$ $(i = 1, 2, \ldots, n)$, where $\mathbf{r}$ is the three-dimensional position vector variable. Consider the corresponding $n \times n$ density matrix $\mathbf{P}$, the electron density $\rho(\mathbf{r})$ expressed in terms of the elements P_{ij} of this density matrix $\mathbf{P}$, and the atomic orbitals $\varphi_i(\mathbf{r})$:

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \,. \tag{4}$$

For each nuclear family f_k an additive fuzzy electron density fragment $\rho^k(\mathbf{r})$ can be defined by a general Additive Fuzzy Density Fragmentation (AFDF) scheme proposed earlier [19]. According to this scheme a fragment density matrix $\mathbf{P}^k$ of dimension $n \times n$ is defined in terms of its matrix elements P_{ij}^k as

$$\begin{aligned} P_{ij}^k &= P_{ij} \text{ if both atomic orbitals } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ are centered on nuclei of} \\ &\quad \text{the } k\text{-th nuclear family,} \\ &= f(k, i, j) P_{ij} \text{ if precisely one of atomic orbitals } \varphi_i(\mathbf{r}) \text{ and } \varphi_j(\mathbf{r}) \text{ is cen-} \\ &\quad \text{tered on a nucleus of the } k\text{-th nuclear family,} \\ &\quad \text{where } f(k, i, j) > 0,\ f(k, i, j) + f(k', i, j) = 1, \text{ and where} \\ &\quad \text{the } k'\text{-th nuclear family contains the other nucleus,} \\ &= 0 \text{ otherwise} \,. \end{aligned} \tag{5}$$

The factors $f(k, i, j)$ can be chosen in a variety of ways; for example, take a scalar property $A(i)$ that can be assigned to atomic orbitals and does not change sign (formal electronegativity is one such possible choice). With respect to the scalar property $A(i)$ a suitable function $f(k, i, j)$ is defined by

$$f(k, i, j) = A(i)/[A(i) + A(j)] \,, \tag{6}$$

where the atomic orbital $\varphi_i(\mathbf{r})$ is assumed to be centered on a nucleus that belongs to the nuclear family of index k.

The k-th fuzzy, fragment electron density is defined in terms of the fragment density matrix $\mathbf{P}^k$, as

$$\rho^k(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}^k \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \,. \tag{7}$$

The general AFDF scheme of Eq. (5) ensures that the sum of the fragment density matrices P_{ij}^k is equal to the density matrix of the molecule, consequently, one can easily show that the sum of the fragment densities $\rho^k(\mathbf{r})$ is equal to the density $\rho(\mathbf{r})$ of Eq. (4) of the molecule:

$$P_{ij} = \sum_{k=1}^{m} P_{ij}^k \tag{8}$$

and

$$\rho(\mathbf{r}) = \sum_{k=1}^{m} \rho^k(\mathbf{r})\,. \tag{9}$$

The earliest and simplest implementation of the general scheme of Eq. (5) has involved the choice of

$$f(k,i,j) = 0.5 \quad \text{for all indices } k,i,j\,, \tag{10}$$

which in fact corresponds to a scheme analogous to Mulliken's population analysis scheme without integration. This scheme has been used in the MEDLA numerical approach to the construction of macromolecular electron densities, including *ab initio* quality electron densities for the proteins crambin, bovine insulin, the gene-5 protein (g5p) of bacteriophage M13, the HIV-1 protease monomer of 1564 atoms, the proto-oncogene tyrosine kinese protein 1ABL containing 873 atoms, as well as some important drug molecules, including the anti-cancer drug taxol [22, 23, 37–39]. One should note that the usual criticism of Mulliken's scheme for the calculation of formal atomic charges stems from the artificial nature of the very concept of atomic charge within a molecule, and the present scheme as applied to electron densities without integration provides a valid, fully additive, fuzzy density fragmentation scheme.

The AFDF local molecular pieces can be used to build high quality approximate electron densities for large molecules, and also to study the local molecular subsystems themselves.

2.2
A Local Density Computational Method based on the Adjustable Density Matrix Assembler (ADMA) Approach

Whereas the first applications of the general AFDF scheme involved the numerical representation and numerical superposition of local electron densities on a rectangular grid [22, 23, 37–39], a more advanced approach involves the analytical construction of density matrices of larger, composite molecular fragments and of macromolecular density matrices. This approach, the Adjustable Density Matrix Assembler (ADMA) method [20, 21], is proposed for the computation of local and intermediate range electron densities within large molecules, if the spatial distribution of electron density above some chemically significant density threshold, say, 0.001 a.u. (atomic unit), is found in an extensive volume of the space. Typical examples of such extensive molecular pieces with chemically significant roles are large segments of protein backbones, or collections of interconnected functional groups in segments of long chain polymers. The main features of the ADMA method are outlined below.

The macromolecular density matrix $\mathbf{P}(\varphi(K))$ of the large molecule M (or large molecular fragment M) of nuclear configuration K is assembled from local, fragment density matrices $\mathbf{P}^k(\varphi(K_k))$ obtained from high quality *ab in-*

itio calculations on small parent molecules M_k of nuclear configurations K_k. Each fragment density matrix is obtained by applying the AFDF fragmentation method for the small parent molecules M_k, where the local surroundings of the actual nuclear family f_k of the k-th fragment in the parent molecule M_k is identical to the local surroundings of the same set f_k of nuclei in the large target molecule M. The presence of these identical "coordination shells" around the nuclei of each fragment used in the construction of the macromolecular density matrix $\mathbf{P}(\varphi(K))$ of large molecule M ensures that the local interactions are properly described by the technique. In other words, each parent molecule M_k contains several additional nuclei, beyond those required for the given fragment associated with the nuclear family f_k, and their presence and the presence of the associated electron density contributions ensure that the k-th fuzzy electron density fragment includes the relevant share of electron density interactions with its local surroundings. The size of the "coordination shell" exactly reproduced in the parent molecules M_k is adjustable; usually, a shell of 5.0 Å thickness appears sufficient.

The following, mutual compatibility conditions simplify the construction of the macromolecular density matrix $\mathbf{P}(\varphi(K))$:

(a) The local coordinate systems of the AO basis sets of all the fragment density matrices $\mathbf{P}^k(\varphi(K_k))$ have axes that are parallel and have matching orientations with the axes of the reference coordinate system of the macromolecule M. This can always be ensured by a simple similarity transformation of the fragment density matrix $\mathbf{P}^k(\varphi(K_k))$, using an orthogonal transformation $\mathbf{T}^{(k)}$ of the AO sets.

(b) Each parent molecule M_k may contain only complete nuclear families from the sets of nuclear families $f_1, f_2, \ldots, f_k, \ldots, f_m$, of the target macromolecule M. Some of these families are those surrounding the family f_k of nuclei representing the "anchor points" of the actual density fragment. In addition, each parent molecule M_k may contain a few more nuclei at their peripheral regions in order to avoid "dangling bonds".

In order to identify the index assignment of various local density matrix elements $P^k_{ij}(\varphi(K_k))$ in the macromolecular density matrix $\mathbf{P}(\varphi(K))$, the following notations are introduced.

The number of atomic orbitals in the nuclear family f_k of the target macromolecule M is denoted by n_k. For each pair $(f_k, f_{k'})$ of nuclear families a quantity $c_{k'k}$ is defined as

$$c_{k'k} = \begin{cases} 1, \text{ if nuclear family } f_{k'} \text{ is present in parent molecule } M_k \\ 0 \text{ otherwise}. \end{cases} \tag{11}$$

A given atomic orbital belongs to several lists of orbitals, depending on whether its contribution to a parent molecule or to the target macromolecule is considered. Accordingly, the serial number of a given atomic orbital $\varphi(\mathbf{r})$ is denoted by a, or i, or x, depending on the actual reference to a given collection

of atomic orbitals, where each collection actually contains the atomic orbital $\varphi(\mathbf{r})$. These indices a, or i, or x are used in the notations

$$\varphi_{a,k'}(\mathbf{r}), \quad \varphi_i^k(\mathbf{r}), \quad \text{and} \quad \varphi_x(\mathbf{r}) \tag{12}$$

with reference to orbital sets

$$\left\{\varphi_{a,k'}(\mathbf{r})\right\}_{a=1}^{n_{k'}} \tag{13}$$

$$\left\{\varphi_i^k(\mathbf{r})\right\}_{i=1}^{n_{P^k}} \tag{14}$$

and

$$\left\{\varphi_x(\mathbf{r})\right\}_{x=1}^{n} \tag{15}$$

respectively. Here $n_{k'}$ is the total number of orbitals within the set (13) of atomic orbitals associated with the nuclear family $f_{k'}$, n_{P^k} is the total number of atomic orbitals in set (14) involved in the k-th fragment density matrix $\mathbf{P}^k(\varphi(K_k))$,

$$n_{P^k} = \sum_{k'=1}^{m} c_{k'k} n_{k'}\,, \tag{16}$$

and n is the total number of atomic orbitals in set (15) involved in the density matrix $\mathbf{P}(K)$ of the target macromolecule M.

Index x for each AO $\varphi_{a,k'}(\mathbf{r}) = \varphi_i^k(\mathbf{r}) = \varphi_x(\mathbf{r})$ is determined by the serial index a in the AO set of nuclear family $f_{k'}$ as follows:

$$x = x_f(k', a) = a + \sum_{b=1}^{k'-1} n_b\,. \tag{17}$$

In this notation the subscript f in $x_f(k', a)$ indicates that indices k' and a refer to a family of nuclei (in the actual case, to family $f_{k'}$).

Three additional quantities are introduced for each index pair k and i, with respect to all nuclear families $f_{k''}$ for which $c_{k''k} \neq 0$:

$$a'_k(k'', i) = i - \sum_{b=1}^{k''} n_b c_{bk}\,, \tag{18}$$

$$k' = k'(i, k) = \min\{k'' : a'_k(k'', i) \leq 0\}\,, \tag{19}$$

and

$$a_k(i) = a'_k(k', i) + n_{k'}\,. \tag{20}$$

Using these auxiliary quantities the index $x = x_P(k, i)$ of an AO basis function in the density matrix $\mathbf{P}(K)$ of target molecule M can be computed from the row (or column) index i and serial index k of fragment density matrix $\mathbf{P}^k(\varphi(K_k))$. The index $x = x_P(k, i)$ depends on indices i and k and can be expressed using index k' and the function $x_f(k', a)$

$$x = x_P(k, i) = x_f(k', a_k(i))\,. \tag{21}$$

The subscript P in the index function $x_P(k, i)$ indicates that indices k and i refer to the fragment density matrix $\mathbf{P}^k(\varphi(K_k))$.

The actual index calculations can be restricted to the nonzero elements of each fragment density matrix $\mathbf{P}^k(\varphi(K_k))$, and the macromolecular (or large fragment) density matrix $\mathbf{P}(K)$ can be assembled iteratively:

$$P_{x_P(k,i),x_P(k,j)}(K) \Leftarrow P_{x_P(k,i),x_P(k,j)}(K) + P^k_{ij}(K_k)\,. \tag{22}$$

Usually, the parent molecules M_k are confined to some limited size that allows rapid determination of the parent molecule density matrices within a conventional *ab initio* Hartree–Fock–Roothaan–Hall scheme, followed by the determination of the fragment density matrices and the assembly of the macromolecular density matrix using the method described above. The entire iterative procedure depends linearly on the number of fragments, that is, on the size of the target macromolecule M. When compared to the conventional *ab initio* type methods of computer time requirements growing with the third or fourth power of the number of electrons, the linear scaling property of the ADMA method is advantageous.

3
Functional Groups and the Density Domain Criterion

The concept of density domains can be introduced in the context of isodensity contour surfaces. A molecular isodensity contour surface, MIDCO $G(K,a)$, of nuclear configuration K and density threshold a is defined as the collection of all points $\mathbf{r}$ of the 3D space where the electronic density is equal to the threshold value a. In the notation for electron density it is useful to specify the nuclear configuration K of the molecule and in the forthcoming discussion the notation $\rho(K,\mathbf{r})$ will be used for the electron density of nuclear configuration K. Accordingly, the MIDCO $G(K,a)$ is defined as

$$G(K,a) = \{\mathbf{r} : \rho(K,\mathbf{r}) = a\}\,. \tag{23}$$

We shall assume that the molecular electronic density $\rho(K,\mathbf{r})$ is a continuous function of the position variable $\mathbf{r}$, implying that the set of all points $\mathbf{r}$ fulfilling the defining equation (23) for any given MIDCO $G(K,a)$ forms a set of continuous closed surfaces. In the simplest case, a MIDCO $G(K,a)$ is a single closed surface.

The density domain $\mathrm{DD}(K,a)$ associated with the MIDCO $G(K,a)$ is the point set that includes all the points $\mathbf{r}$ of the corresponding MIDCO $G(K,a)$ of the same density threshold a, and all points $\mathbf{r}$ enclosed by $G(K,a)$:

$$\mathrm{DD}(K,a) = \{\mathbf{r} : \rho(K,\mathbf{r}) \geq a\}\,. \tag{24}$$

The shape and size of a MIDCO $G(K,a)$ and the associated density domain $\mathrm{DD}(K,a)$ depend on the choice of the threshold value a and on the nuclear configuration K of the molecule.

Consider a subfamily f_k of nuclei of a molecule M of nuclear configuration K with the following property: there exists a density threshold a such that a corresponding density domain $\mathrm{DD}(K,a)$ is connected and contains the family

f_k of nuclei and no other nuclei of the molecule. If this condition is fulfilled, then these nuclei form the nuclear family f_k of a quantum chemical functional group $F_k(K)$. With respect to the given atomic orbital basis set and *ab initio* wavefunction representation, the functional group $F_k(K)$ of molecule M of nuclear configuration K is composed from the nuclear set f_k and the associated fuzzy electron density fragment $\rho^k(K, \mathbf{r})$ that can be computed by the AFDF method:

$$F_k(K) = f_k \cup \rho^k(K, \mathbf{r}) . \tag{25}$$

This definition is consistent with the following general property of individual molecules: a molecule placed in the vicinity of other molecules preserves its separate identity as long as there exists a MIDCO $G(K,a)$ of some density threshold a that separates the nuclei of this molecule from the nuclei of the other molecules. In the case of functional group $F_k(K)$ of molecule M, the existence of a density domain $DD(K,a)$ and the associated MIDCO $G(K,a)$ that separates the nuclei f_k of the functional group $F_k(K)$ from all other nuclei of the molecule M also indicates that this fuzzy piece $F_k(K)$ of the molecule M also has some degree of limited autonomy and separate identity. This separate identity is what qualifies the functional group $F_k(K)$ as a significant and possibly chemically characteristic fuzzy part of the molecule.

In general, formal molecular fragments that are larger than the conventional functional groups can also be represented by fuzzy moieties of electron densities, dominated by several nuclei. The shapes of molecular fragments with density domains indicating separate identity have important chemical consequences, and these shapes can be characterized by topological shape analysis methods [40].

Besides a quantum-chemically motivated definition of functional groups, density domains also provide a detailed description of chemical bonding, based on the interfacing of fuzzy electron density regions. In this description of chemical bonding some of the topological features of molecular electron densities play a prominent role. In particular, there are only a finite number of topologically different density domains within each molecule, a fact that provides a natural, topological characterization of molecular electron densities. These topological sequences form the basis of an algebraic characterization in terms of homology and homotopy groups, describing the essential features of molecular shapes.

According to the electron density threshold a, there are several typical shapes of density domains, falling within various electron density ranges. A list of these ranges and subranges is given below, starting with high density thresholds.

Atomic range: Only individual nuclear neighborhoods appear as disconnected density domains with precisely one nucleus within each density domain.
Strictly atomic range: All the atomic density domains are convex sets.

Prebonding range: At least one density domain is no longer convex as it takes the shape of a breaking droplet, which eventually joins a neighboring density domain.

Functional group range or ***Bonding range:*** Some nuclear neighborhoods are connected and form density domains containing two or more nuclei; however, not all nuclei of the molecule are contained within a common density domain.

Molecular density range: All nuclei of the molecule are found within a common density domain, establishing the essential molecular pattern of bonding.

Skinny molecular range: There exists at least one multiply connected set on the surface of the density domain DD(K,a) along which the density domain is not locally convex (such as a formal "neck region").

Corpulent molecular range: No locally nonconvex multiply connected set on the surface of the density domain (no neck region) occurs, but there exists at least one local nonconvex region along the surface of the density domain DD(K,a).

Quasi-spherical molecular range: All density domains are convex.

Collectively, the atomic and the functional group ranges form the *localized range*, whereas the molecular density range is the *global density range*.

Among the various density domains those defining functional groups are of special significance. In the following section one application of these special density domains DD(K,a) will be discussed.

4
The Role of the Holographic Electron Density Theorem and the Predictability of Differences in the Reactivities of Functional Groups

It is a natural expectation that the local, fuzzy electron density contributions of functional groups, influenced by the interactions with other parts of the molecule, should provide the clues to their reactivity. A given type of functional group, for example, a carbonyl group, shows different degrees and even different types of reactivities depending on the rest of the molecule. The influence of the interactions between a given functional group and the local surroundings or even remote surroundings has been long used by chemists to explain the differences in reactivities and to plan strategies to achieve a given synthetic goal.

For typical functional group electron densities, most of the corresponding electronic charge is enclosed by those density domains DD(K,a) that provide the very criterion for classifying the given fuzzy molecular piece as a functional group. This feature provides a natural connection to the application of the holographic electron density theorem to functional groups: the nonzero volume domain d in the formulation of the theorem can be taken as one of these density domains DD(K,a). Since a large contribution from the fuzzy electron

density associated with the functional group F is found within these density domains, one may expect that the differences in functional groups induced by their surroundings are well detectable and are clearly reflected in the electron densities enclosed by these density domains.

Based on the holographic electron density theorem one can reformulate and quantify the influence of surroundings on functional groups. Consider two molecules A and B, both containing a given type of functional group F. Although the chemical composition of the functional group F, in the limited sense as indicated by the chemical formula, is the same in the two molecules A and B, they nevertheless, have different reactivities as a consequence of the influence of their differing surroundings. Accordingly, the two groups are distinguished as F_A and F_B, indicating that they belong to the two different molecules A and B. The corresponding AFDF density contributions of these functional groups to the total electron densities $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$ of the two molecules are denoted by $\rho_{F_A}(\mathbf{r})$ and $\rho_{F_B}(\mathbf{r})$, respectively.

Though the two functional groups are formally the same, recognized as the functional group F, the two versions, F_A and F_B, are found in the two different molecules. Consequently, it is reasonable to assume that there exists an electron density threshold a_{AB} and two density domains $\mathrm{DD}_{F_A}(K, a_{AB})$ and $\mathrm{DD}_{F_B}(K, a_{AB})$ in the two molecules where these density domains are actually suitable to serve as the density domain in the defining criterion for the functional group. One may take a_{AB} as the average value of all such electron density thresholds. The electron densities $\rho_{F_{A,a_{AB}}}(\mathbf{r})$ and $\rho_{F_{B,a_{AB}}}(\mathbf{r})$ in the corresponding two density domains, $\mathrm{DD}_{F_A}(K, a_{AB})$ and $\mathrm{DD}_{F_B}(K, a_{AB})$, respectively, are necessarily different, otherwise, the unique extension property of local densities, a consequence of the holographic electron density theorem, would imply that the two molecules A and B are also the same; that is,

$$\mathrm{EXT}\ \rho_{F_{A,a_{AB}}}(\mathbf{r}) = \rho_A(\mathbf{r}) \tag{26}$$

$$\mathrm{EXT}\ \rho_{F_{B,a_{AB}}}(\mathbf{r}) = \rho_B(\mathbf{r}), \tag{27}$$

where

$$\rho_{F_{A,a_{AB}}}(\mathbf{r}) = \rho_{F_{B,a_{AB}}}(\mathbf{r}) \tag{28}$$

implies

$$\rho_A(\mathbf{r}) = \rho_B(\mathbf{r}), \tag{29}$$

contrary to the initial assumption.

The holographic electron density theorem does not exclude the possibility that two MIDCOs, for example, the two MIDCOs $G_{F_A}(K, a_{AB})$ and $G_{F_B}(K, a_{AB})$ on the boundaries of the two density domains $\mathrm{DD}_{F_A}(K, a_{AB})$ and $\mathrm{DD}_{F_B}(K, a_{AB})$ coincide,

$$G_{F_A}(K, a_{AB}) = G_{F_B}(K, a_{AB}), \tag{30}$$

even if

$$\rho_{F_{A,a_{AB}}}(\mathbf{r}) \neq \rho_{F_{B,a_{AB}}}(\mathbf{r}). \tag{31}$$

Nevertheless, this is a very unlikely event, and within any open range of density thresholds a, satisfying the condition that these MIDCOs separate the nuclei of the functional group F from the rest of the nuclei of the respective molecules A and B, there must exist a continuum of threshold values a' such that

$$G_{F_A}(K, a') \neq G_{F_B}(K, a') \,. \tag{32}$$

In fact, according to the holographic electron density theorem, no density threshold interval $[a_A, a_A]$ of positive width exists where $G_{F_A}(K, a) = G_{F_B}(K, a)$ for every value a from this interval. In other words, any MIDCO pair fulfilling the condition

$$G_{F_A}(K, a) = G_{F_B}(K, a) \tag{33}$$

must be isolated, if such a pair exists at all.

Consequently, if the selected threshold a_{AB} happens to belong to a coincident MIDCO pair, $G_{F_A}(K, a_{AB}) = G_{F_B}(K, a_{AB})$, then an infinitesimal change of the threshold value is sufficient to avoid coincidence. In the following discussion we shall assume that the threshold a_{AB} does not belong to a "shape degeneracy" expressed by $G_{F_A}(K, a_{AB}) = G_{F_B}(K, a_{AB})$ for $\rho_{F_{A,a_{AB}}}(\mathbf{r}) \neq \rho_{F_{B,a_{AB}}}(\mathbf{r})$.

A function $\rho_{F_{A-B,a_{AB}}}(\mathbf{r})$ is defined as

$$\rho_{F_{A-B,a_{AB}}}(\mathbf{r}) = \rho_{F_{A,a_{AB}}}(\mathbf{r}) - \rho_{F_{B,a_{AB}}}(\mathbf{r}) \,, \tag{34}$$

where the superposition of the two functional groups F_A and F_B, and the implied superposition of the corresponding functions $\rho_{F_{A,a_{AB}}}(\mathbf{r})$ and $[-\rho_{F_{B,a_{AB}}}(\mathbf{r})]$ minimizes the maximum value of the function $|\rho_{F_{A,a_{AB}}}(\mathbf{r}) - \rho_{F_{B,a_{AB}}}(\mathbf{r})|$,

$$\max |\rho_{F_{A,a_{AB}}}(\mathbf{r}) - \rho_{F_{B,a_{AB}}}(\mathbf{r})| = \min \,. \tag{35}$$

The set Z of points $\mathbf{r}$ where the function $\rho_{F_{A-B,a_{AB}}}(\mathbf{r})$ is zero,

$$Z = \left\{ \mathbf{r} : \rho_{F_{A-B,a_{AB}}}(\mathbf{r}) = 0 \right\} \tag{36}$$

provides a topological characterization for the differences between the two versions, F_A and F_B, of functional group F. If shape degeneracy occurs, then a typical set Z subdivides the space into compartments and the pattern of these compartments can be described by three-dimensional homology groups [40]. Alternatively, the patterns these compartments produce on 2D MIDCO surfaces, $G_{F_A}(K, a_{AB})$ and $G_{F_B}(K, a_{AB})$, can be described by the general Shape Group Method [40].

5
Symmetry Deficiency as a Diagnostic Tool for Atypical Functional Groups

The fuzzy electron density cloud of a complete molecule, as well as that of a functional group, can be treated as a formal fuzzy set. Fuzzy sets are defined in terms of their membership functions specifying for each element its "degree of belonging" to the given fuzzy set. In the case of electron densities a scaling procedure is required in order to ensure that the membership values fall within the usual interval of $[0, 1]$, where a value of 1 for the membership function of an

element indicates complete belonging, whereas a membership function value of 0 indicates a compete lack of this element belonging to the given fuzzy set A.

Considering a set of possible functional groups $F_1, F_2, \ldots, F_m$, of a molecule X, and the associated functional group electron densities $\rho_{F_1}(\mathbf{r}), \rho_{F_2}(\mathbf{r}), \ldots, \rho_{F_m}(\mathbf{r})$, we first note that the thresholds are not restricted to any specific value.

For scaling purposes, the maximum value $\rho_{\max,i}$ of the electron density $\rho_{F_i}(\mathbf{r})$ within a spatial domain D_{F_i} containing all the nuclei of functional group F_i is used as a reference:

$$\rho_{\max,i} = \max\left\{\rho_{F_i}(\mathbf{r}), \quad \mathbf{r} \in D_{F_i}\right\} . \tag{37}$$

We shall assume that a point $\mathbf{r} = \mathbf{r}_{\max,i}$ is chosen where this maximum density value $\rho_{\max,i}$ is realized for the given functional group:

$$\rho_{F_i}(\mathbf{r}_{\max,i}) = \rho_{\max,i} . \tag{38}$$

The fuzzy membership function for points $\mathbf{r}$ of the space describing their "degree of belonging" to the functional group F_i is defined as

$$\mu_{F_i}(\mathbf{r}) = \rho_{F_i}(\mathbf{r})/\rho_{\max,i} . \tag{39}$$

In many fuzzy electron density problems the density contributions of the functional groups and other moieties of the molecule X are also considered. In such cases, the membership function $\mu_{F_i}(\mathbf{r})$ defined above is no longer appropriate. An alternative fuzzy set approach, that includes the effects of the electron density contributions of all other functional groups of the molecule, is based on Mezey's additive fuzzy density matrix fragmentation method [20, 21].

The complete molecule X is regarded as a collection of appropriately arranged, mutually interpenetrating electron density clouds of functional groups $F_1, F_2, \ldots, F_j, \ldots, F_m$, where each fuzzy electron density fragment F_j contains the corresponding set of nuclei. As a consequence of the exact additivity property of the AFDF scheme, at each point $\mathbf{r}$ the total electronic density $\rho_X(\mathbf{r})$ of molecule X is given as the sum of functional group electron densities:

$$\rho_X(\mathbf{r}) = \sum_j \rho_{F_j}(\mathbf{r}) . \tag{40}$$

The corresponding fuzzy membership function $\mu_{F_i,X}(\mathbf{r})$ for points $\mathbf{r}$ of the space belonging to functional group F_i of molecule X can be defined as

$$\mu_{F_i,X}(\mathbf{r}) = \mu_{F_i}(\mathbf{r}) \left[\rho_{\max,i}/\rho_X(\mathbf{r}_{\max,i})\right] , \tag{41}$$

equivalent to the density ratio:

$$\mu_{F_i,X}(\mathbf{r}) = \rho_{F_i}(\mathbf{r})/\rho_X(\mathbf{r}_{\max,i}) . \tag{42}$$

These fuzzy membership functions $\mu_{F_i,X}(\mathbf{r})$ describe the "degree of belonging" of individual points $\mathbf{r}$ to various functional groups $F_1, F_2, \ldots, F_i, \ldots, F_m$ of molecule X.

For the fuzzy electron densities of functional groups a conceptually simple computational tool is often useful to diagnose the change of shape due to

interactions of the functional group with the rest of the molecule. If the ideal functional group has some symmetry, which is not uncommon if the functional group involves only a few nuclei, then deviations from the ideal symmetry may serve as an indication of the influence of the surrounding molecular regions. Since the contribution of a functional group to the overall fuzzy electron distribution of the molecule is also a fuzzy electron density cloud, it is natural to evaluate the deviation of the functional group contribution from the ideal symmetry using fuzzy set methods. For this purpose we shall review briefly some of the essential properties of fuzzy symmetry deficiency measures, following a more detailed earlier description [17].

We shall require an auxiliary measure, the cardinality $m(A)$ of a fuzzy set A. This measure $m(A)$ that can be regarded as a generalization of the concept of *mass* for fuzzy sets. The measure $m(A)$ of a fuzzy set A can be defined as the integral

$$m(A) = \int X\mu_A(x)\mathrm{d}x\,. \tag{43}$$

The integration is over the whole domain X of fuzzy set A, and the associated n-dimensional volume element $\mathrm{d}x$ refers to the metric and dimension of the underlying set X. The integration is replaced by summation if the underlying set is a discrete set.

With respect to a family of symmetry elements,

$$R = \{R_1, R_2, ..., R_m\} \tag{44}$$

a fuzzy set A is regarded as an *R-set* by definition if set A has each of the symmetry elements R_i of family R.

With respect to the given family R of symmetry elements, a fuzzy set B is regarded as an *R-deficient set* by definition if B has none of the point symmetry elements of family R. Infinitesimal distortions are sufficient to destroy a given symmetry element, consequently, the total mass difference between a fuzzy set of some symmetry and another fuzzy set that does not have this symmetry can be infinitesimal. As far as mass is concerned, R-deficient fuzzy sets and fuzzy R-sets can be nearly identical.

The concepts of *maximal mass R-subset* and *minimal mass R-superset* provide numerical measures for the evaluation of the degree of symmetry deficiency of various electron density contributions of functional groups.

Maximal R-subset: The fuzzy set B' is a *maximal R-subset* of fuzzy set A if B' is an R-set, $A \supset B'$, and no R-set B'' exists such that $B'' \supset B'$, $B' \neq B''$, and $A \supset B''$. Note that the fuzzy, maximal R-subset of B' is not necessarily unique for any given fuzzy set A.

Maximal mass R-subset: The fuzzy set B is a *maximal mass R-subset* of fuzzy set A if B is a fuzzy R-set, $A \supset B$, and if for all maximal fuzzy R-subsets B' of fuzzy set A, $m(B') \leq m(B)$. Note that, the fuzzy, maximal mass R-subset B

is not necessarily unique for any given fuzzy set A; however, the mass $m(B)$ is already a unique number for each fuzzy set A.

Minimal R-superset: The fuzzy set C' is a *minimal R-superset* of fuzzy set A if C' is an R-set, $C' \supset A$, and if no fuzzy R-set C'' exists such that $C' \supset C''$, $C' \neq C''$, and $C'' \supset A$. The fuzzy, minimal R-superset C' is not necessarily unique for any given fuzzy set A.

Minimal mass R-superset: The fuzzy set C is a *minimal mass R-superset* of fuzzy set A if C is an R-set, $C \supset A$, and if for all minimal fuzzy R-supersets C' of fuzzy set A, $m(C) \leq m(C')$. Note that, a fuzzy minimal mass R-superset C is not necessarily unique for a given fuzzy set A, but the mass $m(C)$ is a unique number for each fuzzy set A.

If the fuzzy set A is an R-set then B and C are the same, unique set, A itself:

$$B = C = A\,. \tag{45}$$

Of course, in such a case there is no symmetry deficiency concerning the symmetry elements in set R.

More interesting are those cases where the ideal functional group has a given family R of symmetry elements, but some of the actual versions of the functional group in a set of diverse molecules miss some or all of these symmetry elements as a result of the influence of the surroundings of the functional group within the molecules. In some instances the functional group may exhibit large deviations from the ideal symmetry, and such large deviations are likely to correlate with unusual reactivity properties. Measures of the degree of symmetry deficiency, when applied to fuzzy electron density contributions, may provide useful clues for such unusual reactivities.

The relations involving a fuzzy set A, a maximal mass fuzzy R-subset B of fuzzy set A, and a minimal mass fuzzy R-superset C of fuzzy set A define various measures for symmetry deficiency.

Specifically, the *internal R-deficiency measure* $\delta_{R,B}(A)$, defined as

$$\delta_{R,B}(A) = 1 - m(B)/m(A) \tag{46}$$

and the *external R-deficiency measure* $\delta_{R,C}(A)$, defined as

$$\delta_{R,C}(A) = 1 - m(A)/m(C) \tag{47}$$

provide numerical tools for the evaluation of symmetry deficiency.

By taking set A as the fuzzy set of electron density of a functional group, the deviations from any ideal symmetry R can be characterized.

If applied to functional groups, then the internal R-deficiency measure $\delta_{R,B}(A)$ refers to electron density contributions actually present within the molecule, whereas the external R-deficiency measure $\delta_{R,C}(A)$ may involve fuzzy sets that do not necessarily occur within the given molecule.

Taking the average of the two measures $\delta_{R,B}(A)$ and $\delta_{R,C}(A)$ defines the following *R-deficiency measure:*

$$\Delta_{R,B}(A) = (\delta_{R,B}(A) + \delta_{R,C}(A))/2\,. \tag{48}$$

Similar to the external *R*-deficiency measure $\delta_{R,C}(A)$, the *R*-deficiency measure $\Delta_{R,B}(A)$ may involve fuzzy sets which are not physically present within the given molecule.

Whereas both the external *R*-deficiency measure $\delta_{R,C}(A)$ and the *R*-deficiency measure $\Delta_{R,B}(A)$ are mathematically correct and valid tools for comparisons of the degrees of symmetry deficiencies and their correlations with unusual reactivities of functional groups, the actual physical existence of fuzzy sets involved in the determination of the internal *R*-deficiency measure $\delta_{R,B}(A)$ nevertheless suggests that this measure is likely to provide more reliable correlations with variations in the chemical reactivity of functional groups.

6 Summary

Specific aspects of the quantum chemical concept of local electron densities and functional groups of chemistry have been discussed, with emphasis on the Additive Fuzzy Density Fragmentation (AFDF) Principle, on the Adjustable Density Matrix Assembler (ADMA) Method of using a local density matrix formalism of fuzzy electron density fragments in macromolecular quantum chemistry, and on the fundamental roles of the holographic electron density theorem, local symmetry, and symmetry deficiency.

7 References

1. Hohenberg P, Kohn W (1964) Phys Rev 136:B864
2. Levy M (1979) Proc Natl Acad Sci USA 76:6062
3. Levy M (1979) Bull Amer Phys Soc 24:626
4. Levy M (1982) Phys Rev A 26:1200
5. Levy M (1990) Adv Quant Chem 21:69
6. Hegerfeldt GC, Ruijsenaars SNM (1980) Phys Rev D 22:377
7. Foster JM, Boys SF (1960) Rev Mod Phys 32:300
8. Boys SF (1966) In: Löwdin P-O (ed) Quantum Theory of Atoms, Molecules and the Solid-state. Academic Press, New York
9. Edmiston C, Ruedenberg K (1963) Rev Mod Phys 35:457
10. Riess J, Münch W (1981) Theor Chim Acta 58:295
11. Pipek J, Mezey PG (1988) Internat J Quantum Chem Symp 22:1
12. Pipek J, Mezey PG (1989) J Chem Phys 90:4916
13. Schmidt MW (1997) GAMESS Quantum Chemistry Program Package Documentation (ISU Quantum Chemistry Group ISUQCG, Ames Laboratory — USDOE, Iowa State University, IA 50011 USA)
14. Mezey PG (1994) Canad J Chem 72:928 (Special issue dedicated to Prof J C Polanyi)

15. Mezey PG (1996) Functional Groups in Quantum Chemistry. In: Advances in Quantum Chemistry, Vol 27, p 163
16. Mezey PG (1996) Molecular Similarity Measures of Conformational Changes and Electron Density Deformations. In: Advances in Molecular Similarity, Vol 1, p 89
17. Mezey PG (1997) Fuzzy Measures of Molecular Shape and Size. In: Rouvray DH (ed) Fuzzy Logic in Chemistry, Academic Press, San Diego
18. Mezey PG (1997) Internat Rev Phys Chem 16:361
19. Mezey PG (1995) Density Domain Bonding Topology and Molecular Similarity Measures. In: Sen K (ed) Topics in Current Chemistry, Vol 173, Molecular Similarity. Springer-Verlag, Heidelberg, pp 63-83
20. Mezey PG (1995) J Math Chem 18:141
21. Mezey PG (1997) Int J Quantum Chem 63:39
22. Walker PD, Mezey PG (1993) J Amer Chem Soc 115:12423
23. Walker PD, Mezey PG (1994) J Amer Chem Soc 116:12022
24. Mezey PG (1999) Mol. Phys 96:169
25. Mezey PG (1998) J Math Chem 23:65
26. Mezey PG (to be published).
27. Luo X, Arteca GA, Zhang C, Mezey PG (1993) J Organometallic Chem 444:131
28. Walker PD, Mezey PG, Maggiora GM, Johnson MA, Petke JD (1995) J Comput Chem 16:1474
29. Walker PD, Maggiora GM, Johnson MA, Petke JD, Mezey PG (1995) J Chem Inf Comp Sci 35:568
30. Mezey PG, Zimpel Z, Warburton P, Walker PD, Irvine DG, Dixon DG, Greenberg B (1996) J Chem Inf Comp Sci 36:602
31. Du Q, Arteca GA, Mezey PG (1997) J Comp -Aided Mol Design 11:503
32. Mezey PG, Zimpel Z, Warburton P, Walker PD, Irvine DG, Huang X-D, Dixon DG, Greenberg BM (1998) Environ Toxicol Chem 17:1207
33. Mezey PG (1998) Molecular Structure — Reactivity — Toxicity Relationships. In: Huang PM (ed) Soil Chemistry and Ecosystem Health. SSSA Publ, Pittsburgh, PA, USA, pp 21-43
34. Du Q, Mezey PG (1998) J Comp-Aided Mol Design (in press)
35. Mezey PG (1999) J Chem Inf Comp Sci 39:224
36. Mezey PG (1999) Quantitative Shape — Activity Relations (QShAR), Molecular Shape Analysis, Charge Cloud Holography, and Computational Microscopy. In: Walker JD (ed) QSARs in Environmental Toxicology — VIII. QSARs for Predicting Endocrine Disruption, Chemical Persistence and Effects. SETAC Publ (in press)
37. Walker PD, Mezey PG (1994) Canad J Chem 72:2531
38. Walker PD, Mezey PG (1995) J Math Chem 17:203
39. Mezey PG, Walker PD (1997) Drug Discovery Today (Elsevier Trend Journal) 2:6
40. Mezey PG (1993) Shape in Chemistry: An Introduction to Molecular Shape and Topology. VCH Publishers, New York

Electron Correlation and Reduced Density Matrices

Carmela Valdemoro

Instituto de Matemáticas y Física Fundamental,
Consejo Superior de Investigaciones Científicas, Serrano 123,
E-28006 Madrid, Spain. *E-mail: c.valdemoro@imaff.cfmac.csic.es*

Electron correlation in atoms and molecules finds a very natural expression in terms of the reduced density matrices within the formalism of second quantization. The fundamental relations of this theory lead here to a discussion of the correlation effects and their connexion with two properties of the group functions frequently studied by Prof. Ede Kapuy: strong orthogonality and the concept of independence. The other aim of this paper is to examine the direct expression of the correlation effects in terms of reduced density matrices.

Keywords: Reduced density matrices, electron correlation, strong orthogonality, contracted Schrödinger equation

Topics in Current Chemistry, Vol. 203

1 Introduction

The study of electron correlation, in terms of reduced density matrices (RDM's) may be approached from many different points of view. My first purpose here is to connect the main relations involving *RDM*'s with some concepts appearing in the theory developed by Prof. Ede Kapuy [1–5]. Thus, after an introductory theoretical background (Sect. two), I look again at the concepts of *independence* and of *strong orthogonality* within an extended second quantization framework.

My second aim here is directly related with the research project which is being developed in our laboratory at present: the iterative solution of the first and second order contracted Schrödinger equations, (1- and 2-CSE respectively).

In 1976 Nakatsuji [6] and Cohen and Frishberg [7], by integrating the Schrödinger equation over the variables of N-1 and N-2 electrons, obtained the first- and second order *density equations* respectively. These *density equations* were elements of a hierarchy of coupled integro-differential equations. Each equation by itself was undetermined [8], since the p-CSE solution was a function of the (p+1)- and of the (p+2)-*CSE*. By applying a matrix contracting mapping (CM) to a matrix representation of the Schrödinger equation, Valdemoro obtained [9] a totally equivalent hierarchy of matrix equations, the *CSE*'s. Since then, the question of how to decouple this hierarchy of equations has been underlying our research. An algorithm for obtaining a satisfactory approximation of a *p-RDM* in terms of the lower order *RDM*'s was found in the early nineties [10–12]. This result opened the way towards an iterative solution of the *CSE* [12–14]. In a similar way, as in the Hartree–Fock theory, there are many possible ways of looking for a rapid convergence and a stable solution in the case of the *CSE*. Several such procedures have been proposed [12–17] but probably this part of the method can still be improved. However, the crucial step for obtaining good results is not the iterative scheme itself but the optimization of the algorithm for approximating a *p-RDM* as a function of the *t-RDM* for $t < p$. The importance of this problem was pointed out by Nakatsuji and Yasuda [15,16], who on the basis of the Green function perturbation theory and its asymptotic limit to an *RDM* expansion, proposed a method for improving the quality of the approximated 3-*RDM* and 4-*RDM*. The same *RDM* expansion was obtained by following an alternative approach described by Mazziotti [17,18].

Notwithstanding the very good results obtained by these authors, I consider that this question is still an open and challenging problem which deserves further study. In this respect, Valdemoro et al. [19], recently reported the exact structure of the pure two- and three-body terms, as well as some basic properties of the two-body terms.

Here, I will focus my attention on an interesting feature of the 1-CSE. Mainly, that the 1-*CSE* accepts at least two solutions [6,7,13]: the FCI and the Hartree–

Fock (HF) one. This property is exploited (Sect. 4) for obtaining an equation in which both the correlation energy and the correlation matrices appear explicitly. These correlation matrices are defined here as the difference between a *FCI-RDM* and the corresponding HF-*RDM*. By applying the arguments given in [19] the exact structure of these correlation matrices can be expressed in terms of the 1-*RDM* and of the first order transition *RDM*'s. A calculation of the ground state of the beryllium atom illustrates the formalism.

2
Theoretical Background

In the occupation number representation within the second quantization formalism, which will be used here, the 1-*RDM* takes the form:

$$ {}^{1}\mathrm{D}^{\Phi}_{ij} \;=\; \left\langle \Phi | \mathrm{b}^{\dagger}_{i} \mathrm{b}_{j} | \Phi \right\rangle \;, \tag{1} $$

where the operators $\mathrm{b}^{\dagger}$ and b create/annihilate an electron on the N-electron state denoted as $|\Phi\rangle$ and i and j represent orthonormal spin-orbitals. Thus, are expressed in a compact and clear formula both the integration over the variables of $N-1$ electrons and the representation of a matrix element in the spin-orbital basis. This formula is easily generalized when the aim is to consider explicitly the variables of more than one electron. Thus, the r-RDM (the integration concerns $N-r$ electron variables) is written in this notation as:

$$ {}^{r}\mathrm{D}^{\Phi}_{i_1,i_2,\ldots,i_r;j_1,j_2,\ldots,j_r} \;=\; \frac{1}{r!} \left\langle \Phi | \mathrm{b}^{\dagger}_{i_1} \mathrm{b}^{\dagger}_{i_2} \ldots \mathrm{b}^{\dagger}_{i_r} \ldots \mathrm{b}_{j_r} \ldots \mathrm{b}_{j_2} \mathrm{b}_{j_1} | \Phi \right\rangle \;. \tag{2} $$

When one imposes a definite order to the spin-orbital indices, i.e. $i_1 < i_2 < \ldots < i_r$ and similarly for the j string, the factor $\frac{1}{r!}$ disappears.

2.1
The Global Operators in Second Quantization

Usually, when working in second quantization, the developments consider the creator/annihilator operators of each single electron explicitly. However, one may define the global operators (GO), which create/annihilate more than one electron at a time. After one becomes familiar with their algebra, the use of the GO's enhances the directness of the deductions and renders the results more compact and thus easier to analyse, unless one wishes to stress an one-electron property. These GO's, which will be used in the following section, are defined as:

$$ {}^{r}\mathrm{B}^{\dagger}_{\Theta} \;\equiv\; \mathrm{b}^{\dagger}_{\vartheta_1} \mathrm{b}^{\dagger}_{\vartheta_2} \ldots \mathrm{b}^{\dagger}_{\vartheta_r} \;, \tag{3} $$

with $\vartheta_1 < \vartheta_2 < \ldots < \vartheta_r$, and

$$ {}^{r}\mathrm{B}_{\Lambda} \;\equiv\; \mathrm{b}_{\lambda_r} \ldots \mathrm{b}_{\lambda_2} \mathrm{b}_{\lambda_1} \;. \tag{4} $$

These definitions may be extended to any basis of orthonormal r-electron states. Thus, if one has a basis of Slater determinants (Λ), then any state $|I\rangle$ is:

$$|I\rangle \;=\; \sum_{\Lambda} \mathcal{C}_{I\Lambda}\, |\Lambda\rangle \tag{5}$$

and

$${}^{r}\mathrm{B}_{I}^{\dagger} \;\equiv\; \sum_{\Lambda} \mathcal{C}_{I\Lambda}\, {}^{r}\mathrm{B}_{\Lambda}^{\dagger}\,. \tag{6}$$

The algebra obeyed by these *GO*'s [20](which follows from that of the single electron operators) is summarized in the relation

$$\boxed{{}^{q}\mathrm{B}_{I}\, {}^{r}\mathrm{B}_{J}^{\dagger} \;-\; (-1)^{rq}\, {}^{r}\mathrm{B}_{J}^{\dagger}\, {}^{q}\mathrm{B}_{I} \;=\; \sum_{k=0}^{q-1} (-1)^{(r-q)q+k} \sum_{\gamma=1}^{\binom{2K}{r-q+k}} \sum_{\eta=1}^{\binom{2K}{k}} \mathrm{D}_{\gamma\eta}^{JI}\, {}^{(r-q+k)}\mathrm{B}_{\gamma}^{\dagger}\, {}^{k}\mathrm{B}_{\eta}}\,, \tag{7}$$

where

$$\mathrm{D}_{\gamma\eta}^{JI} \;=\; \langle J|\mathrm{B}_{\gamma}^{\dagger}\, \mathrm{B}_{\eta}\, |I\rangle\,.$$

In Eq.(7) we have assumed that, $q < r$ (the case $r < q$, which has a very similar structure, is not needed here). We will assume in what follows that the states denoted by the capital Latin letters ($I, J, ..$) correspond to general states of the form described in Eq.(5). The lower case Greek letters always refer to a single determinant unless it is explicitly mentioned. Equation (7) is valid also for $q = r$ and in this case it should be noted that the term corresponding to $k = 0$ is just $\delta_{I;J}$

3
Independence and Strong Orthogonality

Strictly speaking, an independence shown by the probability distributions derived from two states implies an absence of interaction between the systems or subsystems described by these two states; however, in practice, the great success of models which combine the concept of *independence* with a non-negligible interaction shows that "except for a relatively small *correlation* error, **on average** two systems may show an *independent* behaviour even if they interact rather strongly". The most obvious example of this is the Hartree–Fock method.

The aim of this section is to look at the concept of *independence* from the second quantization point of view, and in connection with the anticommutation/commutation properties of the *GO*'s.

Let us consider Eq. (7). If the states I and J were independent, the *r.h.s* of this equation would be zero. That is, B_I and $B_J^{\dagger}$ would exactly commute/anticommute, since the only thing that should be done when shifting their order would be to multiply the new product by $(-1)^{qr}$. The condition

$$\mathrm{D}_{\gamma\eta}^{JI} \;=\; 0 \tag{8}$$

for all the configurations γ and η corresponding to $(r-1)$ and $(q-1)$ electrons respectively, would imply that all the terms of the *r.h.s.* of the equation would also be zero since all the lower order D^{JI} would be obtained by contraction of a matrix equal to zero. A similar analysis, where the *RDM*'s only appeared implicitly, was carried out for geminals by Surján ten years ago [21,22]. Equation (8) expresses, within this formalism, the *strong orthogonality* condition which was introduced in the fifties [1,23–26] and whose applicability continued to be discussed in the sixties [27–36] when the separated groups and pairs theory became one of the main centers of interest in quantum chemistry. It should be noted that the interest in discussing the connection among separability, independence, and strong orthogonality has not lessened, hence McWeeny's recent analysis of the separability of quantum systems [37].

It seems, therefore, that the strong orthogonality between two states implies, according to Eq.(7), that the operators which create and annihilate those states commute, or anticommute, according to whether (qr) is even or odd respectively.

When the states I and J describe groups of the same number r of electrons, the condition expressed by Eq.(8) may be softened by referring, in an averaged way, to a concrete state of a system, Φ. Thus, let us consider $q = r$ and take the expectation value of Eq.(7) by $|\Phi\rangle$,

$$\boxed{{}^r\bar{\mathrm{D}}^{\Phi}_{JI} \; - \; (-1)^{r^2} \; {}^r\mathrm{D}^{\Phi}_{JI} \; = \; \delta_{I;J} \; - \; \sum_{k=1}^{(r-1)} (-1)^k \;\; tr \; ({}^k\mathrm{D}^{JI} \; {}^k\mathrm{D}^{\Phi})} \,, \qquad (9)$$

where ${}^r\bar{\mathrm{D}}^{\Phi}$ is the r-order holes reduced density matrix (r-HRDM), which provides the complementary information to that given by the *r-RDM*.

When $I \neq J$, the condition

$$\sum_{k=1}^{(r-1)} (-1)^k \;\; tr \; ({}^k\mathrm{D}^{JI} \; {}^k\mathrm{D}^{\Phi}) \; = \; 0 \qquad (10)$$

describes an **effective** or **average** independence (or average strong orthogonality) of the states I and J when the N-electron system is in state Φ. Let us suppose that $r = 2$ and that Eq.(10) is fulfilled for all $I \neq J$ and $r = 2$. In this case, **on average**, there would be no correlation between the different geminals.

At this stage, let me just mention another approach which is also concerned with the separability and independence among pairs and groups [38–41] but where the strong orthogonality condition is not explicitly imposed. This method approximates the 2-*RDM*, without a previous knowledge of the wave function, by modeling the N-electron system as $\binom{N}{p}$ effective independent pairs whose states are described by the eigenvectors of the spin-adapted reduced Hamiltonian; which is a two electron matrix which contains all the relevant information about the N-electron system. The similarity of this approach to that of the separated pair methods focused here is, due to the terminology, more

apparent than real; which explains why the conditions set in both approaches are not the same.

Let us now return to Eq.(9) and analyse it for $I = J$. In this case, if the *r.h.s* of Eq.(9) were equal to one, the geminal I would behave as a boson. Let us therefore examine to what extent this is possible under the usual conditions. Since one is free to choose the form of the one electron orthonormal basis set, I will consider that this basis is formed by natural orbitals, thus one may write:

$$1 - tr\,({}^1\mathrm{D}^{II}\,{}^1\mathrm{D}^{\Phi}) \;=\; 1 - \sum_i {}^1\mathrm{D}^{II}_{i\,;\,i}\,{}^1\mathrm{D}^{\Phi}_{i\,;\,i}\,. \tag{11}$$

Clearly all the products which are added up in the second term of Eq. (11) are positive; therefore the total value of these terms may be small but it cannot have the value zero unless the occupation number of the geminal I in Φ is only generated by the non-diagonal contributions

$$\sum_{\lambda\neq\eta} \mathrm{D}^{II}_{\lambda;\eta}\,\langle\Phi|\mathrm{B}^{\dagger}_{\lambda}\,\mathrm{B}_{\eta}\,|\Phi\rangle\,, \tag{12}$$

which usually is not the case. Thus, in general, a positive and not too small value of Eq.(11), corresponds to a small occupation number of I in Φ.

4
The First Order Contracted Schrödinger Equation and the Correlation Energy

It has been shown [14,17,19], that an intermediate step in the deduction of the 1-*CSE* is

$$\langle\Phi|\hat{H}\,b^{\dagger}_{p_\tau}\,b_{q_\tau}|\Phi\rangle \;=\; E_\Phi\,{}^1\mathrm{D}^{\Phi}_{p_\tau q_\tau}\,, \tag{13}$$

where Φ is an eigenstate of $\hat{H}$, and τ represents here the spin variable which is being considered. In what follows, in order to reduce the number of indices and thus simplify the reading, whenever the spin variable is equal to τ it will be omitted. I will also omit the upper index Φ when it can be inferred. From Eq.(13) follows the final form of the 1-*CSE* :

$$\begin{aligned} E\,{}^1\mathrm{D}_{p\,;\,q} \;&=\; (h\,{}^1\mathrm{D}\,)_{p\,;\,q} \;+\; 2\sum_{i,j,\sigma} h_{ij}\,{}^2\mathrm{D}_{pi_\sigma\,;\,qj_\sigma} \\ &+\; 2\sum_{\ell,\sigma}(V\,{}^2\mathrm{D})_{p\ell_\sigma\,;\,q\ell_\sigma} \;+\; 3\sum_{i,j,k,\ell,\sigma,\sigma'} V_{ij;\ell k}\,{}^3\mathrm{D}_{i_\sigma j_{\sigma'} p\,;\;\ell_\sigma k_{\sigma' q}}\,, \end{aligned} \tag{14}$$

where i, j, k, ℓ.. represent the orthonormal orbitals, σ and σ' denote the spin variables, h groups the one electron integrals and V the two electron ones. Equation (14) is an alternative and equivalent form to that of the 1-*CSE* given previously [19]. In Eq.(14), the one electron terms of the Hamiltonian appear explicitly, instead of being included into an effective two-electron matrix, ${}^oH$ [14]. In this way, when the Hartree–Fock solution is taken as a reference, the analysis of the correlation effects is more direct. Note that the indices are free

to take any values, which implies that the 2-*RDM* and the 3-*RDM* contain the factors $\frac{1}{2!}$ and $\frac{1}{3!}$ respectively [see Eq.(2)].

An important feature of the 1-*CSE* is that it is exactly satisfied by at least two sets of *RDM*'s [6,7,13,19], the set corresponding to the Hartree–Fock solution and the set corresponding to the *FCI* solution. In what follows, the set of *HF* matrices, as well as the corresponding energy will be distinguished from those corresponding to the exact *FCI* solution by an upper index $(*)$. Thus the *HF* form of Eq.(14) is written

$$E^* \, {}^1\mathrm{D}^*_{p\,;\,q} \;=\; (h\,{}^1\mathrm{D}^*)_{p\,;\,q} \;+\; 2\sum_{i,j,\sigma} h_{ij}\,{}^2\mathrm{D}^*_{pi_\sigma\,;\,qj_\sigma} + \;2\sum_{\ell,\sigma}(V\,{}^2\mathrm{D}^*)_{p\ell_\sigma\,;\,q\ell_\sigma} \;+\; 3\sum_{i,j,k,\ell,\sigma,\sigma'} V_{ij;\ell k}\,{}^3\mathrm{D}^*{}_{i_\sigma j_{\sigma'} p\,;\;\ell_\sigma k_{\sigma'} q}\,. \qquad (15)$$

Let us now subtract Eq.(15) from Eq.(14). The *r.h.s.* of the new equation has the same form as that of the 1-*CSE*, except that the *RDM* elements are replaced by those of the correlation terms of first, second and third order. These correlation matrices will be denoted in what follows by ${}^1\Upsilon$, ${}^2\Upsilon$ and ${}^3\Upsilon$ respectively which are defined as:

$$^i\Upsilon \;=\; {}^i\mathrm{D} \;-\; {}^i\mathrm{D}^*\,. \qquad (16)$$

The resulting equation may therefore be expressed as:

$$E_{corr}\,({}^1\mathrm{D}^*_{p\,;\,q} \;+\; {}^1\Upsilon_{p\,;\,q}) + E^*\,{}^1\Upsilon_{p\,;\,q} \;=\; (h\,{}^1\Upsilon)_{p\,;\,q} \;+\; 2\sum_{i,j,\sigma} h_{ij}\,{}^2\Upsilon_{pi_\sigma\,;\,qj_\sigma} + 2\sum_{\ell,\sigma}(V\,{}^2\Upsilon)_{p\ell_\sigma\,;\,q\ell_\sigma} \;+\; 3\sum_{i,j,k,\ell,\sigma,\sigma'} V_{ij;\ell k}\,{}^3\Upsilon_{i_\sigma j_{\sigma'} p\,;\;\ell_\sigma k_{\sigma'} q}\,, \qquad (17)$$

that is

$$E_{corr}\,{}^1\mathrm{D}_{p\,;\,q} \;=\; \mathcal{F}({}^1\Upsilon\,,\,{}^2\Upsilon\,,\,{}^3\Upsilon\,,\,E^*)\,, \qquad (18)$$

and since the *HF* solution may be assumed to be known, the correlation energy (E_{corr}) is just a function of the Υ's. Equation (17), which will be referred to as the correlation energy equation (CEE), is very appealing since it aims directly at what is unknown and it can, in principle, be iterated in a similar way to the 1-*CSE*. However, as is later shown, a successful iterative solution of this equation relies on a successful approximation of ${}^2\Upsilon$ and ${}^3\Upsilon$.

4.1 The Correlation Matrices

In the previous paragraph, the ${}^1\Upsilon$, ${}^2\Upsilon$ and ${}^3\Upsilon$ were globally defined by Eq.(16), which emphasizes that the *HF* solution is taken as a reference.

The structure of the ${}^2\Upsilon$ and ${}^3\Upsilon$ will now be examined more closely.

The $^2\Upsilon$. When reordering the four operators appearing in Eq.(2) for $r = 2$ according to the pattern $b^\dagger b b^\dagger b$ and then inserting the resolution of identity after the first two operators, the 2-*RDM* may be decomposed as

$$\begin{aligned} 2!\ {}^2\mathrm{D}^{\Phi}_{ij\,;\,k\ell} &\equiv \langle\Phi|\mathrm{b}^\dagger_i \mathrm{b}^\dagger_j \mathrm{b}_\ell \mathrm{b}_k|\Phi\rangle \\ &= {}^1\mathrm{D}^{\Phi}_{i\,;\,k}\ {}^1\mathrm{D}^{\Phi}_{j\,;\,\ell} - {}^1\mathrm{D}^{\Phi}_{i\,;\,\ell}\ \delta_{j\,;\,k} + \sum_{\Phi'\neq\Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,k}\ {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,\ell} \end{aligned} \tag{19}$$

or, equivalently, by replacing $\delta_{j;k}$ by $({}^1\mathrm{D} + {}^1\bar{\mathrm{D}})_{j\,;\,k}$

$$\begin{aligned} 2!\ {}^2\mathrm{D}^{\Phi}_{ij\,;\,k\ell} &= {}^1\mathrm{D}^{\Phi}_{i\,;\,k}\ {}^1\mathrm{D}^{\Phi}_{j\,;\,\ell} - {}^1\mathrm{D}^{\Phi}_{i\,;\,\ell}\ {}^1\mathrm{D}^{\Phi}_{j\,;\,k} - {}^1\mathrm{D}^{\Phi}_{i\,;\,\ell}\ {}^1\bar{\mathrm{D}}^{\Phi}_{j\,;\,k} \\ &\quad + \sum_{\Phi'\neq\Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,k}\,{}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,\ell}\,, \end{aligned} \tag{20}$$

where

$$\sum_{\Phi'\neq\Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,k}\,{}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,\ell} = \sum_{\Phi'\neq\Phi} \langle\Phi|\mathrm{b}^\dagger_i \mathrm{b}_k|\Phi'\rangle\langle\Phi'|\mathrm{b}^\dagger_j \mathrm{b}_\ell|\Phi\rangle\,, \tag{21}$$

and since there are four equivalent ways of reordering the operators according to this pattern, there are also four equivalent expressions of this type for the 2-*RDM*.

The first two terms of the r.h.s. of Eq.(20) are an antisymmetrized product of one-electron probabilities, therefore they can also be used as a reference in the absence of correlation in the 2-*RDM* [15–19]. The main drawback, in considering these two terms as a reference, is that they do not form an *N*-representable 2-*RDM*. On the other hand, the 1-, 2-, 3-*HF-RDM*'s as zero-correlation references are *N*-representable, well behaved *RDM*'s, which is a strong argument in favor of taking them as references.

Let us replace the different *RDM* elements in Eq.(19), by their expression in terms of the corresponding *HF-RDM* and $^1\Upsilon$ elements. Then

$$\begin{aligned} 2!\ {}^2\Upsilon_{ij\,;\,k\ell} &= {}^1\mathrm{D}^{*}_{i\,;\,k}\ {}^1\Upsilon_{j\,;\,\ell} + {}^1\Upsilon_{i\,;\,k}\ {}^1\mathrm{D}^{*}_{j\,;\,\ell} + {}^1\Upsilon_{i\,;\,k}\ {}^1\Upsilon_{j\,;\,\ell} \\ &\quad - {}^1\Upsilon_{i\,;\,\ell}\ \delta_{j\,;\,k} - {}^1\mathrm{D}^{*}_{i\,;\,\ell}\ {}^1\bar{\mathrm{D}}^{*}_{j\,;\,k} + \sum_{\Phi'\neq\Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,k}\ {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,\ell} \\ &= {}^{1\to 2}\Upsilon_{ij\,;\,k\ell} - {}^1\mathrm{D}^{*}_{i\,;\,\ell}\ {}^1\bar{\mathrm{D}}^{*}_{j\,;\,k} + \sum_{\Phi'\neq\Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,k}\ {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,\ell}\,. \end{aligned} \tag{22}$$

The simplest way to reach this result is to replace directly the Kronecker δ appearing in Eq.(19) by ${}^1\mathrm{D}^* + {}^1\bar{\mathrm{D}}^*$. Note that the possibility of replacing a Kronecker δ by its expression in terms of the 1-*RDM* and the 1-*HRDM*, irrespective of whether the state of reference is an *FCI* or an *HF* one, implies that ${}^1\Upsilon = -{}^1\bar{\Upsilon}$.

In Eq.(22) ${}^{1\to 2}\Upsilon$ represents the contribution of $^1\Upsilon$ to the 2-*RDM*. Another type of term appearing in Eq.(22) is a product of an element of the 1-*HF-HRDM* by an element of the 1-*HF-RDM*. Finally, the last type of term is expressed in

Table 1. Some values of $^2\Upsilon$

$i\bar{j};k\bar{\ell}$	^2D	$^2D^*$	$^{1\to2}\Upsilon$	$\sum_{\Phi'\neq\Phi} {}^1D_{i\,;\,\bar{k}}^{\Phi\Phi'}\,{}^1D_{j\,;\,\bar{\ell}}^{\Phi'\Phi}$
$1\bar{1};1\bar{1}$	0.9993	1	−0.0012	0.0005
$1\bar{2};1\bar{2}$	0.9847	1	−0.0153	0.0000
$1\bar{3};1\bar{3}$	0.0147	0	0.0148	0.0000
$1\bar{2};1\bar{3}$	0.1058	0	0.1058	0.0000
$1\bar{1};3\bar{3}$	−0.0018	0	0.0001	−0.0019
$1\bar{1};4\bar{4}$	−0.0191	0	0.0000	−0.0191
$2\bar{2};2\bar{2}$	0.9726	1	−0.0292	0.0018
$2\bar{3};2\bar{3}$	0.0128	0	0.0146	−0.0018
$2\bar{2};2\bar{3}$	0.1109	0	0.1043	−0.0066
$2\bar{2};3\bar{3}$	−0.0442	0	0.0112	−0.0550
$3\bar{3};3\bar{3}$	0.0020	0	0.0002	0.0018
$2\bar{2};4\bar{4}$	−0.0014	0	0.0000	−0.0014

terms of the first order transition *RDM*'s and describes the simultaneous virtual excitations of two electrons. It is thus a pure two-body correlation term [19].

In order to analyse the relative importance of these terms, a *FCI* calculation for the ground state of the berylium atom was carried out. The basis was a Slater double zeta one [42], and the orbital basis of representation was the Hartree–Fock one (the bar over the orbital index indicates that the spin function is β). The most significant values for the $\alpha\beta$ block of the 2-*RDM* are reported in Table 1. Note that in this case the product of the 1-*HRDM* and 1-*RDM* elements is zero due to the spin orthogonality; which is why it has been chosen to illustrate the relative importance of the pure two-body effects with respect to the pure one-body ones. The results in Table 1 show that the two largest contributions are of the $^{1\to2}\Upsilon$ type. When looking more closely at these data, it appears that the relative importance of $^{1\to2}\Upsilon$ and the pure two-body term in an element depends on the kind of orbitals involved in that element. That is, it depends on whether the orbitals, in the *HF* configuration, are doubly occupied (o), doubly empty (e), the frontier orbital *homo* (h) or *lumo* (l). Thus:

- In the element $1\bar{1};1\bar{1}$, of the type $o\bar{o};o\bar{o}$, the correlation error is small, due to a cancellation of nearly half the $^{1\to2}\Upsilon$ by the two body term.
- In the elements of the type: $o\bar{o};l\bar{l}$, $o\bar{o};e\bar{e}$, where the correlation error is nonnegligible, the dominant contribution is the two-body one. In fact, the $^{1\to2}\Upsilon$ contribution is only significant when the o orbital is the *homo* one. In this case both contributions partially cancel each other.
- The *homo* and *lumo* play an important role, particularly in the elements whose labels are: $x\bar{h};x\bar{h}$, $h\bar{h};l\bar{l}$ and $h\bar{l};h\bar{l}$. In all these elements there is some cancellation between the one- and two-body terms.
- The two-body effect is very important in the $h\bar{h};l\bar{l}$ element.

The $^3\Upsilon$. Considering Eq.(2) for $r = 3$ and proceeding in a similar manner as for the 2-*RDM* case, one finds thirty six possible and equivalent ways of reordering the operators in a sequence $b^\dagger bb^\dagger bb^\dagger b$. Therefore, the formula that will be given here is just one of these 36 possible ones.

$$
\begin{aligned}
3!\ {}^3\mathrm{D}_{ijk\,;\,\ell mn} &= {}^1\mathrm{D}_{i\,;\,\ell}\ {}^1\mathrm{D}_{j\,;\,m}\ {}^1\mathrm{D}_{k\,;\,n} \\
&+ 2\ {}^2\mathrm{D}_{ij\,;\,mn}\ \delta_{k\,;\,\ell} - 2\ {}^2\mathrm{D}_{ik\,;\,mn}\ \delta_{j\,;\,\ell} - {}^1\mathrm{D}_{i\,;\,\ell}\ {}^1\mathrm{D}_{j\,;\,n}\ \delta_{k\,;\,m} \\
&- \delta_{k\,;\,m} \sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,\ell}\, {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,n} + {}^1\mathrm{D}_{i\,;\,\ell} \Big(\sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{j\,;\,m}\ {}^1\mathrm{D}^{\Phi'\Phi}_{k\,;\,n} \Big) \\
&+ \Big(\sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,\ell}\, {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,m} \Big)\ {}^1\mathrm{D}_{k\,;\,n} + \sum_{\Phi', \Phi'' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,\ell}\, {}^1\mathrm{D}^{\Phi'\Phi''}_{j\,;\,m}\, {}^1\mathrm{D}^{\Phi''\Phi}_{k\,;\,n} .
\end{aligned}
\tag{23}
$$

Now the *HF-RDM* is

$$
3!\ {}^3\mathrm{D}^*_{ijk\,;\,\ell mn} = \sum_P (-1)^P\ ({}^1\mathrm{D}^*_{i\,;\,\ell}\ {}^1\mathrm{D}^*_{j\,;\,m}\ {}^1\mathrm{D}^*_{k\,;\,n}) , \tag{24}
$$

where the $\sum_P (-1)^P$ symbol antisymmetrizes the product by permuting either the annihilator indices, ℓ, m and n or, alternatively, the creator indices and multiplying by $\pm$ according to the parity of the permutation. The structure of the $^3\Upsilon$ follows directly by applying Eq.(16); and, after some algebraic operations involving the results of the previous paragraph, one has:

$$
\begin{aligned}
3!\ {}^3\Upsilon_{ijk\,;\,\ell mn} &= {}^{1\to3}\Upsilon_{ijk\,;\,\ell mn} + {}^{2\to3}\Upsilon_{ijk\,;\,\ell mn} + {}^{1\to3}\bar{\mathrm{D}}_{ijk\,;\,\ell mn} \\
&+ {}^{2\to3}\bar{\mathrm{D}}_{ijk\,;\,\ell mn} + \sum_{\Phi', \Phi'' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,\ell}\, {}^1\mathrm{D}^{\Phi'\Phi''}_{j\,;\,m}\, {}^1\mathrm{D}^{\Phi''\Phi}_{k\,;\,n} ,
\end{aligned}
\tag{25}
$$

where the meaning of the new symbols is:

$$
\begin{aligned}
{}^{1\to3}\Upsilon_{ijk\,;\,\ell mn} &= - {}^1\mathrm{D}^*_{i\,;\,\ell}\ \delta_{k\,;\,m}\ {}^1\Upsilon_{j\,;\,n} + \delta_{j\,;\,\ell}\ {}^1\Upsilon_{i\,;\,m}\ {}^1\mathrm{D}_{k\,;\,n} \\
&- \delta_{k\,;\,m}\ {}^1\Upsilon_{i\,;\,\ell}\ {}^1\mathrm{D}_{j\,;\,n} + {}^1\mathrm{D}^*_{i\,;\,\ell}\ {}^1\mathrm{D}^*_{j\,;\,m}\ {}^1\Upsilon_{k\,;\,n} \\
&- {}^{1\to2}\Upsilon_{ik\,;\,mn}\ \delta_{j\,;\,\ell} + {}^{1\to2}\Upsilon_{ij\,;\,mn}\ \delta_{k\,;\,\ell} \\
&+ {}^{1\to2}\Upsilon_{ij\,;\,\ell m}\ {}^1\mathrm{D}_{k\,;\,n} ,
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
{}^{2\to3}\Upsilon_{ijk\,;\,\ell mn} &= - \delta_{j\,;\,\ell} \sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,m}\ {}^1\mathrm{D}^{\Phi'\Phi}_{k\,;\,n} + \delta_{k\,;\,\ell} \sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,m}\ {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,n} \\
&+ {}^1\mathrm{D}_{i\,;\,\ell} \Big(\sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{j\,;\,m}\ {}^1\mathrm{D}^{\Phi'\Phi}_{k\,;\,n} \Big) \\
&+ {}^1\mathrm{D}_{k\,;\,n} \Big(\sum_{\Phi' \neq \Phi} {}^1\mathrm{D}^{\Phi\Phi'}_{i\,;\,\ell}\ {}^1\mathrm{D}^{\Phi'\Phi}_{j\,;\,m} \Big) ,
\end{aligned}
\tag{27}
$$

Table 2. Some values of $^3\Upsilon$

$ij\bar{k};\ell\ m\bar{n}$	$^3D - {}^3D^*$	$^{1\to3}\Upsilon$	$^{2\to3}\Upsilon$	$\sum_{\Phi',\Phi''\neq\Phi} {}^1D_{i;\ell}^{\Phi\Phi'}\,{}^1D_{j;m}^{\Phi'\Phi''}\,{}^1D_{k;n}^{\Phi''\Phi}$	$^3\check{D} - {}^3D^*$
$12\bar{1};12\bar{1}$	−0.01541	−0.01585	0.00002	0.00041	−0.01541
$12\bar{2};12\bar{2}$	−0.02795	−0.02977	0.00180	0.00002	−0.02795
$12\bar{2};12\bar{3}$	0.11083	0.10423	0.00660	−0.00000	0.11083
$12\bar{1};13\bar{1}$	0.10578	0.10573	0.00001	0.00004	0.10579
$12\bar{3};12\bar{3}$	0.01272	0.01455	−0.00180	−0.00003	0.01273
$12\bar{3};13\bar{2}$	0.01264	0.01120	0.00144	0.00000	0.01264
$24\bar{4};12\bar{1}$	0.01884	0.00000	−0.00005	0.01889	0.01884
$34\bar{4};12\bar{1}$	0.00203	0.00000	0.00004	0.00199	0.00203
$23\bar{2};23\bar{2}$	0.00011	0.01435	−0.01600	0.00177	−0.00302
$23\bar{3};23\bar{3}$	0.00000	0.00022	0.00156	−0.00177	0.00314
$23\bar{2};23\bar{3}$	0.00001	0.00154	−0.00803	0.00650	−0.00067

$$
\begin{aligned}
{}^{1\to3}\bar{D}_{ijk;\ell mn} \;=\; & -\,{}^1\bar{D}^*_{j;m}\,{}^1D^*_{i;n}\,{}^1D^*_{k;\ell} \\
& +\,2\,{}^1\bar{D}^*_{k;\ell}\,{}^2D^*_{ij;mn} \;-\; 2\,{}^1\bar{D}^*_{k;m}\,{}^2D^*_{ij;\ell n} \;-\; 2\,{}^1\bar{D}^*_{j;\ell}\,{}^2D^*_{ik;mn}\,, \qquad (28)
\end{aligned}
$$

$$
{}^{2\to3}\bar{D}_{ijk;\ell mn} \;=\; 2\,{}^1D^*_{i;n}\,{}^2\bar{D}^*_{jk;\ell m}\,. \qquad (29)
$$

As previously mentioned, there are thirty-six equivalent ways of writing the $^3\Upsilon$ but the same type of terms are involved in all of them. In a similar way as in the second-order case, the relative importance of the different contributions of the correlation matrices is illustrated in Table 2. The quantities shown correspond to elements of the $\alpha\alpha\beta$ block of the 3-*RDM* obtained for the ground state of the berylium atom in the same calculation as was done previously; and therefore, in this case, and for the the elements reported here, the $^{1\to3}\bar{D}$ and the $^{2\to3}\bar{D}$ contributions are zero. In the last column I have included the difference with respect to the *HF-RDM* of $^3\check{D}$, which denotes the approximate 3-*RDM* before renormalization, obtained by applying the algorithms previously used by Valdemoro et al. [14]. For the present purpose, it is not necessary to enter more fully into the structure of $^3\check{D}$ for which the interested reader may find a detailed discussion in its spin-free form, in [11,12].

There are two results in the data reported in Table 2 which I find both striking and surprising. Unexpectedly it appears that the largest error of $^3\check{D}$ which occurs in the elements simultaneously involving the *homo* and the *lumo* is not due to the omission of three-body correlation effects, as the developments of Nakatsuji and Yasuda [15,16] seemed to imply, but is rather a result of subtracing instead of adding this three-body contribution to the joint effects of the one- and the two-body contributions. That is, in the *FCI* calculation and for these elements, the three correlation matrices practically cancel each

other, whereas in $\breve{D}$ subtraction of the three-body effect practically multiplies by a factor of two the sum of the one- and the two-body correlation effects. The other striking feature of these results is that the really significant three-body correlation effect is shown by the elements $24\bar{4};12\bar{1}$ and $34\bar{4};12\bar{1}$. And, surprisingly enough, these two elements are well estimated by the $^3\breve{D}$. If more extended calculations confirm these two kinds of results, the estimation of the 3-*RDM* could be straightforward, provided that the 2-*RDM* is known.

The special role played by the *homo* and *lumo* was also observed in a recent calculation [43] of the BeH_2, where the molecular orbitals belonged to five different symmetries. But in this case the *empty* and *occupied* concepts must be extended to each symmetry.

As a conclusion to this section, let me remark that what is interesting about the partitioning given here is that the analytical form of all the contributions to each correlation matrix is known exactly, which, in my view, favours our search for their reliable estimation. The other expansions proposed in the literature establish the correspondence of the different terms of the *RDM* expansion with the perturbative series diagrams [15,16] or with the terms obtained through the use of Schwinger probes [17,18]. A given term of these expansions does not necessarily coincide with a *p*-body correlation effect which in some instances may be misleading. For instance, what is denoted here as a pure *p*-body term does not coincide with what Mazziotti defines in [18] as a *connected* term of order *p* although the ideas behind both denominations are very closely related.

5
Concluding Remarks

What has been reported here shows that for the 1-*CSE* or the *CEE* to be operative several questions need to be answered. They may be summarized in one comprehensive one: Can a practical procedure be devised to approximate the $^1\Upsilon$, the $^2\Upsilon$ and the $^3\Upsilon$?

In the $^1\Upsilon$ case there is, I think, no problem since, provided the higher order conditions are fulfilled, each iteration should improve the 1-*RDM*, and consequently the estimation of the $^1\Upsilon$ would also improve.

Knowledge of the exact structure of the $^2\Upsilon$ and of the $^3\Upsilon$ (the $^4\Upsilon$, although not commented on here because it does not play any role in the 1-*CSE*, has also been developed) provides a sound base for devising algorithms for the evaluation of these matrices. In my view, if more complex calculations were to confirm the results reported here in Tables 1 and 2, the problem that remains to be solved will be limited to the evaluation of only a few elements in these matrices, in particular in the $^2\Upsilon$ case.

Evidently, if our objective were to solve the 2-*CSE* instead of the 1-*CSE*, the problem would not be so hard, since at each iteration the $^2\Upsilon$ should improve;

and this matrix together with $^1\Upsilon$ are possibly sufficient to obtain satisfactory estimates of the $^3\Upsilon$ and the $^4\Upsilon$.

On the other hand, the real challenge is to solve the 1-*CSE* or, equivalently, the *CEE*; hence a good algorithm is needed for approximating **all** the elements of $^2\Upsilon$. A partitioning of the pure two-body part of $^2\Upsilon$ according to the spin of the excited states (entering in the pure two-body term) was given in [19]. It was also shown there that there existed a relation linking the different spin parts of this term with products of an element of the correlated 1-*RDM* by an element of the 1-*HRDM*. Unfortunately there were more unknowns than equations, which means that this very important question remains fully open.

Acknowledgements. I wish to thank Prof. Péter Surján for granting me the opportunity of collaborating in this volume in honour of Prof. Ede Kapuy. I also wish to gratefully acknowledge the very helpful remarks and discussions of my colleagues Luis Tel and Encarnación Pérez-Romero, who revised the manuscript of this work. This work has been supported by the Dirección General de Investigación Científica y Técnica del Ministerio de Educación y Ciencia under project PB96-0891.

6 References

1. Kapuy E (1958) Acta Physica Acad Sci Hung 9:237
2. Kapuy E (1960) Acta Physica Hung 12:185
3. Kapuy E (1961) Acta Physica Hung 13:345
4. Kapuy E (1961) Acta Physica Hung 13:461
5. Kapuy E (1962) Acta Physica Hung 15:177
6. Nakatsuji H (1976) Phys Rev A 14:41
7. Cohen l, Frishberg C (1976) Phys Rev A 13:927
8. Harriman JE (1979) Phys Rev A 19:1893
9. Valdemoro C (1987) In: Erdahl R., Smith V. (eds) Density Matrices and Density Functionals (Proceedings of the A.J. Coleman Symposium, Kingston, Ontario, 1985) Reidel, Dordrecht p 275
10. Valdemoro C (1992) Phys Rev A 45:4462
11. Colmenero F, Pérez del Valle C, Valdemoro C (1993) Phys Rev A 47:971
12. Colmenero F, Valdemoro C (1993) Phys Rev A 47:979
13. Colmenero F, Valdemoro C (1994) Int J Quantum Chem 51:369
14. Valdemoro C, Tel LM, Pérez-Romero E (1997) Adv Quantum Chem 28:33
15. Nakatsuji H, Yasuda K (1996) Phys Rev Letts 76:1039
16. Yasuda K, Nakatsuji H (1997) Phys Rev A 56:2648
17. Mazziotti D (1998) Phys Rev A 57:4219
18. Mazziotti D (1998) Chem Phys Let 289:419
19. Valdemoro C, Tel LM, Pérez-Romero E (1999) Adv Quantum Chem 31:37
20. Valdemoro C, Tel LM, Pérez-Romero E (to be published)
21. Surján PR (1989) Second Quantized Approach to Quantum Chemistry, Springer, Heidelberg, p 142
22. Surján PR (1990) Croatica Chemica Acta 62:579
23. Hurley AC, Lennard-Jones J, Pople J A (1953) Proc Roy Soc (London) A220:446
24. Parr RG, Ellison FO, and Lykos PG (1956) J Chem Phys 24:1106
25. Parks JM, Parr RG (1958) J Chem Phys 28:335

26. McWeeny R (1959) Proc Roy Soc (London) A253:242
27. Arai T (1960) J Chem Phys 33:95
28. McWeeny R (1960) Rev Mod Physics 32:35
29. Löwdin PO (1961) J Chem Phys 35:78
30. Allen TL, Shull H (1961) J Chem Phys 35:1644
31. Kapuy E (1962) Phys Letts 1:205
32. Kapuy E (1966) J Chem Phys 44:956
33. Kapuy E (1966) Theoretica Chimica Acta 6:281
34. Kapuy E (1967) The Physical Society of Japan (eds). Applicability of almost strongly orthogonal geminals in many-electron wave function. Series in Selected Papers in Physics, Theory of Molecular Structure II
35. Kapuy E (1969) Acta Physica Hung 27:179
36. Kapuy E (1969) Chem Phys Letts 3:43
37. McWeeny R (1997) Conference at the Second European Workshop in Quantums Systems in Chemistry and Physics. Oxford
38. Valdemoro C (1985) Phys Rev A 31:2123
39. Valdemoro C (1996) Int J Quantum Chem 60:131
40. Karwowski J, Valdemoro C, Laín L (1989) Phys Rev A 39:4967
41. Valdemoro C, de Lara-Castells MP, Bochicchio R, Pérez-Romero E (1997) Int J Quantum Chem 65:107
42. Clementi E, Roetti C (1974) At Data Nucl Data Tables 14:428
43. Tel L, Pérez-Romero E, Valdemoro C (in preparation)

Localization via Density Functionals

N.H. March

Oxford University, Oxford, England and Department of Physics, University of Antwerp, RUCA, Antwerp, Belgium.

After a discussion of the form of low-order density matrices built from strongly orthogonal geminals, density functional theory is utilized to deal with the phenomenon of Wigner electron crystallization. This is followed by a brief treatment of localized versus molecular orbital theories of electrons, illustrated by the stretched H_2 molecule. Then a synthesis of density functional and density matrix theory is effected by deriving, following Holas and March, the exchange-correlation potential in terms of first- and second-order density matrices. Localization of correlated electrons by applied magnetic fields is finally discussed, with the example of magnetically induced Wigner solidification as a focal point.

Keywords: Density matrices, Wigner solid, strongly orthogonal geminals

Topics in Current Chemistry, Vol. 203

1 Introduction, Plus a Little History

Ede Kapuy visited my research group in the University of Sheffield, UK for 6 months around 1966, having been awarded a (then) DSIR Research Fellowship (later Science Research Council, UK: SRC). Many happy thoughts about those months come back as I write this contribution to this volume honouring Ede. As he came alone to England, he spent some time most weekends at our Sheffield home, with my late wife Joan, our two sons (teenagers at that time) and myself. My two sons and Ede were mostly in occupation of the room where we had a record-player, along with our sons' collection of the music popular at that time. Joan and I used to go to the farthest region of the house on these occasions as Eddie (as we all called him!) and our boys renormalized the unit of sound from the decibel to the megabel! But as well as fine scientific collaboration, these six months saw the forging of a firm friendship between Eddie and our family.

Returning to the science, we wrote together as a result of this visit, a paper in the Journal of Mathematical Physics [(1967) 8:1915] entitled 'Two-body orbitals for one-dimensional fermion gas with application to repulsive δ-function interactions'. Some of the quantitative results used in this study will be summarized in Sect. 2 below. However, it is highly relevant to the theme of the present article, namely 'Localization via Density Functional Theory', to note the motivation for the above paper by Ede and myself. This can be summarized by quoting the first part of the Introduction to that work:

> 'While some definite progress has been made in the search for a valid variational method based on low-order density matrices [1–5] rather

than on a many-body wave function, the necessary and sufficient conditions for a trial second-order density matrix (which determines the total energy when only two-body interactions involved) are in a form which, with present techniques, render them rather unsuitable for practical applications [6]'.

The appeal to an admittedly restrictive model of repulsive δ-function interactions between the Fermions was motivated by the desire to compare and contrast with the Overhauser [7] spin density wave state. Strongly orthogonal two-body functions (geminals) were employed by Ede and myself in the above model problem. Such geminals have implications for low-order density matrices, which we shall discuss immediately below.

2
Low-Order Density Matrices Built from Strongly Orthogonal Geminals

The starting point was to take a set of geminals $\psi_I(1,2), I = 1,2,...,N$, which satisfy the strong orthogonality conditions

$$\int \psi_I^*(1,2)\psi_J(1,3)d1 = 0, \quad J \neq I\,. \tag{1}$$

The notation here is such that 1,2,.., refers to both space and spin coordinates, while ψ_I is assumed normalized to unity, and antisymmetrized such that

$$\psi_I(1,2) = -\psi_I(2,1)\,. \tag{2}$$

Then one can construct a normalized antisymmetric $2N$-body wave function of the form

$$\Psi = \left[\frac{2^N}{(2N)!}\right]^{1/2} \sum_P (-1)^P \psi_1(1,2)\psi_2(3,4)...\psi_N(2N-1,2N) \tag{3}$$

where P denotes permutations which interchange the electrons between the geminals. This wave function takes account of the simultaneous correlations of N pairs including the corresponding unlinked clusters such that even if one passes to an infite system $N \to \infty$ one transcends the Hartree–Fock approximation unless the two-body functions are simple products of one-electron spin orbitals.

With restrictions to two-body interactions, the Hamiltonian can formally be written as

$$H = H(0) + \sum_{\alpha=1}^{2N} H(\alpha) + \sum_{\beta>\alpha=1}^{2N} H(\alpha,\beta)\,. \tag{4}$$

The expectation value of H is then readily calculated as

$$E \;=\; H(0) + \sum_{I=1}^{N} \int \psi_I^*(1,2)\,[H(1) + H(2) + H(1,2)]\,\psi_I(1,2)d1\,d2$$

$$+2\sum_{I,J=1}^{N}\int d1\,d2\,d3\,d4\,H(1,3)\,[1-P_{13}]$$
$$\times\psi_I^*(1',2)\psi_I(1,2)\psi_J^*(3',4)\psi_J(3,4)\,. \tag{5}$$

Of considerable interest for what follows is that the first- and second-order density matrices γ and Γ are given by [8]

$$\gamma(1',1)=\sum_{I=1}^{N}\gamma_I(1',1)=2\sum_{I=1}^{N}\int\psi_I^*(1',2)\psi_I(1,2)d2 \tag{6}$$

and

$$\begin{aligned}\gamma_2(1',2';1,2) \;=\; &\sum_{I=1}^{N}\psi_I^*(1',2')\psi_I(1,2)\\ &+\frac{1}{4}\sum_{I,J=1}^{N}\{\gamma_I(1',1)\gamma_J(2',2)+\gamma_I(2',2)\gamma_J(1',1)\\ &-\gamma_I(2',1)\gamma_J(1',2)-\gamma_I(1',2)\gamma_J(2',1)\}\,.\end{aligned} \tag{7}$$

The fact that these 1 DM γ and 2 DM γ_2 have been calculated from the demonstrably antisymmetric many-body wave function (3) using the strong orthogonality conditions (1) is important for what follows in the context of density functional theory (DFT) to which we now turn.

3
Origins of Density Functional Theory: the Thomas–Fermi–Dirac Statistical Method Applied to Heavy Atoms

Thomas (T) [9] and Fermi (F) [10] independently proposed to treat the electrons in a heavy atom as a Fermi gas in which free electron relations were used locally. Then, with the ground-state density $\rho(\mathbf{r})$ taken as the basic variable of the theory, the ground-state energy E_{TF} can be constructed from their statistical method (valid, as already remarked, for heavy atoms: i.e. an assumed large number of electrons) as

$$E_{TF}=c_k\int\rho(\mathbf{r})^{5/3}\,d\mathbf{r}-\int\frac{Ze^2}{r}\rho(\mathbf{r})\,d\mathbf{r}+\frac{1}{2}e^2\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' \tag{8}$$

where c_k is given from the kinetic energy of a uniform Fermi gas as

$$c_k=\frac{3h^2}{10m}\left(\frac{3}{8\pi}\right)^{2/3}, \tag{9}$$

while Z is evidently the atomic number. Minimizing the energy E_{TF} with respect to the density $\rho(\mathbf{r})$, with the constraint that (with N now denoting the total number of electrons)

$$\int\rho(\mathbf{r})\,d\mathbf{r}=N \tag{10}$$

one finds, with μ introduced as the Lagrange multiplier taking care of the normalization constraint (10) as the density is varied, the Euler equation of the variational problem posed by Eq. (8):

$$\mu_{TF} = \frac{5}{3} c_k \rho(\mathbf{r})^{2/3} - \frac{Ze^2}{r} + e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' . \tag{11}$$

This equation has been solved self-consistently for atomic ions with $N \leq Z$ and one of its principal results, due to Milne [11], is that the ground-state energy of neutral atoms is given by

$$E_{TF}^{N=Z} = -0.77\, Z^{7/3} \left(e^2/a_0\right) . \tag{12}$$

While Eq. (12) represents the correct prediction of the non-relativistic Schrödinger equation as $Z \to \infty$, in the range of the Periodic Table corrections are needed. One is due to the fact that Eq. (11) near the point atomic nucleus assumed, shows that $\rho(\mathbf{r})$ diverges as $r^{-3/2}$ and this is due to neglecting density gradients in the Fermi gas model employed. This, as was shown by Scott [12], corrects [13, 14] Eq. (12) with a term $(1/2)Z^2$. Earlier Dirac had introduced the exchange energy A into the Thomas–Fermi atom, with the result

$$A = -c_X \int \rho(\mathbf{r})^{4/3} d\mathbf{r} \;:\; c_X = \frac{3}{4} e^2 \left(\frac{3}{\pi}\right)^{1/3} , \tag{13}$$

again from Fermi gas theory (exchange hole: see below). Adding A on to E_{TF} in Eq. (8) and again performing the minimization with respect to $\rho(\mathbf{r})$ yields the Euler equation of the Thomas–Fermi–Dirac (TFD) method as

$$\mu_{TFD} = \frac{5}{3} c_k \rho^{2/3} - \frac{Ze^2}{r} + e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \frac{4}{3} c_X \rho^{1/3} . \tag{14}$$

Though Eq. (14) is a result of Dirac's work [15], he then solved the quadratic equation presented for $\rho^{1/3}$.

3.1
Dirac–Slater Exchange Potential

Much later Slater [16] emphasised that one could instead view the exchange term $-(4/3)c_X\rho^{1/3}$ as an exchange potential V_X adding to the Hartree atomic potential

$$-\frac{Ze^2}{r} + e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' . \tag{15a}$$

Thus

$$V_X^{DS} = -\frac{4}{3} c_X \rho^{1/3} \tag{15}$$

where the superscripts DS denote Dirac–Slater. Cortona [17] has 'optimized' such a $\rho^{1/3}$ exchange potential for lighter atoms and Table 1 has been constructed from some of his data, in which he connects with Slater $X\alpha$ theory. While the foundations of the optimized potential method and the Harbola–Sahni ap-

Table 1. Absolute atomic total energies (Ry) calculated by Hartree–Fock (HF), optimized potential (OPM), Harbola–Sahni (HS) and using the self-consistent $X\alpha$ (SC–$X\alpha$) method by Cortona [15] (table constructed from data in Table II of Cortona's paper [17])

Atom	HF	OPM	HS	SC–$X\alpha$
He	5.723	5.723	5.724	5.722
Li	14.865	14.865	14.864	14.863
Be	29.146	29.145	29.142	29.141
B	49.058	49.056	49.052	49.051
C	75.377	75.373	75.370	75.370
N	108.802	108.796	108.792	108.794
O	149.619	149.615	149.610	149.609
F	198.819	198.816	198.810	198.799
Ne	257.094	257.091	257.084	257.066
Na	323.718	323.712	323.702	323.691
Mg	399.229	399.222	399.212	399.201
Al	483.753	483.746	483.736	483.725
Si	577.709	577.701	577.688	577.679
P	681.437	681.426	681.414	681.406
S	795.010	795.000	794.986	794.978
Cl	918.964	918.957	918.940	918.929
Ar	1053.635	1053.623	1053.608	1053.596

proach referred to in the table will be briefly touched on below, suffice it to say that the last 3 columns in the table are based on a 'local' exchange potential $V_X(\mathbf{r})$, in contrast to the Hartree–Fock method which has a non-local or energy-dependent potential. The degree of agreement between the alternative local potential $V_X(\mathbf{r})$ methods is almost fully quantitative, testifying to the utility of Slater's original idea [16] based on the TFD Euler equation (14). Of course, the kinetic energy used in Eq. (14) is transcended (the term $(5/3)c_k\rho(\mathbf{r})^{2/3}$ is an electron gas theory, whereas all the methods in Table 1 correctly calculate this single-particle kinetic energy by fully wave-mechanical theories).

However, correlation effects can alter the physics and chemistry of multicentre molecules and solids, and it is important in the present context therefore to ask first how the Fermi gas model, utilized directly in writing the kinetic energy density $c_k\rho(\mathbf{r})^{5/3}$ in Eq. (8) and the corresponding exchange energy density $-c_X\rho(\mathbf{r})^{4/3}$ in Eq. (13) is modified by electron correlation. The answer, anticipated in the pioneering work of Wigner [18, 19] is dramatic. Electrons eventually localize, due to strong electron-electron repulsions, on the sites of a lattice, forming the so-called Wigner electron crystal [18, 19]. This area has been reviewed by Care and March [20] in considerable detail. However the presentation below follows closely that given subsequently by Senatore and March [21].

4
Homogeneous Electron Assembly: Eventual Localization at Sufficiently Low Density in Wigner Electron Crystal

Related to the work of Overhauser [7]; see also Young [22], and of Kapuy and March already referenced in Sect. 1, in the study of electron correlation, it is natural enough to begin with the homogeneous electron fluid (i.e. the so-called jellium model of a metal) in which electrons repelling Coulombically move in a nonresponsive uniform, positive neutralizing background charge.

Whereas the electron density (now constant: ρ_0 say) was used above, it will be convenient in this section to work with the mean interelectronic separation $r_S a_0$, where a_0 is the Bohr radius h^2/me^2, with m the electron mass. This separation is defined precisely by

$$\rho_0 = \frac{3}{4\pi r_S^3 a_0^3} \,. \tag{16}$$

The jellium model just defined affords, in fact, the basis for the so-called theory of the inhomogeneous electron gas [23], which had its origin in the Thomas–Fermi–Dirac method already discussed above. This latter theory was formally completed by the Hohenberg–Kohn theorem [24]. This states that the ground-state energy of an inhomogeneous electron fluid is a unique functional of its electron density $\rho(\mathbf{r})$. Unfortunately, however, the functional is not known. There are a number of reviews already available both on the electron gas [25–27] and on density functional theory [28–35]. Thus the discussion below can be kept relatively brief, focusing in particular on the manner in which Wigner electron crystallization can be treated within the density-functional framework.

Therefore, in what follows the focus will be on the regime of extremely low densities $\rho_0 \to 0$, or equivalently from Eq. (16), $r_S \to \infty$, leaving aside, in the main, the range of high and real metallic densities (for simple s–p metals r_S lies between 2 and 5.5). Here, let us recall though that at high densities, i.e. $r_S \to 0$, the kinetic energy dominates the potential energy and an independent-electron description affords a useful starting point to describe the electron assembly. Indeed, a single plane-wave Slater determinant of spin orbitals is a Hartree–Fock solution for the jellium model yielding a homogeneous density distribution [25] and a ground-state energy per electron given by

$$\frac{E_{HF}}{N} = \left[\frac{2.21}{r_S^2} - \frac{0.916}{r_S}\right] Ry \,. \tag{17}$$

The Thomas–Fermi kinetic energy density $c_k \rho(\mathbf{r})^{5/3}$ derives directly from the first term on the RHS of Eq. (17), the Dirac exchange energy density $-c_X \rho(\mathbf{r})^{4/3}$ coming from the second term. Many-body perturbation theory on this state, in which electrons are fully delocalized, yields a precise result [36, 37] for the correlation energy E_C in the high-density limit as $A \ln r_S + B$, where for present purposes the correlation energy is defined as the difference between the true

ground-state energy E_0 and its Hartree–Fock approximation E_{HF}, namely $E_C = E_0 - E_{HF}$.

4.1 Crystallization (Wigner) of Electrons at Low Density

In the opposite, extremely low-density limit $r_S \to \infty$, it was Wigner [18, 19] who stressed that the above delocalized picture failed completely and that, once the potential energy became large in comparison with the kinetic contribution, the electrons would want to avoid one another maximally. Wigner emphasized that this situation would be reached by electrons becoming localized on the sites of a crystal lattice. He then argued that one must find the stable structure by minimizing the Madelung energy. Of the lattices so far studied, the body-centred-cubic (bcc) structure has the lowest Madelung contribution. This then yields, again for the ground-state energy per electron

$$\lim_{r_s \to \infty} \frac{E}{N} = -\frac{1.792}{r_S} Ry . \tag{18}$$

This demonstrates, when compared with the Hartree–Fock plane wave result (17) that this latter approximation is no longer useful as $r_S \to \infty$, the energy being too high by a factor of about 2. Another way of saying this is that in this extremely low density limit the correlation energy is about equal to the exchange energy in Eq. (17): an evidently strongly correlated assembly leading to electron localization.

Though the above argument can leave no doubt that in the jellium model there will be a localized assembly of electrons, i.e. a Wigner crystal, in the extremely low density limit, the actual analytic calculation of when the electron liquid, at absolute zero of temperature, freezes as the density is lowered has proved very delicate [20]. Eventually, this matter was settled using quantum Monte Carlo computer simulation by Ceperley and Alder [38]. They found in this way that the crystallization first occurred at $r_S \cong 100$. Herman and March [39] subsequently pointed out that, for the Wigner crystal phase, the theoretical expression [40, 41]

$$E(r_S) = \frac{-1.79186}{r_S} + \frac{2.65}{r_S^{3/2}} - \frac{0.73}{r_S^2} Ry \tag{19}$$

without any adjustment at all, reproduces the energies to within 0.124, 0.003, 0.002 and 0.006 mRy respectively. The uncertainties in the corresponding computer calculations are 0.010, 0.003, 0.002 and 0.001 mRy. Prior to the work of Herman and March [39] the Ceperley–Alder ground-state energy data for the electron fluid in the jellium model had been accurately fitted, over the entire range of rS studied, by Vosko et al. [42, 43] and by Perdew and Zunger [44]. Widespread use of these fits has subsequently been made in DFT within the so-called local density approximation (LDA). The density-functional treatment of Wigner crystallization by Senatore and Pastore [45] will be summarized below.

4.2
Density-Functional Treatment of Wigner Crystallization

Let us summarize the underlying assumption of the local density approximation (LDA) as follows:

$$E_{XC}^{LDA}[\rho] = \int \varepsilon_{XC}(\rho(\mathbf{r}))\, d\mathbf{r}\,. \tag{20}$$

The unknown exchange-correlation energy functional $E_{XC}[\rho]$ is thereby approximated in terms of the exchange-correlation energy for the particle, $\varepsilon_{XC}(\rho_0)$ of the electron fluid discussed above at uniform density ρ_0. $\varepsilon_{XC}(\rho_0)$ is accurately known, as referenced earlier.

Of the situations where the above theory, based on the use of uniform electron-gas relations locally, is too crude, the electron Wigner crystal at zero temperature (i.e.in the completely degenerate limit) constitutes one example. This has been explored in the work of Senatore and Pastore [45].

To perform such calculations, one needs the response function $\chi(q)$ of the homogeneous interacting electron liquid: a quantity that is not known exactly. Therefore Senatore and Pastore [45] have employed the so-called STLS (Singwi, Tosi, Land, Sjölander) decoupling scheme [46] to generate $\chi^{-1}(q)$ from the structure factor $S(q)$ obtained from quantum Monte Carlo (QMC) simulations.

Senatore and Pastore [45] study, in fact the coexistence between a fully spin-polarized liquid and a regular crystalline phase. After demonstrating the inadequacy of the LDA (Eq. 20) they propose an approximation for exchange plus correlation in which one considers, rather than the full functional $E_{XC}[\rho]$, the difference

$$\Delta = E_{XC}\left[\rho_S\right] - E_{XC}\left[\rho_0\right] \tag{21}$$

between the solid and the liquid phase. They then appeal to an expansion of Δ about the liquid, in powers of the density difference

$$\rho_Q(\mathbf{r}) = \rho_S(\mathbf{r}) - \rho_0\,. \tag{22}$$

To second order their expansion yields

$$E_{XC}\left[\rho_S\right] = E_{XC}\left[\rho_0\right] + \frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}'\left[-\chi^{-1}(\mathbf{r}-\mathbf{r}') + \chi_0^{-1}(\mathbf{r}-\mathbf{r}')\right]\rho_Q(\mathbf{r})\rho_Q(\mathbf{r}')\,. \tag{23}$$

In Eq. (23), $\chi(r)$ and $\chi_0(r)$ are the static response function of the homogeneous liquid and the response function of the noninteracting electrons (namely the Lindhard function [47]).

The calculation of the ground-state energy of the Wigner electron crystal necessitates the self-consistent solution of the Slater–Kohn–Sham equations for the Bloch orbitals of a single fully occupied energy band, since there is one electron per unit cell and one is concerned with the spin-polarized state [45]. This was accomplished by standard computational routines for energy band-

structure studies. The conclusions of Senatore and Pastore can be summarized as in the following.

The quadratic approximation (23) leads to freezing into the bcc lattice at $r_S = 102$, in good agreement with QMC simulations [38]. The Lindemann ratio γ (rms-deviation about the lattice site divided by the nearest-neighbour separation of the vibrating electron crystal) of 0.34 was obtained [45], where QMC suggests that $\gamma = 0.30 \pm 0.02$ for all quantum assemblies investigated to date. The calculated electron density still proves to be quite localized. This, together with the high symmetry of the periodic lattice, prompted Senatore and Pastore [45] to consider a possibility of a tight-binding approximation in which Bloch orbitals are built from one Gaussian orbital [33] per site with a width determined variationally. By this approximation considerable simplification is achieved, whereas the results for the r_S value for freezing and the Lindemann ratio γ change only slightly. For further details the reader is referred to the original paper [45].

5
Localized versus Molecular Orbital Theories of Electrons: Example of Stretched H_2 Molecule

Correlation in atoms introduces largely quantitative changes. As seen above, qualitative changes can come about in 'condensed' phases. Here, we consider briefly the case of the stretched H_2 molecule, going back to the work of Coulson and Fischer [48] (CF). Their work is a forerunner of the so called generalized valence bond theory.

Suppose instead of atomic orbitals (say H atom 1*s* orbitals in the simplest case) ϕ_a on proton a and $\phi - b$ on proton b, one formed asymmetric orbitals $\phi_a + \lambda\phi_b$, $0 < \lambda \leq 1$ and $\phi_b + \lambda\phi_a$, the former representing, with $\lambda < 1$, the electron primarily but not wholly belonging to nucleus a etc. Then CF formed the (unsymmetrized) variational wave function

$$\Psi_{Coulson-Fischer} = [\phi_a(1) + \lambda\phi_b(1)]\,[\phi_b(2) + \lambda\phi_a(2)]\ . \tag{24}$$

If H denotes the total Hamiltonian of the H_2 molecule, CF found λ as a function of the internuclear distance R by minimizing $< H >$ with respect to the wave function (24). Their findings were: (a) for $R < 1.6\,R_{equilibrium}$, $\lambda = 1$; (b) for $R > 1.6\,R_{equilibrium}$, λ falls quite rapidly to zero as the H_2 molecule is stretched further. For $\lambda = 1$, Eq. (24) is the molecular orbital wave function built as a linear combination of atomic orbitals, whereas for $R > 1.6\,R_{equilibrium}$ one sees that electrons quickly 'go back on to their own atoms'.

It is relevant in this context that, in work on the cleavage force of crystalline Si, Matthai and March [49] have used a coordination-dependent force field due to Tersoff [50] to study related 'bond-breaking' in a condensed phase. Tersoff's force field was fitted to available density-functional calculations, With such a force law, Matthai and March [49] concluded that the Si bond is comparable in

its 'elasticity', but now in the condensed covalent network of solid Si, with the H_2 molecule in free space. Mechanical properties of solids appears therefore to be an area where electron correlation may play an important role, especially in systems such as Si or *d*-electron metals, with rather well defined directional bonding.

Let us conclude this section by returning to the electron density from the Coulson–Fischer wave function. Writing a normalization factor $\aleph$ on the RHS of Eq. (24), one finds

$$\begin{aligned}\rho_{CF}(1,R) &= \int \Psi^2_{Coulson-Fischer}(1,2)\, d\mathbf{r}_2 \\ &= \aleph^2 \left[\phi_a(1) + \lambda(R)\phi_b(1)\right]^2 \left[1 + 2\lambda S(R) + \lambda^2\right] . \end{aligned} \tag{25}$$

Here $S(R)$ is the overlap integral $\int \phi_a \phi_b d\mathbf{r}$ and ϕ_a and ϕ_b are taken as normalized to unity. While S is a smooth function of R, the CF calculations show that a derivative discontinuity exists in $\lambda(R)$ at $R = 1.6\, R_{equilibrium}$. Simple approximations in density functional theory do not reproduce this type of 'bond-breaking' behaviour.

6
Some Remarks on Low-Order Density Matrices and Electron Localization

Let us briefly return to the jellium model, but in connection now with low-order density matrices. Young and March [3] exploited in early work the fact that in this model the energy of the uniform electron fluid can be specified completely in terms of its second-order density matrix (2 DM). Mayer [1] had therefore suggested that trial density matrices satisfying all the usual physical conditions might be used to determine variationally correlation energies and pair functions. It subsequently emerged that the particular choice made by Mayer did not satisfy all the Pauli conditions on the 2 DM. However, the variational 2 DMs set up by Young and March [3], and designed to embrace both the homogeneous electron liquid and the broken symmetry phase on freezing into the Wigner electron crystal, were satisfactory from the above point of view and these workers investigated the consequences of using such forms. Since one the matrices which was proposed [3] was a generalization of the Hartree–Fock expression, a description of correlation effects was assured and it appeared that this should be useful for all densities. Later in the chapter, the relation of such a 2 DM to an antisymmetric N-electron total wave function will be formally discussed and it will then be clear that the 2 DM of Young and March, especially in the localized electron regime $r_S >\cong 100$ of the Wigner crystal, is worthy of further investigation.

Turning from the jellium model to electron correlation in molecules, it is worth noting here that Klein and March [51] have reopened the long-standing question of the relation between first- and second-order density matrices in relation to the approximation of perfect pairing and natural orbitals. Using per-

fect pairing, the examples of H_2, already treated above, and the H_2 dimer are employed to motivate a more formal treatment of $(N/2)$ pairs. For the H_2 dimer and the general case of $(N/2)$ pairs, the strong orthogonality constraint discussed in Sect. 2 is again invoked. A characteristic of the treatment of Klein and March [51] is that natural orbital and natural geminal expansions of first- and second-order density matrices respectively are reduced from generally infinite to finite sums within the framework adopted [51]. As in the complementary treatment of jellium set out immediately above, the procedure of Klein and March allows an approximation transcending Hartree–Fock theory, and in their case a relation between first- and second-order density matrices can be exhibited.

This is the point at which to demonstrate the intimate connection between density functional and density matrix theory. The link will be forged by expressing the exchange-correlation potential $V_{XC}(\mathbf{r})$ of DFT in terms of the 1 and 2 DMs of the atom, molecule, or finite cluster under consideration. The result for $V_{XC}(\mathbf{r})$ in terms of low-order DMs is due to Holas and March [52].

7
Exchange-Correlation Potential of Density Functional Theory in Terms of First- and Second-Order Density Matrices

To motivate the approach of Holas and March [52], let us briefly set out the early work of March and Young [53], who set up the so-called differential virial theorem for independent fermions moving in 1 dimension (x) only in a common potential energy $V(x)$. If $t(x)$ denotes the kinetic energy per unit length at position x, their result was

$$\frac{\partial t}{\partial x} = -\frac{1}{2}\rho(x)\frac{\partial V}{\partial x} + \frac{1}{8}\rho'''(x) \tag{26}$$

That this is indeed the differential form of the customary virial theorem is readily seen by multiplying Eq. (26) throughout by x and then integrating over all x from $-\infty$ to $+\infty$. Some elementary integrations by parts recovers the usual (integral) virial theorem of Clausius, in, of course, now fully quantum-mechanical form [54].

To make contact between Eq. (26) and the study of Holas and March [52], let us rewrite the equation as an expression for the force $-\partial V/\partial x$, to find

$$-\frac{\partial V}{\partial x} = \frac{2}{\rho(x)}\frac{\partial t}{\partial x} - \frac{1}{4}\frac{\rho'''(x)}{\rho(x)}\,. \tag{27}$$

This Eq. (27) may be viewed as a 'force balance' equation, the total force F on the LHS being 'balanced' by a kinetic (K) piece F_K plus a contribution which, in three-dimensions, involves the Laplacian (L) of the electron density $\rho(x)$, and the corresponding force we shall write as F_L. Thus Eq. (27) has the form

$$F = F_K + F_L\,. \tag{28}$$

The first step in the study of Holas and March [52] was to generalize this Eq. (28) to fully interacting electrons moving in three dimensions.

7.1 Force Balance Equation for External Potential

If the external potential energy ($-Ze^2/r$ in an atom having nuclear charge Ze) is written as V_{ext}, then their generalization of Eq. (28) is a vector force balance equation:

$$\mathbf{F}_{ext} = -\frac{\partial V_{ext}}{\partial \mathbf{r}} = \mathbf{F}_K + \mathbf{F}_L + \mathbf{F}_{ee} \tag{29}$$

where $\mathbf{F}_{ee}$ represents the force term from the full electron-electron interactions. However, only one of the three terms on the RHS of Eq. (29) can be written in terms of the ground-state electron density $\rho(\mathbf{r})$, namely $\mathbf{F}_L$, which is given by [52]:

$$\mathbf{F}_L = -\frac{1}{4}\frac{\nabla\nabla^2\rho(\mathbf{r})}{\rho(\mathbf{r})}\,; \tag{30}$$

which evidently is the 3-dimensional generalization of the last term on the RHS of Eq. (27).

If we denote the electron-electron Coulombic interaction $e^2/|\mathbf{r} - \mathbf{r}'|$ by $u(\mathbf{r}, \mathbf{r}')$, then it is not difficult to show [41] that the electron-electron contribution $\mathbf{F}_{ee}(\mathbf{r})$ in Eq. (29) involves the pair correlation function, say $n_2(\mathbf{r}, \mathbf{r}')$, which is the diagonal element of the second order density matrix $\gamma_2(\mathbf{r}, \mathbf{r}', \mathbf{r}, \mathbf{r}')$ through

$$\mathbf{F}_{ee} = 2\frac{\int \nabla u(\mathbf{r}, \mathbf{r}')n_2(\mathbf{r}, \mathbf{r}')\, d\mathbf{r}'}{\rho(\mathbf{r})}\,. \tag{31}$$

Evidently the kinetic term $\mathbf{F}_K$ which remains in Eq. (29) must be determined by the fully interacting first-order density matrix $\gamma_1(\mathbf{r}, \mathbf{r}')$. Holas and March [52] express $\mathbf{F}_K$ in terms of a vector field $\mathbf{z}(\mathbf{r})$ through

$$\mathbf{F}_K = \frac{\mathbf{z}(\mathbf{r})}{\rho(\mathbf{r})}\,. \tag{32}$$

This field is related to the kinetic-energy density tensor by

$$z_\alpha(\mathbf{r}) = 2\sum_\beta \frac{\partial}{\partial r_\beta} t_{\alpha\beta}(\mathbf{r})\,. \tag{33}$$

The final step here is then to relate the kinetic-energy density tensor $t_{\alpha\beta}$ to the fully interacting 1 DM $\gamma_1(\mathbf{r}, \mathbf{r}')$ through [55]

$$t_{\alpha\beta}(\mathbf{r}) = \frac{1}{4}\left(\frac{\partial^2}{\partial r'_\alpha \partial r''_\beta} + \frac{\partial^2}{\partial r'_\beta \partial r''_\alpha}\right)\gamma_1(\mathbf{r}', \mathbf{r}'')|_{\mathbf{r}'=\mathbf{r}''=\mathbf{r}}\,. \tag{34}$$

This is a real symmetric tensor, the trace of which is the non-negative kinetic-energy density

$$t(\mathbf{r}) = \sum_{\alpha} t_{\alpha\alpha}(\mathbf{r}) \geq 0 \tag{35}$$

leading to the global kinetic energy

$$T = \int t(\mathbf{r}) d\mathbf{r} \,. \tag{36}$$

7.2 Gradient of Exchange-Correlation Potential $V_{XC}(\mathbf{r})$

The next step in the Holas–March study was to form from the above, formally exact, theory based on the many-electron Schrödinger equation, the gradient of the exchange-correlation potential energy $V_{XC}(\mathbf{r})$. This is the many-electron part of the one-body potential to be inserted in the one-body Schrödinger equations (the so-called Slater–Kohn–Sham (SKS) equations)

$$\nabla^2 \psi_i(\mathbf{r}) + \frac{2m}{\hbar^2} \left[\varepsilon_i - V(\mathbf{r})\right] \psi_i = 0 \tag{37}$$

which purport to generate the exact ground-state density $\rho(\mathbf{r})$ as

$$\rho(\mathbf{r}) = \sum_{i}^{occupied\ states} \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}) \,, \tag{38}$$

This is, of course, a formal statement, in that $V_{XC}(\mathbf{r})$ enters $V(\mathbf{r})$ through

$$V(\mathbf{r}) = V_{ext} + e^2 \int \frac{\rho(\mathbf{r}')}{\left|\mathbf{r} - \mathbf{r}'\right|} d\mathbf{r}' + V_{XC}(\mathbf{r}) \,. \tag{39}$$

In terms of the exchange-correlation energy functional $E_{XC}[\rho]$, V_{XC} is the functional derivative:

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}\left[\rho\right]}{\delta \rho(\mathbf{r})} \,. \tag{40}$$

Clearly, if an approximation is made to $E_{XC}[\rho]$ (which, as discussed further below, must have in it fingerprints of the antisymmetry of the N-electron fully interacting wave function), this approximation will be exacerbated by the functional differentiation required in Eq. (39).

One of the considerable merits of the Holas–March theory of $V_{XC}(\mathbf{r})$, to be completed immediately below, is that the functional derivative in Eq. (40) is entirely bypassed. The price paid, of course, is that one needs through Eq. (31) the pair function $n_2(\mathbf{r}, \mathbf{r}')$ of the fully interacting system, and the near-diagonal behaviour of the 1 DM $\gamma_1(\mathbf{r}, \mathbf{r}')$.

One can immediately apply the above force balance arguments to the single-particle SKS force $-\partial V(\mathbf{r})/\partial \mathbf{r} \equiv \mathbf{F}_S$. Then

$$\mathbf{F}_S = \{-\frac{1}{4} \nabla \nabla^2 \rho(\mathbf{r}) + \mathbf{z}_S(\mathbf{r})\} / \rho(\mathbf{r}) \tag{41}$$

Hence the desired gradient of the exchange-correlation potential $\mathbf{F}_{XC} = -\partial V_{XC}/\partial \mathbf{r}$ is given by

$$\mathbf{F}_{XC} = \frac{\{z(\mathbf{r}) - z_S(\mathbf{r}) + \int d\mathbf{r}' \left[\nabla u(\mathbf{r}, \mathbf{r}')\right] \left[\rho(\mathbf{r})\rho(\mathbf{r}') - 2n_2(\mathbf{r}, \mathbf{r}')\right]\}}{\rho(\mathbf{r})}. \quad (42)$$

Thus, with $\mathbf{F}_{XC}$ given in Eq. (42) V_{XC} is determined as the line integral:

$$V_{XC}(\mathbf{r}_0) = -\int_{\infty}^{\mathbf{r}_0} d\mathbf{r}.\mathbf{F}_{XC} \quad (43)$$

Eqs. (42) and (43) embody the Holas–March results for the exact exchange-correlation potential in terms of low-order density matrices; both fully interacting and non-interacting matrices built from SKS one-electron wave functions being involved.

7.3 Generalizations and Separation of $V_{XC}(\mathbf{r})$ into Exchange and Correlation Contributions

Subsequent to the study of Holas and March summarized above, Levy and March [56] have effected a generalization which permits a formulation as a function of the electron-electron repulsion coupling constant λ. As will be discussed below, the purpose of their generalization is to allow $V_{XC}(\mathbf{r})$ to be separated into exchange and correlation contributions. This is achieved by then developing relations which are associated with each order in λ. The first-order development yields a formal expression for the exact exchange potential $V_X(\mathbf{r})$, given by

$$V_X = \frac{\delta E_X[\rho]}{\delta \rho(\mathbf{r})}. \quad (44)$$

Holas and March [52] had earlier pointed out that a zero-order approximation to their exact line-integral form (43) was the Harbola–Sahni [57] approximation to V_X. The exchange expression of Levy and March [56] identifies the correction to the Harbola–Sahni approximation to V_X. It is thereby shown that this correction, which is small in atoms, is zero if the Hartree–Fock single determinant, for the given density, is identical to the SKS single determinant for the same density. From the work of Görling and Ernzerhof [58], it is known that these determinants are generally quite close in atoms.

Higher-order terms in the expansion in the coupling constant λ yield formal expressions for the corresponding parts of the exact correlation potential:

$$V_C(\mathbf{r}) = \frac{\delta E_C[\rho]}{\delta \rho(\mathbf{r})}. \quad (45)$$

By way of motivation, it is to be noted [56] that it is especially important to have knowledge of a separate exact expression for V_C when one employs an approximation to V_X which does not include line integrals or if one attaches

an approximation for V_C to an exact optimized effective potential (OEP) study or to a Hartree–Fock calculation. Consequently Levy and March [56] derive an explicit formal expression for V_C, which is given below.

The idea in Ref. [56] is then to expand the pair function $n_2^\lambda(\mathbf{r}, \mathbf{r}')$ for coupling constant λ around the non-interacting SKS point $\lambda = 0$: i.e.

$$n_2^\lambda = n_2^0 + \lambda n_{2,1}^c + \lambda^2 n_{2,2}^c + ... \tag{46}$$

with a similar expansion for the 1 DM. Then, at the exchange only level, with superscript *HS* denoting the Harbola–Sahni result:

$$V_X(\mathbf{r}_0) = V_X^{HS}(\mathbf{r}_0) + \int_\infty^{\mathbf{r}_0} d\mathbf{r}.\mathbf{z}(\gamma_1^c; \mathbf{r})\rho^{-1}(\mathbf{r}) \tag{47}$$

where λ_1^C is the order λ contribution to the 1 DM. Correlation is handled in a related way by Levy and March [56] and the reader is referred to their original study for the details.

An alternative approach to V_{XC} using integral equations involving 1–3 DMs has also been developed by Holas and March [59]. The partition of these integral equations for exchange and correlation potentials paralleling the above separation of line integral formulae into V_X and V_C has been given by Holas and Levy [60].

It is also of relevance in the present context to note that the perturbation theory of Görling and Levy [61] has been applied by Holas and March [62] to provide a calculational scheme for exact exchange and correlation potentials based on the equation of motion for the density matrix.

8
Some Simplified Forms of E_{XC} Based on Adiabatic Connection Method

Levy et al. [63] have used the scaling parameter λ (introduced above in the context of the adiabatic connection method) to propose some very practical modifications of existing $E_{XC}[\rho]$ functionals. These workers first point out that the idea of an adiabatic connection is taking on new importance. The earlier references include those of Harris and Jones [64], Langreth and Perdew [65], Gunnarsson and Lundqvist [66], Levy [67] and Harris [68]. The exchange-correlation energy functional $E_{XC}[\rho]$ (following Refs. 65 and 66, and the review in Parr and Yang [32]) may be written as

$$E_{XC}\left[\rho\right] = \int_0^1 d\lambda\, U_{XC}^\lambda\left[\rho\right] \tag{48}$$

where

$$U_{XC}^\lambda\left[\rho\right] = \left\langle \Psi_\rho^\lambda \left| \hat{V}_{ee} \right| \Psi_\rho^\lambda \right\rangle - J\left[\rho\right] \tag{49}$$

and $J[\rho]$ denotes the Coulomb energy. In Eq. (49) Ψ_ρ^λ is the ground-state wave function of a Hamiltonian H_λ which describes a system of electrons in which the electron-electron Coulombic repulsion is scaled:

$$\hat{H}_\lambda = \hat{T} + \lambda \hat{V}_{ee} + \sum_i V_\lambda(\mathbf{r}_i)\,. \tag{50}$$

It also minimizes the constrained search expression [69, 70], for fixed ρ:

$$\left\langle \Psi_\rho^\lambda \left| \hat{T} + \lambda \hat{V}_{ee} \right| \Psi_\rho^\lambda \right\rangle\,. \tag{51}$$

In the above expressions, $\rho(\mathbf{r})$ is to be interpreted as the exact ground-state density of H_λ. The form (51) was anticipated by Percus [71] who gave an expression for the non-interacting kinetic energy functional.

It is next to be noted that Becke [72, 73] has used the above approach in the context of a two-point integration formula. Thus he replaces Eq. (48) by

$$E_{XC}\left[\rho\right] \sim \frac{1}{2}\left(U_{XC}^0 + U_{XC}^1\right) \tag{52}$$

where U_{XC}^0 denotes the exchange energy of the SKS determinant and U_{XC}^1 is the potential energy contribution to the exchange-correlation energy of the fully interacting system. These considerations led Becke to his half-and-half DFT functional: the forerunner of this fruitful three-parameter functional. However, as discussed in Ref. [63], Becke's three-parameter functional [74], in contrast to the half-and-half functional, really results from a partial abandonment of the adiabatic connection idea.

Levy et al. [63] examine adiabatic scaling in the Thomas–Fermi–Dirac method set out earlier in this chapter, and also propose a formula for U_{xc}^1 in Eq. (52). These two aspects will be taken in turn below.

8.1
Electron-Electron Scaling with Constant Density: Thomas–Fermi–Dirac Limit

For the ensuing discussion, let us write the Hamiltonian H_λ in the form

$$\hat{H}_\lambda = \hat{T} + \lambda \hat{V}_{ee} + \sum_i^N V_\lambda(\left[\rho\right];\mathbf{r}_i)\,. \tag{53}$$

The ground-state density ρ in this approach is fixed independently of λ and in atomic units

$$\hat{T} = -\sum_i^N \frac{1}{2}\nabla_i^2\,,\ \hat{V}_{ee} = \sum_{i<j} \frac{1}{r_{ij}}\,. \tag{54}$$

The Thomas–Fermi–Dirac (TFD) ground-state (GS) energy corresponding to different λ and a definite density $\rho_0(\mathbf{r})$, is now introduced as the minimum obtained from

$$E_{GS,\lambda,\rho_0}^{TFD} = \min_\rho \left[\int V_\lambda(\left[\rho_0\right];\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + F_\lambda^{TFD}\left[\rho\right]\right]$$

$$= \int V_\lambda([\rho_0];\mathbf{r})\rho_0(\mathbf{r})d\mathbf{r} + F_\lambda^{TFD}[\rho_0] \tag{55}$$

where

$$F_\lambda^{TFD}[\rho] = c_k \int \rho^{5/3}(\mathbf{r})d\mathbf{r} + \lambda J[\rho] - \lambda c_X \int \rho^{4/3}(\mathbf{r})d\mathbf{r}\,. \tag{56}$$

Here $J[\rho]$ is explicitly

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1\, d\mathbf{r}_2\,. \tag{57}$$

The stationary principle yields

$$V_\lambda([\rho_0];\mathbf{r}) = -\frac{\delta F_\lambda^{TFD}[\rho]}{\delta\rho(\mathbf{r})} \tag{58}$$

evaluated at $\rho = \rho_0$. It follows that

$$\begin{aligned} V_\lambda([\rho_0];\mathbf{r}) \;=\; & -\frac{5}{3}c_k\{\rho_0(\mathbf{r})\}^{2/3} - \lambda\int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \\ & +\frac{4}{3}\lambda c_X\{\rho_0(\mathbf{r})\}^{1/3}\,. \end{aligned} \tag{59}$$

One now observes that in this TFD limit the scaling of the one-body potential V_λ is such that V_λ is linear in λ. One may write the ground-state energy explicitly as

$$\begin{aligned} E_{TFD} \;=\; & -\frac{2}{3}c_k\int \rho_0^{5/3}(\mathbf{r})d\mathbf{r} - \frac{\lambda}{2}\int\frac{\rho_0(\mathbf{r}_1)\rho_0(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1\, d\mathbf{r}_2 \\ & +\frac{1}{3}\lambda c_X\int \rho_0^{4/3}(\mathbf{r})d\mathbf{r}\,. \end{aligned} \tag{60}$$

Evidently again, as was the case with V_λ, $E^{\mathrm{TFD}}_{GS,\lambda,\rho 0}$, the ground-state energy corresponding to electron density $\rho_0(\mathbf{r})$ at a coupling constant λ, is linear in λ.

8.2 Two-Point Adiabatic Connection Formula Revisited

In Ref. [49], a relook was taken at the expression to be used for U^1_{XC} in Eq. (52). Bearing this need for U^1_{XC} in mind, the starting point was the formula [75, 76]

$$U^1_{XC}[\rho] = 2E_{XC}[\rho] - \frac{\partial E_{XC}[\rho_\lambda]}{\partial\lambda}|_{\lambda=1} \tag{61}$$

where

$$\rho_\lambda(\mathbf{r}) = \lambda^3\rho(\lambda r)\,. \tag{62}$$

Eq. (61) is in fact a special case of the result

$$U^\lambda_{XC}[\rho] = 2\lambda E_{XC}[\rho] + \lambda^2\frac{\partial E_{XC}[\rho_{1/\lambda}]}{\partial\lambda}\,. \tag{63}$$

In Eq. (52), one could use the exact formula for U^0_{XC} and could approximate U^1_{XC} using some model, through the use of Eq. (61). The resulting approximation

to E_{XC} will be denoted by $\overline{E}_{XC}$ and is evidently given by

$$\overline{E}_{XC} = \frac{1}{2}U^0_{XC}[\rho] + \frac{1}{2}U^{mod,1}_{XC}[\rho]\,. \tag{64}$$

Here 'mod' refers to some model for U^1_{XC} and

$$U^{mod,1}_{XC}[\rho] = 2E^{mod}_{XC}[\rho] - \frac{\partial E^{mod}_{XC}[\rho_\lambda]}{\partial\lambda}|_{\lambda=1}\,. \tag{65}$$

Combining Eqs. (64) and (65) evidently yields

$$\overline{E}_{XC}[\rho] = \frac{1}{2}U^0_{XC}[\rho] + E^{mod}_{XC}[\rho] - \frac{1}{2}\frac{\partial E^{mod}_{XC}[\rho_\lambda]}{\partial\lambda}|_{\lambda=1} \tag{66}$$

or

$$\overline{E}_{XC}[\rho] = \frac{1}{2}E_X[\rho] + \frac{1}{2}E^{mod}_X[\rho] + E^{mod}_C[\rho] - \frac{1}{2}\frac{\partial E^{mod}_C[\rho_\lambda]}{\partial\lambda}|_{\lambda=1}\,, \tag{67}$$

the requirement having been imposed that

$$E^{mod}_X[\rho_\lambda] = \lambda E^{mod}_X[\rho]\,, \tag{68}$$

which is the correct scaling of exchange and, use having been made of the fact that $U^0_{XC}[\rho]$ equals the exact exchange functional $E_X[\rho]$. Formulae (66) and (67) are the desired generalized linear (two-point) results [63].

8.3 Validity of Two-Point Formula (Linear Approximation)

Following Ref. [63] again, it is useful to examine the validity of the linear approximation when the exact U^λ_{XC} is employed. It is known that $U^\lambda_{XC}[\rho]$ monotonically decreases [75] as λ increases and that [76]

$$U^\lambda_{XC}[\rho] = \lambda U^1_{XC}[\rho_{1/\lambda}] \tag{69}$$

and also [76, 77]

$$U^1_{XC}[\rho_{1/\lambda}] \sim constant\ + \lambda^{-1}U^0_{XC}[\rho] \tag{70}$$

as $\lambda \to 0$. Thus, it can be inferred that $U^\lambda_{XC}[\rho]$ is linear at sufficiently small λ.

To proceed, let us, in Eq. (61), use

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \tag{71}$$

and also recognize that

$$E_X[\rho] = U^0_{XC}[\rho]\,. \tag{72}$$

Furthermore

$$U^1_{XC}[\rho] = E_X[\rho] + E_C[\rho] - T_C[\rho] \tag{73}$$

and one consequently deduces that Eq. (52) implies

$$E_C[\rho] \simeq \frac{1}{2}\left(E_C[\rho] - T_C[\rho]\right)\,. \tag{74}$$

This Eq. (74) would evidently be exact if

$$E_C[\rho] = -T_C[\rho]\,. \tag{75}$$

It turns out, however, that the exact relation connecting $E_C[\rho]$ and $T_C[\rho]$ is [75]

$$E_C[\rho] = -T_C[\rho] + \frac{\partial E_C[\rho_\lambda]}{\partial \lambda} \, |_{\lambda=1} \, . \tag{76}$$

Equation (75) is a useful approximation only if $\partial E_C[\rho_\lambda]/\partial\lambda|_{\lambda=1}$ is small. Some insights are provided in Ref. [63] concerning improvements of LDA but the interested reader is referred to the original study for the details [63].

9
Rosina's Theorem and (Formal) Reconstruction Procedure of *N*-Electron Wave Function from Second-Order Density Matrix

It has been known for a long time, especially from the work of Dirac [15] and of Löwdin [78], that the (now idempotent) 1 DM is sufficient to determine the N-electron wave function for the case of a single Slater determinant. It has been equally clear to many workers in the field that such knowledge of the 1 DM cannot be adequate to reconstruct the N-body wave function for the fully interacting electron system, without appeal to the total Hamiltonian.

Here, attention will be drawn to the important work of Rosina [79] and to the subsequent discussion of his study by Mazziotti [80]. As summarized in Ref. [80], Rosina showed that the ground-state 2 DM for a quantum system completely determines the exact N-electron ground-state wave function without any specific knowledge of the exact Hamiltonian except that it has no more than two-particle interactions. Mazziotti [80] points out that a consequence of this theorem is that any ground-state electronic 2 DM precisely determines within the ensemble N-representable space a unique series of higher p-DMs where $2 < p \leq N$. He asserts that these results provide important justification for the functional description of the higher DMs in terms of the 2 DM.

With the theoretical underpinning provided by Rosina's theorem, Mazziotti [80] proposes a new reconstruction procedure in which the p-DM is generated from the 2 DM by imposing contraction and p-ensemble representability conditions. For a p-DM to be p-representable, Mazziotti stresses that it must be Hermitian, antisymmetric and positive semidefinite. By his procedure he establishes contact with the earlier studies of Valdemoro [81] (see also the following chapter) and of Nakatsuji and Yasuda [82]. His principal conclusion, based on numerical examples, is that

> 'through its direct determination of the 2 DM without the wave function the contracted Schrödinger equation [83, 84], coupled with a reconstruction strategy, provides a fresh path towards the calculation of electron correlation'.

Though different in philosophy from the above, it is relevant here to note that approaches using the diagonal element of the 2 DM have been proposed by Ziesche [85] and by Gonis [86]. In these studies, generalizations of the original density functional theorems have been effected. These approaches, as well as

that based on the contracted Schrödinger equation [87] are well worth further study in the area of strongly correlated electron assemblies [88] and of electron localization

10
Localization of Electrons Aided by Applied Magnetic Fields

In this penultimate section, consideration will be given to the way in which suitable applied magnetic fields can aid electron localization, driven initially by Coulomb repulsion. Then it is natural enough to start with the Wigner electron crystal, discussed in zero magnetic field $\mathbf{B} = 0$ in Sect. 4.

10.1
Magnetically-Induced Wigner Solidification (MIWS)

In 1968, Durkan et al. [89] discussed the localization of electrons in impure semiconductors by a magnetic field. From this idea, in which a Wigner transition which was aided by a magnetic field in *n*-type In Sb at low temperature was proposed, has grown an area termed MIWS (magnetically-induced Wigner solidification).

Though the above proposal was concerned with a dilute three-dimensional electron gas, much of the interest more recently has been focussed on the two-dimensional electron gas that can be produced in a GaAs/AlGaAs heterojunction. The magnetic field $\mathbf{B}$ is applied perpendicular to the plane of the electron gas. That Wigner electron solidification could occur in such a system at a critical magnetic field was demonstrated in the experiments at He^3 temperatures of Andrei et al. [90]. Their idea was that an electron fluid could not sustain shear, whereas an electron solid could. By an ingenious experiment involving magnetophonons, these workers demonstrated that, beyond a critical magnetic field, the electron assembly could sustain shear waves, thus establishing the existence of a Wigner electron solid. Proof of crystallinity experimentally, should it exist, will require diffraction experiments. Wigner solidification was subsequently confirmed by the photoluminescence experiments of Buhmann et al. [91], on the consequences of which the ensuing discussion is focussed.

10.2
Melting Curve of Wigner Electron Solid with and without Magnetic Field

Ferraz et al. [92] treated the melting of the Wigner electron crystal in three dimensions for $\mathbf{B} = 0$. Figure 1 shows two asymptotes which they established in the (T, ρ) plane. As $\rho \to 0$, the Wigner electrons, because of negligible overlap between electrons on different sites, can no longer tunnel and all fingerprints of quantum mechanics are then lost. One is dealing with the non-degenerate limit of the jellium model, or what is called by plasma physicists the one-

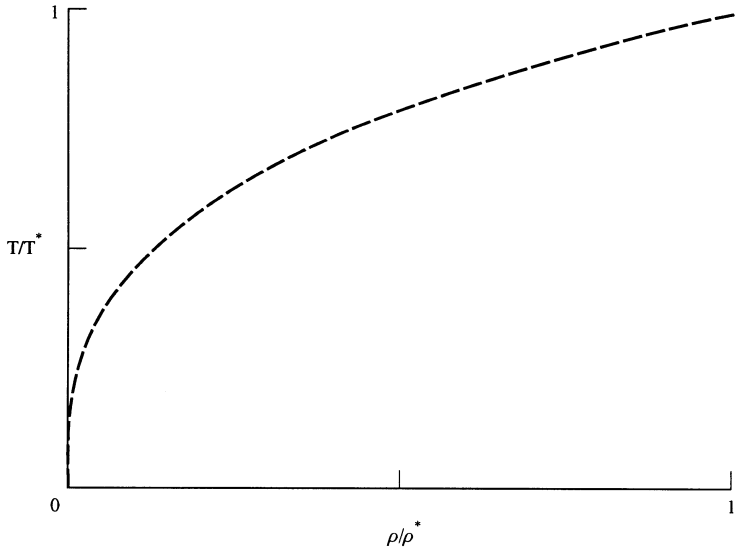

Fig. 1. Reduced melting curve T/T^* (schematic depiction only) of three-dimensional Wigner electron crystal versus reduced density ρ/ρ^*. Upper curve is classical asymptote. Melting curve of quantal crystal is shown coming out from $\rho/\rho^* = 1$, $T = 0$ with vertical slope. The small piece shown will continue steeply upwards until it meets the classical asymptote

component plasma (OCP). It is known from computer simulation, starting with the pioneering study of Brush et al. [93], that this OCP freezes (into a now classical Wigner electron crystal) when the ratio $\Gamma = (e^2/r_S)/k_BT$ reaches a value of ~ 170. This yields the asymptotic $T \propto \rho^{1/3}$ shown schematically in Fig. 1. In Ref. [71], it was also argued that the melting curve came out vertically from the density ρ_C (determined by Ceperley and Alder as corresponding to $r_S \sim 100$) at which the Wigner crystal forms at $T = 0$: i.e. in its ground-state.

One must join these two asymptotes, and in Ref. [92] this was attempted by approximate integration of the appropriate form of the Clausius–Clapeyron equation. However, the definitive work as to how the two asymptotes connect is that of Ceperley et al. [94]. The melting curve of the quantal system closely follows the OCP melting curve $T \propto \rho^{1/3}$ until ρ closely approaches ρ_C and then plunges rather suddenly, consistent with the vertical slope of the melting curve coming away from the point $T = 0$, $\rho = \rho_C$. This situation in Fig. 1 for $\mathbf{B} = 0$ will now be contrasted with that found in the experiments on the two-dimensional electron gas in a strong magnetic field $\mathbf{B}$.

The phase diagram in Fig. 2 is taken from Buhmann et al. [91] and is consistent with their experiments, plus additional theoretical requirements. The most important of these is that the Laughlin liquid is very stable, even down to $T = 0$, for particular values of the Landau level filling factor of the

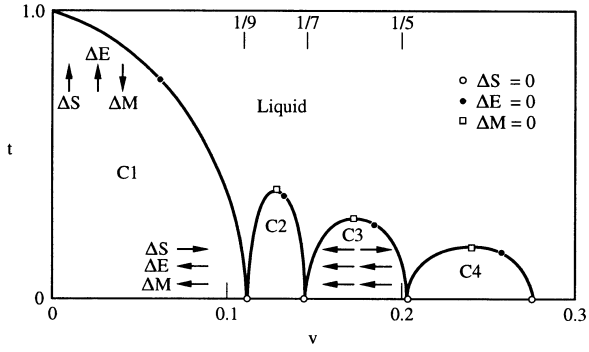

Fig. 2. Schematic phase diagram as proposed by Buchmann et al. [91] for the two-dimensional Wigner electron solidification in a magnetic field perpendicular to electron assembly. Landau filling factor ν is the independent variable [88]. The four Wigner solid phases are labelled C1 to C4. t denotes the reduced melting temperature (compare Fig. 1 which, however, is for three dimensions and zero magnetic field)

form 1/m, where m is an odd integer. As Fig. 2 shows, there are four regions proposed in which the localized Wigner electron solid is stabilized.

Lea, March and Sung [95–97] have analyzed the consequences of such a phase diagram using thermodynamics [98] plus some many-electron concepts such as composite fermions. It would take us too far from our main theme to develop this in detail. However, at the end of the following sub-section, we shall return briefly to the phase-diagram in Fig. 2, but there confining attention to the ground-state, $T = 0$.

10.3
Density Matrices and Density Functionals in Strong Magnetic Field

With a view to future application to localization of electrons already observed in MIWS treated above, this sub-section will summarize the investigation of Holas and March [99] on density matrices and density functionals in strong magnetic fields.

In their study, the equation of motion for the first-order density matrix (1 DM) is first constructed for interacting electrons moving under the influence of applied external scalar and vector potentials. The electron-electron Coulombic repulsion is there the coupling to the 2 DM. This equation of motion is then utilized to establish the differential virial equation for interacting electrons moving in a magnetic field of arbitrary strength. The integral form of the virial theorem, as derived by Erhard and Gross [100] is recovered by suitable integration of this differential form.

The exchange-correlation scalar potential of the current-density functional theory of Vignale and Rasolt [101, 102] is then derived. Again as in the zero

field case treated above, the 1 DM and 2 DM of the fully interacting system, as well as their non-interacting counterparts, are involved. Naturally, for non-zero magnetic field **B**, the vector potentials are also involved. The Hartree–Fock 1 DM for atoms and molecules in magnetic fields is shown in the study of Holas and March [99] to obey the same equation of motion as the fully interacting 1 DM.

If the external scalar potential is denoted by $v(\mathbf{r})$, and the vector potential **A** imposed is related to the application of a strong magnetic field **B**:

$$\mathbf{B} = \operatorname{curl} \mathbf{A} \tag{77}$$

then if T denotes the global kinetic energy the integral virial theorem reads

$$2T[\gamma_1] + E_{ee}[\eta_2] = \int d\mathbf{r} r\rho(\mathbf{r}) \mathbf{r}.\nabla(v(\mathbf{r}) + \frac{e^2}{2mc^2} A^2(\mathbf{r}))) + \int d\mathbf{r}\, \mathbf{j}_p(\mathbf{r}; [\gamma_1])\{\mathbf{r}.\nabla - 1\}\frac{e}{c}\mathbf{A}(\mathbf{r})\,. \tag{78}$$

The kinetic energy T is given in terms of the kinetic energy operator $t(\mathbf{r})$:

$$\hat{t}(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2(\mathbf{r}) \tag{79}$$

and the 1 DM λ_1 by:

$$T[\gamma_1] = \int d\mathbf{r}_1 \hat{t}(\mathbf{r}_1)\gamma_1(\mathbf{r}_1; \mathbf{r}_1^{'})|_{\mathbf{r}_1^{'}=\mathbf{r}_1} = \int d\mathbf{r} \sum_{\alpha} t_{\alpha\alpha}(\mathbf{r}; [\gamma_1]) \tag{80}$$

while the electron-electron interaction E_{ee} is, in terms of the Coulomb interaction u and the pair function n_2:

$$E_{ee}[n_2] = \int d1\, d2 u(12) n_2(12)\,. \tag{81}$$

The gauge-invariant physical current $\mathbf{j}(\mathbf{r})$ is related to the paramagnetic current $\mathbf{j}_p(\mathbf{r})$ entering the above integral virial theorem by

$$\mathbf{j}(\mathbf{r}) = \mathbf{j}_p(\mathbf{r}) + \frac{e}{mc}\rho(\mathbf{r})\mathbf{A}(\mathbf{r})\,. \tag{82}$$

The first achievement of the study of Holas and March [99] is to establish the differential form of the above virial theorem [their Eq. (2.15)]. Again, as in the zero field case treated above, this differential virial theorem is interpreted as a force-balance equation. The well-known Lorentz force of electromagnetism then appears quite naturally in this equation.

One consequence of their equation is that the one-electron result of Amovilli and March [103], namely

$$-\nabla v = \nabla\left(\frac{\hbar^2}{8m}\left(\frac{\nabla\rho}{\rho}\right)^2 - \frac{\hbar^2}{4m}\frac{\nabla^2\rho}{\rho} + \frac{m}{2}\frac{j^2}{\rho^2}\right) \tag{83}$$

is recovered.

Of particular importance for future work on the Wigner electron solid are the forms established by Holas and March for the exchange-correlation potentials. As, however, no such application is available to date, let us conclude this discussion of localization of electrons by applied magnetic fields in referring briefly to the computer simulation study of Ortiz et al. [104] on the stability of the Wigner electron solid relative to the Laughlin liquid, the seat of the fractional quantum Hall effect. This was, of course, already referred to above in the context of the melting curve of the Wigner electron solid in two dimensions.

In the investigation of Ortiz et al. [104], a stochastic method is presented which can handle complex Hermitian Hamiltonians where time-reversal invariance is broken explicitly. These workers fix the phase of the wave function and show that the equation for the modulus can be solved using quantum Monte Carlo techniques. Then, any choice for its phase affords a variational upper bound for the ground-state energy of the system. These authors apply this fixed phase method to the 2D electron fluid in an applied magnetic field with generalized periodic boundary conditions.

Let us turn to summarize the results of applying the above method to study the transition between an incompressible Laughlin liquid and a Wigner solid. Ortiz et al. first note that there are two independent parameters characterizing the 2D electron fluid in the presence of an applied magnetic field. The first is the usual variable r_S, which measures the interelectronic spacing in units of the Bohr radius. In terms of areal density $\rho, r_S = 1/\{\pi\rho\}^{1/2}$. The other parameter is the Landau level filling factor ν. This can be written in terms of the magnetic length l, defined for present purposes by $l = 1/\mathrm{B}^{1/2}$. Then ν is equal to $2e^2/r_S^2$. As Ortiz et al. emphasize, it is the interplay between these two parameters that determines the zero-temperature phase diagram. As ν is decreased, or r_S is increased, one can expect the 2D electron assembly to undergo a first-order transition between a Laughlin liquid and a Wigner solid.

Prior to the fixed phase study of Ortiz et al., the variational Monte Carlo method was used by Zhu and Louie [105] to study the magnetic-field-induced electron localization into a Wigner electron solid in the fractional quantum Hall effect regime. The energy they obtained for this solid phase ($r_S = 20$) compared with the best energy available then for the liquid obtained by Price et al. [106] led Zhu and Louie [105] to conclude that there would be a transition to the Wigner solid for $\nu = 1/5$. The improved energies for the liquid obtained by Ortiz et al. lead to the conclusion 'that further work must be done to show definitively which phase is stable at $\nu = 1/5$.' Perhaps the application of DFT to this magnetically-induced Wigner solid, complementing thereby the work of Senatore and Pastore, can lead to further insight bearing on the $T = 0$ phase diagram of Fig. 2.

11
Summary and Future Directions Involving Electron Localization

In his early researches, Ede Kapuy was already exploiting strongly orthogonal geminals and low-order density matrices (cf. Sect. 2 above). In collaborative work with the present writer, extended systems were also embraced, and much use made of localized Wannier functions, rather than Bloch orbitals: see also March and Stoddart [107]. Though the Hohenberg–Kohn theorem formally completed the Thomas–Fermi–Dirac method, which [cf. Eqs. (8) and (13)] wrote the ground-state energy E quite explicitly in terms of the ground-state density $\rho(\mathbf{r})$, no route exists within the density functional theory itself to determine the exchange-correlation energy function $E_{XC}[\rho]$. The work of Holas and the present writer has, however, expressed the exchange-correlation potential $V_{XC}(\mathbf{r})$ quite explicitly in terms of low-order density matrices.

In jellium, Young and March [3] had already constructed trial density matrices to be used variationally, but many constraints must be applied which lie in the requirement that a trial density matrix describing electrons in molecules and solids must come from some antisymmetric N-electron wave function. This is already appearing as a difficulty with even some of the most successful DFT prescriptions, which are known occasionally to go below an exactly calculated ground-state energy. Obviously, in a variational method, this must reflect the fact that some fermionic constraint has been violated, assuming one's starting point is the exact many-electron Hamiltonian. Of course, as the Thomas–Fermi–Dirac method makes quite clear, the antisymmetry of the N-electron wave function is reflected in the functionals. In the SKS approach [16, 108], the density $\rho(\mathbf{r})$ is N-representable, but the difficulty lies in approximating $E_{XC}[\rho]$ and hence the crucial $V_{XC}(\mathbf{r})$ as its functional derivative. Though $V_{XC}(\mathbf{r})$, in some sense, is an irreducible minimum of information for many-electron theory, it must be tackled from first-principles through hierarchical equations (compare the contracted Schrödinger equation due to Cohen and Frishberg [83] and to Nakatsuji [84]; see also Mazziotti [80] and other references therein, or the reduced Green function approach; see, for example, Kurth [109]).

Key examples of the simplest kind which reflect extreme examples of electron localization are Wigner electron solids, with and without applied magnetic fields, and 'stretched' chemical bonds. An obvious example is the H_2 molecule, where at $1.6\,R_{equilibrium}$ electrons 'go back on to their own atoms'. With directionally bonded solids, such as covalently bonded semiconductors like Si and Ge, one can expect mechanical properties like cleavage force [49] and extended defect propagation [110] (e.g. screw dislocations and also cracks) to involve the breaking of chemical bonds and hence electron correlation and localization. Ede Kapuy was, of course, well ahead of his time in his recognition of the importance of quantitatively dealing with electron localization in both molecules and extended systems. It is a privilege to dedicate this chapter to a firm friend and a fine theoretical scientist.

Acknowledgements. The writer is greatly indebted to Professor A. Holas (Warsaw) for much collaboration and valuable discussions on many aspects covered in this chapter. The opportunity to bring this work to fruition was afforded by a visit to ICTP, Trieste. The author wishes to thank Professor Yu Lu and his colleagues in the Condensed Matter Theory Group for much stimulation and encouragement in the general area of density functional theory and for generous hospitality. Finally, the work surveyed here on matter in intense magnetic fields has been partially supported by ONR. Especial thanks are due to Dr P. Schmidt of that Office for his continuing support.

Appendix: Exchange-Correlation Potential in Exactly Solvable Hooke's Atom

Qian and Sahni [111] have used the Holas–March route to the exchange-correlation potential $V_{XC}(\mathbf{r})$ to construct this quantity in an exactly solvable model. This is the so-called Hooke's atom, in which two electrons, mutually repelling Coulombically, move in an external harmonic potential. For a particular choice of spring constant, the ground-state wave function is analytically known. In the study of Qian and Sahni (see their Fig. 12), the exchange-correlation potential V_{XC} is calculated exactly from the Holas–March [52] approach. In particular, Qian and Sahni make a full study of the contribution to V_{XC} from the correlation kinetic energy.

It is also of importance to record here that the same authors have subsequently considered analytically the asymptotic structure of the correlation-kinetic component of the exchange correlation potential $V_{XC}(\mathbf{r})$ in atoms. Their finding is an r^{-5} decay, with a strength proportional to the product of the square root of the ionization potential and an expectation value associated with the resulting ion. The interested reader is referred to their paper for full details. The final comment in this Appendix is that, for this Hooke's atom, March et al. [112] have achieved one of the long-term aims of DFT by constructing an explicit, third-order differential equation for the ground-state electron density. The shape of this equation has similarities to that for independent electrons moving in a harmonic potential [113].

12 References

1. Mayer JE (1955) Phys Rev 100:1579
2. Tredgold RH (1957) Phys Rev 105:1421
3. Young WH, March NH (1960) Proc Roy Soc (London) A256:62
4. Carlson BC, Keller JM (1961) Phys Rev 121:659
5. Coleman AJ (1963) Rev Mod Phys 35:668
6. Garrod C, Percus JK (1964) J Math Phys 5:1756
7. Overhauser AW (1960) Phys Rev Lett 4:414,462
8. Kapuy E (1958) Acta Phys Acad Sci Hung 9:237
9. Thomas LH (1926) Proc Camb Phil Soc 23:542
10. Fermi E (1928) Z Phys 48:73
11. Milne EA (1927) Proc Camb Phil Soc 23:794

12. Scott JMC (1952) Phil Mag 43:859
13. Ballinger RA, March NH Phil Mag 46:246
14. March NH, Plaskett JS (1956) Proc Roy Soc A235:419
15. Dirac PAM (1930) Proc Camb Phil Soc 26:376
16. Slater JC (1951) Phys Rev 81:385
17. Cortona P (1998) Phys Rev A57:4306
18. Wigner EP (1934) Phys Rev 46:1002
19. Wigner EP (1938) Trans Faraday Soc 34:678
20. Care CM, March NH (1975) Adv Phys 24:101
21. Senatore G, March NH (1994) Rev Mod Phys 66:445
22. Young WH (1961) Phil Mag 6:371
23. See, for example, Lundquist S, March NH (1983) The theory of the inhomogeneous electron gas. Plenum, New York
24. Hohenberg P, Kohn W (1964) Phys Rev 136:B864
25. See, for example, March NH, Young WH, Sampanthar S (1967) The many-body problem in quantum mechanics. Cambridge University Press, Cambridge, UK
26. Singwi KS, Tosi MP (1981) Solid State Phys 36:177
27. Ichimaru S (1982) Rev Mod Phys 54:1017
28. Bamzai AS, Deb BM (1981) Rev Mod Phys 53:95
29. Ghosh SK, Deb BM (1982) Phys Reports 92:1
30. Callaway J, March NH (1984) Solid State Phys 38:136
31. Jones RO, Gunnarsson O (1989) Rev Mod Phys 61:689
32. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, Oxford, UK
33. March NH (1992) Electron density theory of atoms and molecules. Academic, New York
34. Gross EKU, Dreizler RM (1995) Density functional theory NATO ASI Series B Physics vol 337
35. Kryachko ES, Ludena EV (1990) Density functional theory of many-electron systems. Kluwer Press, Dordrecht
36. Macke W (1950) Z Naturforsch 5a:192
37. Gell-Mann M, Brueckner KA (1957) Phys Rev 106:364
38. Ceperley DM, Alder BJ (1980) Phys Rev Lett 45:566
39. Herman F, March NH (1984) Solid State Commun 50:725
40. Coldwell-Horsfall RA, Maradudin AA (1960) J Math Phys 1:395
41. Carr WJ, Coldwell-Horsfall RA, Fein AE (1961) Phys Rev 124:747
42. Vosko SH, Wilk L, Nusair M, (1980) Can J Phys 58:1200
43. See also Painter GS (1981) Phys Rev B24:4264
44. Perdew JP, Zunger A (1981) Phys Rev B23:5048
45. Senatore G, Pastore G (1990) Phys Rev Lett 64:303
46. Singwi KS, Tosi MP, Land RH, Sjölander A (1968) Phys Rev 176:689
47. See, for example, Jones W, March NH (1985) Theoretical Solid State Physics. Dover (Reprint Series), New York, Vol 1
48. Coulson CA, Fischer I (1949) Phil Mag 40:386
49. Matthai CC, March NH (1997) J Phys Chem Solids 58:765
50. Tersoff J (1989) Phys Rev B39:5566
51. Klein DJ, March NH (1995) Theochem (Netherlands) 358:151
52. Holas A, March NH (1995) Phys Rev A51:2040
53. March NH, Young WH (1959) Nuclear Physics 12:237
54. March NH (1996) Molecular Physics 88:1039
55. See also Nagy A, March NH (1997) Molecular Physics 91:597

56. Levy M, March NH (1997) Phys Rev A55:1885
57. Harbola MK, Sahni V (1989) Phys Rev Lett 62:489
58. Görling A, Ernzerhof M (1995) Phys Rev A51:4501
59. Holas A, March NH (1997) Int J Quant Chem 61:263
60. Holas A, Levy M (1997) Phys Rev A56:1031
61. Görling A, Levy M (1993) Phys Rev B47:13105
62. Holas A, March NH (1997) Phys Rev A56:3597
63. Levy M, March NH, Handy NC (1996) J Chem Phys 104:1989
64. Harris J, Jones RO (1974) J Phys F4:1170
65. Langreth DC, Perdew JP (1980) Phys Rev B21:5469
66. Gunnarsson O, Lundquist BI (1976) Phys Rev B13:4274
67. Levy M (1983) In: Keller J, Gasquez JL (eds) Density functional theory. Springer, New York
68. Harris J (1984) Phys Rev A29:1648
69. Levy M (1979) Proc Natl Acad Sci USA 76:6062
70. Levy M (1982) Phys Rev A26:1200
71. Percus JK (1978) Int J Quant Chem 13:89
72. Becke AD (1993) J Chem Phys 98:1372
73. Becke AD (1988) J Chem Phys 88:1053
74. Becke AD (1993) J Chem Phys 98:5648
75. Levy M, Perdew JP (1985) Phys Rev A32:2010
76. Levy M (1991) Phys Rev A43:4637
77. Görling A, Levy M (1992) Phys Rev A45:1509
78. Löwdin PO (1955) Phys Rev 97:1509
79. Rosina M (1968) Reduced density operators with application to physical and chemical systems. In: Coleman AJ, Erdahl RM (eds) Queen's papers in Pure and Applied Mathematics No. 11. Queen's University, Kingston, Ontario
80. Mazziotti DA (1998) Phys Rev A57:4219
81. See, for example, Valdemoro C, Tel LM, Perez-Romero E (1997) Adv Quantum Chem 28:33
82. Nakatsuji H, Yasuda K (1996) Phys Rev Lett 76:1039
83. Cohen L, Frishberg C (1976) Phys Rev A13:927
84. Nakatsuji H (1976) Phys Rev A14:41
85. Ziesche P (1996) Int J Quant Chem 60:1361
86. Gonis A, Schulthess TC, Turchi PEA, van Ek J (1997) Phys Rev B56:9335
87. See also Dawson K, March NH (1984) J Chem Phys 81:5850
88. March NH (1997) Electron correlation in molecules and condensed phases. Plenum, New York
89. Durkan J, Elliott RJ, March NH (1968) Rev Mod Phys 40:812
90. Andrei EY, Deville G, Glattli DC, Williams FIB, Paris E, Etienne B (1988) Phys Rev Lett 60:2765
91. Buhmann H, Joss W, von Klitzing K, Kukuskim IV, Plant AS, Martinez G, Ploos K, Timofeev VB (1991) Phys Rev Lett 6:926
92. Ferraz A, March NH, Suzuki M (1979) Phys Chem Liquids 9:59
93. Brush SG, Sahlin HL, Teller E (1966) J Chem Phys 45:2102
94. Ceperley DM (1998) private communication
95. Lea MJ, March NH, Sung W (1991) J Phys Condens Matter 3:CL4301
96. Lea MJ, March NH (1990) Phys Chem Liquids 21:183
97. Lea MJ, March NH (1991) J Phys Condens Matter3:3493
98. See also, for $\mathbf{B} = 0$, Parrinello M, March NH (1976) J Phys C9:L147
99. Holas A, March NH (1997) Phys Rev A56:4595

100. Erhard S, Gross EKU (1996) Phys Rev A53:R5
101. Vignale G, Rasolt M (1987) Phys Rev Lett 59:2360
102. Vignale G, Rasolt M (1988) Phys Rev B37:10685
103. Amovilli C, March NH (1990) Chem Phys 146:207
104. Ortiz G, Ceperley DM, Martin RM (1993) Phys Rev Lett 71:2777
105. Zhu X, Louie SG (1993) Phys Rev Lett 70:335
106. Price R, Platzman PM, He S (1993) Phys Rev Lett 70:339
107. March NH, Stoddart JC (1968) Repts Prog Physics 31:533
108. Kohn W, Sham LJ (1965) Phys Rev 140:A1133
109. Kurth S (1998) Proc Crete Int Workshop on Electron correlation and material properties (to be published)
110. March NH, Razavy M, Paranjape BV (1998) J Appld Phys, 84:209
111. Qian Z, Sahni V (1998) Phys Rev A57:2527
112. March NH, Gal T, Nagy A (1998) Chem Phys Lett (in press)
113. Lawes GP, March NH, (1979) J Chem Phys 71:1007

Author Index Volume 201–203

Author Index Vols. 26–50 see Vol. 50
Author Index Vols. 51–100 see Vol. 100
Author Index Vols. 101–150 see Vol. 150
Author Index Vols. 151–200 see Vol. 200

The volume numbers are printed in italics