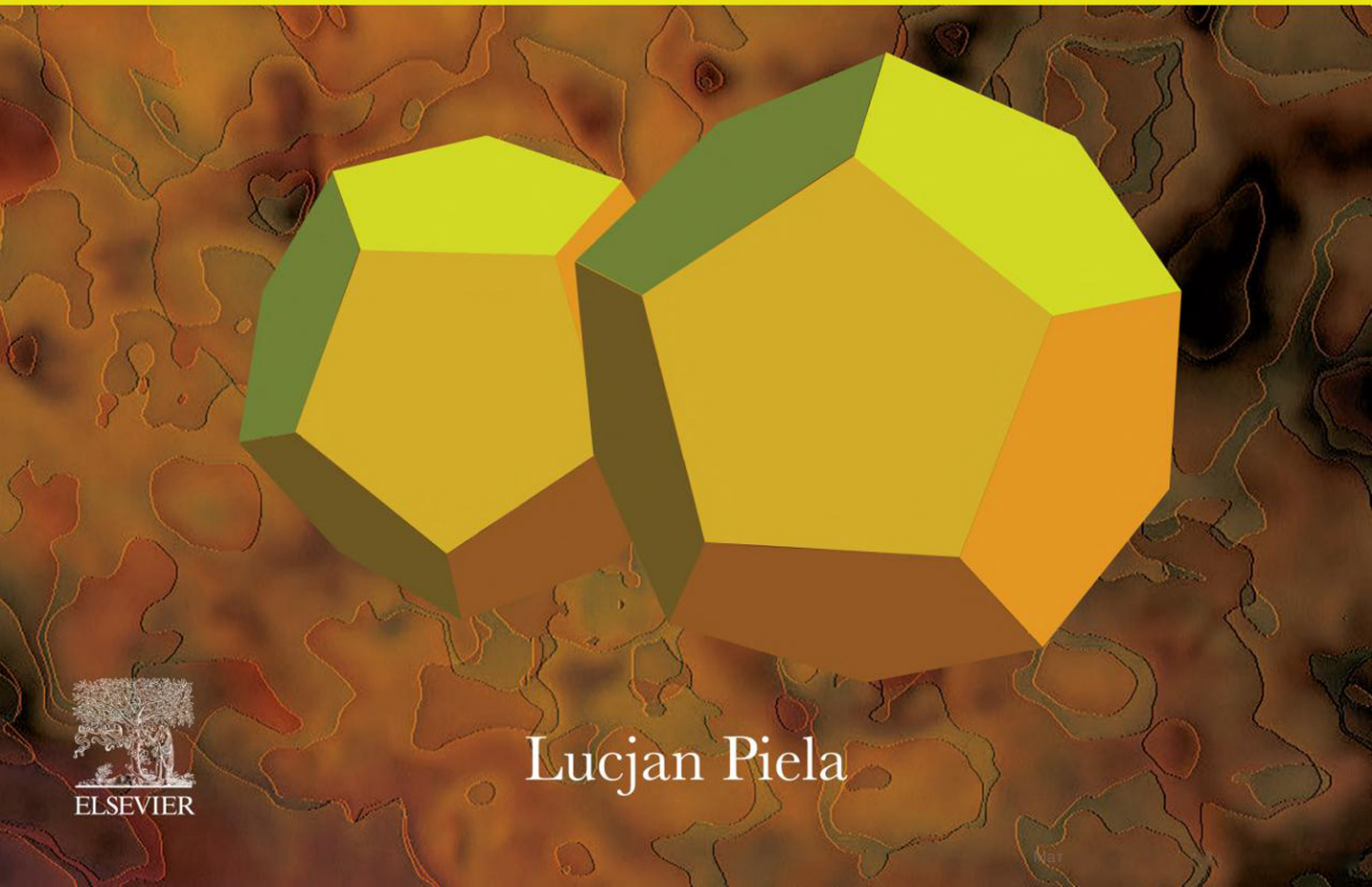
Third Edition

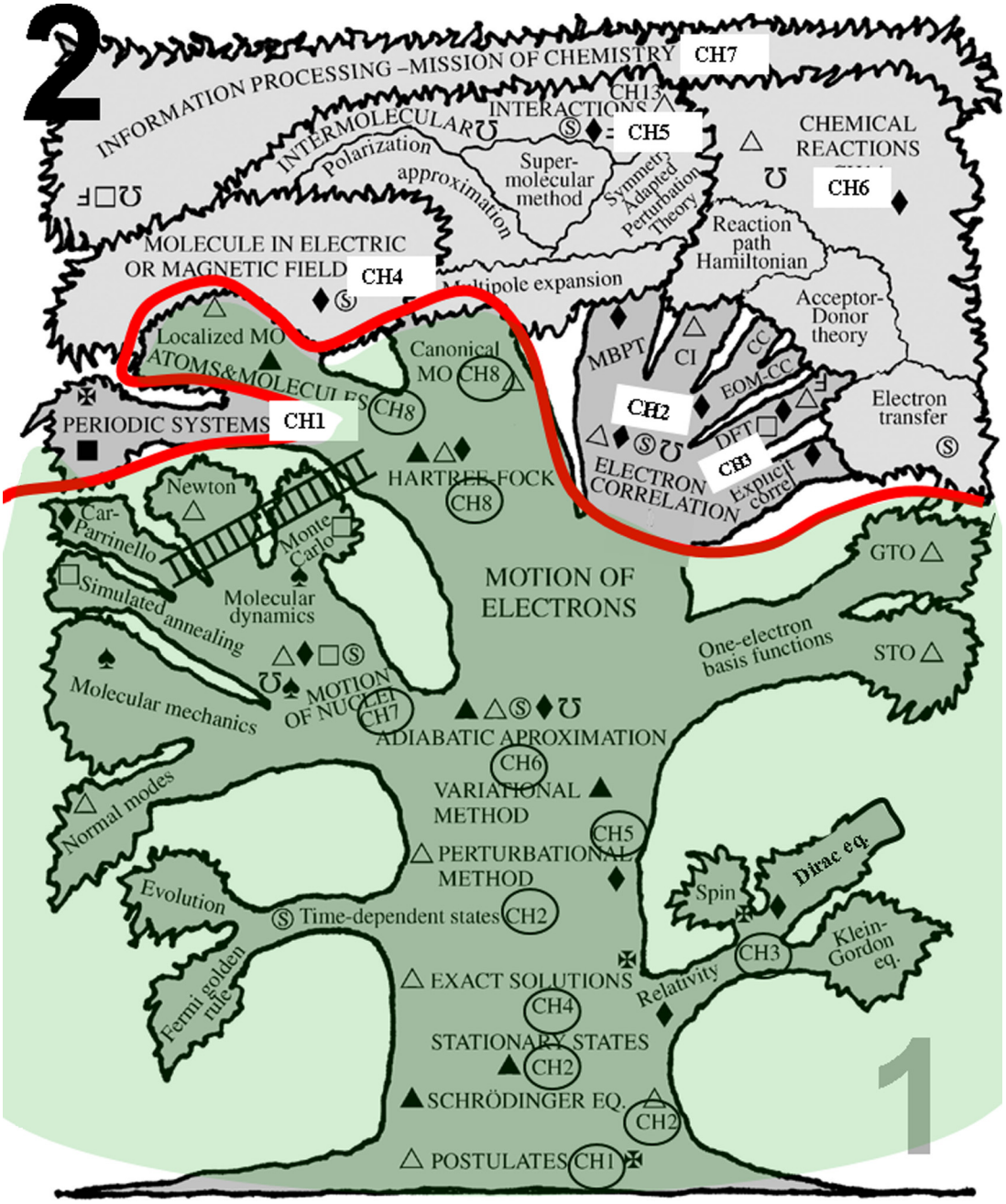
IDEAS OF QUANTUM CHEMISTRY

Volume 2: Interactions



Lucjan Piela

ELSEVIER


2
INFORMATION PROCESSING –MISSION OF CHEMISTRY
CH7
INTERMOLECULAR INTERACTIONS
CH15
CH5
Polarization approximation
Super-molecular method
Symmetry Adapted Perturbation Theory
CHEMICAL REACTIONS
CH6
Reaction path Hamiltonian
MOLECULE IN ELECTRIC OR MAGNETIC FIELD
CH4
Multipole expansion
Acceptor-Donor theory
Localized MO
ATOMS&MOLECULES
Canonical MO
CH8
MBPT
CI
CC
EOM-CC
PERIODIC SYSTEMS
CH1
CH2
DFT
Electron transfer
CH3
ELECTRON CORRELATION
Explicit correl.
HARTREE-FOCK
CH8
Newton
Car-Parrinello
Monte Carlo
Simulated annealing
Molecular dynamics
MOTION OF ELECTRONS
GTO
One-electron basis functions
STO
Molecular mechanics
MOTION OF NUCLEI
CH7
ADIABATIC APROXIMATION
CH6
VARIATIONAL METHOD
Normal modes
CH5
PERTURBATIONAL METHOD
Dirac eq.
Spin
Evolution
Time-dependent states
CH2
Klein-Gordon eq.
CH3
Relativity
Fermi golden rule
EXACT SOLUTIONS
CH4
STATIONARY STATES
CH2
SCHRÖDINGER EQ.
CH2
POSTULATES
CH1
1

WHERE ARE WE?

Compose your own book according to your needs

Choose your own path through the TREE (all paths begin at the basis of the TREE). You may also consider the author's recommendations with two basic paths:

- *minimum minimorum*, with the sign ▲, for those who want to proceed as quickly as possible to get an idea of what quantum chemistry is all about,
- *minimum*, with the signs ▲ *and* △, for those who seek basic information about quantum chemistry,

as well as other paths, which consist of the *minimum path*, i.e., ▲ and △, *and* special excursions (following the corresponding flags, please note the chapter numbering shown as, e.g., CH1 for Chapter 1, etc.) into the subjects related to:

- *large molecules* (□),
- *molecular mechanics and molecular dynamics* (♠),
- *solid state chemistry/physics* (■),
- *chemical reactions* (℧),
- *spectroscopy* (Ⓢ),
- *exact calculations* on atoms or small molecules (⧫),
- *relativistic and quantum electrodynamics effects* (►),
- *most important computational methods of quantum chemistry* (♢),
- *the future of quantum chemistry* (⊣), and
- *"magical" aspects of quantum physics* (✠).

Ideas of Quantum Chemistry

Volume 2: Interactions

Third Edition

Things appear, ideas persist
Plato

Ideas of Quantum Chemistry

Volume 2: Interactions

Third Edition

Lucjan Piela

Department of Chemistry
University of Warsaw
Warsaw, Poland

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2020 Elsevier B.V. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-444-64248-6

For information on all Elsevier publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Joe Hayton
Acquisition Editor: Anneka Hess
Editorial Project Manager: Emerald Li
Production Project Manager: Omer Mukthar
Designer: Christian Bilbow

Typeset by VTeX

To all on the quest for the Truth

Contents

Introduction (2)

Whatever we know results from our *interaction* with the Big Thing, which we call the Universe (and which is largely unknown). What we do know until now, however, seems to support the view that the Universe itself operates through interaction of some small objects known as elementary particles and that this interaction itself *can* be described in a mathematically consistent way. As usual, physics may have problems with what the word "elementary" really means, but for the purpose of this book we may safely accept a less rigorous "elementariness," in which we have to do with nuclei and electrons only. To explain chemistry at the quantum level (which is usually more than satisfactory for chemical practice) *it is sufficient to treat all nuclei and electrons as point charges that interact electrostatically* (through the Coulomb law).

"*Ideas of quantum chemistry*" is composed of volumes 1 and 2. They are aimed to be autonomous (1 on basics, 2 on some specialized topics) for economical as well as ergonomical reasons, but at the same time interrelated and tied by several aspects (topics, structure, logical connections, appendices) thus forming an entity.

"Ideas of quantum chemistry" Volume 1: From quantum physics to chemistry focuses on making *two important approximations* with the goal of being able to explain the very basics of chemistry. The most important is the concept of the molecular three-dimensional structure (resulting from the Born–Oppenheimer approximation, Chapter V1-6). A second approximation made assumes that each electron is described by its own wave function ("electronic orbital," Chapter V1-8). To arrive at these approximations within the nonrelativistic approach we first exploit the fact that even the lightest nucleus (proton) is more than a thousand times heavier than an electron. As a consequence the Born–Oppenheimer approximation allows one to solve the problem of electronic motion assuming that the nuclei are so heavy that they do not move, i.e., they occupy certain positions in space. The orbital approximation means that any electron sees other electrons' motion averaged. From this, however, follows that a moving electron does not react to some particular positions of other electrons, which looks as if they did not see each other's positions.

By making approximations of this kind, we have arrived at a ("minimal") model of the molecule as an entity that has a distinct spatial shape ("geometry") and is kept together because of the

chemical bonds between pairs of neighboring atoms. Using the minimal model we were able to explain why chemical bonds can be formed. Such a system of bonds is stable with respect to relatively small deviations off the equilibrium molecular geometry ("chemical bond pattern stays unchanged") defined as those positions of the nuclei which ensure the lowest electronic ground-state energy of the system (potential energy surface [PES] for the motion of the nuclei). Larger deviations move the system towards other chemical bond patterns (still within the same electronic ground state[1]), which may also have their stable Born–Oppenheimer equilibrium structures. The atoms vibrate about their equilibrium positions, while the molecule as a whole flies in space along a straight line and keeps rotating about its center of mass.

Thus, any molecule can be treated to a first approximation as an object, small, but similar to the objects we encounter in everyday life. It has a stable albeit flexible three-dimensional architecture, we may speak about its details like left- and right-hand side, its front, its back, etc. *This picture represents the very foundation of all branches of chemistry and biology as well as a large portion of physics.* Understanding chemistry would be extremely difficult, if not impossible, without this crucial picture. We have to remember, however, that the reality is more complicated than that, and this is a simplification only, an extremely fertile and fortuitous approximation.

The minimal model works usually with an error of about 1% in total energy and equilibrium atom–atom distances. This seems at first sight as satisfactory, and it is for a big portion of chemistry. However, the model fails spectacularly for several important phenomena, like description of metals, dissociation of a chemical bond, ubiquitous and important intermolecular interactions, etc.

"Ideas of quantum chemistry" Volume 2: Interactions improves the minimal model of a molecule by taking into account the electron–electron mutual correlation of motion (Chapters 2 and 3) as well as the molecule's interaction with the external world: in a crystal (Chapter 1), in external electric and magnetic fields (Chapter 4), and with other molecules (Chapters 5 and 6).

The author is convinced that chemistry faces currently the challenge of information processing, quite different to this performed by our computers. This perspective, addressed mainly to my young students, is discussed in the last chapter (Chapter 7) of this book, which differs very much from other chapters. It shows some exciting possibilities of chemistry and of theoretical chemistry, and poses some general questions as to where the limits are to be imposed on development science in order to be honest with respect to ourselves, to other people, and to our planet.

The idea of orbitals is not only a simple and fruitful description of molecules within the minimal model. Most importantly, the orbital model provides also a map of ideas and a vocabulary,

[1] Higher in energy scale are excited PESs.

which are used beyond the orbital model, in any more sophisticated theory. Strictly speaking, in any advanced theory introduced in volume 2, there is no such thing as a definition of electronic orbital, but still the "orbital language" is most often used in this more complex situation, because it is simple, flexible, and informative.

Volume 1 may be viewed as independent of volume 2, i.e., volume 1 is autonomous, if the minimal model of chemistry is the reader's target. The opposite is not true: volume 2 relies on volume 1, mainly because it needs the minimal model as the starting point. To keep volume 2 as autonomic as possible the orbital model (the Hartree–Fock theory) is briefly repeated (mainly Appendices A and B).

TREE

Any book has a linear appearance and the page numbers remind us of that. However, the *logic* of virtually any book is *nonlinear*, and in many cases can be visualized by a diagram connecting the chapters that (logically) follow from one another. Such a diagram allows for multiple branches emanating from a given chapter, if being on equal footing. The logical connections are illustrated in this book as a TREE diagram, playing an important role in our book and intended to be a study guide. An author leads the reader in a certain direction and the reader expects to be informed what this direction is, why this direction is needed, what will follow then, and what benefits he/she will gain after such a study.

A thick line in the center of the TREE separates volume 1 (bottom part) from volume 2 (upper part).

The trunk represents the present book's backbone and is covered by the content of volume 1.

- It begins by presenting the foundation of quantum mechanics (postulates).
- It continues with the Schrödinger equation for stationary states, so far the most important equation in quantum chemical applications, and
- the separation of nuclear and electronic motion (through the adiabatic approximation, *the central idea of the present book and chemistry in general*).
- It then develops the orbital model of electronic structure.

The trunk thus corresponds to a traditional course in quantum chemistry for undergraduates. This material represents the necessary basis for further extensions into other parts of the TREE (appropriate for graduate students). The trunk makes it possible to reach the crown of the TREE (volume 2), where the reader may find tasty fruit.

The TREE helps tailoring your own book

The TREE serves not only as a diagram of logical chapter connections, but also enables the reader to make important decisions, such as the following:

- the choice of a logical path of study ("itinerary") leading to topics of interest, and
- elimination of chapters that are irrelevant to the goal of study.[2] This means tailoring the reader's own book.

All readers are welcome to design their own itineraries when traversing the TREE, i.e., to create their own reader-tailored books. Some readers might wish to take into account the suggestions for how *Ideas of quantum chemistry* can be shaped.

Minimum minimorum and minimum

First of all, the reader can follow two basic paths:

- *Minimum minimorum*, for those who want to proceed as quickly as possible to get an idea what quantum chemistry is all about, following the chapters designated by (▲) (only 47 pages, volume 1).
- *Minimum*, for those who seek basic information about quantum chemistry, e.g., in order to use popular computer packages for the study of molecular electronic structure. They may follow the chapters designated by the symbols ▲ *and* △ in volumes 1 and 2. One may imagine here a student of chemistry, specializing in, say, analytical or organic chemistry (not quantum chemistry). This path involves reading approximately 300 pages plus the appropriate appendices (if necessary).

Other proposed paths consist of the *minimum itinerary* (i.e., ▲ *and* △) *plus* special excursions, which are termed additional itineraries.

Additional itineraries

Those who want to use the existing computer packages in a knowledgeable fashion or just want to know more about the chosen subject may follow the chapters designated by the following special signs:

- *large molecules* (□),
- *molecular mechanics and molecular dynamics* (♠),
- *solid state chemistry/physics* (■),

[2] It is, therefore, possible to prune some of the branches.

- *chemical reactions* (℧),
- *spectroscopy* (Ⓢ),
- *exact calculations* on atoms or small molecules (⧫),
- *relativistic and quantum electrodynamics effects* (►),
- *most important computational methods of quantum chemistry* (♢).

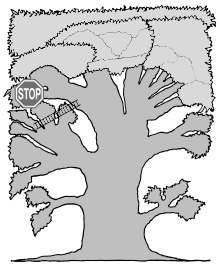
Special itineraries

For readers interested in particular aspects of this book rather than any systematic study, the following suggestions are offered.

- To control the main message of the chapter: "*Where are we*," "*An example*," "*What is it all about*," "*Why is this important*," "*Summary*," "*Questions*," and "*Answers*."
- For those interested in recent progress in quantum chemistry, we suggest sections "*From the research front*" in each chapter.
- For those interested in the future of quantum chemistry we propose the sections labeled "*Ad futurum*" in each chapter, and the chapters designated by (⍊).
- For people interested in the "magical" aspects of quantum physics (e.g., bilocation, reality of the world, teleportation, creation of matter, tunneling, volume 1) we suggest sections with the label (✠).

Your own computations are easy

On the web page www.webmo.net the reader will find a possibility to carry out his/her own quantum mechanical calculations (free up to 60 seconds of CPU time). Nowadays this is a sufficiently long time to perform computations for molecules even with several dozens of atoms.



CHAPTER 1

Electronic Orbital Interactions in Periodic Systems

Beauty of style and harmony and grace and good rhythm depend on simplicity.

Plato

Where are we?

We are on the upper left branch of the TREE.

An example

Polyacetylene[1] represents a *practically* infinite polymeric chain: ···–CH=CH–CH=CH–CH=CH–CH=CH–··· . There is no such a thing in Nature as a truly infinite system.[2] Yet, if we examine larger and larger portions of a homogeneous material, we come to the idea that such quantities as energy per stoichiometric unit, electron excitation energy, vibrational frequencies, etc., depend less and less on system size. This means that a boundary region (polymer ends, crystal surface) contribution to these quantities becomes negligible. Therefore, these (known as *intensive*) quantities attain limit values identical to those for an infinite system. *It pays to investigate the infinite system, because we can use its translational symmetry to simplify its description.* Well, this is what this chapter is all about.

[1] The discovery of conducting polymers (like polyacetylene) was highlighted by the Nobel Prize 2000 for Hideki Shirakawa, who synthesized a crystalline form of polyacetylene, as well as Allan G. MacDiarmid and Allan J. Heeger, who increased its electric conductivity by 18 orders of magnitude by doping the crystal with some electron acceptors and donors. This incredible increase is probably the largest known to humanity in any domain of experimental sciences (H. Shirakawa, E.J. Louis, A.G. MacDiarmid, C.K. Chiang, A.J. Heeger, *Chem. Soc. Chem. Commun*., 578(1977)).

[2] That is, a macromolecule. The concept of polymer was introduced to chemistry by Herman Staudinger.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00009-0
Copyright © 2020 Elsevier B.V. All rights reserved.

Herman Staudinger (1881–1965), German polymer chemist, professor at the University of Freiburg, received the Nobel Prize in 1953 "*for his discoveries in the field of macromolecular chemistry*." However strange it may sound now, as late as in 1926 the concept of polymers was unthinkable in chemistry. It will be encouraging for PhD students to read that a professor advised Staudinger in the late 1920s: "*Dear colleague, leave the concept of large molecules well alone: organic molecules with a molecular weight above 5000 do not exist. Purify your products, such as rubber, then they will crystallize and prove to be lower molecular substances*."

What is it all about?

If a motif (e.g., a cluster of atoms) associated with a unit cell is *regularly translated along three different directions in space*, we obtain an infinite three-dimensional periodic structure (translational symmetry). In 2D (Fig. 1.1a) this means that having a single (special) type of tiles (unit cells) we have been successful in tiling the *complete* two-dimensional space. One of the consequences is that the five-fold symmetry axes have to be absent in the atomic arrangements in such crystals. *It turned out that a vast majority of real crystals can be reliably modeled using this idea.* There were efforts in mathematics to design some nontranslational complete tilings. In 1963 it was first shown that for a number $N = 20\,000$ of square tiles such a nontranslational tiling is possible. This number has been gradually reduced and in 1976 Roger Penrose proposed covering by $N = 2$ kinds of tiles (see Fig. 1.1b). Then, Daniel Shechtman discovered[3] that there are substances (known now as quasicrystals) that indeed show five-fold symmetry axes (also other translationally forbidden symmetry axes). The reason why quasicrystals exist in Nature is quite simple: some strong short-range interactions force unusual five-ligand complexes. It is remarkable that the discovery of quasicrystals has been preceded by ancient artists (Fig. 1.1c).

When applying the Hartree–Fock method to such periodic infinite objects one exploits the translational symmetry of the system, e.g., in calculating integrals. It would be indeed prodigal to compute the integrals many times, the equality of which is *guaranteed* by translational symmetry. When translational symmetry is taken into account, the problem reduces to the calculation of the interaction of a single unit

[3] He received the 2011 chemistry Nobel Prize for "the discovery of quasicrystals."

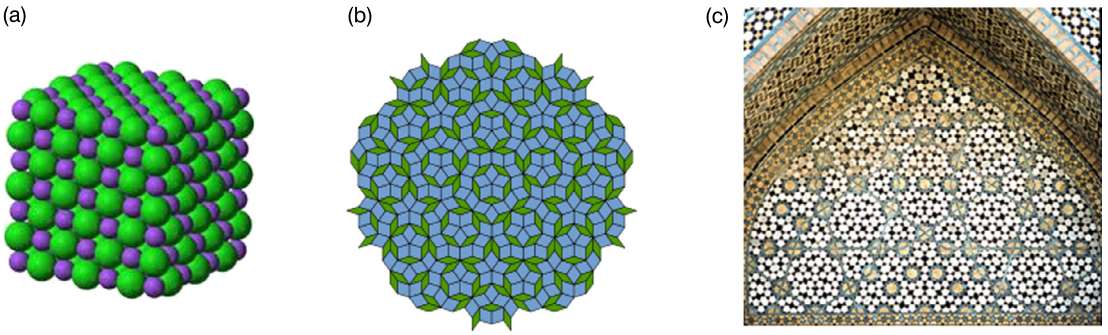


Fig. 1.1. Translational symmetry in crystals and its lack in the quasicrystals despite a perfect long-range order. (a) Translational symmetry in the NaCl crystal build of Na^+ and Cl^- ions. (b) The Penrose tiling as an example of a quasicrystal, no translational symmetry. (c) A medieval Arabian mosaic as an example of a long-range nontranslational order.

cell (reference, labeled by 0) with all other unit cells, the nearest neighbor cells being most important. The infinite size of the system is hidden in the plethora of points (to be taken into account) in what is known as the first Brillouin zone (FBZ). The FBZ represents a unit cell in what is called inverse lattice (associated with a given lattice reflecting the translational symmetry).

The electronic orbital energy becomes a function of the FBZ points and we obtain what is known as a band structure of energy levels. This band structure decides the electronic properties of the system (insulator, semiconductor, metal). We will also show how to carry out the mean-field (Hartree–Fock) computations on infinite periodic systems. The calculations require infinite summations (interaction of the reference unit cell with the infinite crystal) to be made. This creates some mathematical problems, which will be also described in the present chapter.

Why is this important?

The present chapter is particularly important for those readers who are interested in solid state physics and chemistry. Others may treat it as exotic and, if they decide they do not like exotic matter, may go directly to other chapters.

The properties of a polymer or a crystal sometimes differ very widely from those of the atoms or molecules of which they are built. *The same substance* may form *different* periodic structures, which have *different* properties (e.g., graphite and diamond). The properties of periodic structures could be computed by extrapolation of the results obtained for larger and larger clusters of the atoms from which the substance is composed. This avenue is however noneconomic. It is easier to carry out quantum mechanical calculations for an infinite system[4] than for a large cluster.[5]

[4] The surface effects can be neglected and the units the system is composed of are treated as equivalent.

[5] Sometimes we may be interested in a particular cluster, not in an infinite system. Then it may turn out that it is more economic to perform the calculations for the infinite system, and use the results in computations for the clusters (e.g., R.A. Wheeler, L. Piela, R. Hoffmann, *J. Am. Chem. Soc.*, 110(1988)7302).

What is needed?

- Operator algebra (Appendix V1-B, p. V1-595, necessary),
- translation operator (Appendix V1-C, p. V1-605, necessary),
- Hartree–Fock method (Chapter V1-8, necessary),
- multipole expansion (Appendix G, p. 613, advised),
- matrix diagonalization (Appendix V1-L, p. V1-703, advised).

Classical works

At the age of 23, Felix Bloch published an article "*Über die Quantenmechanik der Elektronen in Kristallgittern*" in *Zeitschrift für Physik*, 52(1928)555 (only two years after Schrödinger's historic publication) on the translational symmetry of the wave function. This has also been the first application of LCAO expansion. ★ A book appeared in 1931 by Leon Brillouin entitled *Quantenstatistik* (Springer Verlag, Berlin, 1931), in which the author introduced some of the fundamental notions of band theory. ★ The first *ab initio* calculations for a polymer were carried out by Jean-Marie André in the paper "*Self-Consistent Field Theory for the Electronic Structure of Polymers*," published in the *Journal of the Chemical Physics*, 50(1969)1536.

* * *

1.1 Primitive lattice

Let us imagine an *infinite crystal*, e.g., a system that exhibits the 3D *translational symmetry* of the charge distribution (nuclei and electrons). The translational symmetry will be fully determined by three (linearly independent) basis vectors[6]: $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$, having the property that $\boldsymbol{a}_i$ beginning at any atom extends to the corresponding nearest-neighbor identical atom located in the crystal. The lengths of the basis vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ are called the *lattice constants* along the three periodicity axes.[7]

A lot of such basis sets are possible. Any choice of basis vectors is acceptable from the point of view of mathematics. For economic reasons we choose one of the possible vector sets that give the *least volume parallelepiped*[8] with sides $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$. This parallelepiped (arbitrarily

6 Not necessarily perpendicular though; they determine the periodicity axes.

7 As shown on p. V1-514, a symmetry of the nuclear framework does not guarantee the same symmetry of the electronic charge distribution computed using a mean-field method. We may cope with the period doubling as compared to the period of the nuclear framework (cf. BOAS, p. V1-514). If this happens, then we should choose such lattice constants that ensure the periodicity of both nuclear and electron distributions.

8 We are, however, interested in the smallest unit cell.

shifted in space,[9] Fig. 1.2) represents *our* choice of the *unit cell*,[10] which together with its content (motif) is to be translationally repeated.[11]

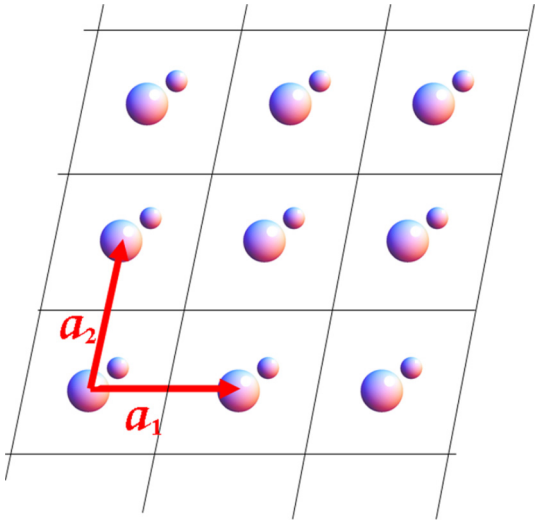


Fig. 1.2. Periodicity in 2D. We choose the unit cell (the parallelogram with vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$) and its content (motif) in such a way as to reproduce the whole infinite crystal by repeating the unit cells through its translation vectors $\mathbf{R}_i = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2$ with integer n_1, n_2. In 3D, instead of the parallelogram, we would have a parallelepiped, which would be repeated by translation vectors $\mathbf{R}_i = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$ with integer n_1, n_2, n_3.

Let us now introduce the *space of translation vectors* $\mathbf{R}_i = \sum_{j=1}^{3} n_{ij}\boldsymbol{a}_j$, where n_{ij} are *arbitrary integer numbers*.

The points indicated by all the translation vectors ("lattice vectors") are called the *crystallographic lattice* or the primitive lattice or simply the lattice.

[9] The choice of the origin of the coordinate system is arbitrary, the basis vectors are determined within the accuracy of an arbitrary translation.

[10] An example of a jigsaw puzzle shows that other choices are possible as well. A particular choice may result from its convenience. This freedom will be used on p. 14.

[11] The motif can be ascribed to the unit cell (i.e., chosen) in many different ways provided that after putting the cells together, we get the same original infinite crystal. Let me propose to disregard this problem for the time being (as well as the problem of the choice of the unit cell) and to think of the unit cell as a space-fixed *parallelepiped with the motif that has been enclosed in it*. We will come back to this complex problem at the end of the present chapter.

Let us introduce the *translation operators* $\widehat{T}(\mathbf{R}_i)$ defined as translations of a *function* on which the operator acts by vector $\mathbf{R}_i$ (cf. Chapter V1-2 and Appendix V1-C on p. V1-605):

$$\widehat{T}(\mathbf{R}_i)\, f(\mathbf{r}) = f(\mathbf{r} - \mathbf{R}_i). \tag{1.1}$$

The function $f(\mathbf{r}) \equiv f(\mathbf{r} - \mathbf{0})$ is centered in the neighborhood of the origin of the coordinate system, while the function $f(\mathbf{r} - \mathbf{R}_i)$ is centered on the point shown by vector $\mathbf{R}_i$.

The crystal periodicity is reflected by the following obvious property of the potential energy V for an electron (V depends on its position in the crystal):

$$V(\mathbf{r}) = V(\mathbf{r} - \mathbf{R}_i), \tag{1.2}$$

for any $\mathbf{R}_i$. The equation simply says that the infinite crystal looks exactly the same close to the origin $\mathbf{0}$ as to the point shown by any lattice vector $\mathbf{R}_i$.

It is easy to see that the operators $\widehat{T}(\mathbf{R}_i)$ form a *group* (Appendix V1-C, p. V1-605) with respect to their multiplication as the group operation.[12,13] In Chapter V1-2 it was shown that the Hamiltonian is invariant with respect to any translation of a molecule. For infinite systems, the proof looks the same for the kinetic energy operator, the invariance of V is guaranteed

[12] Indeed, first a product of such operators represents a translational operator:

$$\widehat{T}(\mathbf{R}_1)\widehat{T}(\mathbf{R}_2)\, f(\mathbf{r}) = \widehat{T}(\mathbf{R}_1)\, f(r - \mathbf{R}_2) = f(\mathbf{r} - \mathbf{R}_1 - \mathbf{R}_2) = f(\mathbf{r} - (\mathbf{R}_1 + \mathbf{R}_2)) = \\ = \widehat{T}(\mathbf{R}_1 + \mathbf{R}_2)\, f(\mathbf{r}).$$

Therefore,

$$\widehat{T}(\mathbf{R}_1)\widehat{T}(\mathbf{R}_2) = \widehat{T}(\mathbf{R}_1 + \mathbf{R}_2). \tag{1.3}$$

The second requirement is to have a unity operator. This role is played by $\widehat{T}(\mathbf{0})$, since

$$\widehat{T}(\mathbf{0})\, f(\mathbf{r}) = f(\mathbf{r} - \mathbf{0}) = f(\mathbf{r}). \tag{1.4}$$

The third condition is the existence (for every $\widehat{T}(\mathbf{R}_i)$) of the inverse operator, which in our case is $\widehat{T}(-\mathbf{R}_i)$, because

$$\widehat{T}(\mathbf{R}_i)\widehat{T}(-\mathbf{R}_i) = \widehat{T}(\mathbf{R}_i - \mathbf{R}_i) = \widehat{T}(\mathbf{0}). \tag{1.5}$$

The group is Abelian (i.e., the operations commute), since

$$\widehat{T}(\mathbf{R}_1)\widehat{T}(\mathbf{R}_2) = \widehat{T}(\mathbf{R}_1 + \mathbf{R}_2) = \widehat{T}(\mathbf{R}_2 + \mathbf{R}_1) = \widehat{T}(\mathbf{R}_2)\widehat{T}(\mathbf{R}_1). \tag{1.6}$$

[13] Besides the translational group, the crystal may also exhibit what is called the *point group*, associated with rotations, reflections in planes, inversion, etc., and the *space group* that results from the translational group and the point group. In such cases, a smaller unit cell may be chosen, because the whole crystal is reproduced not only by translations, but also by other symmetry operations. In the present textbook, we will concentrate on the translational symmetry group only.

by Eq. (1.2). Therefore, the effective one-electron Hamiltonian commutes with any translation operator:

$$\hat{H}\widehat{T}(\mathbf{R}_i) = \widehat{T}(\mathbf{R}_i)\hat{H}.$$

1.2 Wave vector

Since $\widehat{T}(\mathbf{R}_i)$ commutes with the Hamiltonian, its eigenfunctions also represent the eigenfunctions of the translation operator[14] (cf. Chapter V1-2, p. V1-87, also Appendix V1-C on p. V1-605), i.e., in this case $\widehat{H}\psi = E\psi$ *and* $\widehat{T}(\mathbf{R}_j)\psi(\mathbf{r}) = \psi(\mathbf{r} - \mathbf{R}_j) = \lambda_{\mathbf{R}_j}\psi(\mathbf{r})$. The symmetry of V requires the equality of the probability densities

$$|\psi(\mathbf{r} - \mathbf{R}_j)|^2 = |\psi(\mathbf{r})|^2, \tag{1.7}$$

for any lattice vector $\mathbf{R}_j$, which gives $|\lambda_{\mathbf{R}_j}|^2 = 1$, and therefore we may write

$$\lambda_{\mathbf{R}_j} = \exp(-i\theta_{\mathbf{R}_j}), \tag{1.8}$$

where $\theta_{\mathbf{R}_j}$ will be found in a moment.[15]

From equation $\widehat{T}(\mathbf{R}_j)\psi(\mathbf{r}) = \lambda_{\mathbf{R}_j}\psi(\mathbf{r})$ it follows that

$$\lambda_{\mathbf{R}_j}\lambda_{\mathbf{R}_l} = \lambda_{\mathbf{R}_j+\mathbf{R}_l}, \tag{1.9}$$

because

$$\widehat{T}(\mathbf{R}_j + \mathbf{R}_l)\psi(\mathbf{r}) = \lambda_{\mathbf{R}_j+\mathbf{R}_l}\psi(\mathbf{r}). \tag{1.10}$$

However,

$$\widehat{T}(\mathbf{R}_j + \mathbf{R}_l)\psi(\mathbf{r}) = \widehat{T}(\mathbf{R}_j)\widehat{T}(\mathbf{R}_l)\psi(\mathbf{r}) = \lambda_{\mathbf{R}_l}\widehat{T}(\mathbf{R}_j)\psi(\mathbf{r}) = \\ = \lambda_{\mathbf{R}_j}\lambda_{\mathbf{R}_l}\psi(\mathbf{r}).$$

Since this relation has to be satisfied for any $\mathbf{R}_j$ and $\mathbf{R}_l$, it is sufficient to have

$$\theta_{\mathbf{R}_j} = \mathbf{k}\cdot\mathbf{R}_j, \tag{1.11}$$

because a multiplication of λ by λ corresponds to adding the exponents, which results in adding vectors $\mathbf{R}$, which we need to have. The dot product $\mathbf{k}\cdot\mathbf{R}_j$ for simplicity will also be written as $\mathbf{k}\mathbf{R}_j$.

[14] The irreducible representations of an Abelian group are one-dimensional. In our case (translational group) this means that there is no degeneracy and that an eigenfunction of the Hamiltonian is also an eigenfunction of all the translation operators.

[15] The exponent sign is arbitrary, we use "−" following a widely used convention.

CONCLUSION:

The eigenfunctions of the one-electron Hamiltonian and the translation operators correspond to the following eigenvalues of the translation operator: $\lambda_{\mathbf{R}_j} = \exp(-i\mathbf{k}\mathbf{R}_j)$,

where the vector $\mathbf{k}$ characterizes the function, not the direction of $\mathbf{R}_j$. In other words, any one-electron wave function (crystal orbital) which is the eigenfunction of the one-electron Hamiltonian could be labeled by its corresponding vector $\mathbf{k}$, i.e., $\psi(\mathbf{r}) \rightarrow \psi_{\mathbf{k}}(\mathbf{r})$.

BLOCH THEOREM

The value of such a function in the point shifted by the vector $\mathbf{R}_j$ is equal to

$$\psi_{\mathbf{k}}(\mathbf{r} - \mathbf{R}_j) = \exp(-i\mathbf{k}\mathbf{R}_j)\psi_{\mathbf{k}}(\mathbf{r}). \tag{1.12}$$

Felix Bloch (1905–1983), American physicist of Swiss origin, from 1936–1971 professor at Stanford University. Bloch contributed to the electronic structure of metals, superconductivity, ferromagnetism, quantum electrodynamics, and the physics of neutrons. In 1946, independently from E.M. Purcell, he discovered the nuclear magnetic resonance effect. Both scientists received the Nobel Prize in 1952 "*for the development of new methods for nuclear magnetic precision measurements and the discoveries in connection therewith*."

This relation represents a necessary condition to be fulfilled by the eigenfunctions for a perfect periodic structure (crystal, layer, polymer). This equation differs widely from Eq. (1.2) for potential energy. Unlike potential energy, which does not change upon a lattice translation, the wave function undergoes a change of its phase acquiring the factor $\exp(-i\mathbf{k}\mathbf{R}_j)$.

Any linear combination of functions labeled by the same $\mathbf{k}$ represents an eigenfunction of any lattice translation operator and corresponds to the same $\mathbf{k}$. Indeed, from the linearity of the translation operator

$$\widehat{T}(\mathbf{R}_l)\big(c_1\phi_{\mathbf{k}}(\mathbf{r})+c_2\psi_{\mathbf{k}}(\mathbf{r})\big)=c_1\phi_{\mathbf{k}}(\mathbf{r}-\mathbf{R}_l)+c_2\psi_k(\mathbf{r}-\mathbf{R}_l)=$$
$$=c_1\exp(-i\mathbf{k}\mathbf{R}_l)\phi_{\mathbf{k}}(\mathbf{r})+c_2\exp(-i\mathbf{k}\mathbf{R}_l)\psi_{\mathbf{k}}(\mathbf{r})=$$
$$=\exp(-i\mathbf{k}\mathbf{R}_l)\big(c_1\phi_{\mathbf{k}}(\mathbf{r})+c_2\psi_{\mathbf{k}}(\mathbf{r})\big).$$

Let us construct the following function (called a *Bloch function*) from a function $\chi(\mathbf{r})$, which in the future will play the role of an atomic orbital (in this case centered at the origin):

$$\phi(\mathbf{r})=\sum_j \exp(i\mathbf{k}\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j),$$

where the summation extends over all possible $\mathbf{R}_j$, i.e., over the whole crystal lattice. The function ϕ is *automatically* an eigenfunction of any translation operator and may be labeled by the index[16] $\mathbf{k}$.

Our function ϕ represents, therefore, an eigenfunction of the translation operator with the same eigenvalue as that corresponding to $\psi_{\mathbf{k}}$. In the following very often $\psi_{\mathbf{k}}$ will be constructed as a linear combination of Bloch functions ϕ.

A Bloch function is nothing but a symmetry orbital built from the functions $\chi(\mathbf{r}-\mathbf{R}_j)$.

A symmetry orbital is a linear combination of atomic orbitals that transforms according to an irreducible representation Γ of the symmetry group of the Hamiltonian (cf. Appendix V1-C).

[16] Indeed, first

$$\widehat{T}(\mathbf{R}_l)\phi(\mathbf{r})=\widehat{T}(\mathbf{R}_l)\sum_j\exp(i\mathbf{k}\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j)=\sum_j\exp(i\mathbf{k}\mathbf{R}_j)\widehat{T}(\mathbf{R}_l)\chi(\mathbf{r}-\mathbf{R}_j)=$$
$$=\sum_j\exp(i\mathbf{k}\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j-\mathbf{R}_l).$$

Instead of the summation over $\mathbf{R}_j$, let us introduce a summation over $\mathbf{R}_{j'}=\mathbf{R}_j+\mathbf{R}_l$, which means an *identical* summation as before, but we begin to sum the term up from another point of the lattice. Then, we can write

$$\sum_{j'}\exp(i\mathbf{k}(\mathbf{R}_{j'}-\mathbf{R}_l))\chi(\mathbf{r}-\mathbf{R}_{j'})=\exp(-i\mathbf{k}\mathbf{R}_l)\sum_{j'}\exp(i\mathbf{k}\mathbf{R}_{j'})\chi(\mathbf{r}-\mathbf{R}_{j'})=$$
$$=\exp(-i\mathbf{k}\mathbf{R}_l)\phi(\mathbf{r}),$$

which had to be proved.

In order to obtain such a function we may use the corresponding projection operator (see Eq. (V1-C.13)).

There is also another way to construct a function $\phi_{\mathbf{k}}(\mathbf{r})$ of a given $\mathbf{k}$ from an auxiliary function $u(\mathbf{r})$ satisfying an equation similar to Eq. (1.2) for the potential V

$$\widehat{T}(\mathbf{R}_i)u(\mathbf{r}) = u(\mathbf{r}-\mathbf{R}_i) = u(\mathbf{r}). \tag{1.13}$$

Then, $\phi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{kr})u(\mathbf{r})$. Indeed, let us check

$$\widehat{T}(\mathbf{R}_j)\phi_{\mathbf{k}}(\mathbf{r}) = \widehat{T}(\mathbf{R}_j)\exp(i\mathbf{kr})u(\mathbf{r}) = \exp(i\mathbf{k}(\mathbf{r}-\mathbf{R}_j))u(\mathbf{r}-\mathbf{R}_j) = \exp(-i\mathbf{kR}_j)\phi_{\mathbf{k}}(\mathbf{r}). \tag{1.14}$$

1.3 Inverse lattice

Let us now construct the so-called *biorthogonal basis* $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ with respect to the basis vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ of the primitive lattice, i.e., the vectors that satisfy the *biorthogonality relations*

$$\boldsymbol{b}_i\,\boldsymbol{a}_j = 2\pi\,\delta_{ij}. \tag{1.15}$$

The vectors $\boldsymbol{b}_i$ can be expressed by the vectors $\boldsymbol{a}_i$ in the following way:

$$\boldsymbol{b}_i = 2\pi \sum_j \boldsymbol{a}_j \left(\mathbf{S}^{-1}\right)_{ji}, \tag{1.16}$$

$$\mathbf{S}_{ij} = \boldsymbol{a}_i \cdot \boldsymbol{a}_j. \tag{1.17}$$

The vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ form the basis of a lattice in a three-dimensional space. This lattice will be called the *inverse lattice*. The inverse lattice vectors are, therefore,

$$\mathbf{K}_j = \sum_{i=1}^{i=3} g_{ji}\boldsymbol{b}_i, \tag{1.18}$$

where g_{ij} represent arbitrary *integers*. We have $\mathbf{K}_j\mathbf{R}_i = 2\pi M_{ij}$, where M_{ij} are integer numbers.

Indeed,

$$\mathbf{K}_j \cdot \mathbf{R}_i = \sum_{l=1}^{l=3} g_{jl}\boldsymbol{b}_l \cdot \sum_{k=1}^{k=3} n_{ik}\boldsymbol{a}_k = \sum_{l=1}^{l=3}\sum_{k=1}^{k=3} n_{ik}g_{jl}\boldsymbol{b}_l \cdot \boldsymbol{a}_k = \tag{1.19}$$

$$= \sum_{l=1}^{l=3} \sum_{k=1}^{k=3} n_{ik} g_{jl} (2\pi) \delta_{lk} = (2\pi) \sum_{l=1}^{l=3} n_{il} g_{jl} = 2\pi M_{ij} \tag{1.20}$$

with n_{ik}, g_{jl}, and therefore also M_{ij} as *integers*.

The inverse lattice is composed, therefore, from the isolated points indicated from the origin by the vectors $\mathbf{K}_j$. All the vectors that begin at the origin form the *inverse space*.

Let us see how we obtain the inverse lattice (1D, 2D, 3D) in practice.

1D

We have only a single biorthogonality relation: $\boldsymbol{b}_1 \boldsymbol{a}_1 = 2\pi$, i.e., after skipping the index $\boldsymbol{ba} = 2\pi$. Because of the single dimension, we have to have $\boldsymbol{b} = \frac{2\pi}{a} \left(\frac{\boldsymbol{a}}{a}\right)$, where $|\boldsymbol{a}| \equiv a$. Therefore,

the vector $\boldsymbol{b}$ has length $\frac{2\pi}{a}$ and the same direction as $\boldsymbol{a}$.

2D

This time we have to satisfy: $\boldsymbol{b}_1 \boldsymbol{a}_1 = 2\pi$, $\boldsymbol{b}_2 \boldsymbol{a}_2 = 2\pi$, $\boldsymbol{b}_1 \boldsymbol{a}_2 = 0$, $\boldsymbol{b}_2 \boldsymbol{a}_1 = 0$. This means that the game takes place within the plane determined by the lattice vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. The vector $\boldsymbol{b}_1$ has to be perpendicular to $\boldsymbol{a}_2$, while $\boldsymbol{b}_2$ has to be perpendicular to $\boldsymbol{a}_1$, their directions as shown in Fig. 1.3 (each of the $\boldsymbol{b}$ vectors is a linear combination of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ according to Eq. (1.16)).

3D

In the three-dimensional case the biorthogonality relations are equivalent to setting

$$\boldsymbol{b}_1 = \boldsymbol{a}_2 \times \boldsymbol{a}_3 \frac{2\pi}{V}, \tag{1.21}$$

$$\boldsymbol{b}_2 = \boldsymbol{a}_3 \times \boldsymbol{a}_1 \frac{2\pi}{V}, \tag{1.22}$$

$$\boldsymbol{b}_3 = \boldsymbol{a}_1 \times \boldsymbol{a}_2 \frac{2\pi}{V}, \tag{1.23}$$

where

$$V = \boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) \tag{1.24}$$

is the volume of the unit cell of the crystal (Fig. 1.4).

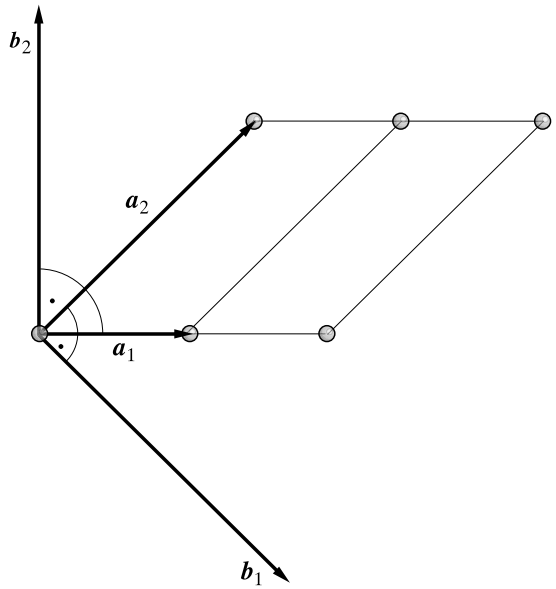


Fig. 1.3. Construction of the inverse lattice in 2D. In order to satisfy the biorthogonality relations (1.15) the vector $\boldsymbol{b}_1$ has to be orthogonal to $\boldsymbol{a}_2$, while $\boldsymbol{b}_2$ must be perpendicular to $\boldsymbol{a}_1$. The lengths of the vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ also follow from the biorthogonality relations $\boldsymbol{b}_1 \cdot \boldsymbol{a}_1 = \boldsymbol{b}_2 \cdot \boldsymbol{a}_2 = 2\pi$.

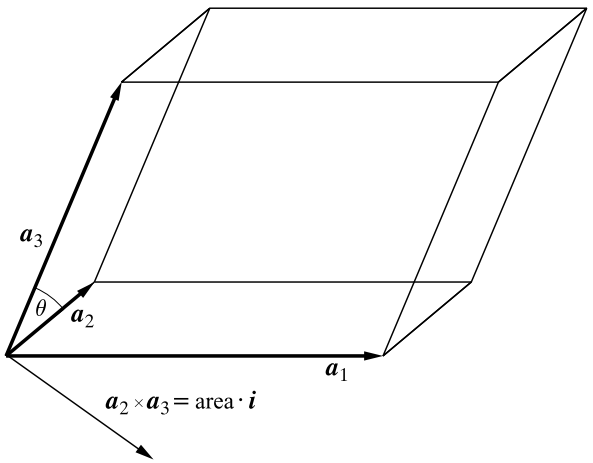


Fig. 1.4. The volume V of the unit cell is equal to $V = \boldsymbol{a}_1 \cdot (\textit{area of the base})\mathbf{i} = \boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)$.

1.4 First Brillouin zone (FBZ)

Léon Nicolas Brillouin (1889–1969), French physicist, professor at the Sorbonnne and College de France in Paris, after 1941 in the USA: at the University of Madison, Columbia University, Harvard University. His contributions included quantum mechanics and solid state theory (he is one of the founders of electronic band theory).

As was remarked at the beginning of this chapter, the example of a jigsaw puzzle shows us that a parallelepiped unit cell does not represent the only choice. Now, we will profit from this extra freedom and will define the so-called *Wigner–Seitz* unit cell. Here follows the prescription for how to construct it (Fig. 1.5).

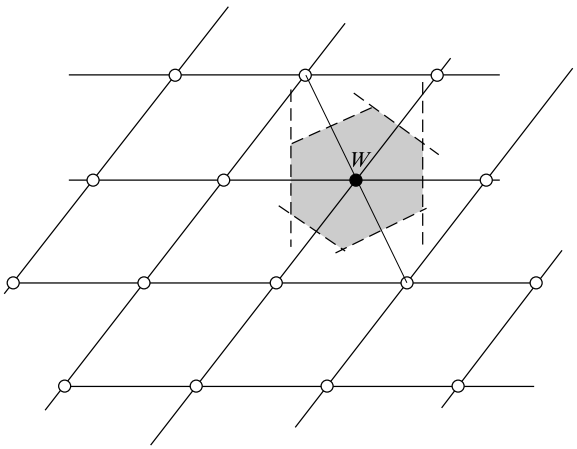


Fig. 1.5. Construction of the First Brillouin zone (FBZ) as a Wigner–Seitz unit cell of the inverse lattice in 2D. The circles represent the nodes of the inverse lattice. We cut the lattice in the middle between the origin node W and all the other nodes (here it turns out to be sufficient to take only the nearest and the next nearest neighbors) and remove all the sawn-off parts that do not contain W. Finally we obtain the FBZ in the form of a hexagon. The Wigner–Seitz unit cells (after performing all allowed translations in the inverse lattice) reproduce the complete inverse space.

We focus on a node W, saw the crystal along the plane that dissects (symmetrically) the distance to a nearest neighbor node, throw the part that *does not* contain W into the fireplace, and then repeat the procedure until we are left with a solid containing W. This solid represents the FBZ.

1.5 Properties of the FBZ

The vectors $\mathbf{k}$, which begin at the origin and end in the interior of the FBZ, label *all different* irreducible representations of the translational symmetry group.

Let us imagine two inverse space vectors $\mathbf{k}'$ and $\mathbf{k}''$ related by the equality $\mathbf{k}'' = \mathbf{k}' + \mathbf{K}_s$, where $\mathbf{K}_s$ stands for an inverse lattice vector. Taking into account the way the FBZ has been constructed, if one of them, say, $\mathbf{k}'$, indicates a point in the interior of the FBZ, then the second, $\mathbf{k}''$, "protrudes" outside the FBZ. Let us try to construct a Bloch function that corresponds to $\mathbf{k}''$. We have

$$\phi_{\mathbf{k}''} = \sum_j \exp(i\mathbf{k}''\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j) = \sum_j \exp\left(i(\mathbf{k}'+\mathbf{K}_s)\mathbf{R}_j\right)\chi(\mathbf{r}-\mathbf{R}_j) = \tag{1.25}$$

$$= \exp(i\mathbf{K}_s\mathbf{R}_j)\sum_j \exp(i\mathbf{k}'\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j) = \tag{1.26}$$

$$= \exp(i2\pi M_{sj})\sum_j \exp(i\mathbf{k}'\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j) = \tag{1.27}$$

$$= \sum_j \exp(i\mathbf{k}'\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j) = \phi_{\mathbf{k}'}. \tag{1.28}$$

It turns out that our function ϕ does behave like corresponding to $\mathbf{k}'$. We say that the two vectors are *equivalent*.

Vector $\mathbf{k}$ outside the FBZ is always equivalent to a vector from inside the FBZ, while two vectors from inside of the FBZ are never equivalent. Therefore, if we are interested in electronic states (the irreducible representation of the translational group is labeled by $\mathbf{k}$ vectors) it is sufficient to limit ourselves to those $\mathbf{k}$ vectors that are enclosed in the FBZ.

1.6 A few words on Bloch functions

1.6.1 Waves in 1D

Let us take a closer look of a Bloch function corresponding to the vector $\mathbf{k}$,

$$\phi_{\mathbf{k}}(\mathbf{r}) = \sum_j \exp(i\mathbf{k}\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j), \tag{1.29}$$

and limit ourselves to one-dimensional periodicity. In such a case, the wave vector $\mathbf{k}$ reduces to a *wave number* k, and the vectors $\mathbf{R}_j$ can all be written as $\mathbf{R}_j = aj\mathbf{z}$, where $\mathbf{z}$ stands for the unit vector along the periodicity axis, a means the lattice constant (i.e., the nearest neighbor distance), while $j = 0, \pm 1, \pm 2, \ldots$. Let us assume that in the lattice nodes we have hydrogen atoms with orbitals $\chi = 1s$. Therefore, in 1D we have

$$\phi_k(\mathbf{r}) = \sum_j \exp(ikja)\chi(\mathbf{r} - aj\mathbf{z}). \tag{1.30}$$

Let me stress that ϕ_k represents a function of position $\mathbf{r}$ in the three-dimensional space; only the periodicity has a one-dimensional character. The function is a linear combination of the hydrogen atom $1s$ orbitals. The coefficients of the linear combination depend exclusively on the value of k. Eq. (1.28) tells us that the allowed $k \in (0, \frac{2\pi}{a})$, or alternatively $k \in (-\frac{\pi}{a}, \frac{\pi}{a})$. If we exceed the FBZ *length* $\frac{2\pi}{a}$, then we would simply repeat the Bloch functions. For $k = 0$ we get

$$\phi_0 = \sum_j \exp(0)\chi(\mathbf{r} - aj\mathbf{z}) = \sum_j \chi(\mathbf{r} - aj\mathbf{z}), \tag{1.31}$$

i.e., simply a sum of the $1s$ orbitals. Such a sum has a large value on the nuclei, and close to a nucleus ϕ_0 will be delusively similar to its $1s$ orbital (Fig. 1.6a).

The function looks like a chain of buoys floating on a perfect water surface. If we ask whether ϕ_0 represents a wave, the answer could be that if any, then its wave length is ∞. What about $k = \frac{\pi}{a}$? In such a case

$$\phi_{\frac{\pi}{a}}(\mathbf{r}) = \sum_j \exp(ij\pi)\chi(\mathbf{r} - aj\mathbf{z}) = \sum_j (\cos \pi j + i \sin \pi j)\chi(\mathbf{r} - aj\mathbf{z}) =$$
$$= \sum_j (-1)^j \chi(\mathbf{r} - aj\mathbf{z}).$$

If we decide to draw the function in space, we would obtain Fig. 1.6b. When asked this time, we would answer that the wave length is equal to $\lambda = 2a$, which by the way is equal to[17] $\frac{2\pi}{|k|}$. There is a problem. Does the wave correspond to $k = \frac{\pi}{a}$ or $k = -\frac{\pi}{a}$? It corresponds to *both* of them. Well, does it contradict the theorem that the FBZ contains all *different* states? No, everything is OK. Both functions are from the border of the FBZ, their k values differ by $\frac{\pi}{2a}$ (one of the inverse lattice vectors), and therefore both functions represent *the same state*.

Now, let us take $k = \frac{\pi}{2a}$. We obtain

$$\phi_k(\mathbf{r}) = \sum_j \exp(\frac{i\pi j}{2})\chi(\mathbf{r} - aj\mathbf{z}) = \sum_j (\cos(\frac{\pi j}{2}) + i \sin(\frac{\pi j}{2}))\chi(\mathbf{r} - aj\mathbf{z}) \tag{1.32}$$

[17] In the preceding case the formula $\lambda = \frac{2\pi}{k}$ also worked, because it gave $\lambda = \infty$.

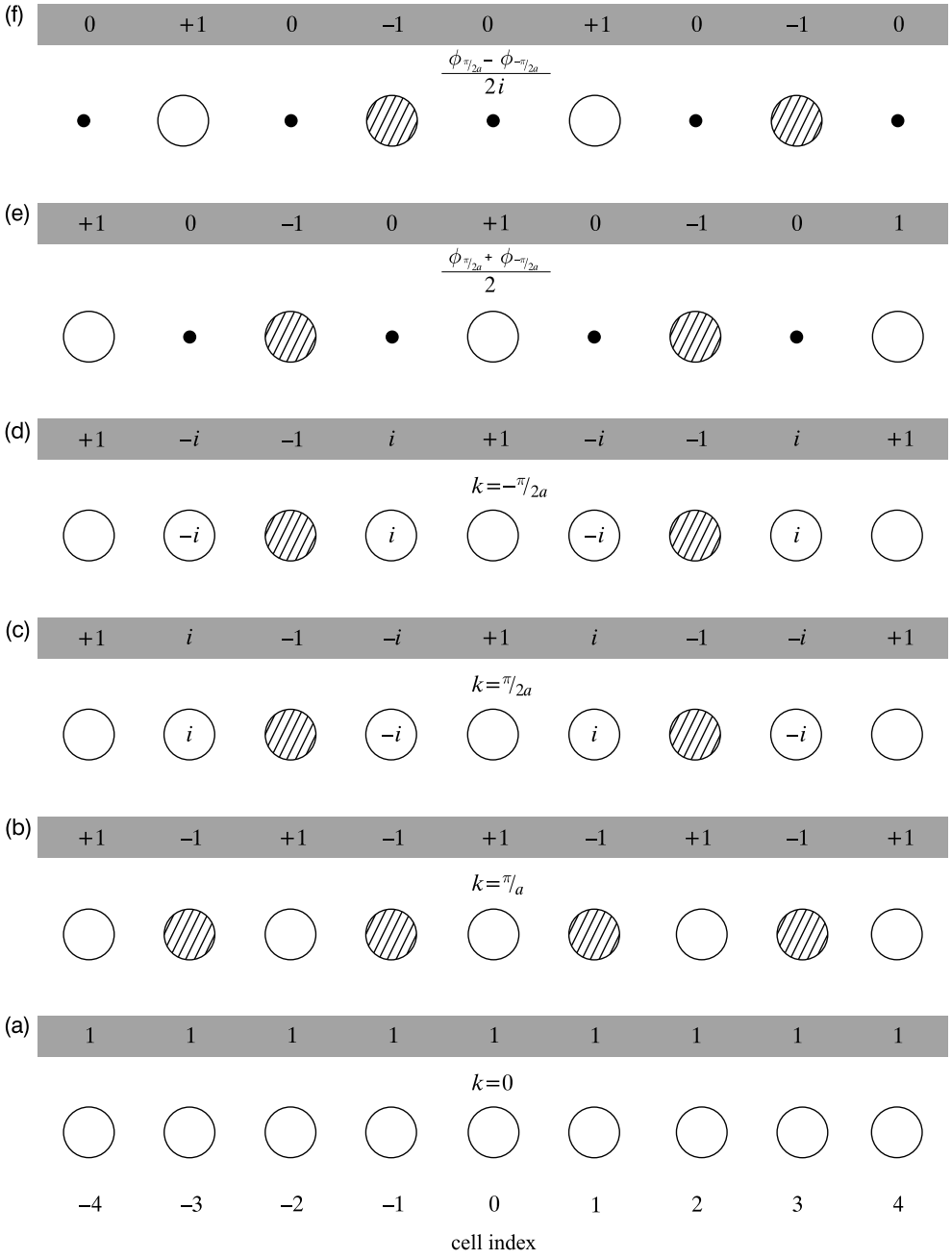


Fig. 1.6. Waves in 1D. Shadowed (white) circles mean negative (positive) value of the function, the coefficients multiplying the hydrogen 1s orbitals are given within the gray stripes. Despite the fact that the waves are complex, in each of the cases (a)–(f) we are able to determine their wave length.

with some coefficients being complex numbers. For $j = 0$ the coefficient is equal to 1, for $j = 1$ it equals i, for $j = 2$ it takes the value -1, for $j = 3$ it attains $-i$, for $j = 4$ it is again 1, and the values repeat periodically. This is depicted in Fig. 1.6c. If this time we ask whether we see any wave there, we have to answer that we do, because after the length $4a$ everything begins to repeat. Therefore, $\lambda = 4a$ and again it is equal to $\frac{2\pi}{k} = \frac{2\pi}{\frac{\pi}{2a}}$. Everything is OK, except that humans like pictures more than schemes. Can we help it somehow? Let us take a look of $\phi_k(\mathbf{r})$ which corresponds to $k = -\frac{\pi}{2a}$. We may easily convince ourselves that this situation corresponds to what we have in Fig. 1.6d.

Let us stress that $\phi_{-k} = \phi_k^*$ represents *another* complex wave. By adding and subtracting $\phi_k(\mathbf{r})$ and $\phi_{-k}(\mathbf{r})$ we receive the real functions, which can be plotted and that is all we need. By adding $\frac{1}{2}(\phi_k + \phi_{-k})$, we obtain

$$\frac{1}{2}(\phi_k + \phi_{-k}) = \sum_j \cos(\frac{\pi j}{2})\chi(\mathbf{r} - aj\mathbf{z}), \tag{1.33}$$

while $\frac{1}{2i}(\phi_k - \phi_{-k})$ results in

$$\frac{1}{2i}(\phi_k - \phi_{-k}) = \sum_j \sin(\frac{\pi j}{2})\chi(\mathbf{r} - aj\mathbf{z}). \tag{1.34}$$

Now, there is no problem with plotting the new functions (Fig. 1.6e,f).[18]

A similar technique may be applied to any k. Each time we will find that the wave we see exhibits the wave length $\lambda = \frac{2\pi}{k}$.

1.6.2 Waves in 2D

Readers confident in their understanding of the wave vector concept may skip this subsection.

This time we will consider the crystal as a two-dimensional rectangular lattice; therefore, the corresponding inverse lattice is also two-dimensional as well as the wave vectors $\mathbf{k} = (k_x, k_y)$.

Let us first take $\mathbf{k} = (0, 0)$. We immediately obtain $\phi_{\mathbf{k}}$, as shown in Fig. 1.7a, which corresponds to infinite wave length (again $\lambda = \frac{2\pi}{k}$), which looks as "no wave" at all.

[18] And what would happen if we took $k = \frac{\pi}{a}\frac{m}{n}$, with the integer $m < n$? We would again obtain a wave with the wave length $\lambda = \frac{2\pi}{k}$, i.e., in this case $\lambda = \frac{n}{m}2a$. It would be quite difficult to recognize such a wave computed at the lattice nodes, because the closest wave maxima would be separated by $n2a$ and this length would have been covered by m wave lengths.

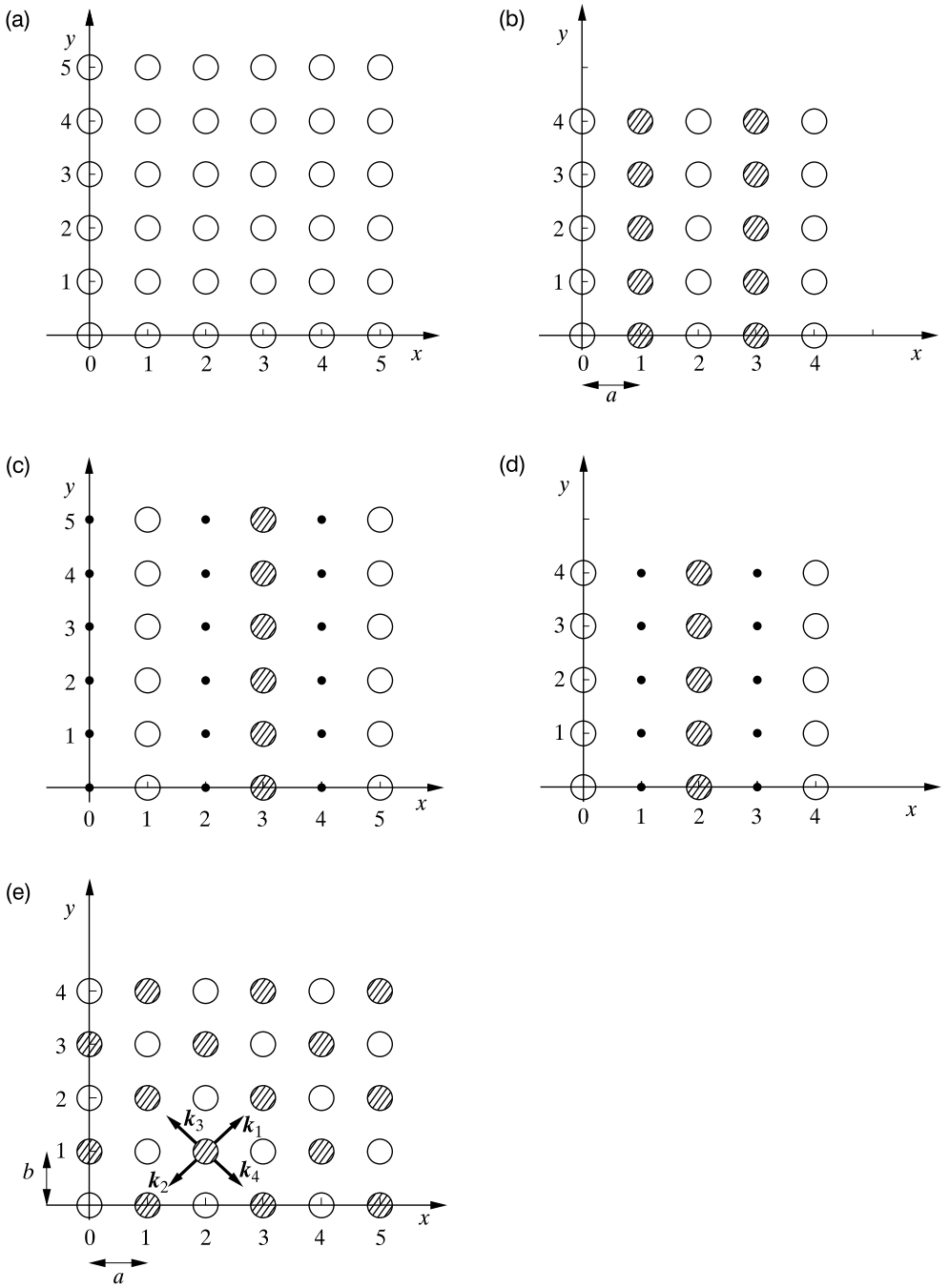


Fig. 1.7. Waves in 2D. In any case $\lambda = \frac{2\pi}{k}$, while the wave vector $\mathbf{k}$ points to the direction of the wave propagation. (a) $\mathbf{k} = (0, 0)$. (b) $\mathbf{k} = \left(\frac{\pi}{a}, 0\right)$. (c) $\mathbf{k} = \left(\frac{\pi}{2a}, 0\right)$, $\frac{1}{2i}(\phi_{\mathbf{k}} - \phi_{-\mathbf{k}})$. (d) $\mathbf{k} = \left(\frac{\pi}{2a}, 0\right)$, $\frac{1}{2}(\phi_{\mathbf{k}} + \phi_{-\mathbf{k}})$. (e) $\mathbf{k} = \left(\frac{\pi}{a}, \frac{\pi}{b}\right)$.

Let us try $\mathbf{k} = (\frac{\pi}{a}, 0)$. The summation over j may be replaced by a double summation (indices m and n along the x and y axes, respectively); therefore, $\mathbf{R}_j = ma\mathbf{x} + nb\mathbf{y}$, where m and n correspond to the unit cell j, a and b denote the lattice constants along the axes shown by the unit vectors $\mathbf{x}$ and $\mathbf{y}$. We have

$$\phi_{\mathbf{k}} = \sum_{mn} \exp\big(i(k_x ma + k_y nb)\big)\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}) = \\ = \sum_{mn} \exp(i\pi m)\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}) = \sum_{mn} (-1)^m \chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}).$$

If we go through all m and n, it is easily seen that moving along x we will meet the signs $+1,-1,+1,-1, \ldots$, while moving along y we have the same sign all the time. This will correspond to Fig. 1.7b.
This is a wave.

> The wave fronts are oriented along y, i.e., the wave runs along the x axis, in the direction of the wave vector $\mathbf{k}$. The same happened in the one-dimensional cases, but we did not express that explicitly: the wave moved along the (one-dimensional) vector $\mathbf{k}$.

Exactly as before the wave length is equal to 2π divided by the length of $\mathbf{k}$. Since we are at the FBZ border, a wave with $-\mathbf{k}$ simply means the same wave as for $\mathbf{k}$.

If we take $\mathbf{k} = [\frac{\pi}{2a}, 0]$, then

$$\phi_{\mathbf{k}} = \sum_{mn} \exp\Big(i\,(\mathbf{k}_x ma + \mathbf{k}_y nb)\Big)\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}) = \\ = \sum_{mn} \exp(\frac{i\pi m}{2})\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}).$$

This case is very similar to that in 1D for $k = \frac{\pi}{2a}$, when we look at the index m and $k = 0$, and when we take into account the index n. We may carry out the same trick with addition and subtraction, and immediately get Fig. 1.6c,d.

Is there any wave over there? Yes, there is. The wave length equals $4a$, i.e., $\lambda = \frac{2\pi}{k}$, and the wave is directed along vector $\mathbf{k}$. When making the figure, we also used the wave corresponding to $-\mathbf{k}$; therefore, neither the sum nor the difference corresponds to $\mathbf{k}$ *or* $-\mathbf{k}$, but rather to both of them (we have two standing waves). The reader may guess the wave length and direction of propagation for $\phi_{\mathbf{k}}$ corresponding to $\mathbf{k} = [0, \frac{\pi}{2b}]$.

Let us see what happens for $\mathbf{k} = [\frac{\pi}{a}, \frac{\pi}{b}]$. We obtain

$$\begin{aligned}\phi_{\mathbf{k}} &= \sum_{mn} \exp\left(i\left(k_x ma + k_y nb\right)\right)\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}) = \\ &= \sum_{mn} \exp\left(i\left(m\pi + n\pi\right)\right)\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}) = \\ &= \sum_{mn} (-1)^{m+n}\chi(\mathbf{r} - ma\mathbf{x} - nb\mathbf{y}),\end{aligned}$$

which produces waves propagating along $\mathbf{k}$. And what about the wave length? We obtain[19]

$$\lambda = \frac{2\pi}{\sqrt{(\frac{\pi}{a})^2 + (\frac{\pi}{b})^2}} = \frac{2ab}{\sqrt{a^2 + b^2}}. \tag{1.35}$$

In the last example there is something that may worry us. As we can see, our figure corresponds not only to $\mathbf{k}_1 = (\frac{\pi}{a}, \frac{\pi}{b})$ and $\mathbf{k}_2 = (-\frac{\pi}{a}, -\frac{\pi}{b})$, which is understandable (as discussed above), but also to the wave with $\mathbf{k}_3 = (-\frac{\pi}{a}, \frac{\pi}{b})$ and to the wave evidently coupled to it, namely, with $\mathbf{k}_4 = (\frac{\pi}{a}, -\frac{\pi}{b})$! What is going on? Again, let us recall that we are on the FBZ border and this identity is natural, because the vectors $\mathbf{k}_2$ and $\mathbf{k}_3$ as well as $\mathbf{k}_1$ and $\mathbf{k}_4$ differ by the inverse lattice vector $(0, \frac{2\pi}{b})$, which makes the two vectors equivalent.

1.7 Infinite crystal as a limit of a cyclic system

1.7.1 Origin of the band structure

Let us consider the hydrogen atom in its ground state (cf. p. V1-232). The atom is described by the atomic orbital $1s$ and corresponds to energy -0.5 a.u. Let us now take two such atoms. Due to their interaction we have two molecular orbitals: bonding and antibonding (cf. p. V1-511), which correspond, respectively, to energies a bit lower than -0.5 and a bit higher than -0.5 (this splitting is larger if the overlap of the atomic orbitals gets larger). We therefore have two energy levels, which stem directly from the $1s$ levels of the two hydrogen atoms. For three atoms we would have three levels, for 10^{23} atoms we would get 10^{23} energy levels, which would

[19] The formula can be easily verified in two limiting cases. The first corresponds to $a = b$. Then, $\lambda = a\sqrt{2}$, and this agrees with Fig. 1.7e. The second case is when $b = \infty$, which gives $\lambda = 2a$, exactly as in the one-dimensional case with $k = \frac{\pi}{a}$. This is what we expected.

be densely distributed along the energy scale, but would not cover the whole scale. There will be a bunch of energy levels stemming from $1s$, i.e., an *energy band* of allowed electronic states. If we had an infinite chain of hydrogen atoms, there would be a band resulting from $1s$ levels, a band stemming from $2s$, $2p$, etc., so the bands might be separated by *energy gaps*.

How dense would the distribution of the electronic levels be? Will the distribution be uniform? Answers to such questions are of prime importance for the electronic theory of crystals. It is always advisable to come to a conclusion by steps, starting from something as simple as possible, which we understand very well.

Fig. 1.8 shows how the energy level distribution looks for longer and longer rings (regular polygon) of hydrogen atoms. One of the important features of the distribution is that

> the levels extend over an energy interval and are more numerous for energy extremes.

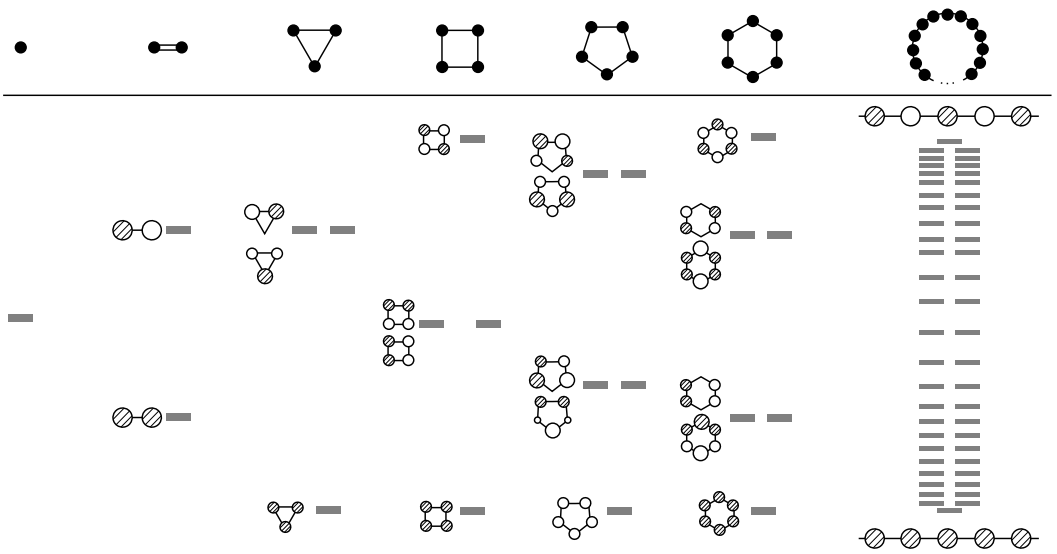


Fig. 1.8. Energy level distribution for a regular polygon built from hydrogen atoms. It is seen that the energy levels are located within an energy band, and are closer to one another at the band edges. The center of the band is close to the binding energy in the isolated hydrogen atom (equal to -0.5 a.u.). Next to energy levels the molecular orbitals are shown schematically (the shadowed circles mean negative values). R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

How do the wave functions that correspond to higher and higher-energy levels in a band look? Let us see the situation in the ring H_n molecules. Fig. 1.8 indicates that the rule is very simple.

The number of nodes of the wave function increases by one when we go to the next level (higher in the energy scale).[20]

1.7.2 Born–von Kármán condition in 1D

How is it in the case of a crystal? Here we are confronted with the first difficulty. Which crystal, and of what shape? Should it be an ideal crystal, i.e., with perfectly ordered atoms? There is nothing like the perfect crystal in Nature. For the sake of simplicity (as well as generality) let us assume, however, that our crystal *is* perfect indeed. And what about its surface (shape)? Even if we aimed at studying the surface of a crystal, the first step would be the infinite crystal (i.e., with no surface). This is the way theoreticians always operate.[21]

> One of the ingenious ideas in this direction is known as the *Born–von Kármán boundary conditions*. The idea is that instead of considering a crystal treated as a stick (let us consider the one-dimensional case), we treat it as a circle, i.e., *the value of the wave function at one end of the stick has to be equal to the wave function value at the other end*. In this way we remove the problem of the crystal ends, and on top of that, all the unit cells become equivalent.

The same may be done in two- and three-dimensional cases. We usually introduce the Born–von Kármán boundary conditions for a finite N and then go with N to ∞. After such a procedure is carried out, we are pretty sure that the solution we are going to obtain will not only be true for an infinite cycle but also for the mass (bulk) of the infinite crystal. This stands to reason, provided that the crystal surface does not influence the (deep) bulk properties at all.[22] In the ideal periodic case, we have to do with the cyclic translational symmetry group (Appendix V1-C on p. V1-605). The group is Abelian and, therefore, all the irreducible representations have dimension 1.

---

[20] They are bound to differ by the number of nodes, because this ensures their mutual orthogonality (required for the eigenfunctions of a Hermitian operator).

[21] People say that when theoreticians attack the problem of stability of a table as a function of the number n of its legs, they do it in the following way. First, they start with $n = 0$, then they proceed with $n = 1$, then they go to $n = \infty$, and after that they have no time to consider other values of n.

[22] We circumvent the difficult problem of the crystal surface. The boundary (surface) problem is extremely important for obvious reasons: we usually *have to do with this, not with the bulk*. The existence of the surface leads to some specific surface-related electronic states.

Theodore von Kármán (1881–1963), American physicist of Hungarian origin, director of the Guggenheim Aeronautical Laboratory at the California Institute of Technology in Pasadena. Von Kármán was also a founder of the NASA Jet Propulsion Laboratory and father of the concept of the first supersonic aeroplane. On the Hungarian stamp one can see the famous "Kármán vortex street" behind an airplane. He was asked by the father of the young mathematical genius John von Neumann to persuade him that the job of a mathematician is

far less exciting than that of a banker. Theodore von Kármán did not accomplish this mission well (to the benefit of science).

Let us assume we have to do with N equidistant atoms located on a circle, the nearest neighbor distance being a. From the Bloch theorem, Eq. (1.12), for the wave function ψ we have

$$\psi(N) = \exp(-ikaN)\psi(0), \tag{1.36}$$

where we have assumed that the wave function ψ corresponds to the wave vector $\mathbf{k}$ (here, in 1D, to the wave number k), the translation has been carried out by Na, and as the argument of the function ψ we have (symbolically) used the number $(0, 1, 2, \ldots N-1)$ of the atom on which the function is computed.

The Born–von Kármán condition means

$$\psi(N) = \psi(0), \tag{1.37}$$

or

$$\exp(-ikaN) = 1. \tag{1.38}$$

From this it follows that

$$kaN = 2\pi J, \tag{1.39}$$

where $J = 0, \pm 1, \pm 2, \ldots$. This means that only *some* k are allowed, namely, $k = \frac{2\pi}{a}\frac{J}{N}$.

The Bloch functions take the form (cf. Eq. (1.29))

$$\sum_j \exp(ikja)\chi_j, \tag{1.40}$$

where χ_j denotes a given atomic orbital (e.g., $1s$) centered on atom j. The summation over j in our case is finite, because we only have N atoms ($j = 0, 1, 2, \ldots, N-1$). Let us consider

$J = 0, 1, 2, \ldots, N-1$ and the corresponding values of $k = \frac{2\pi}{a}\frac{J}{N}$. For each k we have a Bloch function; altogether we have, therefore, N Bloch functions. Now, we may try to increase J and take $J = N$. The corresponding Bloch function may be written as

$$\sum_j \exp(i2\pi j)\chi_j = \sum_j \chi_j, \tag{1.41}$$

which turns out to be identical to the Bloch function with $k = 0$, i.e., with $J = 0$. We are reproducing what we already have. It is clear, therefore, that we have a set of those k that form a *complete set of nonequivalent states*; they correspond to $J = 0, 1, 2, \ldots N-1$. It is also seen that if the limits of this set are shifted by the same integer (like, e.g., $J = -3, -2, -1, 0, 1, 2, ..., N-4$), then we still have the same complete set of nonequivalent states. Staying for the time being with our primary choice of the set, we will get N values of $k \in [0, \frac{2\pi}{a}\frac{N-1}{N}]$, i.e., $k \in \{0, \frac{2\pi}{a}\frac{1}{N}, \frac{2\pi}{a}\frac{2}{N}, \ldots \frac{2\pi}{a}\frac{N-1}{N}\}$. *Those k values are equidistant*. When $N \to \infty$, the section to be divided attains the length $\frac{2\pi}{a}$. Hence

> the nonequivalent states (going with N to infinity) correspond to those k's that are from section $[0, \frac{2\pi}{a}]$ or shifted section $[-\frac{\pi}{a}, +\frac{\pi}{a}]$, called the FBZ. From now on we will adopt this last choice, i.e., $[-\frac{\pi}{a}, +\frac{\pi}{a}]$. We are allowed to make any shift, because, as we have shown, we keep the same nonequivalent values of k. The allowed k values are distributed *uniformly* within the FBZ. The number of the allowed k's is equal to ∞, because $N = \infty$ (and the number of the allowed k's is always equal to N).

1.7.3 k dependence of orbital energy

Note that the higher energy of a molecular orbital (in our case they are identical to the Bloch functions), the more nodes they have. Let us take the example of benzene ($N = 6$, Fig. 1.8) and consider only those molecular orbitals that can be written as linear combinations of the carbon $2p_z$, where z is the axis orthogonal to the plane of the molecule. The wave vectors[23] ($k = \frac{2\pi}{a}\frac{J}{N}$) may be chosen as corresponding to $J = 0, 1, 2, \ldots, 5$, or equivalently to $J = -3, -2, -1, 0, +1, +2$. It is seen that $J = 0$ gives a nodeless function,[24] $J = \pm 1$ lead to a pair of the Bloch functions with a single node, $J = \pm 2$ give a pair of the two-node functions, and finally $J = -3$ corresponds to a three-node function.

It has occasionally been remarked in this book (cf., e.g., Chapter V1-4), that increasing the number of nodes[25] results in higher energy. This rule becomes most transparent in the present

[23] In this case this is a wave number.

[24] We neglect here the node that follows from the reflection in the molecular plane as being shared by all the molecular orbitals considered.

[25] That is, considering another wave function that has a larger number of nodes.

case (see Fig. 1.8). A nodeless Bloch function means all the contacts between the $2p$ orbitals are π *bonding*, which results in *low energy*. A single node means introducing two nearest neighbor π *antibonding* interactions, and this causes an energy increase. Two nodes result in four antibonding interactions, and the energy goes up even more. Three nodes already give all the nearest neighbor contacts of antibonding character and the energy is the highest possible.

1.8 A triple role of the wave vector

As has already been said, the wave vector (in 1D, 2D, and 3D) plays several roles. They are the following:

1. The wave vector $\mathbf{k}$ tells us which type of plane wave arranged from certain objects (like atomic orbitals) we are concerned with. The direction of $\mathbf{k}$ is the propagation direction, the wave length is $\lambda = \frac{2\pi}{|\mathbf{k}|}$.

2. The wave vector may also be treated as a label for the irreducible representation of the translational group.

   In other words, $\mathbf{k}$ determines which irreducible representation we are dealing with (Appendix V1-C on p. V1-605). This means that $\mathbf{k}$ tells us *which permitted rhythm* is exhibited by the coefficients at atomic orbitals in a particular Bloch function (permitted, i.e., ensuring that the square has the symmetry of the crystal). There are a lot of such rhythms, e.g., all the coefficients equal each other ($k = 0$), or one node introduced, two nodes, etc. The FBZ represents a set of such $\mathbf{k}$, which corresponds to *all possible rhythms*, i.e., nonequivalent Bloch functions.[26] In other words, the FBZ gives us all the possible symmetry orbitals that can be formed from an atomic orbital.

3. The longer the $\mathbf{k}$, the more nodes the Bloch function $\phi_{\mathbf{k}}$ has: $|\mathbf{k}| = 0$ means no nodes, and at the boundary of the FBZ there is the maximum number of nodes.

1.9 Band structure

1.9.1 Born–von Kármán boundary condition in 3D

The Hamiltonian $\widehat{H}$ we were discussing represents an effective one-electron Hamiltonian, its form not yet given. From Chapter V1-8, we know that it may be taken as the Fock operator.

[26] That is, linearly independent.

A crystal represents nothing but a huge (quasiinfinite) molecule, and assuming the Born–von Kármán condition, a huge cyclic molecule.

> This is how we will get the Hartree–Fock solution for the crystal – by preparing the Hartree–Fock solution for a cyclic molecule and then letting the number of unit cells N go to infinity.

Hence, let us take a large piece of crystal – a parallelepiped with the number of unit cells *in each of the periodicity directions* (i.e., along the three basis vectors) equal to $2N + 1$ (the reference cell 0, N cells on the right, N cells on the left). The particular number, $2N + 1$, is not very important; we have only to be sure that such a number is large. We assume that the Born–von Kármán condition is fulfilled. This means that we treat the crystal like a snake eating its tail, and this will happen on each of the three periodicity axes. This enables us to treat the translational group as a cyclic group, which gives an enormous simplification to our task (no end effects, all cells equivalent). The cyclic group of the lattice constants a, b, c implies that (cf. Eq. (1.38))

$$\exp\left(ik_x a(2N+1)\right) = 1, \tag{1.42}$$
$$\exp\left(ik_y b(2N+1)\right) = 1, \tag{1.43}$$
$$\exp\left(ik_z c(2N+1)\right) = 1, \tag{1.44}$$

which can be satisfied only for some special vectors $\mathbf{k} = (k_x, k_y, k_z)$ satisfying

$$k_x = \frac{2\pi}{a}\,\frac{J_x}{2N+1}, \tag{1.45}$$
$$k_y = \frac{2\pi}{b}\,\frac{J_y}{2N+1}, \tag{1.46}$$
$$k_z = \frac{2\pi}{c}\,\frac{J_z}{2N+1}, \tag{1.47}$$

with any of J_x, J_y, J_z taking $(2N + 1)$ consecutive integer numbers. We may, for example, assume that $J_x, J_y, J_z \in \{-N, -N+1, \ldots, 0, 1, 2, \ldots, N\}$. Whatever N is, $\mathbf{k}$ will always satisfy

$$-\frac{\pi}{a} < k_x < \frac{\pi}{a}, \tag{1.48}$$
$$-\frac{\pi}{b} < k_y < \frac{\pi}{b}, \tag{1.49}$$
$$-\frac{\pi}{c} < k_z < \frac{\pi}{c}, \tag{1.50}$$

which is what we call the FBZ. We may therefore say that before letting $N \to \infty$

> the FBZ is filled with the allowed vectors $\mathbf{k}$ in a *grain-like way*, the number being equal to the number of unit cells, i.e., $(2N+1)^3$. Note that the distribution of the vectors allowed in the FBZ is *uniform*. This is ensured by the numbers J, which divide the axes k_x, k_y, k_z in the FBZ into equal pieces.

1.9.2 Crystal orbitals from Bloch functions (LCAO CO method)

What we expect to obtain finally in the Hartree–Fock method for an infinite crystal are the molecular orbitals, which in this context will be called the *crystal orbitals* (COs). As usual we will plan to expand the CO as a linear combination of atomic orbitals (cf. p. V1-499). Which atomic orbitals? Well, those which we consider appropriate[27] for a satisfactory description of the crystal, e.g., the atomic orbitals of all the atoms of the crystal. We feel, however, that we are going to suffer a big defeat trying to perform this task.

> There will be a lot of atomic orbitals, and therefore also an astronomic number of integrals to compute (infinite for the infinite crystal) and that is it, we cannot help this. However, if we begin such a hopeless task, the value of any integral would repeat an infinite number of times. This indicates a chance to simplify the problem. Indeed, we have not yet used the translational symmetry of the system.

If we are going to use the symmetry, then *we may create the Bloch functions representing the building blocks that guarantee the proper symmetry in advance*. Each Bloch function is built from an atomic orbital χ, i.e.,

$$\phi_{\mathbf{k}} = (2N+1)^{-\frac{3}{2}} \sum_j \exp(i\mathbf{k}\mathbf{R}_j)\chi(\mathbf{r}-\mathbf{R}_j). \tag{1.51}$$

The function is identical to that of Eq. (1.29), except it has a factor $(2N+1)^{-\frac{3}{2}}$, which makes the function approximately normalized.[28]

Any CO will be a linear combination of such Bloch functions, each corresponding to a given χ. This is equivalent to the LCAO expansion for molecular orbitals; the only difference is

[27] As for molecules.

[28] The function without this factor is of class Q, i.e., normalizable for any finite N, but nonnormalizable for $N = \infty$. The approximate normalization makes the function square integrable, even for $N = \infty$. Let us see. We

that we have cleverly preorganized the atomic orbitals (of one type) into symmetry orbitals (Bloch functions). Hence, it is indeed appropriate to call this approach the LCAO CO method *(linear combination of atomic orbitals, crystal orbitals*), analogous to the LCAO MO method (cf. p. V1-503). There is, however, a problem. Each CO should be a linear combination of the $\phi_{\mathbf{k}}$ for various types of χ and for *various* $\mathbf{k}$. Only then would we have the full analogy: a molecular orbital is a linear combination of all the atomic orbitals belonging to the atomic basis set.[29]

It will be shown below that the situation is far better.

> Each CO corresponds to a single vector $\mathbf{k}$ from the FBZ and is a linear combination of the Bloch functions, each characterized by $\mathbf{k}$.

There are, however, only a few Bloch functions – their number is equal to the number of the atomic orbitals per unit cell (denoted by ω).[30]

It is easy to show that, indeed, we can limit ourselves to a single vector $\mathbf{k}$. Imagine this is false, and our CO is a linear combination of all the Bloch functions corresponding to a given $\mathbf{k}$. Then, it is of all the Bloch functions corresponding to the next $\mathbf{k}$, etc., up to the exhaustion of all the allowed $\mathbf{k}$. When, in the next step, we solve the orbital equation with the effective (i.e., Fock) Hamiltonian using the Ritz method, then we will end up computing the integrals $\langle \phi_{\mathbf{k}} | \widehat{F} \phi_{\mathbf{k}'} \rangle$ and

---

have

$$\langle \phi_{\mathbf{k}} | \phi_{\mathbf{k}} \rangle = (2N+1)^{-3} \sum_j \sum_{j'} \exp\left(i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_{j'})\right) \int \chi(\mathbf{r} - \mathbf{R}_j)\chi(\mathbf{r} - \mathbf{R}_{j'}) d\tau =$$

$$= (2N+1)^{-3} \sum_j \sum_{j'} \exp\left(i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_{j'})\right) \int \chi(\mathbf{r})\chi\left(\mathbf{r} - (\mathbf{R}_j - \mathbf{R}_{j'})\right) d\tau,$$

because the integral does depend on a *relative* separation in space of the atomic orbitals. Further,

$$\langle \phi_{\mathbf{k}} | \phi_{\mathbf{k}} \rangle = \sum_j \exp(i\mathbf{k}\mathbf{R}_j) \int \chi(\mathbf{r})\chi(\mathbf{r} - \mathbf{R}_j) d\tau, \tag{1.52}$$

because we can replace a double summation over j and j' by a double summation over j and $j'' = j - j'$ (both double summations exhaust all the lattice nodes), and the latter summation always gives the same independent of j; the number of such terms is equal to $(2N+1)^3$. Finally, we may write $\langle \phi_{\mathbf{k}} | \phi_{\mathbf{k}} \rangle = 1 +$ various integrals. The largest of these integrals is the nearest neighbor overlap integral of the functions χ. For normalized χ each of these integrals represents a fraction of 1. Additionally the contributions for further neighbors decay exponentially (cf. p. V1-735). As a result, $\langle \phi_{\mathbf{k}} | \phi_{\mathbf{k}} \rangle$ is a number of the order of 1 or 2. This is what we have referred to as an approximate normalization.

29 Indeed, for any $\mathbf{k}$ the number of distinct Bloch functions is equal to the number of atomic orbitals per unit cell. The number of allowed vectors, $\mathbf{k}$, is equal to the number of unit cells in the crystal. Hence, using the Bloch functions for all allowed $\mathbf{k}$ would be justified, any CO would represent a linear combination of all the atomic orbitals of the crystal.

30 Our optimism pertains, of course, to taking a modest atomic basis set (small ω).

$\langle\phi_{\mathbf{k}}|\phi_{\mathbf{k}'}\rangle$. For $\mathbf{k} \neq \mathbf{k}'$ such integrals *are equal to zero* according to group theory (Appendix V1-C on p. V1-605), because $\widehat{F}$ transforms according to the fully symmetric irreducible representation of the translational group,[31] while $\phi_{\mathbf{k}}$ and $\phi_{\mathbf{k}'}$ transform according to *different* irreducible representations.[32] Therefore the secular determinant in the Ritz method will have a *block form* (cf. Appendix V1-C). The first block will correspond to the first **k**, the second to the next **k**, etc., where every block[33] would look as if in the Ritz method we used the Bloch functions corresponding uniquely to that particular **k**. The conclusion is the following: since a CO has to be a wave with a *given* **k**, let us construct it with Bloch functions, which already have just this type of behavior with respect to translation operators, i.e., have just this **k**. This is fully analogous with the situation in molecules, if we used atomic symmetry orbitals.[34]

Thus, each vector **k** from the FBZ is associated with a crystal orbital, and therefore with a set of LCAO CO coefficients.

The number of such CO sets (each **k** – one set) *in principle* has to be equal to the number of unit cells, i.e., infinite.[35] The only profit we may expect could be associated with the hope that the computed quantities do not depend on **k** too much, but will rather change smoothly when **k** changes. This is indeed what will happen, *then a small number of vectors* **k** will be used, *and the quantities requiring other* **k** *will be computed by interpolation*.

Only a part of the computed COs will be occupied, and this depends on the orbital energy of a given CO, the number of electrons, and the corresponding **k**, similarly to what we had for molecules.

The set of SCF LCAO CO equations will be very similar to the set for the molecular orbital method (SCF LCAO MO). In principle, the only difference will be that in the crystal case we will consequently use symmetry orbitals (Bloch functions) instead of atomic orbitals.

31 Unit cells (by definition) are identical.

32 Recall that **k** also has the meaning of the irreducible representation index (of the translational group).

33 The whole problem can be split into independent problems for individual blocks.

34 A symmetry atomic orbital (SAO) represents such linear combination of equivalent-by-symmetry atomic orbitals that transforms according to one of the irreducible representations of the symmetry group of the Hamiltonian. Then, when molecular orbitals (MOs) are formed in the LCAO MO procedure, any given MO is a linear combination of the SAOs belonging to a particular irreducible representation. For example, a water molecule exhibits the symmetry plane (σ) that is perpendicular to the plane of the molecule. An MO which is symmetric with respect to σ contains the symmetry atomic orbital $1s_a + 1s_b$, which is symmetric with respect to σ, but does not contain the symmetry atomic orbital $1s_a - 1s_b$.

35 Well, we cannot fool Mother Nature! Was there an infinite molecule (crystal) to be computed or not? Then the number of such sets of computations has to be infinite full stop.

That is it. The rest of this section is associated with some technical details accompanying the operation $N \to \infty$.

1.9.3 SCF LCAO CO equations

Let us write down the SCF LCAO CO equations as if they corresponded to a large molecule (Bloch functions will be used instead of atomic orbitals). Then the n-th CO may be written as

$$\psi_n(\mathbf{r}, \mathbf{k}) = \sum_q c_{qn}(\mathbf{k}) \phi_q(\mathbf{r}, \mathbf{k}), \tag{1.53}$$

where ϕ_q is the Bloch function corresponding to the atomic orbital χ_q, i.e.,

$$\phi_q(\mathbf{r}, \mathbf{k}) = (2N+1)^{-\frac{3}{2}} \sum_j \exp(i\mathbf{k}\mathbf{R}_j) \chi_q^j, \tag{1.54}$$

with $\chi_q^j \equiv \chi_q(\mathbf{r} - \mathbf{R}_j)$ (for $q = 1, 2, \ldots, \omega$).

The symbol χ_q^j means the q-th atomic orbital (from the set we prepared for the unit cell motif) located in the cell indicated by vector $\mathbf{R}_j$ (j-th cell).

In the expression for ψ_n, we have taken into account that there is no reason whatsoever that the coefficients c were $\mathbf{k}$-independent, since the expansion functions ϕ depend on $\mathbf{k}$. This situation does not differ from that which we encountered in the Hartree–Fock–Roothaan method (cf. p. V1-506), with one technical exception: instead of the atomic orbitals we have symmetry orbitals, in our case Bloch functions.

The secular equations for the Fock operator will have, of course, the form of the Hartree–Fock–Roothaan equations (cf. Chapter V1-8, p. V1-505):

$$\sum_{q=1}^{\omega} c_{qn}[F_{pq} - \varepsilon_n S_{pq}] = 0$$

for $p = 1, 2, \ldots, \omega$,

where the usual notation has been applied. For the sake of simplicity, we have not highlighted the $\mathbf{k}$ dependence of c, F, and S. Whenever we decide to do this in future, we will put it in the form $F_{pq}(\mathbf{k})$, $S_{pq}(\mathbf{k})$, etc. Of course, ε_n will become a function of $\mathbf{k}$, as will be stressed by the symbol $\varepsilon_n(\mathbf{k})$. Theoretically, the secular equation has to be solved for every $\mathbf{k}$ of the FBZ.

Therefore, despite the fact that the secular determinant is of rather low rank (ω), the infinity of the crystal forces us to solve this equation an infinite number of times. For the time being, though, do not worry too much.

1.9.4 Band width

The number of secular equation solutions is equal to ω, and let us label them using index n. If we focus on one such solution and check how $\varepsilon_n(\mathbf{k})$ and $\psi_n(\mathbf{r}, \mathbf{k})$ are sensitive to a tiny change of $\mathbf{k}$ within the FBZ, it turns out that $\varepsilon_n(\mathbf{k})$ and $\psi_n(\mathbf{r}, \mathbf{k})$ change smoothly. This may not be true when $\mathbf{k}$ passes through the border of the FBZ.

> The function $\varepsilon_n(\mathbf{k})$ is called the n-th electronic band.

If we traveled in the FBZ, starting from the origin and continuing along a straight line, then ε_1, ε_2, ..., etc., would change as functions of $\mathbf{k}$ and we would be concerned with several energy bands. If $\varepsilon_n(\mathbf{k})$ changes very much during our travel over the FBZ, we would say that the n-th band has large width or *dispersion*.

As shown on p. 22 for hydrogen atoms, an energy band forms due to the bonding and antibonding effects, the energy splitting being of the order of the overlap integral between the nearest neighbor $1s$ atomic orbitals. If instead of hydrogen atoms, we put a unit cell with a few atoms inside (motif), then the story is similar: the motif has some one-electron energy levels (orbital energies), putting together the unit cells makes these energy levels change into energy bands, and the number of levels in any band is equal to the number of unit cells, or the number of allowed $\mathbf{k}$ vectors in the FBZ.

> The band width is related to interactions among the unit cell contents, and is roughly proportional to the overlap integral between the orbitals of the interacting unit cells.

How do we plot the band structure? For the one-dimensional crystal, e.g., a periodic polymer, there is no problem: the wave vector $\mathbf{k}$ means the number k changes from $-\frac{\pi}{a}$ to $\frac{\pi}{a}$, and we plot the function $\varepsilon_n(k)$. For each n we have a single plot, e.g., for the hydrogen atom the band ε_1 collects energies resulting from the $1s$ atomic orbital interacting with other atoms, the band ε_2 those from $2s$, etc. In the three-dimensional case we usually choose a path in the FBZ. We start from the point Γ defined as $\mathbf{k} = \mathbf{0}$. Then, we continue to some points located on the faces and edges of the FBZ surface. It is impossible to go through the whole FBZ. The band structure in the three-dimensional case is usually shown by putting the described itinerary through the FBZ on the abscissa (Fig. 1.9), and $\varepsilon_n(\mathbf{k})$ on the ordinate. Fig. 1.9 shows an example of what we might obtain from such calculations.

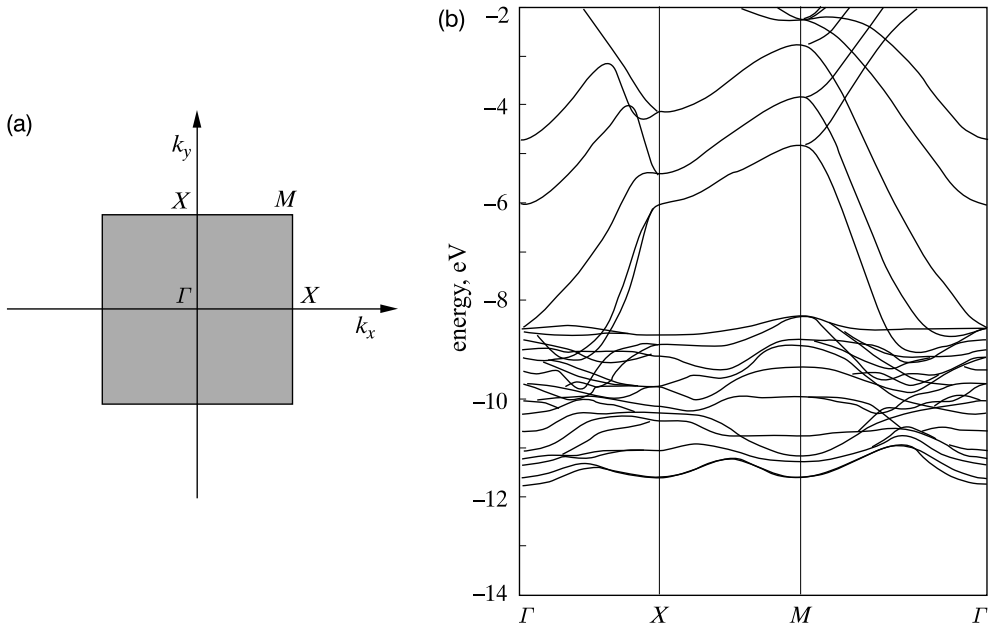


Fig. 1.9. (a) FBZ for four regular layers of nickel atoms (a crystal surface model), four characteristic points in the FBZ are shown. (b) The band structure for this system (for a particular itinerary within the FBZ). We see that we cannot understand much: just a horrible irregular mess of lines. All the band structures look equally clumsy. In spite of this, from such a plot we may determine the electrical and optical properties of the nickel slab. We will see later on why the bands have such a mysterious form. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

1.9.5 Fermi level and energy gap: insulators, metals, and semiconductors

Insulators

How many electrons do we have in a crystal? The answer is simple: the infinite crystal contains an infinite number of electrons. But infinities are often different. The decider is the number of electrons per unit cell. Let us denote this number by n_0.

If this means a double occupation of the molecular orbitals of the unit cell, then the corresponding band in the crystal will also be fully occupied, because the number of energy levels in a band is equal to the number of unit cells, and each unit cell contributes two electrons from the abovementioned molecular orbital. The bands that come from the valence orbitals of the motif are called *valence bands*. Therefore,

doubly occupied orbitals of the motif, related usually to the inner electronic shells, lead to fully occupied bands. Accordingly, singly occupied orbitals lead to bands that are *half-occupied*, while empty (virtual) orbitals lead to empty bands (unoccupied, or conduction bands).

The highest occupied crystal orbital is known as the Fermi level; it is equivalent to the HOMO of the crystal.[36] The HOMO and LUMO levels, fundamental as always, decide about the chemistry of the system, in our case the chemical and physical properties of the crystal.

The gap between the HOMO and LUMO of the crystal means the gap between the top of the occupied valence band and the bottom of the conduction band (Fig. 1.10). When the gap is large we have to do with insulators.

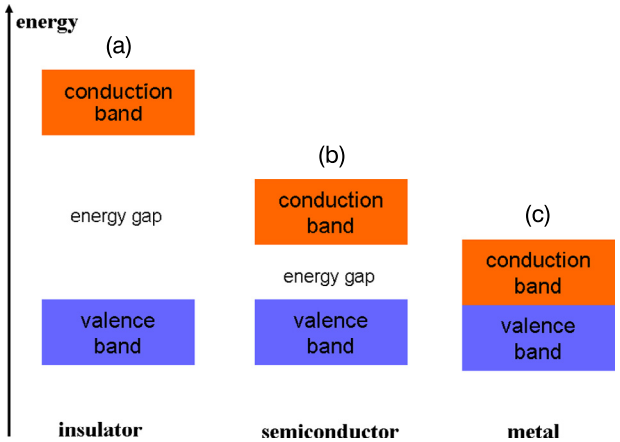


Fig. 1.10. Valence bands (highest occupied by electrons) and conduction bands (empty). The electric properties of a crystal depend on the energy gap between them (i.e., HOMO-LUMO separation). (a) A large gap is typical of insulators. (b) A medium gap means a semiconductor. (c) A zero gap is typical of metals.

Metals, one-dimensional metals, and Peierls distortion

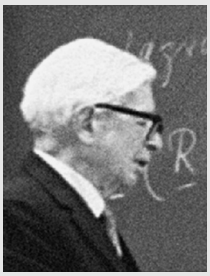
A partially filled band may lead to the situation with the band gap equal to zero.

[36] We sometimes find a thermodynamic definition of the Fermi level, but in this book it will always be the highest occupied crystal orbital.

A metal is characterized by having empty levels (conduction band) immediately (zero distance) above doubly occupied valence ones.

Metals, because of the zero gap, are conductors of electric current.[37]

The conductivity of the metallic systems is typically orientation-independent. In the last decades two-dimensional and one-dimensional metals (with anisotropy of conductivity) have been discovered. The latter are called *molecular wires* and may have unusual properties, but are difficult to prepare for they often undergo spontaneous dimerization of the lattice (known as the *Peierls transition*).



Rudolph Peierls (1907–1995), British physicist, professor at the universities of Birmingham and Oxford. Peierls participated in the Manhattan Project (atomic bomb) as leader of the British group.

As Fig. 1.11a shows, dimerization makes the bonding (and antibonding) effects stronger a little below (and above) the middle of the band, whereas at the band edges the effect is almost zero (since dimerization makes the bonding as well as antibonding effects cancel *within a pair of consecutive bonds*). As a result, the degeneracy is removed in the middle of the band (Fig. 1.11b), i.e., a band gap appears and the system undergoes metal–insulator transition (Fig. 1.11c). This is why polyacetylene, instead of having all the CC bonds equivalent (Fig. 1.11d), which would make it a metal, exhibits alternation of bond lengths (Fig. 1.11e), and it becomes an insulator or semiconductor.

To a chemist, the Peierls transition is natural. The hydrogen atoms will not stay equidistant in a chain, but will simply react and form hydrogen molecules, i.e., will dimerize like lightning. Also the polyacetylene will try to form π bonds by binding the carbon atoms in *pairs*. There is simply a shortage of electrons to keep *all* the CC bonds strong; *there are only enough for every second*, which means simply dimerization through creating π bonds. On the other hand, the Peierls transition may be seen as the Jahn–Teller effect: there is a degeneracy of the occupied and empty levels at the Fermi level, and it is therefore possible to lower the energy by removing the degeneracy through a distortion of geometry (i.e., dimerization). Both pictures are correct and represent the thing.

[37] When an electric field is applied to a crystal, its energy levels change. If the field is weak, then the changes may be computed by perturbation theory (treating the zero-field situation as the unperturbed one). This means that the perturbed states acquire some admixtures of the excited states (cf. Chapter V1-5). The lower the energy gap, the more mixing takes place. For metallic systems (gap zero), such perturbation theory certainly would not be applicable, but real excitation to the conduction band may take place.

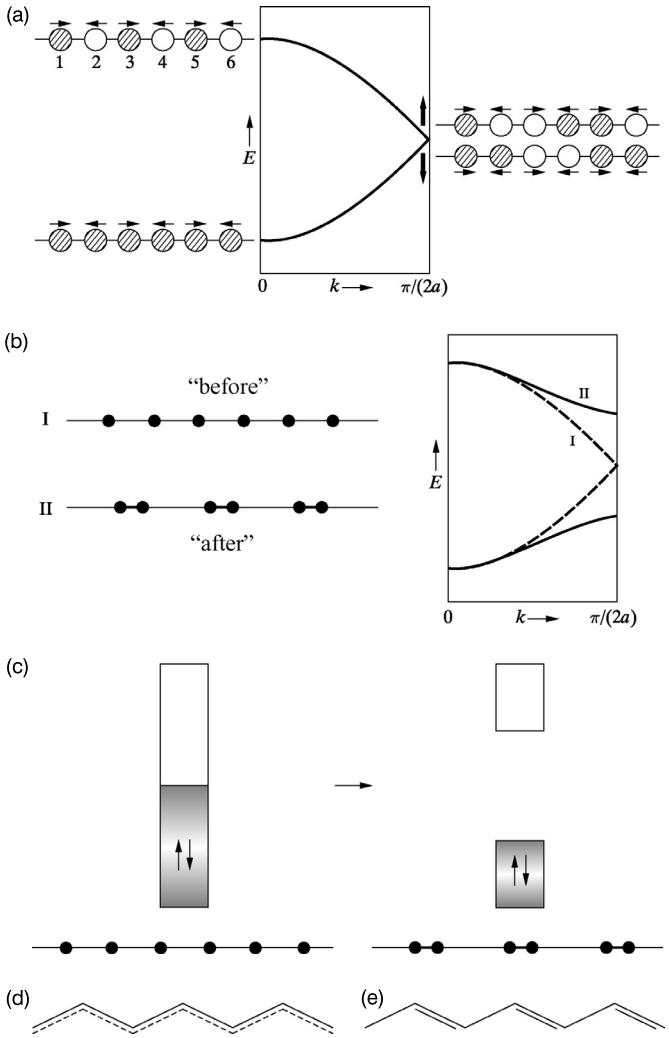


Fig. 1.11. The Peierls effect has the same origin as the Jahn–Teller effect in removing the electronic level degeneracy by distorting the system (H.A. Jahn, E. Teller, *Proc. Roy. Soc. A*, 161(1937)220). The electrons occupy half the FBZ, i.e., $-\frac{\pi}{2a} \leq k \leq \frac{\pi}{2a}$, with a standing for the nearest neighbor distance. The band has been plotted assuming that the period is equal to $2a$, hence a characteristic back folding of the band (similarly as we would fold a sheet of paper with band structure drawn, the period equal to a). (a) Lattice dimerization, shown by little arrows, *amplifies* the bonding and antibonding effects close to the middle of the FBZ, i.e., in the neighborhood of $k = \pm\frac{\pi}{2a}$. At the same time close to $k = 0$ there is a *cancellation* of the opposite effects: within bonding (bottom) and also within antibonding (top) interactions. (b) As a result, the degeneracy at $k = \frac{\pi}{2a}$ is removed and the band gap appears, which corresponds to lattice dimerization. (c) The system lowers its energy when undergoing metal–insulator or metal–semiconductor transition. (d) The polyacetylene chain, forcing equivalence of all CC bonds, represents a metal. (e) However, due to the Peierls effect, the system undergoes dimerization and becomes an insulator. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

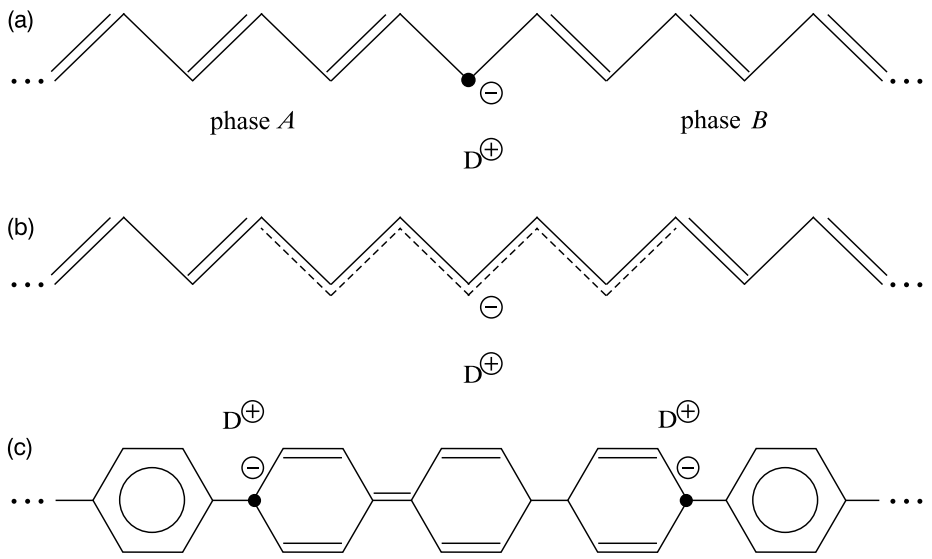


Fig. 1.12. Solitons and bipolarons as a model of electric conductivity in polymers. (a) Two phases of polyacetylene separated by a defect. Originally the defect was associated with an unpaired electron, but when a donor, D, gave its electron to the chain, the defect became negatively charged. (b) The energy of such a defect is independent of its position in the chain (important for charge transportation); in reality the change of phase takes place in sections of about 15 CC bonds, not two bonds as suggested in (a). Such a situation is sometimes modeled by a nonlinear differential equation, which describes a soliton motion ("solitary wave") that represents the traveling phase boundary. (c) In the polyparaphenylene chain two phases (low-energy aromatic and high-energy quinoid) are possible as well, but in this case they are of different energies. Therefore, the energy of a single defect (aromatic structures–kink–quinoid structures) depends on its position in the chain (hence, no charge transportation). However, a *double* defect with a (higher-energy) section of a quinoid structure has a position-independent energy, and when charged by dopants (*bipolaron*) can conduct electricity. The abovementioned polymers can be doped either by electron donors (e.g., arsenium, potassium) or electron acceptors (iodine), which results in a spectacular increase in their electric conductivity.

Polyacetylene (Fig. 1.12a,b), after doping, becomes ionized if the dopants are electron acceptors, or receives extra electrons if the dopant is an electron donor (symbolized by D^+ in Fig. 1.12). The perfect polyacetylene exhibits the bond alternation discussed above, but it may be that we have a defect that is associated with a region of "changing rhythm" (or "phase"): from[38] $(= - = - =)$ to $(- = - = -)$. Such a kink is sometimes described as a *soliton* wave (Fig. 1.12a,b), i.e., a "solitary wave" first observed in the 19th century in England on a water channel, where it preserved its shape while moving over a distance of several kilometers.

[38] This possibility was first recognized by J.A. Pople, S.H. Walmsley, *Mol. Phys.*, 5(1962)15, 15 years before the experimental discovery of this effect.

The soliton defects cause some new energy levels ("solitonic levels") to appear within the gap. These levels too form their own solitonic band.

Charged solitons may travel when subject to an electric field, and therefore the doped polyacetylene turns out to be a good conductor (organic metal).

In polyparaphenylene, soliton waves are impossible, because the two phases (aromatic and quinoid, Fig. 1.12c) differ in energy (low-energy aromatic phase and high-energy quinoid phase). However, when the polymer is doped, a charged double defect (*bipolaron,* Fig. 1.12c) may form, and such a defect may travel when an electric field is applied. Hence, the doped polyparaphenylene, similarly to the doped polyacetylene, is an "organic metal."

Controlling the metal Fermi level – an electrode

The Fermi level (i.e., HOMO level) is especially interesting in metals, because there are ways to change its position on the energy scale. We may treat the metal as a container for electrons: we may pump the electrons into it or make the electron deficiency in it by using it as a cathode or anode, respectively. Having a tunable HOMO level, we decide if and when our reactant (i.e., the electrode) acts as an electron donor or electron acceptor! This opens avenues like, e.g., polarography, when scanning the electrode potential results in consecutive electrode reactions occurring whenever the electrode Fermi level matches the LUMO of the substances present in the solution. Since the matching potentials are characteristic of the substances, this is a way of performing chemical identification together with quantitative analysis.

Semiconductors

An intrinsic semiconductor exhibits a conduction band separated by a small energy gap (band gap) from the valence band. (See Fig. 1.13a.)

If the empty energy levels of the dopant are located just over the occupied band of an intrinsic semiconductor, the dopant may serve as an electron acceptor for the electrons from the occupied band (thus introducing its own conduction band), and we have a *p-type semiconductor*, Fig. 1.13b. If the dopant energy levels are occupied and located just under the conduction band, the dopant may serve as an *n-type semiconductor*, Fig. 1.13c.

Among these three fundamental classes of materials (insulators, metals, and semiconductors), the semiconductors are most versatile as to their properties and practical applications. The metals just conduct electricity, and the carriers of the electric current are electrons. The metals' conductivity spans only one order of magnitude. The insulators are useful, because they do not conduct electric current at all. In contrast to the metals and the insulators, the conductivity of

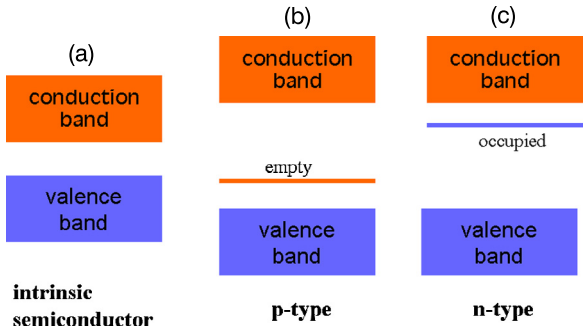


Fig. 1.13. Energy bands for semiconductors. (a) Intrinsic semiconductor (small gap). (b) *p*-type semiconductor (electron acceptor levels close to the occupied band). (c) *n*-type semiconductor (electron donor levels close to the conduction band).

semiconductors can be controlled within many orders of magnitude (mainly by doping, i.e., admixture of other materials). The second extraordinary feature is that only in semiconductors the conductivity can be tuned by using two types of the charge carriers: (negative) electrons and (positive) electron holes. The results of such tuning depends on temperature, light, and the electric and magnetic fields. In contrast to metals and insulators, the semiconductors are able to emit visible light. All these features make it possible to tailor functional semiconductor devices with versatile electric and photonic properties. This is why in practically any electric or photonic equipment a semiconductor device is operating.

Additional reasons why organic metals and semiconductors are of practical interest include their versatility and tunability (precision) offered by the kingdom of organic chemistry, easy processing typical of the plastics industry, the ability to literally bend the device without loosing its properties, and last but not least low weight.

What kind of substances are semiconductors? Well, their most important class can be derived directly from a section of the Mendeleev periodic table (the first row shows the group number, Table 1.1): (i) *IV–IV semiconductors*: the elemental semiconductors C, Si, Ge, as well as the compounds SiGe, SiC, (ii) *III–V semiconductors*: GaS, GaN, GaP, InP, InSb, etc., (iii) *II–IV semiconductors*: CdSe, CdS, CdTe, ZnO, ZnS, etc.

1.10 Solid state quantum chemistry

A calculated band structure, with information about the position of the Fermi level, tells us a lot about the electric properties of the material under study (insulator, semiconductor, metal). They tell us also about basic optical properties; e.g., the band gap indicates what kind of absorption

Table 1.1. A "semiconductor section" of the Mendeleev periodic table.

II	III	IV	V	VI
	B	C	N	O
	Al	Si	P	S
Zn	Ga	Ge	As	Se
Cd	In	Sn	Sb	Te

spectrum we may expect. We can calculate any measurable quantity, because we have at our disposal the computed (though approximate) wave function.

However, despite this very precious information, which is present in the band structure, there is a little worry. When we stare at any band structure, such as that shown in Fig. 1.9, the overwhelming feeling is a kind of despair. All band structures look similar, well, just a tangle of plots. Some go up, some down, some stay unchanged, some, it seems without any reason, change their direction. Can we understand this? What is the theory behind this band behavior?

1.10.1 Why do some bands go up?

Let us take our beloved chain of hydrogen atoms in the 1*s* state, to which we already owe so much (Fig. 1.14).

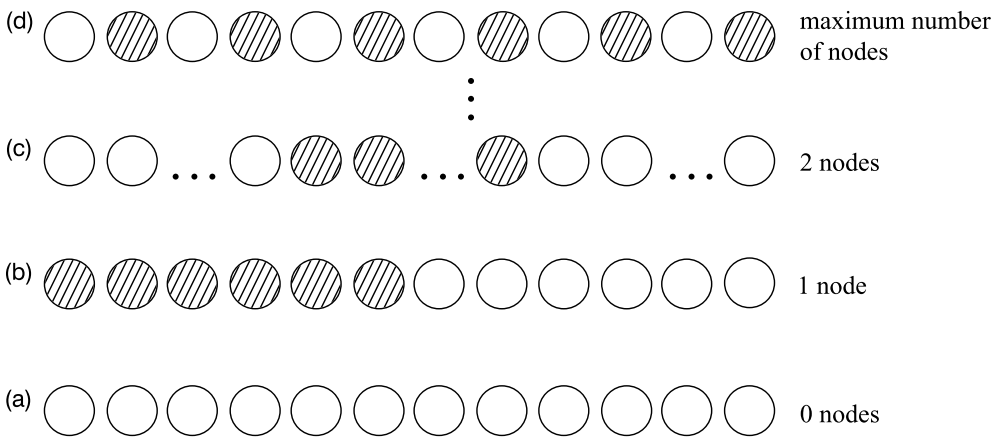


Fig. 1.14. The infinite chain of ground-state hydrogen atoms and the role of bonding and antibonding effects. (a) All interactions are bonding. (b) Introduction of a single node results in an energy increase. (c) Two nodes increase the energy even more. (d) Maximum number of nodes – the energy is the highest possible.

When will the state of the chain have the lowest energy possible? Of course, when all the atoms interact in a bonding, and not antibonding, way. This corresponds to Fig. 1.14a (no nodes of the wave function). When in this situation we introduce a single nearest neighbor antibonding interaction, the energy will for sure increase a bit (Fig. 1.14b). When two such interactions are introduced (Fig. 1.14c), the energy goes up even more, and the plot corresponds to two nodes. Finally, in the highest-energy situation all nearest neighbor interactions are antibonding (maximum number of nodes, Fig. 1.14d). Let us recall that the wave vector was associated with the number of nodes. Hence, if k increases from zero to $\frac{\pi}{a}$, the energy increases from the energy corresponding to the nodeless wave function to the energy characteristic for the maximum-node wave function. We understand, therefore, that some band plots are such as in Fig. 1.15a.

1.10.2 Why do some bands go down?

Sometimes the bands go in the opposite direction: the lowest energy corresponds to $k = \frac{\pi}{a}$, the highest energy to $k = 0$. What happens over there? Let us once more take the hydrogen atom chain, this time, however, in the $2p_z$ state (z is the periodicity axis). This time the Bloch function corresponding to $k = 0$, i.e., a function that follows just from locating the orbitals $2p_z$ side by side, describes the highest-energy interaction – the *nearest neighbor interactions are all antibonding*. Introduction of a node (increasing k) means a relief for the system – instead of one painful antibonding interaction we get a soothing bonding one. The energy goes down. No wonder, therefore, some bands look like those shown in Fig. 1.15b.

1.10.3 Why do some bands stay constant?

According to numerical rules (p. V1-503) inner shell atomic orbitals do not form effective linear combinations (crystal orbitals). Such orbitals have large exponential coefficients and resulting overlap integral, and therefore the band width (bonding versus antibonding effect) is negligible. This is why the nickel 1s orbitals (deep-energy level) result in a low-energy band of almost zero width (Fig. 1.15c), i.e., staying flat as a pancake all the time. Since they are always of very low energy, they are doubly occupied and their plot is so boring, they are not even displayed (they are absent in Fig. 1.9).

1.10.4 More complex behavior explainable – examples

We understand, therefore, at least why some bands are monotonically going down, some up, and some stay constant. In explaining these cases, we have assumed that a given CO is dominated by a single Bloch function. Other behaviors can be explained as well by detecting what kind of Bloch function *combination* we have in a given crystal orbital.

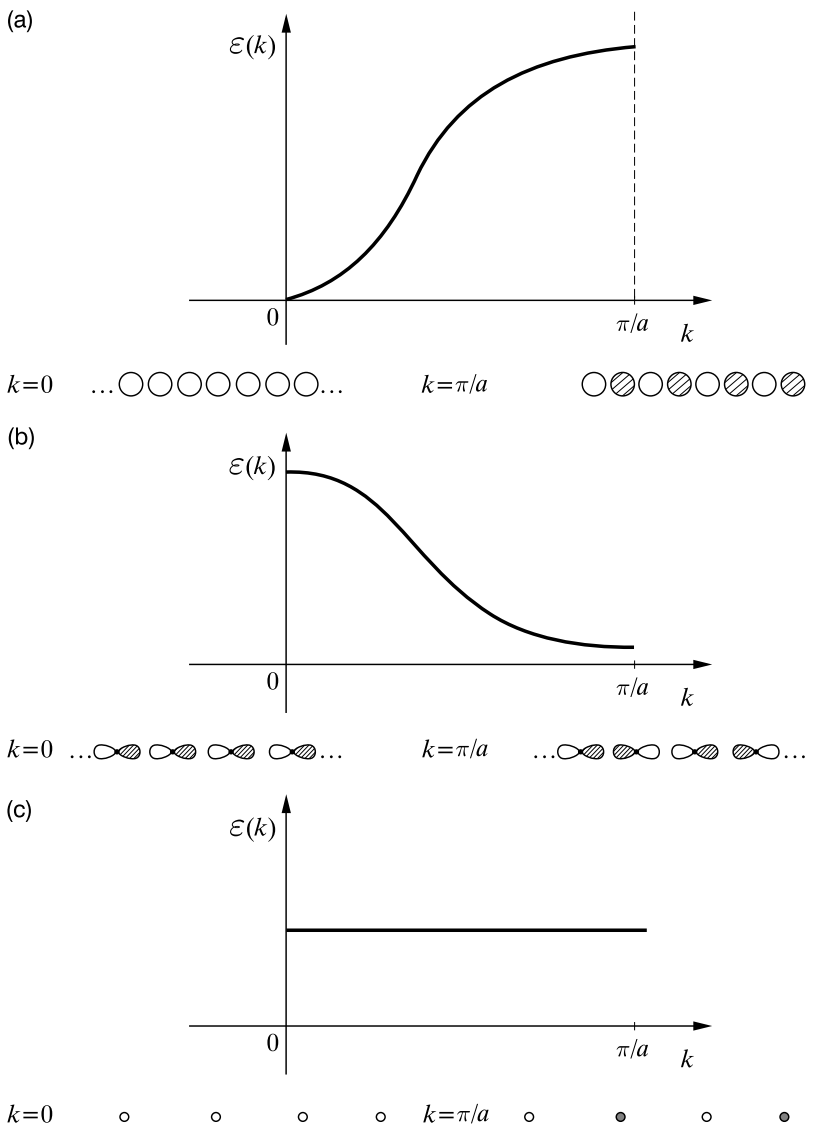


Fig. 1.15. Three typical band plots in the FBZ. (a) $1s$ orbitals. Increasing k is accompanied by an *increase* of the antibonding interactions and this is why the energy goes up. (b) $2p_z$ orbitals (z denotes the periodicity axis). Increasing k results in *decreasing* the number of antibonding interactions and the energy goes down. (c) Inner shell orbitals. The overlap is small as it is, and therefore the band width is practically zero.

Two-dimensional regular lattice of hydrogen atoms

Let us take a planar regular lattice of hydrogen atoms in their ground state.[39] Fig. 1.9 shows the FBZ of a similar lattice. We (arbitrarily) choose as the itinerary through the FBZ: Γ–X–M–Γ. From Fig. 1.7a we easily deduce that the band energy for the point Γ has to be the lowest, because it corresponds to all the interaction bonding. What will happen at the point X (Fig. 1.9a), which corresponds to $\mathbf{k} = (\pm\frac{\pi}{a}, 0)$ or $\mathbf{k} = (0, \pm\frac{\pi}{a})$? This situation is related to Fig. 1.7b. If we focus on any of the hydrogen atoms, it has four nearest neighbor interactions: two bonding and two antibonding. This, to a good approximation, corresponds to the nonbonding situation (hydrogen atom ground-state energy), because the two effects nearly cancel each other out. Halfway between Γ and X, we go through the point that corresponds to Fig. 1.7c,d. For such a point, any hydrogen atom has two bonding and two nonbonding interactions, i.e., the energy is the average of the Γ and X energies. The point M is located in the corner of the FBZ, and corresponds to Fig. 1.7e. All the nearest neighbor interactions are antibonding there, and the energy will be very high. We may, therefore, anticipate a band structure of the kind sketched in Fig. 1.16a. The figure has been plotted to reflect the fact that the density of states for the band edges is the largest, and therefore the slope of the curves has to reflect this. Fig. 1.16b shows the results of the computations.[40] It is seen that even very simple reasoning may rationalize the main features of band structure plots.

Trans-polyacetylene (regular one-dimensional polymer)

Polyacetylene already has quite a complex band structure, but as usual the bands close to the Fermi level (valence bands) are the most important in chemistry and physics. All these bands are of the π type, i.e., their COs are antisymmetric with respect to the plane of the polymer. Fig. 1.17 shows how the valence bands are formed. We can see the principle is identical to that for the chain of hydrogen atoms: the more nodes, the higher the energy. The highest energy corresponds to the band edge.

The resulting band is only *half-filled* (metallic regime), because each of the carbon atoms offers one electron, and the number of COs is equal to the number of carbon atoms (each CO can accommodate two electrons). Therefore, the Peierls mechanism (Fig. 1.11) is bound to enter into play, and in the middle of the band a gap will open. The system is, therefore, predicted to be an insulator (or semiconductor) and indeed it is. It may change to a metal when doped. Fig. 1.17 shows a situation analogous to the case of a chain of the ground-state hydrogen atoms.

---

39 A chemist's first thought would be that this could never stay like this when the system is isolated. We are bound to observe the formation of hydrogen molecules.

40 R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*," VCH Publishers, New York, 1988.

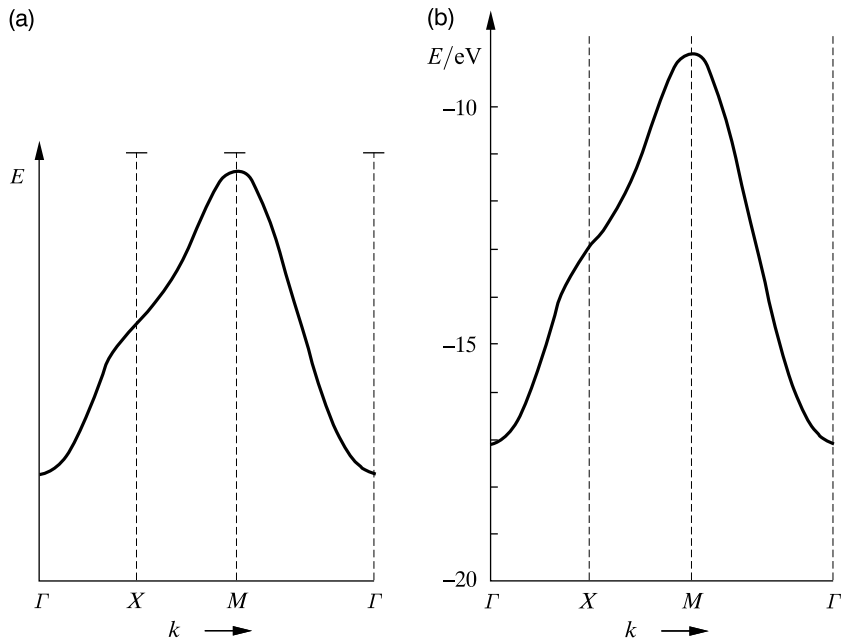


Fig. 1.16. (a) A sketch of the valence band for a regular planar lattice of ground-state hydrogen atoms. (b) The valence band as computed in the laboratory of Roald Hoffmann for a nearest neighbor distance equal to 2 Å. The similarity of the two plots confirms that we are able, at least in some cases, to predict band structure. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

Polyparaphenylene

The extent to which the COs conform to the rule of increasing number of nodes with energy (or k) will be seen in the example of a planar conformation of polyparaphenylene.[41] On the left-hand side of Fig. 1.18 we have the valence π orbitals of benzene:

- first, the lowest-energy nodeless[42] doubly occupied molecular orbital φ_1,
- then, we have a doubly degenerate and fully occupied level with the corresponding orbitals, φ_2 and φ_3, each having a single node,
- next, a similar doubly degenerate empty level with orbitals φ_4 and φ_5 (each with two nodes),
- and finally, the highest-energy empty three-node orbital φ_6.

[41] J.-M. André, J. Delhalle, J.-L. Brédas, "*Quantum Chemistry Aided Design of Organic Polymers*," World Scientific, Singapore, 1991.

[42] Besides the nodal plane of the nuclear framework.

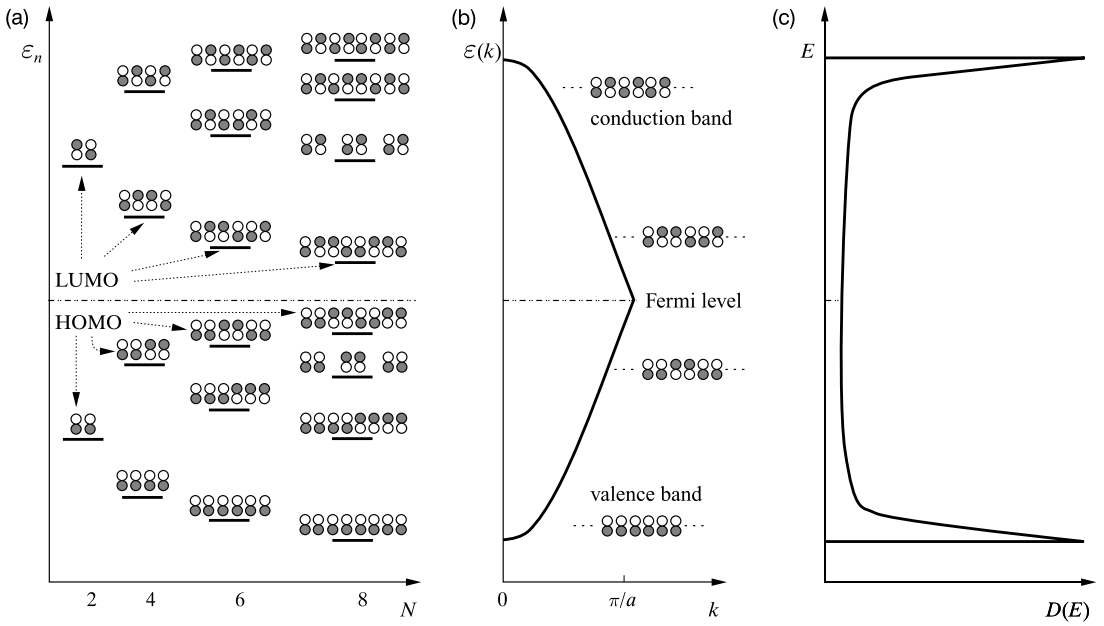


Fig. 1.17. (a) π band formation in polyenes (N stands for the number of carbon atoms) with the assumption of CC bond equivalence (each has length $a/2$). For $N = \infty$ this gives the metallic solution (no Peierls effect). As we can see, the band formation principle is identical to that which we have seen for hydrogen atoms. (b) Band structure. (c) density of states $D(E)$, i.e., the number of states per energy unit at a given energy E. The density has maxima at the extremal points of the band. If we allowed the Peierls transition, at $k = \pm\pi/a$ we would have a gap. J.-M. André, J. Delhalle, J.-L. Brédas, "*Quantum Chemistry Aided Design of Organic Polymers*", World Scientific, Singapore, 1991. Reprinted with permission from the World Scientific Publishing Co. Courtesy of the authors.

Thus, even in the single monomer we have the rule (of energy increasing with the number of nodes) fulfilled.

Binding phenyl rings by using CC σ bonds results in polyparaphenylene. Let us see what happens when the wave number k increases (the middle and the right-hand side of Fig. 1.18). What counts now is how two *complete monomer orbitals* combine: in-phase or out-of-phase. The lowest-energy π orbitals of benzene (φ_1) arranged in-phase ($k = 0$) give point Γ, the lowest energy in the polymer, while out-of-phase, point $k = \frac{\pi}{a}$, the highest energy. At $k = \frac{\pi}{a}$ there is a degeneracy of this orbital and of φ_3 arranged out-of-phase. The degeneracy is quite interesting because, despite a superposition of the orbitals with the different number of nodes, the result, for obvious reasons, corresponds to the same number of nodes. Note the extremely small dispersion of the band which results from the arrangement of φ_2. The figure shows that it is bound to be small, because it is caused by the arrangement of two molecular orbitals that are further away in space than those considered so far (the overlap results from the overlap of the atomic

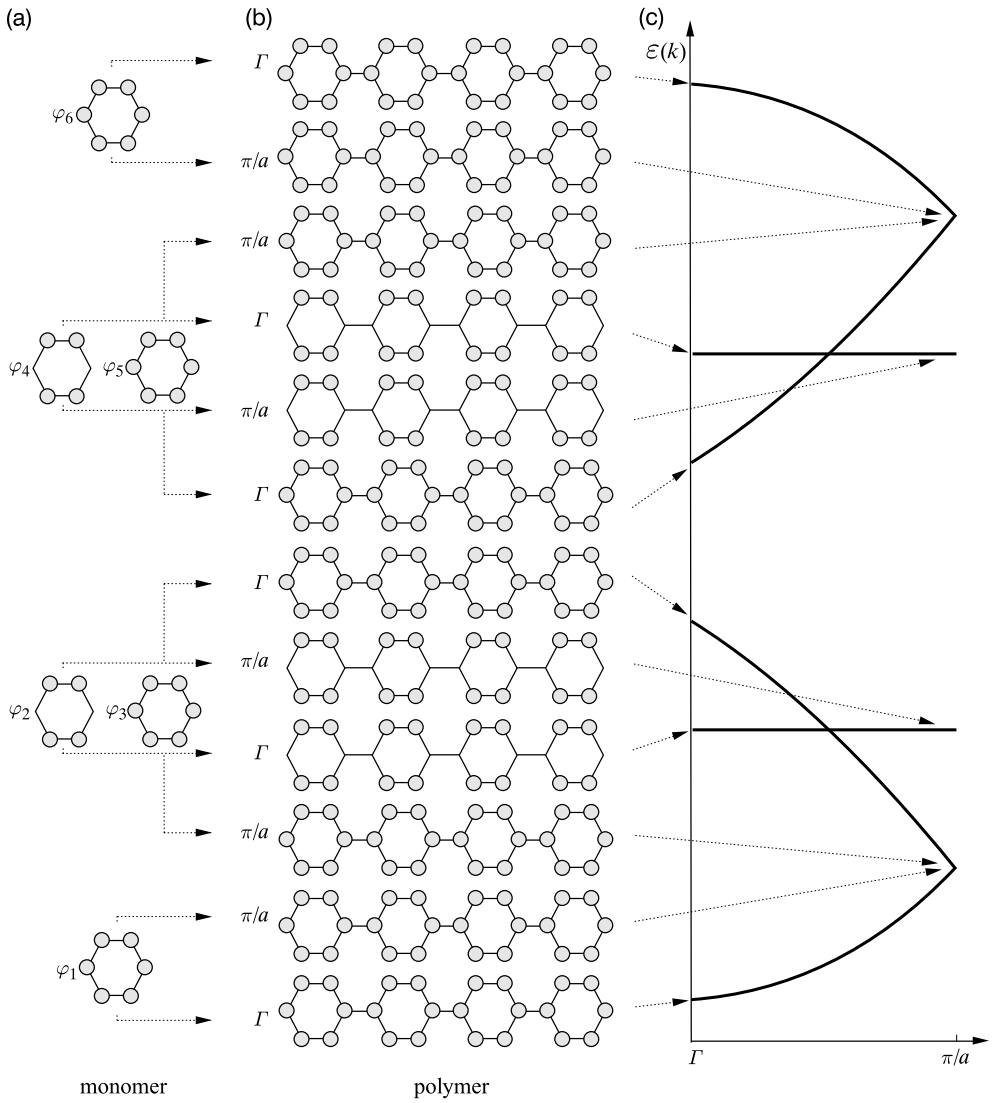


Fig. 1.18. Rationalizing the band structure of polyparaphenylene (π bands). The COs (in the center) built as in-phase or out-of-phase combinations of the benzene π molecular orbitals (left-hand side). It is seen that energy of the COs for $k = 0$ and $k = \frac{\pi}{a}$ agree with the rule of increasing number of nodes. A small band width corresponds to small overlap integrals of the monomer orbitals. J.-M. André, J. Delhalle, J.-L. Brédas, "*Quantum Chemistry Aided Design of Organic Polymers*", World Scientific, Singapore, 1991. Reprinted with permission from the World Scientific Publishing Co. Courtesy of the authors.

orbitals separated by three bonds, and not by a single bond, as it has been). We see a similar regularity in the conduction bands that correspond to the molecular orbitals φ_4, φ_5, and φ_6. The rule works here without any exception and results from the simple statement that a bonding superposition has a lower energy than the corresponding antibonding one.

Thus, when looking at the band structure for polyparaphenylene we stay cool: we understand every detail of this tangle of bands.

A stack of Pt(II) square planar complexes

Let us try to predict[43] qualitatively (without making calculations) the band structure of a stack of platinum square planar complexes, typically $[Pt(CN^-)_4^{2-}]_\infty$. Consider the parallel eclipsed configuration of all the monomeric units. Let us first simplify our task. Who likes cyanides? Let us throw them away and take something theoreticians really love: H^-. This is a little less than just laziness. If needed, we are able to make calculations for cyanides too, but to demonstrate that we really understand the machinery we are always recommended to make the system as simple as possible (but not simpler). We suspect that the main role of CN^- is just to interact electrostatically, and H^- does this too (being much smaller). In reality, it turns out that the decisive factor is the Pauli exclusion principle, rather than the ligand charge.[44]

The electronic dominant configuration of the platinum atom in its ground state is[45] $(Xe)(4f^{14})5d^96s^1$. As we can see, we have the xenon-like closed shell and also the full closed subshell $4f$. The orbital energies corresponding to these closed shells are much lower than the orbital energy of the hydrogen anion (they are to be combined to). This is why they will not participate in the Pt–H bonds. They will of course contribute to the band structure, but this contribution will be trivial: flat bands (because of small overlap integrals) with energies very close to the energies characterizing the corresponding atomic orbitals. The Pt valence shell is therefore $5d^96s^16p^0$ (for Pt^0) and $5d^86s^06p^0$ (for Pt^{2+}), the latter we have in our stack. The corresponding orbital energies are shown on the left-hand side of Fig. 1.19a.

Let us choose a Cartesian coordinate system with the origin on the platinum atom and the four ligands at equal distances on the x and y axes. In the Koopmans approximation (cf. Chapter V1-8, p. V1-544) an orbital energy represents the electron energy on a given orbital. We see that because of the ligands' pushing (Pauli exclusion principle operating), all the platinum

[43] R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*," VCH publishers, New York, 1988.

[44] When studying complexes of Fe^{2+} and Co^{2+} (of planar and tetrahedral symmetry) it turned out that splitting the d energy levels by negative point charges (simulating ligands) has been very ineffective even when making the negative charges excessively large and pushing them closer to the ion. In contrast, replacing the point charges by some closed shell entities resulted in strong splitting.

[45] Xe denotes the xenon-like configuration of electrons.

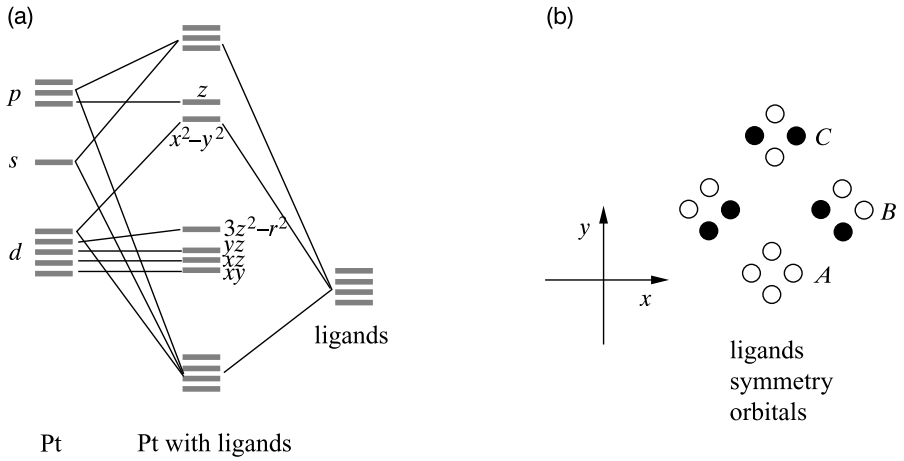


Fig. 1.19. Predicting the band structure of $\left(PtH_4^{2-}\right)_\infty$, monomer analysis. (a) Left-hand side. The Pt atomic orbitals of Pt^{2+} ion (eight valence electrons): four doubly occupied $5d$ orbitals and empty $6s$ and $6p$ orbitals. Right-hand side of (a): four ligand ($L=H^-$) atomic orbitals (symmetrized as shown in (b)). Center of (a): mixing of the orbitals when forming PtH_4^{2-}. Most important orbital interactions are: mixing of the empty $d_{x^2-y^2}$ with the ligand occupied C combination, mixing of the empty p_x and p_y orbitals with the ligand doubly-occupied two B combinations and mixing of the empty s orbital with the ligand doubly-occupied A combination (see (b)). This results in four low-energy doubly occupied ligand-like orbitals (center, bottom) and four doubly occupied platinum atomic $5d$ orbitals (center), while $5d_{x^2-y^2}$ (mostly platinum centered) is the monomer's LUMO (of high energy, because it protrudes right across to the ligands). R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

atom orbital energies will go up (destabilization; in Fig. 1.19a this shift is not shown, only a relative shift is given). The largest shift up will be undergone by the $5d_{x^2-y^2}$ orbital energy, because the orbital lobes protrude right across to the ligands. Eight electrons of Pt^{2+} will therefore occupy four other d orbitals[46] ($5d_{xy}$, $5d_{xz}$, $5d_{yz}$, $5d_{3z^2-r^2}$), while $5d_{x^2-y^2}$ will become the LUMO. The four ligand atomic orbitals practically do not overlap (long distance) and this is why in Fig. 1.19a they are depicted as a quadruply degenerate level. We organize them as the ligand symmetry orbitals, and they are shown in Fig. 1.19b: the nodeless orbital (A), two single-node orbitals (B) corresponding to the same energy, and the two-node orbital (C). The effective linear combinations (cf. p. V1-503, what counts most is symmetry) are formed by the following

[46] Of these four the lowest-energy will correspond to the orbitals $5d_{xz}$, $5d_{yz}$, because their lobes just avoid the ligands. The last two orbitals $5d_{xy}$ and $5d_{3z^2-r^2} = 5d_{z^2-x^2} + 5d_{z^2-y^2}$ will go up somewhat in the energy scale (each to a different extent), because they aim in part at the ligands. However, these splits will be smaller when compared to the fate of the orbital $5d_{x^2-y^2}$ and therefore, these levels are shown in the figure as a single degenerate level.

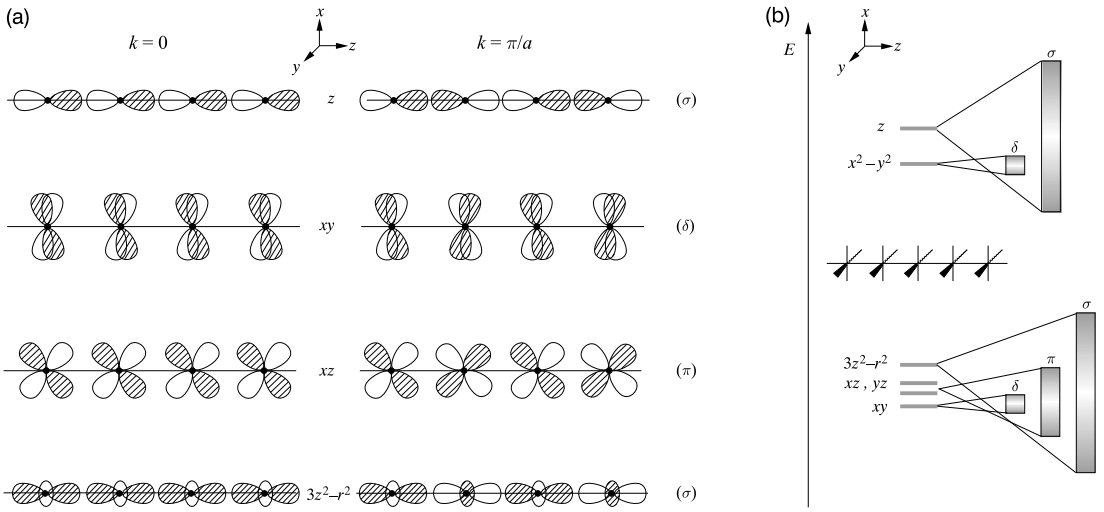


Fig. 1.20. Predicting the band structure of $\left(PtH_4^{2-}\right)_\infty$. (a) The Bloch functions for $k = 0$ and $k = \frac{\pi}{a}$ corresponding to the atomic orbitals $6p_z$ (σ-type orbitals), $5d_{xy}$ (δ-type orbitals), $5d_{xz}$ (π-type orbitals, similarly for $5d_{yz}$), and $5d_{3z^2-r^2}$ (σ-type orbitals). (b) The band width is very sensitive to the overlap of the atomic orbitals. The band widths in $\left(PtH_4^{2-}\right)_\infty$ result from the overlap of the $\left(PtH_4^{2-}\right)$ orbitals. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

pairs of orbitals: $6s$ with A, $6p_x$ and $6p_y$ with B, and the orbital $5d_{x^2-y^2}$ with C (in each case we obtain the bonding and the antibonding orbital); the other platinum orbitals, $5d$ and $6p_z$, do not have partners of the appropriate symmetry (and therefore their energy does not change). Thus we obtain the energy level diagram of the monomer in Fig. 1.19a.

Now, we form a stack of PtH_4^{2-} along the periodicity axis z. Let us form the Bloch functions (Fig. 1.20a) for each of the valence orbitals at two points of the FBZ: $k = 0$ and $k = \frac{\pi}{a}$. The results are given in Fig. 1.20b. Because of large overlap of the $6p_z$ orbitals with themselves, and $3d_{3z^2-r^2}$ also with themselves, these σ bands will have very large dispersions. The smallest dispersion will correspond to the $5d_{xy}$ band (as well as to the empty band $5d_{x^2-y^2}$), because the orbital lobes of $5d_{xy}$ (also of $5d_{x^2-y^2}$) are oriented perpendicularly to the periodicity axis. Two bands $5d_{xz}$ and $5d_{yz}$ have a common fate (i.e., the same plot) due to the symmetry, and a medium band width (Fig. 1.20b). We predict therefore the band structure shown in Fig. 1.21a. It is to be compared (Fig. 1.21b) with the calculated band structure for $(PtH_4^{2-})_\infty$. As we can see, the prediction turns out to be correct.

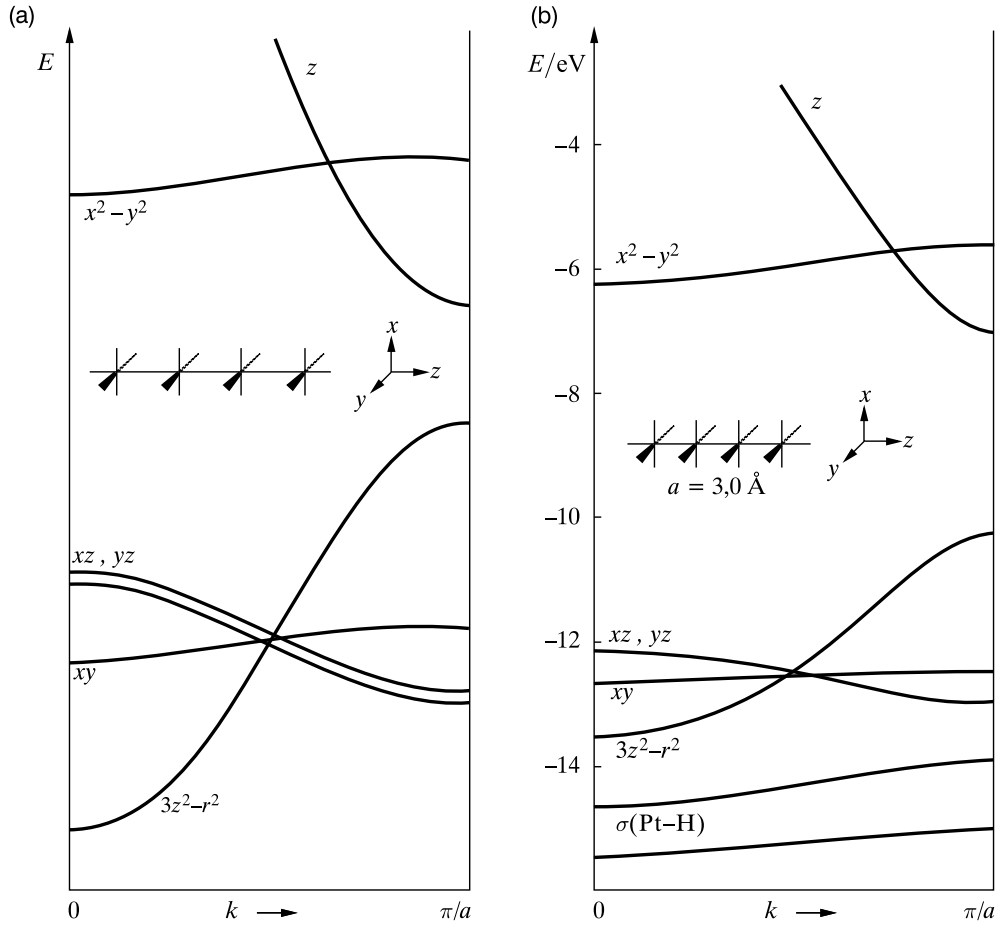


Fig. 1.21. Predicting the band structure of $\left(PtH_4^{2-}\right)_\infty$. (a) The predicted band structure. (b) The computed band structure (by Roald Hoffmann) for $a = 3$ Å. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © VCH Publishers. Reprinted with permission of John Wiley&Sons, Inc.

1.11 The Hartree–Fock method for crystals

1.11.1 Secular equation

What has been said previously about the Hartree–Fock method is only a sort of general theory. The time has now arrived to show how the method works in practice. We have to solve the Hartree–Fock–Roothaan equation (cf. Chapter V1-8, pp. V1-506 and 31).

The Fock matrix element is equal to (noting that $(\chi_p^j|\widehat{F}\chi_q^{j'}) \equiv F_{pq}^{jj'}$ depends on the *difference*[47] between the vectors $\mathbf{R}_j$ and $\mathbf{R}_{j'}$)

$$F_{pq} = (2N+1)^{-3} \sum_{jj'} \exp\left(i\mathbf{k}(\mathbf{R}_j - \mathbf{R}'_j)\right)(\chi_p^{j'}|\widehat{F}\chi_q^j) = \tag{1.55}$$

$$= \sum_j \exp(i\mathbf{k}\mathbf{R}_j) F_{pq}^{0j}. \tag{1.56}$$

The same can be done with S_{pq} and therefore the Hartree–Fock–Roothaan secular equation (see p. 31) has the form

$$\sum_{p=1}^{\omega} c_{pn}(\mathbf{k}) \Big(\sum_j \exp(i\mathbf{k}\mathbf{R}_j) \big(F_{pq}^{0j}(\mathbf{k}) - \varepsilon_n(\mathbf{k}) S_{pq}^{0j}(\mathbf{k}) \big) \Big) = 0, \tag{1.57}$$

for $q = 1, 2, \ldots \omega$.

The integral S_{pq} equals

$$S_{pq} = \sum_j \exp(i\mathbf{k}\mathbf{R}_j) S_{pq}^{0j}, \tag{1.58}$$

where the summation goes over the lattice nodes and $S_{pq}^{0j} \equiv \left(\chi_p^0|\chi_q^j\right)$. In order to be explicit, let us see what is inside the Fock matrix elements $F_{pq}^{0j}(\mathbf{k})$. *We have* to find a dependence there on the Hartree–Fock–Roothaan solutions (determined by the coefficients c_{pn}), and more precisely on the bond-order matrix.[48] Any CO, according to Eq. (1.53), has the form

$$\psi_n(\mathbf{r}, \mathbf{k}) = (2N+1)^{-\frac{3}{2}} \sum_q \sum_j c_{qn}(\mathbf{k}) \exp(i\mathbf{k}\mathbf{R}_j) \chi_q^j(\mathbf{r}), \tag{1.59}$$

where we promise to use such c_{qn} that ψ_n are normalized. For molecules the bond-order matrix element (for the atomic orbitals χ_p and χ_q) has been defined as $2\sum c_{pi} c_{qi}^*$ (the summation is over the doubly occupied orbitals), where the factor 2 results from the double occupation of the

[47] As a matter of fact, all depends on how distant the unit cells j and j' are. We have used the fact that $\widehat{F}$ exhibits crystal symmetry.

[48] We have met the same in the Hartree–Fock method for molecules, where the Coulomb and exchange operators depended on the solutions to the Fock equation (cf. p. 484).

closed shell. We have exactly the same for the crystal, where we define the bond-order matrix element corresponding to atomic orbitals χ_q^j and χ_p^l as

$$P_{pq}^{lj} = 2(2N+1)^{-3} \sum_{\text{occupied}} c_{pn}(\mathbf{k}) \exp(i\mathbf{k}\mathbf{R}_l) c_{qn}(\mathbf{k})^* \exp(-i\mathbf{k}\mathbf{R}_j), \tag{1.60}$$

where the summation goes over all the occupied COs (we assume double occupation, hence factor 2). This means that in the summation we have to go over all the occupied bands (index n), and in each band over all allowed COs, i.e., all the allowed $\mathbf{k}$ vectors in the FBZ. Thus,

$$P_{pq}^{lj} = 2(2N+1)^{-3} \sum_{n} \sum_{\mathbf{k}}^{FBZ} c_{pn}(\mathbf{k}) c_{qn}(\mathbf{k})^* \exp\left(i\mathbf{k}(\mathbf{R}_l - \mathbf{R}_j)\right). \tag{1.61}$$

This definition of the $\mathbf{P}$ matrix is exactly what we should have for a large closed shell molecule. The matrix element has to have four indices (instead of the two indices in the molecular case), because we have to describe the atomic orbitals indicating that atomic orbital p is from unit cell l, and atomic orbital q from unit cell j. It is easily seen that P_{pq}^{lj} depends on the *difference* $\mathbf{R}_l - \mathbf{R}_j$, not on $\mathbf{R}_l$ and $\mathbf{R}_j$ themselves. The reason for this is that in a crystal everything is repeated and the important thing are the *relative* distances. Thus the $\mathbf{P}$ matrix is determined by all the elements P_{pq}^{0j}.

1.11.2 Integration in the FBZ

There is a problem with $\mathbf{P}$, because it requires a summation over $\mathbf{k}$. We do not like this, because the number of the permitted vectors $\mathbf{k}$ is huge for large N (and N has to be large, because we are dealing with a crystal). We have to do something with it.

Let us try a small exercise. Imagine we have to perform a summation $\sum_{\mathbf{k}} f(\mathbf{k})$, where f represents a smooth function in the FBZ. Let us denote the sum to be found by X. Let us multiply X by a small number $\Delta = \frac{V_{FBZ}}{(2N+1)^3}$, where V_{FBZ} stands for the FBZ volume. Then we have

$$X\Delta = \sum_{\mathbf{k}}^{FBZ} f(\mathbf{k})\Delta. \tag{1.62}$$

In other words, we just cut the FBZ into tiny segments of volume Δ, their number equal to the number of the permitted $\mathbf{k}$'s. It is clear that if N is large (as in our case), then a very good approximation of $X\Delta$ would be

$$X\Delta = \int_{FBZ} f(\mathbf{k})\, d\mathbf{k}. \tag{1.63}$$

Hence,

$$X = \frac{(2N+1)^3}{V_{FBZ}} \int_{FBZ} f(\mathbf{k})\, d\mathbf{k}. \tag{1.64}$$

After applying this result to the bond-order matrix we obtain

$$P_{pq}^{lj} = \frac{2}{V_{FBZ}} \int \sum_{n}^{FBZ} c_{pn}(\mathbf{k}) c_{qn}(\mathbf{k})^* \exp\big(i\mathbf{k}(\mathbf{R}_l - \mathbf{R}_j)\big) d\mathbf{k}. \tag{1.65}$$

For a periodic *polymer* (in 1D: $V_{FBZ} = \frac{2\pi}{a}$, $\Delta = \frac{V}{2N+1}$) we would have

$$P_{pq}^{lj} = \frac{a}{\pi} \int \sum_{n} c_{pn}(k) c_{qn}(k)^* \exp\big(ika(l-j)\big) dk. \tag{1.66}$$

1.11.3 Fock matrix elements

In full analogy with the formula on p. V1-506, we can express the Fock matrix elements by using the bond-order matrix **P** for the crystal as follows:

$$F_{pq}^{0j} = T_{pq}^{0j} - \sum_{h}\sum_{u} Z_u V_{pq}^{0j}(\mathbf{A}_u^h) + \sum_{hl}\sum_{rs} P_{sr}^{lh}\big(\left(\begin{smallmatrix} 0h \\ pr \end{smallmatrix} \middle| \begin{smallmatrix} jl \\ qs \end{smallmatrix}\right) - \frac{1}{2}\left(\begin{smallmatrix} 0h \\ pr \end{smallmatrix} \middle| \begin{smallmatrix} lj \\ sq \end{smallmatrix}\right)\big), \tag{1.67}$$

where **P** satisfies the normalization condition[49]

[49] The **P** matrix satisfies the normalization condition, which we obtain in the following way. As in the molecular case the normalization of COs means

$$\begin{aligned} 1 &= \langle \psi_n(\mathbf{r},\mathbf{k}) | \psi_n(\mathbf{r},\mathbf{k}) \rangle \\ &= (2N+1)^{-3} \sum_{pq}\sum_{jl} c_{pn}(\mathbf{k})^* c_{qn}(\mathbf{k}) \exp[i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_l)] S_{pq}^{lj} \\ &= (2N+1)^{-3} \sum_{pq}\sum_{jl} c_{pn}(\mathbf{k})^* c_{qn}(\mathbf{k}) \exp[i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_l))] S_{pq}^{o(j-l)} \\ &= \sum_{pq}\sum_{j} c_{pn}(\mathbf{k})^* c_{qn}(\mathbf{k}) \exp(i\mathbf{k}\mathbf{R}_j) S_{pq}^{0j}. \end{aligned}$$

Now let us do the same for all the occupied COs and sum the results. On the left-hand side we sum just 1, and therefore we obtain the number of doubly occupied COs, i.e., $n_0(2N+1)^3$, because n_0 denotes the number of doubly occupied bands, and in each band we have in 3D $(2N+1)^3$ allowed vectors **k**. Therefore, we have

$$\begin{aligned} n_0(2N+1)^3 &= \sum_{pq}\sum_{j}\Big(\sum_{n}\sum_{\mathbf{k}}^{FBZ} c_{pn}(\mathbf{k})^* c_{qn}(\mathbf{k}) \exp(i\mathbf{k}\mathbf{R}_j)\Big) S_{pq}^{0j} \\ &= \sum_{pq}\sum_{j} \frac{1}{2}(2N+1)^3 P_{qp}^{j0} S_{pq}^{0j}, \end{aligned}$$

$$\sum_{j}\sum_{pq} P_{qp}^{j0} S_{pq}^{0j} = 2n_0, \tag{1.68}$$

where $2n_0$ means the number of electrons in the unit cell.

The first term on the right-hand side of Eq. (1.67) represents the kinetic energy matrix element

$$T_{pq}^{0j} = (\chi_p^0 | -\frac{1}{2}\Delta | \chi_q^j), \tag{1.69}$$

the second term is a sum of matrix elements, each corresponding to the nuclear attraction of an electron and the nucleus of index u and charge Z_u in the unit cell h, i.e.,

$$V_{pq}^{0j}(\mathbf{A}_u^h) = (\chi_p^0 | \frac{1}{|\mathbf{r} - \mathbf{A}_u^h|} | \chi_q^j), \tag{1.70}$$

where the upper index of χ denotes the cell number, the lower index denotes the number of the atomic orbital in a cell, the vector $\mathbf{A}_u^h$ indicates nucleus u (numbering within the unit cell) in unit cell h (from the coordinate system origin), and the third term is connected to the Coulombic operator (the first of two terms) and the exchange operator (the second of two terms). The summations over h and l go over the unit cells of the whole crystal and are therefore very difficult and time consuming.

The definition of the two-electron integral

$$\left(\begin{smallmatrix} 0h \\ pr \end{smallmatrix} \middle| \begin{smallmatrix} jl \\ qs \end{smallmatrix}\right) = \int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \chi_p^0(\mathbf{r}_1)^* \chi_r^h(\mathbf{r}_2)^* \frac{1}{r_{12}} \chi_q^j(\mathbf{r}_1) \chi_s^l(\mathbf{r}_2) \tag{1.71}$$

is in full analogy to the notation of Chapter V1-8 and Appendix V1-N (p. V1-707).

---

where from (1.61) after exchanging $p \leftrightarrow q$, $j \leftrightarrow l$ we have

$$P_{qp}^{jl} = 2(2N+1)^{-3} \sum_{n}^{} \sum_{\mathbf{k}}^{FBZ} c_{qn}(\mathbf{k}) c_{pn}(\mathbf{k})^* \exp\left(i\mathbf{k}(\mathbf{R}_j - \mathbf{R}_l)\right).$$

And then

$$P_{qp}^{j0} = 2(2N+1)^{-3} \sum_{n}^{} \sum_{\mathbf{k}}^{FBZ} c_{qn}(\mathbf{k}) c_{pn}(\mathbf{k})^* \exp\left(i\mathbf{k}\mathbf{R}_j\right).$$

Hence,

$$\sum_{pq}\sum_{j} P_{qp}^{j0} S_{pq}^{0j} = 2n_0.$$

1.11.4 Iterative procedure (SCF LCAO CO)

To solve Eq. (1.57) one uses the SCF LCAO MO technique as applied for molecules (Chapter V1-8) and now adapted for crystals. This particular method will be called SCF LCAO CO, because the linear combinations (LCs) of the symmetry atomic orbitals are used as the expansion functions for the COs in a self-consistent procedure (SCF stands for self-consistent field) described below. How does the SCF LCAO CO method work?

- First (zeroth iteration) we start from a guess[50] for **P**.
- Then we calculate the elements F_{pq}^{0j} for all atomic orbitals p, q for unit cells $j = 0, 1, 2, ...$ $j_{\max}$. What is $j_{\max}$? The answer is certainly nonsatisfactory: $j_{\max} = \infty$. In practice, however, we often take $j_{\max}$ as being of the order of a few cells,[51] most often we take $j_{\max} = 1$.
- For each **k** from the FBZ we calculate the elements F_{pq} and S_{pq} of Eqs. (1.56) and (1.58), and then solve the secular equations within the Hartree–Fock–Roothaan procedure. This step requires diagonalization[52] (see Appendix V1-L, p. V1-703). As a result, for each **k** we obtain a set of coefficients c for the COs and the energy eigenvalue $\varepsilon_n(\mathbf{k})$.
- We repeat all this for the values of **k** covering in some optimal way (some recipes exist) the FBZ. We are then all set to carry out the numerical integration in the FBZ and we calculate an approximate matrix **P**.
- This enables us to calculate a new approximation to the matrix **F**, and so on, until the procedure converges in a self-consistent way, i.e., produces **P** very close to that matrix **P** which has been inserted into the Fock matrix **F**. In this way we obtain the band structure $\varepsilon_n(\mathbf{k})$ and all the corresponding COs.

1.11.5 Total energy

How do we calculate the total energy for an infinite crystal? We know the answer without any calculation: $-\infty$. Indeed, since the energy represents an extensive quantity, for an infinite number of unit cells we get $-\infty$, because a single cell usually represents a bound state (negative energy). Therefore, the question has to be posed in another way.

How to calculate the total energy *per unit cell*? Aha, this is a different story. Let us denote this quantity by E_T. Since a crystal only represents a *very* large molecule, we may use the

[50] The result is presumed to be independent of this choice.

[51] The "nearest neighbor approximation." We encounter a similar problem *inside* the F_{pq}^{0j}, because we somehow have to truncate the summations over h and l. These problems will be discussed later in this chapter.

[52] Unlike the molecular case, this time the matrix to diagonalize is Hermitian, and not necessarily symmetric. Methods of diagonalization exist for such matrices, and there is a guarantee that their eigenvalues are real.

expression for the total energy of a molecule (p. V1-506). In the three-dimensional case we have

$$(2N+1)^3 E_T = \frac{1}{2}\sum_{pq}\sum_{lj} P_{qp}^{jl}\left(h_{pq}^{lj} + F_{pq}^{lj}\right) + \frac{1}{2}\sum_{lj}{}'\sum_{uv}\frac{Z_u Z_v}{R_{uv}^{lj}}, \tag{1.72}$$

where the summation over p and q extends over the ω atomic orbitals that any unit cell offers, and l and j tell us in which cells these orbitals are located. The last term on the right-hand side refers to the nuclear repulsion of all the nuclei in the crystal, u, v number the nuclei in a unit cell, while l, j indicate the cells (a prime means that there is no contribution from the charge interaction with itself). Since the summations over l and j extend over the whole crystal, we have

$$(2N+1)^3 E_T = \frac{1}{2}(2N+1)^3\sum_{pq}\sum_{j} P_{qp}^{j0}[h_{pq}^{0j} + F_{pq}^{0j}] + (2N+1)^3\frac{1}{2}\sum_{j}{}'\sum_{uv}\frac{Z_u Z_v}{R_{uv}^{0j}}, \tag{1.73}$$

because each term has an equal contribution, and the number of such terms is equal to $(2N+1)^3$.

Therefore, the total energy per unit cell amounts to

$$E_T = \frac{1}{2}\sum_{j}\sum_{pq} P_{qp}^{j0}\left(h_{pq}^{0j} + F_{pq}^{0j}\right) + \frac{1}{2}\sum_{j}\sum_{u}{}'\sum_{v}\frac{Z_u Z_v}{R_{uv}^{0j}}. \tag{1.74}$$

The formula is correct, but we can easily see that we are to be confronted with some serious problems. For example, the summation over nuclei represents a divergent series and we will get $+\infty$. *This problem appears only because we are dealing with an infinite system*. We have to manage the problem somehow.

1.12 Long-range interaction problem

What is left to be clarified are some problems about how to go with N to infinity.[53] It will be soon shown how dangerous this problem is.

[53] Let me tell you about my adventure with this problem, because I remember how as a student I wanted to hear about struggles with understanding matter and ideas instead of some dry summaries.
The story began quite accidentally. In 1977, at the University of Namur (Belgium) Professor Joseph Delhalle asked the PhD student Christian Demanet to perform a numerical test. The test consisted of taking a simple infinite polymer (the infinite chain $\cdots$LiH–LiH–LiH$\cdots$ had been chosen), to use the simplest atomic basis set possible and to see what we should take as N, to obtain the Fock matrix with sufficient accuracy. Demanet first

We see from Eqs. (1.67) and (1.74) that thanks to the translational symmetry, we may treat each **k** separately, but infinity continues to make us a little nervous. In the expression for F_{pq}^{0j} we have a summation (over the whole infinite crystal) of the interactions of an electron with all the nuclei, and in the next term we have a summation over the whole crystal of the electron–electron interactions. This is of course natural, because our system is infinite. The problem is, however, that both summations diverge: the first tends to $-\infty$, the second to $+\infty$. On top of this, to compute the bond-order matrix **P** we have to perform another summation in Eq. (1.65) over the FBZ of the crystal. We have a similar, very unpleasant, situation in the total energy expression, where the first term tends to $-\infty$, while the nuclear repulsion term goes $+\infty$.

The routine approach in the literature was to replace the infinity by taking the first neighbor interactions. This approach is quite understandable, because any attempt to take further neighbors ends up with an exorbitant bill to pay.[54]

1.12.1 Fock matrix corrections

A first idea we may think of is to separate carefully the long-range part of the Fock matrix elements and of the total energy from these quantities as calculated in a traditional way, i.e., by limiting the infinite-range interactions to those for the N neighbors on the left and N neighbors on the right of cell 0. For the Fock matrix element we would have

$$F_{pq}^{0j} = F_{pq}^{0j}(N) + C_{pq}^{0j}(N), \tag{1.75}$$

took $N = 1$, then $N = 2$, $N = 3$ – the Fock matrix changed all the time. He got impatient, took $N = 10$, $N = 15$ – the matrix continued to change. Only when he used $N = 200$ did the Fock matrix elements stabilize within the accuracy of six significant figures. We could take $N = 200$ for an extremely poor basis set and for a few such tests, but never in good quality calculations as their cost would become astronomic. Even for the case in question the computations had to be done overnight. In a casual discussion at the beginning of my six-week stay at the University of Namur, Joseph Delhalle told me about the problem. He also said that in a recent paper the Austrian scientists Alfred Karpfen and Peter Schuster noted that the results depend strongly on the chosen value of N. They made a correction *after* the calculations with a small N had been performed. They added the dipole–dipole electrostatic interaction of cell 0 with a few hundred neighboring cells, and as the dipole moment of a cell they took the dipole moment of the isolated LiH molecule. As a result the Fock matrix elements changed much less with N. This information made me think about implementing the multipole expansion right from the beginning of the self-consistent Hartree–Fock–Roothaan procedure for a polymer. Below you will see what has been done. The presented theory pertains to a regular polymer (a generalization to two and three dimensions is possible).

[54] The number of two-electron integrals, which quantum chemists positively dislike, increases with the number of neighbors to take (N) and the atomic basis set size per unit cell (ω) as $N^3\omega^4$. Besides, the nearest neighbors *are* indeed the most important.

where $C_{pq}^{0j}(N)$ stands for the long-range correction, while for $F_{pq}^{0j}(N)$ we assume interactions with the N right and N left neighbors of cell 0. They are calculated as follows:

$$F_{pq}^{0j}(N) = T_{pq}^{0j} + \sum_{h=-N}^{h=+N} \left(-\sum_u Z_u V_{pq}^{0j}(\mathbf{A}_u^h) + \sum_{l=h-N}^{l=h+N} \sum_{rs} P_{sr}^{lh} \left(\left({}^{0h}_{pr} | {}^{jl}_{qs} \right) - \frac{1}{2} \left({}^{0h}_{pr} | {}^{lj}_{sq} \right) \right) \right), \quad (1.76)$$

$$C_{pq}^{0j}(N) = \sum_h^{\#} \left(-\sum_u Z_u V_{pq}^{0j}(\mathbf{A}_u^h) + \sum_{l=h-N}^{l=h+N} \sum_{rs} P_{sr}^{lh} \left({}^{0h}_{pr} | {}^{jl}_{qs} \right) \right), \quad (1.77)$$

where the symbol $\sum_h^{\#}$ means a summation over all the unit cells *except* the section of unit cells with numbers $-N, -N+1, \ldots, 0, 1, \ldots N$, i.e., the neighborhood of cell 0 ("short-range"). The nuclear attraction integral[55]

$$V_{pq}^{0j}(\mathbf{A}_u^h) = (\chi_p^0 | \frac{1}{|\mathbf{r} - (\mathbf{A}_u + ha\mathbf{z})|} | \chi_q^j), \quad (1.78)$$

where the vector $\mathbf{A}_u$ shows the position of the nucleus u in cell 0, while $\mathbf{A}_u^h \equiv \mathbf{A}_u + ha\mathbf{z}$ points to the position of the equivalent nucleus in cell h ($\mathbf{z}$ denotes the unit vector along the periodicity axis).

The expression for $C_{pq}^{0j}(N)$ has a clear physical interpretation. The first term represents interaction of the charge distribution $-\chi_p^0(1)^* \chi_q^j(1)$ (of electron 1, hence the minus sign) with *all nuclei*[56] except those enclosed in the short-range region (i.e., extending from $-N$ to $+N$). The second term describes the interaction of the same electronic charge distribution with the *total electronic distribution* outside the short-range region. How do we see this? The integral $\left({}^{0h}_{pr} | {}^{jl}_{qs} \right)$ means the Coulombic interaction of the distribution under consideration $-\chi_p^0(1)^* \chi_q^j(1)$ with its partner distribution $-\chi_r^h(2)^* \chi_s^l(2)$, does it not? This distribution is multiplied by P_{sr}^{lh} and then summed over all possible atomic orbitals r and s in cell h and its neighborhood (the sum over cells l from the neighborhood of cell h), which gives the total partner electronic distribution $-\sum_{l=h-N}^{l=h+N} \sum_{rs} P_{sr}^{lh} \chi_r^h(2)^* \chi_s^l(2)$. This, however, simply represents the electronic charge distribution of cell h. Indeed, the distribution, when integrated, gives (just look at Eq. (1.68)) $-\sum_{l=h-N}^{l=h+N} \sum_{rs} P_{sr}^{lh} S_{rs}^{hl} = 2n_0$. Therefore, our electron distribution $-\chi_p^0(1)^* \chi_q^j(1)$ interacts electrostatically with the charge distribution of all cells except those enclosed in the short-range region, because Eq. (1.77) contains the summation over all cells h except the short-range region. Finally,

55 Without the minus sign in the definition the name is not quite adequate.

56 See the interpretation of the integral $-V_{pq}^{0j}(\mathbf{A}_u^h) = -(\chi_p^0(\mathbf{r}) | \frac{1}{|\mathbf{r} - \mathbf{A}_u^h|} | \chi_q^j(\mathbf{r}))$.

the long-range correction to the Fock matrix elements $C_{pq}^{0j}(N)$ represents the Coulombic interaction of the charge distribution $-\chi_p^0(1)^*\chi_q^j(1)$ with all the unit cells (nuclei and electrons) from outside the short-range region.

In the $C_{pq}^{0j}(N)$ correction, in the summation over l, *we have neglected the exchange term* $-\frac{1}{2}\sum_h^{\#}\sum_{l=h-N}^{l=h+N}\sum_{rs}P_{sr}^{lh}\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}|\begin{smallmatrix}lj\\sq\end{smallmatrix}\right)$. The reason for this was that we have been *convinced* that P_{sr}^{0h} vanishes fast with h. Indeed, the largest integral in the summation over l is $-\frac{1}{2}\sum_h^{\#}\sum_{rs}P_{sr}^{0h}\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}|\begin{smallmatrix}0j\\sq\end{smallmatrix}\right)$. This term is supposed to be small not because of the integral $\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}|\begin{smallmatrix}0j\\sq\end{smallmatrix}\right)$, which can be quite important (like, e.g., $\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}|\begin{smallmatrix}0,h-1\\pq\end{smallmatrix}\right)$), but because of P_{sr}^{0h}. We will come back to this problem.[57]

1.12.2 Total energy corrections

The total energy per unit cell could similarly be written as

$$E_T = E_T(N) + C_T(N), \tag{1.79}$$

where $E_T(N)$ means the total energy per unit cell as calculated by the traditional approach, i.e., with truncation of the infinite series on the N left and N right neighbors of cell 0. The quantity $C_T(N)$ therefore represents the error, i.e., the long-range correction. The detailed formulae for $E_T(N)$ and $C_T(N)$ are the following:

$$E_T(N) = \frac{1}{2}\sum_{j=-N}^{j=+N}\sum_{pq}P_{qp}^{j0}\left(h_{pq}^{0j} + F_{pq}^{0j}(N)\right) + \frac{1}{2}\sum_{j=-N}^{j=+N}\sum_{u}\sum_{v}{}'\frac{Z_uZ_v}{R_{uv}^{0j}}, \tag{1.80}$$

$$C_T(N) = \frac{1}{2}\sum_j\sum_{pq}P_{qp}^{j0}C_{pq}^{0j}(N) + \frac{1}{2}\sum_h^{\#}\left(\sum_j\sum_{pq}P_{qp}^{j0}\sum_u\left[-Z_uV_{pq}^{0j}(\mathbf{A}_u^h)\right] + \sum_u\sum_v{}'\frac{Z_uZ_v}{R_{uv}^{0h}}\right), \tag{1.81}$$

where from F_{pq}^{0j} we have already separated its long-range contribution $C_{pq}^{0j}(N)$, so that $C_T(N)$ contains *all* the long-range corrections.

[57] Matrix element P_{sr}^{0h}, i.e., the bond-order contribution from the atomic orbital product $\chi_s^0(1)\chi_r^h(1)^*$ pertaining to distant cells 0 and h, seems to be a small number. This will turn out to be delusive. We have to stress, however, that trouble will come only in some "pathological" situations.

Eq. (1.81) for $C_T(N)$ may be obtained *just by looking* at Eq. (1.80). The first term with $C_{pq}^{0j}(N)$ is evident[58]; it represents the Coulombic interaction of the *electronic distribution* (let us recall condition (1.68)) associated with cell 0 with *the whole polymer chain* except the short-range region. What, therefore, is yet to be added to $E_T(N)$? What it lacks is the Coulombic interaction of the *nuclei* of cell 0 with *the whole polymer chain*, except the short-range region. Let us see whether we have it in Eq. (1.81). The last term means the Coulombic interaction of the nuclei of cell 0 with *all the nuclei of the polymer* except the short-range region (and again we know why we have the factor $\frac{1}{2}$). What, therefore, is represented by the middle term[59]? It is clear that it has to be (with the factor $\frac{1}{2}$) the Coulombic interaction of the *nuclei of cell* 0 with the total electronic distribution outside the short-range region. We look at the middle term. We have the minus sign, and that is very good indeed, because we have to have an attraction. Further, we have the factor $\frac{1}{2}$, and that is also OK; then we have $\sum_h^{\#}$, and that is perfect, because we expect a summation over the long range only, and finally we have $\sum_j \sum_{pq} P_{qp}^{j0} \sum_u \left[-Z_u V_{pq}^{0j}(\mathbf{A}_u^h) \right]$, and we do not like this. This is the Coulombic interaction of the total *electronic distribution of cell* 0 with the *nuclei of the long-range region*, while we expected the interaction of the *nuclei*

[58] The factor $\frac{1}{2}$ may worry us a little. Why just $\frac{1}{2}$? Let us see. Imagine N *identical objects*, $i = 0, 1, 2, \ldots N-1$, *playing identical roles in a system* (like our unit cells). We will be interested in the energy per object, E_T. The total energy may be written as (let us assume here pairwise interactions only)

$$NE_T = \sum_j E_j + \sum_{i<j} E_{ij},$$

where E_j and E_{ij} are the isolated object energy and the pairwise interaction energy, respectively. Since the objects are identical, we have

$$NE_T = NE_0 + \frac{1}{2}\sum_{i,j}{}' E_{ij} = NE_0 + \frac{1}{2}\sum_i \left(\sum_j{}' E_{ij} \right) = NE_0 + \frac{1}{2} N \left(\sum_j{}' E_{0j} \right),$$

where the prime means excluding self-interaction and the term in parentheses means the interaction of object 0 with all others. Finally,

$$E_T = E_0 + \frac{1}{2} \left(\sum_j{}' E_{0j} \right),$$

where we have the factor $\frac{1}{2}$ before the interaction of one of the objects with the rest of the system.

[59] As we can see, we have to sum (over j) to infinity the expressions h_{pq}^{0j}, which contain T_{pq}^{0j} (but these terms decay very fast with j and can all be taken into account in $E_T(N)$) and the long-range terms, the Coulombic interaction of the electronic charge distribution of cell 0 with the nuclei beyond the short-range region (the middle term in $C_T(N)$). The argument about fast decay with j of the kinetic energy matrix elements mentioned before follows from the double differentiation with respect to the coordinates of the electron. Indeed, this results in another atomic orbital, but with the same center. This leads to the overlap integral of the atomic orbitals centered like those in $\chi_p^0 \chi_q^j$. Such an integral decays exponentially with j.

of cell 0 with the *electronic charge distribution of the long-range region*. What is going on? Everything is OK. Just count the interactions pairwise and at each of them reverse the locations of the interacting objects – the two interactions mean the same. Therefore,

> the long-range correction to the total energy per cell $C_T(N)$ represents the Coulombic interaction of cell 0 with all the cells from outside the short-range region.

We are now all set to calculate the long-range corrections $C_{pq}^{0j}(N)$ and $C_T(N)$. It is important to realize that all the interactions to calculate pertain to objects that are *far away in space*.[60] This is what we have carefully prepared. This is the condition that enables us to apply the multipole expansion to each of the interactions (Appendix G).

1.12.3 Multipole expansion applied to the Fock matrix

Let us first concentrate on $C_{pq}^{0j}(N)$. As seen from Eq. (1.77) there are two types of interactions to calculate: the nuclear attraction integrals $V_{pq}^{0j}(\mathbf{A}_u^h)$ and the electron repulsion integrals $(^{0h}_{pr}|^{jl}_{qs})$. In the second term, we may use the multipole expansion of $\frac{1}{r_{12}}$ given in Appendix G (p. 615). In the first term, we will do the same, but this time one of the interacting particles will be the nucleus indicated by vector $\mathbf{A}_u^h$. The corresponding multipole expansion (in a.u.; the nucleus u of charge Z_u interacts with the electron of charge -1, $n_k = n_l = \infty$, $S = min(k,l)$) reads

$$-\frac{Z_u}{r_{u1}} = \sum_{k=0}^{n_k}\sum_{l=0}^{n_l}\sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} M_a^{(k,m)}(1)^* M_b^{(l,m)}(u), \tag{1.82}$$

---

60 Let us check this. What objects are we talking about? Let us begin from $C_{pq}^{0j}(N)$. As seen from the formula, one of the interacting objects is the charge distribution of the first electron $\chi_p^0(1)^*\chi_q^j(1)$. The second object is the whole polymer except the nuclei and electrons of the neighborhood of cell 0. The charge distributions $\chi_p^0(1)^*\chi_q^j(1)$ with various j are always close to cell 0, because the orbital $\chi_p^0(1)$ is anchored at cell 0, and such a distribution decays exponentially when cell j goes away from cell 0. The fact that the nuclei with which the distribution $\chi_p^0(1)^*\chi_q^j(1)$ interacts are far apart is evident, but less evident is the fact that the electrons with which the distribution interacts are also far away from cell 0. Let us have a closer look at the electron–electron interaction. The charge distribution of electron 2 is $\chi_r^h(2)^*\chi_s^l(2)$, and the summation over cells h excludes the neighborhood of cell 0. Hence, because of the exponential decay there is a guarantee that the distribution $\chi_r^h(2)^*\chi_s^l(2)$ is bound to be close to cell h, if this distribution is to be of any significance. Therefore, the charge distribution $\chi_r^h(2)^*\chi_s^l(2)$ is certainly far away from cell 0.

Similar reasoning may be used for $C_T(N)$. The interacting objects are of the type $\chi_p^0(1)^*\chi_q^j(1)$, i.e., always close to cell 0, with the nuclei of cell h, and there is a guarantee that h is far away from cell 0. The long distance of the interacting nuclei (second term) is evident.

where R stands for the distance between the origins of the coordinate system centered in cell 0 and the coordinate system in cell h, which, of course, is equal to $R = ha$. The multipole moment operator of electron 1, $M_a^{(k,m)}(1)$, reads as

$$M_a^{(k,m)}(1) = -r_a^k P_k^{|m|}(\cos\theta_{a1}) \exp(im\phi_{a1}), \tag{1.83}$$

while

$$M_b^{(l,m)}(u) = Z_u r_u^l P_l^{|m|}(\cos\theta_u) \exp(im\phi_u) \tag{1.84}$$

denotes the multipole moment operator of nucleus u computed in the coordinate system of cell h. When this expansion and the expansion for $\frac{1}{r_{12}}$ are inserted into Eq. (1.77) for $C_{pq}^{0j}(N)$, we obtain

$$\begin{aligned} C_{pq}^{0j}(N) = &\sum_h^{\#} \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} \Big((\chi_p^0 | \hat{M}_a^{(k,m)}(1)^* | \chi_q^j) \cdot \\ &\Big[\sum_u M_b^{(l,m)}(\mathbf{A}_u^h) \Big] + (\chi_p^0 | \hat{M}_a^{(k,m)}(1)^* | \chi_q^j) \cdot \\ &\sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} \big(\chi_r^h | \hat{M}_b^{(l,m)}(2) | \chi_s^{l'} \big) \Big) \\ = &\sum_h^{\#} \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} \big(\chi_p^0 | \hat{M}_a^{(k,m)}(1)^* | \chi_q^j \big) \cdot \\ &\Big[\sum_u M_b^{(l,m)}(\mathbf{A}_u^h) + \sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} \big(\chi_r^h | \hat{M}_b^{(l,m)}(2) | \chi_s^{l'} \big) \Big]. \end{aligned}$$

Let us note that in the square brackets we have nothing but a multipole moment of *unit cell* h. Indeed, the first term represents the multipole moment of all the nuclei of cell h, while the second term is the multipole moment of electrons of unit cell h. The latter can be best seen if we recall the normalization condition (1.68), i.e., $\sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} S_{rs}^{hl'} = \sum_{l'=-N}^{l'=+N} \sum_{rs} P_{sr}^{l'0} S_{rs}^{0l'} = 2n_0$, with $2n_0$ denoting the number of electrons per cell. Hence, we can write

$$C_{pq}^{0j}(N) = \sum_h^{\#} \sum_{k=0} \sum_{l=0} \sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} \big(\chi_p^0 | \hat{M}_a^{(k,m)}(1)^* | \chi_q^j \big) M^{(l,m)}(h), \tag{1.85}$$

where the dipole moment of cell h is given by

$$M^{(l,m)}(h) = \Big[\sum_u M_b^{(l,m)}(\mathbf{A}_u^h) + \sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} \big(\chi_r^h | \hat{M}_b^{(l,m)}(2) | \chi_s^{l'} \big) \Big], \tag{1.86}$$

because the summation over u goes over the nuclei belonging to cell h, and the coordinate system b is anchored in cell h. Now it is time to say something most important.

> Despite the fact that $M^{(l,m)}(h)$ depends formally on h, in reality it is h-independent, because all the unit cells are identical.

Therefore, we may safely write that $M^{(l,m)}(h) = M^{(l,m)}$.

Now we will avoid a well-hidden trap, and then we will be all set to prepare ourselves to pick the fruit from our orchard. The trap is that $A_{kl|m|}$ depends on h. How is this? Well, in the $A_{kl|m|}$ there is $(-1)^l$, while the corresponding $(-1)^k$ is absent, i.e., there is a thing that is associated with the 2^l-pole in coordinate system b, and there is no analogous expression for its partner, the 2^k-pole of coordinate system a. Remember, however (Appendix G), that the z axes of both coordinate systems have been chosen in such a way that a "shoots" towards b, and b *does not* shoot towards a. Therefore, the two coordinate systems are not equivalent, and hence one may have $(-1)^l$, and not $(-1)^k$. Coordinate system a is associated with cell 0, the coordinate system b is connected to cell h. If $h > 0$, then it is true that a shoots to b, but if $h < 0$ their roles are exchanged. In such a case, in $A_{kl|m|}$ we should not put $(-1)^l$, but $(-1)^k$. If we do this then in the summation over h in Eq. (1.85) the only dependence on h appears in a simple term $(ha)^{-(k+l+1)}$!

> It appears, therefore, to be a possibility of exactly summing the electrostatic interaction along an infinite polymer chain.

Indeed, the sum

$$\sum_{h=1}^{\infty} h^{-(k+l+1)} = \zeta(k+l+1), \tag{1.87}$$

where $\zeta(n)$ stands for the Riemann zeta function, which is known to a high accuracy and available in mathematical tables.[61]

[61] For example, M. Abramovitz, I. Stegun, eds, "*Handbook of Mathematical Functions*," Dover, New York, 1968, p. 811.

Georg Friedrich Bernhard Riemann (1826–1866), German mathematician and physicist, professor at the University of Göttingen. Nearly all his papers gave rise to a new mathematical theory. His life was full of personal tragedies, and he lived only 40 years, but nevertheless he made a giant contribution to mathematics, mainly in non-Euclidean geometries (his geometry plays an important role in the general theory of relativity), in the theory of integrals (Riemann integral), and in the theory of trigonometric series.

The interactions of cell 0 with all the other cells are enclosed in this number. When this is inserted into $C_{pq}^{0j}(N)$, we obtain

$$C_{pq}^{0j}(N) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} U_{pq}^{0j(k,l)} \frac{\Delta_N^{(k+l+1)}}{a^{(k+l+1)}}, \tag{1.88}$$

where

$$U_{pq}^{0j(k,l)} = \sum_{m=-S}^{m=+S} (-1)^m [(-1)^k + (-1)^l] \frac{(k+l)!}{(k+|m|)!(l+|m|)!} M_{pq}^{0j(k,m)*} M^{(l,m)}, \tag{1.89}$$

$$\Delta_N^{(n)} = \zeta(n) - \sum_{h=1}^{N} h^{-n}. \tag{1.90}$$

Note that the formula for $C_{pq}^{0j}(N)$ represents a sum of the multipole–multipole interactions. The formula also shows that

> electrostatic interactions in a regular polymer come from a multipole–multipole interaction with different parity of the multipoles,

which can be seen from the term[62] $[(-1)^k + (-1)^l]$.

62 The term appears due to the problem discussed above of "who shoots to whom" in the multipole expansion. What happens is that the interaction of an even (odd) multipole of cell 0 with an odd (even) multipole on the right-hand side of the polymer cancels out with a similar interaction with the left-hand side. It is easy to understand. Let us imagine the multipoles as nonpoint-like objects built of the appropriate point charges. We look along the

According to the discussion in Appendix G, to preserve the invariance of the energy with respect to translation of the coordinate system, when computing $C_{pq}^{0j}(N)$ we have to add all the terms with $k+l+1=const$, i.e.,

$$C_{pq}^{0j}(N) = \sum_{n=3,5,..}^{\infty} \frac{\Delta_N^{(n)}}{a^n} \sum_{l=1}^{n-1} U_{pq}^{0j(n-l-1,l)}. \tag{1.91}$$

The above expression is equivalent to Eq. (1.88), but ensures automatically the translational invariance by taking into account all the necessary multipole–multipole interactions.[63]

What should we know, therefore, to compute the long-range correction $C_{pq}^{0j}(N)$ to the Fock matrix[64]? From Eq. (1.91) it is seen that one has to know how to calculate three numbers: $U_{pq}^{0j(k,l)}$, a^{-n}, and $\Delta_N^{(n)}$. The equation for the first one is given in Table 1.2, the other two are trivial, Δ is easy to calculate knowing the Riemann zeta function (from tables): in fact we have to calculate the multipole moments, and these are *one-electron* integrals (easy to calculate). Originally, before the multipole expansion method was designed we also had a large number of *two-electron* integrals (expensive to calculate). Instead of overnight calculations, the computer time was reduced to about 1 second and the results were more accurate.

1.12.4 Multipole expansion applied to the total energy

As shown above, the long-range correction to the total energy means the interaction of cell 0 with all the cells from the long-range region multiplied by $\frac{1}{2}$. The reasoning pertaining to its computation may be repeated exactly in the way we have shown in the previous subsection. We have, however, to remember a few differences:

- what interacts is not the charge distribution $\chi_p^{0*}\chi_q^j$, but the complete cell 0;
- the result has to be multiplied by $\frac{1}{2}$ for reasons discussed earlier.

periodicity axis. An even multipole has the same signs at both ends, an odd one has the opposite signs. Thus, when the even multipole is located in cell 0, and the odd one on its right-hand side, this interaction will cancel out exactly with the interaction of the odd one located on the left-hand side (at the same distance).

63 Indeed, $\sum_{l=1}^{n-1} U_{pq}^{0j(n-l-1,l)} = U_{pq}^{0j(n-2,1)} + U_{pq}^{0j(n-3,2)} + \ldots + U_{pq}^{0j(0,n-1)}$, i.e., a review of all terms with $k+l+1=n$ except $U_{pq}^{0j(n-1,0)}$. This term is absent, because it requires calculation of $M^{(0,0)}$, i.e., of the *charge of the elementary cell*, which has to stay electrically neutral (otherwise the polymer falls apart), and therefore $M^{(0,0)}=0$. Why, however, does the summation over n not simply represent $n=1,2,...\infty$, but contains only odd n's except $n=1$? What would happen if we took $n=1$? Look at Eq. (1.88). The value $n=1$ requires $k=l=0$. This leads to the "monopole–monopole" interaction, but this is 0, since the whole unit cell (and one of the multipoles is that of the unit cell) carries no charge. The summation in (1.91) does not contain any even n's, because they would correspond to k and l of different parity, and such interactions (as we have shown before) are equal to 0. Therefore, indeed, (1.91) contains all the terms that are necessary.

64 L. Piela, J. Delhalle, *Intern. J. Quantum Chem.*, 13(1978) 605.

Table 1.2. The quantities $U^{(k,l)}$ for $k+l<7$ necessary for calculating the most important long-range corrections to the Fock matrix elements $U_{pq}^{0j\,(k,l)}$ and to the total energy per unit cell $U_T^{(k,l)}$. The brackets [ ] mean the corresponding Cartesian multipole moment. When computing the Fock matrix correction, the first multipole moment [ ] stands for the multipole moment of the charge distribution $\chi_p^0\chi_q^j$, the second to that of the unit cell. For example, $U^{(0,2)}$ for the correction $C_{pq}^{0j}(N)$ is equal to $({}_p^0|{}_q^j)\big(\sum_u Z_u(3z_u^2-r_u^2)-\sum_{l'=-N}^{l'=+N}\sum_{rs}P_{sr}^{l'0}(\chi_r^0|3z^2-r^2|\chi_s^{l'})\big)$, while $U^{(0,2)}$ for $C_T(N)$ is equal to 0, because [1] means the *charge* of the unit cell, which is equal to zero. In the table only U's for $k\le l$ are given. If $l<k$, then the formula is the same, but the order of the moments is reversed.

n	$U^{(k,l)}, k+l+1=n$
3	$U^{(0,2)}=[1][3z^2-r^2]$
	$U^{(1,1)}=2[x][x]+2[y][y]-4[z][z]$
5	$U^{(0,4)}=\frac{1}{4}[1][35z^4-30z^2r^2+3r^4]$
	$U^{(1,3)}=4[z][3r^2z-5z^3]+3[x][5xz^2-r^2x]+3[y][5yz^2-r^2y]$
	$U^{(2,2)}=3[3z^2-r^2][3z^2-r^2]-24[xz][xz]-24[yz][yz]+\frac{3}{2}[x^2-y^2][x^2-y^2]+6[xy][xy]$
7	$U^{(0,6)}=\frac{1}{8}[1][231z^6-315z^4r^2+105z^2r^4-5r^6]$
	$U^{(1,5)}=-\frac{3}{2}[z][63z^5-70z^3r^2+15zr^4]$
	$\quad+\frac{15}{4}[x][21z^4x-14z^2xr^2+xr^4]+\frac{15}{4}[y][21z^4y-14z^2yr^2+yr^4]$
	$U^{(2,4)}=\frac{15}{8}[3z^2-r^2][35z^4-30z^2r^2+3r^4]-30[xz][7z^3x-3xzr^2]-30[yz][7z^3y-3yzr^2]$
	$\quad+\frac{15}{4}[x^2-y^2][7z^2(x^2-y^2)-r^2(x^2-y^2)]+15[xy][7z^2xy-xyr^2]$
	$U^{(3,3)}=-10[5z^3-3zr^2][5z^3-3zr^2]+\frac{45}{4}[5z^2x-xr^2][5z^2x-xr^2]+\frac{45}{4}[5z^2y-yr^2][5z^2y-yr^2]$
	$\quad-45[zx^2-zy^2][zx^2-zy^2]-180[xyz][xyz]+\frac{5}{4}[x^3-3xy^2][x^3-3xy^2]$
	$\quad+\frac{5}{4}[y^3-3x^2y][y^3-3x^2y]$

Finally we obtain

$$C_T(N)=\frac{1}{2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}U_T^{(k,l)}\frac{\Delta_N^{(k+l+1)}}{a^{k+l+1}}, \tag{1.92}$$

where

$$U_T^{(k,l)}=\sum_{m=-S}^{m=+S}\big((-1)^k+(-1)^l\big)\frac{(k+l)!(-1)^m}{(k+|m|)!(l+|m|)!}M^{(k,m)*}M^{(l,m)}. \tag{1.93}$$

Let us note that (for the same reasons as before)

interaction of multipoles of different parity gives zero

and this time we have to do with the interaction of the multipoles of complete cells. The quantities $U_T^{(k,l)}$ are given in Table 1.2.

Do the Fock matrix elements and the total energy per cell represent finite values?

If the Fock matrix elements were infinite, then we could not manage to carry out the Hartree–Fock–Roothaan self-consistent procedure. If E_T were infinite, the periodic system could not exist at all. It is, therefore, important to know when we can safely *model* an infinite system.

For any finite system there is no problem: the results are always finite. The only danger, therefore, is summation to infinity ("lattice sums"), which always ends with the interaction of a part or whole unit cell with an infinite number of distant cells. Let us take such an example in the simplest case of a single atom per cell. Let us assume that the atoms interact by the Lennard-Jones pairwise potential (p. V1-406), $E = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right]$, where r means the interatomic distance, r_0 means the equilibrium distance, and ε denotes the depth of the potential well. Let us try to compute the lattice sum $\sum_j E_{0j}$, where E_{0j} means the interaction energy of the cells 0 and j. We see that, due to the form of the potential, for long distances what counts is the uniquely attractive term $-2\varepsilon\left(\frac{r_0}{r}\right)^6$. When we take such interactions which pertain to a sphere of the radius R (with the origin located on atom 0), each individual term (i.e., its absolute value) decreases with increasing R. This is very important, because when we have a three-dimensional lattice, the number of such interactions within the sphere *increases* as R^3. We see that the decay rate of the interactions will finally win and the lattice sum will converge. We can, however, easily see that if the decay of the pairwise interaction energy were slower, then we might have had trouble calculating the lattice sum. For example, if, instead of the neutral Lennard-Jones atoms, we took ions of the same charge, the interaction energy would explode to ∞. It is evident, therefore, that for each periodic system there are some conditions to be fulfilled if we want to have finite lattice sums.

These conditions are more severe for the Fock matrix elements because each of the terms represents the interaction of a *charge* with complete distant unit cells. The convergence depends on the asymptotic interaction energy of the potential. In the case of the multipole–multipole interaction, we know what the asymptotic behavior looks like, i.e., it is $R^{-(0+l+1)} = R^{-(l+1)}$, where R stands for the intercell distance. The lattice summation in an n-dimensional lattice ($n = 1, 2, 3$) gives the partial sum dependence on R as $\frac{R^n}{R^{l+1}} = R^{n-l-1}$. This means that[65]

> in 1D the unit cell cannot have any nonzero net charge ($l = 0$), in 2D it cannot have a nonzero charge and dipole moment ($l = 1$), and in 3D it cannot have a nonzero charge, dipole moment, and quadrupole moment ($l = 2$).

[65] L.Z. Stolarczyk, L. Piela, *Intern. J. Quantum Chem.*, 22(1982)911.

1.13 Back to the exchange term

The long-range effects discussed so far result from the Coulomb interaction in the Fock equation for a regular polymer. There is, however, also an exchange contribution, which has been postponed in the long-range region (p. 58). It is time now to reconsider this contribution. The exchange term in the Fock matrix element F_{pq}^{0j} had the form (see (1.67))

$$-\frac{1}{2}\sum_{h,l}\sum_{rs} P_{sr}^{lh}\left(\begin{smallmatrix} 0h \\ pr \end{smallmatrix}\middle|\begin{smallmatrix} lj \\ sq \end{smallmatrix}\right) \tag{1.94}$$

and gave the following contribution to the total energy per unit cell:

$$E_{exch} = \sum_{j} E_{exch}(j), \tag{1.95}$$

where the cell 0–cell j interaction has the form (see (1.81))

$$E_{exch}(j) = -\frac{1}{4}\sum_{h,l}\sum_{pqrs} P_{qp}^{j0} P_{sr}^{lh}\left(\begin{smallmatrix} 0h \\ pr \end{smallmatrix}\middle|\begin{smallmatrix} lj \\ sq \end{smallmatrix}\right). \tag{1.96}$$

It would be very nice to have the exchange contribution $E_{exch}(j)$ decaying fast when j increases, because it could be enclosed in the short-range contribution. Do we have good prospects for this? The above formula shows (the integral) that the summation over l is *safe*: the contribution of those cells l that are far from cell 0 is negligible due to differential overlaps of type $\chi_p^0(1)^*\chi_s^l(1)$. The summation over cells h is safe as well (for the same reasons), because it is bound to be limited to the neighborhood of cell j (see the integral).

> In contrast, the only guarantee of a satisfactory convergence of the sum over j is that *we hope* the matrix element P_{qp}^{j0} decays fast if j increases.

So far, exchange contributions have been neglected, and there has been an indication that suggested this was the right procedure. This was the magic word "exchange." All the experience of my colleagues and myself in intermolecular interactions whispers "this is surely a short-range type." In a manuscript by Sandor Suhai, I read that the exchange contribution is of a long-range type. To our astonishment this turned out to be right (after a few numerical experiments). We have a long-range exchange. After analysis was performed it turned out that

> the long-range exchange interaction appears if and only if the system is metallic.

A metallic system is notorious for its HOMO-LUMO quasidegeneracy; therefore, we began to suspect that when the HOMO-LUMO gap decreases, the P_{qp}^{j0} coefficients do not decay with j.

Such things are most clearly seen when the simplest example is taken, and the hydrogen molecule at long internuclear distance is the simplest prototype of a metal. Indeed, this is a system with half-filled orbital energy levels when the LCAO MO method is applied (in the simplest case: two atomic orbitals). Note that, after subsequently adding two extra electrons, the resulting system (let us not worry too much that such a molecule could not exist!) would model an insulator, i.e., all the levels are doubly occupied.[66]

Analysis of these two cases convinces us that indeed our suspicions were justified. The bond-order matrices we obtain in both cases (see Appendix D, p 595, S denotes the overlap integral of the $1s$ atomic orbitals of atoms a and b) are

$$\mathbf{P} = (1+S)^{-1}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} \quad \text{for } H_2, \tag{1.97}$$

$$\mathbf{P} = (1-S^2)^{-1}\begin{pmatrix}1 & -S\\ -S & 1\end{pmatrix} \quad \text{for } H_2^-. \tag{1.98}$$

We see[67] how profoundly these two cases differ in the off-diagonal elements (they are analogs of P_{qp}^{j0} for $j \neq 0$).

In the second case the proportionality of P_{qp}^{j0} and S ensures an *exponential, therefore very fast, decay* if j tends to ∞, in the first case there is no decay of P_{qp}^{j0} at all.

A detailed analysis for an infinite chain of hydrogen atoms ($\omega = 1$) leads to the following formula[68] for P_{qp}^{j0}:

$$P_{11}^{j0} = \frac{2}{\pi j}\sin\left(\frac{\pi j}{2}\right). \tag{1.99}$$

This means an extraordinarily slow decay of these elements (and therefore of the exchange contribution) with j. When the metallic regime is even slightly removed, the decay gets much, much faster.

---

66 Of course, we could take two helium atoms. This would be also good. However, the first principle in research is "*in a single step only change a single parameter, analyze the result, draw the conclusions, and make the second step*."
Just *en passant*, a second principle also applies here. If we do not understand an effect, what should we do? Just divide the system in two parts and look where the effect persists. Keep dividing until the effect disappears. Take the simplest system in which the effect still exists, analyze the problem, understand it and go back slowly to the original system (this is why we have H_2 and H_2^{2-} here).

67 L. Piela, J.-M. André, J.G. Fripiat, J. Delhalle, *Chem. Phys. Lett.*, 77(1981)143.

68 I.I. Ukrainski, *Theor. Chim. Acta*, 38(1975)139, $q = p = 1$ means that we have a single $1s$ hydrogen orbital per unit cell.

This result shows that the long-range character of the exchange interactions does not exist in reality. It seems to represent an indication that the Hartree–Fock method fails in such a case.

1.14 Choice of unit cell

The concept of the unit cell has been important throughout the present chapter. The unit cell represents an object that, when repeated by translations, gives an infinite crystal. In this simple definition almost every word can be a trap. Is it feasible? Is the choice unique? If not, then what are the differences among them? How is the motif connected to the unit cell choice? Is the motif unique? Which motifs may we think about? As we have already noted, the choice of unit cell as well as of motif is not unique. This is easy to see. Indeed Fig. 1.22 shows that the unit cell and the motif can be chosen in many different and equivalent ways. Moreover, there is no chance of telling in a responsible way which of the choices are reasonable and which are not. And it happens that in this particular case we really have a plethora of choices. Putting no limits to our fantasy, we may choose a unit cell in a particularly capricious way (Figs. 1.22b and 1.23).

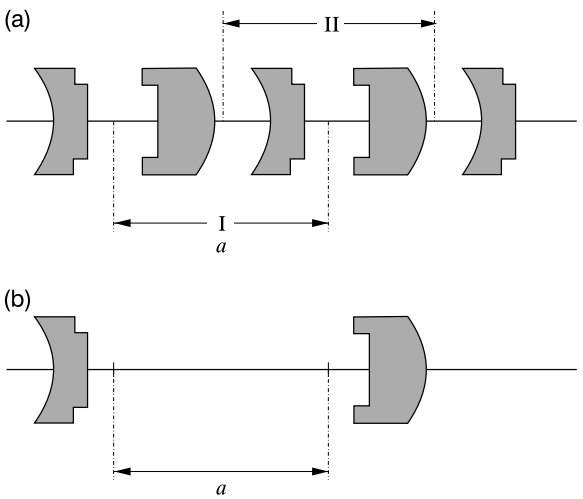


Fig. 1.22. Three of many possible choices of the unit cell motif. (a) Choices I and II differ, but both look "reasonable." (b) Choice III might be called strange. Despite this strangeness, choice III is as legal (from the point of view of mathematics) as I or II.

Fig. 1.23 shows six different, fully legitimate choices of motifs associated with a unit cell in a one-dimensional "polymer" $(LiH)_\infty$. Each motif consists of the lithium nucleus, a proton,

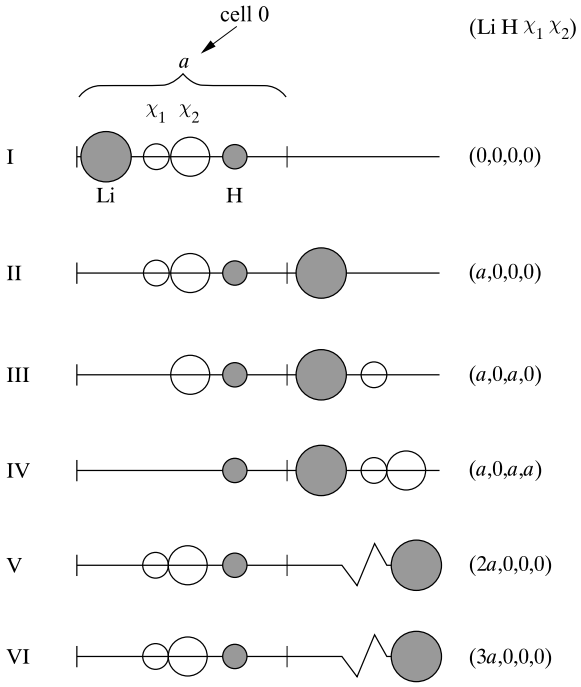


Fig. 1.23. Six different choices (I–VI) of unit cell content (motifs) for a linear chain $(LiH)_\infty$. Each cell has the same length $a = 6.3676$ a.u. There are two nuclei, Li^{3+} and H^+, and two Gaussian doubly occupied $1s$ atomic orbitals (denoted by χ_1 and χ_2 with exponents 1.9815 and 0.1677, respectively) *per cell*. Motif I corresponds to a "common sense" situation: both nuclei and electron distribution determined by χ_1 and χ_2 are within the section $(0,a)$. The other motifs (II–VI), all corresponding to the same unit cell $(0,a)$ of length a, are very strange. Each motif is characterized by the symbol (k,l,m,n), which means that the Li nucleus, the H nucleus, χ_1, and χ_2 are shifted to the right by ka, la, ma, na, respectively. *All the unit cells with their contents (motifs) are fully justified, equivalent from a mathematical point of view, and, therefore, "legal" from the point of view of physics*. Note that the nuclear framework and the electronic density corresponding to a cell are very different for all the choices.

and an electronic charge distribution in the form of two Gaussian $1s$ orbitals that accommodate four electrons altogether. By repeating any of these motifs we reconstitute the same original chain.

We may say there may be many legal choices of motif, but this is without any theoretical meaning, because all choices lead to the same infinite system. Well, this is true with respect to theory, but in practical applications the choice of motif may be of prime importance. We can see this from Table 1.3, which corresponds to Fig. 1.23.

The results without taking into account the long-range interactions depend very strongly on the choice of unit cell motif.

Table 1.3. Total energy per unit cell E_T in the "polymer" LiH as a function of unit cell definition (Fig. 1.23). For each choice of unit cell this energy is computed in four ways: (1) without long-range forces (long range = 0), i.e., unit cell 0 interacts with $N = 6$ unit cells on its right-hand side and N unit cells on its left-hand-side and (2), (3), (4) with the long range computed with multipole interactions up to the a^{-3}, a^{-5}, and a^{-7} terms, respectively. The bold figures are exact. The corresponding dipole moment μ of the unit cell (in Debyes) is also given.

Unit cell	Long range	μ	$-E_T$
I	0	6.6432	**6.61**0869
	a^{-3}	6.6432	**6.612794**92
	a^{-5}	6.6432	**6.612794**87
	a^{-7}	6.6432	**6.612794674**
II	0	−41.878	**6**.524802885
	a^{-3}	−41.878	**6.612**519674
	a^{-5}	−41.878	**6.61279**0564
	a^{-7}	−41.878	**6.612794**04
III	0	−9.5305	**6.60**7730984
	a^{-3}	−9.5305	**6.61278**8446
	a^{-5}	−9.5305	**6.612794**633
	a^{-7}	−9.5305	**6.6127946**73
IV	0	22.82	**6.5**7395630
	a^{-3}	22.82	**6.612**726254
	a^{-5}	22.82	**6.612793**807
	a^{-7}	22.82	**6.612794**62
V	0	−90.399	6.148843431
	a^{-3}	−90.399	**6.60**7530384
	a^{-5}	−90.399	**6.61**2487745
	a^{-7}	−90.399	**6.612**74317

Use of the multipole expansion greatly improves the results and, to very high accuracy, makes them *independent of the choice of unit cell motif*, as it should be.

Note that the larger the dipole moment of the unit cell, the worse the results with the short-range forces only. This is understandable, because the first nonvanishing contribution in the multipole

expansion is the dipole–dipole term (cf. Appendix G). Note how considerably the unit cell dependence drops after this term is switched on (a^{-3}).

The conclusion is that in the standard (i.e., short-range) calculations we should always choose the unit cell motif that corresponds to the smallest dipole moment. It seems however that such a motif is what everybody would choose using their "common sense."

1.14.1 Field compensation method

In a moment we will unexpectedly find a quite different conclusion. The logical chain of steps that led to it has, in my opinion, a didactic value, and contains a considerable amount of optimism. When this result was obtained by Leszek Stolarczyk and myself, we were stunned by its simplicity.

> Is it possible to design a unit cell motif with a dipole moment of zero? This would be a great unit cell, because its interaction with other cells would be weak and it would decay fast with intercellular distance. We could therefore compute the interaction of a few cells like this and the job would be over: we would have an accurate result at very low cost.

There is such a unit cell motif.

Imagine we start from the concept of the unit cell with its motif (with lattice constant a). This motif is, of course, electrically neutral (otherwise the total energy would be $+\infty$), and its dipole moment component along the periodicity axis is equal to μ. Let us put its symbol in the unit cell (Fig. 1.24a).

Now let us add to the motif two extra (i.e., fictitious) point-like opposite charges ($+q$ and $-q$), located on the periodicity axis and separated by a. The charges are chosen in such a way ($q = \frac{\mu}{a}$) that they alone give the dipole moment component along the periodicity axis equal to $-\mu$ (Fig. 1.24b).

In this way the new unit cell dipole moment (with the additional fictitious charges included) is equal to zero. Is this an acceptable choice of motif? Well, what does acceptable mean? The only requirement is that by repeating the new motif with period a, we have to reconstruct the whole crystal. What will we get when repeating the new motif? Let us see (Fig. 1.24c).

We get the original periodic structure, because the charges all along the polymer, except the boundaries, have canceled each other out. Simply, the pairs of charges $+q$ and $-q$ when located at a point result in nothing.

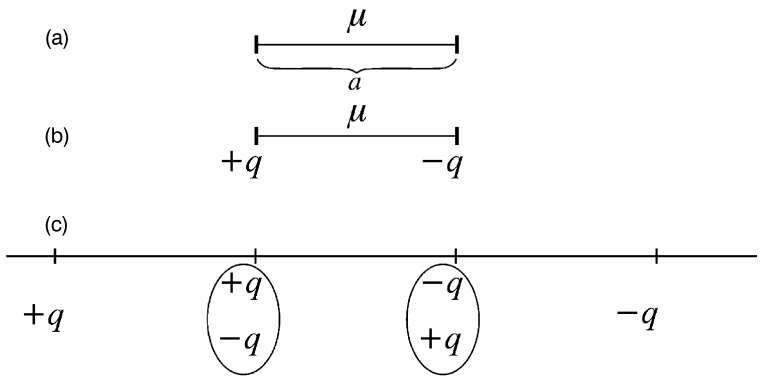


Fig. 1.24. Field compensation method. (a) The unit cell with length a and dipole moment $\mu > 0$. (b) The modified unit cell with additional fictitious charges ($|q| = \frac{\mu}{a}$) which cancel the dipole moment. (c) The modified unit cells (with $\mu' = 0$) give the original polymer when added together.

In practice we would like to repeat just a few neighboring unit cell motifs (a cluster) and then compute their interaction. In such case, we will observe the charge cancellation inside the cluster, but no cancellation on its boundaries ("surface").

> Therefore we get a sort of point charge distribution at the boundaries.

If the boundary charges did not exist, it would correspond to the traditional calculations of the original unit cells without taking any long-range forces into account. The boundary charges therefore play the important role of replacing the electrostatic interaction with the rest of the *infinite* crystal, by the boundary charge interactions with the cluster ("field compensation method").

This is all. The consequences are simple.

> Let us not only kill the dipole moment, but also other multipole moments of the unit cell content (up to a maximum moment), and the resulting cell will be unable to interact electrostatically with anything. Therefore, interaction within a small cluster of such cells will give us an accurate result for the energy per cell.

This multipole killing (field compensation) may be carried out in several ways.[69]

Application of the method is extremely simple. Imagine unit cell 0 and its neighbor unit cells (a cluster). Such a cluster is sometimes treated as a molecule and its role is to represent a bulk

[69] L. Piela, L.Z. Stolarczyk, *Chem. Phys. Letters*, 86(1982)195.

crystal. This is a very expensive way to describe the bulk crystal properties, for the cluster surface atom ratio to the bulk atom is much higher than we would wish (the surface still playing an important role). What is lacking is the crystal field that will change the cluster properties. In the field compensation method we do the same, but there are some fictitious charges at the cluster boundaries that take care of the crystal field. This enables us to use a smaller cluster than before (low cost) and still get the influence of the infinite crystal. The fictitious charges are treated the same way in computations as the nuclei (even if some of them are negatively charged). However artificial it may seem, the results are far better when using the field compensation method than without it.

1.14.2 The symmetry of subsystem choice

The example described above raises an intriguing question, pertaining to our understanding of the relation between a part and the whole.

There are an infinite number of ways to reconstruct the same system from parts. *These ways are not equivalent in practical calculations*, if for any reason we are unable to compute all the interactions in the system. However, if we have a theory (in our case the multipole method) that is able to compute the interactions,[70] including the long-range forces, then it turns out the final result is virtually independent of the choice of unit cell motif. This arbitrariness of choice of subsystem looks analogous to the arbitrariness of the choice of coordinate system. The final results do not depend on the coordinate system used, but still the numerical results (as well as the effort to get the solution) do.

The separation of the whole system into subsystems is of key importance to many physical approaches, but we rarely think of the freedom associated with the choice. For example, an atomic nucleus does not in general represent an elementary particle, and yet in quantum mechanical calculations we treat it as a point particle, without an internal structure, and we are successful.[71] Further, in the *Bogolyubov*[72] *transformation*, the Hamiltonian is represented by creation and annihilation operators, each being a linear combination of the creation and annihilation operators for electrons (described in Appendix C, p. 587). The new operators also fulfill the anticommutation rules, only the Hamiltonian contains more additional terms than before. A particular Bogolyubov transformation may describe the creation and annihilation of quasiparticles, such as the electron hole. We are dealing with the same physical system as before, but we

70 With controlled accuracy, i.e., we still neglect the interactions of higher multipoles.

71 This represents only a fragment of the story-like structure of science (cf. p. V1-76), one of its most intriguing features. It makes science operate, otherwise when considering the genetics of peas in biology we have had to struggle with the quark theory of matter.

72 Nicolai Nicolaevitch Bogolyubov (1909–1992), Russian physicist, director of the Dubna Nuclear Institute, outstanding theoretician.

look at it from a different point of view, by considering it composed of something else. Is there any theoretical (i.e., *serious*) reason for preferring one division into subsystems over another? Such a reason may be only of practical importance. Any correct theory should give the same description of the total system *independently of subsystems we decide to choose*.

SYMMETRY WITH RESPECT TO DIVISION INTO SUBSYSTEMS

The symmetry of *objects* is important for the description of *them*, and therefore may be viewed as of limited interest. The symmetry of the laws of Nature, i.e., of the theory that describes all objects (whether symmetric or not), is much more important. This has been discussed in detail in Chapter V1-2 (cf. p. V1-77), but it seems that we did not list there a fundamental symmetry of any correct theory: the *symmetry with respect to the choice of subsystems. A correct theory has to describe the total system independently of what we decide to treat as subsystems.*

We will meet this problem once more in intermolecular interactions (Chapter V1-5). However, in the periodic system it has been possible to use, in computational practice, the symmetry described above.

Our problem resembles an excerpt found in "*Dreams of a Final Theory*" by Steven Weinberg[73] pertaining to *gauge symmetry*: "*The symmetry underlying it has to do with changes in our point of view about the identity of the different types of elementary particle. Thus it is possible to have a particle wave function that is neither definitely an electron nor definitely a neutrino, until we look at it*." Here also we have freedom in the choice of subsystems and a correct theory has to reconstitute the description of the whole system.

An intriguing problem.

Summary

- A crystal is often approximated by an infinite crystal (*primitive) lattice,* which leads to the concept of the *unit cell*. By translationally repeating a chosen atomic *motif* associated with a unit cell, we reconstruct the whole infinite crystal.
- The one-electron Hamiltonian is invariant with respect to translations by any lattice vector. Therefore its eigenfunctions (crystal orbitals) are simultaneously eigenfunctions of the translation operators (Bloch theorem) $\hat{T}(\mathbf{R}_j)\phi_{\mathbf{k}}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r}-\mathbf{R}_j) = \exp\left(-i\mathbf{k}\mathbf{R}_j\right)\phi_{\mathbf{k}}(\mathbf{r})$ and transform according to the irreducible representation of the translational group labeled by the *wave vector* $\mathbf{k}$.
- *Bloch functions* may be treated as atomic symmetry orbitals $\phi = \sum_j \exp\left(i\mathbf{k}\mathbf{R}_j\right)\chi\left(\mathbf{r}-\mathbf{R}_j\right)$ formed from the atomic orbital $\chi(\mathbf{r})$. Their symmetry is determined by $\mathbf{k}$.

73 Pantheon Books, New York (1992), Chapter 6.

- The crystal lattice basis vectors allow the formation of the basis vectors of the *inverse lattice*. Linear combinations of them (with integer coefficients) determine the *inverse lattice* subject to translational symmetry.
- A special (*Wigner–Seitz)* unit cell of the inverse lattice is called the first Brillouin zone (FBZ).
- The vectors **k** inside the FBZ label possible nonequivalent irreducible representations of the translational group.
- The wave vector plays a triple role:
  - it indicates the *direction of the wave*, which is an eigenfunction of $\hat{T}(\mathbf{R}_j)$ with eigenvalue $\exp\left(-i\mathbf{k}\mathbf{R}_j\right)$;
  - it labels the *irreducible representations* of the translational group;
  - the longer the wave vector **k**, the more nodes the wave has.
- In order to neglect the crystal surface, we apply the *Born–von Kármán boundary condition*: "instead of a stick-like system we take a circle."
- In full analogy with molecules, we can formulate the SCF LCAO CO Hartree–Fock–Roothaan method (CO instead MO). Each CO is characterized by a vector $\mathbf{k} \in$ FBZ and is a linear combination of the Bloch functions with the same **k**.
- The orbital energy dependence on $\mathbf{k} \in$ FBZ is called the *energy band*. The stronger the intercell interaction, the wider the band width (dispersion).
- Electrons occupy the *valence bands*, the *conduction bands* are empty. The *Fermi level* is the HOMO energy of the crystal. If the HOMO-LUMO energy difference (*energy gap* between the valence and conduction bands) is zero, we have a *metal*; if it is large, we have an *insulator*; if it is medium, we have a *semiconductor*.
- Semiconductors may be intrinsic, *n*-type (if the donor dopant levels are slightly below the conduction band), or *p*-type (if the acceptor dopant levels are slightly above the occupied band).
- Metals when cooled may undergo what is known as the Peierls transition, which denotes lattice dimerization and band gap formation. In this way the system changes from a metal to a semiconductor or insulator. This transition corresponds to the Jahn–Teller effect in molecules.
- Polyacetylene is an example of a Peierls transition ("dimerization"), which results in shorter bonds (a little "less-multiple" than double ones) and longer bonds (a little "more multiple" than single ones). Such a dimerization introduces the possibility of a defect separating two rhythms ("phases") of the bonds: from "double-single" to "single-double." This defect can move within the chain, which may be described as a solitonic wave. The soliton may become charged and, in this case, participates in electric conduction (increasing it by many orders of magnitude).
- In polyparaphenylene, a soliton wave is not possible, because the two phases, quinoid and aromatic, are not of the same energy, excluding free motion. A double effect is possible though, a bipolaron. Such a defect represents a section of the quinoid structure (in the aromatic-like chain) at the end of which we have two unpaired electrons. The electrons, when paired with extra electrons from donor dopants or when removed by acceptor dopants, form a double ion (bipolaron), which may contribute to electric conductance.
- The band structure may be foreseen in simple cases and logically connected to the subsystem orbitals.
- To compute the Fock matrix elements or the total energy per cell, we have to calculate the interaction of cell 0 with all other cells.
- The interaction with neighboring cells is calculated without approximations, while that with distant cells uses multipole expansion. Multipole expansion applied to the electrostatic interaction gives accurate results, while the numerical effort is dramatically reduced.

- In some cases (metals), we find long-range exchange interaction, which disappears as soon as the energy gap emerges. This indicates that the Hartree–Fock method is not applicable in this case.
- The choice of unit cell motif is irrelevant from the theoretical point of view, but leads to different numerical results when the long-range interactions are omitted. If all interactions (including the long-range ones) are taken into account, the theory becomes independent of the division of the whole system into arbitrary motifs.

Main concepts, new terms

band (p. 21)
band gap (p. 35)
band structure (p. 26)
band width (p. 32)
biorthogonal basis (p. 11)
bipolaron (p. 37)
Bloch function (p. 10)
Bloch theorem (p. 10)
Born–von Kármán boundary condition (p. 23)
cyclic group (p. 24)
conduction band (p. 35)
crystal orbitals (p. 28)
energy gap (p. 33)
Fermi level (p. 33)
field compensation method (p. 74)
first Brillouin zone (p. 14)
Hartree–Fock method (p. 51)
insulators (p. 35)
intrinsic semiconductor (p. 35)
inverse lattice (p. 11)
Jahn–Teller effect (p. 35)
lattice constant (p. 6)
LCAO CO (p. 31)
long-range exchange (p. 68)
long-range problem (p. 56)
metals (p. 35)
motif (p. 6)
multipole expansion (p. 61)
multipole moment (p. 62)
n-type semiconductor (p. 38)
Peierls transition (p. 35)
primitive lattice (p. 6)
p-type semiconductor (p. 38)
quasicrystal (p. 4)
Riemann zeta function (p. 63)
SCF LCAO CO (p. 55)
soliton (p. 37)
symmetry of division into subsystems (p. 75)
symmetry orbital (p. 10)
translational symmetry (p. 6)
translation operator (p. 6)
translation vector (p. 6)
unit cell (p. 6)
valence band (p. 35)
wave vector (p. 8)
Wigner–Seitz cell (p. 14)

From the research front

The Hartree–Fock method for periodic systems nowadays represents a routine approach coded in several *ab initio* computer packages. We may analyze the total energy, its dependence on molecular conformation, the density of states, the atomic charges, etc. Also calculations of first-order responses to the electric field (polymers are of interest for optoelectronics) have been successful in the past. However, nonlinear problems (like the second harmonic generation, see Chapter V1-4) still represent a challenge.

Ad futurum

Probably there will be no problem in carrying out the Hartree–Fock or DFT (see Chapter V1-3) calculations soon, even for complex polymers and crystals. What will remain for a few decades is the very important problem of lowest-energy crystal packing and of solid state reactions and phase transitions. Post-Hartree–Fock calculations will be more and more important, taking into account electronic correlation effects. The real challenge will start in designing nonperiodic materials, where the polymer backbone will serve as a molecular rack for installing some functions (transport, binding, releasing, signal transmitting). The functions will be expected to cooperate ("intelligent materials," cf. Chapter V1-7).

Additional literature

R. Hoffmann, "Solids and Surfaces. A Chemist's View of Bonding in Extended Structures," VCH publishers, New York, 1988.

A masterpiece written by a Nobel Prize winner, one of the founders of solid state quantum chemistry. More oriented towards chemistry than Levin's book. Solid state theory was traditionally the domain of physicists, some concepts typical of chemistry as, e.g., atomic orbitals, bonding and antibonding effects, chemical bonds, and localization of orbitals were usually absent in such descriptions.

J.-M. André, J. Delhalle, J.-L. Brédas, "Quantum Chemistry Aided Design of Organic Polymers," World Scientific, Singapore, 1991.

A well-written book oriented mainly towards the response of polymers to the electric field.

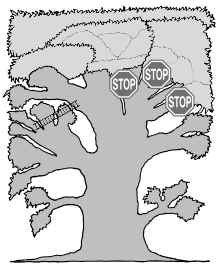
Questions

1. The Bloch theorem (where $\phi_{\mathbf{k}}(\mathbf{r})$ stands for a crystal orbital, $\mathbf{R}_j$ is a lattice vector, and $\mathbf{k}$ is a wave vector):
   a. pertains to the eigenvalue (corresponding to $\phi_{\mathbf{k}}$) of the translation operator by a lattice vector.
   b. $|\phi_{\mathbf{k}}(\mathbf{r})| = \left|\phi_{\mathbf{k}}(\mathbf{r} - \mathbf{R}_j)\right|$.
   c. $|\phi_{\mathbf{k}}(\mathbf{r})|^2$ exhibits the same symmetry as the potential energy $V(\mathbf{r})$.
   d. if $\phi_{\mathbf{k}}(\mathbf{r})$ corresponds to the wave vector $\mathbf{k}$, then $\phi_{\mathbf{k}}(\mathbf{r} - \mathbf{R}_j) = \exp\left(-i\mathbf{k} \cdot \mathbf{R}_j\right) \phi_{\mathbf{k}}(\mathbf{r})$.
2. The FBZ:
   a. means the smallest unit cell of a primitive lattice.
   b. means a smallest motif to be repeated in a crystal.
   c. does not contain in its inner part any pair of equivalent wave vectors.
   d. the wave vectors that correspond to the surface of the FBZ differ by an inverse lattice vector.
3. A function $\phi_{\mathbf{k}}$ corresponding to the wave vector $\mathbf{k}$:
   a. for $\mathbf{k} = \mathbf{0}$ the function $\phi_{\mathbf{k}}$, built of $1s$ atomic orbitals, does not have any nodal planes.
   b. represents a wave with the front perpendicular to $\mathbf{k}$.
   c. the larger $|\mathbf{k}|$, the greater the number of nodes.
   d. has to be a CO.

4. A CO:
   a. represents any linear combination of the atomic orbitals of the atoms the crystal is composed of.
   b. is characterized by its wave vector **k**.
   c. with $\mathbf{k} = \mathbf{0}$ corresponds to the lowest energy for a given electronic band.
   d. two crystal orbitals that differ by any inverse lattice vector are identical.
5. An infinite polyacetylene chain:
   a. represents a conductor.
   b. exhibits an alternation of the CC bond lengths.
   c. exhibits a nonzero energetic gap between the valence band and the conduction band.
   d. conducts the electric current thanks to the solitonic defects that result from donor or acceptor doping.
6. The band width increases if:
   a. one goes from the COs that correspond to the inner electronic shells to the COs corresponding to valence electrons.
   b. one increases a pressure.
   c. the distance between atoms gets larger.
   d. the atomic orbitals overlap more.
7. A semiconductor:
   a. exhibits a small energy gap.
   b. has about a half of the conductivity of copper.
   c. has the energy gap equal to zero.
   d. conducts electricity, but only in one direction.
8. The Fermi level:
   a. represents an electronic energy level from which removing an electron needs the least energy.
   b. has the energy corresponding to the HOMO orbital of the crystal.
   c. means the mean energy of the occupied electronic states.
   d. is the lowest energy of the conducting band.
9. The dipole–quadrupole interaction per unit cell in a regular polymer is:
   a. 0.
   b. equal to the difference between the dipole–dipole and quadrupole–quadrupole interactions.
   c. equal to the mean value of the dipole–dipole and quadrupole–quadrupole interactions.
   d. a sum of the interactions of the dipole of the unit cell 0 with the quadrupoles from beyond the section of cells $-N, \ldots N$.
10. The unit cell dipole moment in a regular polymer:
    a. is uniquely defined for an electrically neutral polymer.
    b. does not depend on the position of the cell with respect to cell 0.
    c. must be equal to 0, otherwise the total dipole–dipole interaction energy would be equal to ∞.
    d. does depend on the choice of the unit cell.

Answers

1a,b,c,d, 2c,d, 3a,b,c, 4b,d, 5b,c,d, 6a,b,d, 7a, 8a,b, 9a,d, 10b,d



CHAPTER 2

Correlation and Anticorrelation of Electronic Motions

God does not care about our mathematical difficulties, he integrates empirically.
Albert Einstein

Where are we?

The main road on the trunk leads us to the right-hand side part of the crown of the tree.

An example

As usual let us consider the simplest example: the hydrogen molecule. The normalized Hartree–Fock determinant with double occupancy of the normalized molecular orbital φ reads as

$$\psi_{RHF}(1,2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(1) & \phi_1(2) \\ \phi_2(1) & \phi_2(2) \end{vmatrix} = \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi(\mathbf{r}_1)\alpha(\sigma_1) & \varphi(\mathbf{r}_2)\alpha(\sigma_2) \\ \varphi(\mathbf{r}_1)\beta(\sigma_1) & \varphi(\mathbf{r}_2)\beta(\sigma_2) \end{vmatrix}.$$

The key quantity here is $|\psi_{RHF}(1,2)|^2$, since it tells us about the probability density of the occurrence of certain coordinates of the electrons. We will study the fundamental problem for the motion of electrons: whether the electrons react to their presence.

Let us ask very important questions. What is the probability density of occurrence of the situation when:

- electron 1 has a position (x_1, y_1, z_1) and the spin coordinate $\sigma_1 = 1/2$, while electron 2 has spin coordinate $\sigma_2 = -1/2$ and its space coordinates are (x_2, y_2, z_2)? Let us calculate the (conditional) probability density of finding such situation. We have

$$|\psi_{RHF}(1,2)|^2 = \left[\varphi(1)\varphi(2)\frac{1}{\sqrt{2}}\{\alpha(\sigma_1)\beta(\sigma_2) - \beta(\sigma_1)\alpha(\sigma_2)\}\right]^2 =$$
$$\left[\varphi(x_1,y_1,z_1)\ \varphi(x_2,y_2,z_2)\frac{1}{\sqrt{2}}\{\alpha(1/2)\beta(-1/2) - \beta(1/2)\alpha(-1/2)\}\right]^2 =$$
$$\left[\varphi(x_1,y_1,z_1)\varphi(x_2,y_2,z_2)\frac{1}{\sqrt{2}}\{1\times 1 - 0\times 0\}\right]^2 = \frac{1}{2}\varphi^2(x_1,y_1,z_1)\times\varphi^2(x_2,y_2,z_2).$$

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00010-7
Copyright © 2020 Elsevier B.V. All rights reserved.

Now imagine that electron 1 changes its position but along an orbital contour surface $\varphi = const$. *The distribution of the probability density* $\frac{1}{2}const^2\varphi^2\,(x_2, y_2, z_2)$ *of electron* 2 *does not change a bit, although electron* 2 *should move away from its partner, since the electrons repel each other*. Electron 2 is not afraid to approach electron 1. The latter can even touch electron 2 and it does not react at all. For such a deficiency we have to pay through the high average value of the Hamiltonian (since there is a high average energy of the electron repulsion). The Hartree–Fock method therefore has an obvious shortcoming.

- we leave everything the same as before, but now electron 2 has spin coordinate $\sigma_2 = 1/2$ (so this is the situation where both electrons have identical projections of spin angular momentum[1])? What will the response to this change be of $|\psi_{RHF}\,(1, 2)|^2$ as a function of the position of electron 2?

Again we calculate

$$|\psi_{RHF}\,(1, 2)|^2 = \left[const\ \varphi\,(x_2, y_2, z_2)\,\frac{1}{\sqrt{2}}\{\alpha\left(\frac{1}{2}\right)\beta\left(\frac{1}{2}\right) - \beta\left(\frac{1}{2}\right)\alpha\left(\frac{1}{2}\right)\}\right]^2 =$$
$$\left[const\ \varphi\,(x_2, y_2, z_2)\,\frac{1}{\sqrt{2}}\{1 \times 0 - 0 \times 1\}\right]^2 = 0.$$

We ask about the distribution of the electron of the same spin. The answer is that this distribution is *everywhere equal to zero*, i.e., we do not find electron 2 with spin coordinate $\frac{1}{2}$ independently of the position of electron 1 with spin coordinate $\frac{1}{2}$ (in whatever point on the contour line or beyond it).

The second conclusion can be accepted, since it follows from the pairing of the spins,[2] but the first conclusion is just absurd. Such nonsense is admitted by the Hartree–Fock method. In this chapter we will ponder how can we introduce a correlation of electronic motions.

We define the electronic *correlation energy* as

$$E_{corel} = E - E_{RHF},$$

where E is the energy entering the Schrödinger equation[3] and E_{RHF} is the restricted Hartree–Fock energy.[4] One has to note that the Hartree–Fock procedure takes into account the Pauli exclusion principle

[1] We may ask: "How come?" After all, we consider a singlet state, hence the spin projections are opposite. We will not find the situation with parallel spin projections. Take it easy. If, in fact, we are right then we will get 0 as the density of the respective conditional probability. Let us see whether it will really be so.

[2] And this is ensured by the singlet form of the spin part of the function.

[3] This is the rigorous nonrelativistic energy of the system in its ground state. This quantity is not available experimentally; we can *evaluate* it by subtraction of the calculated relativistic corrections from the energy of the total ionization of the system.

[4] Usually we define the correlation energy for the case of double occupancy of the molecular orbitals (the RHF method, see Chapter V1-8). In the case of open shells, especially when the multideterminantal description is required, the notion of correlation energy still remains to be defined. These problems will not be discussed in this book.

and therefore also the correlation of electrons of the same spin coordinate. Hence, the correlation energy E_{corel} is defined here with respect to the "Hartree–Fock level of electron correlation".

What is it all about?

The outline of the chapter is as follows:

- First we will discuss the methods which explicitly (via the form of the suggested wave function) allow the electrons to control their mutual distance ("a correlation of motions").
- In the second part of the chapter the correlation will be less visible, since it will be accounted for by application of linear combinations of the Slater determinants. First we will discuss the variational methods (VB, CI, MC SCF), and then the nonvariational ones (CC, EOM-CC, MBPT).

- Example of H_2O
- Which excitations are most important?
- Natural orbitals (NOs) a way to shorter expansions
- Size inconsistency of the CI expansion

Direct CI method (♦) p. 142

Multireference CI method (♦) p. 143

Multiconfigurational self-consistent field (MC SCF) method (△♦Ⓢ℧) p. 144

- Classical MC SCF approach (△)
- Unitary MC SCF method (♦)
- Complete active space (CAS SCF) method is size-consistent (♦Ⓢ℧◊)

NONVARIATIONAL METHODS WITH SLATER DETERMINANTS (△♦Ⓢ)

Coupled cluster (CC) method (♦◊) p. 149

- Wave and cluster operators
- Relationship between CI and CC methods
- Solution of the CC equations
- Example: CC with double excitations
- Size consistency of the CC method

Equation of motion (EOM-CC) method (♦) p. 159

- Similarity transformation
- Derivation of the EOM-CC equations

Many-body perturbation theory (MBPT) (♦) p. 162

- Unperturbed Hamiltonian
- Perturbation theory – slightly different presentation
- MBPT machinery – part one: energy equation
- Reduced resolvent or the "almost" inverse of $(E_0^{(0)} - \hat{H}^{(0)})$
- MBPT machinery – part two: wave function equation
- Brillouin–Wigner perturbation theory
- Rayleigh–Schrödinger perturbation theory

Møller–Plesset version of Rayleigh–Schrödinger perturbation theory (△◊) p. 169

- Expression for MP2 energy
- Is the MP2 method self-consistent?

- Convergence of the Møller–Plesset perturbational series
- Special status of double excitations

NONVARIATIONAL METHODS USING EXPLICITLY CORRELATED WAVE FUNCTIONS (⧫)

Møller–Plesset R12 method (MP2-R12) **p. 176**

- Resolution of identity (RI) method or density fitting (DF)
- Other RI methods

In chapter V1-8 we dealt with the description of electronic motion in the mean-field approximation. Now *we use this approximation as a starting point towards methods accounting for electron correlation*. Each of the methods considered in this chapter, when rigorously applied, should give an exact solution of the Schrödinger equation. Thus this chapter will give us access to methods providing accurate solutions of the Schrödinger equation.

Why is this important?

Perhaps, in our theories, the electrons do not need to correlate their motion and the chemistry will be still all right?

Unfortunately, this is not so. The mean-field method provides, to be sure, ca. 99% of the total energy of the system. This is certainly a lot; in many cases the mean-field method gives quite satisfactory results, but it still falls short of treating several crucial problems correctly. For example,

- *It is only because of electron correlation* the hydride ion (H^-) exists and the noble gas atoms attract each other (in accordance with experiment, liquefaction of gases).
- According to the Hartree–Fock method, the F_2 molecule *does not exist* at all, whereas the fact is that it exists, and is doing not bad (bonding energy amounts to about 38 kcal/mol).[5]
- About *half* the interaction energy of large molecules (often of biological importance) calculated at the equilibrium distance originates purely from the correlation effect.
- The RHF method used to describe the dissociation of the chemical bond gives simply tragic results (cf. Chapter V1-8, p. V1-511), which are *qualitatively wrong* (here the UHF method gives a qualitatively correct description).

We see that in many cases electronic correlation must be taken into account.

What is needed?

- Operator algebra (Appendix V1-B, necessary),

[5] Yet this is not a strong bond. For example, the bonding energy of the H_2 molecule equals 104 kcal/mol, that of HF 135 kcal/mol.

- Hartree–Fock method (Chapter V1-8, necessary),
- eigenvalue problem (Appendix V1-M, p. V1-705, necessary),
- variational method (Chapter V1-5, necessary),
- perturbation theory (Chapter V1-5, recommended),
- matrix diagonalization (Appendix V1-L, p. V1-703, recommended),
- second quantization (Appendix C, p. 587, necessary).

Classical works

The first calculations with electron correlation for molecules were performed by Walter Heitler and Fritz Wolfgang London in a paper "*Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik*," published in *Zeitschrift für Physik*, 44(1927)455. The covalent bond (in the hydrogen molecule) could be correctly described only after the electron correlation has been included. June 30, 1927, when Heitler and London submitted the paper, is the birth date of quantum chemistry. ★ The first calculations incorporating electron correlation in an atom (helium) were published by Egil Andersen Hylleraas in the article "*Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium*," *Zeitschrift für Physik*, 54(1929)347. ★ Later, significantly more accurate results were obtained for the hydrogen molecule by Hubert M. James and Albert S. Coolidge in the article "*The Ground State of the Hydrogen Molecule*," *Journal of the Chemical Physics*, 1(1933)825, and a contemporary reference point for that molecule are papers by Włodzimierz Kołos and Lutosław Wolniewicz, among others an article entitled "*Potential Energy Curves for the* $X^1\Sigma_g^+$, $B^3\Sigma_u^+$, $C^1\Pi_u$ *States of the Hydrogen Molecule*," published in *Journal of Chemical Physics*, 43(1965)2429. ★ Christian Møller and Milton S. Plesset in *Physical Review*, 46(1934)618 published a paper "*Note on an Approximation Treatment for Many-Electron Systems*," where they presented a perturbational approach to electron correlation. ★ The first calculations with the multiconfigurational self-consistent field (MC SCF) method for atoms were published by Douglas R. Hartree, his father William Hartree, and Bertha Swirles in the paper "*Self-Consistent Field, Including Exchange and Superposition of Configurations, with some Results for Oxygen*," *Philosophical Transactions of the Royal Society (London) A*, 238(1939)229, and the general MC SCF theory was presented by Roy McWeeny in the work "*On the Basis of Orbital Theories*," *Proceedings of the Royal Society (London) A*, 232(1955)114. ★ As a classic paper in electronic correlation we also consider an article by Per-Olov Löwdin, "*Correlation Problem in Many-Electron Quantum Mechanics*," in *Advances in Chemical Physics*, 2(1959)207. ★ The idea of the coupled cluster (CC) method was introduced by Fritz Coester in a paper in *Nuclear Physics*, 7(1958)421 entitled "*Bound States of a Many-Particle System*." ★ Jiří Čížek introduced the (diagrammatic) CC method into electron correlation theory in the paper "*On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-type Expansion Using Quantum-Field Theoretical Methods*," published in the *Journal of Chemical Physics,* 45(1966)4256. ★ The book edited by Oktay Sinanoğlu and Keith A. Brueckner "*Three Approaches to Electron Correlation in Atoms*," Yale Univ. Press, New Haven and London, 1970, contains several reprints of the papers which cleared the path towards the CC method. ★ A derivation of the CC equations for interacting nucleons was presented by Herman Kümmel and Karl-Heinz Lührmann, *Nuclear Physics A*, 191(1972)525 in a paper entitled "*Equations for Linked Clusters and the Energy Variational Principle*." ★ Werner Kutzelnigg in a paper "r_{12}-*Dependent*

terms in the wave function as closed sums of partial wave amplitudes for large l," published in *Theoretica Chimica Acta*, 68(1985)445 was the first to introduce the explicitly correlated wave functions in perturbational calculations for atoms and molecules.

* * *

Size consistency requirement

The methods presented in this chapter will take into account the electronic correlation. A particular method may deal better or worse with this difficult problem. The better it deals with it, the more convincing its results are.

There is however one requirement which we feel to be a natural one for any method that pretends to be reasonable. Namely,

> any reliable method when applied to a system composed of very distant (i.e., noninteracting) subsystems should give the energy, which is a sum of the energies for the individual subsystems. A method having this feature is known as size-consistent.[6]

Before we consider other methods, let us check whether our fundamental method, i.e., the Hartree–Fock method, is size-consistent or not.

Hartree–Fock method

As shown on p. V1-490, the Hartree–Fock electronic energy reads as (the summations go over the occupied spin orbitals) $E'_{HF} = \sum_i^{\mathrm{SMO}} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} [\langle ij|ij\rangle - \langle ij|ji\rangle]$, while the total energy is $E_{HF} = E'_{HF} + V_{nn}$, where the last term represents a constant repulsion of the nuclei. When the intersubsystem distances are infinite (they are then noninteracting), one can divide the spin orbitals $|i\rangle$, $i = 1, 2, ...N$, into nonoverlapping sets $i \in A$, $i \in B$, $i \in C$, ..., where $i \in A$ means the molecular spin orbital $|i\rangle$ is localized on subsystem A and represents a Hartree–Fock spin orbital of molecule A, etc. Then, in the limit of large distances (symbolized by lim, V_B stands for the operator of the interaction of the nuclei of molecule B with an electron,

[6] The size consistency has some theoretical issues to be solved. One may define the subsystems and their distances in many different ways, some of them quite weird. For example, one may consider all possible dissociation channels (with different products) with unclear electronic states to assume. Here we consider the simplest cases: the closed shell character of the total system and of the subsystems. Even this is not unique...

while $\lim V_{nn} = \sum_A V_{nn,A}$, with $V_{nn,A}$ representing the nuclear repulsion within molecule A, $E_{HF}(A)$ denotes the Hartree–Fock energy of molecule A), we have

$$\lim E_{HF} = \sum_{i}^{\text{SMO}} \lim \langle i|\hat{h}|i\rangle + \frac{1}{2}\lim \sum_{i,j=1}^{\text{SMO}} [\langle ij|ij\rangle - \langle ij|ji\rangle] + \lim V_{nn} =$$

$$\sum_A \left[\sum_{i\in A}^{\text{SMO}} \langle i|\hat{h}^A|i\rangle + \lim \sum_{i\in A}^{\text{SMO}} \langle i| \sum_{B\neq A} V_B|i\rangle \right] + \frac{1}{2}\sum_{i,j\in A}^{\text{SMO}} [\langle ij|ij\rangle - \langle ij|ji\rangle] +$$

$$\frac{1}{2}\lim \sum_{i\in A, j\in B}^{\text{SMO}} [\langle ij|ij\rangle - \langle ij|ji\rangle] + \sum_A V_{nn,A} =$$

$$\sum_A \left[\sum_{i\in A}^{\text{SMO}} \langle i|\hat{h}^A|i\rangle + 0 + \frac{1}{2}\sum_{i,j\in A}^{\text{SMO}} [\langle ij|ij\rangle - \langle ij|ji\rangle] + 0 + V_{nn,A} \right] = \sum_A E_{HF}(A).$$

The zeros in the above formula appeared instead of the terms that vanish because of the Coulombic interaction of the objects that are further and further from one another. For example, in the mixed terms $\frac{1}{2}\sum_{i\in A, j\in B}^{\text{SMO}}[\langle ij|ij\rangle - \langle ij|ji\rangle]$ the spin orbitals $|i\rangle$ and $|j\rangle$ belong to different molecules, and all integrals of the type $\langle ij|ij\rangle$ vanish, because they correspond to the Coulomb interaction of electron 1 with the probability density distribution $\phi_i^*(1)\,\phi_i(1)$ in molecule A with electron 2 with the distribution $\phi_j^*(2)\,\phi_j(2)$ centered on molecule B. Such an interaction vanishes as the inverse of the AB distance, i.e., goes to zero in the limit under consideration. The integrals $\langle ij|ji\rangle$ vanish even faster, because they correspond to the Coulombic interaction of $\phi_i^*(1)\,\phi_j(1)$ with $\phi_j^*(2)\,\phi_i(2)$ and each of these distributions itself vanishes exponentially if the distance AB goes to infinity. Hence, all the mixed terms tend to zero.

Thus,

> the Hartree–Fock method is size-consistent.

We have learned, from the example given at the beginning of this chapter, that the "genetic defect" of mean-field methods is that they describe electrons that ignore the fact that they are close to or far from each other (Fig. 2.1a,b).

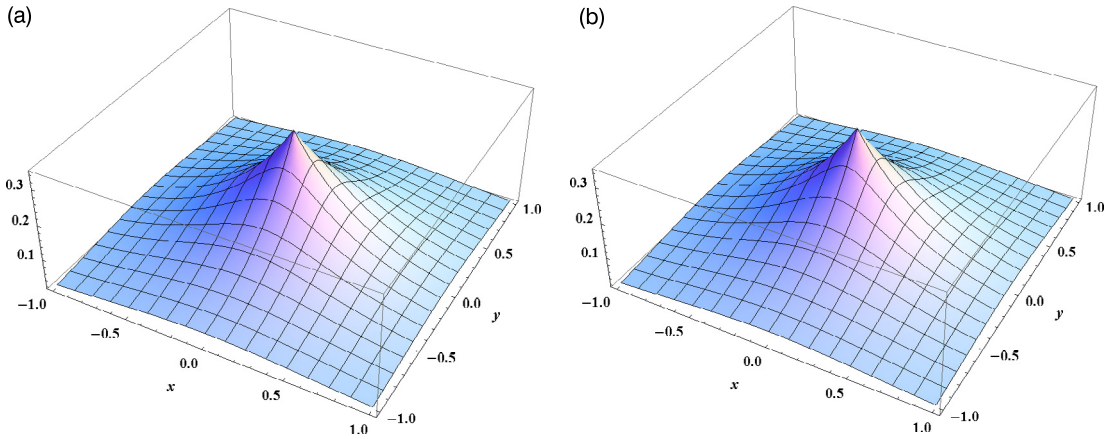


Fig. 2.1. Absence of electronic correlation in the helium atom as seen by the Hartree–Fock method. Visualization of the cross-section of the square of the wave function (probability density distribution) describing electron 2 within the xy plane provided electron 1, the nucleus at $(0, 0, 0)$, is located in a certain point in space. (a) At $(-1, 0, 0)$. (b) At $(1, 0, 0)$. *Note that in both cases the conditional probability density distributions of electron 2 are identical.* This means electron 2 does not react to the motion of electron 1, i.e., there is no correlation whatsoever of the electronic motions (when the total wave function is the Hartree–Fock one).

VARIATIONAL METHODS USING EXPLICITLY CORRELATED WAVE FUNCTIONS

2.1 Correlation cusp condition

The explicitly correlated wave function (we will get to it in a moment) *has the interelectronic distance built in its mathematical form*. We may compare this to making the electrons wear spectacles.[7] Now they avoid each other. One of my students said that it would be the best if the electrons moved apart to infinity. Well, they cannot. They are attracted by the nucleus (energy gain), and being close to it, are necessarily close to each other too (energy loss). There is a compromise to achieve.

Short distances are certainly most important for the Coulombic interaction of two charges, although obviously, the regions of configurational space connected with the long interelectronic

[7] Of course, the methods described further also provide their own "spectacles" (otherwise they would not give the solution of the Schrödinger equation), but the spectacles in the explicitly correlated functions are easier to construct with a small number of parameters.

distances are much larger. Thus the region is not large, but *very* importantly, within it the "collisions" take place. It turns out that the wave function calculated *in the region of collision* must satisfy some very simple mathematical condition (called *correlation cusp* condition). This is what we want to demonstrate. The derived formulae[8] are universal, they apply to any pair of charged particles.

Let us consider *two* particles with charges q_i and q_j and masses m_i and m_j *separated from other particles*. This makes sense since simultaneous collisions of three or more particles occur very rarely in comparison with two-particle collisions. Let us introduce a Cartesian system of coordinates (say, in the middle of the beautiful market square in Brussels), so that the system of two particles is described with six coordinates. Then (atomic units are used) the sum of the kinetic energy operators of the particles is

$$\hat{T} = -\frac{1}{2m_i}\Delta_i - \frac{1}{2m_j}\Delta_j. \tag{2.1}$$

Tosio Kato (1917–1999) was an outstanding Japanese physicist and mathematician. His studies at the University of Tokyo have been interrupted by the Second World War. After the war he obtained his PhD at the same university (about convergence of the perturbational series), and then the title of professor in 1958.
In 1962 Kato became professor at the University of Berkeley, California. He admired the botanic garden over there, knew a lot of Latin names of plants, and very much appreciated the Charles Linnaeus classification of plants.

Now we separate the motion of the center of mass of the two particles with position vectors $\mathbf{r}_i$ and $\mathbf{r}_j$. The center of mass in our coordinate system is indicated by the vector $\mathbf{R}_{CM} = (X_{CM}, Y_{CM}, Z_{CM})$, i.e.,

$$\mathbf{R}_{CM} = \frac{m_i\,\mathbf{r}_i + m_j\,\mathbf{r}_j}{m_i + m_j}. \tag{2.2}$$

Let us also introduce the total mass of the system $M = m_i + m_j$, the reduced mass of the two particles $\mu = \frac{m_i m_j}{m_i + m_j}$, and the vector of their relative positions $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j$. Introducing the

[8] T. Kato, *Commun. Pure Appl. Math.*, 10(1957)151.

three coordinates of the center of mass measured with respect to the market square in Brussels and the x, y, z coordinates which are components of the vector $\mathbf{r}$, we get (Appendix V1-J on p. V1-691, Example 1)

$$\hat{T} = -\frac{1}{2M}\Delta_{\mathbf{CM}} - \frac{1}{2\mu}\Delta, \tag{2.3}$$

$$\Delta_{CM} = \frac{\partial^2}{\partial X_{CM}^2} + \frac{\partial^2}{\partial Y_{CM}^2} + \frac{\partial^2}{\partial Z_{CM}^2}, \tag{2.4}$$

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{2.5}$$

After this operation, the Schrödinger equation for the system is separated (as always in the case of two particles, Appendix V1-J) into two equations: the first describing the motion of the center of mass (seen from Brussels) and the second describing the *relative* motion of the two particles (with Laplacian of x, y, z and reduced mass μ). We are not interested in the first equation, but the second one (Brussels-independent) is what we are after. Let us write down the Hamiltonian corresponding to the second equation,

$$\hat{H} = -\frac{1}{2\mu}\Delta + \frac{q_i q_j}{r}. \tag{2.6}$$

We are interested in how the wave function looks when the distance of the two particles r becomes very small. If r is small, it makes sense to expand the wave function in a power series[9] of r, i.e., $\psi = C_0 + C_1 r + C_2 r^2 + \ldots$. Let us calculate $\hat{H}\psi$ close to $r = 0$. The Laplacian expressed in the spherical coordinates has the form

$$\Delta = \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\sin\theta\frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2},$$

Since we have assumed the function to be dependent on r only, upon the action of the Laplacian only the first term gives a nonzero contribution.

[9] Assuming such a form we exclude the possibility that the wave function goes to $\pm\infty$ for $r \to 0$. This must be so, since otherwise either the respective probability would go to infinity or the operators would become non-Hermitian (cf. p. V1-90). Both possibilities are unacceptable. We covertly assumed also (to simplify our considerations) that the wave function does not depend on the angles θ and ϕ. This dependence can be accounted for by making the constants C_0, C_1, C_2 the functions of θ and ϕ. Then the final results still holds, but for the coefficients C_0 and C_1 averaged over θ and ϕ.

We obtain

$$\hat{H}\psi = \left(-\frac{1}{2\mu}\Delta + \frac{q_i q_j}{r}\right)\psi = \tag{2.7}$$

$$-\frac{1}{2\mu}(\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + ...)(C_0 + C_1 r + C_2 r^2 + ...) + \tag{2.8}$$

$$\frac{q_i q_j}{r}(C_0 + C_1 r + C_2 r^2 + ...) \tag{2.9}$$

$$= 0 - \frac{1}{2\mu}\left(\frac{2C_1}{r} + 6C_2 + 12C_3 r + ...\right) \tag{2.10}$$

$$+ C_0\frac{q_i q_j}{r} + C_1 q_i q_j + C_2 q_i q_j r + ... \tag{2.11}$$

The wave function cannot go to infinity when r goes to zero, while in the above expression we have two terms ($-\frac{1}{2\mu}\frac{2C_1}{r}$ and $C_0\frac{q_i q_j}{r}$) which would then "explode" to infinity.

These two terms must cancel each other.

Hence, we obtain

$$C_0 q_i q_j = \frac{C_1}{\mu}. \tag{2.12}$$

This condition is usually expressed in another way. We use the fact that $\psi(r=0) = C_0$ and $(\frac{\partial\psi}{\partial r})_{r=0} = C_1$ and obtain the cusp condition as

$$(\frac{\partial\psi}{\partial r})_{r=0} = \mu q_i q_j \psi(r=0).$$

- *The case of two electrons*
  In the case of two electrons we have $m_i = m_j = 1$, hence $\mu = \frac{1}{2}$ and $q_i = q_j = -1$. We get the cusp condition for the collision of two electrons as

$$(\frac{\partial\psi}{\partial r})_{r=0} = \frac{1}{2}\psi(r=0)$$

  or (introducing variable $r = r_{12}$ together with particles' position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$)

the wave function should be of the form

$$\psi = \phi(\mathbf{r}_1, \mathbf{r}_2)[1 + \frac{1}{2}r_{12} + \ldots],$$

where $+\ldots$ means higher powers of r_{12}.

- *The nucleus–electron case*
  When one of the particles is a nucleus of charge Z, $\mu \simeq 1$ and we get

$$(\frac{\partial \psi}{\partial r})_{r=0} = -Z\psi(r = 0).$$

Thus

the correct wave function for the electron in the vicinity of a nucleus should have an expansion $\psi = const(1 - Zr_{a1} + \ldots)$, where r_{a1} replacing r is the distance from the nucleus.

Let us see how it is with the $1s$ function for the hydrogen-like atom (the nucleus has charge Z) expanded in a Taylor series in the neighborhood of $r = 0$. We have $1s = N\exp(-Zr) = N(1 - Zr + \ldots)$. It works.

The correlation cusp makes the wave function not differentiable at $r = 0$.

2.2 The Hylleraas CI method

In 1929, two years after the birth of quantum chemistry, a paper by Hylleraas[10] appeared, where, for the ground state of the helium atom, a trial variational function, containing the interelectronic distance explicitly, was applied. This was a brilliant idea, since it showed that already a small number of terms provide very good results. Even though no fundamental difficulties were encountered for larger atoms, the enormous numerical problems were prohibitive for atoms with larger numbers of electrons. In this case, the century-long progress meant going from two- to ten-electron systems. This, however, changed recently.

[10] E.A. Hylleraas, *Zeit. Phys.*, 54(1929)347. Egil Andersen Hylleraas arrived in 1926 in Göttingen, to collaborate with Max Born. His professional experience was related to crystallography and to the optical properties of quartz. When one of the employees fell ill, Born told Hylleraas to continue his work on the helium atom in the context of the newly developed quantum mechanics. The helium atom problem had already been attacked by Albrecht Unsöld in 1927 using first-order perturbation theory, but Unsöld obtained an ionization potential equal to 20.41 eV, while the experimental value was equal to 24.59 eV. In the reported calculations (done on a recently installed calculator) Hylleraas obtained a value of 24.47 eV (cf. contemporary accuracy, p. V1-173).

In the Hylleraas-CI method[11] the Hylleraas idea has been exploited when designing a method for larger systems. The electronic wave function is proposed as a linear combination of Slater determinants, and in front of each determinant $\Phi_i(1, 2, 3, \ldots, N)$ we insert, next to the variational coefficient c_i, correlational factors with some powers $(v, u, ...)$ of the interelectronic distances (r_{mn} between electron m and electron n, etc.). Then we have

$$\psi = \sum_i c_i \hat{A}[r_{mn}^{v_i} r_{kl}^{u_i} ... \Phi_i(1, 2, 3, \ldots, N)], \tag{2.13}$$

where $\hat{A}$ denotes an antisymmetrization operator (see Appendix V1-N, p. V1-707). If $v_i = u_i = 0$, we have the CI expansion, i.e., $\psi = \sum_i c_i \Phi_i$ (we will discuss it on p. 134). If $v_i \neq 0$ or $u_i \neq 0$, we include a variationally proper treatment of the appropriate distances r_{mn} or r_{kl}, i.e., correlation of the motions of the electrons m and n, or k and l, etc. The antisymmetrization operator ensures the symmetry of the wave function with respect to the exchange of any two electrons. The method described was independently proposed in 1971 by Wiesław Woźnicki[12] and by Sims and Hagstrom.[13] The Hylleraas-CI method has a nice feature, in that even a short expansion should give a very good total energy for the system, since we combine the power of the CI method with the great success of the explicitly correlated approaches. Unfortunately, the method has also a serious drawback. To make practical calculations, it is necessary to evaluate the integrals occurring in the variational method, and they were very difficult to calculate. It is enough to realize that, in the matrix element of the Hamiltonian containing two terms of the above expansion, we may find, e.g., a term $1/r_{12}$ (from the Hamiltonian) and r_{13} (from the factor in front of the determinant), as well as the product of six spin orbitals describing electrons 1, 2, 3. Such integrals have to be computed and the existing algorithms were inefficient. A big progress in this field has been made in recent years thanks to the resolution of identity (RI) technique, which will be described later on (p. 177).

2.3 Two-electron systems

2.3.1 Harmonium – the harmonic helium atom

An unpleasant feature of the electron correlation is that we deal either with intuitive concepts or, if our colleagues want to help us, they bring wave functions with formulae as long as the

[11] CI, configuration interaction.

[12] W. Woźnicki, in "*Theory of Electronic Shells in Atoms and Molecules*" (ed. A. Yutsis), Mintis, Vilnius, 1971, p. 103.

[13] J.S. Sims, S.A. Hagstrom, *Phys. Rev. A*, 4(1971)908.

distance from Cracow to Warsaw (or longer[14]) and say: look, this is what *really* happens. It would be good to analyze such formulae term by term, but this does not make sense, because there are too many terms. Even the helium atom, when we write down the formula for its ground-state wave function, becomes a mysterious object. Correlation of motion of whatever seems to be so difficult to grasp mathematically that we easily give up. A group of scientists published a paper in 1993 which can arouse enthusiasm. They obtained a rigorous solution of the Schrödinger equation for harmonium – a helium-like atom with harmonic electron–nucleus attraction (described in Chapter V1-4, p. V1-244). For some particular spring force constant, one gets an exact analytical solution,[15] the only one obtained so far for correlational problems.

Note that the exact wave function (its spatial part[16]) is a *geminal* (i.e., two-electron function), i.e.,

$$\psi\left(\mathbf{r}_1, \mathbf{r}_2\right) = N\left(1 + \frac{1}{2}r_{12}\right) e^{-\frac{1}{4}\left(r_1^2 + r_2^2\right)}. \tag{2.14}$$

Let me be naive. Do we have two harmonic springs here? Yes, we do (see Fig. V1-4.27, p. V1-244). Then, let us treat them first as independent oscillators and take the *product* of the ground-state functions of both oscillators: $\exp[-\frac{1}{4}\left(r_1^2 + r_2^2\right)]$. Well, it would be good to account for the cusp condition $\psi = \phi(\mathbf{r}_1, \mathbf{r}_2)[1 + \frac{1}{2}r_{12} + \ldots]$ and take care of it even in a naive way. Let us just implement the crucial correlation factor $\left(1 + \frac{1}{2}r_{12}\right)$, *the simplest* that satisfies the cusp condition (see p. 92). It turns out that such a recipe leads to a *rigorous* wave function[17]!

From Eq. (2.14) we see that for $r_1 = r_2 = const$ (in such a case both electrons move on the surface of the sphere), the larger value of the function (and *eo ipso* of the probability) is obtained for *larger* r_{12}. This means that it is most probable that the electrons prefer to occupy opposite sides of a nucleus. This is a practical manifestation of the existence of the Coulomb hole around electrons, i.e., the region of the reduced probability of finding a second electron: the electrons simply repel each other. They cannot move apart to infinity since both are held by the nucleus.

---

14 This is a very conservative estimate. Let us calculate – half jokingly. Writing down a single Slater determinant would easily take 10 cm of space. The current world record amounts to several billion such determinants in the CI expansion. Say, three billion. Now let us calculate: $10\text{ cm} \times 3 \times 10^9 = 3 \times 10^{10}\text{ cm} = 3 \times 10^8\text{ m} = 3 \times 10^5\text{ km} = 300\,000\text{ km}$. So, this is not from Warsaw to Cracow, but from Earth to the Moon.

15 S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, G.J. Laming, *J. Chem. Phys.*, 99(1993)417.

16 For one- and two-electron systems the wave function is *a product* of the spatial and spin factors. A normalized spin factor for two-electron systems $\frac{1}{\sqrt{2}}\{\alpha(1)\,\beta(2) - \beta(1)\,\alpha(2)\}$ guarantees that the state in question is a singlet (see Appendix V1-R, p. V1-731). Since we will only manipulate the spatial part of the wave function, the spin part is the default. Since the total wave function has to be antisymmetric, and the spin function is antisymmetric, the spatial function should be symmetric, and it is.

17 As a matter of fact, only for a single force constant. Nevertheless, the striking simplicity of that analytic formula is most surprising.

The only thing they can do is to be close to the nucleus and to avoid each other – this is what we observe in Eq. (2.14).

2.3.2 High accuracy: the James–Coolidge and Kołos–Wolniewicz functions

One-electron problems are the simplest. For systems with *two* electrons[18] we can apply certain mathematical tools which allow very accurate results. We are going to talk about such calculations in a moment.

Kołos and Wolniewicz applied the Ritz variational method (see Chapter V1-5) to the hydrogen molecule with the following trial function:

$$\begin{aligned}\Psi &= \frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2)-\alpha(2)\beta(1)\right]\sum_{i}^{M} c_i\Big(\Phi_i(1,2)+\Phi_i(2,1)\Big),\\ \Phi_i(1,2) &= \exp(-A\xi_1-\bar{A}\xi_2)\xi_1^{n_i}\eta_1^{k_i}\xi_2^{m_i}\eta_2^{l_i}\left(\tfrac{2r_{12}}{R}\right)^{\mu_i}\cdot\\ &\cdot\Big(\exp(B\eta_1+\bar{B}\eta_2)+(-1)^{k_i+l_i}\exp(-B\eta_1-\bar{B}\eta_2)\Big),\end{aligned} \tag{2.15}$$

where the elliptic coordinates of the electrons with index $j = 1, 2$ are given by

$$\xi_j = \frac{r_{aj}+r_{bj}}{R}, \tag{2.16}$$

$$\eta_j = \frac{r_{aj}-r_{bj}}{R}, \tag{2.17}$$

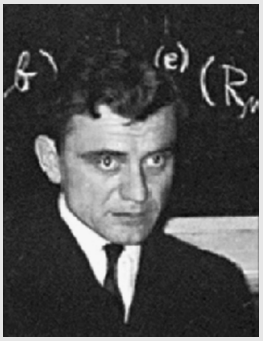
where R denotes the internuclear distance, r_{aj} and r_{bj} are nucleus–electron distances (the nuclei are labeled by a, b), r_{12} is the (crucial to the method) interelectronic distance, c_i, A, $\bar{A}$, B, $\bar{B}$ are variational parameters, and n, k, l, m, μ are integers (smaller than selected limiting values).

The simplified form of this function with $A = \bar{A}$ and $B = \bar{B} = 0$ is the James–Coolidge[19] function, thanks to which the later authors enjoyed the most accurate result for the hydrogen molecule for 27 years.

[18] For a larger number of electrons it is much more difficult.

[19] H.M. James, A.S. Coolidge, *J. Chem. Phys.*, 1(1933)825. Hubert M. James in the 1960s was professor at Purdue University (USA).

Włodzimierz Kołos (1928–1996), Polish chemist, professor at Warsaw University and the tutor of the present author. His calculations on small molecules (with Roothaan, Wolniewicz, Rychlewski) were of unprecedented accuracy in quantum chemistry.
The Department of Chemistry of Warsaw University and the Polish Chemical Society established the Włodzimierz Kołos Medal accompanying a Lecture (the first lecturers were: Roald Hoffmann, Richard Bader, and Paul von Ragué Schleyer). In the Ochota quarter in Warsaw there is a Włodzimierz Kołos Street.



Lutosław Wolniewicz (born 1930), Polish physicist, professor at the Nicolas Copernicus University in Toruń.

Kołos and Roothaan,[20] and later on Kołos and Wolniewicz,[21] as well as Kołos and Rychlewski and others,[22] applied longer and longer expansions (computer technology was improving fast) up to M of the order of thousands (see Table 2.1).

As can be seen from Table 2.1, there was a competition between theoreticians and the experimental laboratory of Herzberg. When, in 1964, Kołos and Wolniewicz obtained 36 117.3 cm^{-1} (Table 2.1, bold face) for the dissociation energy of the hydrogen molecule, quantum chemists held their breath. The experimental result of Herzberg and Monfils, obtained four years earlier (Table 2.1, bold face), was *smaller*, and this seemed to contradict the variational principle (Chapter V1-5, i.e., as if the theoretical result were below the ground-state energy), the foundation of quantum mechanics. There were only three possibilities: either the theoretical or experimental results are wrong or quantum mechanics has internal inconsistency. Kołos and Wolniewicz increased the accuracy of their calculations in 1968 and excluded the first possibility. It soon turned out that the problem lay in the accuracy of the experiment.[23] When Herzberg

20 W. Kołos, C.C.J. Roothaan, *Rev. Modern Phys*., 32(1960)205.

21 For the first time in quantum chemical calculations relativistic corrections and corrections resulting from quantum electrodynamics were included. This accuracy is equivalent to hitting, from Earth, an object on the Moon the size of a car. These results are cited in nearly all textbooks on quantum chemistry to demonstrate that the theoretical calculations have a solid background.

22 The description of these calculations is given in the review article by Piszczatowski et al. cited in the table.

23 At that time Herzberg was hosting them in Canada and treated them to a home-made fruit liquor, the latter event considered by his coworkers to be absolutely exceptional. This is probably the best time to give the recipe for this exquisite drink, which is known in the circles of quantum chemists as "kolosovka."
Pour a pint of pure spirit into a beaker. Hang an orange on a piece of gauze directly over the meniscus. Cover tightly and wait for two weeks. Then throw the orange away – there is nothing of value left in it, and turn your

Table 2.1. Dissociation energy of H_2 in the ground state (in cm^{-1}). Comparison of the results of theoretical calculations and experimental measurements. The references to the cited works can be found in the paper by K. Piszczatowski, G. Łach, M. Przybytek, J. Komasa, K. Pachucki, B. Jeziorski, *J. Chem. Theory and Comput.*, 5(2009)3039. The figures in parentheses mean the error in units of the last digit reported. Bold numbers are used to indicate the values connected with the Herzberg–Kołos–Wolniewicz controversy.

Year	Author	Experiment	Theory
1926	Witmer	35 000	
1927	Heitler–London		23 100[a]
1933	James–Coolidge		36 104[a]
1935	Beutler	36 116(6)	
1960	Kołos–Roothaan		36 113.5[a]
1960	Herzberg–Monfils	**36 113.6**(**3**)	
1964	Kołos–Wolniewicz		**36 117.3**[a]
1968	Kołos–Wolniewicz		**36 117.4**[a]
1970	Herzberg	**36 118.3**[c]	
1970	Stwalley	36 118.6(5)	
1975	Kołos–Wolniewicz		36 118.0
1978	Kołos–Rychlewski		36 118.12[b]
1978	Bishop–Cheung		36 117.92
1983	Wolniewicz		36 118.01
1986	Kołos–Szalewicz–Monkhorst		36 118.088
1991	McCormack–Eyler	36 118.26(20)	
1992	Balakrishnan–Smith– Stoicheff	36 118.11(8)	
1992	Kołos–Rychlewski		36 118.049
1995	Wolniewicz		36 118.069
2009	Piszczatowski et al.		36 118.0695(10)
2009	Liu et al.	36 118.0696(4)	

[a] Obtained from calculated binding energy by subtracting the energy of zero vibrations
[b] Obtained by treating the improvement of the binding energy as an additive correction to the dissociation energy
[c] Upper bound

increased the accuracy, he obtained 36 118.3 cm^{-1} as the dissociation energy (Table 2.1, bold face), which was then consistent with the variational principle.

attention to the spirit. It should contain now all the flavors from the orange, add some sugar to your taste. Next, slowly pour some spring water until the liquid becomes cloudy and some spirit to make it clear again. Propose a toast to the future of quantum chemistry!

Gerhard Herzberg (1904–1999), Canadian chemist of German origin, professor at the National Research Council and at the University of Saskatchewan in Saskatoon and the University of Ottawa. The greatest spectroscopist of the 20th century. Herzberg laid the foundations of molecular spectroscopy, is author of the fundamental monograph on this subject, and received a Nobel prize in 1971 "*for his contribution to knowledge of the electronic structure and geometry of molecules, particularly free radicals*."

The theoretical result of 2009 given in the table includes nonadiabatic, relativistic, and quantum electrodynamic (QED) corrections. The relativistic and QED corrections have been calculated assuming the adiabatic approximation and by taking into account all the terms up to $(\frac{1}{c})^3$ and the leading term in the QED $(\frac{1}{c})^4$ contribution, some effects never taken into account before for any molecule. To get an idea about the importance of the particular levels of theory, let me report their contributions to the H_2 dissociation energy (the number in parentheses means the error in the units of the last digit given). The $(\frac{1}{c})^0$ contribution, i.e., the solution of the Schrödinger equation, gives 36 118.7978(2) cm^{-1}, $(\frac{1}{c})^1$ is equal to zero, $(\frac{1}{c})^2$ is the Breit correction (see p. V1-170) and turned out to be $-0.5319(3)$ cm^{-1}, the QED (see p. V1-175) $(\frac{1}{c})^3$ correction is $-0.1948(2)$ cm^{-1}, while the $(\frac{1}{c})^4$ contribution is $-0.0016(8)$ cm^{-1}. We see that to obtain such agreement with the experimental value as shown in Table 2.1, one needs to include all the abovementioned corrections.

2.3.3 High accuracy: neutrino mass

Calculations like those above required unique software, especially in the context of the nonadiabatic effects included. Additional gains appeared unexpectedly when Kołos and others[24] initiated work aiming at explaining whether the electronic neutrino has a nonzero

[24] W. Kołos, B. Jeziorski, H.J. Monkhorst, K. Szalewicz, *Int. J. Quantum Chem. S*, 19(1986)421.

mass.[25] In order to interpret the expensive experiments, precise calculations were required for the β-decay of the tritium molecule as a function of the neutrino mass. The emission of the antineutrino (ν) in the process of β-decay,

$$T_2 \rightarrow HeT^+ + e + \nu,$$

should have consequences for the final quantum states of the HeT^+ molecule. To enable determination of the neutrino mass by the experimentalists, Kołos et al. performed precise calculations of all possible final states of HeT^+ and presented them as a function of the hypothetical mass of the neutrino. There is such a large number of neutrinos in the Universe that if its mass exceeded a certain, even very small threshold value of the order of[26] 1 eV, the mass of the Universe would exceed the critical value predicted by Alexander Friedmann in his cosmological theory (based on the general theory of relativity of Einstein). This would mean that the currently occurring expansion of the Universe (discovered by Hubble) would finally stop and its collapse would follow. If the neutrino mass would turn out to be too small, then the Universe would continue its expansion. Thus, quantum chemical calculations for the HeT^+ molecule might be helpful in predicting our fate (unfortunately, being crushed or frozen). So far, the estimate of the neutrino mass gives a value smaller than 1 eV, which indicates Universe expansion.[27]

[25] Neutrinos are stable fermions of spin $\frac{1}{2}$. Three types of neutrinos exist (each has its own antiparticle): electronic, muonic, and taonic. The neutrinos are created in the weak interactions (e.g., in β-decay) and do not participate neither in the strong, nor in electromagnetic interactions. The latter feature expresses itself in an incredible ability to penetrate matter (e.g., crossing the Earth as though through a vacuum). The existence of the electronic neutrino was postulated in 1930 by Wolfgang Pauli and discovered in 1956 by F. Reines and C.L. Cowan; the muonic neutrino was discovered in 1962 by L. Lederman, M. Schwartz, and J. Steinberger.

[26] The mass of the elementary particle is given in the form of its energetic equivalent mc^2.

[27] At this moment there are other candidates for contributing significantly to the mass of the Universe, mainly the mysterious "dark matter." This constitutes the major part of the mass of the Universe. We know very little about it.

Recently it turned out that neutrinos undergo what are called oscillations, e.g., an electronic neutrino travels from the Sun and on its way spontaneously changes to a muonic neutrino. The oscillations indicate that the mass of the neutrino is nonzero. According to current estimations, it is much smaller, however, than the accuracy of the tritium experiments.

Alexandr Alexandrovitch Friedmann (1888–1925), Russian mathematician and physicist, in his article in *Zeit. Phys*., 10(1922)377 proved on the basis of Einstein's general theory of relativity that the curvature of the Universe must change, which became the basis of cosmological models of the expanding Universe. During the First World War, Friedman was a pilot in the Russian army and made bombing raids over my beloved Przemyśl.

In one of his letters he asked his friend cheerfully, the eminent Russian mathematician Steklov, for advice about the integration of equations he derived to describe the trajectories of his bombs. Later, in a letter to Steklov of February 28, 1915 he wrote: "*Recently I had an opportunity to verify my theory during a flight over Przemyśl, the bombs fell*

exactly in the places predicted by the theory. To get the final proof of my theory I intend to test it in flights during the next few days."

More information can be found on http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Friedmann.html.

2.4 Exponentially correlated Gaussian functions

In 1960, Boys[28] and Singer[29] noticed that the functions which are products of Gaussian orbitals and correlational factors of Gaussian type, $\exp(-br_{ij}^2)$, where r_{ij} is the distance between electron i and electron j, generate relatively simple integrals in quantum chemical calculations. A product of two Gaussian orbitals with positions shown by the vectors $\mathbf{A}, \mathbf{B}$ and of an exponential correlation factor is called *an exponentially correlated Gaussian geminal*,[30] i.e.,

$$g\left(\mathbf{r}_i, \mathbf{r}_j; \mathbf{A}, \mathbf{B}, a_1, a_2, b\right) = Ne^{-a_1(\mathbf{r}_i-\mathbf{A})^2} e^{-a_2(\mathbf{r}_j-\mathbf{B})^2} e^{-br_{ij}^2}.$$

A geminal represents a generalization of an orbital – there we have a one-electron function, here we have a two-electron one. A single geminal is very rarely used in computations[31]; we apply hundreds or even thousands of Gaussian geminals. When we want to find out what the optimal positions $\mathbf{A}, \mathbf{B}$ and the optimal exponents a and b are in these thousands of geminals, it

28 S.F. Boys, *Proc. Royal Soc. A*, 258(1960)402.

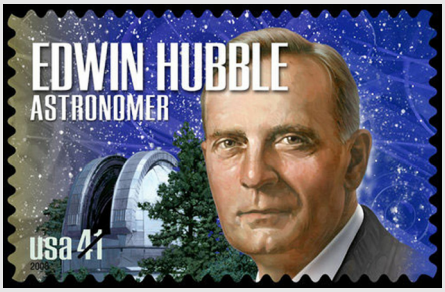
29 K. Singer, *Proc. Royal Soc. A*, 258(1960)412.

30 This is an attempt to go beyond the two-electron systems with the characteristic for these systems approach of James, Coolidge, Hylleraas, Kołos, Wolniewicz, and others.

31 Ludwik Adamowicz introduced an idea of the minimal basis of the Gaussian geminals (equal to the number of the electron pairs) and applied to the LiH and HF molecules; L. Adamowicz, A.J. Sadlej, *J. Chem. Phys.*, 69(1978)3992.

turns out that nothing sure is known about them; the **A**, **B** positions are scattered chaotically,[32] and in the $a > 0$ and $b > 0$ exponents there is no regularity either.

Edwin Powell Hubble (1889–1953), American astronomer, explorer of galaxies, found in 1929 that the distance between galaxies is proportional to the infrared shift in their spectrum caused by the Doppler effect, which is consequently interpreted as expansion of the Universe. A surprise from recent astronomical studies is that the expansion is faster and faster (for unknown reasons).



2.5 Electron holes

2.5.1 Coulomb hole ("correlation hole")

It is always good to count "on fingers" to make sure that everything is all right. Let us see how a *single* Gaussian geminal describes the correlation of the electronic motion. Let us begin with the helium atom with the nucleus in the position $\mathbf{A} = \mathbf{B} = \mathbf{0}$. The geminal takes the form

$$g_{He} = Ne^{-a_1 r_1^2} e^{-a_1 r_2^2} e^{-b r_{12}^2}, \tag{2.18}$$

where $N > 0$ is a normalization factor. Let us assume[33] that electron 1 is at $(x_1, y_1, z_1) = (1, 0, 0)$. Where does electron 2 prefer to be in such situation? We will find it out (Fig. 2.2) from the position of electron 2 for which g_{He} assumes the largest value.

Just to get an idea, let us try to restrict the motion of electron 2. For instance, let us demand that it moves only on the sphere of radius equal to 1 centered at the nucleus, so we insert $r_1 = r_2 = 1$. Then, $g_{He} = const \exp\left[-b r_{12}^2\right]$ and we will find out easily what electron 2 likes most. With $b > 0$ the latter factor tells us that what electron 2 likes best is just to sit on electron 1! Is it what the correlation is supposed to mean that one electron sits on the other? Here we have rather an anticorrelation. Something is going wrong. According to this analysis we should rather take the geminal of, e.g., the form

$$g_{He} = Ne^{-a_1 r_1^2} e^{-a_1 r_2^2} \left[1 - e^{-b r_{12}^2}\right].$$

[32] The methods in which those positions are selected at random scored great success.

[33] We use atomic units.

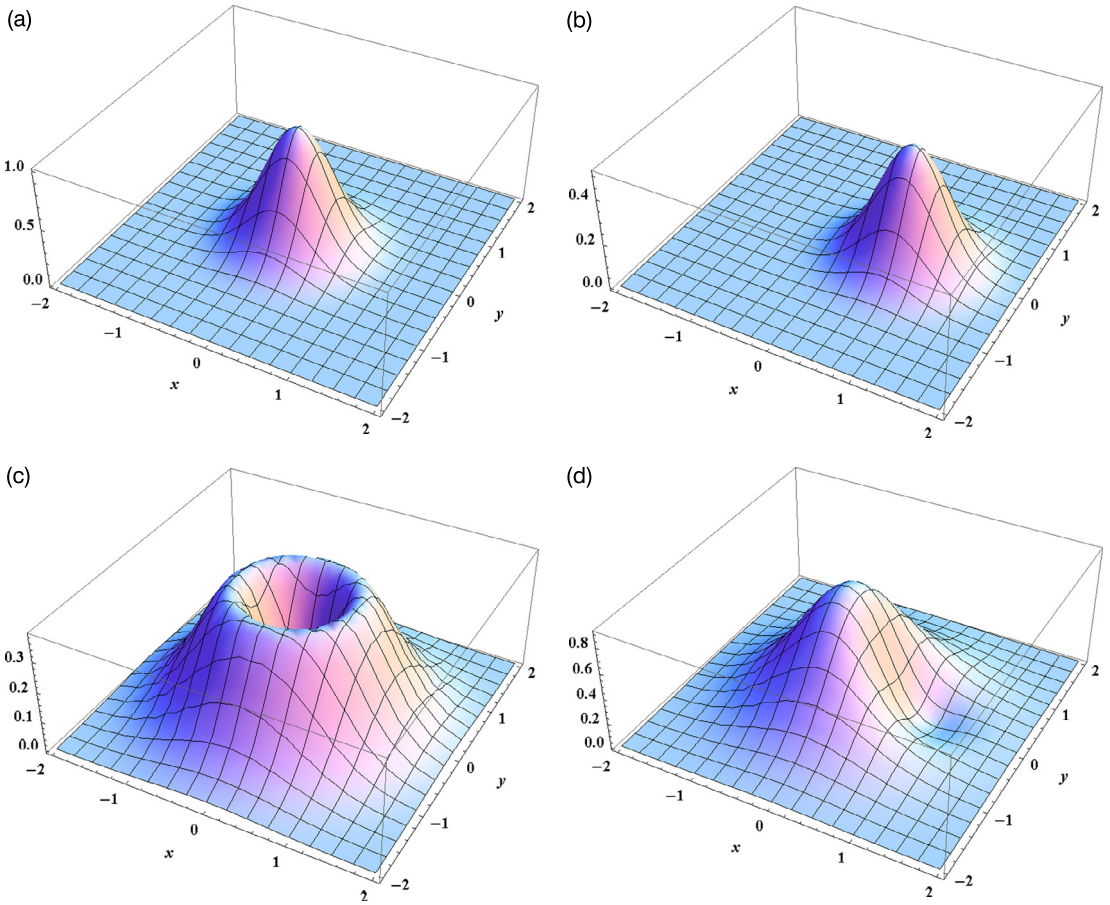


Fig. 2.2. Illustration of the correlation and anticorrelation of the electrons in a model helium atom. The machinery of the "anticorrelation" connected with the (not normalized) geminal $g_{He} = \exp\left[-r_1^2\right]\exp\left[-r_2^2\right]\exp\left[-2r_{12}^2\right]$. (a) Electron 1 has position $(0,0,0)$. (b) Electron 1 is at point $(1,0,0)$. It can be seen that electron 2 *holds on to electron 1*, i.e., it behaves in a completely unphysical manner (since the electrons repel each other). (c,d) Electron 2 responds to such two positions of electron 1 in case the wave function represents the geminal $g_{He} = \exp\left[-r_1^2\right]\exp\left[-r_2^2\right]\left[1-\exp\left[-2r_{12}^2\right]\right]$. (c) Electron 2 runs away (the hollow in the middle) from electron 1 placed at $(0,0,0)$, and we have correlation. (d) Similarly, if electron 1 is in point $(1,0,0)$, then it causes a slight depression for electron 2 in this position, and we have correlation. However, the graph is different from the one in (c). This is understandable since the nucleus is all the time in the point $(0,0,0)$. (e,f) The same displacements of electron 1, but this time the wave function is equal to $\psi(\mathbf{r}_1,\mathbf{r}_2) = \left(1+\frac{1}{2}r_{12}\right)\exp[-\left(r_1^2+r_2^2\right)]$, i.e., is similar to the wave function of the harmonic helium atom. It can be seen (particularly in (e)) that there *is* a correlation, although much less visible than in the previous examples. (g,h) The same as in (e,f), but (to amplify [artificially] the correlation effect) for the function $\psi(\mathbf{r}_1,\mathbf{r}_2) = (1+25r_{12})\exp[-\left(r_1^2+r_2^2\right)]$, which (unlike (e,f)) does not satisfy the correlation cusp condition.

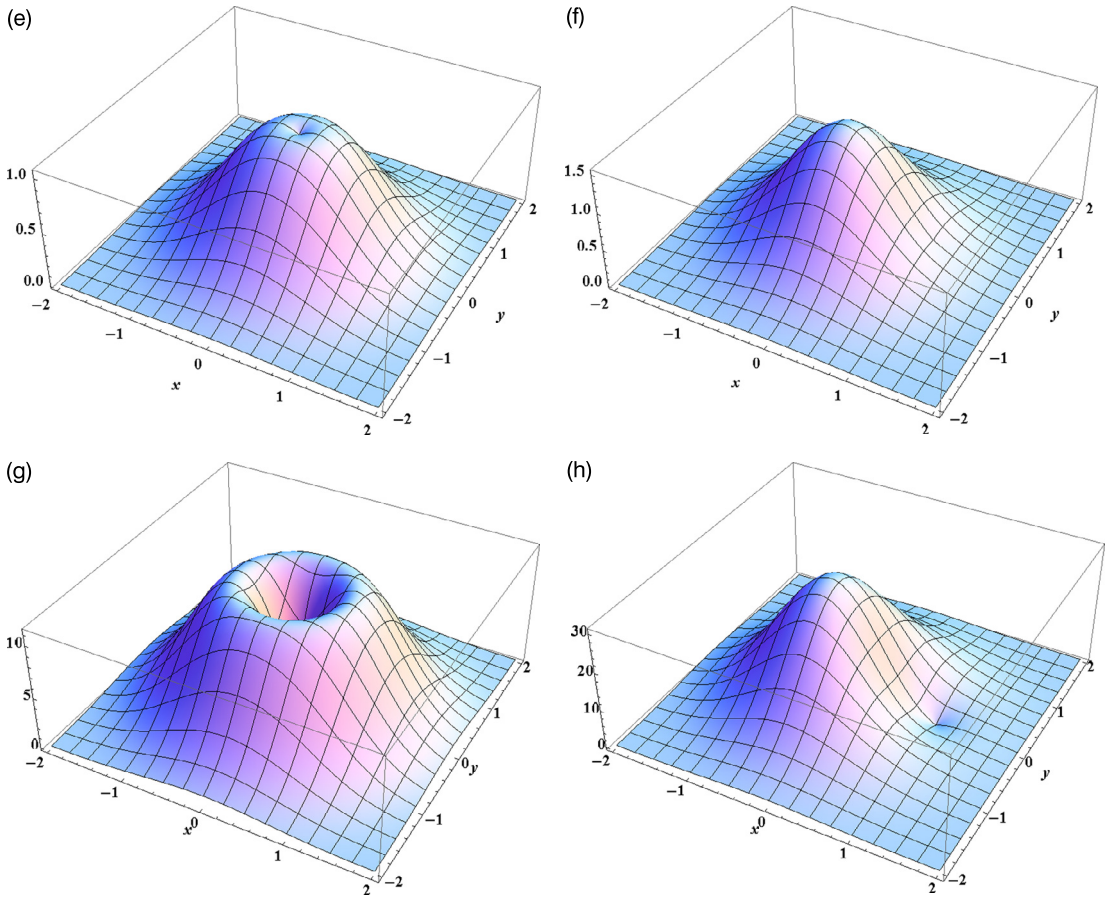


Fig. 2.2. (*continued*)

Now everything is qualitatively in order. When the interelectronic distance increases, the value of the g_{He} function also increases, which means that such a situation is *more* probable than that corresponding to a short distance. If the electrons become too agitated and begin to think that it would be better when their distance gets very large, they would be called to order by the factors $\exp\left[-a_1 r_1^2\right]\exp\left[-a_1 r_2^2\right]$. Indeed, in such a case, the distance between the nucleus and at least one of the electrons is long and the probability of such a situation is quenched by one or both exponential factors. For large r_{12} distances, the factor $\left[1-\exp\left[-br_{12}^2\right]\right]$ is practically equal to 1. This means that for large interelectronic distances g_{He} is practically equal to $N\exp\left[-a_1 r_1^2\right]\exp\left[-a_1 r_2^2\right]$, i.e., to the product of the orbitals (no correlation of motions at long interelectronic distances and rightly so).

Around electron 1 there is a region of low probability of finding electron 2. This region is called the Coulomb hole.

The Gaussian geminals do not satisfy the correlation cusp condition (p. 92), because of the factor $\exp(-br_{ij}^2)$. It is required (for simplicity we write $r_{ij} = r$) that $(\frac{\partial g}{\partial r})_{r=0} = \frac{1}{2}g(r=0)$, whereas the left-hand side is equal to 0, while the right-hand side $\frac{1}{2}N \exp\left[-a_1 (\mathbf{r}_i - \mathbf{A})^2\right] \exp\left[-a_2 \left(\mathbf{r}_j - \mathbf{B}\right)^2\right]$ is not equal to zero. This is not a disqualifying feature, since the region of space in which this condition should be fulfilled is very small.

The method of Gaussian geminals has been applied in extremely accurate calculations but for three- and four-electron systems only.[34]

2.5.2 Exchange hole ("Fermi hole")

The mutual avoidance of electrons in a helium atom or in a hydrogen molecule is caused by Coulombic repulsion of electrons ("Coulomb hole," see above). As we have shown in this chapter, in the Hartree–Fock method the Coulomb hole is absent, whereas methods which account for electron correlation generate such a hole. However, electrons avoid each other not only because of their charge. The Pauli principle is additional reason. One of the consequences is the fact that electrons with the same spin coordinate cannot reside in the same place (see p. V1-40). The continuity of the wave function implies that the probability density of them staying *in the vicinity* of each other is small, i.e.,

around the electron there is a NO PARKING area for other electrons with the same spin coordinate ("exchange hole" or "Fermi hole").

Let us see how such exchange holes arise. We will try to make the calculations as simple as possible.

We have shown above that the Hartree–Fock function does not include any electron correlation. We must admit, however, that we have come to this conclusion on the basis of the two-electron closed shell case. This is a special situation, since both electrons have *different* spin coordinates ($\sigma = \frac{1}{2}$ and $\sigma = -\frac{1}{2}$). Is it really true that the Hartree–Fock function does not include any correlation of electronic motion?

[34] W. Cencek, PhD Thesis, Adam Mickiewicz University, Poznań, 1993; also J. Rychlewski, W. Cencek, J. Komasa, *Chem. Phys. Letters*, 229(1994)657; W. Cencek, J. Rychlewski, *Chem. Phys. Letters*, 320(2000)549.

We take the H_2^- molecule in the simplest formulation of the LCAO MO method.[35] We have three electrons. As a wave function we will take the single (normalized) Hartree–Fock determinant (UHF, or rather ROHF type) with the following orthonormal spin orbitals occupied: $\phi_1 = \varphi_1\alpha$, $\phi_2 = \varphi_1\beta$, $\phi_3 = \varphi_2\alpha$, i.e.,

$$\psi_{UHF}(1,2,3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \phi_1(3) \\ \phi_2(1) & \phi_2(2) & \phi_2(3) \\ \phi_3(1) & \phi_3(2) & \phi_3(3) \end{vmatrix}.$$

Example 1 (The great escape, Fig. 2.3a). We are interested in electron 3, with electron 1 residing at nucleus a with space coordinates $(0,0,0)$ and spin coordinate $\sigma_1 = \frac{1}{2}$ and with electron 2 located at nucleus b with coordinates $(R,0,0)$ and $\sigma_2 = -\frac{1}{2}$, whereas electron 3 itself has spin coordinate $\sigma_3 = \frac{1}{2}$. The square of the absolute value of the function calculated for these values depends on x_3, y_3, z_3 and represents the *conditional probability* density distribution for finding electron 3 (provided electrons 1 and 2 have the fixed coordinates given above and denoted by $1_0, 2_0$). So, let us calculate individual elements of the determinant $\psi_{UHF}(1_0, 2_0, 3)$, taking into account the properties of spin functions α and β (cf. p. 33). We have

$$\psi_{UHF}(1_0,2_0,3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \varphi_1(0,0,0) & 0 & \varphi_1(x_3,y_3,z_3) \\ 0 & \varphi_1(R,0,0) & 0 \\ \varphi_2(0,0,0) & 0 & \varphi_2(x_3,y_3,z_3) \end{vmatrix}.$$

Using the Laplace expansion (Appendix V1-A on p. V1-589) we get

$$\psi_{UHF}(1_0,2_0,3) = \frac{1}{\sqrt{3!}}[\varphi_1(0,0,0)\,\varphi_1(R,0,0)\,\varphi_2(x_3,y_3,z_3) - \varphi_1(x_3,y_3,z_3)\,\varphi_1(R,0,0)\,\varphi_2(0,0,0)].$$

The plot of this function (the overlap integral S is included in normalization factors of the molecular orbitals) is given in Fig. 2.3a.

Qualitatively, however, everything is clear even without the calculations. Due to the forms of the molecular orbitals (S is small) $\varphi_1(0,0,0) = \varphi_1(R,0,0) \approx \varphi_2(0,0,0) = \text{const}$, we get

$$\psi_{UHF}(1_0,2_0,3) \approx -\text{const}^2 \frac{1}{\sqrt{3}} \chi_b(3),$$

[35] Two atomic orbitals only: $1s_a = \chi_a$ and $1s_b = \chi_b$, two molecular orbitals: bonding $\varphi_1 = \frac{1}{\sqrt{2(1+S)}}(\chi_a + \chi_b)$ and antibonding $\varphi_2 = \frac{1}{\sqrt{2(1-S)}}(\chi_a - \chi_b)$ (cf. p. V1-511); the overlap integral $S \equiv (\chi_a|\chi_b)$.

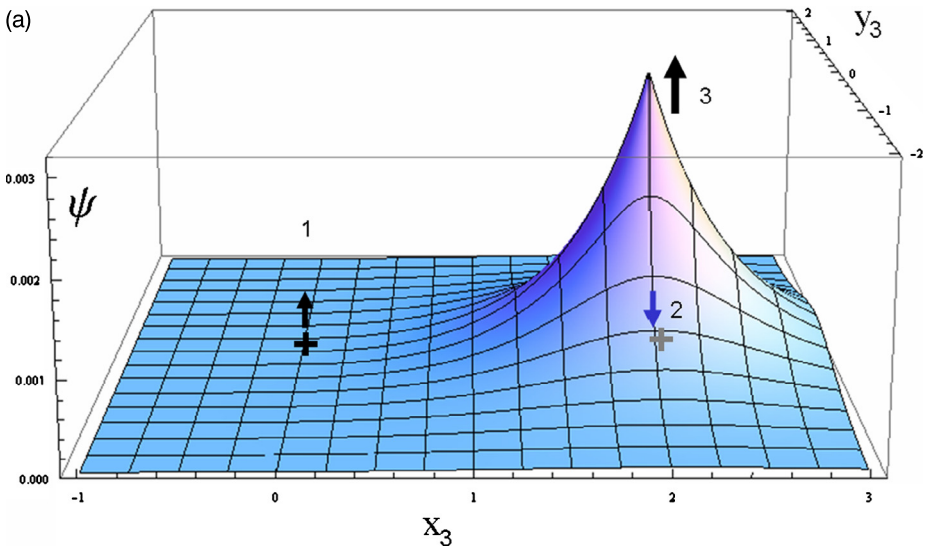


Fig. 2.3. Demonstration of the power of the Pauli exclusion principle, or the Fermi hole formation for the H_2^- molecule in the restricted open shell Hartree–Fock model (ROHF, a wave function in the form of a single Slater determinant). The two protons (a and b) indicated by "+" occupy positions $(0, 0, 0)$ and $(2, 0, 0)$, respectively, in a.u. The space and spin coordinates (the latter shown as arrows) of electrons 1 and 2, as well as the spin coordinate of electron 3 ($\sigma_3 = \frac{1}{2}$), will be fixed at certain values: electron 2 will always sit on nucleus b, electron 1 will occupy some chosen positions on the x axis (i.e., we keep $y_1 = 0, z_1 = 0$). In this way we will have to do with a section $\psi\,(x_3, y_3, z_3)$ of the wave function, visualized in the figure by setting additionally $z_3 = 0$. The square of the resulting function represents a conditional probability density of finding electron 3, if electrons 1 and 2 have the assigned coordinates. (a) Example 1: electron 1 sits on nucleus a. Electron 3 runs away to nucleus b, despite the fact that electron 2 is already there! (b) Example 2: electron 1 sits on nucleus b together with electron 2. Electron 3 runs away to nucleus a. (c) Example 3 – a dilemma for electron 3: electron 1 sits in the middle between the nuclei. Electron 3 chooses the antibonding molecular orbital (c1), because it offers a node exactly at the position of electron 1 (with same spin), thus creating there a Fermi hole (c2)! (d1) An even tougher case: electron 1 sits at $\frac{1}{3}$ of the internuclear distance; what is electron 3 going to do? Electron 3 chooses such a combination of the bonding and antibonding molecular orbitals that creates a node (and a Fermi hole (d2)) precisely at the position of electron 1 with the same spin! Clearly, with a single Slater determinant as the wave function, electrons with the same spin hate one another (Fermi hole), while electrons with the opposite spin just ignore each other (no Coulomb hole).

so the conditional probability of finding electron 3 is

$$\rho\,(3) \approx \frac{1}{3}\mathrm{const}^4\left[\chi_b\,(3)\right]^2. \tag{2.19}$$

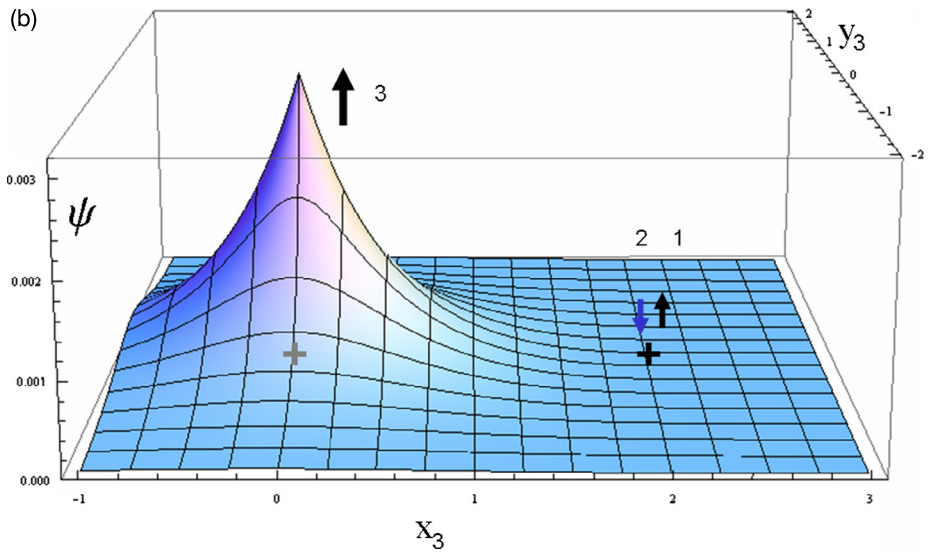


Fig. 2.3. (*continued*)

We can see that *for some reason electron 3 has moved in the vicinity of nucleus b*. What scared it so much, when we placed one of the two electrons at *each* nucleus? Electron 3 ran to be as far away as possible from electron 1 residing on a. It hates electron 1 so much that it has just ignored the Coulomb repulsion with electron 2 sitting on b and jumped on it![36] What the hell has happened? Well, we have some suspicions. Electron 3 could have been scared only by the spin coordinate of electron 1, *the same as its own*.

This is just an indication of the exchange hole around each electron.

Example 2 (Another great escape, Fig. 2.3b). Maybe electron 3 does not run away from anything, but simply always resides at nucleus b? Let us make sure of that. Let us move electron 1 to nucleus b (electron 2 already sits there, but that does not matter). What then will electron 3 do? Let us see. We have electrons 1 and 2 at nucleus b with space coordinates $(R, 0, 0)$ and spin coordinates $\sigma_1 = \frac{1}{2}$, $\sigma_2 = -\frac{1}{2}$, whereas electron 3 has spin coordinate $\sigma_3 = \frac{1}{2}$. To calculate the conditional probability we have to calculate the value of the wave function.

This time

$$\psi_{UHF}(1_0, 2_0, 3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \varphi_1(R,0,0) & 0 & \varphi_1(x_3, y_3, z_3) \\ 0 & \varphi_1(R,0,0) & 0 \\ \varphi_2(R,0,0) & 0 & \varphi_2(x_3, y_3, z_3) \end{vmatrix} = \text{const}^2 \frac{1}{\sqrt{3}} \chi_a(3)$$

[36] In fact it does not even see electron 2 (because of the one-determinantal wave function).

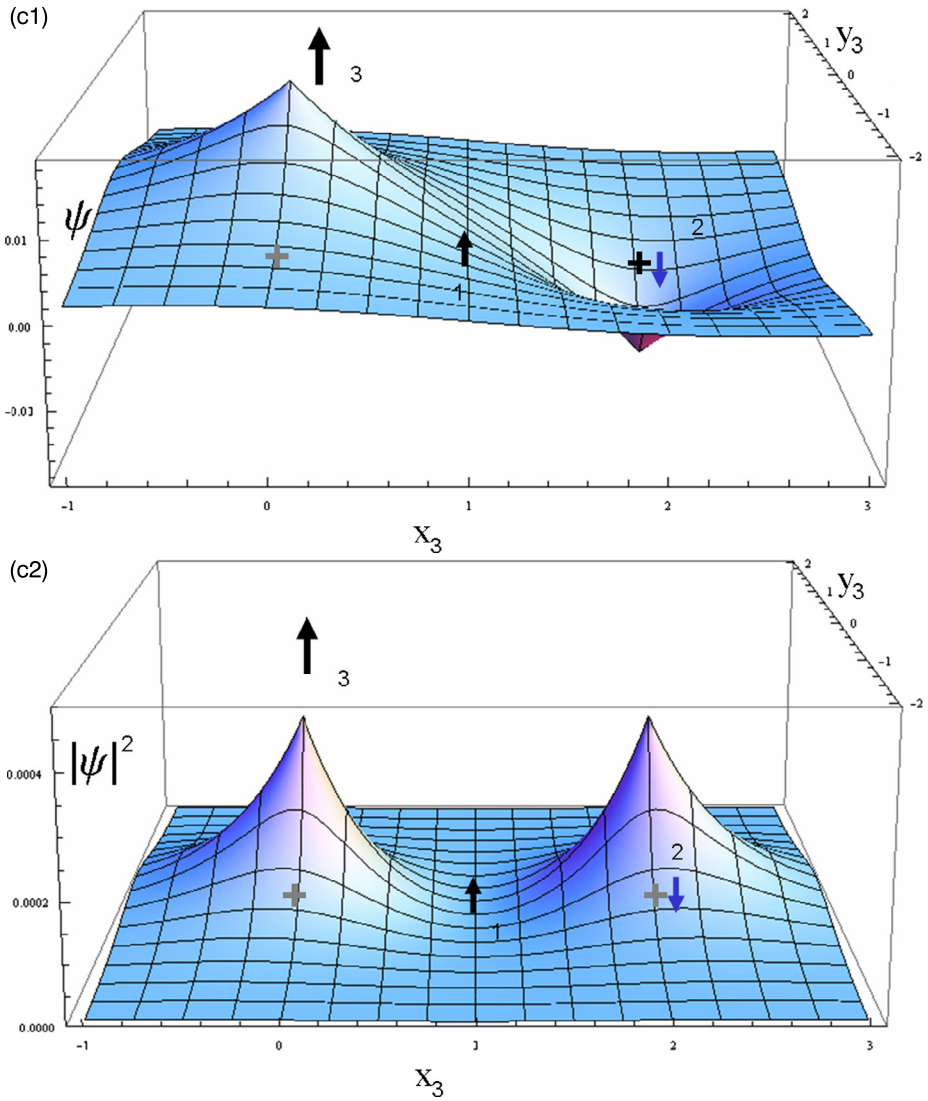


Fig. 2.3. (*continued*)

or

$$\rho\,(3) \approx \frac{1}{3}\text{const}^4\left[\chi_a\,(3)\right]^2. \tag{2.20}$$

We see that electron 3 with spin coordinate $\sigma_3 = \frac{1}{2}$ runs in panic to nucleus a, because it is as scared of electron 1 with spin $\sigma_1 = \frac{1}{2}$ as the devil is of holy water.

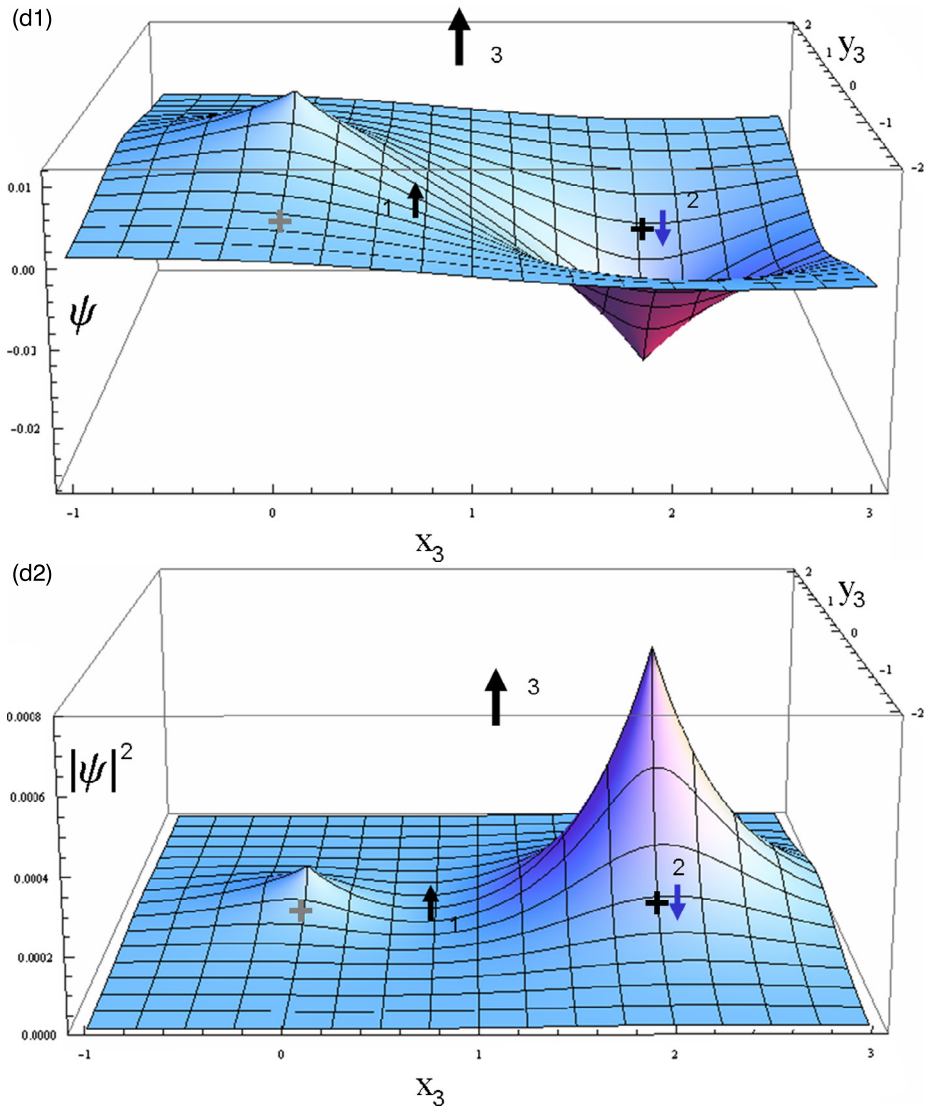


Fig. 2.3. (*continued*)

Example 3 (A dilemma, Fig. 2.3c). And what would happen if we made the decision for electron 3 more difficult? Let us put electron 1 ($\sigma_1 = \frac{1}{2}$) *in the center* of the molecule and electron 2 ($\sigma_2 = -\frac{1}{2}$) as before, at nucleus b. According to what we think about the whole machinery, electron 3 (with $\sigma_3 = \frac{1}{2}$) should run away from electron 1, because both electrons have the same spin coordinates, and this is what they hate most. But *where* should it run? Will electron 3 select nucleus a or nucleus b? The nuclei do not look equivalent. There is

electron sitting at b, while the a center is empty. Maybe electron 3 will jump to a then? Well, the function analyzed is Hartree–Fock – electron 3 ignores the Coulomb hole (it does not see electron 2 sitting on b) and therefore will not prefer the empty nucleus a to sit at. It looks like electron 3 will treat both nuclei on the same basis. In the case of two atomic orbitals, electron 3 has to choose: either bonding orbital φ_1 or antibonding orbital φ_2 (either of these situations corresponds to equal electron densities on a and on b). Out of the two molecular orbitals, φ_2 looks much more attractive to electron 3, because it has a node[37] exactly where electron 1 with its nasty spin is. This means that there is a chance for electron 3 to take care of the Fermi hole of electron 1: we predict that electron 3 will "select" only φ_2. Let us check this step by step. We have

$$\psi_{UHF}(1_0, 2_0, 3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \varphi_1\left(\frac{R}{2}, 0, 0\right) & 0 & \varphi_1(x_3, y_3, z_3) \\ 0 & \varphi_1(R, 0, 0) & 0 \\ \varphi_2\left(\frac{R}{2}, 0, 0\right) & 0 & \varphi_2(x_3, y_3, z_3) \end{vmatrix} =$$

$$\frac{1}{\sqrt{3!}} \begin{vmatrix} \varphi_1\left(\frac{R}{2}, 0, 0\right) & 0 & \varphi_1(x_3, y_3, z_3) \\ 0 & \varphi_1(R, 0, 0) & 0 \\ 0 & 0 & \varphi_2(x_3, y_3, z_3) \end{vmatrix} =$$

$$\frac{1}{\sqrt{3!}} \varphi_1\left(\frac{R}{2}, 0, 0\right) \varphi_1(R, 0, 0) \varphi_2(x_3, y_3, z_3) = \text{const}_1 \varphi_2(x_3, y_3, z_3).$$

And it does exactly so.

In (d1) we give also an example with electron 1 at $\frac{1}{3}R$. The result is similar: a Fermi hole over there (Fig. 2.3d).

Which hole is more important: Coulomb or exchange? This question will be answered in Chapter 3.

VARIATIONAL METHODS WITH SLATER DETERMINANTS

In all of these methods the variational wave function will be sought in the form of a linear combination of Slater determinants. As we have seen a while ago even a single Slater determinant ensures a very serious avoiding of electrons with the same spin coordinate. Using a linear combination of Slater determinants means an automatic (based on variational principle) optimization of the exchange hole (Fermi hole).

[37] That is, low probability of finding electron 3 over there.

What about the Coulomb hole? If also this hole were optimized, a way to the solution of the Schrödinger equation would be open. However, as we have carefully checked before, a single Slater determinant does not know anything about the Coulomb hole. If it does not know, then probably a linear combination of guys, each of them not knowing anything, will not do any better... Wrong! A linear combination of Slater determinants *is* able to describe the Coulomb hole![38]

2.6 Static electron correlation

Some of these Slater determinants are necessary for fundamental reasons. For example, consider the carbon atom ground state, its (triplet) ground state corresponding to the $1s^2 2s^2 2p^2$ configuration. The configuration does not define *which* of the triply *degenerate* $2p$ orbitals have to be included in the Slater determinant. Any choice of the $2p$ orbitals will therefore be nonsatisfactory: one is forced to go beyond a single Slater determinant. A similar situation occurs if an obvious *quasidegeneracy* occurs, like, e.g., for the hydrogen molecule at large distances (see Chapter V1-8). In such a case we are also forced to include in calculations another Slater determinant. One may say that

> what is known as a static correlation represents an energy gain coming from considering in the wave function (in the form of a linear combination of Slater determinants) low-energy Slater determinants, which follow from occupying a set of degenerate or quasidegenerate orbitals.

2.7 Dynamic electron correlation

The dynamic electron correlation means the rest of the correlation effect, beyond the static one. It corresponds also to occupying orbital energies, but not those related to the degeneracy or quasidegeneracy of the ground state. As we see the distinction between the static and the dynamic correlation is a bit arbitrary.

Example of beryllium

Let us take a beryllium atom. The beryllium atom has four electrons ($1s^2 2s^2$ configuration). Beryllium represents a tough case in quantum chemistry, because the formally occupied $2s$

[38] Not all linear combinations of Slater determinants describe the Coulomb hole. Indeed, e.g., a Hartree–Fock function in the LCAO MO approximation may be expanded in a series of Slater determinants (Appendix V1-A) with the atomic orbitals, but no Coulomb hole is described by this function.

orbital energy is quite close to the formally unoccupied orbital energy of $2p$. In the present example we will claim this as a dynamic correlation, but to tell the truth it is just between the static and dynamic correlation. One may therefore suspect that the excited configurations $2s^1 2p^1$ and $2p^2$ will be close in energy scale to the ground-state configuration $2s^2$. There is therefore no legitimate argument for neglecting these excited configurations in the wave function (what the Hartree–Fock method does). Since the Hartree–Fock method is poor in this case, this means the electronic correlation energy must be large for the beryllium atom.[39]

Why worry then about the closed shell electrons $1s^2$? Two of the electrons are bound very strongly ($1s^2$), so strongly that we may treat them as passive observers that do not react to anything that may happen. Let us just ignore the inner shell[40] in such a way that we imagine an "effective nucleus of the pseudoatom" of beryllium as a genuine beryllium nucleus surrounded by the electronic cloud $1s^2$. The charge of this "nucleus" is $4 - 2 = 2$. Then the ground-state Slater determinant for such a pseudoatom reads as

$$\psi_0 = \frac{1}{\sqrt{2!}} \begin{vmatrix} 2s(1)\alpha(1) & 2s(2)\alpha(2) \\ 2s(1)\beta(1) & 2s(2)\beta(2) \end{vmatrix}, \tag{2.21}$$

where we decide to approximate the function $2s$ as a normalized Slater orbital[41] ($\zeta > 0$)

$$2s = \sqrt{\frac{\zeta^5}{3\pi}} r \exp(-\zeta r).$$

Since the Hartree–Fock method looks as a poor tool for beryllium, we propose a more reasonable wave function in the form of a linear combination of the ground-state configuration (2.21) and the configuration given by the following Slater determinant:

$$\psi_1 = \frac{1}{\sqrt{2!}} \begin{vmatrix} 2p_x(1)\alpha(1) & 2p_x(2)\alpha(2) \\ 2p_x(1)\beta(1) & 2p_x(2)\beta(2) \end{vmatrix}, \tag{2.22}$$

where just to keep things as simple as possible we use the $2p_x$ orbital.

Such a function, being a linear combination of antisymmetric functions, is itself antisymmetric with respect to the electron exchange (as it should be, see Chapter V1-1). Just to grasp the essence of the problem we omit all other excitations, including $2s^2 \to 2p^2$ with the orbitals

[39] This is why we took the beryllium atom and not just the helium atom, in which the energy difference between the orbital levels $1s$ and $2s$ is much larger, i.e., the correlation energy much smaller.

[40] The reasoning below may be repeated with the $1s^2$ shell included; the calculations will be a bit more complicated, but the final result very similar.

[41] Let us check whether the normalization coefficient is correct: $\int (2s)^2 \mathrm{d}V = \frac{\zeta^5}{3\pi} \int r^2 \exp(-2\zeta r)\mathrm{d}V = \frac{\zeta^5}{3\pi} \int_0^\infty r^4 \exp(-2\zeta r)\mathrm{d}r \int_0^\pi \sin\theta \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi = \frac{4\pi\zeta^5}{3\pi} \int_0^\infty r^4 \exp(-2\zeta r)\mathrm{d}r = \frac{4\pi\zeta^5}{3\pi} 4!(2\zeta)^{-5} = 1$, as it should be.

$2p_y, 2p_z$ as well as the excitations of the type $2s^2 \rightarrow 2s^1 2p^1$. The latter excitation looks as requiring low energy, and therefore potentially important. However, it will be shown later on in this chapter that there are arguments for neglecting it (because of a weak coupling with the ground-state configuration). The x axis has been highlighted by us (through taking $2p_x$ orbitals only) for purely didactic reason, because soon we are going to frighten electron 2 by putting electron 1 in certain points on the x axis (therefore this axis is expected to be the main direction of escaping for electron 2). We have

$$2p_x = \zeta \sqrt{\frac{\zeta^3}{\pi}} x \exp(-\zeta r) = \zeta x\,(2s).$$

The drastically simplified wave function reads therefore as

$$\psi = \psi_0 + \kappa \psi_1, \tag{2.23}$$

where κ stands for a coefficient to be determined, which measures how much of the $2p^2$ configuration has to be added to the $2s^2$ configuration in order to describe correctly the physical behavior of the electrons[42] (this is forced, e.g., by the variational method or by a perturbational approach, see Chapter V1-5). Let us use a perturbational approach, in which we assume ψ_0 as an unperturbed wave function. Eq. (V1-5.26) (p. V1-279) says that with our current notation the coefficient κ may be estimated as

$$\kappa = \frac{\left\langle \psi_1 | \hat{H}^{(1)} \psi_0 \right\rangle}{E_0 - E_1}, \tag{2.24}$$

where the energies E_0 and E_1 correspond to the ground-state configuration (ψ_0) and the excited-state configuration (ψ_1), while $\hat{H}^{(1)}$ stands for the perturbation. Right now we have no idea what this perturbation is, but this is not necessary since (see Chapter V1-5) $\left\langle \psi_1 | \hat{H}^{(1)} \psi_0 \right\rangle = \left\langle \psi_1 | (\hat{H} - \hat{H}^{(0)}) \psi_0 \right\rangle = \left\langle \psi_1 | \hat{H} \psi_0 \right\rangle - E_0 \langle \psi_1 | \psi_0 \rangle = \left\langle \psi_1 | \hat{H} \psi_0 \right\rangle - 0 = \left\langle \psi_1 | \hat{H} \psi_0 \right\rangle$, where we have used that $\hat{H}^{(0)} \psi_0 = E_0 \psi_0$ and $\langle \psi_1 | \psi_0 \rangle = 0$ (the latter because of the orthogonality of $2s$ and $2p_x$).

It is seen therefore that we have to do with a matrix element of the Hamiltonian calculated with two Slater determinants containing orthonormal spin orbitals: $2s\alpha, 2s\beta, 2p_x\alpha, 2p_x\beta$, the first

[42] We are not intending to get a perfect description of the system, because with such a trial function there is no chance to solve the Schrödinger equation anyway. We are here rather for grasping a qualitative picture: will it be a Coulomb hole or not?

two composing ψ_0, the last ones present in ψ_1. Hence, all necessary conditions are satisfied for operating Slater–Condon rule III (Appendix V1-N, p. V1-707). We get

$$\left\langle \psi_1 | \hat{H} \psi_0 \right\rangle = \langle 2s\alpha\, 2s\beta | 2p_x\alpha\, 2p_x\beta \rangle - \langle 2s\alpha\, 2s\beta | 2p_x\beta\, 2p_x\alpha \rangle = \langle 2s\alpha\, 2s\beta | 2p_x\alpha\, 2p_x\beta \rangle - 0 =$$
$$(2s\,2s|2p_x\,2p_x) \equiv \int \left[2s(1)\,2p_x(1)\right] \frac{1}{r_{12}} \left[2s(2)\,2p_x(2)\right] \mathrm{d}V_1 \mathrm{d}V_2 > 0.$$

We have got a key inequality,[43] because from Eq. (2.24) and $E_0 < E_1$ it follows that

$$\kappa < 0. \tag{2.25}$$

Our qualitative conclusions will depend only on the sign of κ, *not on its particular value*. Let us make a set of exercises listed below (all distances in a.u.), first with ψ_0, then with ψ_1, and finally with $\psi = \psi_0 + \kappa\psi_1$. In all of them:

- the nucleus is immobilized at $(0, 0, 0)$;
- let us put electron 1, having the spin coordinate $\sigma_1 = \frac{1}{2}$, at $(-1, 0, 0)$;
- we will search the probability distribution of finding electron 2 with the spin coordinate $\sigma_2 = -\frac{1}{2}$;
- then, we will repeat the two last points with electron 1 at $(+1, 0, 0)$, i.e., on the opposite side of the nucleus and at the same electron–nucleus distance;
- we will compare the two probability distributions; if they were identical, there would be no correlation whatsoever, otherwise there would be a correlation.

To this end we will need three numbers to be calculated (the three numbers in parentheses represent x, y, z):

$$2s(-1,0,0) = 2s(1,0,0) = \sqrt{\frac{\zeta^5}{3\pi}} \exp(-\zeta) \equiv A > 0,$$
$$2p_x(1,0,0) = \zeta\sqrt{\frac{\zeta^3}{\pi}} \exp(-\zeta) = B > 0, \quad 2p_x(-1,0,0) \equiv \zeta\sqrt{\frac{\zeta^3}{\pi}}(-1)\exp(-\zeta) \equiv -B.$$

Function ψ_0

We expand the determinant (2.21) for electron 1 being at position $(-1, 0, 0)$ and obtain a function of the position of electron 2 in the form[44] $\frac{1}{\sqrt{2}} A \cdot 2s(2)$. Therefore, the (conditional)

[43] The inequality follows from evident repulsion of two identical electron clouds (of electron 1 and of electron 2), because they sit one on top of the other.

[44] Only the diagonal elements of the Slater determinant are nonzero (the rest of the elements vanish because of the spin functions), so we get the result right away.

probability density distribution of electron 2 is $\frac{1}{2}A^2\,[2s(2)]^2$ (Fig. 2.4a1). We repeat the same for position $(1,0,0)$ of electron 1 and get the identical result (Fig. 2.4a2). Conclusion: no Coulomb hole for the ground-state Slater determinant. Well, this is what we should expect. However, it may be the result depends on a type of Slater determinant. Let us take the Slater determinant ψ_1.

Function ψ_1

Expanding (2.22) for a fixed position $(-1,0,0)$ of electron 1 one gets a function depending on the position of electron 2 in the form $\frac{1}{\sqrt{2}}(-B)\cdot 2p_x(2)$, and therefore the conditional probability of finding electron 2 is $\frac{1}{2}B^2\left[2p_x(2)\right]^2$ (Fig. 2.4b1). Repeating the same for position $(1,0,0)$ of electron 1 we obtain the function $\frac{1}{\sqrt{2}}B\cdot 2p_x(2)$, but still we get the same probability distribution, i.e., $\frac{1}{2}B^2\left[2p_x(2)\right]^2$ (Fig. 2.4b2). Once again we obtain no Coulomb hole.

Function $\psi=\psi_0+\kappa\psi_1$

We calculate $\psi=\psi_0+\kappa\psi_1$ for position $(-1,0,0)$ of electron 1 and we obtain a function of the position of electron 2 in the form $\frac{1}{\sqrt{2}}A\cdot 2s(2)+\kappa\left[\frac{1}{\sqrt{2}}(-B)\cdot 2p_x(2)\right]$ with the corresponding conditional probability distribution of electron 2 as $\rho_-(2)=\frac{1}{2}A^2\,[2s(2)]^2+\frac{1}{2}\kappa^2B^2\left[2p_x(2)\right]^2-\kappa AB\cdot 2s(2)\cdot 2p_x(2)$ (Fig. 2.4c1). When repeating the same for position $(1,0,0)$ of electron 1 we obtain a *different* result, i.e., $\frac{1}{\sqrt{2}}A\cdot 2s(2)+\kappa\left[\frac{1}{\sqrt{2}}B\cdot 2p_x(2)\right]$, and therefore a *different* probability distribution: $\rho_+(2)=\frac{1}{2}A^2\,[2s(2)]^2+\frac{1}{2}\kappa^2B^2\left[2p_x(2)\right]^2+\kappa AB\cdot 2s(2)\cdot 2p_x(2)$ (Fig. 2.4c2). So, there *is* a correlation of the electronic motion. It would be even better to have this correlation reasonable.[45] Figures (c1) and (c2) show that indeed the correlation stands to reason: *the two electrons avoid one another if electron 1 is on the left-hand side and electron 2 is on the right-hand side, and* vice versa.

If we did not have inequality (2.25), this conclusion could not be derived. For $\kappa>0$, electron 2 would accompany electron 1, which means "a completely nonphysical" behavior. For $\kappa=0$ or $\kappa=\pm\infty$ there would be no correlation.[46] All therefore depends on the coefficients of the linear combination of Slater determinants. This is the variational principle or the perturbational theory that takes care the wave function is close to the solution of the Schrödinger equation for the ground state. This forces a physics-based description of the electronic correlation, in our case $\kappa<0$.

[45] An unreasonable *correlation* would be, e.g., when the two electrons were sticking to one other!

[46] All these cases correspond to a single determinant (for $\kappa=0$) and ψ_1 or $-\psi_1$ (for $\kappa=\pm\infty$).

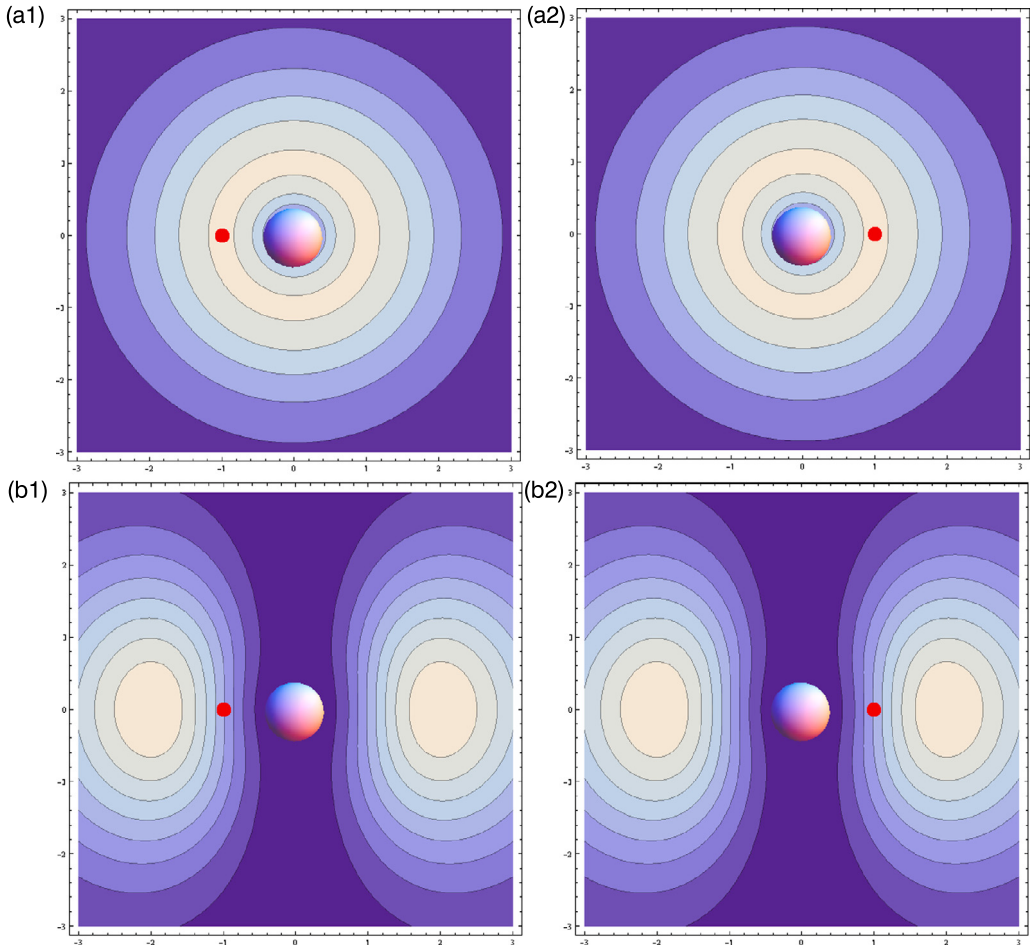


Fig. 2.4. A single Slater determinant cannot describe any Coulomb correlation, but a linear combination of Slater determinants can. The figure pertains to the beryllium atom, with a pseudonucleus (of charge $+2$) shown as a large sphere in the center of each figure. All figures show the sections ($z = 0$) of the (conditional) probability density distribution of finding electron 2 ((a), upper row – for the single Slater determinant ψ_0; (b), second row – for the single Slater determinant ψ_1; (c), bottom row – for a two-determinantal wave function $\psi = \psi_0 + \kappa\psi_1$), when electron 1, symbolized by a small sphere, resides at $(-1, 0, 0)$ (left-hand side figures with the symbol 1) or at $(1, 0, 0)$ (right-hand side figures with the symbol 2). Only in the case of the two-determinantal wave function $\psi = \psi_0 + \kappa\psi_1$ one obtains any difference between the probability distributions, when electron 1 occupies two positions: $(-1, 0, 0)$ and $(1, 0, 0)$! The values $\kappa < 0$ correspond to mutual avoiding of the two electrons (in such a case the wave function takes into account the Coulomb hole), $\kappa = 0$ means mutual ignoring of the two electrons, $\kappa > 0$ would correspond to a very bad wave function that describes the two electrons sticking to one another (see the text). In order to highlight the correlation effect in the figure (purely didactic reasons) we took quite arbitrarily $\kappa = -0.7$ and $\zeta = 1$.

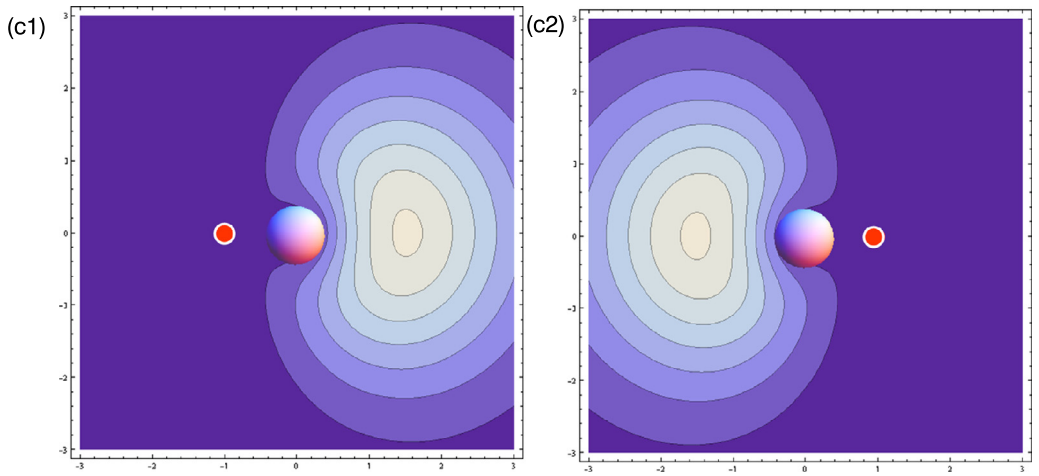


Fig. 2.4. (*continued*)

A two-determinantal function $\psi = \psi_0 + \kappa\psi_1$ with $\kappa < 0$ can (in contrast with the single-determinantal functions ψ_0 and ψ_1) approximate the effect of the dynamic correlation (Coulomb hole). Of course, a combination of many Slater determinants with appropriate coefficients can do it better.

2.8 Anticorrelation, or do electrons stick together in some states?[47]

What about electronic correlation in excited electronic states? Not much is known for excited states in general. In our case of function (2.23) the Ritz variational method would give two solutions: one of lower energy corresponding to $\kappa < 0$ (this solution has been approximated by us using the perturbational approach) and one (the excited electronic state) of the form $\psi_{exc} = \psi_0 + \kappa'\psi_1$. In such a simple two-state model as we have, the coefficient κ' can be found just from the (necessary) orthogonality of the two solutions: $\langle\psi_{exc}|\psi\rangle = \langle\psi_0 + \kappa'\psi_1|\psi_0 + \kappa\psi_1\rangle = 1 + \kappa\kappa'^* + \kappa'^* \langle\psi_1|\psi_0\rangle + \kappa \langle\psi_0|\psi_1\rangle = 1 + \kappa\kappa'^* = 0$.

Hence $\kappa'^* = -\frac{1}{\kappa} > 0$. We have therefore $\kappa' > 0$ and it is quite intriguing that our excited state corresponds now to what we might call here an "anticorrelation." In the excited state we got the two electrons just sticking to one another! This result certainly cannot be thought as of general value for excited states. It is probable that in excited electronic states the electronic correlation gets weaker, but according to what we have found in our two-state model, *some excited*

[47] L. Piela, *Sci. China Chem.*, 57(2014)1383.

states might exhibit the electronic anticorrelation! This indication may be less surprising than it sounds. For example, the hydrogen molecule has not only the covalent states, but also the excited states of ionic character (see below). In the ionic states the two electrons prefer to occupy the same space (still repelling each other), as if there were a kind of "attraction" between them.

Electrons may attract themselves – two rings

Do the electrons repel each other? Of course. Does it mean the electrons try to be as far from each other as possible? Yes, but the words "as possible" are important. Usually this means a game between the electrons strongly attracted by a nucleus and their important repulsion through the Pauli exclusion principle (Fermi hole) together with much less important Coulomb repulsion (Coulomb hole).

Let us try to simplify the situation. First, let us remove the presence of the nuclei and see what electrons like without them. Then, while all the time keeping the Coulomb repulsion ("Coulomb hole"), we will either switch on the Fermi hole by considering the triplet states with the two electrons having opposite spins or switch off the Fermi hole by taking the singlet states of these two electrons.

Let us take two coaxial coplanar rings 1 and 2 (of radii R_1 and R_2, $R_2 \geq R_1$) with a single electron on each of them, and with the electron's position defined by the angle φ_i, for the rings $i = 1, 2$ (Fig. 2.5).

The two electrons repel each other according to the Coulomb law, $\frac{e^2}{r_{12}}$. The Hamiltonian for the system reads as

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2mR_1^2}\frac{\partial^2}{\partial\varphi_1^2} - \frac{\hbar^2}{2mR_2^2}\frac{\partial^2}{\partial\varphi_2^2} + V, \tag{2.26}$$

where

$$V = \begin{cases} \frac{e^2}{r_{12}} & \text{for each electron residing on its ring,} \\ \infty & \text{otherwise.} \end{cases} \tag{2.27}$$

Let us introduce new coordinates

$$\Phi = \frac{R_1^2\varphi_1 + R_2^2\varphi_2}{R_1^2 + R_2^2}, \tag{2.28}$$

$$\varphi = \varphi_1 - \varphi_2. \tag{2.29}$$

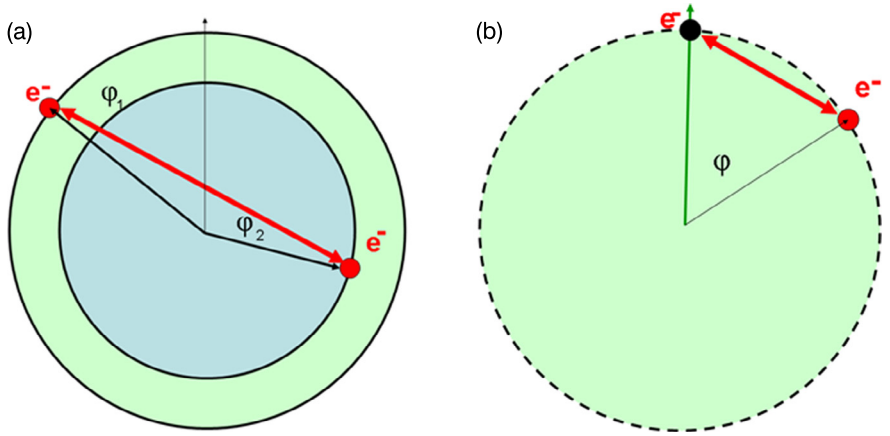


Fig. 2.5. Two electrons in two rings. (a) Two coplanar and concentric rings with a single electron on each of them, indicated by the angles φ_1 and φ_2, respectively. The electrons interact (through space) by Coulombic force. (b) One can separate a single variable out (similarly to elimination of the center-of-mass motion) and obtain the Schrödinger equation that describes the relative motion of electron 1 with respect to electron 2, that is immobilized at the origin (with the only variable left: $\varphi = \varphi_1 - \varphi_2$).

We obtain

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2mR_1^2}\frac{\partial^2}{\partial\varphi_1^2} - \frac{\hbar^2}{2mR_2^2}\frac{\partial^2}{\partial\varphi_2^2} + V = \hat{\mathcal{H}}_{CM}(\Phi) + \hat{H}(\varphi), \tag{2.30}$$

where

$$\hat{\mathcal{H}}_{CM}(\Phi) = -\frac{\hbar^2}{2m\left(R_1^2 + R_2^2\right)}\frac{\partial^2}{\partial\Phi^2}, \tag{2.31}$$

$$\hat{H}(\varphi) = -\frac{\hbar^2}{2m}\left(\frac{1}{R_1^2} + \frac{1}{R_2^2}\right)\frac{\partial^2}{\partial\varphi^2} + \frac{e^2}{r_{12}}. \tag{2.32}$$

The solution of the Schrödinger equation with the Hamiltonian $\hat{\mathcal{H}}$ has the form $\psi_{CM}(\Phi)\,\psi(\varphi)$ with

$$\psi_{CM}(\Phi) = \frac{1}{\sqrt{2\pi}}\exp[i J_{CM}\Phi], \tag{2.33}$$

$$\mathcal{E} = \mathcal{E}_{CM} + E,$$

$$\mathcal{E}_{CM} = \frac{\hbar^2}{2m\left(R_1^2 + R_2^2\right)}J_{CM}^2, \qquad J_{CM} = 0, \pm1, \pm2, ..., \tag{2.34}$$

$$\hat{H}(\varphi)\,\psi_n(\varphi) = E_n\psi_n(\varphi). \tag{2.35}$$

We separate out the Φ variable with the Hamiltonian $\hat{\mathcal{H}}_{CM}(\Phi)$ and we are left with Eq. (2.35) only.

If the Coulombic repulsion were absent, the solutions would be: $const$, $\exp(in\varphi)$, and $\exp(-in\varphi)$, $n = 1, 2, ...$ which means a nondegenerate nodeless ground state and all other states doubly degenerate. Note that *for all these states* $|\psi_n(\varphi)|^2$ *is a constant (does not depend on* φ*)*.

Now we reconsider the Coulombic repulsion. In fact, after separation, we may treat electron 1 as sitting all the time at $\varphi = 0$ and electron 2 (with the coordinate φ) moving. The eigenfunctions for this problem lead to the probability densities shown in Fig. 2.6:

- The nodeless ground state ψ_0, which, because of the Coulombic term, will not be a constant, but have a maximum at $\varphi = 180°$ (i.e., farthest away from electron 1, Fig. 2.6b). The spatial function is a symmetric function of φ, so this describes the singlet ground state.
- The first excited state ψ_1 has one node, and this nodal line should be along the straight line from electron 1 to position $\varphi = 180°$. This function is antisymmetric with respect to exchange of the electrons ($\varphi \to -\varphi$), so it represents the triplet state. This state is of low energy, because it takes care of the Fermi hole; the wave function for electron 2 equals zero at the position of electron 1.
- The second excited state (ψ_2) will also have one node, but the nodal plane has to be orthogonal to that of ψ_1 (symmetric function, i.e., the first excited singlet) and function ψ_2 has to be orthogonal to ψ_0 and ψ_1. *The orthogonality to* ψ_0 *means it has to have a larger absolute amplitude at the position of electron* 1 than on the opposite site ($\varphi = 180°$). So we see that already such a low-energy state as ψ_2 is of the kind that electron 2 prefers to be closer to electron 1!
- Similar phenomenon will appear for higher states.

There are excited states with electrons keeping close to one another.

Electrons may "attract themselves" – two boxes

Let us consider two one-dimensional boxes (see Fig. 2.7a), parallel to each other, each of length L and with an electron inside: electron 1 with position x_1 in box 1, and electron 2 with position x_2 in box 2.

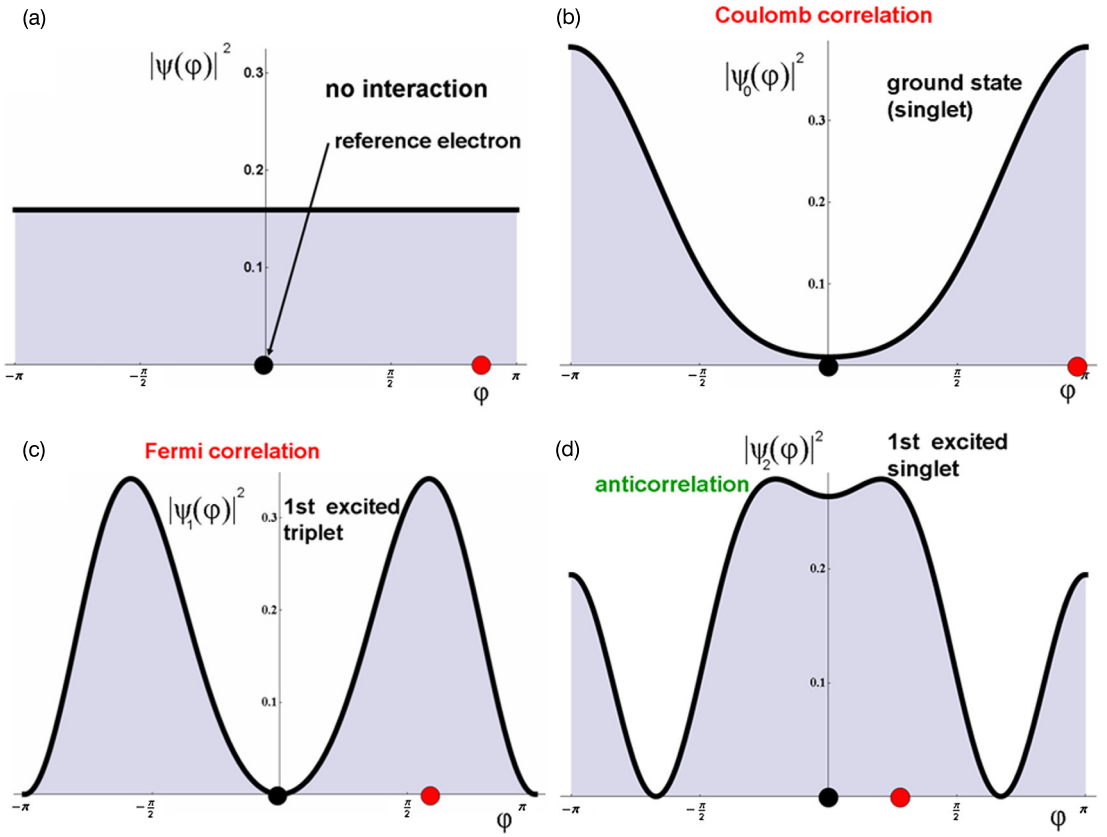


Fig. 2.6. Two electrons within two rings. The probability density distributions in the low-energy quantum states as a function of φ (measuring the position of electron 1 with respect to the immobilized electron 2). (a) The Coulombic interaction switched off. Every position of electron 1 is equally probable. (b) After switching the electron-electron interaction on, in the ground state electron 1 prefers to be on the opposite side with respect to the position of the reference electron 2 (forming the Coulomb hole, both electrons have the opposite spins [singlet state]). (c) The lowest-energy triplet state results in the Fermi hole around electron 2 (the node at the position of electron 2). (d) In the first excited singlet state, electron 1, in spite of a lot of empty space to move and in spite of the Coulomb repulsion with electron 2, prefers to reside close to electron 2. This is what is termed in the present textbook anticorrelation.

The Hamiltonian reads as

$$\hat{H} = \hat{H}^{(0)} + V, \tag{2.36}$$

$$\hat{H}^{(0)} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x_2^2}, \tag{2.37}$$

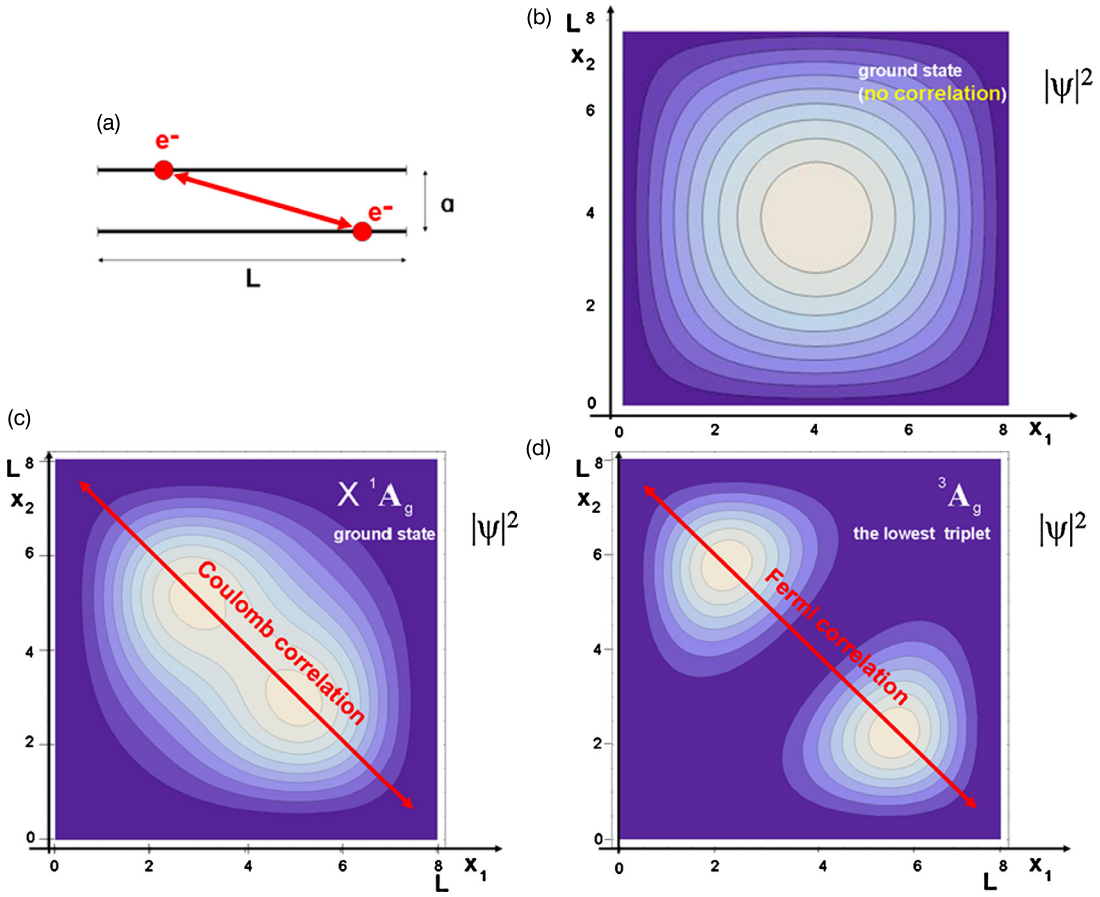


Fig. 2.7. Two electrons within two boxes, with the potential energy being the electron–electron interaction. (a) The two boxes (of length $L = 8$ a.u.), with one electron in each of them, are parallel, separated by the distance $a = 2$ a.u., everywhere else the potential energy is equal to $V = \infty$. The position of the i-th electron in the i-th box is given by $0 \le x_i \le L,\ i = 1, 2$. (b)–(g) The probability density distributions as functions of (x_1, x_2). (b) The unperturbed ground state (the Coulomb repulsion switched off). (c) After switching the electron–electron repulsion on, the ground state (singlet) as a result of the Coulomb correlation is shown. (d) The lowest triplet state with the Fermi correlation visible. (e) The first excited singlet shows the anticorrelation: electron 1, in spite of a lot of empty space to move and in spite of the Coulomb repulsion, prefers to reside close to electron 2. In fact the largest probability density corresponds to the electrons occupying the same position in space. (f) The third excited singlet and (g) the fifth excited singlet states also show the anticorrelation effect.

$$V = \begin{cases} \dfrac{e^2}{\sqrt{(x_1 - x_2)^2 + a^2}} & \text{for } 0 \le x_1 \le L \text{ and } 0 \le x_2 \le L, \\ \infty & \text{otherwise.} \end{cases} \tag{2.38}$$

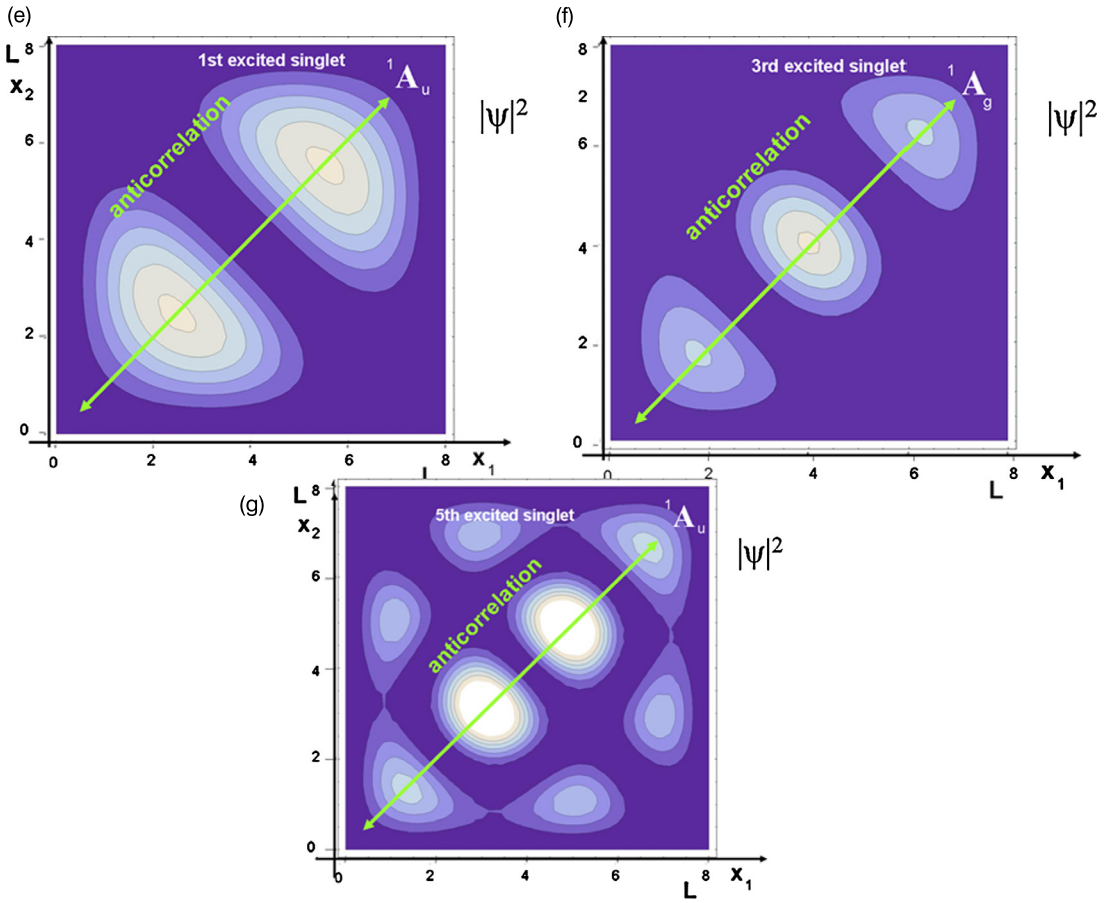


Fig. 2.7. (*continued*)

The solution of the Schrödinger equation

$$\hat{H}\psi(x_1, x_2) = E\psi(x_1, x_2)$$

will be sought as a linear combination of the eigenfunctions of $\hat{H}^{(0)}$ (configuration interaction [CI] method) with the expansion coefficients $c_{n_1 n_2}$, i.e.,

$$\psi(x_1, x_2) = \sum_{n_1=1,2,..,n_{\max}} \sum_{n_2=1,2,..,n_{\max}} c_{n_1 n_2} \phi_{n_1 n_2}(x_1, x_2),$$

and expansion functions

$$\phi_{n_1 n_2}(x_1, x_2) = \frac{2}{L} \sin \frac{n_1 \pi}{L} x_1 \sin \frac{n_2 \pi}{L} x_2.$$

In the Ritz matrix problem $(\mathbf{H} - E\mathbf{S})\,\mathbf{c} = \mathbf{0}$ (the atomic units are used) one has $S_{n_1,n_2,n_1',n_2'} = \delta_{n_1,n_1'}\delta_{n_2,n_2'}$ and $H_{n_1,n_2,n_1',n_2'} = \delta_{n_1,n_1'}\delta_{n_2,n_2'}\left[\left(n_1^2 + n_2^2\right)\frac{\pi^2}{2L^2}\right] + V_{n_1,n_2,n_1',n_2'}$, with $V_{n_1,n_2,n_1',n_2'} = \int_0^L \mathrm{d}x_1 \int_0^L \mathrm{d}x_2 \frac{\phi_{n_1 n_2}(x_1,x_2)\phi_{n_1' n_2'}(x_1,x_2)}{\sqrt{(x_1-x_2)^2+a^2}}$.

Fig. 2.7b shows an approximation $\left|\tilde{\psi}_0\right|^2 = [\frac{2}{L}\sin\frac{\pi}{L}x_1 \sin\frac{\pi}{L}x_2]^2$ to the probability density distribution $|\psi_0|^2$ corresponding to the ground-state wave function ψ_0 (calculated for $L = 8$ a.u., $a = 2$ a.u.). Such $\left|\tilde{\psi}_0\right|^2$ represents a product of the one-electron probability density distributions for the individual electrons, $\frac{2}{L}\sin^2\frac{\pi}{L}x_1$ and $\frac{2}{L}\sin^2\frac{\pi}{L}x_2$; therefore, $\tilde{\psi}_0$ does not contain any electronic correlation (the electrons move independently). For example, from Fig. 2.7b it is seen that irrespective of whether electron 1 is at $x_1 = 1$ a.u. or at $x_1 = L - 1 = 7$ a.u., the probability density for finding electron 2 is the same, always with a maximum at $x_2 = L/2 = 4$ a.u. The calculated $\frac{1}{V_{1111}} \equiv \tilde{r}_{12} = 2.56$ a.u. may be taken as a measure of the interelectronic distance in the situation that the electrons do not correlate their motion (this is not the optimal mean-field approximation though).

The computed (with $n_{\max} = 10$, i.e., 100 expansion functions) ground-state energy is $E_0 = 0.527$ a.u. and the ground-state wave function ψ_0 gives $\tilde{r}_{12} = \left\langle \psi_0 \middle| \frac{1}{r_{12}} \psi_0 \right\rangle^{-1} = 2.83$ a.u. (Fig. 2.7c). This is a singlet function, because its spatial form is symmetric with respect to the operation $x_1 \leftrightarrow x_2$, equivalent to the reflection with respect to the axis $(0,0) - (L,L)$ (the spin function is antisymmetric). The increasing of the estimated interelectronic distance from 2.56 a.u. to 2.83 a.u. shows that ψ_0 describes a correlation of the two electrons (the Coulomb hole), which in their motion avoid each other. This is seen from $|\psi_0|^2$ in Fig. 2.7c, e.g., if electron 1 is at $x_1 = 3$ a.u., the maximum of the probability distribution for finding electron 2 is at $x_2 = 5$ a.u., and *vice versa*. This is an example of what may be called the correlation-dominated state.

The next state in the energy scale ($E_1 = 0.665$ a.u.) represents a triplet state ψ_1 (is antisymmetric with respect to the operation $x_1 \leftrightarrow x_2$, Fig. 2.7d). The estimated electron–electron distance in this state is equal to 3.68 a.u., i.e., much larger than that for the independent-electron model (2.56), and also much larger than for the ground state (2.83), where the Coulomb hole effect was operating. So, this state is definitely also a correlation-dominated one, a result of cooperation of the Fermi and Coulomb holes in this case. One can also notice (from (c,d) and the numerical values of $\tilde{r}_{12}$) that the Fermi hole is much more important than the Coulomb hole.

The next state in the energy scale is the first excited singlet with $E_2 = 0.798$ a.u. This state pertains to the main goal of the present section. Its probability density distribution $|\psi_2|^2$ has

two maxima (Fig. 2.7e), *both corresponding to the same positions of the two electrons*: the first at $x_1 = x_2 \approx 2.25$ a.u., the second at $x_1 = x_2 \approx L - 2.25 = 5.75$ a.u. This suggests that the two electrons, despite their Coulombic repulsion, keep close in space. This is seen also from the estimated electron–electron distance in this state: one gets $\tilde{r}_{12} = 2.48$ a.u., i.e., a *smaller* value than 2.56 a.u., which corresponds to the independent-electron model. Thus, in this state we have what may be called the *electron–electron anticorrelation*. The same anticorrelation-dominated character have also the third excited singlet state (Fig. 2.7f), with energy $E_5 = 1.172$ a.u. and $\tilde{r}_{12} = 2.42$ a.u., and the fifth excited singlet (Fig. 2.7g), of energy $E_9 = 1.687$ a.u. and $\tilde{r}_{12} = 2.55$ a.u.

Other excited states exhibit in most cases a complex behavior of electrons, often with an electron position-dependent intriguing coexistence of the correlation and anticorrelation patterns.

2.9 Valence bond (VB) method

2.9.1 Resonance theory – hydrogen molecule

Slater determinants are usually constructed from *molecular* spin orbitals. If, instead, we use *atomic* spin orbitals and the Ritz variational method (Slater determinants as the expansion functions) we would get the most general formulation of the valence bond (VB) method. Note that since the atomic orbitals form a nonorthogonal basis set, the same is true for the resulting Slater determinants. The beginning of the VB theory goes back to papers by Heisenberg (he was first to use the term "resonance," a slightly misleading name), the first application was made by Heitler and London, and later theory was generalized by Hurley, Lennard-Jones, and Pople.[48]

The essence of the VB method can be explained by an example. Let us take the hydrogen molecule with atomic spin orbitals of type $1s_a\ \alpha$ and $1s_b\ \beta$, denoted shortly as $a\alpha$ and $b\beta$, centered at two points. Let us construct from them several (nonnormalized) Slater determinants, for instance,

$$\psi_1 = \frac{1}{\sqrt{2}} \begin{vmatrix} a(1)\alpha(1) & a(2)\alpha(2) \\ b(1)\beta(1) & b(2)\beta(2) \end{vmatrix} = \frac{1}{\sqrt{2}} \big[a(1)\alpha(1)b(2)\beta(2) - a(2)\alpha(2)b(1)\beta(1) \big],$$

$$\psi_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} a(1)\beta(1) & a(2)\beta(2) \\ b(1)\alpha(1) & b(2)\alpha(2) \end{vmatrix} = \frac{1}{\sqrt{2}} \big[a(1)\beta(1)b(2)\alpha(2) - a(2)\beta(2)b(1)\alpha(1) \big],$$

[48] W. Heisenberg, *Zeit. Phys.*, 38(1926)411, *ibid.* 39(1926)499, *ibid.* 41(1927)239; W. Heitler, F. London, *Zeit. Phys.*, 44(1927)455; A.C. Hurley, J.E. Lennard-Jones, J.A. Pople, *Proc. Roy. Soc. London A*, 220(1953)446.

$$\psi_3 = \frac{1}{\sqrt{2}} \begin{vmatrix} a(1)\alpha(1) & a(2)\alpha(2) \\ a(1)\beta(1) & a(2)\beta(2) \end{vmatrix} = \frac{1}{\sqrt{2}} \left[a(1)\alpha(1)a(2)\beta(2) - a(2)\alpha(2)a(1)\beta(1) \right] =$$

$$a(1)a(2) \cdot \frac{1}{\sqrt{2}} [\alpha(1)\beta(2) - \alpha(2)\beta(1)] \equiv \psi_{H^-H^+},$$

$$\psi_4 = \frac{1}{\sqrt{2}} \begin{vmatrix} b(1)\alpha(1) & b(2)\alpha(2) \\ b(1)\beta(1) & b(2)\beta(2) \end{vmatrix} = b(1)b(2) \cdot \frac{1}{\sqrt{2}} [\alpha(1)\beta(2) - \alpha(2)\beta(1)] \equiv \psi_{H^+H^-}.$$

The functions ψ_3, ψ_4 and the normalized difference $(\psi_1 - \psi_2) \equiv \psi_{HL}$,

HEITLER–LONDON FUNCTION

$$\psi_{HL} = [a(1)b(2) + a(2)b(1)] \cdot \frac{1}{\sqrt{2}} \left[\alpha(1)\beta(2) - \alpha(2)\beta(1) \right], \tag{2.39}$$

are eigenfunctions of the operators $\hat{S}^2$ and $\hat{S}_z$ (cf. Appendix V1-R, p. V1-731) corresponding to the singlet state. The functions ψ_3, ψ_4 for obvious reasons are called *ionic structures* (H^-H^+ and H^+H^-),[49] whereas the function ψ_{HL} is called a Heitler–London function or a *covalent structure*.[50]

The VB method relies on optimization of the expansion coefficients c in front of these structures in the Ritz procedure (p. V1-271),

$$\psi = c_{cov}\psi_{HL} + c_{ion1}\psi_{H^-H^+} + c_{ion2}\psi_{H^+H^-}. \tag{2.40}$$

Since the expansion functions are mutually nonorthogonal, the $|c|^2$ values give only an indication about the magnitude of the expansion functions' contributions. This is a strictly mathematical problem, and the chemistry relationship of the expansion functions is in fact irrelevant if the expansion is infinite. This does not mean, however, that chemistry is irrelevant for short expansions. For short expansions, as usually, the physical/chemical basis of the expansion functions (i.e., their ability to describe the main phenomena) counts, and may represent valuable information for a chemist.

[49] Since both electrons reside at the same nucleus.

[50] Since both electrons belong to the same extent to each of the nuclei and the function describes the main effect of covalent chemical bonding.

Fritz Wolfgang London (1900–1954) was born in Breslau (now Wrocław) and studied in Bonn, Frankfurt, Göttingen, Munich (PhD at 21), and Paris. Later he worked in Zurich, Rome, and Berlin. He escaped from Nazism to the UK, where he worked at Oxford University (1933–1936). In 1939 London emigrated to the USA, where he became professor of theoretical chemistry at Duke University in Durham. Fritz London rendered great services to quantum chemistry. He laid the foundations of the theory of the *chemical* (covalent) bond and, in addition, introduced dispersion interactions, one of the most important *intermolecular* interactions. This is nearly all of what chemistry is about. He also worked in the field of superconductivity.

The covalent structure itself, ψ_{HL}, was one great success of Walter Heitler[51] and Fritz London. For the first time the qualitatively correct description of the chemical bond was obtained. The crucial point turned out to be an inclusion – in addition to the product function $a(1)b(2)$ – its counterpart *with exchanged electron numbers* $a(2)b(1)$, since the electrons *are* indistinguishable. If we expand the Hartree–Fock determinant with doubly occupied bonding orbital $a + b$, we also obtain a certain linear combination of the three structures mentioned,[52] but with *the constant coefficients independent of the interatomic distance*, i.e.,

$$\psi_{RHF} = N(\psi_{HL} + \psi_{H^-H^+} + \psi_{H^+H^-}). \tag{2.41}$$

[51] Walter Heitler (1904–1981), German chemist, professor at the University in Göttingen, later in Bristol and Zurich.

[52] Indeed, the normalized Hartree–Fock determinant (double occupation of the molecular orbital $\varphi_1 = \frac{1}{\sqrt{2(1+S)}}(a+b)$, where the overlap integral between the atomic orbitals $S = (a|b)$) can be rewritten as

$$\psi_{RHF} = \frac{1}{\sqrt{2!}} \begin{vmatrix} \varphi_1(1)\alpha(1) & \varphi_1(2)\alpha(2) \\ \varphi_1(1)\beta(1) & \varphi_1(2)\beta(2) \end{vmatrix} =$$

$$\frac{1}{2(1+S)}[a(1)a(2) + b(1)b(2) + a(1)b(2) + a(2)b(1)] \frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)] =$$

$$\frac{1}{2(1+S)}\left[\psi_{H^--H^+} + \psi_{H^+-H^-} + \psi_{HL}\right].$$

This leads to a very bad description of the H_2 molecule at long internuclear distances with the Hartree–Fock method. Indeed, for long internuclear distances the Heitler–London function should dominate, because it corresponds to the (correct) dissociation limit (two ground-state hydrogen atoms). The trouble is that, with fixed coefficients, *the Hartree–Fock function overestimates the role of the ionic structure* for long interatomic distances. Fig. 2.8 shows that the Heitler–London function describes the electron correlation (Coulomb hole), whereas the Hartree–Fock function does not.

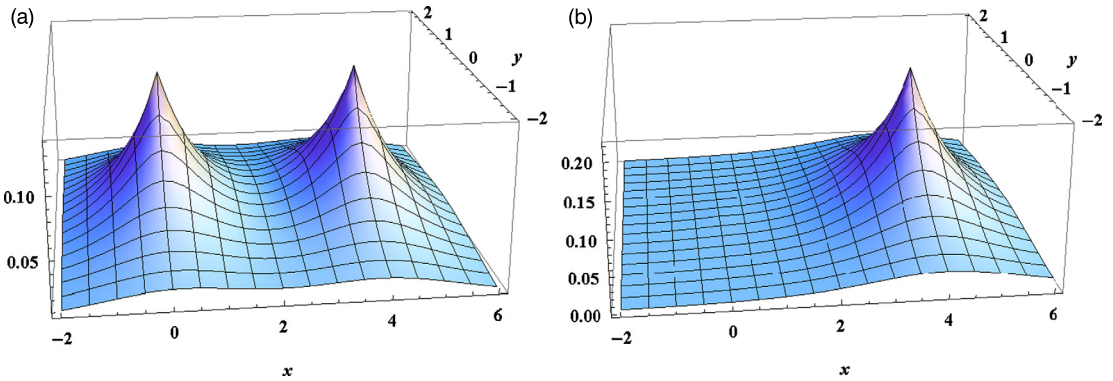


Fig. 2.8. Illustration of electron correlation in the hydrogen molecule. The nuclear positions are $(0,0,0)$ and $(4,0,0)$ in a.u. Slater orbitals of $1s$ type have orbital exponents equal to 1. (a) Visualization of the xy cross-section of the wave function of electron 2, assuming that electron 1 resides on the nucleus (either the first or the second one) and has spin coordinate $\sigma_1 = \frac{1}{2}$, whereas electron 2 has spin coordinate $\sigma_2 = -\frac{1}{2}$ and the total wave function is equal to $\psi = N\{ab + ba + aa + bb\}\{\alpha\beta - \beta\alpha\}$, i.e., it is a Hartree–Fock function. The plot is *the same* independently of which nucleus electron 1 resides at, i.e., we observe the *lack of any correlation* of the motions of electrons 1 and 2. (b) A similar plot, but for the Heitler–London function $\psi_{HL} = \mathcal{N}_{HL}\,[a(1)b(2) + a(2)b(1)]\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big]$ and with electron 1 residing at nucleus $(0,0,0)$. Electron 2 runs to the nucleus in position $(4,0,0)$. We have the correlation of the electronic motion.

2.9.2 Resonance theory – polyatomic case

The VB method was developed in organic chemistry by Linus Pauling under the name of *theory of resonance*. The method can be applied to all molecules, although a particularly useful field of applications of resonance theory can be found in organic chemistry of aromatic systems. For example, the total electronic wave function of the benzene molecule is presented as a linear combination of resonance structures[53] Φ_I,

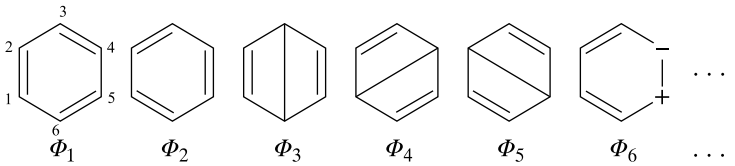
[53] Similar to the original applications, we restrict ourselves to the π electrons; the σ electrons are treated as inactive in each structure, forming, among other things, the six C–C bonds presented below.

Linus Carl Pauling (1901–1994), American physicist and chemist, in the years 1931–1964 professor at the California Institute of Technology in Pasadena, in 1967–1969 professor at the University of California, San Diego, from 1969–1974 professor at Stanford University. He received the 1954 Nobel Prize "*for his research into the nature of the chemical bond and its application to the elucidation of the structure of complex substances*." In 1962 he received the Nobel Peace Prize. His major achievements include the development of the theory of chemical bonds, i.e., the VB method

(also called resonance theory), and the determination of the structure of one of the fundamental structural elements of proteins, the α-helix.

$$\psi = \sum_I c_I \Phi_I, \tag{2.42}$$

to each of which a chemical bond pattern, in analogy with the ionic and covalent structures discussed above, is assigned (in addition to its mathematical form). For example, six π electrons can participate in the following "adventures" (forming covalent and ionic bonds).



The first two structures are the famous Kekulé structures, the next three are Dewar structures, the sixth is an example of the possible mixed covalent-ionic structures (see Fig. 2.9). It is impressive that allowing only the valences 4 and 1 for the carbon and hydrogen atoms, respectively, we are left with as many as 217 basic (nonionic) isomeric skeletal structures...[54] About a half of them may be considered as a doable synthetic target in organic chemistry (as a part of more extended molecules). From Fig. 2.9 one can see how this may happen. Indeed, a chemical

54 After Gopalpur Nagendrappa, *Resonance*, 6(2001)74. If one takes into account their possible diastereomers and enantiomers, the total number of isomers is 328.

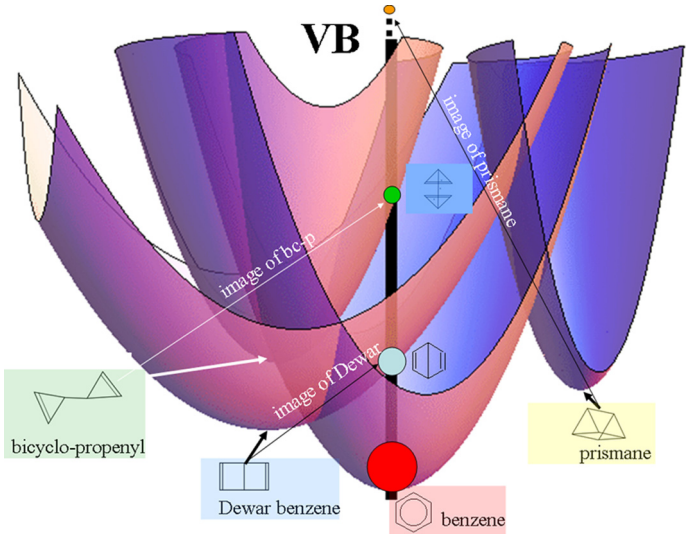


Fig. 2.9. Concept of the VB method, with the example of the C_6H_6 system. Each parabola-like surface for a diabatic state represents the mean value of electronic energy of a resonance structure (characterized by its particular chemical bonding pattern, here the Kekulé structure, the Dewar structure, the prismane, bicyclo-propenyl, etc.) as a function of the configuration of the nuclei. The nuclear configurations corresponding to the minima of these energies mean the equilibrium geometries of the corresponding C_6H_6 isomeric species. As one can see the ground-state PES (bottom surface) is composed of the diabatic-like domains of very different chemical character. The VB method is applied for a fixed geometry of the nuclear framework (shown as the vertical thick line), the same for all the resonance structures. The total wave function is constructed as a linear combination of the resonance structure wave functions, all taken at this particular geometry, and therefore resembling ("being images" of) the isomers' wave functions. The contribution of a given resonance structure to the full wave function is indicated as the size of the corresponding circle located on the thick vertical line. The higher the "image" energy, the smaller the circle. Thus, the VB ground state resembles mainly a superposition of the two Kekulé structures (big circle at the bottom), and to a smaller extent the other resonance structures. This is what is known as benzene in our test tubes. Note that the molecule, being under some steric strain of the neighborhood, when deforming in its ground state, may change dramatically its chemical character.

neighborhood may introduce a steric strain and distort the molecule's nuclear framework to such an extent, that the character of its electronic ground state changes qualitatively. For example, it is no longer a ring with identical bonds as people imagine benzene (a superposition of the two Kekulé structures), but resembles more a molecule with the Dewar bond pattern. Thus, in the VB method we use "images" as the expansion functions, while the same images may become real under a neighborhood's steric strain.

Now, a few words follow about the mathematical form of Φ_I. From the corresponding chemical patterns, we may deduce which atomic orbitals take part in the covalent bonds. For example, for a π-type example (approximation) as far as the Φ_1 structure is concerned, we can write it as the antisymmetrized[55] product of three Heitler–London functions (involving the proper pairs of $2p_z$ carbon atomic orbitals), the first for electrons 1, 2, the second for electrons 3, 4, and the third for electrons 5, 6. We may continue in this way for other Φ_I.

In chemical practice resonance theory operates as a set of heuristic rules of thumb, not as a quantum mechanical computational tool.[56] The rules are based on a reasonable expectation (Fig. 2.9) that

> a molecule gets an energy gain (with the accompanied electronic delocalization), known as the resonance energy, if it has *several* low-energy VB structures possible, each based on the corresponding Lewis electronic structure (electronic octets).

The problem, however, was how to guess which of two VB structures would have lower energy. In organic chemistry practice some simple rules were operating for choosing low-energy VB structures (no calculations...). The VB structure (usually of Lewis type) would presumably be of lower energy if it had:

- a greater number of covalent bonds,
- the least number of formal charges left,
- the least separation of the formal charges,
- a negative charge on the more electronegative atom,
- positive charges on the least electronegative atom,
- resonance structures that are equivalent and contribute equally.

It is remarkable and encouraging that such, in fact primitive, rules continue to represent a great help for organic chemists, who design their reaction paths with enormous synthetic successes.

---

55 See antisymmetrization operator (p. V1-707).

56 One of the reasons is probably a difficulty in programming the procedure with so many Φ_I functions of very distinct internal mathematical structure that depend so much on the molecule considered.

There is another aspect of the resonance theory though. Such rules for writing the structures were not clear, and the electrons were located to some extent in an arbitrary manner, making the impression that it is up to theoretical chemists to use their imaginations and draw imaginary pictures and – next – to translate them into mathematical form to obtain – in the best case after applying the variational method – an approximation to the wave function (and to the energy).

In fact, the deepest sense of the problem is purely mathematical. Although it may seem very strange to students (fortunately), many people were threatened for supporting the theory of resonance. Scientists serving the totalitarian regime decided to attack Eq. (2.42). How, was this possible[57]? The Stalinists did not like the idea that "*the sum of fictitious structures can describe reality*," and especially people who dare to think in an independent way.

[57] Of course, the *true* reason was not a convergence of a series in the Hilbert space, but their personal careers *at any price*. Totalitarian systems never have problems finding such "scientists." In chemistry, there was the danger of losing one's job, in biology, of losing one's life.
To encourage young people to protect the freedom – currently jeopardized by what is presented cleverly as "political correctness" – and to reflect on human nature in general, some exempts from the resolution adopted by the All-Soviet Congress of Chemists of the Soviet Union are reported below. The resolution pertains, i.a., to the theory of resonance (after the disturbing and reflective book by S.E. Shnoll, "*Gieroi i zlodiei rossijskoj nauki*," Kron-Press, Moscow, 1997, p. 297).

"Dear Joseph Vissarionovich [Stalin],
The participants of the All-Soviet Congress send to you, the Great Leader and Teacher of all progressive mankind, our warm and cordial greetings. We Soviet chemists gathered together to decide, by means of broad and free discussion, the fundamental problems of the contemporary theory of the structure of molecules, want to express our deepest gratitude to you for the everyday attention you pay to Soviet science, particularly to chemistry. Our Soviet chemistry is developing in the Stalin era, which offers unlimited possibilities for the progress of science and industry. Your brilliant work in the field of linguistics put the tasks for still swifter progress in front of all scientists of our fatherland [...]. Motivated by the resolutions of the Central Committee of the Bolshevik Communist Party concerning ideological matters and by your instructions, Comrade Stalin, the Soviet chemists wage war against the ideological concepts of bourgeois science. The lie of the so-called "resonance theory" has been disclosed, and the remains of this idea will be thrown away from Soviet chemistry. We wish you, our dear Leader and Teacher, good health and many, many years of famous life to the joy and happiness of the whole of progressive mankind [...]."

The events connected with the theory of resonance started in the autumn of 1950 at Moscow University. Quantum chemistry lecturers Yakov Kivovitch Syrkin and Mirra Yefimovna Diatkina were attacked. The accusation was about diffusion of the theory of resonance and was launched by former assistants of Syrkin. Since everything was in the hands of the professionals, Syrkin and Diatkina confessed guilty with respect to each of the charges.

2.10 Configuration interaction (CI) method

In the configuration interaction (CI) method[58]

> the variational wave function is a linear combination of Slater determinants constructed from *molecular* spin orbitals (Eq. (2.42)): $\psi = \sum_{I=0}^{M} c_I \Phi_I$.

In most cases we are interested in the function ψ for *the electronic ground state of the system*. In addition when solving the CI equations we also get approximations to the excited states with different values of the c_I coefficients.

Generally we construct the Slater determinants Φ_I by placing electrons on the molecular spin orbitals obtained with the Hartree–Fock method,[59] in most cases the set of determinants is additionally limited by imposing an upper bound for the orbital energy. In that case, the expansion in Eq. (2.42) is finite. The Slater determinants Φ_I are obtained by the replacement of occupied spin orbitals with virtual ones in the single Slater determinant, which is – in most cases – the Hartree–Fock function (Φ_0, i.e., ψ_{RHF}). When one spin orbital is replaced,

---

58 Also called the method of superposition of configurations.

59 In this method we obtain M molecular orbitals, i.e., $2M$ molecular spin orbitals, where M is the number of atomic orbitals employed. The Hartree–Fock determinant Φ_0 is the best form of wave function as long as the electronic correlation is not important. The criterion of this "goodness" is the mean value of the Hamiltonian. If we want to include the electron correlation, we may consider another form of the one-determinantal function more suitable as the starting point. We do not change our definition of correlation energy, i.e., we consider the RHF energy as that which does not contain any correlation effects. For instance, we may ask which of the normalized single-determinant functions Φ is closest to the normalized exact function ψ. As a measure of this we might use

$$|\langle \psi | \Phi \rangle| = \text{maximum}. \tag{2.43}$$

The single-determinantal function $\Phi = \Phi_B$, which fulfills the above condition, is called a Brueckner function (O. Sinanoğlu, K.A. Brueckner, "*Three Approaches to Electron Correlation in Atoms*," Yale Univ. Press, New Haven and London, 1970).

the resulting determinant is called singly excited, when two are replaced, doubly excited, etc.[60,61]

The virtual spin orbitals form an orthonormal basis in *the virtual space*. If we carry out any nonsingular linear transformation (cf. p. V1-547) of virtual spin orbitals, each "new" n-tuply excited Slater determinant becomes a linear combination of all "old" n-tuply excited determinants and only n-tuply excited ones.[62] In particular, the unitary transformation would preserve the mutual orthogonality of the n-tuply excited determinantal functions.

Thus, the total wave function (2.42) is a linear combination of the *known* Slater determinants (we assume that the spin orbitals are always known) with *unknown* c coefficients.

The name of the CI methods refers to the linear combination of the configurations rather than to the Slater determinants.

A *configuration* (CSF, i.e., configuration state function) is a linear combination of determinants which is an eigenfunction of the operators $\hat{S}^2$ and $\hat{S}_z$ and belongs to the proper irreducible representation of the symmetry group of the Hamiltonian. We say that this is a linear combination of the (spatial and spin) symmetry-adapted determinants. Sometimes we refer to the spin-adapted configurations which are eigenfunctions only of the $\hat{S}^2$ and $\hat{S}_z$ operators.

---

60 In the language of the second quantization (see Appendix C, p. 587) the wave function in the CI method has the form (the Φ_0 function is a Slater determinant which does not necessarily need to be a Hartree–Fock determinant)

$$\psi = c_0\Phi_0 + \sum_{a,p} c_p^a \hat{p}^\dagger \hat{a}\Phi_0 + \sum_{a<b,p<q} c_{pq}^{ab} \hat{q}^\dagger \hat{p}^\dagger \hat{a}\hat{b}\Phi_0 + \text{higher excitations}, \tag{2.44}$$

where c are the expansion coefficients, the creation operators $\hat{q}^\dagger, \hat{p}^\dagger, \ldots$ refer to the virtual spin orbitals $\phi_p, \phi_q, \ldots$, and the annihilation operators $\hat{a}, \hat{b}, \ldots$ refer to occupied spin orbitals $\phi_a, \phi_b, \ldots$ (the operators are denoted with the same indices as spin orbitals but the former are equipped with hat symbols), and the inequalities satisfied by the summation indices ensure that the given Slater determinant occurs only once in the expansion.

61 The Hilbert space corresponding to N electrons is the sum of the orthogonal subspaces $\Omega_n, n = 0, 1, 2, \ldots N$, which are spanned by the n-tuply excited (orthonormal) Slater determinants. Elements of the space Ω_n are all linear combinations of n-tuply excited Slater determinants. It does not mean, of course, that each element of this space is an n-tuply excited Slater determinant. For example, the sum of two doubly excited Slater determinants is a doubly excited Slater determinant only when one of the excitations is common to both determinants.

62 Indeed, the Laplace expansion (Appendix V1-A) along the row corresponding to the first new virtual spin orbital leads to the linear combination of the determinants containing new (*virtual, which means that the rank of excitation is not changed by this*) orbitals in this row. Continuing this procedure with the Slater determinants obtained, we finally get a linear combination of n-tuply excited Slater determinants expressed in old spin orbitals.

The particular terms in the CI expansion may refer to the respective CSFs or to the Slater determinants. Both versions lead to the same results, but using CSFs may be more efficient if we are looking for a wave function which transforms itself according to a single irreducible representation.

Next, this problem is reduced to the Ritz method (see Appendices V1-M, p. V1-705, and V1-L, p. V1-703), and subsequently to the secular equations $(\mathbf{H} - \epsilon\mathbf{S})\mathbf{c} = \mathbf{0}$. It is worth noting here that, e.g., the CI wave function for the ground state of the helium atom would be linear combinations of the determinants where the largest c coefficient occurs in front of the Φ_0 determinant constructed, e.g., from the spin orbitals $1s\alpha$ and $1s\beta$, but the nonzero contribution would also come from the other determinants constructed from the $2s\alpha$ and $2s\beta$ spin orbitals (one of the doubly excited determinants). The CI wave functions for all states (ground and excited) are linear combinations of *the same Slater determinants*; they differ only in the c coefficients.

The state energies obtained from the solution of the secular equations always approach the exact values from above.

2.10.1 Brillouin theorem

In the CI method we have to calculate matrix elements H_{IJ} of the Hamiltonian.

The Brillouin theorem says that

$$\langle \Phi_0 | \hat{H} \Phi_1 \rangle = 0 \tag{2.45}$$

if Φ_0 is a solution of the Hartree–Fock problem ($\Phi_0 \equiv \psi_{RHF}$), and Φ_1 is a singly excited Slater determinant in which the spin orbital $\phi_{i'}$ is orthogonal to all spin orbitals used in Φ_0.

Proof:

From Slater–Condon rule II (Appendix V1-N, p. V1-707) we have

$$\langle \Phi_0 | \hat{H} \Phi_1 \rangle = \left\langle i | \hat{h} i' \right\rangle + \sum_j \left[\langle ij | i'j \rangle - \langle ij | ji' \rangle \right]. \tag{2.46}$$

On the other hand, considering the integral $\langle i | \hat{F} i' \rangle$, where $\hat{F}$ is a Fock operator, and using the definition of the Coulomb and exchange operators from p. V1-474, from (V1-8.28) we obtain

$$\langle i|\hat{F}i'\rangle = \left\langle i|\hat{h}i'\right\rangle + \sum_j \left[\left\langle i|\hat{J}_j i'\right\rangle - \left\langle i|\hat{K}_j i'\right\rangle\right] = \left\langle i|\hat{h}i'\right\rangle + \sum_j \left[\langle ij|i'j\rangle - \langle ij|ji'\rangle\right] = \langle\Phi_0|\hat{H}\Phi_1\rangle.$$

From the Hermitian character of $\hat{F}$ it follows that

$$\langle i|\hat{F}i'\rangle = \langle\hat{F}i|i'\rangle = \varepsilon_i\delta_{ii'} = 0. \tag{2.47}$$

We have proved the theorem.

The Brillouin theorem is sometimes useful in discussions of the importance of particular terms in the CI expansion for the ground state.

2.10.2 Convergence of the CI expansion

Increasing the number of expansion functions by adding a new function lowers the energy (due to the variational principle). It often happens that the inclusion of only two determinants gives qualitative improvement with respect to the Hartree–Fock method; however, when going further, the situation becomes more difficult. The convergence of the CI expansion is slow, i.e., to achieve a good approximation to the wave function, the number of determinants in the expansion must usually be large. Theoretically, the shape of the wave function ensures solution of the Schrödinger equation $H\psi = E\psi$, but in practice we are *always* limited by the basis of the atomic orbitals employed.

To obtain satisfactory results, we need to increase the number M of atomic orbitals in the basis. The number of molecular orbitals produced by the Hartree–Fock method is also equal to M, hence the number of spin orbitals is equal to $2M$. In this case, the number of all determinants is equal to $\binom{2M}{N}$, where N refers to the number of electrons.

2.10.3 Example of H_2O

We are interested in the ground state of the water molecule, which is a singlet state ($S = 0, M_S = 0$).

The minimal basis set, composed of seven atomic orbitals (two $1s$ orbitals of the hydrogen atoms, $1s, 2s$, and three $2p$ orbitals of the oxygen atom), is considered too poor; therefore we prefer what is called a *double dzeta basis*, which provides two functions with different exponents for each orbital of the minimal basis. This creates a basis of $M = 14$ atomic orbitals. There are 10 electrons, hence $\binom{28}{10}$ gives 13 million Slater determinants. For a matrix of that size to be diagonalized is certainly impressive. Even more impressive is that we achieve only *an approximation* to the correlation energy which amounts to about 50% of the exact correlation

energy,[63] since M is only equal to 14, but in principle it should be equal to ∞. Nevertheless, for *comparative* purposes we assume the correlation energy obtained is 100%.

The simplest remedy is to get rid of some determinants in such a way that the correlation energy is not damaged. Which ones? Well, many of them correspond to the incorrect projection S_z of the total spin or the incorrect total spin S. For instance, we are interested in the singlet state (i.e., $S = 0$ and $S_z = 0$), but some determinants are built of spin orbitals containing exclusively α spin functions. This is a pure waste of resources, since the nonsinglet functions do not make any contributions to the singlet state. When we remove these and other *incorrect* determinants, we obtain a smaller matrix to be diagonalized. The number of Slater determinants with $S_z = 0$ is equal to $\binom{M}{N/2}^2$. In our case, this makes slightly over 4 million determinants (instead of 13 million). What would happen if we diagonalized the huge original matrix anyway? Well, nothing would happen. There would be more work, but the computer would create *the block form*[64] from our enormous matrix, and each would correspond to the particular S^2 and S_z, *while the whole contribution to the correlation energy of the ground state comes from the block corresponding to* $S = 0$ and $S_z = 0$.

Let us continue throwing away determinants. This time, however, we have to make a compromise, i.e., some of the Slater determinants are arbitrarily considered not to be important (which will worsen the results, if they are rejected). Which of the determinants should be considered as not important? The general opinion in quantum chemistry is that the multiple excitations are less and less important (when the multiplicity increases). If we take only the singly, doubly, triply, and quadruply excited determinants, the number of determinants will reduce to 25 000 and we will obtain 99% of the approximate correlation energy defined above. If we take the singly and doubly excited determinants only, there are only 360 of them, and 94% of the correlation effect is obtained. This is why this *CI singles and doubles* (CISD) method is used so often.

For larger molecules this selection of determinants becomes too demanding; therefore, we have to decide individually for each configuration: to include or reject it? The decision is made either on the basis of the perturbational estimate of the importance of the determinant[65] or by a test calculation with inclusion of the determinant in question (Fig. 2.10).

To obtain good results, we need to include a large number of determinants, e.g., of the order of thousands, millions, or even billions. This means that contemporary quantum chemistry has

63 We see here how vicious the dragon of electron correlation is.

64 These square blocks would be easily noticed after proper ordering of the expansion functions.

65 The perturbational estimate mentioned relies on the calculation of the weight of the determinant based on the first-order correction to the wave function in perturbation theory (p. V1-279). In such an estimate the denominator contains the excitation energy evaluated as the difference in orbital energies between the Hartree–Fock determinant and the one in question. In the numerator there is a respective matrix element of the Hamiltonian calculated with the help of the known Slater–Condon rules (Appendix V1-N, p. V1-707).

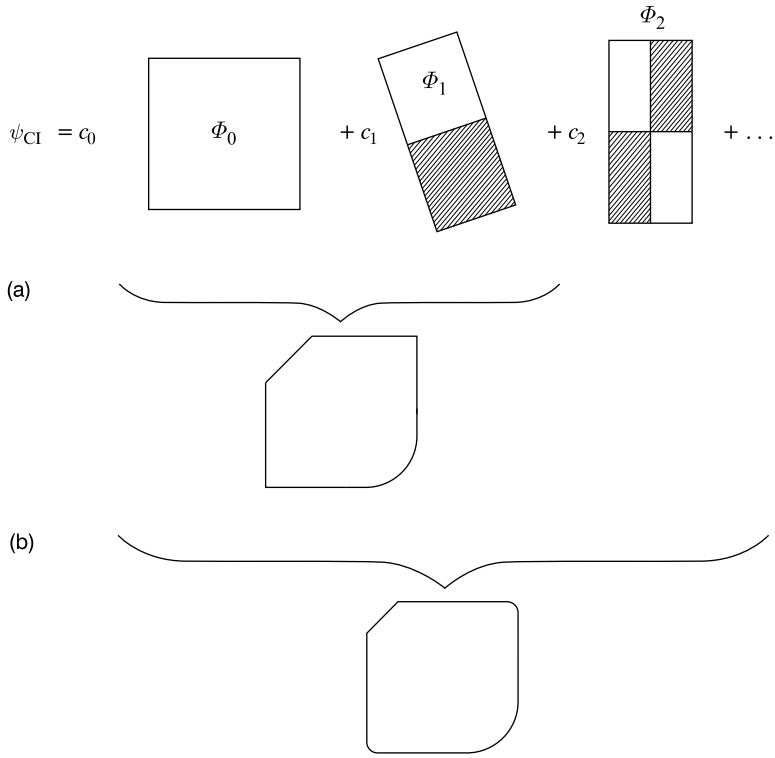


Fig. 2.10. Symbolic illustration of the principle of the CI method with one Slater determinant ψ_0 dominant in the ground state (this is a problem of the many-electron wave function; the picture cannot be understood literally). The purpose of this diagram is to emphasize a relatively small role of electronic correlation (more exactly, of what is known as the dynamical correlation, i.e., correlation of electronic motion). The function ψ_{CI} is a linear combination (the c coefficients) of the determinantal functions of different shapes in the many-electron Hilbert space. The shaded regions correspond to the negative sign of the function; the nodal surfaces of the added functions allow for the effective deformation of ψ_0 to have lower and lower average energy. (a) Since c_1 is small in comparison to c_0, the result of the addition of the first two terms is a slightly deformed ψ_0. (b) Similarly, the additional excitations just make cosmetic changes in the function (although they may substantially affect the quantities calculated with it).

made enormous technical progress.[66] This, however, is a sign, not of the strength of quantum chemistry, but of its weakness. What are we going to do with such a function? We may load it back into the computer and calculate all the properties of the system with high accuracy (although this cannot be guaranteed). To answer a student's question about why we obtained some particular numbers, we have to answer that we do not know, it is the computer which

[66] To meet such needs, quantum chemists have had to develop entirely new techniques of applied mathematics.

knows. This is a trap. It would be more instructive to have, say, two Slater determinants, which describe the system to a reasonable approximation and we can understand what is going on in the molecule.

2.10.4 Which excitations are most important?

The convergence can be particularly bad if we use the virtual spin orbitals obtained by the Hartree–Fock method. Not all excitations are equally important. It turns out that usually, although this is not a rule, low excitations dominate the ground-state wave function.[67] The single excitations *themselves* do not contribute anything to the ground state *energy* (if the spin orbitals are generated with the Hartree–Fock method, then the Brillouin theorem mentioned above applies). *They are crucial, however, for excited states or in dipole moment calculations.* For the ground state, only when coupled to other types of excitation do they assume nonzero (although small) values. Indeed, if in the CI expansion we only use the Hartree–Fock determinant and the determinants corresponding to single excitations, then, due to the Brillouin theorem, the secular determinant would be factorized.[68] This factorization (Fig. 2.11) pertains to the single-element determinant corresponding to the Hartree–Fock function and to the determinants corresponding exclusively to single excitations. Since we are interested in the ground state, only the first determinant is of importance to us, and it does not change whether we include or not a contribution coming from single excitations into the wave function.

Usually, performing CI calculations with the inclusion of all excitations (for the assumed value of M), i.e., the FCI (full CI), is not possible in practical calculations due to the extremely long expansion. We are forced to truncate the CI basis somewhere. It would be good to terminate it in such a way that all *essential* (the problem is what we mean by essential though) terms are retained. *The most significant terms for the correlation energy come from the double excitations, since these are the first excitations coupled to the Hartree–Fock function.* Smaller, although important, contributions come from other excitations (usually of low excitation rank). We certainly wish that it would be like this for large molecules. Nobody knows what the truth is.

2.10.5 Natural orbitals (NOs) – a way to shorter expansions

A fast convergence is achieved in the basis set of *natural orbitals* (NOs), i.e., when we construct spin orbitals with *these* orbitals and from them the Slater determinants. The NO is defined *a*

[67] That is, requiring the lowest excitation energies. Later, a psychological mechanism began to work supported by economics: the *high-energy* excitations are numerous and, because of that, very expensive and they correspond to high *excitation ranks* (the number of electrons excited). Due to this, a reasonable restriction for the number of configurations in the CI expansion is excitation rank. We will come back to this problem later.

[68] That is, could be written out in block form, which would separate the problem into several subproblems of smaller size.

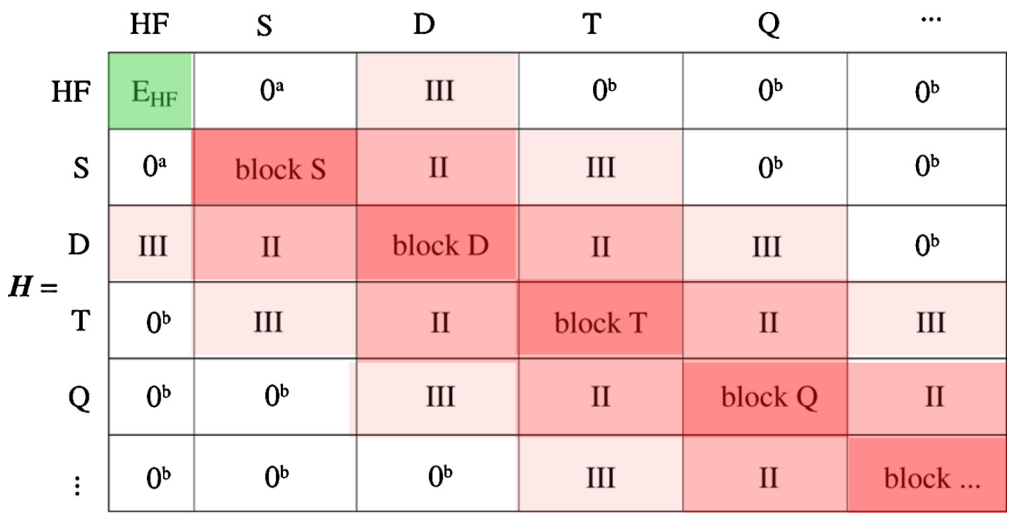


Fig. 2.11. The block structure of the Hamiltonian matrix (**H**) is the result of the Slater–Condon rules (Appendix V1-N, p. V1-707). S, single excitations; D, double excitations; T, triple excitations; Q, quadruple excitations. (a) A block of zero values due to the Brillouin theorem. (b) The block of zero values due to Slater–Condon rule IV. (II) The nonzero block obtained according to Slater–Condon rule II. (III) The nonzero block obtained according to Slater–Condon rule III. In real calculations all the nonzero blocks are sparse matrices dominated by zero values, which is important in the diagonalization process.

posteriori in the following way. After carrying out the CI calculations, we construct the density ρ (see Appendix D, p. 595) as follows:

$$\begin{aligned}\rho(1) &= N\int \psi^*(1,2,3,\ldots N)\psi(1,2,3,\ldots,N)\mathrm{d}\tau_2\mathrm{d}\tau_3\ldots\mathrm{d}\tau_N \\ &= \sum_{ij} D_{ji}\phi_i^*(1)\phi_j(1), \quad D_{ij} = D_{ji}^*,\end{aligned} \tag{2.48}$$

where the summation runs over all the spin orbitals. By diagonalization of matrix **D** (a rotation in the Hilbert space spanned by the spin orbitals) we obtain the density expressed in the natural spin orbitals (NOs) transformed by the unitary transformation

$$\rho(1) = \sum_i (D_{diag})_{ii}\phi_i'^*(1)\phi_i'(1). \tag{2.49}$$

The most important ϕ_i' from the viewpoint of the correlation are the NOs with large *occupancies*, i.e., $(D_{diag})_{ii}$ values. Inclusion of only the most important ϕ_i' in the CI expansion creates a short and quite satisfactory wave function.[69]

[69] Approximate natural orbitals can also be obtained directly without performing the CI calculations.

2.10.6 Size inconsistency of the CI expansion

A truncated CI expansion violates the size-consistency.

Let us imagine we want to calculate the interaction energy of two beryllium atoms. Let us suppose that we decide that to describe the beryllium atom we have to include not only the $1s^2 2s^2$ configuration, but also the doubly excited one, $1s^2 2p^2$. In the case of beryllium, this is a very reasonable step, since both configurations have close energies. Let us assume now that we calculate the wave function for *two* beryllium atoms. If we want this function to describe the system correctly, also at large interatomic distances, we have to make sure that the departing atoms have appropriate excitations at their disposition, i.e., in our case $1s^2 2p^2$ for each. To achieve this we *must incorporate quadruple excitations into the method*.[70]

> If we include quadruples, we have a chance to achieve (an approximate) size consistency, i.e., the energy will be proportional to good accuracy to the number of atoms; otherwise our results will not be size-consistent.

Let us imagine 10 beryllium atoms. In order to have size consistency we need to include 20-fold excitations. This would be very expensive. We clearly see that, for many systems, the size consistency requires inclusion of multiple excitations. If we carried out CI calculations for all possible (for a given number of spin orbitals) excitations, such a FCI method (i.e., *Full CI*) would be size-consistent.

2.11 Direct CI method

We have already mentioned that the CI method converges slowly. Due to this, the Hamiltonian matrices and overlap integral matrices are sometimes so large that they cannot fit into the computer memory. In practice, such a situation occurs in all high-quality calculations for small systems and in all calculations for medium and large systems. Even for a quite large atomic orbital basis, the number of integrals is much smaller than the number of Slater determinants in the CI expansion.

Björn Roos[71] first noticed that to find the lowest eigenvalues and their eigenvectors we do not need to store a huge $\mathbf{H}$ matrix in the computer memory. Instead, we need to calculate the *residual vector* $\boldsymbol{\sigma} = (\mathbf{H} - E\mathbf{1})\mathbf{c}$, where $\mathbf{c}$ is a trial vector (defining the trial function in the variational method, p. V1-265). If $\boldsymbol{\sigma} = \mathbf{0}$, it means that the solution is found. Knowing $\boldsymbol{\sigma}$, we

[70] See J.A. Pople, R. Seeger, R. Krishnan, *Intern. J. Quantum Chem.*, *S*11(1977)149; also p. 47 of the book by P. Jørgensen and J. Simons "*Second Quantization-Based Methods in Quantum Chemistry*," Acad. Press., 1981.

[71] B.O. Roos, *Chem. Phys. Letters*, 15(1972)153.

may find (on the basis of first-order perturbation theory) a slightly improved **c**, etc. The product **Hc** can be obtained by going through the set of integrals and assigning to each a coefficient resulting from **H** and **c**, and adding the results to the new **c** vector. Then the procedure is repeated. Until 1971, CI calculations with 5000 configurations were considered a significant achievement. After Roos's paper, there was a leap of several orders of magnitude, bringing the number of configurations to the range of billions. For the computational method this was a revolution.

2.12 Multireference CI method

Usually in the CI expansion, the dominant determinant is the Hartree–Fock one. We construct the CI expansion, replacing the spin orbitals in this determinant (*single reference method*). We can easily imagine a situation in which taking one determinant is not justified, since the shell is not well closed (e.g., four hydrogen atoms). We already know that certain determinants (or, in other words, configurations) need to be present ("static correlation") in the correct wave function. To be sure, we are the judges, deciding which is good or bad. This set of determinants is a basis in the *model space*.

In the single reference CI method, the model space (Fig. 2.12) is formed by a single Slater determinant. In the multireference CI method, the set of determinants constitute the model space. This time, the CI expansion is obtained by replacement of the spin orbitals participating in the model space by other virtual orbitals. We proceed further as in CI.

There is no end to the problems yet, since again we have billions of possible excitations.[72] We do other tricks to survive in this situation (Fig. 2.12). We may, for instance, get the idea not to excite the inner shell orbitals, since the numerical effort is serious, the lowering of the total energy can also be large, but the effect on the energy *differences* (this is what chemists are usually interested in) is negligible. We say that such orbitals are *frozen*. Some of the orbitals are kept doubly occupied in all Slater determinants but we optimize their shape. Such orbitals

[72] There is another trouble too, called *intruder states*, i.e., states which are of unexpectedly low energy. How could these states appear? Firstly, the CI states known as "front door intruders" appear, if some important (low-energy) configurations were for some reason not included into the model space. Secondly, we may have the "back door intruder" states, when the energy gap between the model space and the other configurations were too small (quasidegeneracy), and some CI states became of low energy (enter the model space energy zone) even if they are not composed of the model space configurations.

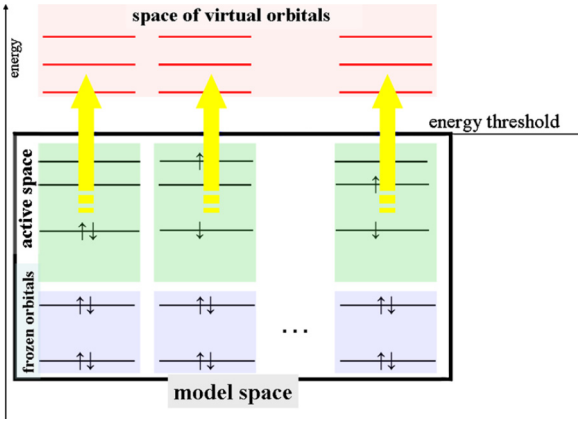


Fig. 2.12. Illustration of the model space in the multireference CI method used mainly in the situation when no single Slater determinant dominates the CI expansion. In the figure the orbital levels of the system are presented. Part of them are occupied in all Slater determinants considered ("frozen spin orbitals"). Above them is a region of closely spaced orbital levels called *active space*. In the optimal case, a significantly large energy gap occurs between the latter and unoccupied levels lying higher. The model space is spanned by all or some of the Slater determinants obtained by various occupancies of the active space levels.

are called *inactive*. Finally, the orbitals of varied occupancy in different Slater determinants are called *active*. The frozen orbitals are, in our method, important spectators of the drama, the inactive orbitals contribute a little towards lowering the energy, but the most efficient work is done by the active orbitals.

2.13 Multiconfigurational self-consistent field (MC SCF) method

In the CI method, it is sometimes obvious that certain determinants of the CI expansion *must* contribute to the wave function, if the latter is to correctly describe the system. For example, if we want to describe the system in which a bond is being broken (or is being formed), for its description we need several determinants for sure (cf. description of the dissociation of the hydrogen molecule on p. V1-511).

Why is this? In the case of the dissociation, with which we are dealing here, there is a quasidegeneracy of the bonding and the antibonding orbital of the bond in question, i.e., the approximate equality of their energies (the bond energy is of the order of the overlap integral and the latter goes to zero when the bond is being broken). The determinants, which can be constructed

by various occupancies of these orbitals, have very close energies and, consequently, their contributions to the total wave function are of similar magnitude and *should be included* in the wave function.

In the MC SCF method, as in CI, it is *up to us* to decide which set of determinants we consider sufficient for the description of the system.

Each of the determinants is constructed from molecular spin orbitals which are not fixed (as in the CI method) but are modified in such a way as to have the total energy as low as possible.

The MC SCF method is the most general scheme of the methods that use a linear combination of Slater determinants as an approximation to the wave function. In the limiting case of the MC SCF, when the number of determinants is equal to 1, we have, of course, the Hartree–Fock method.

2.13.1 Classical MC SCF approach

We will describe first the classical MC SCF approach. This is a variational method. As was mentioned, the wave function in this method has the form of a finite linear combination of Slater determinants Φ_I,

$$\psi = \sum_I d_I \Phi_I, \tag{2.50}$$

where d are variational coefficients.

In the classical MC SCF method we:

1. take a finite CI expansion (the Slater determinants and the orbitals for their construction are fixed),
2. calculate the coefficients for the determinants by the Ritz method (the orbitals do not change),
3. vary the LCAO coefficients in the orbitals at the fixed CI coefficients to obtain the best molecular orbitals, and
4. return to point 1 until self-consistency is achieved.

2.13.2 Unitary MC SCF method

Another version of the MC SCF problem, a *unitary method* suggested by Lévy and Berthier[73] and later developed by Daalgaard and Jørgensen,[74] is gaining increasing importance. The eigenvalue problem does not appear in this method.

We need two mathematical facts to present the unitary MC SCF method. The first is the following theorem.

If $\hat{A}$ is a Hermitian operator, i.e., $\hat{A}^\dagger = \hat{A}$, then $\hat{U} = \exp(i\hat{A})$ is a unitary operator satisfying $\hat{U}^\dagger \hat{U} = 1$.

Let us see how $\hat{U}^\dagger$ looks. We write

$$\begin{aligned}\hat{U}^\dagger = \Big(\exp(i\hat{A})\Big)^\dagger &= \left(1 + i\hat{A} + \frac{1}{2!}(i\hat{A})^2 + \frac{1}{3!}(i\hat{A})^3 + \ldots\right)^\dagger \\ &= \left(1 + (-i)\hat{A}^\dagger + \frac{1}{2!}(-i\hat{A}^\dagger)^2 + \frac{1}{3!}(-i\hat{A}^\dagger)^3 + \ldots\right) \\ &= \left(1 + (-i)\hat{A} + \frac{1}{2!}(-i\hat{A})^2 + \frac{1}{3!}(-i\hat{A})^3 + \ldots\right) = \exp(-i\hat{A}).\end{aligned}$$

Hence, $\hat{U}\hat{U}^\dagger = 1$, i.e., $\hat{U}$ is a unitary operator.[75]

Now follows the second mathematical fact. This is a commutator expansion:

$$e^{-\hat{A}}\hat{H}e^{\hat{A}} = \hat{H} + [\hat{H}, \hat{A}] + \frac{1}{2!}[[\hat{H}, \hat{A}], \hat{A}] + \frac{1}{3!}[[[\hat{H}, \hat{A}], \hat{A}], \hat{A}] + \ldots. \tag{2.51}$$

This theorem can be proved by induction, expanding the exponential functions.

Now we are all set to describe the unitary method. We introduce two new operators, i.e.,

$$\hat{\lambda} = \sum_{ij} \lambda_{ij} \hat{\imath}^\dagger \hat{\jmath}, \tag{2.52}$$

where $\hat{\imath}^\dagger$ and $\hat{\jmath}$ are the creation and annihilation operators, respectively, associated to spin orbitals i, j (see Appendix C), and

$$\hat{S} = \sum_{IJ} S_{IJ} |\Phi_I\rangle\langle\Phi_J|. \tag{2.53}$$

---

73 B. Lévy, G. Berthier, *Intern. J. Quantum Chem.*, 2(1968)307.

74 E. Dalgaard, P. Jørgensen, *J. Chem. Phys.*, 69(1978)3833.

75 Is an operator ($\hat{C}$) of multiplication by a constant Hermitian? Let us see: $\langle\varphi|\hat{C}\psi\rangle \stackrel{?}{=} \langle\hat{C}\varphi|\psi\rangle$; left-hand side $= \langle\varphi|c\psi\rangle = c\langle\varphi|\psi\rangle$; right-hand side $= \langle c\varphi|\psi\rangle = c^*\langle\varphi|\psi\rangle$. We have left-hand side = right-hand side if $c = c^*$. An operator conjugate to c is, therefore, c^*. Further, if $\hat{B} = i\hat{A}$, what is a form of $\hat{B}^\dagger$? We have $\langle\hat{B}^\dagger\varphi|\psi\rangle = \langle\varphi\hat{B}|\psi\rangle$. Then $\langle\varphi|i\hat{A}|\psi\rangle = \langle -i\hat{A}^\dagger\varphi|\psi\rangle$ and finally $\hat{B}^\dagger = -i\hat{A}^\dagger$.

We assume that λ_{ij} and S_{IJ} are elements of the Hermitian matrices[76] (their determination is the goal of the whole method), Φ_I are determinants from the MC SCF expansion (2.50).

It can be seen that the $\hat{\lambda}$ operator replaces a single spin orbital in a Slater determinant and forms a linear combination of such modified determinantal functions; the $\hat{S}$ operator replaces such a combination with another. The "knobs" which control these changes are coefficients λ_{ij} and S_{IJ}.

We will need the unitary transformations $\exp(i\hat{\lambda})$ and $\exp(i\hat{S})$. They are very convenient, since when starting from some set of the orthonormal functions (spin orbitals or Slater determinants) and applying this transformation, we always retain orthonormality of new spin orbitals (due to $\hat{\lambda}$) and of linear combinations of determinants (due to $\hat{S}$). This is an analogy to the rotation of the Cartesian coordinate system. It follows from the above equations that $\exp(i\hat{\lambda})$ modifies spin orbitals, i.e., operates in the one-electron space, and $\exp(i\hat{S})$ rotates the determinants in the space of many-electron functions.

Now we suggest the following form of our variational function:

$$|\tilde{0}\rangle = \exp(i\hat{\lambda})\exp(i\hat{S})|0\rangle, \tag{2.54}$$

where $|0\rangle$ denotes a starting combination of determinants with specific spin orbitals and the matrices $\boldsymbol{\lambda}$ and $\mathbf{S}$ contain the variational parameters. So, we modify the spin orbitals and change the coefficients in front of the determinants to obtain a new combination of the modified determinants, $|\tilde{0}\rangle$. The mean energy value for that function is[77]

$$E = \langle\tilde{0}|\hat{H}|\tilde{0}\rangle = \langle 0|\exp(-i\hat{S})\exp(-i\hat{\lambda})\hat{H}\exp(i\hat{\lambda})\exp(i\hat{S})|0\rangle. \tag{2.55}$$

Taking advantage of the commutator expansion (2.51), we have

$$E = \langle 0|\hat{H}|0\rangle - i\langle 0|[\hat{S}+\hat{\lambda},\hat{H}]|0\rangle + \frac{1}{2}\langle 0|[\hat{S},[\hat{H},\hat{S}]]|0\rangle + \frac{1}{2}\langle 0|[\hat{\lambda},[\hat{H},\hat{\lambda}]]|0\rangle + \\ + \langle 0|[\hat{S},[\hat{H},\hat{\lambda}]]|0\rangle + \ldots.$$

It follows from the last equation that in order to calculate E, we have to know the result of the operation of $\hat{\lambda}$ on $|0\rangle$, i.e., on the linear combination of determinants, which comes down to the operation of the creation and annihilation operators on the determinants, which is simple. It

[76] Considering the matrix elements of the operators $\hat{\lambda}$ and $\hat{S}$, we would easily be convinced that both operators are also Hermitian.

[77] Here we use the equality $[\exp(i\hat{A})]^{\dagger} = \exp(-i\hat{A})$.

can also be seen that we need to apply the operator $\hat{S}$ to $|0\rangle$, but its definition shows that this is trivial. This expression[78] can now be optimized, i.e., the best Hermitian matrices $\boldsymbol{\lambda}$ and $\mathbf{S}$ can be selected. It is done in the same step (this distinguishes the current method from the classical one). Usually the calculations are carried out in the matrix form, neglecting the higher terms and retaining only the quadratic ones in $\hat{S}$ and $\hat{\lambda}$. Neglecting the higher terms is equivalent to allowing for very small rotations in the transformation (2.54), but instead we have a large number of rotations (iterative solution).[79]

The success of the method depends on the starting point. The latter strongly affects the energy and its hypersurface (in the space of the parameters of the matrices $\boldsymbol{\lambda}$ and $\mathbf{S}$) is very complicated, it has many local minima. This problem is not yet solved, but various procedures accelerating the convergence are applied, e.g., the new starting point is obtained by averaging the starting points of previous iterations. The method also has other problems, since the orbital rotations partially replace the rotation in the space of the Slater determinants (the rotations do not commute and are not independent). In consequence, linear dependencies may appear.

2.13.3 Complete active space (CAS SCF) method is size-consistent

An important special case of the MC SCF method is the complete active space self-consistent field (CAS SCF, Fig. 2.13) of Roos, Taylor, and Siegbahn.[80] Let us assume that we are dealing with a closed shell molecule. The RHF method (p. V1-464) provides the molecular orbitals and the orbital energies. From them we select the low-energy orbitals. Part of them are *inactive*, i.e., are doubly occupied in all determinants, but they are varied, which results in lowering the mean value of the Hamiltonian (some of the orbitals may be frozen, i.e., kept unchanged). These are the spin orbitals corresponding to the inner shells. The remaining spin orbitals belong to the *active space*. Now we consider all possible occupancies and excitations of the active spin orbitals (this is where the adjective "*complete*" comes from) to obtain the set of determinants in the expansion of the MC SCF. By taking all possible excitations within the active space, we achieve a *size consistency*, i.e., when dividing the system into subsystems and separating them

[78] The term with i gives a real number,

$$i \cdot \langle 0|[\hat{S}+\hat{\lambda}, \hat{H}]|0\rangle = i \cdot \Big(\langle (\hat{S}+\hat{\lambda})0|\hat{H}0\rangle - \langle \hat{H}0|(\hat{S}+\hat{\lambda})0\rangle \Big) \rightarrow i \cdot (z - z^*) = i(2i\,\mathrm{Im}z) \in R,$$

with R a set of real numbers.

[79] In the classical MC SCF method when minimizing the energy with respect to the parameters, we use only linear terms in the expansion of the energy with respect to these parameters. In the unitary formulation we use both linear and quadratic terms. This implies much better convergence of the unitary method.

[80] B.O. Roos, P.E.M. Siegbahn, in "*Modern Theoretical Chemistry*," vol. III, ed. H.F. Schaefer, Plenum Press, New York, 1977; P.E.M. Siegbahn, *J. Chem. Phys.*, 70(1979)5391; B.O. Roos, P.R. Taylor, P.E.M. Siegbahn, *Chem. Phys.*, 48(1980)157.

(infinite distances) we obtain the sum of the energies calculated for each subsystem separately. By taking the complete set of excitations we also achieve that the results do not depend on any (nonsingular) linear transformation of the molecular spin orbitals within the given subgroup of orbitals, i.e., within the inactive or active spin orbitals. This makes the result invariant with respect to the localization of the molecular orbitals.

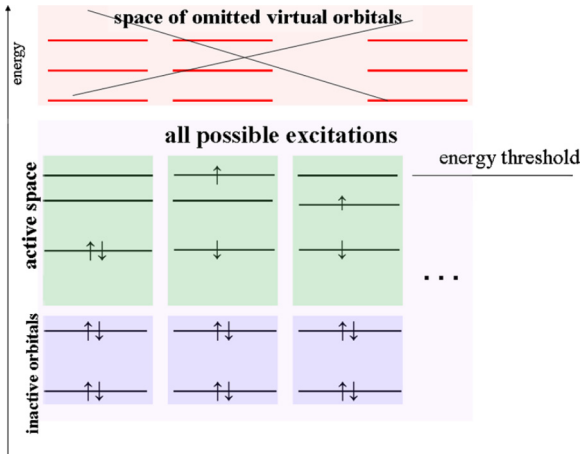


Fig. 2.13. CAS SCF, a method of construction of the Slater determinants in the MC SCF expansion. The inner shell orbitals are usually inactive, i.e., are doubly occupied in each Slater determinant (some of them may be frozen, i.e., their form does not change). From the active space plus inactive spin orbitals we create the complete set of possible Slater determinants to be used in the MC SCF calculations. The spin orbitals of the energy higher than a certain selected threshold are entirely ignored in the calculations.

NONVARIATIONAL METHODS WITH SLATER DETERMINANTS

2.14 Coupled cluster (CC) method

The CC method represents now the most reliable one among quantum mechanical methods applied to chemistry.

The problem of many-body correlation of motion of anything is extremely difficult and so far unresolved (e.g., weather forecasting). The problem of electron correlation also seemed to be hopelessly difficult. It still remains so; however, it turns out that we can exploit a certain ob-

servation made by Sinanoğlu.[81] This author noticed that the major portion of the correlation is included through the introduction of correlation within electron pairs, next through pair–pair interactions, then pair–pair–pair interactions, etc. The canonical molecular spin orbitals, which we can use, are in principle delocalized over the whole molecule, but practically the delocalization is not so large. Even in the case of canonical spin orbitals, and certainly when using localized molecular spin orbitals, we can think about an electron excitation as a transfer of an electron from one place in the molecule to another. Inclusion of the correlation of electronic motion represents, in the language of electron excitations, the following philosophy: when electron 1 jumps from an orbital localized in place A to an orbital localized in place B, it would be good – from the point of view of the variational principle – if electron 2 jumped from the orbital localized at C to the orbital localized at D (Fig. 2.14).

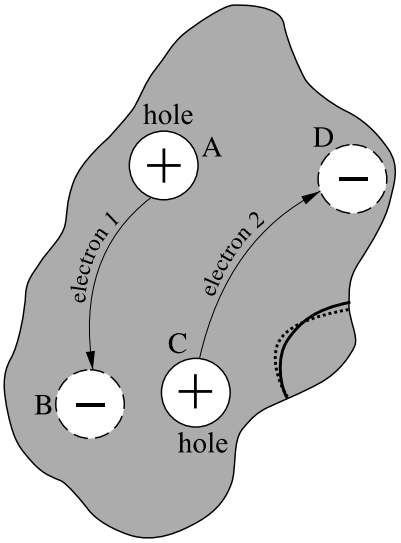


Fig. 2.14. In order to include the electron correlation, the wave function should somehow reflect the fact that electrons avoid each other. Electron 1 jumping from A (an orbital) to B (another orbital) should make electron 2 escape from C (close to B) to D (close to A). This is the very essence of electron correlation. The other orbitals play a role of spectators. However, the spectators change upon the excitations described above. These changes are performed by allowing their own excitations (symbolized by changing from the solid line to the dashed line on the right-hand side of the figure). This is how triple, quadruple, and higher excitations contribute to electronic correlation.

The importance of a given double excitation depends on the energy connected with the electron relocation and the arrangement of points A, B, C, and D. Yet this simplistic reasoning suggests

[81] O. Sinanoğlu, K.A. Brueckner, "*Three Approaches to Electron Correlation in Atoms*," Yale Univ. Press, New Haven and London, 1970.

single excitations do not carry any correlation (this is confirmed by the Brillouin theorem) and this is why their role is very small (in the ground state). Moreover, it suggests that double excitations should be very important.

2.14.1 Wave and cluster operators

First we introduce a special Slater determinant, the *reference determinant* (called the *vacuum state*, it can be the Hartree–Fock determinant) Φ_0 and we write that the exact wave function for the ground state is

$$\psi = \exp(\hat{T})\Phi_0, \tag{2.56}$$

where the unknown $\exp(\hat{T})$ is a *wave operator* and $\hat{T}$ itself is a *cluster operator*. In the CC method *an intermediate normalization*[82] of the function ψ is assumed, i.e.,

$$\langle\psi|\Phi_0\rangle = 1.$$

Eq. (2.56) represents a very ambitious task. It assumes that we will find an operator $\hat{T}$ such that the wave operator $(e^{\hat{T}})$, as with the touch of a wizard's wand, will make an ideal solution of the Schrödinger equation from the Hartree–Fock function. The formula with $\exp(\hat{T})$ is an *Ansatz*. The charming sounding word Ansatz[83] can be translated as arrangement or order, but in mathematics it refers to the construction assumed.

One encounters sometimes the argument that the wave operator ensures the size consistency of the CC. According to this reasoning, for an infinite distance between molecules A and B, both ψ and Φ_0 functions can be expressed in the form of the product of the wave functions for A and B. When the cluster operator is assumed to be of the form (obvious for infinitely separated systems) $\hat{T} = \hat{T}_A + \hat{T}_B$, then the exponential form of the wave operator $\exp(\hat{T}_A + \hat{T}_B)$ ensures a desired form of the product of the wave function $[\exp(\hat{T}_A + \hat{T}_B)]\Phi_0 = \exp\hat{T}_A \exp\hat{T}_B\Phi_0$. *If we took a finite CI expansion*, $(\hat{T}_A + \hat{T}_B)\Phi_0$, *then we would not get the product but the sum, which is incorrect*. In this reasoning there is an error, since due to the Pauli principle (antisymmetry of the wave function with respect to the electron exchange) for long distance neither the function ψ nor the function Φ_0 is the product of the functions for the subsystems.[84] Although the reasoning is not quite correct, the conclusion is correct, as will be shown at the end of the description of the CC method (p. 155).

---

82 It contributes significantly to the numerical efficiency of the method.

83 This word has survived in the literature in its original German form.

84 For instance, the RHF function for the hydrogen molecule is not a product function for long distances (see p. 126).

The CC method is automatically size-consistent.

As a cluster operator $\hat{T}$ we assume a sum of the excitation operators (see Appendix C)

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... + \hat{T}_{l_{\max}}, \tag{2.57}$$

where

$$\hat{T}_1 = \sum_{a,r} t_a^r \hat{r}^\dagger \hat{a} \tag{2.58}$$

is an operator for single excitations,

$$\hat{T}_2 = \frac{1}{4} \sum_{\substack{ab \\ rs}} t_{ab}^{rs} \hat{s}^\dagger \hat{r}^\dagger \hat{a} \hat{b} \tag{2.59}$$

is an operator for double excitations, etc. The subscript $l = 1, 2, ..., l_{\max}$ in $\hat{T}_l$ indicates the rank of the excitations involved (with respect to the vacuum state). The symbols $a, b, ...$ refer to the spin orbitals occupied in Φ_0, $p, q, r, s, ...$ refer to the virtual ones, and

t represents *amplitudes*, i.e., the numbers whose determination is the goal of the CC method. The rest of this chapter will be devoted to the problem how we can obtain these miraculous amplitudes.

In the CC method we want to obtain correct results with the assumption that $l_{\max}$ *of Eq.* (2.57) *is relatively small* (usually $2 \div 5$). If $l_{\max}$ were equal to N, i.e., to the number of electrons, then the CC method would be identical to the full (usually unfeasible) CI method (FCI).

2.14.2 Relationship between CI and CC methods

Obviously, there is a relation between the CI and CC methods. For instance, if we write $\exp(\hat{T})\Phi_0$ in such a way as to resemble the CI expansion, i.e.,

$$\begin{aligned} \exp(\hat{T})\Phi_0 &= [1 + (\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots) + \frac{1}{2}(\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots)^2 + \ldots]\Phi_0 \\ &= (1 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \ldots)\Phi_0, \end{aligned} \tag{2.60}$$

the operators $\hat{C}_i$ (index i denoting the excitation rank; $i = 1$ for singles, $i = 2$ for double, etc.), pertaining to the CI method, have the following structure:

$$\hat{C}_1 = \hat{T}_1,$$
$$\hat{C}_2 = \hat{T}_2 + \frac{1}{2!}\hat{T}_1^2,$$
$$\hat{C}_3 = \hat{T}_3 + \frac{1}{3!}\hat{T}_1^3 + \hat{T}_1\hat{T}_2,$$
$$\hat{C}_4 = \hat{T}_4 + \frac{1}{4!}\hat{T}_1^4 + \frac{1}{2!}\hat{T}_2^2 + \hat{T}_3\hat{T}_1 + \frac{1}{2!}\hat{T}_1^2\hat{T}_2, \tag{2.61}$$
$$\ldots \tag{2.62}$$

We see that the multiple excitations $\hat{C}_l$ result from mathematically distinct terms, e.g., $\hat{C}_3$ is composed of triple excitations $\hat{T}_3$, $\hat{T}_1^3$, and $\hat{T}_1\hat{T}_2$.

On the basis of current numerical experience,[85] we believe (Fig. 2.14) that within the excitation of a given rank, the contributions coming from the correlational interactions within the electron pairs are the most important, e.g., within C_4 the $\frac{1}{2!}\hat{T}_2^2$ excitations containing the product of amplitudes for two electron pairs are the most important, $\hat{T}_4$ (which contains the amplitudes of quadruple excitations) is of little importance, since they correspond to the coupling of the motions of four electrons, and the terms $\hat{T}_1$, $\hat{T}_3\hat{T}_1$, and $\hat{T}_1^2\hat{T}_2$ can be made small by using the MC SCF orbitals. Contemporary quantum chemists use diagrammatic language (following Richard Feynman). The point is that the mathematical terms (the energy contributions) appearing in CC theory can be translated – one by one – into the figures according to certain rules. It turns out that it is much easier to think in terms of diagrams than to speak about the mathematical formulae or to write them out. The CC method, terminated at $\hat{T}_2$ in the cluster operator, automatically includes $\hat{T}_2^2$, etc. We may see in it some resemblance to a group of something (excitations), or in other words to a cluster (Fig. 2.15).

2.14.3 Solution of the CC equations

The strategy of the CC method is the following. First, we make a decision with respect to l_{max} in the cluster expansion (2.57) ($l_{\max}$ should be small[86]).

The exact wave function $\exp(\hat{T})\Phi_0$ satisfies the Schrödinger equation, i.e.,

$$\hat{H}\exp(\hat{T})\Phi_0 = E\exp(\hat{T})\Phi_0, \tag{2.63}$$

which, after operating from the left with $\exp(-\hat{T})$, gives

$$\exp(-\hat{T})\hat{H}\exp(\hat{T})\Phi_0 = E\Phi_0. \tag{2.64}$$

[85] This is a contribution by Oktay Sinanoğlu; O. Sinanoğlu, K.A. Brueckner (eds.), "*Three Approaches to Electron Correlation in Atoms*," Yale Univ. Press, New Haven and London, 1970.

[86] Only then is the method cost-effective.

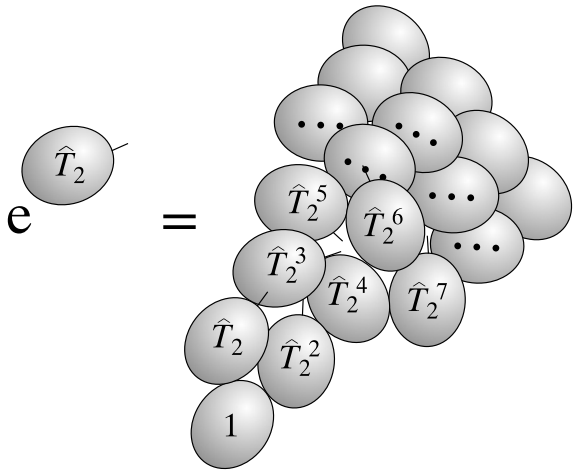


Fig. 2.15. Why such a name? An artistic impression on *coupled clusters*.

The $\exp(-\hat{T})\hat{H}\exp(\hat{T})$ operator can be expressed in terms of the commutators (see (2.51))[87]

$$e^{-\hat{T}}\hat{H}e^{\hat{T}} = \hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}]. \tag{2.65}$$

The expansion (2.65) is *finite* (justification can be only diagrammatic) since in the Hamiltonian $\hat{H}$ we have only two-particle interactions.

Multiplying Eq. (V1-2.64) from the left by the function $\langle^{mn\cdots}_{ab\ldots}|$ representing the determinant obtained from the vacuum state by the action of the annihilators $\hat{a}, \hat{b}, \ldots$ and creators $\hat{n}^\dagger, \hat{m}^\dagger, \ldots$ and integrating over all electronic coordinates, we obtain one equation for each function used[88]:

$$\langle^{mn\cdots}_{ab\ldots}|\exp(-\hat{T})\hat{H}\exp(\hat{T})|\Phi_0\rangle = 0, \tag{2.66}$$

where we have zero on the right-hand side due to the orthogonality. The Slater determinants $|^{mn}_{ab}\rangle$ represent all excitations from Φ_0 resulting from the given cluster expansion $\hat{T} =$

[87] It is straightforward to demonstrate the correctness of the first few terms by expanding the wave operator in the Taylor series.

[88] Therefore, the number of equations is equal to the number of the amplitudes t to be determined.

$\hat{T}_1 + \hat{T}_2 \ldots + \hat{T}_{l_{\max}}$. *This is the fundamental equation of the CC method*. For such a set of excited configurations the number of CC equations is equal to the number of the amplitudes sought.

The CC amplitudes $t_{ab\ldots}^{mn\cdots}$ are numbers to be found by solving Eqs. (2.66), and, consequently, the wave operator (2.56) and therefore wave function for the ground state $\Psi = \Psi_0$ is obtained.

The equations we get in the CC method are nonlinear, since the t amplitudes occur at higher powers than the first (it can be seen from Eq. (2.65) that the highest power of t is 4), which, on the one hand, requires much more demanding and capricious (than linear ones) numerical procedures, and, on the other hand, contributes to the greater efficiency of the method. In practice the number of such equations often exceeds 100 000 or a million.[89] These equations are solved iteratively assuming certain starting amplitudes t and iterating the equations until self-consistency.

We hope that in such a procedure an approximation to the ground-state wave function is obtained, although sometimes an unfortunate starting point may lead to some excited state.[90]

One often uses as a starting point that which is obtained from the linear version (reduced to obtain a linearity) of the CC method. We will write down these equations as $t_{ab}^{mn} = \ldots$ various powers of all t amplitudes. First, we neglect the nonlinear terms; this represents an initial guess for the amplitudes. The amplitudes obtained like that are substituted into the right-hand side and we iterate until self-consistency. When all the amplitudes are found, we obtain the energy E by projecting the equations (2.66) against Φ_0 function instead of $|_{ab}^{mn}\rangle$, i.e.,

$$E = \langle \Phi_0 | e^{-\hat{T}} \hat{H} e^{\hat{T}} \Phi_0 \rangle. \tag{2.67}$$

The nonvariational character of the method

The operator $(e^{-T})^{\dagger}$, conjugate to e^{-T}, is $e^{-T^{\dagger}}$, i.e., the energy

$$E = \langle e^{-\hat{T}^{\dagger}} \Phi_0 | \hat{H} e^{\hat{T}} \Phi_0 \rangle \tag{2.68}$$

89 This refers to calculations with $\hat{T} = \hat{T}_2$ for ca. 10 occupied orbitals (for instance, two water molecules) and 150 virtual orbitals. These are not calculations for large systems...

90 The first complete analysis of all CC solutions was performed by K. Jankowski and K. Kowalski, *Phys. Rev. Letters*, 81(1998)1195; *J. Chem. Phys.*, 110(1999)37, 93; *ibid.* 111(1999)2940, 2952. Recapitulation can be found in K. Jankowski, K. Kowalski, I. Grabowski, H.J. Monkhorst, *Intern. J. Quantum Chem.*, 95(1999)483.

does not represent the mean value of the Hamiltonian. Hence, the CC method is not variational. If we multiplied Eq. (2.63) from the left by $e^{\hat{T}^\dagger}$ we would obtain the variational character of E, i.e.,

$$E = \frac{\langle \Phi_0 | e^{\hat{T}^\dagger} \hat{H} e^{\hat{T}} \Phi_0 \rangle}{\langle \Phi_0 | e^{\hat{T}^\dagger} e^{\hat{T}} \Phi_0 \rangle} = \frac{\langle e^{\hat{T}} \Phi_0 | \hat{H} | e^{\hat{T}} \Phi_0 \rangle}{\langle e^{\hat{T}} \Phi_0 | e^{\hat{T}} \Phi_0 \rangle}. \tag{2.69}$$

However, it would not be possible to apply the commutator expansion (2.65) and instead of the four terms in Eq. (2.65) we would have an infinite number. Thus, the (nonvariational) CC method benefits from the very economical condition of the intermediate normalization (cf. p. V1-275). For this reason, we prefer the nonvariational approach.

2.14.4 Example: CC with double excitations

How does the CC machinery work? Let us show it for a relatively simple case, $\hat{T} = \hat{T}_2$. Eq. (2.66), written without the commutator expansion, has the form

$$\langle {}^{mn}_{ab} | e^{-\hat{T}_2} \hat{H} e^{\hat{T}_2} \Phi_0 \rangle = 0. \tag{2.70}$$

Taking advantage of the commutator expansion (2.65) we have

$$\langle {}^{mn}_{ab} | e^{-\hat{T}_2} \hat{H} e^{\hat{T}_2} \Phi_0 \rangle = \langle {}^{mn}_{ab} | \left(1 - \hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \dots\right) \hat{H} \left(1 + \hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \dots\right) \Phi_0 \rangle = \langle {}^{mn}_{ab} | \hat{H} \Phi_0 \rangle$$
$$+ \langle {}^{mn}_{ab} | \hat{H} \hat{T}_2 \Phi_0 \rangle + \frac{1}{2} \langle {}^{mn}_{ab} | \hat{H} \hat{T}_2^2 \Phi_0 \rangle - \langle {}^{mn}_{ab} | \hat{T}_2 \hat{H} \Phi_0 \rangle - \langle {}^{mn}_{ab} | \hat{T}_2 \hat{H} \hat{T}_2 \Phi_0 \rangle + A = 0.$$

However,

$$A = -\frac{1}{2} \langle {}^{mn}_{ab} | \hat{T}_2 \hat{H} \hat{T}_2^2 \Phi_0 \rangle + \frac{1}{2} \langle {}^{mn}_{ab} | \hat{T}_2^2 \hat{H} \Phi_0 \rangle + \frac{1}{2} \langle {}^{mn}_{ab} | \hat{T}_2^2 \hat{H} \hat{T}_2 \Phi_0 \rangle + \frac{1}{4} \langle {}^{mn}_{ab} | \hat{T}_2^2 \hat{H} \hat{T}_2^2 \Phi_0 \rangle = 0.$$

The last equality follows from the fact that each term is equal to zero. The first vanishes since both determinants differ by four excitations. Indeed, $\langle \left(\hat{T}_2^\dagger\right)^{mn}_{ab} |$ denotes a double *deexcitation*[91] of the doubly excited function, i.e., something proportional to $\langle \Phi_0 |$. For similar reasons (too strong deexcitations give zero) the remaining terms in A also vanish. As a result we need to solve a set of equations (for all possible excitations)

$$\langle {}^{mn}_{ab} | \hat{H} \Phi_0 \rangle + \langle {}^{mn}_{ab} | \hat{H} \hat{T}_2 \Phi_0 \rangle + \frac{1}{2} \langle {}^{mn}_{ab} | \hat{H} \hat{T}_2^2 \Phi_0 \rangle - \langle {}^{mn}_{ab} | \hat{T}_2 \hat{H} \Phi_0 \rangle - \langle {}^{mn}_{ab} | \hat{T}_2 \hat{H} \hat{T}_2 \Phi_0 \rangle = 0.$$

[91] Opposite to excitation.

After *several days*[92] of algebraic manipulations, we get the following equations for the t amplitudes (for each t_{ab}^{mn} amplitude one equation):

$$
\begin{aligned}
&(\varepsilon_m + \varepsilon_n - \varepsilon_a - \varepsilon_b)\, t_{ab}^{mn} \\
&= \langle mn|ab\rangle - \sum_{p>q} \langle mn|pq\rangle t_{ab}^{pq} - \sum_{c>d} \langle cd|ab\rangle t_{cd}^{mn} \\
&\quad + \sum_{c,p} \left[\langle cn|bp\rangle t_{ac}^{mp} - \langle cm|bp\rangle t_{ac}^{np} - \langle cn|ap\rangle t_{bc}^{mp} + \langle cm|ap\rangle t_{bc}^{np} \right] + \qquad (2.71) \\
&\quad \sum_{c>d,\, p>q} \langle cd|pq\rangle \left[t_{ab}^{pq} t_{cd}^{mn} - 2\left(t_{ab}^{mp} t_{cd}^{nq} + t_{ab}^{nq} t_{cd}^{mp}\right) - 2\left(t_{ac}^{mn} t_{bd}^{pq} + t_{ac}^{pq} t_{bd}^{mn}\right) + 4\left(t_{ac}^{mp} t_{bd}^{nq} + t_{ac}^{nq} t_{bd}^{mp}\right) \right]. \qquad (2.72)
\end{aligned}
$$

It can be seen that the last expression includes the term independent of t, the linear terms, and the quadratic terms.

How can we find the t's that satisfy Eq. (2.72)? We do it with the help of the iterative method. First, we substitute zeros for all t amplitudes on the right-hand side of the equation. Thus, from the left-hand side the first approximation to t_{ab}^{mn} is[93] $t_{ab}^{mn} \cong \frac{\langle mn|ab\rangle}{(\varepsilon_m+\varepsilon_n-\varepsilon_a-\varepsilon_b)}$. We have now an estimate of each amplitude – we are making progress. The approximation to t obtained in this way is substituted into the right-hand side to evaluate the left-hand side, and so forth. Finally, we achieve a self-consistency of the iterative process and obtain the CC wave function for the ground state of our system. With the amplitudes we calculate the energy of the system with Eq. (2.67).

This is how the CCD (the CC with double excitations in the cluster operator) works from the practical viewpoint. It is more efficient when the initial amplitudes are taken from a short CI expansion,[94] with subsequent linearization (as above) of terms containing the initial (known) amplitudes.

The computational cost of the CCD and CCSD (singles and doubles) methods scales as N^6, where N is a number of molecular orbitals (occupied and virtual[95]), whereas the analogous

---

92 Students – more courage!

93 As we see we would have trouble if $(\varepsilon_m + \varepsilon_n - \varepsilon_a - \varepsilon_b)$ is close to 0 (quasidegeneracy of the vacuum state with some other state), because then $t_{ab}^{mn} \to \infty$.

94 The CI method with inclusion of single and double excitations only.
CCD: J.A. Pople, R. Krishnan, H.B. Schlegel, J.S. Binkley, *Intern. J. Quantum Chem.*, S14(1978)545; R.J. Bartlett, G.D. Purvis III, *Intern. J. Quantum Chem.*, S14(1978)561;
CCSD: G.D. Purvis III, *J. Chem. Phys.*, 76(1982)1910.

95 These estimations are valid for the same relative increase of the number of occupied and virtual orbitals, as it is, e.g., for going from a molecule to its dimer. In the case of calculations for the same molecule, but two atomic basis sets (that differ in size), the cost increases only as N^4.

cost of the CCSDT (singles, doubles, and triples) method requires N^8 scaling. This means that if we increase the orbital basis twice, the increase in the computational cost of the CCSDT method will be four times larger than that of the CCSD scheme. This is a lot, and because of this wide-spread popularity has been gained for the CCSD(T) method, in which the role of the triple excitations is only estimated by using perturbation theory.

2.14.5 Size consistency of the CC method

The size consistency of the CC method can be proved on the basis of Eqs. (2.64) and (2.66). Let us assume that the system dissociates into two[96] noninteracting subsystems A and B (i.e., at infinite distance). Then the orbitals can be also divided into two separable (mutually orthogonal) subsets. We will show[97] that the cluster amplitudes, having *mixed* indices (from the first and second groups of orbitals), are equal to 0.

Let us note first that, for infinite distance, the Hamiltonian $\hat{H} = \hat{H}_A + \hat{H}_B$. In such a situation the wave operator can be expressed as

$$\hat{T} = \hat{T}_A + \hat{T}_B + \hat{T}_{AB}, \tag{2.73}$$

where $\hat{T}_A, \hat{T}_B, \hat{T}_{AB}$ include the operators corresponding to spin orbitals from the subsystems A, B and from the system AB, respectively. Of course, in this situation we have the following commutation condition:

$$[\hat{H}_A, \hat{T}_B] = [\hat{H}_B, \hat{T}_A] = 0. \tag{2.74}$$

Then, owing to the commutator expansion in Eq. (2.65), we obtain

$$e^{-\hat{T}}(\hat{H}_A + \hat{H}_B)e^{\hat{T}} = e^{-\hat{T}_A}\hat{H}_A e^{\hat{T}_A} + e^{-\hat{T}_B}\hat{H}_B e^{\hat{T}_B} + O(\hat{T}_{AB}), \tag{2.75}$$

where $O(\hat{T}_{AB})$ denotes the linear and higher terms in $\hat{T}_{AB}$. Substituting this into Eq. (2.66) with *bra* ⟨mixed| vector representing mixed excitation, we observe that the first two terms on the right-hand side of the last equation give zero. It means that we get the equation

$$\left\langle \text{mixed}|O(\hat{T}_{AB})\Phi_0 \right\rangle = 0, \tag{2.76}$$

which, due to the linear term in $O(\hat{T}_{AB})$, is fulfilled by $\hat{T}_{AB} = 0$. Conclusion: for the infinite distance between the subsystems we do not have mixed amplitudes and the energy of the AB system is bound to be the sum of the energies of subsystem A and subsystem B (size consistency).

[96] This can be generalized to many noninteracting subsystems.

[97] B. Jeziorski, J. Paldus, P. Jankowski, *Intern. J. Quantum Chem.*, 56(1995)129.

2.15 Equation of motion method (EOM-CC)

The CC method is used to calculate the *ground-state* energy and wave function. What about the excited states? This is a task for the equation of motion CC (EOM-CC) method, the primary goal being not the excited states themselves, but the excitation energies with respect to the ground state.

2.15.1 Similarity transformation

Let us note that for the Schrödinger equation $\hat{H}\psi = E\psi$, we can perform an interesting sequence of transformations based on the wave operator $e^{\hat{T}}$. We have

$$e^{-\hat{T}}\hat{H}\psi = Ee^{-\hat{T}}\psi,$$
$$e^{-\hat{T}}\hat{H}e^{\hat{T}}e^{-\hat{T}}\psi = Ee^{-\hat{T}}\psi.$$

We obtain the eigenvalue equation again, but for the *similarity-transformed Hamiltonian*[98]

$$\hat{\mathcal{H}}\bar{\psi} = E\bar{\psi},$$

where $\hat{\mathcal{H}} = e^{-\hat{T}}\hat{H}e^{\hat{T}}$, $\bar{\psi} = e^{-\hat{T}}\psi$, and the energy E does not change at all after this transformation. This result will be very useful in a moment.

2.15.2 Derivation of the EOM-CC equations

As the reference function in the EOM-CC method, we take the CC wave function for the ground state,

$$\psi_0 = \exp(\hat{T})\Phi_0, \tag{2.77}$$

where Φ_0 is usually a Hartree–Fock determinant. Now, we define the operator $\hat{U}_k$, which ("EOM-CC Ansatz") performs a miracle: from the wave function of the ground state ψ_0 it creates the wave function ψ_k for the k-th excited state of the system:

$$\psi_k = \hat{U}_k\psi_0.$$

The operators $\hat{U}_k$ change the coefficients in front of the configurations (see p. 135). The operators $\hat{U}_k$ are (unlike the wave operator $\exp(\hat{T})$) linear with respect to the excitations, i.e., the

[98] In contrast to the Hamiltonian $\hat{H}$ the similarity-transformed Hamiltonian does not represent a Hermitian operator. Moreover, it contains not only the one- and two-electron terms as in $\hat{H}$, but also all other many-electron operators up to the total number of electrons in the system.

excitation amplitudes occur there in first powers. For the case of the single and double excitations (EOM-CCSD) we have $\hat{T}$ in the form of the sum of single and double excitations:

$$\hat{T} = \hat{T}_1 + \hat{T}_2$$

and

$$\hat{U}_k = \hat{U}_{k,0} + \hat{U}_{k,1} + \hat{U}_{k,2},$$

where the task for the $\hat{U}_{k,0}$ operator is to change the coefficient in front of the function Φ_0 to that appropriate to the $|k\rangle$ function and the role of the operators $\hat{U}_{k,1}, \hat{U}_{k,2}$ is an appropriate modification of the coefficients in front of the singly and doubly excited configurations, respectively. These tasks are carried out by the excitation operators with τ amplitudes (they have to be distinguished from the amplitudes of the CC method), i.e.,

$$\hat{U}_{k,0} = \tau_0(k),$$
$$\hat{U}_{k,1} = \sum_{a,p} \tau_a^p(k)\, \hat{p}^\dagger \hat{a},$$
$$\hat{U}_{k,2} = \sum_{a,b,p,q} \tau_{ab}^{pq}(k)\, \hat{q}^\dagger \hat{p}^\dagger \hat{a}\hat{b},$$

where the amplitudes $\tau(k)$ are numbers, which are the targets of the EOM-CC method. The amplitudes give the wave function ψ_k and the energy E_k.

We write down the Schrödinger equation for the excited state,

$$\hat{H}\psi_k = E_k\psi_k.$$

Now we substitute the EOM-CC Ansatz:

$$\hat{H}\hat{U}_k\psi_0 = E_k\hat{U}_k\psi_0,$$

and from the definition of the CC wave operator we get[99]

$$\hat{H}\hat{U}_k \exp(\hat{T})\Phi_0 = E_k\hat{U}_k \exp(\hat{T})\Phi_0.$$

Due to the missing *deexcitation* part (i.e., that which lowers the excitation rank, e.g., from doubles to singles) the operators $\hat{U}_k$ and $\hat{T}$ commute,[100] hence the operators $\hat{U}_k$ and $\exp(\hat{T})$

[99] By neglecting higher than single and double excitations the equation represents an approximation.

[100] If $\hat{U}_k$ contains true excitations, then it does not matter whether excitations are performed by $\hat{U}_k\hat{T}$ or $\hat{T}\hat{U}_k$ (commutation), because both $\hat{U}_k$ and $\hat{T}$ mean going up in the energy scale. If, however, $\hat{U}_k$ contains deexcitations, then it may happen that there is an attempt in $\hat{T}\hat{U}_k$ to deexcite the ground-state wave function – that makes immediately 0, whereas $\hat{U}_k\hat{T}$ may be still OK, because the excitations in $\hat{T}$ may be more important than the deexcitations in $\hat{U}_k$.

also commute, so we have

$$\hat{U}_k \exp(\hat{T}) = \exp(\hat{T})\hat{U}_k.$$

Substituting this we have

$$\hat{H} \exp(\hat{T})\hat{U}_k\Phi_0 = E_k \exp(\hat{T})\hat{U}_k\Phi_0,$$

and multiplying from the left with $\exp(-\hat{T})$ we get

$$[\exp(-\hat{T})\hat{H} \exp(\hat{T})]\hat{U}_k\Phi_0 = E_k\hat{U}_k\Phi_0,$$

or introducing the similarity-transformed Hamiltonian

$$\hat{\mathcal{H}} = e^{-\hat{T}}\hat{H}e^{\hat{T}},$$

we obtain

$$\hat{\mathcal{H}}\hat{U}_k\Phi_0 = E_k\hat{U}_k\Phi_0.$$

From the last equation we subtract the CC equation for the ground state,

$$[\exp(-\hat{T})\hat{H} \exp(\hat{T})]\Phi_0 = E_0\Phi_0$$

multiplied from the left with $\hat{U}_k$, i.e., $\hat{U}_k\hat{\mathcal{H}}\Phi_0 = E_0\hat{U}_k\Phi_0$, and we get

$$\hat{\mathcal{H}}\hat{U}_k\Phi_0 - \hat{U}_k\hat{\mathcal{H}}\Phi_0 = E_k\hat{U}_k\Phi_0 - E_0\hat{U}_k\Phi_0.$$

Finally, we obtain an important result:

$$[\hat{\mathcal{H}}, \hat{U}_k]\Phi_0 = (E_k - E_0)\,\hat{U}_k\Phi_0.$$

The operator $\hat{U}_k$ contains the sought amplitudes $\tau\,(k)$.

We find them in a similar manner as in the CC method. For that purpose we make a scalar product of the left- and right-hand sides of that equation with each excitation $|^{mn\cdots}_{ab\ldots}\rangle$ used in $\hat{U}_k$. We get the set of the EOM-CC equations whose number is equal to the number of sought amplitudes plus one more equation due to the normalization condition of ψ_k. The unknown parameters are amplitudes and for the excitation energies $E_k - E_0$ we have

$$\left\langle {}^{mn\ldots}_{ab\ldots} \middle| [\hat{\mathcal{H}}, \hat{U}_k] \middle| \Phi_0 \right\rangle = (E_k - E_0) \left\langle {}^{mn\ldots}_{ab\ldots} \middle| \hat{U}_k \middle| \Phi_0 \right\rangle.$$

We solve these equations and the problem is over.

It is important that the excitations $|^{mn\cdots}_{ab\cdots}\rangle$ used in $\hat{U}_k$ include not only the regular singles and doubles, but also the function with no excitation,[101] i.e., the function Φ_0, and the states with different numbers of electrons, i.e., with the ionized states or the states with extra electrons. It turned out that the last possibility offers an intriguing way of determination of a particular electronic state starting from "several distinct points of view." Indeed, one may carry out the EOM-CC computations for a given state (with N electrons) starting first from function $\Phi_0(1, 2, ...N)$ and then repeating the calculations with different functions $\Phi_0(1, 2, ...N - M)$, where $M = \pm 1, \pm 2, ...$, and compare the results. As shown by Kucharski and Musiał[102] such a possibility is especially fruitful if $\Phi_0(1, 2, ...N)$ were a very bad approximation to the ground-state wave function, e.g., in case of dissociation of a chemical bond. This approach may offer an elegant avenue to circumvent the serious problem of bond dissociation.

2.16 Many-body perturbation theory (MBPT)

The majority of routine calculations in quantum chemistry are done with variational methods (mainly the Hartree–Fock scheme). If we consider post-Hartree–Fock calculations, then nonvariational (CCSD, CCSD(T)), as well as perturbational (among them MBPT), approaches take the lead. The perturbational methods are based on the simple idea that the system in slightly modified conditions is similar to that before the perturbation is applied (cf. p. V1-274).

In the formalism of perturbation theory, knowing the unperturbed system and the perturbation we are able to provide successive corrections to obtain the solution of the perturbed system. Thus, for instance, the energy of the perturbed system is the energy of the unperturbed system plus the first-order correction plus the second-order correction, etc. If the perturbation is small then we *hope*[103] the series is convergent; even then however, there is no guarantee that the series converges fast.

2.16.1 Unperturbed Hamiltonian

In the perturbational approach (cf. p. V1-274) to the electron correlation the Hartree–Fock function, Φ_0, is treated as the zero-order approximation to the true ground-state wave function, i.e., $\Phi_0 = \psi_0^{(0)}$. Thus, the Hartree–Fock wave function stands at the starting point, while the goal is the exact ground-state electronic wave function.

101 More precisely: to get only the excitation energy we do not need the coefficient next to Φ_0.

102 S. Kucharski, M. Musiał, *Proc. Conference HITY*, Cracow, 2011.

103 There is not much known concerning the convergence of series occurring in quantum chemistry. Commonly, only a few perturbational corrections are computed.

In the majority of cases this is a reasonable approximation, since the Hartree–Fock method usually provides as much as 98%–99% of the total energy.[104] A Slater determinant Φ_I is constructed from the spin orbitals obeying the Fock equation. How to construct the operator for which the Slater determinant is an eigenfunction? We will find out in a moment that this operator is the sum of the Fock operators (cf. Appendix C)

$$\hat{H}^{(0)} = \sum_i \hat{F}(i) = \sum_i \epsilon_i \hat{\imath}^\dagger \hat{\imath}. \tag{2.78}$$

Indeed,

$$\hat{H}^{(0)} \Phi_I = \sum_i \epsilon_i \hat{\imath}^\dagger \hat{\imath} \cdot \Phi_I = (\sum_i \epsilon_i) \cdot \Phi_I, \tag{2.79}$$

since the annihilation of one spin orbital in the determinant and the creation of the same spin orbital leaves the determinant unchanged. This is so under the condition that the spin orbital ϕ_i *is present* in $\psi_0^{(0)}$.

> The eigenvalue of $\hat{H}_0 = \sum_i \epsilon_i \hat{\imath}^\dagger \hat{\imath}$ is always the sum of the orbital energies corresponding to all spin orbitals in the Slater determinant Φ_I.

This means that the sum of several Slater determinants, each built from a different (in the sense of the orbital energies) set of spin orbitals, is not an eigenfunction of $\hat{H}^{(0)}$.

2.16.2 Perturbation theory – slightly different presentation

We have to solve the Schrödinger equation for the ground state[105] $\hat{H}\psi_0 = E\psi_0$, with $\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)}$, where $\hat{H}^{(0)}$ denotes the unperturbed Hamiltonian given by Eq. (2.78), and $\hat{H}^{(1)}$ is a perturbation operator. The eigenfunctions and the eigenvalues of $\hat{H}^{(0)}$ are given by Eq. (2.79), but reminding the perturbation theory formulae we will denote the Slater determinants $\Phi_I \equiv \psi_I^{(0)}$.

For the ground state, we expand the energy E_0 and the wave function ψ_0 in a power series[106]: we put $\lambda \hat{H}^{(1)}$ instead of $\hat{H}^{(1)}$ in the Hamiltonian and expand the energy and the wave function

104 Sometimes, as we know, the method fails and then the perturbation theory based on the Hartree–Fock starting point is a risky business, since the perturbation is very large.

105 We use the notation from Chapter V1-5.

106 This is an old trick of perturbation theory equivalent to saying that the shape of a bridge loaded with a car is the shape of the bridge without the car plus the deformation proportional to the mass of the car plus the deformation proportional to the square of the mass of the car, etc. This works if the bridge is solid and the car is light (the perturbation is small).

in a power series[107] with respect to λ, so we have

$$E_0 = E_0^{(0)} + \lambda E_0^{(1)} + \lambda^2 E_0^{(2)} + \ldots, \tag{2.80}$$

$$\psi_0 = \psi_0^{(0)} + \lambda \psi_0^{(1)} + \lambda^2 \psi_0^{(2)} + \ldots. \tag{2.81}$$

The Schrödinger equation does not force the normalization of the function. It is convenient then to use the *intermediate normalization* (Fig. 2.16a), i.e., to require that $\langle \psi_0 | \psi_0^{(0)} \rangle = 1$.

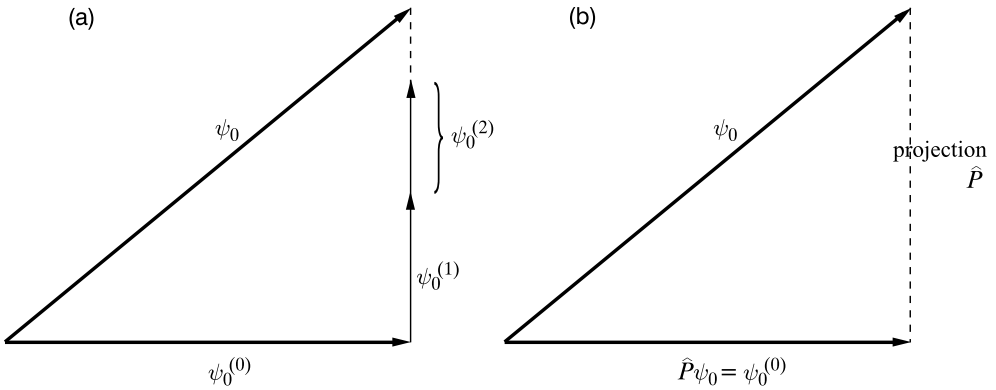


Fig. 2.16. Pictorial presentation of (a) the intermediate normalization $\langle \psi | \psi_0^{(0)} \rangle = 1$ and (b) the projection onto the axis $\psi_0^{(0)}$ in the Hilbert space using the operator $\hat{P} = |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|$; $\psi_0^{(n)}$, $n = 1, 2$, represents a correction of the n-th order to the ground-state wave function. The picture can only be simplistic and schematic: the orthogonality of $\psi_0^{(n)}$ to ψ_0 is shown correctly, but the apparent parallelism of $\psi_0^{(1)}$ and $\psi_0^{(2)}$ is false.

This means that the (nonnormalized) ψ_0 must include the normalized function of zero-order $\psi_0^{(0)}$ *and, possibly, something orthogonal to it*.

2.16.3 MBPT machinery – part one: energy equation

Let us write $\hat{H}\psi_0$ as $\hat{H}\psi_0 = (\hat{H}^{(0)} + \hat{H}^{(1)})\psi_0$, or, in another way, as $\hat{H}^{(1)}\psi_0 = (\hat{H} - \hat{H}^{(0)})\psi_0$. Multiplying this equation by $\psi_0^{(0)}$ and integrating (and taking advantage of the intermediate normalization), we get

$$\langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0 \rangle = \langle \psi_0^{(0)} | (\hat{H} - \hat{H}^{(0)}) \psi_0 \rangle = E_0 \langle \psi_0^{(0)} | \psi_0 \rangle - \langle \psi_0^{(0)} | \hat{H}^{(0)} \psi_0 \rangle = E_0 - E_0^{(0)} = \Delta E_0. \tag{2.82}$$

Thus,

[107] So we assume that the respective functions are analytic in the vicinity of $\lambda = 0$.

$$\Delta E_0 = \langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0 \rangle. \tag{2.83}$$

2.16.4 Reduced resolvent or the "almost" inverse of $(E_0^{(0)} - \hat{H}^{(0)})$

Let us define several useful tools – we need to get familiar with them now – which will introduce a certain elegance into our final equations.

Let the first be *a projection operator* on the ground-state zero-order function (see Fig. 2.16b),

$$\hat{P} = |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|. \tag{2.84}$$

This means that $\hat{P}\chi$ is, within accuracy to a constant, equal to $\psi_0^{(0)}$ for an arbitrary function χ, or zero. Indeed, if χ is expressed as a linear combination of the eigenfunctions $\psi_n^{(0)}$ (these functions form an orthonormal complete set as eigenfunctions of the Hermitian operator)

$$\chi = \sum_n c_n \psi_n^{(0)}, \tag{2.85}$$

then (Fig. 2.16b)

$$\hat{P}\chi = \sum_n c_n \hat{P}\psi_n^{(0)} = \sum_n c_n |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|\psi_n^{(0)}\rangle = \sum_n c_n \delta_{0n} \psi_0^{(0)} = c_0 \psi_0^{(0)}. \tag{2.86}$$

Let us now introduce a projection operator

$$\hat{Q} = 1 - \hat{P} = \sum_{n=1}^{\infty} |\psi_n^{(0)}\rangle\langle\psi_n^{(0)}| \tag{2.87}$$

on the space orthogonal to $\psi_0^{(0)}$. Obviously, $\hat{P}^2 = \hat{P}$ and $\hat{Q}^2 = \hat{Q}$. The latter holds since $\hat{Q}^2 = (1 - \hat{P})^2 = 1 - 2\hat{P} + \hat{P}^2 = 1 - \hat{P} = \hat{Q}$.

Now we define *a reduced resolvent*

$$\hat{R}_0 = \sum_{n=1}^{\infty} \frac{|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|}{E_0^{(0)} - E_n^{(0)}}. \tag{2.88}$$

The definition says that the reduced resolvent represents an operator that, from an arbitrary vector ϕ of the Hilbert space:

- cuts out its components along all the unit (i.e., normalized) basis vectors $\psi_n^{(0)}$ *except* $\psi_0^{(0)}$,
- weighs the projections by the factor $\frac{1}{E_0^{(0)}-E_n^{(0)}}$, so they become less and less important for higher and higher energy states,
- adds all the weighed vectors together.

We easily obtain[108]

$$\hat{R}_0(E_0^{(0)} - \hat{H}^{(0)}) = (E_0^{(0)} - \hat{H}^{(0)})\hat{R}_0 = \hat{Q}. \tag{2.89}$$

For functions ϕ orthogonal to $\psi_0^{(0)}$, i.e., satisfying $\hat{Q}\phi = \phi$, the action of the operator $\hat{R}_0$ is identical to that of the operator $(E_0^{(0)} - \hat{H}^{(0)})^{-1}$; $\hat{R}_0$ does not represent however the inverse of $(E_0^{(0)} - \hat{H}^{(0)})$, because for $\phi = \psi_0^{(0)}$ we get $\hat{R}_0(E_0^{(0)} - \hat{H}^{(0)})\phi = 0$, and not the unchanged ϕ.

2.16.5 MBPT machinery – part two: wave function equation

Our goal now will be to present the Schrödinger equation in a different and useful form. Let us first write it down as follows:

$$(E_0 - \hat{H}^{(0)})\psi_0 = \hat{H}^{(1)}\psi_0. \tag{2.90}$$

We aim at having $(E_0^{(0)} - \hat{H}^{(0)})\psi_0$ on the left-hand side, because we plan to apply to it the reduced resolvent. Let us add $(E_0^{(0)} - E_0)\psi_0$ to both sides of that equation to obtain

$$(E_0^{(0)} - \hat{H}^{(0)})\psi_0 = (E_0^{(0)} - E_0 + \hat{H}^{(1)})\psi_0. \tag{2.91}$$

Let us now operate on both sides of this equation with the reduced resolvent $\hat{R}_0$. Then we have

$$\hat{R}_0(E_0^{(0)} - \hat{H}^{(0)})\psi_0 = \hat{R}_0(E_0^{(0)} - E_0 + \hat{H}^{(1)})\psi_0. \tag{2.92}$$

On the left-hand side we have $\hat{Q}\psi_0$ (as follows from Eq. (2.89)), but $\hat{Q}\psi_0 = (1 - \hat{P})\psi_0 = \psi_0 - |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|\psi_0\rangle = \psi_0 - \psi_0^{(0)}$, due to the intermediate normalization. As a result, the equation takes the form

$$\psi_0 - \psi_0^{(0)} = \hat{R}_0(E_0^{(0)} - E_0 + \hat{H}^{(1)})\psi_0. \tag{2.93}$$

---

[108] Let us make sure of this. We have
$\hat{R}_0(E_0^{(0)} - \hat{H}^{(0)})\phi = \sum_{n=1}^{\infty}(E_0^{(0)} - E_n^{(0)})^{-1}|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|(E_0^{(0)} - \hat{H}^{(0)})|\phi\rangle = \sum_{n=1}^{\infty}(E_0^{(0)} - E_n^{(0)})^{-1}(E_0^{(0)} - E_n^{(0)})|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|\phi\rangle = \sum_{n=1}^{\infty}|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|\phi\rangle = \hat{Q}\phi.$
Let us now operate on the same function with the operator $(E_0^{(0)} - \hat{H}^{(0)})\hat{R}_0$ (i.e., the operators are in reverse order). Then we have
$(E_0^{(0)} - \hat{H}^{(0)})\hat{R}_0\phi = (E_0^{(0)} - \hat{H}^{(0)})\sum_{n=1}^{\infty}(E_0^{(0)} - E_n^{(0)})^{-1}|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|\phi\rangle = \sum_{n=1}^{\infty}(E_0^{(0)} - E_n^{(0)})^{-1}(E_0^{(0)} - \hat{H}^{(0)})|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|\phi\rangle = \sum_{n=1}^{\infty}|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|\phi\rangle = \hat{Q}\phi.$

Thus, we obtain

$$\psi_0 = \psi_0^{(0)} + \hat{R}_0(E_0^{(0)} - E_0 + \hat{H}^{(1)})\psi_0. \tag{2.94}$$

At the same time, based on the expression for ΔE in perturbation theory (Eq. (2.83)), we have

$$E_0 = E_0^{(0)} + \langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0 \rangle. \tag{2.95}$$

These are the equations of the MBPT, in which the exact wave function and energy are expressed in terms of the unperturbed functions and energies plus certain corrections. The problem is that, as can be seen, these corrections involve the unknown function and unknown energy.

Let us not despair in this situation, but try to apply an iterative technique. First substitute for ψ_0 in the right-hand side of Eq. (2.94) that which most resembles ψ_0, i.e., $\psi_0^{(0)}$. We obtain

$$\psi_0 \cong \psi_0^{(0)} + \hat{R}_0(E_0^{(0)} - E_0 + \hat{H}^{(1)})\psi_0^{(0)}, \tag{2.96}$$

and then the new approximation to ψ_0 should again be plugged into the right-hand side, and this procedure is continued until convergence. It can be seen that the successive terms form a series (let us hope that it is convergent). We have

$$\psi_0 = \sum_{n=0}^{\infty} [\hat{R}_0(E_0^{(0)} - E_0 + \hat{H}^{(1)})]^n \psi_0^{(0)}. \tag{2.97}$$

Now only known quantities occur on the right-hand side except for E_0, the exact energy. Let us pretend that its value is known and insert into the energy expression (2.95) the function ψ_0. Then we have

$$E_0 = E_0^{(0)} + \langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0 \rangle = E_0^{(0)} + \langle \psi_0^{(0)} | \hat{H}^{(1)} \sum_{n=0}^{M} [\hat{R}_0(E_0^{(0)} - E_0 + \hat{H^{(1)}})]^n | \psi_0^{(0)} \rangle. \tag{2.98}$$

Let us go back to our problem: we want to have E_0 on the left-hand side of the last equation, while – for the time being – E_0 occurs on the right-hand sides of both equations. To exit the

situation we will treat E_0 occurring on the right-hand side as a parameter manipulated in such a way as to obtain equality in both above equations. We may do it in two ways. One leads to Brillouin–Wigner perturbation theory, the other to Rayleigh–Schrödinger perturbation theory.

2.16.6 Brillouin–Wigner perturbation theory

Let us decide first at what $n = M$ we terminate the series, i.e., to what order of perturbation theory the calculations will be carried out. Say, $M = 3$. Let us now take any reasonable value[109] as a parameter of E_0. We insert this value into the right-hand side of Eq. (2.98) for E_0 and calculate the left-hand side, i.e., E_0. Then let us again insert the new E_0 into the right-hand side and continue in this way until self-consistency, i.e., until (2.98) is satisfied. When E_0 is known we go to Eq. (2.97) and compute ψ_0 (through a certain order, e.g., M).

Brillouin–Wigner perturbation theory has, as seen, the somewhat unpleasant feature that successive corrections to the wave function depend on the M assumed at the beginning.

We may suspect[110] – and this is true – that *the Brillouin–Wigner perturbation theory is not size-consistent*.

2.16.7 Rayleigh–Schrödinger perturbation theory

As an alternative to Brillouin–Wigner perturbation theory, we may consider Rayleigh–Schrödinger perturbation theory, which *is size-consistent*. In this method the total energy is computed in a stepwise manner,

$$E_0 = \sum_{k=0}^{\infty} E_0^{(k)}, \tag{2.99}$$

in such a way that we first calculate the first-order correction $E_0^{(1)}$, i.e., of the order of $\hat{H}^{(1)}$, then the second-order correction, $E_0^{(2)}$, i.e., of the order of $(\hat{H}^{(1)})^2$, etc. If we insert into the right-hand sides of (2.97) and (2.98) the expansion $E_0 = \sum_{k=0}^{\infty} E_0^{(k)}$ and then, by applying the usual perturbation theory argument, we equalize the terms of the same order, we get:

- for $n = 0$,

$$E_0^{(1)} = \langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0^{(0)} \rangle; \tag{2.100}$$

[109] An "unreasonable" value will lead to numerical instabilities. *Then* we will learn that it was unreasonable indeed to take it.

[110] Due to the iterative procedure.

- for $n = 1$,

$$E^{(2)} = \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 (E_0^{(0)} - E_0 + \hat{H}^{(1)}) \psi_0^{(0)} \rangle = \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 \hat{H}^{(1)} \psi_0^{(0)} \rangle, \tag{2.101}$$

since $\hat{R}_0 \psi_0^{(0)} = 0$;

- for $n = 2$,
$E^{(3)} =$ the third-order terms from the expression

$$\langle \psi_0^{(0)} | \hat{H}^{(1)} [\hat{R}_0 (E_0^{(0)} - E_0^{(0)} - E_0^{(1)} - E_0^{(2)} - \ldots + \hat{H}^{(1)})]^2 \psi_0^{(0)} \rangle$$
$$= \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 (-E_0^{(1)} - E_0^{(2)} - \ldots + \hat{H}^{(1)}) \hat{R}_0 (-E_0^{(1)} - E_0^{(2)} - \ldots + \hat{H}^{(1)}) \psi_0^{(0)} \rangle$$

and the only terms of the third order are

$$E^{(3)} = \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 \hat{H}^{(1)} \hat{R}_0 \hat{H}^{(1)} \psi_0^{(0)} \rangle - E_0^{(1)} \langle \psi_0^{(0)} | \hat{H}^{(1)} R_0^2 \hat{H}^{(1)} \psi_0^{(0)} \rangle; \tag{2.102}$$

- etc.

Unfortunately, we cannot give a general expression for the k-th correction to the energy although we can give an algorithm for the construction of such an expression.[111] Rayleigh–Schrödinger perturbation theory (unlike the Brillouin–Wigner approach) has the nice feature that the corrections of the particular orders are independent of the maximum order chosen.

2.17 Møller–Plesset version of Rayleigh–Schrödinger perturbation theory

Let us consider the case of a closed shell.[112] In the Møller–Plesset perturbation theory we assume as $\hat{H}^{(0)}$ the sum of the Hartree–Fock operators (from the RHF method, see Eq. (2.78)) and $\psi_0^{(0)} = \psi_{RHF}$, i.e.,

$$\hat{H}^{(0)} = \sum_i^N \hat{F}(i) = \sum_i^\infty \epsilon_i i^\dagger i,$$

$$\hat{H}^{(0)} \psi_{RHF} = E_0^{(0)} \psi_{RHF}, \tag{2.103}$$

$$E_0^{(0)} = \sum_i \epsilon_i \tag{2.104}$$

(the last summation is over spin orbitals occupied in the RHF function); hence the perturbation, known in the literature as *a fluctuation potential*, is equal to

$$\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}. \tag{2.105}$$

111 J. Paldus, J. Čížek, *Adv. Quantum Chem.,* 105(1975).

112 Møller–Plesset perturbation theory also has its multireference formulation when the function Φ_0 is a linear combination of determinants (K. Woliński, P. Pulay, *J. Chem. Phys*., 90(1989)3647).

For such a perturbation we may carry out calculations through a given order n: we have a sequence of approximations MPn. A very popular method relies on the inclusion of the perturbational corrections to the energy through the second order (known as MP2 method) and through the fourth order (MP4).

2.17.1 Expression for MP2 energy

What is the expression for the total energy in the MP2 method?

Let us note first that when calculating the mean value of the Hamiltonian in the standard Hartree–Fock method, we automatically obtain the sum of the zero-order energies $\sum_i \epsilon_i$ and the first-order correction to the energy $\left\langle \psi_{RHF} | \hat{H}^{(1)} \psi_{RHF} \right\rangle$. Indeed,

$$E_{RHF} = \left\langle \psi_{RHF} | \hat{H} \psi_{RHF} \right\rangle = \left\langle \psi_{RHF} | (\hat{H}^{(0)} + \hat{H}^{(1)}) \psi_{RHF} \right\rangle = (\sum_i \epsilon_i) + \left\langle \psi_{RHF} | \hat{H}^{(1)} \psi_{RHF} \right\rangle.$$

So what is left to be done (in the MP2 approach) is the addition of the second-order correction to the energy (p. V1-279, the prime in the summation symbol indicates that the term making the denominator equal to zero is omitted), where, as the complete set of functions, we assume the Slater determinants $\psi_k^{(0)}$ corresponding to the energy $E_k^{(0)}$ (they are generated by various spin orbital occupancies), i.e.,

$$E_{MP2} = E_{RHF} + \sum_k{}' \frac{\left| \left\langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_{RHF} \right\rangle \right|^2}{E_0^{(0)} - E_k^{(0)}} = E_{RHF} + \sum_k{}' \frac{\left| \left\langle \psi_k^{(0)} | \hat{H} \psi_{RHF} \right\rangle \right|^2}{E_0^{(0)} - E_k^{(0)}}. \qquad (2.106)$$

The last equality holds, because ψ_{RHF} is an eigenfunction of $\hat{H}^{(0)}$, and $\psi_k^{(0)}$ and ψ_{RHF} are orthogonal. It can be seen that among possible functions $\psi_k^{(0)}$, we may ignore all but doubly excited ones. Why? This is because:

- the single excitations give $\left\langle \psi_k^{(0)} | \hat{H} \psi_{RHF} \right\rangle = 0$ due to the Brillouin theorem, and
- the triple and higher excitations differ by more-than-two excitations from the functions ψ_{RHF} and, due to Slater–Condon rule IV (see Appendix V1-N, p. V1-707), give a contribution equal to 0.

In such a case, we take as the functions $\psi_k^{(0)}$ only doubly excited Slater determinants ψ_{ab}^{pq}, which means that we replace the occupied spin orbitals $a \to p, b \to q$, and to avoid repetitions $a < b, p < q$. These functions are eigenfunctions of $\hat{H}^{(0)}$ with the eigenvalues being the sum

of the respective orbital energies, Eq. (2.79). Thus, using Slater–Condon rule III, we obtain the energy correct through the second order,

$$E_{MP2} = E_{RHF} + \sum_{a<b,p<q} \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q}, \tag{2.107}$$

and hence, the MP2 scheme viewed as an approximation to the correlation energy gives[113]

$$E_{corel} \approx E_{MP2} - E_{RHF} = \sum_{a<b,p<q} \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q}. \tag{2.108}$$

The MP2 energy is an easy to calculate post-Hartree–Fock improvement. Well, how effective is the MP method in computing the electron correlation? Fig. 2.17 shows a comparison of the RHF, MP2, MP3, and CISD (in this case equivalent to CI) methods applied to the hydrogen molecule for several values of the internuclear distance R. The results of CI are better than those of the RHF method – a feature guaranteed by the variational principle. As one can see the RHF method indicates quite accurately the position of the minimum, although it makes there a clearly visible error in energy. In contrast to this, for large R, the method creates a kind of disaster as expected (see Section V1-8.5.1, p. V1-510). The duty of the perturbational MP2 and MP3 methods is to improve the RHF mess by adding some corrections. This difficult job is done very well for distances R close to the minimum. The duty is, however, too demanding for large internuclear distances, although even there the improvement is important, especially for the MP3 method.

2.17.2 Is the MP2 method size-consistent?

Let us see. From Eq. (2.108) (p. 171) we have $E_{MP2} = E_{HF} + \sum'_{a<b,p<q} \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a+\varepsilon_b-\varepsilon_p-\varepsilon_q}$. On the right-hand side, the E_{HF} energy *is* size-consistent, as shown at the beginning of this chapter. It is therefore sufficient to prove that the second term is also size-consistent. For separated subsystems the excitations $a \to p$ and $b \to q$ must correspond to the spin orbitals a and p belonging to the same molecule (and represent the Hartree–Fock orbitals for the subsystems).

[113] The MP2 method usually gives satisfactory results, i.a., the frequencies of the normal modes. There are indications, however, that the deformations of the molecule connected with some vibrations strongly affecting the electron correlation (vibronic coupling) create too severe a test for the method – the error may amount to 30%–40% for frequencies of the order of hundreds of cm^{-1}, as has been shown by D. Michalska, W. Zierkiewicz, D.C. Bieńko, W. Wojciechowski, T. Zeegers-Huyskens, *J. Phys. Chem. A*, 105(2001)8734).

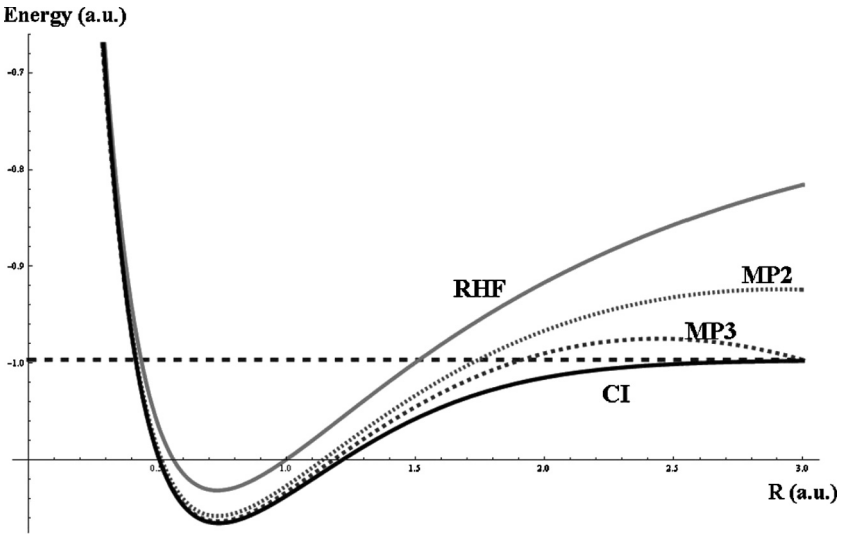


Fig. 2.17. The electronic energy of the hydrogen molecule as a function of the internuclear distance R. The energy is computed by using the RHF (gray solid line), MP2 (lighter dotted line), MP3 (darker dotted line), and CI (black solid line) methods. The energy of the two isolated hydrogen atoms is shown as a horizontal dashed line. The computations have been carried out by using the Gaussian program with a standard basis of atomic orbitals, 6-311G(d,p). Energies and distances are given in a.u.

The same can be said for the spin orbitals b and q. We have therefore (lim denotes the limit corresponding to all distances among the subsystems equal to infinity, $E_{RHF}(A)$ stands for the Hartree–Fock energy of molecule A)

$$\lim E_{MP2} = \sum_A E_{RHF}(A) + \lim \sum_{a<b,p<q}{}' \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q} =$$

$$\sum_A E_{RHF}(A) + \sum_A \sum_{a,b,p,q\in A}{}' \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q} +$$

$$\sum_{A<B} \sum_{a,p\in A,b,q\in B}{}' \lim \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q} =$$

$$\sum_A E_{MP2}(A) + 0,$$

because in the last term the integral $\langle ab|pq\rangle$ vanishes as $\frac{1}{R_{AB}}$, while the integral $\langle ab|qp\rangle$ vanishes even faster (exponentially, because of the overlap of spin orbitals belonging to different molecules).

The result obtained means that

> the MP2 method is size-consistent.

Example 4. The proof of the size consistency should be reflected by numerical results in practical applications. Let us perform some routine calculations for two helium atoms[114] by using the HF, MP2, CISD, and CCSD methods. We perform the calculations for a single helium atom, and then for two separated helium atoms, but with the internuclear distance so large that there are serious grounds for rejecting any suspicion about their significant mutual interaction. Then, we will see whether the energy for the two noninteracting atoms is twice the energy of a single isolated atom (as it should be for size consistency). Well, how to decide about such a safe distance? A helium atom is an object of a diameter of about 2 Å (in a simple and naive view). A distance of about 30 Å should be sufficiently large to have the interaction energy negligible. The numerical results are collected in the table below.

	2 He	He_2 ($R = 30$ Å)
HF	**−5.7103209**	**−5.7103209**
MP2	**−5.7327211**	**−5.7327211**
CISD	**−5.7403243**	**−5.740**1954
CCSD	**−5.7403243**	**−5.7403243**

The numbers given confirm the theoretical considerations. The numbers in the second column (twice the energy of the isolated helium atom) and the third column (the energy of the two distant atoms) are identical to eight significant figures (bold) for the HF, MP2, and CCSD methods. In contrast, according to what we know about the CI method, the CISD method is size-inconsistent (the difference is much more important – on the fifth significant figure).

2.17.3 Convergence of the Møller–Plesset perturbation series

Does the Møller–Plesset perturbational series converge? Very often this question can be considered surrealist, since *most frequently* we carry out calculations through the second, third, and – at most – fourth order of perturbation theory. Such calculations usually give a quite satisfactory description of the physical quantities considered and we do not think about going to high orders requiring major computational effort. There were, however, scientists interested to see how fast the convergence is if very high orders are included (MPn) for $n < 45$. And there was a surprise (see Fig. 2.18).

[114] One may use, e.g., the public domain www.webmo.net, which offers several quantum chemistry programs; we use here the Gaussian program with the atomic orbital basis set 6-31G(d).

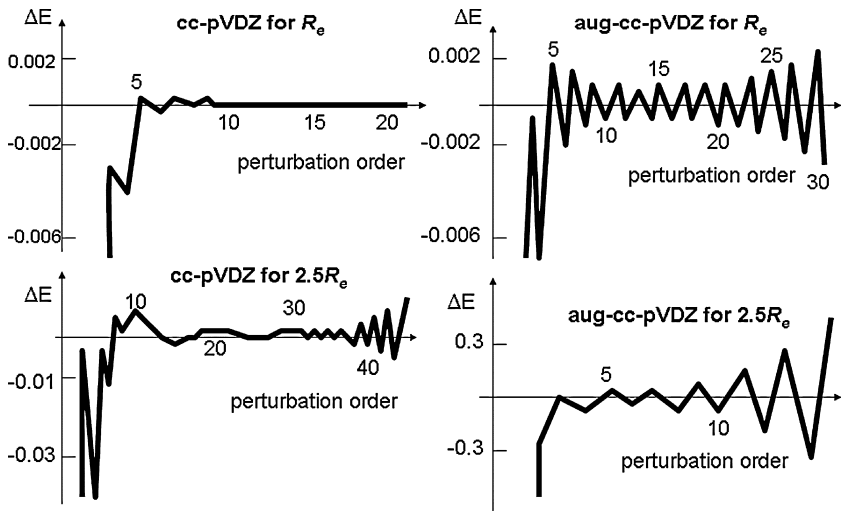


Fig. 2.18. Convergence of the Møller–Plesset perturbation theory (deviation from the exact value, a.u.) for the HF molecule as a function of the basis set used (cc-pVDZ and augmented cc-pVDZ) and assumed bond length, R_e denotes the HF equilibrium distance (following T. Helgaker, P. Jørgensen, J. Olsen, "*Molecular Electronic-Structure Theory*," Wiley, Chichester, 2000, p. 780, Fig. 14.6. Reproduced with permission of John Wiley and Sons Ltd.).

It is true that the first few orders of the MP perturbation theory give reasonably good results, but later, the accuracy of the MP calculations gets worse. A lot depends on the atomic orbital basis set adopted and wealthy people (using the augmented basis sets – which is much more rare) encounter some difficulties whereas poor ones (modest basis sets) do not. Moreover, for long bond lengths (2.5 of the equilibrium distance R_e) the MPn performance is worse. For high orders, the procedure is heading for a catastrophe[115] of the kind already described on p. V1-284. The reason for this is the highly excited and diffuse states used as the expansion functions.[116]

2.17.4 Special status of double excitations

In Møller–Plesset perturbation theory

$$\Delta E = E_0 - E_0^{(0)} = E_0 - E_{RHF} - E_0^{(0)} + E_{RHF} = E_{corel} + \left(E_{RHF} - E_0^{(0)}\right).$$

[115] Except for the smaller basis set and the equilibrium bond length, but the problem has been studied up to $n = 21$.
[116] An analysis of this problem is given in the book cited in the caption to Fig. 2.18.

However,[117]

$$\Delta E = E_0 - E_0^{(0)} = \left\langle \psi_0^{(0)} | \hat{H} \psi_0 \right\rangle - E_0^{(0)}.$$

The function ψ_0 can be expanded in Slater determinants of various excitation rank (we use intermediate normalization): $\psi_0 = \psi_0^{(0)} + \textit{excitations}$. Then, by equalizing the two expressions for ΔE obtained above, we have

$$E_{corel} + E_{RHF} = \left\langle \psi_0^{(0)} | \hat{H} \psi_0 \right\rangle = \left\langle \psi_0^{(0)} | \hat{H} (\psi_0^{(0)} + \textit{excitations} \right\rangle = E_{RHF} + \left\langle \psi_0^{(0)} | \hat{H} (\textit{excitations}) \right\rangle,$$

hence

$$E_{corel} = \left\langle \psi_0^{(0)} | \hat{H} (\textit{excitations}) \right\rangle. \tag{2.109}$$

The Slater–Condon rules (Appendix V1-N, p. V1-707) show immediately that the only excitations which give nonzero contributions are the single and double excitations. Moreover, taking advantage of the Brillouin theorem, we obtain single excitation contributions exactly equal to zero. So we get the result that

> the exact correlation energy can be obtained using only that part of a formula for the CI wave function ψ_{CI}, which contains exclusively double excitations, i.e., $E_{corel} = \left\langle \psi_0^{(0)} | \hat{H} (\textit{double excitations only}) \right\rangle$.

The problem, however, lies in the fact that these doubly excited determinants are equipped with coefficients obtained in the *full* CI method, i.e., with all possible excitations. How is this? We should draw attention to the fact that, in deriving the formula for ΔE, intermediate normalization is used. If someone gave us the normalized FCI wave functions as a Christmas gift,[118] then the coefficients occurring in the formula for ΔE would not be the double excitation coefficients we need. We would have to denormalize this function to have the coefficient for the Hartree–Fock determinant equal to 1. We cannot do this without knowledge of the coefficients for higher excitations.

It is as if somebody said: the treasure is hidden in our room, but to find it we have to solve a very difficult problem in the Kingdom of Far Far Away. Imagine a compass which leads us

---

117 Also taking advantage of the intermediate normalization $\left\langle \psi_0^{(0)} | \psi_0 \right\rangle = 1$ and $\left\langle \psi_0^{(0)} | \psi_0^{(0)} \right\rangle = 1$ and the fact that ψ_0 is an eigenfunction of $\hat{H}$.

118 Dreams...

unerringly to that place in our room where the treasure is hidden. Perhaps a functional exists whose minimization would provide us directly with the solution, but we do not know it yet.[119]

NONVARIATIONAL METHODS USING EXPLICITLY CORRELATED WAVE FUNCTIONS

2.18 Møller–Plesset R12 method (MP2-R12)

As shown by Kato, the electronic wave function should take care of the cusp condition (p. 92). For example, the singlet wave function[120] for a two-electron system, when explicitly dependent on the interelectronic distance r_{12}, has to behave as

$$\psi_0 = \Phi(\mathbf{r}_1, \mathbf{r}_2)(1 + \frac{1}{2} r_{12} + ...), \tag{2.110}$$

where Φ is a symmetric function of the radius vectors $\mathbf{r}_1, \mathbf{r}_2$ for the two electrons, while "+..." means a polynomial with higher powers of r_{12}. Werner Kutzelnigg was first to remark[121] that this expression may be treated as a perturbational expansion for the ground-state wave function, and proposed to use the perturbation theory accurate through the second order in energy[122]

$$\psi_0 \simeq \psi_0^{(0)} + \psi_0^{(1)}, \tag{2.111}$$

$$E_0 \simeq E_0^{(0)} + E_0^{(1)} + E_0^{(2)}. \tag{2.112}$$

In computational practice the main problem of any workable perturbation theory is to select a suitable starting point, i.e., the zero-order wave function $\psi_0^{(0)}$, which has to satisfy the zero-order Schrödinger equation $\hat{H}^{(0)}\psi_0^{(0)} = E_0^{(0)}\psi_0^{(0)}$, with the unperturbed Hamiltonian $\hat{H}^{(0)}$. This defines the perturbation $\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}$. From that point on all the perturbational formulae necessary to calculate E_0 and ψ_0 are defined (see Chapter V1-5). In particular, the perturbational corrections to energy (accurate through the second order) are

$$E_0^{(1)} = \left\langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0^{(0)} \right\rangle, \tag{2.113}$$

---

119 It looks like the work by H. Nakatsuji, *Phys. Rev. A*, 14(1976)41 and M. Nooijen, *Phys. Rev. Letters*, 84(2000)2108 go in this direction.

120 Only its space part (symmetric) is given here. To obtain the total antisymmetric wave function one has to multiply Ψ by the spin factor $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$.

121 W. Kutzelnigg, *Theor. Chim. Acta*, 68(1985)445.

122 In the perturbational theory the first-order correction to the wave function ($\psi_0^{(1)}$) is sufficient to calculate the energy through the second order.

$$E_0^{(2)} = \left\langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0^{(1)} \right\rangle. \tag{2.114}$$

The first-order energy correction of Eq. (2.113) requires simply the calculation of the integral $\left\langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0^{(0)} \right\rangle$. To get $E_0^{(2)}$ one has first to have $\psi_0^{(1)}$. Kutzelnigg proposed to solve the problem for the helium-like series of atoms by assuming Φ in the form of a single Slater determinant, while taking the first-order correction to the wave function in the form

$$\psi_0^{(1)} = \frac{1}{2} r_{12} \Phi + \theta \tag{2.115}$$

with Φ and θ to be found. Since this corresponds to the second-order (in energy) perturbation theory with the Hartree–Fock wave function $\psi_0^{(0)}$, this approach is known as the Møller–Plesset r_{12} method (MP2-R12, see p. 169).

Well, but how to make this in practice? This form of $\psi_0^{(1)}$ takes care of the cusp only, but does not determine θ, which we should find somehow. There is a good news though: the best approximation to the $\psi_0^{(1)}$ function can be obtained by using the Hylleraas functional (Chapter V1-5)

$$\mathcal{E}\left[\tilde{\chi}\right] = \left\langle \tilde{\chi} \left| \left(\hat{H}^{(0)} - E_0^{(0)} \right) \tilde{\chi} \right\rangle + \left\langle \tilde{\chi} \left| \left(\hat{H}^{(1)} - E_0^{(1)} \right) \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} \right| \left(\hat{H}^{(1)} - E_0^{(1)} \right) \tilde{\chi} \right\rangle. \tag{2.116}$$

The functional attains its minimum precisely at $\tilde{\chi} = \psi_0^{(1)}$ and, moreover, for this function the value of the functional is equal to $E_0^{(2)}$, which is to be sought. The function θ is coupled to Φ by a "residual interaction" that has no singularity (already taken into account by $\frac{1}{2} r_{12} \Phi$) for $r_{12} = 0$. At a given choice of Φ one expands θ as a sum of products of one-electron basis functions with unknown coefficients. Minimization of $\mathcal{E}\left[\tilde{\chi}\right]$ enables one to calculate these coefficients, and then to calculate the corrections to the energy from Eqs. (2.113) and (2.114). In this way the target of the MP2-R12 method is achieved.

There are, however, two practical problems left. One is easy, the other is very difficult. The relatively easy one is how to choose the starting Φ. It turned out that the best reference function Φ for the helium-like atoms shows very little shielding and resembles closely the eigenstate of the bare nuclear Hamiltonian!

2.18.1 Resolution of identity (RI) method or density fitting (DF)

The second problem is how to calculate the appearing integrals. This problem becomes increasingly difficult for many-electron systems due to the r_{ij} from the correlation factors and $\frac{1}{r_{mn}}$ as well as the kinetic energy operators from the Hamiltonian. One gets molecular integrals

that no longer involve one (over three spatial coordinates) or two electrons (over six spatial coordinates) only, but instead we may have to do with very many-electron integrals.

A way to reduce the computation time by an order of magnitude (at the same accuracy) is what is known as the resolution of identity (RI) approximation or density fitting (DF). The main idea of the RI approximation is that products of orbital basis functions (here the coordinates of electron 1 are given as an example),

$$\varphi_i(1)\varphi_j(1) = \sum_n C_{nij}\chi_n(1), \tag{2.117}$$

can be approximated by a linear expansion of some auxiliary basis functions χ_n, usually centered on a single atom. The basis functions $\{\chi_n\}$ should in principle form a complete set, but in practice a truncation of the sum introduces an additional error. There is however an important profit: the integrals appearing are easy, because of a reduction in the number of centers.

2.18.2 Other RI methods

This idea has then been generalized for larger systems and other numerical methods[123]; e.g., Noga, Kutzelnigg, and Klopper[124] have presented the first formalism for MP4-R12, CCSD-R12, and CCSD[T]-R12. DF is used to approximate all of the four-index two-electron integrals. The resulting method, DF-MP2-R12, requires only two- and three-index integrals over various two-electron operators, and is extremely efficient.[125] The errors arising from the fitting process can be made small by using robust fitting formulae. Sample calculations on glycine (the smallest amino acid) reveal that for large basis sets DF-MP2-R12 is faster than a standard MP2 calculation and takes only a small fraction of the time required for the Hartree–Fock calculation. This may be viewed as a foundation of hope, judged unimaginable even a few years ago, that the explicitly correlated methods may become a draft horse of quantum chemical computations.

Ten-no introduced a frozen superposition of Gaussian geminals as the correlation factor[126] in the MP2-R12 method, which was fixed by the analytic electron–electron cusp conditions. Numerical stability in the DF expansions has been achieved by Valeev.[127] It turned out that the DF technique enables one to replace the frozen Gaussian geminal correlation factor by the Slater one.[128] In this way not only the electron–electron, but also the electron–nucleus cusp conditions were taken into account simultaneously.

123 See, e.g., W. Klopper, W. Kutzelnigg, *Chem. Phys. Lett.*, 134(1987)17; W. Kutzelnigg, W. Klopper, *J. Chem. Phys.*, 94(1991)1985.
124 J. Noga, W. Kutzelnigg, W. Klopper, *Chem. Phys. Lett.*, 199(1992)497.
125 F.R. Manby, *J. Chem. Phys.*, 119(2003)4607.
126 S. Ten-no, *J. Chem. Phys.*, 121(2004)117.
127 E.F. Valeev, *Chem. Phys. Lett.*, 395(2004)190.
128 S. Ten-no, *Chem. Phys. Lett.*, 398(2004)56.

Summary

- In the Hartree–Fock method *electrons of opposite spins do not correlate their motion*,[129] which is an absurd situation (in contrast to this, electrons of the same spins avoid each other, which is reasonable). In many cases (like the F_2 molecule, description of dissociation of chemical bonds, intermolecular interaction) this leads to wrong results. In this chapter we have learned about the methods which do take into account a correlation of electronic motions.

Variational methods using explicitly correlated wave functions

- rely on employing in the variational method a trial function which contains the explicit distance between the electrons. This improves the results significantly, but requires evaluation of *very* complex integrals.
- The correlation cusp condition $(\frac{\partial\psi}{\partial r})_{r=0} = \mu q_i q_j \psi(r=0)$ can be derived, where r is the distance of two particles with charges q_i and q_j, and μ is the reduced mass of the particles. This condition helps to determine the correct form of the wave function ψ. For example, for the two electrons the correct wave function has to satisfy (in a.u.) $(\frac{\partial\psi}{\partial r})_{r=0} = \frac{1}{2}\psi(r=0)$.
- The family of variational methods with explicitly correlated functions includes the Hylleraas method, the Hylleraas CI method, the James–Coolidge and Kołos–Wolniewicz approaches, and the method with exponentially correlated Gaussian functions. The method of explicitly correlated functions is very successful for two-, three-, and four-electron systems. For larger systems, due to the excessive number of complicated integrals, variational calculations are not yet feasible.

Variational methods with Slater determinants

- The configuration interaction (CI) approach is a *Ritz method* (Chapter V1-5) which uses the expansion in terms of *known* Slater determinants. These determinants are constructed from the molecular spin orbitals, usually occupied and virtual ones, produced by the Hartree–Fock method.
- Full CI expansion usually contains an enormous number of terms and is not feasible. Therefore, the CI expansion must be truncated somewhere. Usually we truncate it at a certain maximum rank of excitations with respect to the Hartree–Fock determinant (i.e., the Slater determinants corresponding to single, double, etc., up to some maximal excitations are included).
- Truncated (limited) CI expansion *is not size-consistent*, i.e., the energy of the system of noninteracting objects is not equal to the sum of the energies of the individual objects (calculated separately with the same truncation pattern).
- The multiconfiguration self-consistent field (MC SCF) method is similar to the CI scheme, but we vary not only *the coefficients in front of the Slater determinants, but also the Slater determinants themselves* (changing the analytical form of the orbitals in them). We have learned about two versions: the classic one (we optimize alternatively coefficients of Slater determinants and the orbitals) and a unitary one (we optimize simultaneously the determinantal coefficients and orbitals).
- The complete active space self-consistent field (CAS SCF) method is a special case of the MC SCF approach and relies on *selection* of a set of spin orbitals (usually separated energetically from others)

[129] Although they repel each other (mean field) as if they were electron clouds.

and on the construction from them of all possible Slater determinants within the MC SCF scheme. Usually low-energy spin orbitals are "*inactive*" during this procedure, i.e., they all occur in *each* Slater determinant (and are either frozen or allowed to vary). Most important active spin orbitals correspond to HOMO and LUMO.

Nonvariational method based on Slater determinants

- The coupled cluster (CC) method is an attempt to find such an expansion of the wave function in terms of the Slater determinants which would preserve size consistency. In this method the wave function for the electronic ground state is obtained as a result of the operation of the wave operator $\exp(\hat{T})$ on the Hartree–Fock function (this *ensures* size consistency). The wave operator $\exp(\hat{T})$ contains the cluster operator $\hat{T}$, which is defined as the sum of the operators for the l-tuple excitations, $\hat{T}_l$ up to a certain maximum $l = l_{\max}$. Each $\hat{T}_l$ operator is the sum of the operators each responsible for *a particular l-tuple excitation* multiplied by its *amplitude t*. The aim of the CC method is to find the t values, since they determine the wave function and energy. The method generates nonlinear (with respect to unknown t amplitudes) equations. The CC method usually provides very good results.
- The equation of motion coupled cluster (EOM-CC) method is based on the CC wave function obtained for the ground state and is designed to provide the electronic excitation energies and the corresponding excited-state wave functions.
- The many-body perturbation theory (MBPT) method is a perturbation theory in which the unperturbed system is usually described by a single Slater determinant. We obtain two basic equations of the MBPT approach for the ground-state wave function, i.e., $\psi_0 = \psi_0^{(0)} + \hat{R}_0 \left(E_0^{(0)} - E_0 + \hat{H}^{(1)} \right) \psi_0$ and $E_0 = E_0^{(0)} + \left\langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0 \right\rangle$, where $\psi_0^{(0)}$ is usually the Hartree–Fock function, $E_0^{(0)}$ is the sum of the orbital energies, $\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}$ is the fluctuation potential, and $\hat{R}_0$ is the reduced resolvent (i.e., "almost" inverse of the operator $E_0^{(0)} - \hat{H}^{(0)}$). These equations are solved in an iterative manner. Depending on the iterative procedure chosen, we obtain either the Brillouin–Wigner or the Rayleigh–Schrödinger perturbation theory. The latter is applied in the Møller–Plesset (MP) method.
- *One of the basic computational methods for the correlation energy is the MP2 method*, which gives the result correct through the second order of the Rayleigh–Schrödinger perturbation theory (with respect to the energy) at low cost.

Nonvariational methods using explicitly correlated wave functions

The road map in this direction has been designed by Werner Kutzelnigg, who proposed to use the Møller–Plesset perturbational approach (coined as MP-R12). One starts from the Hartree–Fock wave function and calculates the energy accurate through the second order using the Hylleraas variational principle. The method has been extended later to higher orders of perturbation theory. The idea of the explicitly correlated factors has been also exploited within the CC method. The notorious difficulty in calculating molecular integrals, which was the main obstacle in the past, has been diminished much by using a concept of the resolution of identity (that replaces a difficult integral by a sum of easier ones). This makes these methods currently among the most accurate ones in the field of quantum chemistry.

Main concepts, new terms

active orbitals (p. 144)
anticorrelation (p. 118)
Brillouin theorem (p. 136)
Brillouin–Wigner perturbation theory (p. 168)
Brueckner function (p. 135)
CC amplitudes (p. 152)
cluster operator (p. 151)
commutator expansion (p. 146)
complete active space (CAS) (p. 148)
configuration (p. 135)
configuration interaction (p. 135)
configuration mixing (p. 135)
correlation energy (p. 82)
Coulomb hole (p. 104)
coupled cluster (CC) (p. 149)
covalent structure (p. 127)
cusp condition (p. 91)
deexcitations (p. 160)
density matrix (p. 140)
direct method (p. 142)
EOM-CC method (p. 159)
exchange hole (p. 105)
explicit correlation (p. 89)
exponentially correlated function (p. 101)
Fermi hole (p. 105)
frozen orbitals (p. 144)
full CI method (p. 140)
geminal (p. 101)
harmonic helium atom (p. 94)
Heitler–London function (p. 127)
Hylleraas CI (p. 93)
Hylleraas function (p. 93)
inactive orbital (p. 144)
intermediate normalization (p. 151)
ionic structure (p. 127)
James–Coolidge function (p. 96)
Kołos–Wolniewicz function (p. 96)
many-body perturbation theory (MBPT) method (p. 162)
Møller–Plesset perturbation theory (p. 169)
multiconfigurational SCF methods (p. 145)
multireference methods (p. 142)
natural orbitals (p. 140)
Rayleigh–Schrödinger perturbation theory (p. 168)
reduced resolvent (p. 165)
resonance theory (p. 127)
similarity transformation (p. 159)
size consistency (p. 142)
unitary MC SCF method (p. 146)
vacuum state (p. 151)
valence bond (VB) method (p. 126)
wave operator (p. 151)

From the research front

The computational cost in the Hartree–Fock method scales with the size N of the atomic orbital basis set as N^4 and, while using devices similar to *direct* CI, even[130] as N^3. However, after making the Hartree–Fock computations for small (say, up to 10 electrons) systems, we perform more and more frequently calculations of the electronic correlation. The main approaches used to this end are the MP2 method and the CC method with single and double excitations in $\hat{T}$ and partial inclusion of triple ones (the so-called CCSD(T) approach). The CC method has been generalized for important cases involving chemical

[130] This reduction is caused mainly by a preselection of the two-electron integrals. The preselection allows us to estimate the value of the integral without its computation and to reject the large number of integrals of values close to zero.

bond breaking.[131] The state of the art in CC theory currently includes the full CCSDTQP model, which incorporates into the cluster expansion all the operators through pentuple excitations.[132] The formulae in these formalisms become monstrous to such an extent that scientists desperately invented an "anti-weapon" (artificial intelligence): first, automatic (computer-based) derivation of the formulae is used, followed by automatic coding of the derived formulae into executable programs (usually using Fortran). In such an approach we do not need even to see our formulae...

The computational cost of the CCSD scheme scales as N^6. The computational strategy often adopted relies on obtaining the optimum geometry of the system with a less sophisticated method (e.g., Hartree–Fock) and, subsequently, calculating the wave function for that geometry with a more sophisticated approach (e.g., the MP2 that scales as N^5, MP4 or CCSD(T) scaling as N^7). In the next chapter we will learn about density functional theory (DFT), which joins to some extent circumvents the abovementioned methods and is used for large systems.

Since the 2000s much progress has been made in methods using explicitly correlated functions. The progress is both conceptual and numerical. As an example may serve the calculations on glycine,[133] where it turned out that, for large basis sets, a version of the MP2-R12 method is faster than a standard MP2 calculation and takes only a small fraction of the time for the Hartree–Fock calculation!

Recoupling quantum chemistry with nuclear forces

The CC method has been designed first in the field of nuclear physics. This fact, however, had no consequences until recent years, since the numerical procedure has been judged by the community as intractable. Only because the quantum chemist Jiří Čížek accidentally looked up a nuclear physics journal, the idea diffused to the quantum chemistry community and after some spectacular developments turned out to become the most successful in studying atoms and molecules. Imagine that the idea went back to nuclear physics from quantum chemistry. The quantum chemistry CC technique has been applied to compute the energy levels for nucleons in several nuclei with much higher precision than that accessible by the nuclear physics theoretical methods.[134]

Nakatsuji strategy

Hiroshi Nakatsuji looked at the Schrödinger equation from an unexpected side.[135] He wrote two equations, i.e.,

$$\left\langle \delta\psi | (\hat{H} - E)\psi \right\rangle = 0, \tag{2.118}$$

$$\left\langle \psi | (\hat{H} - E)^2\psi \right\rangle = 0, \tag{2.119}$$

and asked what their relation is to the Schrödinger equation $(\hat{H} - E)\psi = 0$.

131 P. Piecuch, M. Włoch, *J. Chem. Phys.*, 123(2005)224105.

132 M. Musiał, S.A. Kucharski, R.J. Bartlett, *J. Chem. Phys.,* 116(2002)4382.

133 F.R. Manby, *J. Chem. Phys*., 119(2003)4607.

134 M. Włoch, D.J. Dean, J.R. Gour, M. Hjorth-Jensen, K. Kowalski, T. Papenbrock, P. Piecuch, *Phys. Rev. Letters*, 94(2005)212501.

135 H. Nakatsuji, *J. Chem. Phys*., 113(2000)2949.

Hiroshi Nakatsuji, professor at Kyoto University, Japan, then professor at Quantum Chemistry Research Institute, Kyoto. We had the following conversation in Warsaw. Me: "You are a mathematician I presume?" Professor Nakatsuji: "No, I am just an ordinary organic chemist!"

Note that (2.118) follows from minimizing the functional $\left\langle \psi | \hat{H} \psi \right\rangle$ under a normalization constraint[136] ($\langle \psi | \psi \rangle = 1$) of the trial function ψ. This is the essence of the variational method described in Chapter V1-5. Satisfaction of (2.118) may happen *either* because ψ fulfills the Schrödinger equation, *or*, if ψ does not satisfy the Schrödinger equation, but is optimal within the variational method restricted to a class of variations[137] $\delta\psi$. Anyway, if ψ satisfies (2.118), it not necessarily represents a solution to the Schrödinger equation, it does with no restrictions imposed on $\delta\psi$.

Eq. (2.119) has a different status: it is satisfied only for the solution ψ of the Schrödinger equation.[138] Unfortunately, it contains the square of the Hamiltonian. This smells like difficult integrals to be calculated in the future, but for the time being we are going forward courageously.

Imagine that the variation of ψ in Eq. (2.118) was chosen to have a very special form

$$\delta\psi = (\hat{H} - E)\psi \cdot \delta C, \tag{2.120}$$

where C is a variational parameter in ψ. Then from Eq. (2.118) we have the precious Eq. (2.119),

$$\left\langle (\hat{H} - E)\psi | (\hat{H} - E)\psi \right\rangle \cdot \delta C^* = 0,$$

and in such a case[139] $(\hat{H} - E)\psi = 0$ (solution of the Schrödinger equation). It is seen, therefore, that the right-hand side of Eq. (2.120) in a sense "forces" a correct structure of the wave function, hopefully also when we take an approximation instead of the exact (and unknown) energy E. Having this in mind, let us construct a variational function satisfying (2.120). But how to get this? Well, let us begin an iteration game with functions (n stands for the number of iterations, $\delta\psi$ represents an analog of $\psi_{n+1} - \psi_n$, we define $\bar{E}_n \equiv \left\langle \psi_n | \hat{H} \psi_n \right\rangle$)

$$\psi_{n+1} = \left[1 + C_n(\hat{H} - \bar{E}_n)\right] \psi_n. \tag{2.121}$$

We start from an arbitrary function ψ_0 (not to be mixed with the exact ground-state wave function) and in each iteration we determine variationally the value of the coefficient C_n. We hope the procedure

136 A conditional minimum can be found by using the Lagrange multipliers method (Appendix V1-O).

137 If no restriction is imposed, the function found satisfies the Schrödinger equation.

138 Indeed, $\left\langle \psi | (\hat{H} - E)^2 \psi \right\rangle = \left\langle \left(\hat{H} - E\right) \psi | (\hat{H} - E)\psi \right\rangle = ||(\hat{H} - E)\psi||^2 = 0$, where $||(\hat{H} - E)\psi||$ is the vector length. The latter is equal to 0 only if all the components of the vector are equal to 0. This means that in any point of space we have $(\hat{H} - E)\psi = 0$.

139 Because of the arbitrariness of δC.

converges, i.e., what we get as the left-hand side is the function inserted into the right-hand side. If this happens, we achieve the satisfaction of

$$\psi = \left[1 + C(\hat{H} - E)\right]\psi, \tag{2.122}$$

where we have removed the lower indices, because they do not matter at convergence. For $C \neq 0$ this means the achievement of our aim, i.e., $(\hat{H} - E)\psi = 0$.

As it turned out this recipe needs some corrections when applied in practical calculations. In order to be able to calculate the integrals $\bar{E}_n = \left\langle \psi_n | \hat{H} \psi_n \right\rangle$ safely,[140] Nakatsuji considered what is known as the scaled Schrödinger equation[141]

$$g(\hat{H} - E)\psi = 0, \tag{2.123}$$

instead of the original one (satisfied by the same ψ), where the arbitrary function (of the electronic coordinates) g does not commute with the Hamiltonian and must be positive everywhere, except points of singularity, but even approaching a singularity it still has to be $\lim gV \neq 0$. Thus, the philosophy behind function g is to destroy the "singularity character in singularities" and at the same time not to destroy the precious information about these singular points, present in the potential energy V. Several possibilities have been tested, e.g., $g = \frac{1}{-V_{ne}+V_{ee}}$ or $g = -\frac{1}{V_{ne}V_{ee}}$, etc., where V_{ne} and V_{ee} are the Coulomb potential energy of the electron–nucleus and electron–electron interactions.[142]

The results witness about great effectiveness of this iterative method. For example,[143] in a little more than 20 iterations the Schrödinger equation was practically solved (with nearly 100% of the correlation energy within finite basis sets) for the molecules HCHO, CH_3F, HCN, CO_2, and C_2H_4. Analytical calculations[144] for H_2 within four to six iterations gave the electronic energy (at the equilibrium distance) with 15 significant figures (independently of several tested starting functions ψ_0). Similar calculations for the helium atom gave over 40 digits accuracy.[145]

No doubt, Nakatsuji's idea not only represents a fresh look on quantum theory, but also has significant practical power. It remains to learn what the complicated final form of the wave function is telling us. This however pertains also to wave functions produced by many other methods.

Full CI quantum Monte Carlo (FCIQMC) and CC

The full CI expansion may be of practical value only for very small systems and quite poor basis sets. The reason is an exponential growth of the number of expansion functions (Slater determinants). Are these

140 We have to calculate the mean values of higher and higher powers of the Hamiltonian. These integrals are notorious for diverging.

141 H. Nakatsuji, *Phys. Rev. Lett.*, 93(2004)30403. In this reference the Nakatsuji standard method is described.

142 The integration difficulty can be circumvented also by considering satisfaction (in points of space) of the Schrödinger equation in the form $\frac{\hat{H}\psi}{\psi} = \text{const}$, as described in H. Nakatsuji, H. Nakashima, Y. Kurokawa, A. Ishikawa, *Phys. Rev. Lett.*, 99(2007)240402.

143 H. Nakatsuji, *Bull. Chem. Soc. Japan*, 78(2005)1705.

144 Iterations result in a (nested) analytical form of the wave function.

145 H. Nakashima, H. Nakatsuji, *J. Chem. Phys.*, 127(2007)224104.

limitations possible to be overcome? It turned out that there is a relatively low-cost avenue to achieve this goal.

The imaginary-time Schrödinger equation

Let us suppose that we have a system with the time-independent Hamiltonian $\hat{H}$.

We are free to add an arbitrary number to the potential energy, this does not change physics or chemistry. Let us do so. The exact energies will form an ascending order: $0 \leq E_0 \leq E_1 \leq E_2 \ldots$. They come from the solutions of the Schrödinger equation (and are real and nonnegative)

$$\hat{H}\psi_n(x) = E_n\psi_n(x). \tag{2.124}$$

The corresponding time-dependent stationary states are (see Chapter V1-2)

$$\Psi_n(x,t) = \psi_n(x)\exp(-iE_nt). \tag{2.125}$$

The $\psi_n(x)$ and E_n are unknown, and *our target will be the ground-state energy E_0 and the ground-state wave function $\psi_0(x)$*.

Now, let us write the time-dependent Schrödinger equation (in a.u.), Chapter V1-2

$$\hat{H}\psi(x,t) = i\frac{\partial\psi(x,t)}{\partial t}, \tag{2.126}$$

with $\psi(x,0) = \phi_0(x)$. If one introduces the imaginary time $t = -i\tau$ (i.e., $it = \tau$), one changes the Schrödinger equation into a propagation of a function $\psi(x,\tau)$ (when changing the parameter τ)

$$\hat{H}\psi(x,\tau) = -\frac{\partial\psi(x,\tau)}{\partial\tau}, \tag{2.127}$$

with $\psi(x,\tau=0) \equiv \phi_0(x)$.

Formally its solution reads as[146] $\psi(x,\tau) = \exp(-\tau\hat{H})\phi_0(x)$. Eq. (2.127) will be treated by us as a machinery to produce the ground-state ψ_0. The corresponding proof is simple.[147] By expanding $\phi_0(x) = \sum_n c_n\psi_n$ one gets $\psi(x,\tau) = \exp(-\tau\hat{H})\sum_n c_n\psi_n = \sum_n c_n\exp(-\tau\hat{H})\psi_n = \sum_n c_n\exp(-\tau E_n)\psi_n$.

If $\tau \to \infty$, we obtain

$\psi(x,\tau=\infty) = c_0\psi_0$, because other terms $\exp(-\tau E_n)$ vanish with increasing τ.

Thus, when using imaginary time evolution, Eq. (2.127), we are able to get an exact ground-state ψ_0 starting from $\phi_0(x)$ and going with τ to infinity.

---

146 Indeed, taking the right-hand side of Eq. (2.127) we have $-\frac{\partial\psi(x,\tau)}{\partial\tau} = -\frac{\partial\exp(-\tau\hat{H})\phi_0(x)}{\partial\tau} = -\exp(-\tau\hat{H})(-\hat{H})\phi_0(x) = \hat{H}\exp(-\tau\hat{H})\phi_0(x) = \hat{H}\psi(x,\tau)$.

147 It involves the unknown stationary wave functions, which form the complete set.

Full CI quantum Monte Carlo[148]

We decide to obtain the solution of Eq. (2.127) by using a kind of stochastic game. First, let us consider the Hilbert space for the system under consideration as spanned by all possible Slater determinants $\{\chi_j\}$. Function ψ_0 represents a particular vector in this space $\psi_0 = \sum_j a_j^{(0)} \chi_j$ and our goal is to find the coefficients $a_j^{(0)}$. We do not have them. Instead we have an initial set (an "educated guess") of $a_j^{\text{init}}(\tau = 0)$ in the expansion

$$\psi(x, \tau) = \sum_j a_j(\tau) \chi_j. \tag{2.128}$$

What to do? Well, we have an indication how $\psi(x, \tau)$ and therefore $a_j(\tau)$ should change if τ increases a bit. After inserting (2.128) in Eq. (2.127), multiplying by χ_i^*, and integrating, we get (the matrix $H_{ij} \equiv \left\langle \chi_i | \hat{H} \chi_j \right\rangle$ is supposed to be precalculated)

$$-\frac{\mathrm{d}a_i(\tau)}{\mathrm{d}\tau} = a_i(\tau) H_{ii} + \sum_{j(\neq i)} a_j(\tau) H_{ij}. \tag{2.129}$$

From the left-hand side we know whether to increase or decrease each of the coefficients $a_i(\tau)$ to obtain $a_i(\tau + \Delta\tau)$.

This may be done in a stochastic way. The Slater determinant coefficients $a_i(\tau)$ are discretized assigning to each determinant a certain number of "walkers," i.e., δa contributions (positive or negative). The walkers "occupy" each Slater determinant, which means a coarse-grained snapshot of the wave function amplitudes. The walkers inhabit Slater determinant space and evolve according to a simple set of rules, which include spawning, death, and annihilation processes (walker annihilation plays a key role). When the imaginary time τ goes on, the occupation by the walkers changes according to the master equation (2.129) until it converges, i.e., statistically stops changing. This means the computed amplitudes $a_i(\tau = \infty) = a_i^{(0)}$ and therefore ψ_0 are found. No diagonalization – typical for the CI method – is involved in the procedure, and no tedious CC equations are to be solved. The wave function can be resolved to arbitrary accuracy (in principle FCI quality, as verified for the small molecules CO, O_2, CH_4, and NaH).

The QMC method certainly represents a conceptual breakthrough, but the way to FCI quality is still quite difficult. The method described below makes this way easier and more precise.[149] The improving procedure is shown below in three steps.

Step 1. The QMC procedure with solving the imaginary-time Schrödinger equation is used for producing the main excitations in the CI expansion fast ("initial selection P," nonexpensive).

Step 2. This selection of excitations is then used in the CC method, when constructing the $\hat{T}$ cluster operator of Eq. (2.57) on p. 152 (thus approximating $\hat{T}$ by $\hat{T}^{(P)}$) and one proceeds along the CC formalism, i.e., finding the corresponding $t^{(P)}$ amplitudes (see p. 155).

148 G.H. Booth, A.J.W. Thom, A. Alavi, *J. Chem. Phys.*, 131(2009)054106.

149 J.E. Deustua, J. Shen, P. Piecuch, *Phys. Rev. Letters*, 119(2017)223003.

Step 3. Now one concentrates on a set Q of other excitations, not taken into account so far. We plan to include them, but in an approximate and inexpensive way. We estimate their energy contribution through the second-order perturbation theory (i.e., noniteratively, see p. V1-279), treating the wave function of step 1 as the zero-order approximation.

The calculated perturbational approximate contribution is then added to the ground-state energy of step 1, which produces the ground-state approximate energy $E_0^{(P+Q)}$, thus completing the CC procedure denoted as CC(P;Q).

The FCI method is sometimes given as an example showing that quantum mechanics is prohibitively fast complex in its very nature, i.e., the numerical effort grows exponentially with the system's size (because of the "exploding number" of excitations and therefore of the expansion functions). This point of view changes nowadays. The very reason is the Hamiltonian matrix sparsity (most elements are equal to zero), which dumps very effectively the exponential effort, still being able to give chemically accurate results. First the stochastic FCIQMC gives the main ingredients of the wave function at low cost. Then, as it turns out, it is sufficient afterwards to grasp perturbationally the remainder of the energy at a cost growing only as N^6. This approach designed by Piotr Piecuch and collaborators seems to be currently the most efficient among accurate quantum chemical methods.

Ad futurum

Experimental chemistry is focused, in most cases, on molecules of *larger* size than those for which fair calculations with correlation are possible. However, recently this opinion is changing due to enormous progress achieved in quantum chemistry. The FCIQMC methods, based on the CI concept with a stochastically determined first guess and then perturbational improvement, turn out to be very accurate and still feasible. Also, quite unexpectedly, the explicitly correlated wave functions are becoming applicable in chemical practice. Also, after thorough analysis of the situation, it turns out that the cost of the calculations does not necessarily increase very fast with the size of a molecule.[150] Employing localized molecular orbitals and using the multipole expansion (see Appendix G) of the integrals involving the orbital separated in space causes, for elongated molecules, the cost of the post-Hartree–Fock calculations to scale linearly with the size of a molecule.

There is one more problem which will probably be faced by quantum chemistry when moving to larger molecules containing heteroatoms. Nearly all the methods, including electron correlation, described so far (with the exception of the explicitly correlated functions) are based on the silent and pretty "obvious" assumption that the higher the excitation we consider, the higher the configuration energy we get. This assumption seems to be satisfied so far, but the molecules considered were always small, and the method has usually been limited to a small number of excited electrons. This assumption can be challenged in certain cases.[151] The multiple excitations in large molecules containing easily polarizable fragments can result in electron transfers which cause energetically favorable strong electrostatic interactions ("mnemonic effect"[152]) which lower the energy of the configuration. The reduction can be large

150 H.-J. Werner, *J. Chem. Phys.*, 104(1996)6286.

151 There are exceptions though; see A. Jagielska, L. Piela, *J. Chem. Phys.*, 112(2000)2579.

152 L.Z. Stolarczyk, L. Piela, *Chem. Phys. Letters*, 85(1984)451.

enough to make the energy of the *formally* multiply excited determinant close to that of the Hartree–Fock determinant. Therefore, it should be taken into account on the same footing as the Hartree–Fock. This is rather unfeasible for the methods discussed above.

The explicitly correlated functions have a built-in adjustable and efficient basic mechanism accounting for the correlation within the interacting electronic pair. The mechanism is based on the obvious notion that electrons should avoid each other.[153]

Let us imagine the CH_4 molecule. Let us look at it from the viewpoint of localized orbitals. With the method of explicitly correlated geminal functions for bonds we would succeed in making the electrons avoid each other within the same bond. And what should happen if the center of mass of the electron pair of one of the bonds shifts towards the carbon atom? The centers of mass of the electron pairs of the remaining three bonds should move away along the C–H bonds. The wave function must be designed in such a way that it accounts for this. In current theories, this effect is either deeply hidden or entirely neglected. A similar effect may happen in a polymer chain. One of the natural correlations of electronic motions should be a shift of electron pairs of all bonds in the same phase. As a highly many-electron effect the latter is neglected in current theories. However, the purely correlational Axilrod–Teller effect in the case of linear configuration, discussed in Chapter 5 (three-body dispersion interaction in the third order of perturbation theory), suggests clearly that the correlated motion of many electrons should occur.

Additional literature

A. Szabo, N.S. Ostlund, "Modern Quantum Chemistry," McGraw-Hill, New York, 1989, p. 231–378.

Excellent book.

T. Helgaker, P. Jørgensen, J. Olsen, "Molecular Electronic-Structure Theory," Wiley, Chichester, 2000, p. 514.

Practical information on the various methods accounting for electron correlation presented in a clear and competent manner.

Questions

1. The Hartree–Fock method:
   a. describes the electrons with their positions being completely independent.
   b. introduces correlation of motion of electrons with the same spin coordinate.
   c. does not introduce any correlation of motion of electrons with the opposite spin coordinates.
   d. ignores the Coulomb hole, but takes care of the Fermi hole.
2. For the ground state of the helium atom in the Hartree–Fock method:

[153] In special conditions one electron can follow the other together forming a Cooper pair. The Cooper pairs are responsible for the mechanism of superconductivity. This will be a fascinating field of research for chemist-engineered materials in the future.

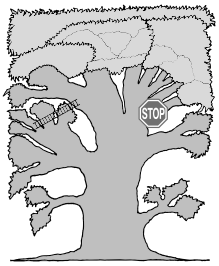
a. if one electron is on the nucleus, the probability of finding the second one in a small volume dV is also the largest on the nucleus.
b. if electron 1 is on one side of the nucleus, electron 2 is easiest to find on the nucleus.
c. if both electrons are at the same distance from the nucleus, it is equally easy to find them in the same point as in two points opposite to each other with respect to the nucleus.
d. if both electrons are at the same distance from the nucleus, they will tend to be on opposite sides of the nucleus.

3. The CI method truncated at double excitations gives energy E_{BeBe} for two beryllium atoms at large distance R. In calculations by using this method:
   a. if $R \longrightarrow \infty$ we will have $E_{BeBe} = 2E_{Be}$.
   b. if $R \longrightarrow \infty$ one will obtain $E_{BeBe} - 2E_{Be} = const \neq 0$.
   c. if $R \longrightarrow \infty$ one will get $E_{BeBe} = 2E_{Be}$, but under the condition that the CI calculation for the individual beryllium atom was limited to single excitations.
   d. the result obtained contains an error coming from the size inconsistency.
4. The CC method (with the cluster operator truncated at double excitations) gives the energy E_{BeBe} for two beryllium atoms at very large internuclear distance R and the energy E_{Be} for a single beryllium atom. In the calculations using this method:
   a. if $R \longrightarrow \infty$ there will be $E_{BeBe} = 2E_{Be}$.
   b. if $R \longrightarrow \infty$ one will obtain $E_{BeBe} - 2E_{Be} = const \neq 0$.
   c. if $R \longrightarrow \infty$ one will get $E_{BeBe} = 2E_{Be}$, but under the condition that the CC calculation for the individual beryllium atom was limited to single excitations.
   d. the result obtained contains an error coming from the size inconsistency.
5. The cusp condition for collision of two charged particles (μ means the reduced mass, all quantities are given in a.u.):
   a. follows from the requirement that a wave function cannot acquire infinite values.
   b. for an electron and an atomic nucleus of charge Z reads as $(\frac{\partial \psi}{\partial r})_{r=0} = Z\psi(r=0)$.
   c. for two electrons $(\frac{\partial \psi}{\partial r})_{r=0} = \frac{1}{2}\psi(r=0)$.
   d. for any two particles with charges q_1 and q_2, $(\frac{\partial \psi}{\partial r})_{r=0} = \mu q_1 q_2 \psi(r=0)$.
6. The wave function $\psi(\mathbf{r}_1, \mathbf{r}_2) = N\left(1 + \frac{1}{2}r_{12}\right)\exp[-\frac{1}{4}\left(r_1^2 + r_2^2\right)]$ (N stands for the normalization constant, $\mathbf{r}_1$ and $\mathbf{r}_2$ denote the radius vectors for two electrons, and r_{12} means their distance) represents:
   a. an exact wave function for harmonium ("harmonic helium atom") with the force constant equal to $\frac{1}{4}$.
   b. an orbital occupied by electrons 1 and 2.
   c. a product of two orbitals.
   d. a geminal that takes into account the Coulomb hole.
7. A helium atom with an approximate wave function (see question 6): $\psi(\mathbf{r}_1, \mathbf{r}_2) = N\left(1 + \frac{1}{2}r_{12}\right)\exp[-\frac{1}{4}\left(r_1^2 + r_2^2\right)]$. From this function it follows that:
   a. if the nucleus–electron distance is the same for the two electrons, the electrons will have a tendency to be more often on the opposite sides of the nucleus.
   b. finding the electrons at the same point in space is more probable for smaller nucleus–electron distances.
   c. it takes into account the Fermi hole.
   d. the electrons are always on the opposite sides of the nucleus.

8. An intermediate normalization of the wave function ψ_0 and the normalized function $\psi_0^{(0)}$ means that:
   a. $\left\langle \psi_0 | \psi_0 - \psi_0^{(0)} \right\rangle = 1.$
   b. the Hilbert space vector ψ_0 is composed of the unit vector $\psi_0^{(0)}$ plus some vectors that are orthogonal to $\psi_0^{(0)}$.
   c. $\langle \psi_0 | \psi_0 \rangle \neq 1.$
   d. $\left\langle \psi_0 | \psi_0^{(0)} \right\rangle = 1.$
9. The Møller–Plesset method known as MP2:
   a. is equivalent to the Ritz variational method (CI procedure) with the double excitations only.
   b. is based on the perturbational approach with the Hartree–Fock wave function as the unperturbed.
   c. represents a perturbational approach with calculation of the electronic energy up to the second order; the zero-order plus the first-order energies give the Hartree–Fock energy.
   d. in this method the zero-order electronic energy represents a sum of the orbital energies of all spin orbitals present in the Hartree–Fock Slater determinant.
10. To calculate the electronic correlation energy:
   a. it is sufficient to carry out calculation within the Hartree–Fock method, and then to perform the full CI computation.
   b. it is sufficient to know the Hartree–Fock energy and all ionization potentials for the system.
   c. one has to use an explicitly correlated variational wave function.
   d. it is sufficient to know a wave function expansion containing only the double excitations, but with their CI coefficients obtained in the presence of all excitations.

Answers

1b,c,d, 2a,b,c, 3b,d, 4a, 5a,c,d, 6a,d, 7a,b, 8b,c,d, 9b,c,d, 10a,d



CHAPTER 3

Chasing the Correlation Dragon: Density Functional Theory (DFT)

As I observe, meditate and pray
Are we like clouds on a summer's day?...
Ruth Oliver, "Clouds"

Where are we?

We are on an upper right-hand side branch of the TREE.

An example

A metal represents a system that is very difficult to describe using the quantum chemistry methods given so far. The restricted Hartree–Fock method here offers a very bad, if not pathological, approximation (cf. Chapter V1-8, p. V1-511), because the HOMO-LUMO gap is equal to zero in metals. The methods based on the Slater determinants (CI, MC SCF, CC, etc., Chapter 2) are ruled out as involving a giant number of excited configurations to be taken into account, because of the continuum of the occupied and virtual energy levels (see Chapter 1). Meanwhile, in the past some properties of metals could be obtained, from simple theories that assumed that the electrons in a metal behave similarly to a homogeneous electron gas (also known as jellium), and the nuclear charge (to make the whole system neutral) has been treated as smeared out uniformly in the metal volume. There has to be something physically important captured in such theories.

What is it all about?

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00011-9
Copyright © 2020 Elsevier B.V. All rights reserved.

- Correlation dragon resides in electron density: equivalence of Ψ_0 and ρ_0
- A secret of the correlation dragon: the existence of energy functional minimized by ρ_0

The preceding chapter has shown how difficult it is to calculate the correlation energy. Basically there are two approaches: either to follow configuration interaction-type methods (CI, MC SCF, CC, etc.), or to go in the direction of explicitly correlated functions. The first means a barrier of more and more numerous excited configurations to be taken into account, the second means very tedious and time consuming integrals. In both cases we know the Hamiltonian and fight for a satisfactory wave function (often using the variational principle, Chapter V1-5). It turns out that there is also a third direction (presented in the present chapter) that does not regard electronic configurations (except a single special one) and does not have the bottleneck of difficult integrals. Instead, we have the kind of wave function in the form of a single Slater determinant, but we have a serious problem in defining the proper Hamiltonian.

The ultimate goal of the DFT method is calculation of the total energy of the system and the ground-state electron density distribution without using the wave function of the system.

Why is this important?

DFT calculations (despite taking electronic correlation into account) are not expensive; their cost is comparable with that of the Hartree–Fock method. Therefore, the same computer power allows us to explore much larger molecules than with other post-Hartree–Fock (correlation) methods.

What is needed?

- The Hartree–Fock method (Chapter V1-8, necessary),
- the perturbational method (Chapter V1-5, advised),
- Lagrange multipliers (Appendix V1-O, p. V1-719, advised).

Classical works

The idea of treating electrons in metal as an electron gas was conceived in 1900, independently by Lord Kelvin[1] and by Paul Drude.[2] ★ The concept explained the electrical conductivity of metals, and was then used by Llewellyn Hilleth Thomas in "*The Calculation of Atomic Fields*," published in *Proceedings of the Cambridge Philosophical Society*, 23(1926)542 as well as by Enrico Fermi in "*A Statistical Method for the Determination of Some Atomic Properties and the Application of this Method to the Theory of the Periodic System of Elements*" in *Zeitschrift für Physik*, 48(1928)73. They (independently) calculated the electronic kinetic energy per unit volume of the electron gas (this is therefore the *kinetic energy density*) as a function of the local electron density ρ. ★ In 1930 Paul Adrien Maurice Dirac presented a similar result in "*Note on the Exchange Phenomena in the Thomas Atom*," *Proceedings of the Cambridge Philosophical Society*, 26(1930)376 for the *exchange energy* as a function of ρ. ★ In a classic paper "*A Simplification of the Hartree–Fock Method*," published in *Physical Review*, 81(1951)385, John Slater showed that the Hartree–Fock method applied to metals gives the exchange energy density proportional to $\rho^{\frac{1}{3}}$. ★ For classical positions specialists often use a book by Pál Gombas "*Die statistische Theorie des Atoms und ihre Anwendungen*," Springer Verlag, Vienna, 1948. ★ The contemporary theory was born in 1964–1965, when two fundamental works appeared: Pierre Hohenberg and Walter Kohn in *Physical Review*, 136(1964)*B*864 entitled "*Inhomogeneous Electron Gas*" and Walter Kohn and Lu J. Sham in *Physical Review A*, 140(1965)1133 under the title "*Self-Consistent Equations including Exchange and Correlation Effects*." ★ Mel Levy in "*Electron Densities in Search of Hamiltonians*," published in *Physical Review A*, 26(1982)1200, proved that the variational principle in quantum chemistry can be equivalently presented as a minimization of the Hohenberg–Kohn functional that depends on the electron density ρ. ★ Richard F.W. Bader in 1994 wrote a book on mathematical analysis of the electronic density "*Atoms in Molecules. A Quantum Theory*," Clarendon Press, Oxford, that enabled chemists to look at molecules in a synthetic way, independently of the level of theory that has been used to describe it. ★ Erich Runge and Eberhard K.U. Gross in a paper "*Density-Functional Theory for Time-Dependent Systems*," published in *Physical Review Letters*, 52(1984)997, have extended the Hohenberg–Kohn–Sham formalism to the time domain.

* * *

[1] Or William Thomson (1824–1907), British physicist and mathematician, professor at the University of Glasgow. His main contributions are in thermodynamics (the second law, internal energy), theory of electric oscillations, theory of potentials, elasticity, and hydrodynamics. His great achievements were honored by the title of Lord Kelvin (1892).

[2] Paul Drude (1863–1906), German physicist, professor at the universities in Leipzig, Giessen, and Berlin.

3.1 Electronic density – the superstar

In the DFT method we begin from the Born–Oppenheimer approximation, which allows us to consider the electronic motion assuming some fixed positions of the nuclei. We will be interested in the ground state of the system.

Density matrix and electron density

Let us introduce a notion of the *first-order density matrix*[3] (*label* α stands here for the spin coordinate $\sigma = \frac{1}{2}$, β means $\sigma = -\frac{1}{2}$)

$$\rho(\mathbf{r};\mathbf{r}') = \rho_\alpha(\mathbf{r};\mathbf{r}') + \rho_\beta(\mathbf{r};\mathbf{r}'), \tag{3.1}$$

which we define as

$$\rho_\sigma(\mathbf{r};\mathbf{r}') = N\int \mathrm{d}\tau_2\mathrm{d}\tau_3...\mathrm{d}\tau_N \Psi^*(\mathbf{r}',\sigma,\mathbf{r}_2,\sigma_2,...,\mathbf{r}_N,\sigma_N)\Psi(\mathbf{r},\sigma,\mathbf{r}_2,\sigma_2,...,\mathbf{r}_N,\sigma_N). \tag{3.2}$$

Thus, in Eq. (3.2) we integrate $N\Psi^*\Psi$ over all spatial and spin electron coordinates except of those of electron number 1, for which just to preserve an additional mathematical freedom, we assign *two distinct positions*, $\mathbf{r}$ and $\mathbf{r}'$, while still fixing the same $\sigma_1 = \sigma$.

The key quantity in this chapter will be the (physically observable) electron density $\rho(\mathbf{r})$

$$\rho(\mathbf{r}) = \rho_\alpha(\mathbf{r}) + \rho_\beta(\mathbf{r}) \equiv \sum_\sigma \rho_\sigma(\mathbf{r};\mathbf{r}), \tag{3.3}$$

which represents the diagonal element of $\rho(\mathbf{r};\mathbf{r}')$, i.e., $\rho(\mathbf{r};\mathbf{r}) \equiv \rho(\mathbf{r})$ with $\rho_\sigma(\mathbf{r};\mathbf{r}) \equiv \rho_\sigma(\mathbf{r})$.

The wave function Ψ is antisymmetric with respect to the exchange of the coordinates of any two electrons, and, therefore, $|\Psi|^2$ is symmetric with respect to such an exchange. Hence, the definition of ρ is independent of the label of the electron we do not integrate over. According to this definition,

[3] The "indices" of this "matrix element" are $\mathbf{r}$ and $\mathbf{r}'$.

ρ represents nothing else but the density of the electron cloud carrying N electrons, because from the normalization condition of function Ψ we have (the integration is over the whole three-dimensional space)

$$\int \rho(\mathbf{r})\mathrm{d}\mathbf{r} = N. \tag{3.4}$$

Therefore, the electron density distribution $\rho(\mathbf{r})$ is given for a point $\mathbf{r}$ in the unit of number of electrons per volume unit. Since $\rho(\mathbf{r})$ represents an integral of a nonnegative integrand, $\rho(\mathbf{r})$ is always nonnegative. Let us check that ρ may be also defined as the mean value of the *electron density operator*

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}), \tag{3.5}$$

being a sum of the Dirac delta operators (cf. Appendix V1-E on p. V1-659) for individual electrons at position $\mathbf{r}$. We have

$$\langle \Psi | \hat{\rho} \Psi \rangle = \langle \Psi | \Big(\sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \Big) \Psi \rangle = \sum_{i=1}^{N} \langle \Psi | \big(\delta(\mathbf{r}_i - \mathbf{r}) \big) \Psi \rangle = \sum_{i=1}^{N} \frac{\rho(\mathbf{r})}{N} = \frac{\rho(\mathbf{r})}{N} N = \rho(\mathbf{r}). \tag{3.6}$$

Indeed, each of the integrals in the summation[4] is equal to $\rho(\mathbf{r})/N$, and the summation over i gives N; therefore, we obtain $\rho(\mathbf{r})$.

In orbital approximation

If the function Ψ is taken as a normalized Slater determinant built of N spin orbitals ϕ_i, from Slater–Condon rule I (Appendix V1-N, we replace there $\hat{h}$ by δ) for $\langle \Psi | \big(\sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \big) \Psi \rangle$ we obtain (after renaming the electron coordinates in the integrals, all the integrations are over the spatial and spin coordinates of electron 1)[5]

[4] Remember $\langle | \rangle$ means integration over space coordinates and summation over spin coordinates.

[5] This expression is invariant with respect to any unitary transformation of the molecular orbitals (cf. Chapter V1-8).

$$\begin{aligned}\rho(\mathbf{r}) &= \langle\phi_1(1)\,|\delta(\mathbf{r}_1-\mathbf{r})\phi_1(1)\rangle_1 + \langle\phi_2(1)\,|\delta(\mathbf{r}_1-\mathbf{r})\phi_2(1)\rangle_1 + \\ &\quad \ldots\langle\phi_N(1)\,|\delta(\mathbf{r}_1-\mathbf{r})\phi_N(1)\rangle_1 \\ &= \sum_{i=1}^{N}\sum_{\sigma=-\frac{1}{2},+\frac{1}{2}}|\phi_i(\mathbf{r},\sigma)|^2 = \sum_{i=1}^{N}\left|\phi_i\left(\mathbf{r},\sigma=\frac{1}{2}\right)\right|^2 + \sum_{i=1}^{N}\left|\phi_i\left(\mathbf{r},\sigma=-\frac{1}{2}\right)\right|^2 = \\ &= \sum_{i_\alpha}\left|\phi_i\left(\mathbf{r},\sigma=\frac{1}{2}\right)\right|^2 + \sum_{i_\beta}\left|\phi_i\left(\mathbf{r},\sigma=-\frac{1}{2}\right)\right|^2 = \\ &= \sum_{i_\alpha}^{N_\alpha}|\varphi_i(\mathbf{r})|^2 + \sum_{i_\beta}^{N_\beta}|\varphi_i(\mathbf{r})|^2 \equiv \rho_\alpha(\mathbf{r}) + \rho_\beta(\mathbf{r}). \end{aligned} \tag{3.7}$$

In the last row the summation (i_α, i_β) is over *electrons (occupying an orbital diagram)* having the spin coordinate $\sigma = \frac{1}{2}$ and the spin coordinate $\sigma = -\frac{1}{2}$.

If we assume the *double occupancy* of the molecular orbitals, we have

$$\rho(\mathbf{r}) = \sum_{i=1}^{N/2}|\varphi_i(\mathbf{r})|^2 + \sum_{i=1}^{N/2}|\varphi_i(\mathbf{r})|^2 = \sum_{i=1}^{N/2}2\,|\varphi_i(\mathbf{r})|^2, \tag{3.8}$$

where φ_i stand for the molecular orbitals. We see that admitting the open shells we have

in the one-determinantal approximation

$$\rho(\mathbf{r}) = \sum_i n_i\,|\varphi_i(\mathbf{r})|^2, \tag{3.9}$$

with $n_i = 0, 1, 2$ denoting orbital occupancy in the Slater determinant.

3.2 Electron density distributions – Bader analysis

3.2.1 Overall shape of ρ

Imagine an electron cloud with a charge distribution[6] that carries the charge of N electrons. Unlike a storm cloud, the electron cloud does not change in time (for stationary state), but has density $\rho(\mathbf{r})$ that changes in space (similar to the storm cloud). Inside the cloud the nuclei are located. The function $\rho(\mathbf{r})$ exhibits nonanalytical behavior (discontinuity of its gradient) at the

[6] Similar to a storm cloud in the sky.

positions of the nuclei, which results from the poles ($-\infty$) of the potential energy at these positions. Recall the shape of the $1s$ wave function for the hydrogen-like atom (see Fig. V1-4.21); it has a spike at $r = 0$. In Chapter 2 it was shown that the correct electronic wave function has to satisfy the cusp condition in the neighborhood of each of the nuclei, where ρ changes as $\exp(-2Zr)$ (p. 91). This condition results in spikes of $\rho(\mathbf{r})$ exactly at the positions of the nuclei (Fig. 3.1a). How sharp such a spike is depends on the charge of nucleus Z: an infinitesimal deviation from the position of the nucleus (p. 92)[7] has to be accompanied by such a decrease in density[8] that $\frac{\partial\rho}{\partial r}/\rho = -2Z$.

Thus, because of the Coulombic interactions, the electrons will concentrate close to the nuclei, and therefore we will have maxima of ρ right on them. At long distances from the nuclei the density ρ will decay to practically zero with the asymptotics $\exp[-2\sqrt{2I}r]$, where I is the first ionization potential. Further details will be of great interest, e.g., are there any concentrations of ρ besides the nuclei, e.g., in the regions *between* nuclei? If yes, will it happen for every pair of nuclei or for some pairs only? This is of obvious importance for chemistry, which deals with the idea of chemical bonds between some atoms and a model of the molecule as the nuclei kept together by a chemical bond pattern.

3.2.2 Critical points

For analysis of any smooth function, including the electronic density as a function of the position in space, the *critical (or stationary) points* are defined as those for which we have vanishing of the gradient

$$\nabla\rho = \mathbf{0}.$$

These are maxima, minima, and saddle points. If we start from an arbitrary point and follow the direction of $\nabla\rho$, we end up at a maximum of ρ. Its position may correspond to any of the nuclei or to a nonnuclear concentration distribution (Fig. 3.2). Formally, positions of the nuclei are not

[7] If nonzero-size nuclei were considered, the cusps would be rounded (within the size of the nuclei), the discontinuity of the gradient would be removed, and regular maxima would be observed.

[8] It has been shown (P.D. Walker, P.G. Mezey, *J. Am. Chem. Soc*., 116(1994)12022) that despite the nonanalytical character of ρ (because of the spikes) the function ρ has the following remarkable property: *if we know* ρ even in the smallest volume, this determines ρ in the whole space. A by-product of this theorem is of interest for chemists. Namely, this means that a functional group in two different molecules or in two conformations of the same molecule cannot have an identical ρ characteristic for it. If it had, from ρ in its neighborhood we would be able to reproduce the whole density distribution $\rho(\mathbf{r})$, but for which of the molecules or conformers? Therefore, by *reductio ad absurdum* we have the result: it is impossible to define (with all details) the notion of a functional group in chemistry. This is analogous to the conclusion drawn in Chapter V1-8 about the impossibility of a rigorous definition of a chemical bond (p. V1-561). This also shows that chemistry and physics (relying on mathematical approaches) profit very much, and further, are heavily based on some ideas that mathematics destroys in a second. Nevertheless, without these ideas natural sciences would lose their generality, efficiency, and beauty.

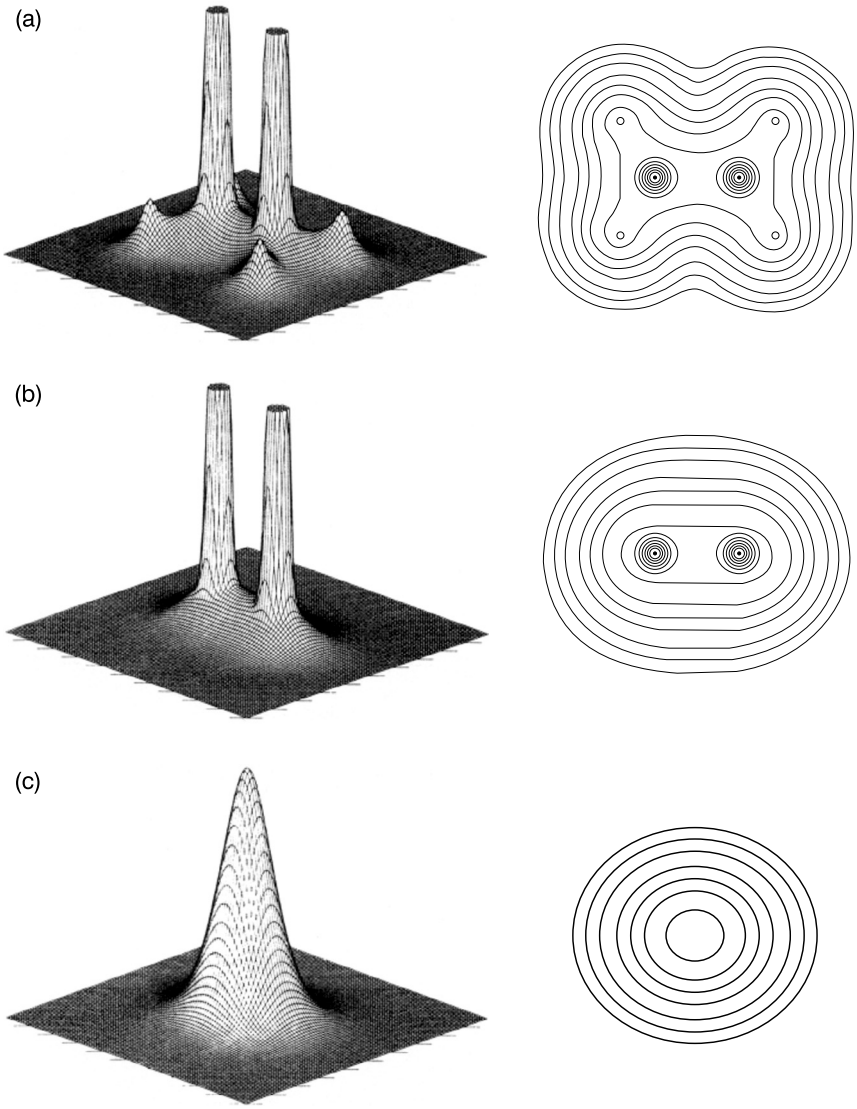


Fig. 3.1. Electron density ρ for the planar ethylene molecule shown in three cross-sections. We have $\int \rho(\mathbf{r})d\mathbf{r} = 16$, the number of electrons in the molecule. (a) The cross-section *within the molecular plane*. The positions of the nuclei can be easily recognized by the "spikes" of ρ (obviously much more pronounced for the carbon atoms than for the hydrogens atoms), their charges can be computed from the slope of ρ. (b) The cross-section *along the C–C bond perpendicular to the molecular plane*; therefore, only the maxima at the positions of the carbon nuclei are visible. (c) The cross-section *perpendicular to the molecular plane and intersecting the C–C bond* (through its center). It is seen that ρ decays monotonically with the distance from the bond center. Most interesting, however, is that the cross-section resembles an ellipse rather than a circle. Note that we do not see any separate σ or π densities. This is what the concept of π bond is all about, just to reflect the bond cross-section ellipticity. R.F.W. Bader, T.T. Nguyen-Dang, Y. Tal, *Rep. Progr. Phys.*, 44(1981)893, courtesy of Professor Richard Bader.

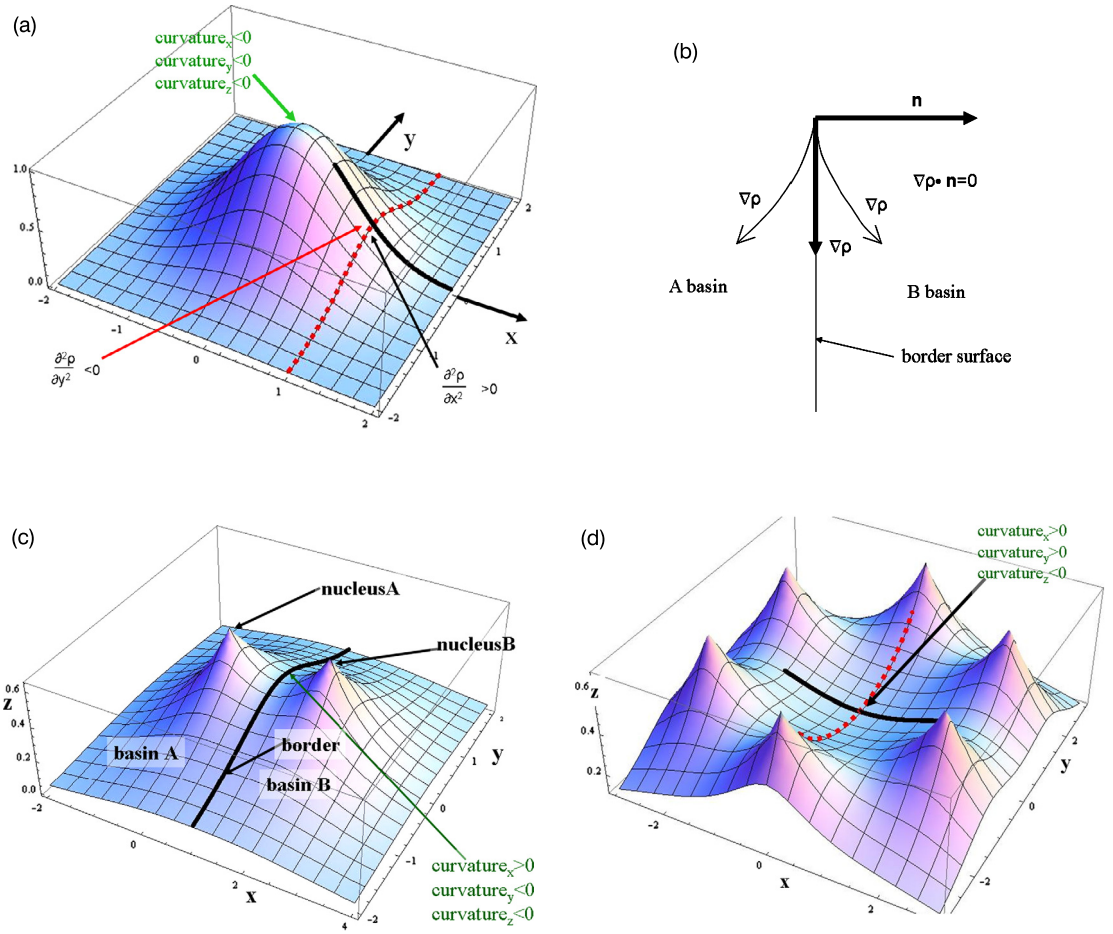


Fig. 3.2. How does the electronic density change when we leave a critical point? (a) The nonnuclear attractor (maximum of ρ). Note that we can tell the signs of some second derivatives (curvatures) computed at the intersection of the thick lines (slope), the radial curvature $\frac{\partial^2\rho}{\partial x^2}$ is positive, while the two lateral ones (only one of them: $\frac{\partial^2\rho}{\partial y^2}$ is shown) are negative. If for the function shown the curvatures were computed at the maximum, all three curvatures would be negative. (b) The idea of the border surface separating two basins of ρ corresponding to two nuclei: A and B. Right at the border between the two basins the force lines of $\nabla\rho$ diverge: if you make a step left from the border, you end up in nucleus A, and if you make a step right, you get into the basin of B. Just at the border you have to have $\nabla\rho \cdot \mathbf{n} = 0$, because the two vectors $\nabla\rho$ and $\mathbf{n}$ are perpendicular. (c) The same showing additionally the density function for the chemical bond A–B. The border surface is shown as a black line. Two of three curvatures are negative (one of them shown), the third one is positive. (d) The electronic density distribution in benzene. In the middle of the ring two curvatures are positive (shown), the third curvature is negative (not shown). If the curvatures were computed in the center of the fullerene (not shown), all three curvatures would be positive (because the electron density increases when going out of the center).

the stationary points, because $\nabla\rho$ has a discontinuity here connected to the cusp condition (see Chapter 2, p. 91), but the largest maxima correspond to the positions of the nuclei. Maxima may appear not only at the positions of the nuclei, but also elsewhere[9] (*nonnuclear attractors*, Fig. 3.2a). The compact set of starting points which converge in this way (i.e., following $\nabla\rho$) to the same maximum is called the *basin of attraction of this maximum*, and the position of the maximum is known as *attractor*. We have therefore the nuclear and nonnuclear attractors and basins. A basin has its neighbor basins and the border between the basins represents a surface satisfying $\nabla\rho \cdot \mathbf{n} = 0$, where $\mathbf{n}$ is a unit vector perpendicular to the surface (Fig. 3.2b).

In order to tell whether a particular critical point represents a maximum (nonnuclear attractor), a minimum, or a saddle point, we have to calculate at this point the Hessian, i.e., the matrix of the second derivatives, $\{\frac{\partial^2\rho}{\partial\xi_i\partial\xi_j}\}$, where $\xi_1 = x$, $\xi_2 = y$, $\xi_3 = z$. Now, the stationary point is used as the origin of a local Cartesian coordinate system, which will be rotated in such a way as to obtain the Hessian matrix (computed in the rotated coordinate system) diagonal. This means that the rotation has been performed in such a way that the axes of the new local coordinate system are colinear with the principal axes of a quadratic function's section that approximates ρ in the neighborhood of the stationary point (this rotation is achieved simply by diagonalization of the Hessian $\{\frac{\partial^2\rho}{\partial\xi_i\partial\xi_j}\}$, cf. Appendix V1-L). The diagonalization gives three eigenvalues. We have the following possibilities (the case when the Hessian matrix is singular will be considered later on):

- *All three eigenvalues are negative – we have a maximum of* ρ (nonnuclear attractor, Fig. 3.2a).
- *All three eigenvalues are positive – we have a minimum of* ρ. The minimum appears when we have a cavity, e.g., in the center of fullerene. When we leave this point, independently of the direction of this motion, the electron density increases.
- *Two eigenvalues are positive, one is negative – we have a first-order saddle point of* ρ. The center of the benzene ring may serve as an example (Fig. 3.2d). If we leave this point in the molecular plane in any of the two independent directions, ρ increases (thus, a minimum of ρ within the plane, the two eigenvalues positive), but when leaving perpendicularly to the plane the electronic density decreases (thus a maximum of ρ along the axis, the negative eigenvalue).
- *One eigenvalue is positive, while two are negative – we have a second-order saddle point of* ρ. This is a very important case, because this is what happens at any covalent chemical

[9] For example, imagine a few dipoles with their positive poles oriented towards a point in space. If the dipole moments exceed some value, it may turn out that around this point there will be a concentration of electron density having a maximum there. This is what happens in certain dipoles, in which an electron is far away from the nuclear framework (sometimes as far as 50 Å) and keeps following the positive pole of the dipole (a "dipole-bound electron") when the dipole rotates in space (see, e.g., J. Smets, D.M.A. Smith, Y. Elkadi, L. Adamowicz, *Pol. J. Chem*., 72(1998)1615).

bond (Figs. 3.1 and 3.2c). In the region between *some*[10] nuclei of a polyatomic molecule we may have such a critical point. When we go perpendicularly to the bond in any of the two possible directions, ρ decreases (two eigenvalues negative), while going towards any of the two nuclei ρ increases (to achieve maxima at the nuclei; a single minimum along the way, i.e., one eigenvalue positive). The critical point needs not be located along the straight line going through the nuclei ("banana" bonds are possible), also its location may be closer to one of the nuclei (polarization). Thus the nuclei are connected by a kind of electronic density "rope" (most dense at its core and decaying outside) extending from one nucleus to the other along (in general; the rope's density changing) a curved line, having a single critical point on it, a density minimum where the rope is the weakest, its cross-section perpendicular to the bond for some bonds is circular, for others elliptic-like.[11]

- Some of the eigenvalues may be equal to zero. The set of parameters (like the internuclear distance) at which $\det\{\frac{\partial^2\rho}{\partial\xi_i\partial\xi_j}\} = 0$ (corresponding to an eigenvalue equal to 0) is called the catastrophe set. Calculations have shown that when the two nuclei separate, the rope elongates and *suddenly*, at a certain internuclear distance, it breaks down (this corresponds to zeroing one of the eigenvalues). Thus the catastrophe theory of René Thom turns out to be instrumental in chemistry.

Richard Bader (1931–2012), Canadian chemist, professor at McMaster University in Canada. After his PhD at the Massachusetts Institute of Technology he won an international fellowship to study at Cambridge University in UK under Christopher Longuet-Higgins. At their first meeting Bader was given the titles of two books together with the remark: "*When you have read these books, maybe we can talk again*." From these books Bader found out about theories of electron density. From that time on he became convinced that electron density was the quantity of prime importance for the theory. Photo reproduced thanks to courtesy of Richard Bader.

[10] Only *some* pairs of atoms correspond to chemical bonds.

[11] All the details may be computed nowadays by using quantum mechanical methods, often most demanding ones (with the electronic correlation included). Contemporary crystallography is able to measure the same quantities in some fine X-ray experiments. Therefore, the physicochemical methods are able to indicate precisely which atoms are involved in a chemical bond, if it is strong or not, if it is straight or curved ("rope-like"), the thickness of the "rope," whether it has a cylindrical or oval cross-section (connected to its σ or π character), etc. A good review is available: T.S. Koritsanszky, P. Coppens, *Chem. Rev.*, 101(2001)1583.

3.2.3 Laplacian of the electronic density as a "magnifying glass"

Fig. 3.3a shows the functions $f(x)$, $f' = \frac{df}{dx}$, and $f'' = \frac{d^2 f}{dx^2}$, where $f(x)$ is a function with a well-visible maximum at $x = 0$ and a hump close to $x = 0.9$. The hump is hardly visible, it is so small that there is no local maximum of $f(x)$ over there. Such a function resembles somewhat the electron density decay for an atom when we go off the nucleus (position of the maximum[12]).

We may say that $-\frac{d^2 f}{dx^2}$ can detect some subtle features of the $f(x)$ plot and gives maxima where the original function $f(x)$ has only almost invisible "humps."

There is a similar story with the function $-\Delta\rho(x, y, z) = -\left(\frac{\partial^2 \rho}{\partial x^2} + \frac{\partial^2 \rho}{\partial y^2} + \frac{\partial^2 \rho}{\partial z^2}\right)$, except that here we have three Cartesian coordinates. The way we choose the directions of the Cartesian axes is irrelevant, because at any point of space $-\Delta\rho(x, y, z)$ does not depend on such a choice. Indeed, the coordinate systems, which we may choose, differ by an orthogonal transformation, which is peculiar for it does leave the *trace* of the Hessian invariant.

Imagine now ρ of an atom decaying with the distance to the nucleus as $f(x)$, similar to the decay of a smoke cloud (Fig. 3.4a), dense in the center and vanishing outward.

Fig. 3.4 displays ρ and $-\Delta\rho$ for the argon atom. Despite an apparent lack of any internal structure of the function ρ (left figure), the function $-\Delta\rho$ detected three concentrations of charge similar to the hump of the function $f(x)$ at x_2. We may say that $-\Delta\rho(x, y, z)$ plays the role of a "magnifying glass": these are the K, L, M shells of the argon atom, seen very clearly.

Fig. 3.5 shows $-\Delta\rho$ for the systems N_2, Ar_2, and F_2. The figure highlights the shell character of the electronic structure of each of the atoms.[13] Fig. 3.2c shows that the electronic density is the greatest along the bond and drops outside in *each of the two orthogonal* directions. If, however, we went along the bond the *density would have a minimum*. This means that there is a saddle point of the second order, because one eigenvalue of the Hessian is positive and two are negative.

[12] Well, no cusp, so we have a nonzero size of the nucleus and/or Gaussian-type orbitals used.

[13] Note that the nitrogen as well as the fluorine have two shells (K and L), while the argon atom has three shells (K, L, M) (cf. Chapter V1-8).

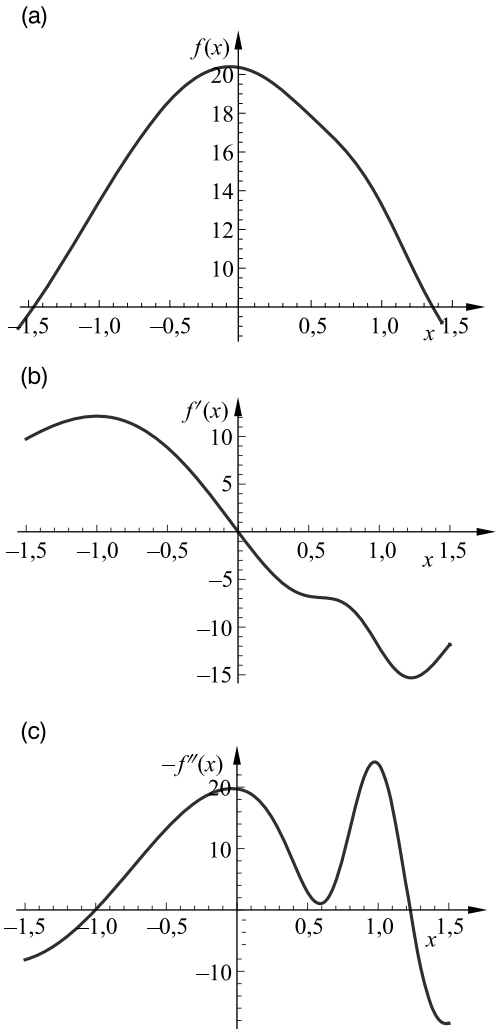


Fig. 3.3. The Laplacian $-\Delta\rho$ represents a kind of "magnifying glass." Here we illustrate this in a one-dimensional case: instead of $-\Delta\rho\,(x, y, z)$ we have $-f''(x) \equiv -\frac{\mathrm{d}^2 f}{\mathrm{dx}^2}$. (a) A function $f(x)$ with a single maximum. One can see a small asymmetry of the function resulting from a hardly visible hump on the right-hand side. (b) The first derivative $f'(x)$. (c) The plot of $-f''(x)$ shows two maxima. One of them (at $x = 0$) indicates the maximum of f, the second one (close to $x = 1$) makes the small hump of $f(x)$ clearly visible.

If there were no covalent bond at all (nonbonded atoms or ionic bond: no electron density "rope" connecting the nuclei), the curvatures along y and z would be zero, and this means that $-\Delta\rho < 0$. Thus, if it happens that for a bond $-\Delta\rho > 0$, this means a large perpendicular contribution, i.e., a strong "rope-like" covalent bond.

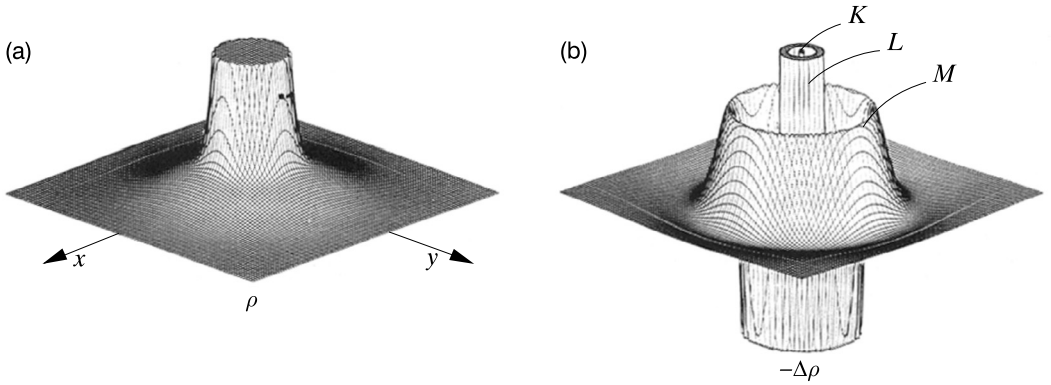


Fig. 3.4. A cross-section of (a) ρ and a cross section of (b) $-\Delta\rho$ for the argon atom. The three humps (b) correspond to the K, L, M electron shells (cf. p. V1-522). R.F.W. Bader, "*Atoms in Molecules. A Quantum Theory*," Clarendon Press, Oxford, 1994, courtesy of Professor Richard Bader.

For the N_2 molecule we have a large value of $-\Delta\rho > 0$ between the nuclei, which means an electronic charge concentrated in a strong bond, Fig. 3.5a. The nuclei have, therefore, a dilemma, whether to run off, because they repel each other, or to run only a little, because there in the middle of the bond is such a beautiful negative charge (the nuclei choose the second possibility). This dilemma is absent in the Ar_2 system (Fig. 3.5b): the electronic charge runs off the middle of the bond, the nuclei get uncovered and they run off. The molecule F_2 sticks together but not very strongly; just look at the internuclear region (Fig. 3.5c), $-\Delta\rho$ is quite low over there.[14]

3.3 Two important Hohenberg–Kohn theorems

3.3.1 Correlation dragon resides in electron density: equivalence of Ψ_0 and ρ_0

Hohenberg and Kohn proved in 1964 an interesting theorem.[15]

> The ground-state electronic density $\rho_0(\mathbf{r})$ and the ground-state wave function Ψ_0 can be used alternatively as full descriptions of the ground state of the system.

[14] We see now why the F_2 molecule does not represent an easy task for the Hartree–Fock method (Chapter V1-8, the method indicated that the molecule does not exist).

[15] P. Hohenberg, W. Kohn, *Phys. Rev.*, 136(1964)*B*864.

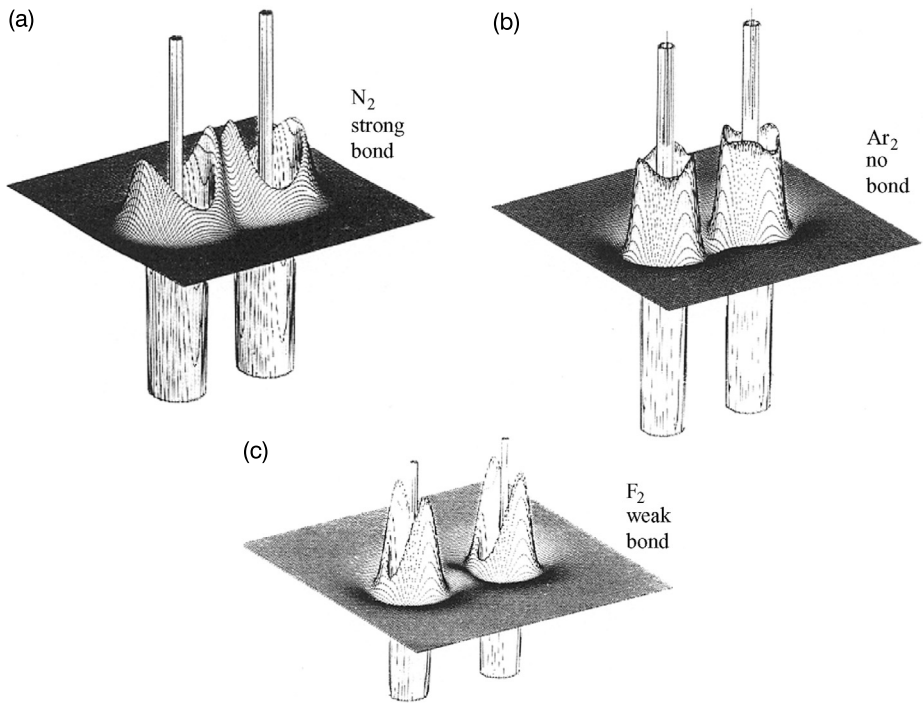
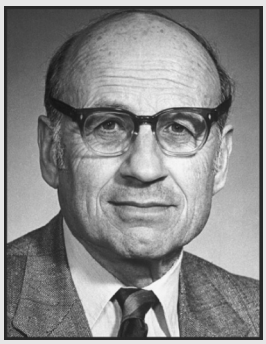


Fig. 3.5. A cross-section of the quantity $-\Delta\rho$ for N_2, Ar_2, and F_2. We will focus now on the $-\Delta\rho$ value, computed in the middle of the internuclear distance. We can see that (a) for N_2 the value of $-\Delta\rho > 0$ (chemical bond), (b) for Ar–Ar, $-\Delta\rho < 0$ (no chemical bond), and (c) for F_2 a very small positive $-\Delta\rho$ (weak chemical bond). R.F.W. Bader, "*Atoms in Molecules. A Quantum Theory*," Clarendon Press, Oxford, 1994, courtesy of Professor Richard Bader.



Walter Kohn (born 1923), American physicist of Austrian origin, professor at the University of California, Santa Barbara. His conviction about the primary role the electronic density plays led him to fundamental theoretical discoveries. Kohn shared the Nobel Prize with John A. Pople in 1998, receiving it "*for his development of the density-functional theory*."

This theorem is sometimes proved in a quite special way. Imagine somebody gave us $\rho_0(\mathbf{r})$ without a single word of explanation. We have no idea which system it corresponds to. First, we calculate $\int \rho_0(\mathbf{r})\mathrm{d}\mathbf{r}$, where the integration goes over the whole 3D space. This gives a natural number N, which is the number of electrons in the system. We did not know it, now we do. Next, we investigate the function $\rho_0(\mathbf{r})$ looking point by point at its val-

ues. We are searching for the "spikes" (cusps), because every cusp tells us where a nucleus is.[16] After this is done, we know all the positions of the nuclei. Now, we concentrate on each of the nuclei and look how fast the density drops when leaving the nucleus. The calculated slope has to be equal to a negative even number[17]: $-2Z$ (see p. 197), and Z gives us the charge of the nucleus. Thus, we have deduced the composition of our system. Now we are in a position to write down the Hamiltonian for the system and *solve the Schrödinger equation*. After that we know the ground-state wave function.

> We started, therefore, from $\rho_0(\mathbf{r})$, and we got the ground-state wave function Ψ_0. According to Eqs. (3.1) and (3.2), from the wave function by integration we obtain the density distribution $\rho_0(\mathbf{r})$. Hence, $\rho_0(\mathbf{r})$ contains the same precise information about the system as Ψ_0.

Thus, if we know ρ_0, we also know[18] Ψ_0, and, if we know Ψ_0, we also know ρ_0.[19]

The proof we carried out pertains only to the case when the *external potential* (everything except the interelectronic interaction) acting on the electrons stems from the nuclei. The Hohenberg–Kohn theorem can be proved for an arbitrary external potential – this property of the density is called *v-representability*. The arbitrariness mentioned above is necessary in order to define in the future the functionals for more general densities (than for isolated molecules). We will need that generality when introducing the functional derivatives (p. 215) in which $\rho(\mathbf{r})$ has to result from any external potential (or to be a *v-representable density*). Also, we will be interested in a *non-Coulombic* potential corresponding to the harmonic helium atom (cf. harmonium, p. 94) to see how exact the DFT method is. We may imagine ρ, which is

---

16 $\rho(\mathbf{r})$ represents a cloud similar to those that float in the sky. This "spike," therefore, means simply a high density in the cloud.

17 The $|1s|^2$ electron density $\sim \exp(-2Zr) = 1 - 2Zr + \ldots$.

18 And all the excited states wave functions as well! This is an intriguing conclusion, supported by experts (see W. Koch, M.C. Holthausen, "A Chemist's Guide to Density Functional Theory," Second ed., Wiley, Weinheim, 2001). On p. 59 we read: "*the DFT is usually termed a ground state theory. The reason for this is not that the ground state density does not contain the information on the excited states – it actually does! – but because no practical way to extract this information is known so far.*"

19 The theorem just proved shines in its simplicity. People thought that the wave function, usually a very complicated mathematical object (that depends on $3N$ space and N spin coordinates) is indispensable for computing the properties of the system. Moreover, the larger the system, the worse the difficulties in calculating it (please recall Chapter 2 with billions of excitations, nonlinear parameters, etc.). Besides, how to interpret such a complex object? Horror. And it turns out that everything about the system just sits in $\rho(\mathbf{r})$, a function of position in our well-known three-dimensional space. It turns out that information about nuclei is hidden in such a simple object, it seems trivial (cusps), but also much more subtle information about how electrons avoid each other due to Coulombic repulsion and the Pauli exclusion principle can be found.

not v-representable, e.g., discontinuous (in one, two, or even in every point like the Dirichlet function). The density distributions that are not v-representable are outside our field of interest.

3.3.2 A secret of the correlation dragon: the existence of energy functional minimized by ρ_0

Hohenberg and Kohn also proved an analog of the variational principle (p. V1-265).

HOHENBERG–KOHN THEOREM

For a given number of electrons (the integral over ρ equals N) and external potential v, there *exists* a functional of ρ, denoted by $E_v^{\mathrm{HK}}[\rho]$, for which the following variational principle is satisfied:

$$E_v^{\mathrm{HK}}[\rho] \geq E_v^{\mathrm{HK}}[\rho_0] = E_0,$$

where ρ_0 stands for the (ideal) ground-state electronic density distribution corresponding to the ground-state energy E_0.

We will prove this theorem using the variational principle in a way shown first by Levy.[20] The variational principle states that

$$E_0 = \min\langle\Psi|\hat{H}|\Psi\rangle,$$

where we search among the wave functions Ψ normalized to 1 and describing N electrons.

This minimization may be carried out in two steps:

$$E_0 = \min_{\rho, \int \rho dV = N} \min_{\Psi \to \rho} \langle\Psi|\hat{T} + U + V|\Psi\rangle, \tag{3.10}$$

where $\hat{T}, U, V$ represent the kinetic energy, the electron repulsion, and the electron–nuclei attraction operators, respectively, for all the N electrons of our system (the hat in operators will be omitted if the operator has a multiplicative character).

The two minimization steps have the following meaning:

- The internal minimization is performed at the condition labeled as "$\Psi \to \rho$," which means minimization of the integral among the N-electron functions that are normalized to 1, and

[20] M. Levy, *Phys. Rev. A*, 26(1982)1200.

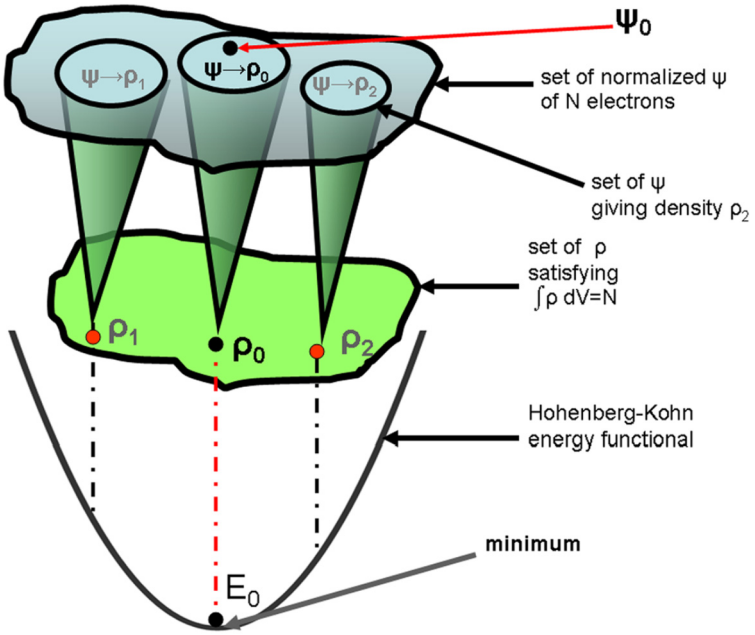


Fig. 3.6. Levy variational principle (scheme). The task of the internal minimization is the following. At a *given fixed* density distribution ρ carrying N electrons, choose among those normalized functions Ψ, *that all produce* ρ (we will denote this by the symbol "$\Psi \to \rho$"), *such a function that minimizes* $\langle\Psi|\hat{T} + U + V|\Psi\rangle$ of Eq. (3.10). In the upper part of the figure three sets of such functions Ψ are shown: one set gives ρ_1, the second ρ_0, the third ρ_2. The external minimization symbolized by "$\rho, \int \rho dV = N$ " chooses among all possible electron distributions ρ (that correspond to N electrons, the center part of the figure) such a distribution $\rho = \rho_0$ that gives the lowest value (the ground-state energy E_0, see the bottom part of the figure) of the Hohenberg–Kohn functional E_v^{HK}, i.e., $E_0 = \min_{\rho, \int \rho dV = N} E_v^{\mathrm{HK}}$. Note that among the functions Ψ that give ρ_0 there is the exact ground-state wave function Ψ_0.

any of them giving a *fixed* density distribution ρ "carrying" N electrons (the minimum attained at $\Psi = \Psi_{\min}$). As a result of this minimization we obtain a functional of ρ given as $\min_{\Psi\to\rho}\langle\Psi|\hat{T} + U + V|\Psi\rangle = \langle\Psi_{\min}(\rho)|\hat{T} + U + V|\Psi_{\min}(\rho)\rangle$, because $\langle\Psi_{\min}|\hat{T} + U + V|\Psi_{\min}\rangle$ depends on what we have taken as ρ.

- In the external minimization symbolized by "$\rho, \int \rho dV = N$" we go over all the density distributions ρ that integrate to N (i.e., describe N electrons), and we choose that $\rho = \rho_0$ which minimizes the functional $\langle\Psi_{\min}(\rho)|\hat{T} + U + V|\Psi_{\min}(\rho)\rangle$. According to the variational principle (p. V1-266), this minimum is bound to be the exact ground-state energy E_0, while ρ_0 is the exact ground-state density distribution.

Therefore, both minimizations do the same as the variational principle.

The external potential

It is easy to show that $\langle\Psi|V\Psi\rangle$ may be expressed as an integral involving the density distribution ρ instead of Ψ. Indeed, since

$$V = \sum_{i=1}^{N} v(\mathbf{r}_i), \quad \text{where} \quad v(\mathbf{r}_i) = \sum_A -\frac{Z_A}{|\mathbf{r}_i - \mathbf{r}_A|}, \tag{3.11}$$

in each of the resulting integrals $\langle\Psi|v(\mathbf{r}_i)\Psi\rangle$ we may carry out the integration over all the electrons except one, and for this single one we sum over its spin coordinate. It is easy to see that every such term (their number is N) gives[21] *the same result* $\langle\Psi|v(\mathbf{r}_i)\Psi\rangle = \frac{1}{N}\int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}$, because the electrons are indistinguishable (this is why we omit the index i). Because of this we will get

$$\langle\Psi|V\Psi\rangle = \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{3.12}$$

Therefore, the Levy minimization may be written as

$$E_0 = \min_{\rho,\int\rho dV=N}\{\int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} + \min_{\Psi\to\rho}\langle\Psi|(\hat{T}+U)\Psi\rangle\}. \tag{3.13}$$

The universal potential

At this point we define the auxiliary functional[22] F^{HK} as follows:

$$F^{\mathrm{HK}}[\rho] = \min_{\Psi\to\rho}\langle\Psi|(\hat{T}+U)\Psi\rangle \equiv \langle\Psi_{\min}(\rho)|(\hat{T}+U)\Psi_{\min}(\rho)\rangle, \tag{3.14}$$

where $\Psi_{\min}$ stands for a normalized function which has been chosen among those that produce a given ρ and makes the smallest value of $\langle\Psi|\hat{T}+U|\Psi\rangle$. This functional is often called universal, because it does not depend on any external potential – it pertains to interacting electrons only.

---

21 $\langle\Psi|V\Psi\rangle = \langle\Psi|\sum_i v(\mathbf{r}_i)\Psi\rangle = \sum_i\langle\Psi|v(\mathbf{r}_i)\Psi\rangle = N\langle\Psi|v(\mathbf{r}_1)\Psi\rangle = N\frac{1}{N}\int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} = \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}$.

22 A functional is always defined in a domain, in our case a domain of the allowed ρ's. How do allowed ρ's look? Here are the conditions to fulfill: (i) $\rho \geq 0$, (ii) $\int \rho dV = N$, (iii) $\nabla\rho^{1/2}$ square integrable. Among these conditions we do not find any that would require the existence of such an antisymmetric Ψ of N electrons that would correspond (in the sense of Eq. (3.2)) to the density ρ under consideration (this is known as N-representability). It turns out that such a requirement is not needed, since it was proved by Thomas Gilbert (the proof may be found in the book by R.G. Parr and W. Yang, "*Density-functional theory of atoms and molecules*," Oxford University Press, New York, 1989) that every ρ that satisfies the above conditions is N-representable, because it corresponds to at least one antisymmetric N-electron Ψ.

The Hohenberg–Kohn potential

In the DFT we define the crucial

Hohenberg–Kohn functional $E_v^{\mathrm{HK}}[\rho]$ as

$$E_v^{\mathrm{HK}}[\rho] = \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} + F^{\mathrm{HK}}[\rho], \tag{3.15}$$

and the minimum of this functional is the ground-state energy

$$E_0 = \min_{\rho, \int \rho dV = N} E_v^{\mathrm{HK}}[\rho], \tag{3.16}$$

while ρ that minimizes $E_v^{\mathrm{HK}}[\rho]$ represents the exact ground-state density distribution ρ_0 (Fig. 3.6). Each ρ corresponds to at least one antisymmetric electronic wave function (the "N-representability" mentioned above), and there is no better wave function than the ground-state one, which, of course, corresponds to the density distribution ρ_0. This is why we have the following.

HOHENBERG–KOHN FUNCTIONAL

The Hohenberg–Kohn functional $E_{\mathrm{v}}^{\mathrm{HK}}[\rho]$ attains minimum $E_0 = E_v^{\mathrm{HK}}[\rho_0]$ for the ideal ground-state density distribution. Now our job will be to find out what mathematical form the functional $E[\rho]$ could have. And here we meet the basic problem of the DFT method: nobody has so far been able to give such a formula. The best which has been achieved are some approximations. These approximations, however, are so good that they begin to supply results that satisfy chemists.

Therefore, when the question is posed, "*Is it possible to construct a quantum theory, in which the basic concept would be electronic density?*", we have to answer, "*Yes, it is*." This answer, however, has only an existential value ("*yes, it exists*"). We have no information about how such a theory could be constructed.

An indication may come from the concept of the wave function. In order to proceed towards the abovementioned unknown functional, we will focus on the ingenious idea of a *fictitious Kohn–Sham system of noninteracting electrons*.

3.4 The Kohn–Sham equations

3.4.1 A Kohn–Sham system of noninteracting electrons

Let us consider an electron subject to some "external" potential $v(\mathbf{r})$, *for example* coming from the Coulombic interaction with the nuclei (with charges Z_A in a.u. and positions $\mathbf{r}_A$)

$$v(\mathbf{r}) = \sum_A -\frac{Z_A}{|\mathbf{r} - \mathbf{r}_A|}. \tag{3.17}$$

In this system we have N electrons, which also interact by Coulombic forces between themselves. All these interactions produce the ground-state electronic density distribution ρ_0 (ideal, i.e., that we obtain from the exact, 100% correlated wave function). Now let us consider

FICTITIOUS KOHN–SHAM SYSTEM

the fictitious Kohn–Sham system of N model electrons (fermions), that *do not interact* at all (as if their charge were equal to zero), but, instead of the interaction with the nuclei, they are subject to an external potential $v_0(\mathbf{r})$ so ingeniously tailored that ρ does not change, i.e., we still have the ideal ground-state electronic density $\rho = \rho_0$.

Let us assume for a while that we have found such a wonder potential $v_0(\mathbf{r})$. We will worry later about how to find it in reality. Now we assume this problem has been solved. Can we find ρ? Of course, we can. Since the Kohn–Sham electrons do not interact between themselves, we only have to solve the one-electron equation (with the wonder v_0)

$$\left(-\frac{1}{2}\Delta + v_0\right)\phi_i = \varepsilon_i \phi_i, \tag{3.18}$$

where ϕ_i are the solutions – some spin orbitals, of course, called the Kohn–Sham spin orbitals.[23]

[23] If the electrons do not interact, the corresponding wave function can be taken as a *product* of the spin orbitals for individual electrons. Such a function for electrons is not antisymmetric, and, therefore, is "illegal." Taking the *Kohn–Sham determinant* (instead of the product) helps, because it is antisymmetric and represents an eigenfunction of the total Hamiltonian of the fictitious system (i.e., the sum of the one-electron operators given in Eq. (3.18)). This is easy to show, because a determinant represents a sum of products of the spin orbitals, the products differing only by permutation of electrons. If the total Hamiltonian of the fictitious system acts on such a sum, each term (product) is its eigenfunction, and each eigenvalue amounts to $\sum_{i=1}^{N} \varepsilon_i$, i.e., is the same for each product. Hence, the Kohn–Sham determinant represents an eigenfunction of the fictitious system. Scientists compared the Kohn–Sham orbitals with the canonical Hartree–Fock orbitals with great interest. It turns out that the differences are small.

The total wave function is a Slater determinant, which in our situation should rather be called the Kohn–Sham determinant. The electronic density distribution of such a system is given by Eq. (3.9) and the density distribution ρ means *exact*, i.e., correlated 100% (thanks to the "wonder" operator v_0).

3.4.2 Chasing the correlation dragon into an unknown part of the total energy

Let us try to write down a general expression for the electronic ground-state energy of the system under consideration. Obviously, it must include the kinetic energy of the electrons, their interaction with the nuclei, and their repulsion among themselves. However, in the DFT approach we write the following:

$$E = T_0 + \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} + J[\rho] + E_{\mathrm{xc}}[\rho], \tag{3.19}$$

where:

- instead of the electronic kinetic energy of the system we write down (in cold blood) the electronic kinetic energy *of the fictitious Kohn–Sham system of (noninteracting) electrons* T_0 (please recall the Slater–Condon rules, p. V1-707),

$$T_0 = -\frac{1}{2}\sum_{i=1}^{N}\langle\phi_i|\Delta\phi_i\rangle; \tag{3.20}$$

- next, there is the correct electron–nuclei interaction term, $\int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}$;
- then, there is an interaction of the electron cloud with itself,[24]

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2. \tag{3.21}$$

[24] How to compute the Coulombic interaction within a storm cloud exhibiting certain charge distribution ρ? At first sight it looks like a difficult problem, but remember we know how to calculate the Coulombic interaction of two *point* charges. Let us divide the whole cloud into tiny cubes, each of volume $\mathrm{d}V$. The cube that is pointed by the vector $\mathbf{r}_1$ contains a tiny charge $\rho(\mathbf{r}_1)\mathrm{d}V \equiv \rho(\mathbf{r}_1)\mathrm{d}\mathbf{r}_1$. We know that when calculating the Coulombic interaction of two such cubes we have to write $\frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2$. This has to be summed over all possible positions of the first and the second cube, i.e., $\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2$, but in such a way each interaction is computed twice, whereas they represent parts of the same cloud. Hence, the final self-interaction of the storm cloud is $\frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2$. The expression for the self-interaction of the electron cloud is the same.

No doubt, such an idea looks reasonable for the energy expression should contain an interaction of the electron cloud with itself (because the electrons repel each other). However, there is a trap in this concept – an illness hidden in $J[\rho]$. The illness is seen best if one considers the simplest system: the hydrogen atom ground state. In Eq. (3.21) we have an interelectronic self-repulsion, which for sure does not exist, because we have one electron only! So, whatever reasonable remedy is to be designed in the future, it should reduce this unwanted self-interaction in the hydrogen atom to zero. Of course, the problem is not limited to the hydrogen atom. When taking $J[\rho]$, *an electron is interacting with itself and this self-interaction has to be somehow excluded from* $J[\rho]$ *by introducing a correction*. Two electrons repel each other electrostatically and therefore around each of them there has to exist a kind of no-parking zone for the other one ("Coulomb hole," cf. p. 104). A no-parking zone also results, because electrons of the same spin coordinate hate one another[25] ("exchange hole" or "Fermi hole," cf. p. 105). The integral J does not take such a correlation of motions into account.

Thus, we have written a few terms and we do not know what to write next. Well,

> in the DFT in the expression for E we write in (3.19) the lacking remainder as E_{xc}, and we call it the *exchange-correlation energy* (label x stands for "exchange," c is for "correlation") and declare, courageously, that we will manage somehow to get it.

The above formula represents a definition of the *exchange-correlation energy*, although it is rather a strange definition – it requires us to know E. We should not forget that in E_{xc} a correction to the kinetic energy has also to be included (besides the exchange and correlation effects) that takes into account that kinetic energy has to be calculated for the true (i.e., interacting) electrons, not for the noninteracting Kohn–Sham ones. The next question is connected to what kind of mathematical form E_{xc} might have. Let us assume, for the time being, we have no problem with this mathematical form. For now we will establish a relation between our wonder external potential v_0 and our mysterious E_{xc}, both quantities performing miracles, but not known.

3.4.3 Derivation of the Kohn–Sham equations

Now we will make a variation of E, i.e., we will find the linear effect of changing E due to a variation of the spin orbitals (and therefore also of the density). We make a spin orbital variation denoted by $\delta\phi_i$ (as before, p. V1-472, it is justified to vary either ϕ_i or ϕ_i^*, the result is the same;

[25] A correlated density and a noncorrelated density differ in that in the correlated one we have smaller values in the high-density regions, because the holes make the overcrowding of space by electrons less probable.

we choose, therefore, $\delta\phi_i^*$) and see what effect it will have on E keeping only the linear term. We have (see Eq. (3.7))

$$\phi_i^* \to \phi_i^* + \delta\phi_i^*, \tag{3.22}$$

$$\rho \to \rho + \delta\rho, \tag{3.23}$$

$$\delta\rho\left(\mathbf{r}\right) = \sum_{\sigma}\sum_{i=1}^{N} \delta\phi_i^*\left(\mathbf{r},\sigma\right)\phi_i\left(\mathbf{r},\sigma\right). \tag{3.24}$$

We insert the right-hand sides of the above expressions into E, and identify the variation, i.e., the linear part of the change of E. The variations of the individual terms of E look like (note that the symbol $\langle|\rangle$ stands for an integral over space coordinates and a summation over the spin coordinates, see p. V1-469)

$$\delta T_0 = -\frac{1}{2}\sum_{i=1}^{N}\langle\delta\phi_i|\Delta\phi_i\rangle, \tag{3.25}$$

$$\delta\int v\rho\mathrm{d}\mathbf{r} = \int v\delta\rho\mathrm{d}\mathbf{r} = \sum_{i=1}^{N}\langle\delta\phi_i|v\phi_i\rangle, \tag{3.26}$$

$$\begin{aligned}
\delta J &= \frac{1}{2}\left[\int\frac{\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 + \int\frac{\delta\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\right] \\
&= \int\frac{\rho(\mathbf{r}_2)\delta\rho(\mathbf{r}_1)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \\
&= \sum_{i=1}^{N}\int\sum_{\sigma_1}\delta\phi_i^*\left(\mathbf{r}_1,\sigma_1\right)\phi_i\left(\mathbf{r}_1,\sigma_1\right)\frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \\
&= \sum_{i=1}^{N}\sum_{\sigma_1}\int_1\mathrm{d}\mathbf{r}_1\delta\phi_i^*\left(\mathbf{r}_1,\sigma_1\right)\phi_i\left(\mathbf{r}_1,\sigma_1\right)\int_2\frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 \\
&= \sum_{i=1}^{N}\sum_{\sigma_1}\int_1\mathrm{d}\mathbf{r}_1\delta\phi_i^*\left(\mathbf{r}_1,\sigma_1\right)\phi_i\left(\mathbf{r}_1,\sigma_1\right)\int_2\frac{\sum_j\sum_{\sigma_2}\phi_j^*\left(\mathbf{r}_2,\sigma_2\right)\phi_j\left(\mathbf{r}_2,\sigma_2\right)}{|\mathbf{r}_1-\mathbf{r}_2|}\mathrm{d}\mathbf{r}_2 \\
&= \sum_{i,j=1}^{N}\langle\delta\phi_i(\mathbf{r}_1,\sigma_1)|\hat{J}_j(\mathbf{r}_1)\phi_i(\mathbf{r}_1,\sigma_1)\rangle_1,
\end{aligned} \tag{3.27}$$

where $\langle\ldots|\ldots\rangle_1$ means integration over spatial coordinates and the summation over the spin coordinate of electron 1 ($\int_1$ means the integration only), with the Coulomb operator $\hat{J}_j$ associated

with the spin orbital ϕ_j

$$\hat{J}_j(\mathbf{r}_1) = \sum_{\sigma_2} \int \frac{\phi_j(\mathbf{r}_2, \sigma_2)^* \phi_j(\mathbf{r}_2, \sigma_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \mathrm{d}\mathbf{r}_2. \tag{3.28}$$

Finally, we come to the variation of E_{xc}, i.e., δE_{xc}. We are in a quite difficult situation, because we do not know the mathematical dependence of the functional E_{xc} on ρ, and therefore also on $\delta\phi_i^*$. Nevertheless, we somehow have to get the linear part of E_{xc}, i.e., the variation.

A change of functional F (due to $f \to f + \delta f$) contains a part linear in δf denoted by δF plus some higher powers[26] of δf denoted by $O((\delta f)^2)$, i.e.,

$$F[f + \delta f] - F[f] = \delta F + O((\delta f)^2). \tag{3.29}$$

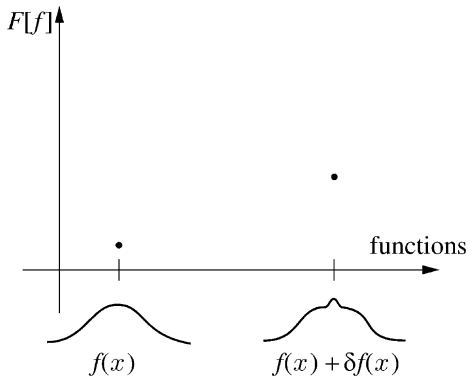


Fig. 3.7. A scheme showing what a functional derivative is about. The ordinate represents the values of a functional $F[f]$, while each point of the horizontal axis represents a function $f(x)$. The functional $F[f]$ depends, of course, on details of the function $f(x)$. If we consider a *small local change* of $f(x)$, this change may result in a large change of F – then the derivative $\frac{\delta F}{\delta f}$ is large – or in a small change of F – then the derivative $\frac{\delta F}{\delta f}$ is small (this depends on the particular functional).

The δF is defined through the *functional derivative* (Fig. 3.7) of F with respect to the function f (denoted by $\frac{\delta F[f]}{\delta f(x)}$), for a single variable[27] x, i.e.,

$$\delta F = \int_a^b \mathrm{d}x \frac{\delta F[f]}{\delta f(x)} \delta f(x). \tag{3.30}$$

[26] If δf is very small, the higher terms are negligible.

[27] Just for the sake of simplicity. The functional derivative itself is a functional of f and a function of x. An example of a functional derivative may be found in Eq. (3.27), when looking at $\delta J = \int \frac{\rho(\mathbf{r}_2)\delta\rho(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 = \int \mathrm{d}\mathbf{r}_1 \{\int \mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\} \delta\rho(\mathbf{r}_1)$. Indeed, as we can see from Eq. (3.30) $\int \mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \equiv \frac{\delta J[\rho]}{\delta\rho(\mathbf{r}_1)}$, which is a three-dimensional equivalent of $\frac{\delta F[f]}{\delta f(x)}$. Note that $\int \mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}$ is a functional of ρ and a function of $\mathbf{r}_1$.

Indeed, in our case we obtain as $\delta E_{\rm xc}$

$$\delta E_{\rm xc} = \int {\rm d}\mathbf{r} \frac{\delta E_{\rm xc}}{\delta \rho(\mathbf{r})} \delta\rho(\mathbf{r}) = \sum_{i=1}^{N} \langle \delta\phi_i | \frac{\delta E_{\rm xc}}{\delta \rho} \phi_i \rangle. \tag{3.31}$$

Therefore, an unknown quantity $E_{\rm xc}$ is replaced by another unknown quantity $\frac{\delta E_{\rm xc}}{\delta \rho}$, but there is profit from this: the functional derivative enables us to write an equation for spin orbitals. The variations of the spin orbitals are not arbitrary in this formula – they have to satisfy the orthonormality conditions (because our formulae, e.g., (3.7), are valid only for such spin orbitals) for $i, j = 1, \ldots N$, which gives

$$\langle \delta\phi_i | \phi_j \rangle = 0 \quad \text{for} \quad i, j = 1, 2, \ldots N. \tag{3.32}$$

Let us multiply each of Eqs. (3.32) by a Lagrange multiplier ε_{ij}, add them together, subtract from the variation δE, and then write the result as equal to zero.[28] We obtain

$$\delta E - \sum_{i,j}^{N} \varepsilon_{ij} \langle \delta\phi_i | \phi_j \rangle = 0 \tag{3.33}$$

or (note that $\langle \delta\phi_i | \phi_j \rangle_1$, i.e., integration over electron 1 is equal to $\langle \delta\phi_i | \phi_j \rangle$)

$$\sum_{i=1}^{N} \langle \delta\phi_i | \{ [-\frac{1}{2}\Delta + v + \sum_{j=1}^{N} \hat{J}_j + \frac{\delta E_{\rm xc}}{\delta\rho}] \phi_i - \sum_{j=1}^{N} \varepsilon_{ij} \phi_j \} \rangle_1 = 0. \tag{3.34}$$

After inserting the Lagrange multipliers, the variations of ϕ_i^* are already *independent* and the only possibility to have zero on the right-hand side is that every individual ket $|\ \rangle$ is zero (Euler equation, cf. p. V1-720), i.e.,

$$\{ -\frac{1}{2}\Delta + v + v_{\rm coul} + v_{\rm xc} \} \phi_i = \sum_{j=1}^{N} \varepsilon_{ij} \phi_j, \tag{3.35}$$

$$v_{\rm coul}(\mathbf{r}) \equiv \sum_{j=1}^{N} \hat{J}_j(\mathbf{r}), \tag{3.36}$$

$$v_{\rm xc}(\mathbf{r}) \equiv \frac{\delta E_{\rm xc}}{\delta \rho(\mathbf{r})}. \tag{3.37}$$

It would be good now to get rid of the nondiagonal Lagrange multipliers in order to obtain a beautiful one-electron equation analogous to the Fock equation. To this end we need the

[28] Appendix V1-O (p. V1-719) explains why such a procedure corresponds to minimization with constraints.

operator in the curly brackets in (3.35) to be invariant with respect to an arbitrary unitary transformation of the spin orbitals. The sum of the Coulomb operators (v_{coul}) is invariant, as has been demonstrated on p. V1-477. As to the unknown functional derivative $\delta E_{xc}/\delta\rho$, i.e., potential v_{xc}, its invariance follows from the fact that it is a functional of ρ (and ρ of Eq. (3.7) is invariant). Finally, after applying such a unitary transformation that diagonalizes the matrix of ε_{ij}, we obtain the Kohn–Sham equation ($\varepsilon_{ii} \equiv \varepsilon_i$), i.e.,

KOHN–SHAM EQUATION

$$\{-\frac{1}{2}\Delta + v + v_{\text{coul}} + v_{\text{xc}}\}\phi_i = \varepsilon_i \phi_i. \tag{3.38}$$

The equation is analogous to the Fock equation (p. V1-478).[29] We solve the Kohn–Sham equation by an iterative method. We start from any zero-iteration orbitals. This enables us to calculate a zero approximation to ρ, and then the zero approximations to the operators v_{coul} and v_{xc} (in a moment we will see how to compute E_{xc}, and then using (3.37), we obtain v_{xc}). The solution to the Kohn–Sham equation gives new orbitals and new ρ. The procedure is then repeated until consistency is achieved.

Hence, finally we "know" what the wonder operator v_0 looks like:

$$v_0 = v + v_{\text{coul}} + v_{\text{xc}}. \tag{3.39}$$

As in the Hartree–Fock method there is no problem with v_{coul}, a serious difficulty arises with the exchange-correlation operator v_{xc}, or (equivalent) with the energy E_{xc}. The second Hohenberg–Kohn theorem says that the functional $E_v^{HK}[\rho]$ exists, but it does not guarantee that it is simple. For now we worry about this potential, but we courageously go ahead.

Kohn–Sham equations with spin polarization

Before searching for v_{xc} let us generalize the Kohn–Sham formalism and use Eq. (3.3) for splitting ρ into the α and β spin functions. If these contributions are not equal (even for some $\mathbf{r}$),

[29] There is a difference in notation: the one-electron operator $\hat{h}$ and the Coulomb operator $\hat{J}$ from the Fock equation are now replaced by $-\frac{1}{2}\Delta + v \equiv \hat{h}$ and $\hat{J} \equiv v_{coul}$. There is, however, a serious difference: instead of the exchange operator $-\hat{K}$ in the Fock equation, we have here the exchange-correlation potential v_{xc}.

we will have a *spin polarization*. In order to reformulate the equations, we consider two non-interacting fictitious electron systems: one described by the spin functions α, and the other by functions β, with the corresponding density distributions $\rho_\alpha(\mathbf{r})$ and $\rho_\beta(\mathbf{r})$ exactly equal to ρ_α and ρ_β, respectively, in the (original) interacting system. Then, we obtain two coupled[30] Kohn–Sham equations, for $\sigma = \alpha$ and $\sigma = \beta$ playing here the role of labels only, with potential v_0 that depends on the spin coordinate σ, i.e.,

$$v_0^\sigma = v + v_{\text{coul}} + v_{\text{xc}}^\sigma. \tag{3.40}$$

The situation is analogous to the unrestricted Hartree–Fock (UHF) method (cf. p. V1-464).

This extension of the DFT is known as spin density functional theory (SDFT).

3.5 Trying to guess the appearance of the correlation dragon

We approach the point where we promised to write down the mysterious exchange-correlation energy. Well, how to tell you the truth? Let me put it straightforwardly: we do not know the analytical form of this quantity. Nobody knows what the exchange-correlation is; there are only guesses. The number of formulae will be, as usual with guesses, almost unlimited.[31] Let us take the simplest ones to show the essence of the procedure.

3.5.1 Local density approximation (LDA)

The electrons in a molecule are in quite a complex situation, because they not only interact between themselves, but also with the nuclei. However, a simpler system has been elaborated theoretically for years: a homogeneous gas model in a box,[32] an electrically neutral system (the nuclear charge is smeared out uniformly). It does not represent the simplest system to study, but it turns out that theory is able to determine (exactly) some of its properties. For example, it has been deduced how E_{xc} depends on ρ, and even how it depends on ρ_α and ρ_β. Since the gas is homogeneous and the volume of the box is known, we could easily work out how the E_{xc} per unit volume depends on these quantities.

Then, the reasoning is the following.[33]

---

30 Through the common operator v_{coul}, a functional of $\rho_\alpha + \rho_\beta$ and through v_{xc}, because the latter is in general a functional of both ρ_α and ρ_β.

31 Some detailed formulae are reported in the book by J.B. Foresman and A. Frisch "*Exploring Chemistry with Electronic Structure Methods*," Gaussian, Pittsburgh, PA, USA, p. 272.

32 With periodic boundary conditions. This is a common trick to avoid the surface problem. We consider a box having such a property, that if something goes out through one wall it enters through the opposite wall (cf. p. 23).

33 W. Kohn, L.J. Sham, *Phys. Rev.*, 140(1965)*A*1133.

The electronic density distribution in a molecule is certainly inhomogeneous, but locally (within a small volume) we may assume its homogeneity. Then, if someone asks about the exchange-correlation energy contribution from this small volume, we would say that in principle we do not know, but to a good approximation the contribution could be calculated as a product of the small volume and the exchange-correlation energy density from the homogeneous gas theory (with the electronic gas density as calculated inside the small volume). Then, we sum up such contributions over the whole molecule and our task is over.

Thus, everything is decided locally: we have a sum of contributions from each infinitesimally small element of the electron cloud with the corresponding density. This is why it is called the local density approximation (LDA, when the ρ dependence is used) or the local spin density approximation (LSDA, when the ρ_α and ρ_β dependencies are exploited).

3.5.2 Nonlocal density approximation (NLDA)

Gradient expansion approximation

There are approximations that go beyond the LDA. They consider that the dependence $E_{\mathrm{xc}}[\rho]$ may be *nonlocal*, i.e., E_{xc} may depend on ρ at a given point (locality), but also on ρ nearby (nonlocality). When we are at a point, what happens further off depends not only on ρ at that point, but also on the gradient of ρ at the point, etc..[34] This is how the idea of the gradient expansion approximation (GEA) appeared, which is written

$$E_{\mathrm{xc}}^{\mathrm{GEA}} = E_{\mathrm{xc}}^{\mathrm{LSDA}} + \int B_{\mathrm{xc}}(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)\mathrm{d}\mathbf{r}, \tag{3.41}$$

where the exchange-correlation function B_{xc} is carefully selected as a function of ρ_α, ρ_β, and their gradients, in order to maximize the successes of the theory/experiment comparison. However, this recipe was not so simple and some strange unexplained discrepancies were still taking place.

Perdew–Wang functional (PW91)

A breakthrough in the quality of results is represented by the following proposition of Perdew and Wang:

$$E_{\mathrm{xc}}^{\mathrm{PW91}} = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)\mathrm{d}\mathbf{r}, \tag{3.42}$$

[34] As in a Taylor series, we may need not only the gradient, but also the Laplacian, etc.

where the function f of ρ_α,ρ_β, and their gradients has been tailored in an ingenious way. It sounds unclear, but it will be shown later on that their approximation used some fundamental properties and this enabled them *without introducing any parameters* to achieve a much better agreement between the theory and experiment.

The famous B3LYP hybrid functional

The B3LYP approach belongs to the *hybrid (i.e., mixed) approximations* for the exchange-correlation functional. The approximation is famous, because it gives very good results and, therefore, is extremely popular. So far so good, but there is a danger of Babylon-type science.[35] It smells like a witch's brew for the B3LYP exchange-correlation potential E_{xc}: *take the exchange-correlation energy from the LSDA method (a unit), and add a pinch (*0.20 *unit) of the difference between the Hartree–Fock exchange energy*[36] $E_{\mathrm{x}}^{\mathrm{KS}}$ *and the LSDA* $E_{\mathrm{x}}^{\mathrm{LSDA}}$. *Then, mix well* 0.72 *unit of Becke exchange potential* E_{x}^{B88} *which includes the 1988 correction*, *and then strew in* 0.81 *unit of the Lee–Young–Parr correlation potential* $E_{\mathrm{c}}^{\mathrm{LYP}}$, both E_{x}^{B88} and $E_{\mathrm{c}}^{\mathrm{LYP}}$ meaning their nonlocal parts only. Then we end up with

$$E_{\mathrm{xc}} = E_{\mathrm{xc}}^{\mathrm{LSDA}} + 0.20\left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{LSDA}}\right) + 0.72E_{\mathrm{x}}^{B88} + 0.81E_{\mathrm{c}}^{\mathrm{LYP}}. \tag{3.43}$$

If you do it this way, satisfaction is (almost) guaranteed, and your results will agree very well with experiment.

3.5.3 The approximate character of the DFT versus apparent rigor of ab initio computations

There are lots of exchange-correlation potentials in the literature. There is an impression that their authors worried most about the agreement between theory and experiment. We can hardly admire this kind of science, but the alternative, i.e., the practice of *ab initio* methods with the intact and "holy" Hamiltonian operator, has its own dark sides and smells a bit like witch brew too. It does, because finally we have to choose a given atomic basis set, and this influences the results. It is true that we have the variational principle at our disposal, and it is possible to tell which result is more accurate. But more and more often in quantum chemistry we use some nonvariational methods (cf. Chapter 2). Besides, the Hamiltonian holiness disappears when the theory becomes relativistic (cf. Chapter V1-3).

[35] The Chaldean priests working out "Babylonian science" paid attention to making their small formulae efficient. The ancient Greeks (contemporary science owes them so much) were in favor of crystal clear reasoning.

[36] In fact, this is Kohn–Sham exchange energy (see Eq. (3.74)), because the Slater determinant wave function, used to calculate it, is the Kohn–Sham determinant, not the Hartree–Fock one.

All theoreticians would like to have agreement with experiment, and it is no wonder people tinker with the exchange-correlation enigma. This tinkering, however, is by no means arbitrary. There are some serious physical restraints behind it, which will be shown in a while.

3.6 On the physical justification for the exchange-correlation energy

We have to introduce several useful concepts, such as the "electron pair distribution function" and the "electron hole" (in a more formal way than in Chapter 2, p. 104).

3.6.1 The electron pair distribution function

From the N-electron wave function we may compute what is called the *electron pair correlation function* $\Pi(\mathbf{r}_1, \mathbf{r}_2)$, or in short, a pair function defined as[37]

$$\Pi(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \sum_{\sigma_1, \sigma_2} \int |\Psi|^2 \mathrm{d}\tau_3 \mathrm{d}\tau_4 \ldots \mathrm{d}\tau_N, \tag{3.44}$$

where the summation over spin coordinates pertains to all electrons (for electrons $3, 4, \ldots N$ the summation is hidden in the integrals over $d\tau$), while the integration is over the space coordinates of electrons $3, 4, \ldots N$.

> The function $\Pi(\mathbf{r}_1, \mathbf{r}_2)$ measures the probability density of finding one electron at the point indicated by $\mathbf{r}_1$ and another at $\mathbf{r}_2$, and tells us how the motions of two electrons are correlated. If Π were a *product* of two functions $\rho_1(\mathbf{r}_1) > 0$ and $\rho_2(\mathbf{r}_2) > 0$, then this motion is not correlated (because the probability of two events represents a product of the probabilities for each of the events only for *independent, i.e., uncorrelated events*).

Note that (see Eqs. (3.1) and (3.2) on p. 194)

$$\int \Pi(\mathbf{r}_1, \mathbf{r}_2) \mathrm{d}V_2 = N(N-1) \sum_{\sigma_1} \int \mathrm{d}\tau_2 \int |\Psi|^2 \mathrm{d}\tau_3 \mathrm{d}\tau_4 \ldots \mathrm{d}\tau_N = (N-1)\rho(\mathbf{r}_1) \tag{3.45}$$

and

$$\int \int \Pi(\mathbf{r}_1, \mathbf{r}_2) \mathrm{d}V_1 \mathrm{d}V_2 = (N-1) \int \rho(\mathbf{r}_1) \mathrm{d}V_1 = N(N-1). \tag{3.46}$$

[37] The function represents the diagonal element of the *two-particle electron density matrix*: $\Gamma(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') = N(N-1) \sum_{\text{all } \sigma} \int \Psi^* \left(\mathbf{r}_1' \sigma_1, \mathbf{r}_2', \sigma_2, \mathbf{r}_3, \sigma_3, \ldots \mathbf{r}_N, \sigma_N\right) \Psi \left(\mathbf{r}_1, \sigma_1, \mathbf{r}_2, \sigma_2, \mathbf{r}_3, \sigma_3, \ldots \mathbf{r}_N, \sigma_N\right) \mathrm{d}\mathbf{r}_3 \mathrm{d}\mathbf{r}_4 \ldots \mathrm{d}\mathbf{r}_N$, $\Pi(\mathbf{r}_1, \mathbf{r}_2) \equiv \Gamma(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1, \mathbf{r}_2)$.

Function Π appears in a natural way when we compute the mean value of the total electronic repulsion $\langle\Psi|U|\Psi\rangle$ with the Coulomb operator $U = \sum_{i<j}^{N} \frac{1}{r_{ij}}$ and a normalized N-electron wave function Ψ. Indeed, we have ("prime" in the summation corresponds to omitting the diagonal term)

$$\begin{aligned}
\langle\Psi|U\Psi\rangle &= \frac{1}{2}\sum_{i,j=1}^{N}{}'\langle\Psi|\frac{1}{r_{ij}}\Psi\rangle \\
&= \frac{1}{2}\sum_{i,j=1}^{N}{}'\{\sum_{\sigma_i,\sigma_j}\int \mathrm{d}\mathbf{r}_i\ \mathrm{d}\mathbf{r}_j \frac{1}{r_{ij}}\int|\Psi|^2\frac{\mathrm{d}\tau_1\mathrm{d}\tau_2\ldots\mathrm{d}\tau_N}{\mathrm{d}\tau_i\mathrm{d}\tau_j}\} \\
&= \frac{1}{2}\sum_{i,j=1}^{N}{}'\int \mathrm{d}\mathbf{r}_i\mathrm{d}\mathbf{r}_j\frac{1}{r_{ij}}\frac{1}{N(N-1)}\Pi(\mathbf{r}_i,\mathbf{r}_j) \\
&= \frac{1}{2}\frac{1}{N(N-1)}\sum_{i,j=1}^{N}{}'\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\frac{1}{r_{12}}\Pi(\mathbf{r}_1,\mathbf{r}_2) \\
&= \frac{1}{2}\frac{1}{N(N-1)}\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\frac{\Pi(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}\sum_{i,j=1}^{N}{}'1 \\
&= \frac{1}{2}\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\frac{\Pi(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}.
\end{aligned} \tag{3.47}$$

We will need this result in a moment. We see that to determine the contribution of the electron repulsion to the total energy we need the two-electron function Π. The first Hohenberg–Kohn theorem tells us that it is sufficient to know something simpler, namely, the electronic density ρ. How to reconcile these two demands?

The further DFT story will pertain to the following question: how can we change the potential in order to replace Π by ρ?

3.6.2 Adiabatic connection: from what is known towards the target

To begin let us write two Hamiltonians that are certainly very important for our goal. The first is the total Hamiltonian of our system (of course, with the Coulombic electron–electron interactions U). Let us denote the operator for some reasons as $\hat{H}(\lambda = 1)$, where we use the

abbreviation $v(\mathbf{r}_i) \equiv v(i)$, i.e.,

$$\hat{H}(\lambda = 1) = \sum_{i=1}^{N} [-\frac{1}{2}\Delta_i + v(i)] + U. \tag{3.48}$$

The second Hamiltonian $H(\lambda = 0)$ pertains to the Kohn–Sham fictitious system of the *noninteracting* electrons (it contains our wonder v_0, which we solemnly promise to search for, and the kinetic energy operator and nothing else), i.e.,

$$\hat{H}(\lambda = 0) = \sum_{i=1}^{N} [-\frac{1}{2}\Delta_i + v_0(i)]. \tag{3.49}$$

We will try to connect these two important systems by generating some intermediate Hamiltonians $\hat{H}(\lambda)$ for λ *intermediate* between 0 and 1, i.e.,

$$\hat{H}(\lambda) = \sum_{i=1}^{N} [-\frac{1}{2}\Delta_i + v_\lambda(i)] + U(\lambda), \tag{3.50}$$

where

$$U(\lambda) = \lambda \sum_{i<j}^{N} \frac{1}{r_{ij}}.$$

Our electrons are not real electrons for intermediate values of λ; each electron carries the electric charge $-\sqrt{\lambda}$.

> The intermediate Hamiltonian $\hat{H}(\lambda)$ contains a mysterious v_λ, which generates the exact density distribution ρ that corresponds to the Hamiltonian $\hat{H}(\lambda = 1)$, i.e., with all interactions in place. The same exact ρ corresponds to $\hat{H}(\lambda = 0)$.

We have, therefore, the ambition to go from the $\lambda = 0$ situation to the $\lambda = 1$ situation, all the time guaranteeing that the antisymmetric ground-state eigenfunction of $\hat{H}(\lambda)$ for any λ gives *the same electron density distribution* ρ, *the ideal (exact) one*. The way chosen represents a kind of "path of life" for us, because by sticking to it we do not lose the most precious of our treasures: the ideal density distribution ρ. We will call this path the *adiabatic connection*, because all the time we will adjust the correction computed to our actual position on the path.

Our goal will be the total energy $E(\lambda = 1)$. The adiabatic transition will be carried out by tiny steps. We will start with $E(\lambda = 0)$, and end up with $E(\lambda = 1)$. We write

$$E(\lambda = 1) = E(\lambda = 0) + \int_0^1 E'(\lambda)\mathrm{d}\lambda, \tag{3.51}$$

where the increments $\mathrm{d}E(\lambda) = E'(\lambda)\mathrm{d}\lambda$ will be calculated as the first-order perturbation energy correction (Eq. (V1-5.22)). The first-order correction is sufficient, because we are going to apply only infinitesimally small λ increments.[38] Each time when λ changes from λ to $\lambda + \mathrm{d}\lambda$, the situation at λ (i.e., the Hamiltonian $\hat{H}(\lambda)$ and the wave function $\Psi(\lambda)$) will be treated as unperturbed. What, therefore, does the perturbation operator look like? Well, when we go from λ to $\lambda + \mathrm{d}\lambda$, the Hamiltonian changes by $\hat{H}^{(1)}(\lambda) = \mathrm{d}\hat{H}(\lambda)$. Then the first-order perturbation correction to the energy given by (V1-5.22) represents the mean value of $\mathrm{d}\hat{H}(\lambda)$ with the unperturbed function $\Psi(\lambda)$, i.e.,

$$\mathrm{d}E(\lambda) = \langle\Psi(\lambda)|\mathrm{d}\hat{H}(\lambda)\Psi(\lambda)\rangle, \tag{3.52}$$

where in $\mathrm{d}\hat{H}$ we only have a change of v_λ and of $U(\lambda)$ due to the change of λ, i.e.,

$$\mathrm{d}\hat{H}(\lambda) = \sum_{i=1}^{N} \mathrm{d}v_\lambda(i) + \mathrm{d}\lambda \sum_{i<j}^{N} \frac{1}{r_{ij}}. \tag{3.53}$$

Note that we have succeeded in writing such a simple formula *because the kinetic energy operator stays unchanged all the time (it does not depend on* λ). Let us insert this into the first-order correction to the energy in order to get $\mathrm{d}E(\lambda)$ and use Eqs. (3.12) and (3.47). We obtain

$$\mathrm{d}E(\lambda) = \langle\Psi(\lambda)|\mathrm{d}\hat{H}(\lambda)\Psi(\lambda)\rangle = \int \rho(\mathbf{r})\mathrm{d}v_\lambda(\mathbf{r})\mathrm{d}\mathbf{r} + \frac{1}{2}\mathrm{d}\lambda \int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \tag{3.54}$$

In the last formula we introduced a function Π_λ that is an analog of the pair function Π, but pertains to the electrons carrying the charge $\sqrt{\lambda}$ (we have used the formula (3.47), noting that we have a λ-dependent wave function $\Psi(\lambda)$).

[38] λ plays a different role here than the perturbational parameter λ on p. V1-276.

In order to go from $E(\lambda = 0)$ to $E(\lambda = 1)$, it is sufficient just to integrate this expression from 0 to 1 over λ (this corresponds to the infinitesimally small increments of λ as mentioned before). Note that (by definition) ρ *does not depend on* λ, which is of fundamental importance in the success of the integration $\int \rho(\mathbf{r})\mathrm{d}v_\lambda(\mathbf{r})\mathrm{d}\mathbf{r}$ and gives the result

$$E(\lambda = 1) - E(\lambda = 0) = \int \rho(\mathbf{r})\{v - v_0\}(\mathbf{r})\mathrm{d}\mathbf{r} + \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \tag{3.55}$$

The energy for $\lambda = 0$, i.e., for the noninteracting electrons in an unknown external potential v_0, will be written as (cf. the formulae (3.18) and (3.20))

$$E(\lambda = 0) = \sum_i \varepsilon_i = T_0 + \int \rho(\mathbf{r})v_0(\mathbf{r})\mathrm{d}\mathbf{r}. \tag{3.56}$$

Inserting this into (3.55) we obtain $E(\lambda = 1)$, i.e., the energy of our original system,

$$E(\lambda = 1) = T_0 + \int \rho(\mathbf{r})v(\mathbf{r})\mathrm{d}\mathbf{r} + \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \tag{3.57}$$

Note that according to Eq. (3.45) we get $\int \Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = (N-1)\rho(\mathbf{r}_1)$, because $\rho(\mathbf{r}_1)$ does not depend on λ due to the nature of our adiabatic transformation.

The expression for $E(\lambda = 1)$ may be simplified by introducing the pair distribution function Π_{aver}, which is the $\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)$ averaged over $\lambda = [0, 1]$,

$$\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2) \equiv \int_0^1 \Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\lambda. \tag{3.58}$$

Here also (we will use this result in a moment)

$$\int \Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = \int_0^1 \int \Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\lambda\mathrm{d}\mathbf{r}_2 = (N-1)\rho(\mathbf{r}_1)\int_0^1 \mathrm{d}\lambda = (N-1)\rho(\mathbf{r}_1). \tag{3.59}$$

Finally we obtain the following expression for

the total energy E:

$$E(\lambda = 1) = T_0 + \int \rho(\mathbf{r}) v(\mathbf{r}) \mathrm{d}\mathbf{r} + \frac{1}{2} \int\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \frac{\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \tag{3.60}$$

Note that this equation is similar to the total energy expression appearing in traditional quantum chemistry[39] (without repulsion of the nuclei),

$$E = T + \int \rho(\mathbf{r}) v(\mathbf{r}) \mathrm{d}\mathbf{r} + \frac{1}{2} \int\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \frac{\Pi(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \tag{3.61}$$

As we can see, the DFT total energy expression, instead of the mean kinetic energy of the fully interacting electrons T, contains T_0, i.e., the mean kinetic energy of the noninteracting (Kohn–Sham) electrons.[40] We pay, however, a price, namely, we need to compute the function Π_{aver} somehow. Note, however, that the correlation energy dragon has been driven into the problem of finding a two-electron function Π_{aver}.

3.6.3 Exchange-correlation energy and the electron pair distribution function

What is the relation between Π_{aver} and the exchange-correlation energy E_{xc} introduced earlier? We find it immediately, comparing the total energy given in Eqs. (3.19) and (3.21) and now in (3.60). It is seen that the exchange-correlation energy

$$E_{\mathrm{xc}} = \frac{1}{2} \int\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \frac{1}{r_{12}} \{\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\}. \tag{3.62}$$

The energy looks as if it were a potential energy, but it implicitly incorporates (in Π_{aver}) the kinetic energy correction for changing the electron noninteracting to the electron interacting system.

Now let us try to get some information about the integrand, i.e., Π_{aver}, by introducing the notion of the electron hole.

[39] It is evident from the mean value of the total Hamiltonian (taking into account the mean value of the electron–electron repulsion, Eq. (3.12) and (3.47)).

[40] As a matter of fact the whole Kohn–Sham formalism with the fictitious system of the noninteracting electrons has been designed precisely because of this.

3.6.4 The correlation dragon hides in the exchange-correlation hole

Electrons do not like each other, which manifests itself by Coulombic repulsion. On top of that, two electrons having the same spin coordinates hate each other (Pauli exclusion principle) and also try to get out of the other electron's way. This has been analyzed in Chapter 2 (p. 105). We should somehow highlight these features, because both concepts are basic and simple.

Let us introduce the definition of the exchange-correlation hole h_{xc} as satisfying the equation

$$\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) + \rho(\mathbf{r}_1)h_{\mathrm{xc}}(\mathbf{r}_1; \mathbf{r}_2). \tag{3.63}$$

Thus, in view of Eqs. (3.61) and (3.47) we have the electron repulsion energy

$$\frac{1}{2}\int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}} = \frac{1}{2}\int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} + \frac{1}{2}\int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)h_{\mathrm{xc}}(\mathbf{r}_1; \mathbf{r}_2)}{r_{12}} \tag{3.64}$$

as the self-interaction of the electron cloud of the density distribution $\rho(\mathbf{r})$, Eq. (3.21), plus a correction $\frac{1}{2}\int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)h_{\mathrm{xc}}(\mathbf{r}_1;\mathbf{r}_2)}{r_{12}}$, which takes into account all necessary interactions, i.e., our complete correlation dragon is certainly hidden in the unknown hole function $h_{\mathrm{xc}}(\mathbf{r}_1; \mathbf{r}_2)$. Note that the hole charge distribution integrates over $\mathbf{r}_2$ to the charge -1 irrespectively of the position $\mathbf{r}_1$ of electron 1. Indeed, integrating (3.63) over $\mathbf{r}_2$ and using (3.45) we get[41]

$$\int \mathrm{d}\mathbf{r}_2 h_{\mathrm{xc}}(\mathbf{r}_1; \mathbf{r}_2) = -1. \tag{3.65}$$

3.6.5 Electron holes in spin resolution

First, we will decompose the function Π_{aver} into the components related to the spin functions[42] of electrons 1 and 2, $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, $\beta\beta$, i.e.,

$$\Pi_{\mathrm{aver}} = \Pi^{\alpha\alpha}_{\mathrm{aver}} + \Pi^{\alpha\beta}_{\mathrm{aver}} + \Pi^{\beta\alpha}_{\mathrm{aver}} + \Pi^{\beta\beta}_{\mathrm{aver}}, \tag{3.66}$$

[41] $$\int \mathrm{d}\mathbf{r}_2 h_{\mathrm{xc}}(\mathbf{r}_1; \mathbf{r}_2) = \int \mathrm{d}\mathbf{r}_2 h_{\mathrm{xc}}(\mathbf{r}_1; \mathbf{r}_2) = \int \mathrm{d}\mathbf{r}_2 \frac{\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{\rho(\mathbf{r}_1)} =$$
$$= \frac{1}{\rho(\mathbf{r}_1)}\int \mathrm{d}\mathbf{r}_2 \left(\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\right) = \frac{1}{\rho(\mathbf{r}_1)}[(N-1)\,\rho(\mathbf{r}_1) - N\rho(\mathbf{r}_1)] = -1.$$

[42] Such a decomposition follows from Eq. (3.44). We average all the contributions $\Pi^{\sigma\sigma'}$ separately and obtain the formula.

where $\Pi^{\alpha\beta}_{\text{aver}} dV_1 dV_2$ represents a measure of the probability density[43] that two electrons are in their small boxes indicated by the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, the boxes have the volumes dV_1 and dV_2, and the electrons are described by the spin functions α and β (the other components of Π_{aver} are defined in a similar way). Since $\rho = \rho_\alpha + \rho_\beta$, the exchange-correlation energy can be written as[44]

$$E_{\text{xc}} = \frac{1}{2} \sum_{\sigma\sigma'} \int\int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\Pi^{\sigma\sigma'}_{\text{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho_\sigma(\mathbf{r}_1)\rho_{\sigma'}(\mathbf{r}_2)}{r_{12}}, \tag{3.67}$$

where the summation goes over the spin coordinates. It is seen that

> a nonzero value of E_{xc} tells us whether the behavior of electrons deviates from their *independence* (the latter is described by the product of the probability densities, i.e., the second term in the nominator). This means that E_{xc} has to contain the electron–electron correlation resulting from Coulombic interaction and their avoidance from the Pauli exclusion principle.

By using the abbreviation for the exchange-correlation hole

$$h^{\sigma\sigma'}_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2) \equiv \frac{\Pi^{\sigma\sigma'}_{\text{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho_\sigma(\mathbf{r}_1)\rho_{\sigma'}(\mathbf{r}_2)}{\rho_\sigma(\mathbf{r}_1)},$$

we obtain

$$E_{\text{xc}} = \frac{1}{2} \sum_{\sigma\sigma'} \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\rho_\sigma(\mathbf{r}_1)}{r_{12}} h^{\sigma\sigma'}_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2). \tag{3.68}$$

The final expression for the exchange-correlation hole is

---

43 λ-averaged.

44 Indeed,

$$\begin{aligned} E_{\text{xc}} &= \frac{1}{2} \int\int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\Pi_{\text{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} = \\ &= \frac{1}{2} \int\int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\sum_{\sigma\sigma'} \Pi^{\sigma\sigma'}_{\text{aver}}(\mathbf{r}_1, \mathbf{r}_2) - (\sum_\sigma \rho_\sigma(\mathbf{r}_1))(\sum_{\sigma'} \rho_{\sigma'}(\mathbf{r}_2))}{r_{12}} = \\ &= \frac{1}{2} \sum_{\sigma\sigma'} \int\int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\Pi^{\sigma\sigma'}_{\text{aver}}(\mathbf{r}_1, \mathbf{r}_2) - \rho_\sigma(\mathbf{r}_1)\rho_{\sigma'}(\mathbf{r}_2)}{r_{12}}. \end{aligned}$$

EXCHANGE-CORRELATION HOLE

$$h_{\mathrm{xc}}^{\sigma\sigma'}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\Pi_{\mathrm{aver}}^{\sigma\sigma'}(\mathbf{r}_1, \mathbf{r}_2)}{\rho_\sigma(\mathbf{r}_1)} - \rho_{\sigma'}(\mathbf{r}_2). \tag{3.69}$$

The hole represents that part of the pair distribution function that is inexplicable by a product-like dependence. Since a product function describes independent electrons, the hole function grasps the "intentional" avoidance of the two electrons.

We have, therefore, four exchange-correlation holes: $h_{\mathrm{xc}}^{\alpha\alpha}$, $h_{\mathrm{xc}}^{\alpha\beta}$, $h_{\mathrm{xc}}^{\beta\alpha}$, $h_{\mathrm{xc}}^{\beta\beta}$.

3.6.6 The dragon's ultimate hide-out: the correlation hole!

Dividing the exchange-correlation hole into the exchange hole and the correlation hole

The restrictions introduced come from the Pauli exclusion principle, and hence have been related to the exchange energy. So far no restriction has appeared that would stem from the Coulombic interactions of electrons.[45] This made people think of differentiating the holes into two contributions: exchange hole h_{x} and correlation hole h_{c} (so far called the Coulombic hole). Let us begin with a formal division of the exchange-correlation energy into the exchange and the correlation parts. We write

EXCHANGE-CORRELATION ENERGY

$$E_{\mathrm{xc}} = E_{\mathrm{x}} + E_{\mathrm{c}}, \tag{3.70}$$

and *we will say that we know what the exchange part is*.

The DFT exchange energy (E_x) is calculated in the same way as in the Hartree–Fock method, but with the Kohn–Sham determinant. The correlation energy E_{c} represents just a rest.

This is again the same strategy of chasing the electronic correlation dragon into a hole, this time into the correlation hole. When we do not know a quantity, we write down what we know plus

[45] This is the role of the Hamiltonian.

a remainder. And the dragon with a hundred heads sits in the latter. Because of this division, the Kohn–Sham equation will contain the sum of the exchange and correlation potentials instead of v_{xc}, i.e.,

$$v_{\mathrm{xc}} = v_{\mathrm{x}} + v_{\mathrm{c}} \tag{3.71}$$

with

$$v_{\mathrm{x}} \equiv \frac{\delta E_{\mathrm{x}}}{\delta \rho}, \tag{3.72}$$

$$v_{\mathrm{c}} \equiv \frac{\delta E_{\mathrm{c}}}{\delta \rho}. \tag{3.73}$$

Let us recall what the Hartree–Fock exchange energy[46] looks like (Chapter V1-8, Eq. (V1-8.38)). The Kohn–Sham exchange energy looks, of course, the same, except that the spin orbitals are now Kohn–Sham, not Hartree–Fock. Therefore, we have the exchange energy E_{x} as (the sum is over the molecular spin orbitals[47])

$$\begin{aligned} E_{\mathrm{x}} &= -\frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} K_{ij} = -\frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} \langle ij|ji\rangle \\ &= -\frac{1}{2}\sum_{\sigma}\int \frac{\left\{\sum_{i=1}^{N}\phi_i^*(1)\phi_i(2)\right\}\left\{\sum_{j=1}^{N}\phi_j^*(2)\phi_j(1)\right\}}{r_{12}}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \\ &= -\frac{1}{2}\sum_{\sigma}\int \frac{|\rho_\sigma(\mathbf{r}_1;\mathbf{r}_2)|^2}{r_{12}}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2, \end{aligned} \tag{3.74}$$

where

$$\rho_\sigma(\mathbf{r}_1;\mathbf{r}_2) \equiv \sum_{i=1}^{N}\phi_i(\mathbf{r}_1,\sigma)\phi_i^*(\mathbf{r}_2,\sigma) \tag{3.75}$$

represents the *one-particle* density matrix for the σ subsystem (Eq. (3.2)) and ρ_σ is obtained from the Kohn–Sham determinant. Note that density $\rho_\sigma(\mathbf{r})$ is its diagonal, i.e., $\rho_\sigma(\mathbf{r}) \equiv \rho_\sigma(\mathbf{r};\mathbf{r})$.

46 The one which appeared from the exchange operator, i.e., containing the exchange integrals.

47 Note that spin orbital i has to have the same spin function as spin orbital j (otherwise $K_{ij} = 0$).

The above may be incorporated into the exchange energy E_{x} equal to

$$E_{\mathrm{x}} = \frac{1}{2}\sum_{\sigma\sigma'}\int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho_\sigma(\mathbf{r}_1)}{r_{12}} h_{\mathrm{x}}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2) \tag{3.76}$$

if the exchange hole (also known as Fermi hole) h is proposed as

$$h_{\mathrm{x}}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2) = \delta_{\sigma\sigma'}\left\{-\frac{|\rho_\sigma(\mathbf{r}_1;\mathbf{r}_2)|^2}{\rho_\sigma(\mathbf{r}_1)}\right\}. \tag{3.77}$$

It is seen that the exchange hole is negative everywhere[48] and diagonal in the spin index. Let us integrate the exchange hole over $\mathbf{r}_2$ for an arbitrary position of electron 1. First we have

$$|\rho_\sigma(\mathbf{r}_1;\mathbf{r}_2)|^2 = \sum_i \phi_i^*(\mathbf{r}_2,\sigma)\phi_i(\mathbf{r}_1,\sigma)\sum_j \phi_j^*(\mathbf{r}_1,\sigma)\phi_j(\mathbf{r}_2,\sigma) =$$
$$\sum_{ij}\phi_i^*(\mathbf{r}_2,\sigma)\phi_i(\mathbf{r}_1,\sigma)\phi_j^*(\mathbf{r}_1,\sigma)\phi_j(\mathbf{r}_2,\sigma).$$

Then, the integration gives[49]

$$\int h_{\mathrm{x}}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = -\delta_{\sigma\sigma'}. \tag{3.78}$$

Therefore,

the exchange hole $h_{\mathrm{x}}^{\sigma\sigma}(\mathbf{r}_1,\mathbf{r}_2)$ is negative everywhere and when integrated over $\mathbf{r}_2$ at any position $\mathbf{r}_1$ of electron 1 gives -1, i.e., exactly the charge of one electron is expelled from the space around electron 1.

[48] Which has its origin in the minus sign before the exchange integrals in the total energy expression.

[49] Indeed,

$$\int h_{\mathrm{x}}^{\sigma\sigma'}(\mathbf{r}_1,\mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = -\delta_{\sigma\sigma'}\frac{1}{\rho_\sigma(\mathbf{r}_1)}\int |\rho_\sigma(\mathbf{r}_1;\mathbf{r}_2)|^2\,\mathrm{d}\mathbf{r}_2 =$$
$$= -\delta_{\sigma\sigma'}\frac{1}{\rho_\sigma(\mathbf{r}_1)}\sum_{ij}\phi_i(\mathbf{r}_1,\sigma)\phi_j^*(\mathbf{r}_1,\sigma)\int \phi_i^*(\mathbf{r}_2,\sigma)\phi_j(\mathbf{r}_2,\sigma)\mathrm{d}\mathbf{r}_2 =$$
$$= -\delta_{\sigma\sigma'}\frac{1}{\rho_\sigma(\mathbf{r}_1)}\sum_{ij}\phi_i(\mathbf{r}_1,\sigma)\phi_j^*(\mathbf{r}_1,\sigma)\delta_{ij} = -\delta_{\sigma\sigma'}\frac{1}{\rho_\sigma(\mathbf{r}_1)}\sum_i \phi_i(\mathbf{r}_1,\sigma)\phi_i^*(\mathbf{r}_1,\sigma) =$$
$$= -\delta_{\sigma\sigma'}\frac{1}{\rho_\sigma(\mathbf{r}_1)}\rho_\sigma(\mathbf{r}_1) = -\delta_{\sigma\sigma'}.$$

What, therefore, does the correlation hole look like? According to the philosophy of dragon chasing it is the rest,

$$h_{\mathrm{xc}}^{\sigma\sigma'} = h_{\mathrm{x}}^{\sigma\sigma'} + h_{\mathrm{c}}^{\sigma\sigma'}. \tag{3.79}$$

The correlation energy from Eq. (3.70) has, therefore, the following form:

$$E_{\mathrm{c}} = \frac{1}{2}\sum_{\sigma\sigma'}\int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \frac{\rho_\sigma(\mathbf{r}_1)}{r_{12}} h_{\mathrm{c}}^{\sigma\sigma'}(\mathbf{r}_1, \mathbf{r}_2). \tag{3.80}$$

Since the exchange hole has already fulfilled the boundary conditions (3.65)–(3.78) forced by the Pauli exclusion principle, the correlation hole satisfies a simple boundary condition,

$$\int h_{\mathrm{c}}^{\sigma\sigma'}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = 0. \tag{3.81}$$

Thus, the correlation hole means that electron 1 is pushing electron 2 (i.e., other electrons) off, but this means only the pushed electrons are moved further out.

The dragon of electronic correlation has been chased into the correlation hole. Numerical experience turns out to reveal (below an example will be given) that

> the exchange energy E_{x} is much more important than the correlation energy E_{c} and, therefore, scientists managed in their chasing to replace the terrible exchange-correlation dragon to a quite small beast hiding in the correlation hole (to be found).

3.6.7 Physical grounds for the DFT functionals

LDA

The LDA is not as primitive as it looks. The electron density distribution for the homogeneous gas model satisfies the Pauli exclusion principle and, therefore, this approximation gives the Fermi holes that fulfill the boundary conditions with Eqs. (3.65), (3.78), and (3.81). The LDA is often used because it is rather inexpensive, while still giving reasonable geometry of molecules and vibrational frequencies.[50] The quantities that the LDA fails to reproduce are the binding energies,[51] ionization potentials, and the intermolecular dispersion interaction.

[50] Some colleagues of mine sometimes add a malicious remark that the frequencies are so good that they even take into account the anharmonicity of the potential.

[51] The average error in a series of molecules may even be of the order of 40 kcal/mol; this is a lot, since the chemical bond energy is of the order of about 100 kcal/mol.

The Perdew–Wang functional (PW91)

Perdew noted a really dangerous feature in an innocent and reasonable looking GEA potential. It turned out that in contrast to the LDA the boundary conditions for the electron holes were not satisfied. For example, the exchange hole was not negative everywhere as Eq. (3.77) requires. Perdew and Wang corrected this deficiency in a way similar to that of Alexander the Great when he cut the Gordian knot. They tailored the formula for E_{xc} in such a way as to change the positive values of the function just to zero, while the peripheral parts of the exchange holes were cut to force the boundary conditions to be satisfied anyway. The authors noted an important improvement in the results.

The functional B3LYP

It was noted that the LDA and even GEA models systematically give too large chemical bond energies. On the other hand, it is known that the Hartree–Fock method is notorious for making the bonds too weak. What are we to do? Well, just mix the two types of potential and hope to have an improvement with respect to any of the models. Recall formula (3.58) for Π_{aver}, where the averaging extended from $\lambda = 0$ to $\lambda = 1$. The contribution to the integral for λ close to 0 comes from the situations similar to the fictitious model of noninteracting particles, where the wave function has the form of the Kohn–Sham determinant. Therefore, those contributions contain the exchange energy E_{x} corresponding to such a determinant. We may conclude that a contribution from the Kohn–Sham exchange energy $E_{\mathrm{x}}^{\mathrm{HF}}$ might look quite natural.[52] This is what the B3LYP method does (Eq. (3.43)). Of course, it is not possible to justify the particular proportions of the B3LYP ingredients. Such things are justified only by their success.[53]

3.7 Visualization of electron pairs: electron localization function (ELF)

One of the central ideas of general chemistry is the notion of electron pair, i.e., two electrons of opposite spins, which occupy a certain region of space. Understanding chemistry means knowing what is the role of these electron pairs in individual molecules (which is directly related to their structure) and what may happen to them when two molecules are in contact (chemical reactions). Where in a molecule do electron pairs prefer to be? This is the role of the ELF, which may be seen as an idea of visualization helping chemists to elaborate what is known as chemical intuition ("understanding"), an important qualitative generalization staying behind any practical chemist's action like, e.g., planning chemical synthesis.

In Chapter 2 (p. 105) and in the present chapter we were dealing with the Fermi hole that characterized quantitatively the strength of the Pauli exclusion principle: two electrons with same

[52] The symbol HF pertains rather to Kohn–Sham than to Hartree–Fock.

[53] As in herbal therapy.

spin coordinate avoid each other. We have given several examples showing that a (probe) electron with the same spin as a reference electron tries to be as far as possible from the later one – a very strong effect. And what about electrons of opposite spins? Well, they are not subject to this restriction and can approach each other, but not too close, because of the Coulomb interaction. As a result, in molecules (and atoms as well) *we have to do with a shell-like electronic structure*: an electron pair (while keeping a reasonable electron–electron distance) may profit from occupying a domain very close to the nuclei. There is no future over there for any other electron or electron pair, because of the Pauli exclusion principle. Other electron pairs have to occupy separately other domains in space.

The strength of the Pauli exclusion principle will certainly depend on the position in space with respect to the nuclear framework. Testing this strength represents our goal now.

Let us take a reference electron at position $\mathbf{r}$ in a global coordinate system (Fig. 3.8a) and let us try to approach to it a "probe electron" of the same spin coordinate and shown by the radius vector $\mathbf{r}+\mathbf{r}_p$ (i.e., the probe electron would have the radius vector $\mathbf{r}_p$, when seen from the reference electron shown by $\mathbf{r}$). We will consider only such $\mathbf{r}_p$, which ensure the probe electron be enclosed around the reference one in a sphere of radius R_p, i.e., $r_p \leq R_p$. The key function is the Fermi hole function, $h_x^{\sigma\sigma}(\mathbf{r}, \mathbf{r}+\mathbf{r}_p)$ of Eq. (3.77) on p. 231. We are interested in what fraction of the probe electron is outside the abovementioned sphere. For small r_p one can certainly write the following Taylor expansion about point $\mathbf{r}$:

$$h_x^{\sigma\sigma}(\mathbf{r}, \mathbf{r}+\mathbf{r}_p) = -\rho_\sigma(\mathbf{r}) + (\nabla h_x^{\sigma\sigma})_{\mathbf{r}_p=\mathbf{0}} \cdot \mathbf{r}_p + C(\mathbf{r})\, r_p^2 + ... \tag{3.82}$$

Since function $h_x^{\sigma\sigma}(\mathbf{r}, \mathbf{r}+\mathbf{r}_p)$ has a minimum for[54] $\mathbf{r}_p = \mathbf{0}$ (the most improbable scenario: two electrons of the same spin coordinate would sit one on top of the other!), we have vanishing of the gradient, i.e., $(\nabla h_x^{\sigma\sigma})_{\mathbf{r}_p=\mathbf{0}} = \mathbf{0}$. As to the last term shown, instead of the usual second derivatives calculated at the minimum we simplified things by putting a rotation-averaged constant $C(\mathbf{r}) > 0$. Truncating in Eq. (3.82) all terms beyond quadratic ones, we get

$$h_x^{\sigma\sigma}(\mathbf{r}, \mathbf{r}+\mathbf{r}_p) = -\rho_\sigma(\mathbf{r}) + C(\mathbf{r})\, r_p^2. \tag{3.83}$$

Thus, $h_x^{\sigma\sigma}(\mathbf{r}, \mathbf{r}+\mathbf{r}_p)$ is nothing but a paraboloidal well with the minimum equal to $-\rho_\sigma(\mathbf{r})$ at position $\mathbf{r}$ of the reference electron. The well is controlled by the value of $C(\mathbf{r})$; for small $C(\mathbf{r})$ the well is wide, for large $C(\mathbf{r})$ the well is narrow (Fig. 3.8b1,b2).

[54] This is true, but for those positions $\mathbf{r}$ of the reference electron for which the density $\rho(\mathbf{r})$ is not too small. For $\mathbf{r}$ belonging to peripheries of molecule, a substantial exchange hole cannot be dug out at $\mathbf{r}$, because the ground there is shallow. In such a case the exchange hole "stays behind" $\mathbf{r}$, in the region of appreciable values of ρ.

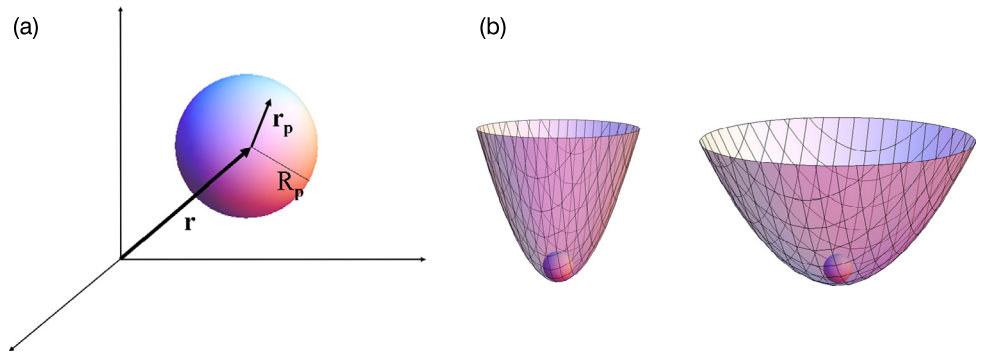


Fig. 3.8. A gear to test the power of the Pauli exclusion principle (Fermi hole). (a) The reference electron with the radius vector $\mathbf{r}$ and probe electron with the radius vector $\mathbf{r}+\mathbf{r}_p$ (both with the same spin coordinate). From the sphere shown an electron of the same spin as that of the reference one is expelled. Therefore, the same sphere is a residence for two electrons of the *opposite* spins (electron pair). (b) Two parabolic Fermi holes – a result of the Pauli exclusion principle. In each hole the reference electron is shown (represented by a small ball). A narrow well means a large value of $C(\mathbf{r})$ and therefore a small value of ELF, which means a small propensity to host an electron pair. In contrast, a wide well corresponds to a large propensity to home there an electron pair.

What fraction of the probe electron charge is expelled from the sphere of radius R_p? This can be calculated from the hole function by integrating it over all possible θ_p, ϕ_p, and $r_p < R_p$. We have

$$\int d\mathbf{r}_p h_x^{\sigma\sigma}(\mathbf{r}, \mathbf{r}+\mathbf{r}_p) = -\rho_\sigma(\mathbf{r}) \int d\mathbf{r}_p + C(\mathbf{r}) \int r_p^2 d\mathbf{r}_p =$$

$$-\rho_\sigma(\mathbf{r})\frac{4}{3}\pi R_p^3 + C(\mathbf{r}) \int_0^{R_p} dr_p r_p^2 r_p^2 \int_0^{\pi} d\theta_p \sin\theta_p \int_0^{2\pi} d\phi_p = -\rho_\sigma(\mathbf{r})\frac{4}{3}\pi R_p^3 + C(\mathbf{r})\frac{4\pi R_p^5}{5}.$$

Now we have to decide what to choose as R_p if the reference electron has position $\mathbf{r}$. It is reasonable to make R_p dependent on the position in space, because the Fermi hole should be created easier (i.e., would be larger) for small values of the electron density $\rho_\sigma(\mathbf{r})$, and harder for a larger value. Quite arbitrarily we choose such a function $R_p(\mathbf{r})$ that the first term on the right-hand side satisfies, i.e.,

$$-\rho_\sigma(\mathbf{r})\frac{4}{3}\pi R_p^3 = -1, \tag{3.84}$$

which means that close to the point shown by vector $\mathbf{r}$ the volume of the sphere of radius R_p is equal to the mean volume per single electron of the spin coordinate σ in the uniform electron gas of density $\rho_\sigma(\mathbf{r})$. If we have the closed shell case ($\rho_\alpha(\mathbf{r}) = \rho_\beta(\mathbf{r})$), the same volume contains

also an electron of the opposite spin. This means such a volume contains on average a complete *electron pair*. As a consequence, using Eq. (3.84) one may write

$$C(\mathbf{r})\frac{4\pi R_p^5}{5} \sim C(\mathbf{r})\,(\rho_\sigma(\mathbf{r}))^{-\frac{5}{3}}.$$

Becke and Edgecombe used this expression to construct a function ELF($\mathbf{r}$) that reflects a tendency for the electron pair[55] to reside at point $\mathbf{r}$ (a large value of the ELF for a strong tendency, a small one for a weak tendency), i.e.,

$$\mathrm{ELF}(\mathbf{r}) = \frac{1}{1+\kappa C(\mathbf{r})\,(\rho_\sigma(\mathbf{r}))^{-\frac{5}{3}}},$$

where $\kappa > 0$ represents an arbitrary scaling constant. Since $\infty > C(\mathbf{r}) > 0$, at any κ we have $0 \leq \mathrm{ELF}(\mathbf{r}) \leq 1$.

A large ELF($\mathbf{r}$) value at position $\mathbf{r}$ means a large Fermi hole over there, or a lot of space for an electron pair.

ELF($\mathbf{r}$) represents a function in a three-dimensional space. How to visualize such a function inside a molecule? Well, one way is to look at an "iso-ELF" surface. But, which one (we have to decide among the ELF values ranging from 0 to 1)? There is a general problem with isosurfaces, because one has to choose an ELF value that returns interesting information. A unfortunate value may give a useless result.[56]

Let us take a series of diatomic molecules, F_2, Cl_2, Br_2; their electronic structure is believed to be well known in chemistry. The meaning of such a sentence is in reality that whoever we have met in the past, when asked about it, said that there is a single covalent bond in all of these molecules. Students were told about the VSEPR algorithm (Chapter V1-8), which also predicts on each atom a tetrahedral configuration of the three lone pairs and a single chemical bond with the partner. If one were interested in the electron density coming from these three lone pairs, one would discover that instead of a "tripod-like" density we would get an object having *cylindrical* symmetry. This would reflect the fact that the tripod's legs can be positioned anywhere on the ring with the center on the atom–atom axis, and perpendicular to the axis.

Let us begin from Br_2. The above conviction seems to be confirmed by the obtained ELF($\mathbf{r}$) function (Fig. 3.9a). Indeed, as one can see the ELF isosurface shows two tori, each behind the corresponding bromine atom, as could be expected for the electron density coming from the

[55] A.D. Becke, K.E. Edgecombe, *J. Chem. Phys.*, 92(1990)5397.

[56] For example, a section of Himalaya mountains at 10 000 m altitude brings a function that is zero everywhere.

three corresponding atomic lone pairs, and each conserving cylindrical symmetry. What about the bond electron pair? Well, we see (Fig. 3.9a) that right in the middle of the Br–Br distance there is a preferred place for an electron pair, also conserving cylindrical symmetry of the total system.

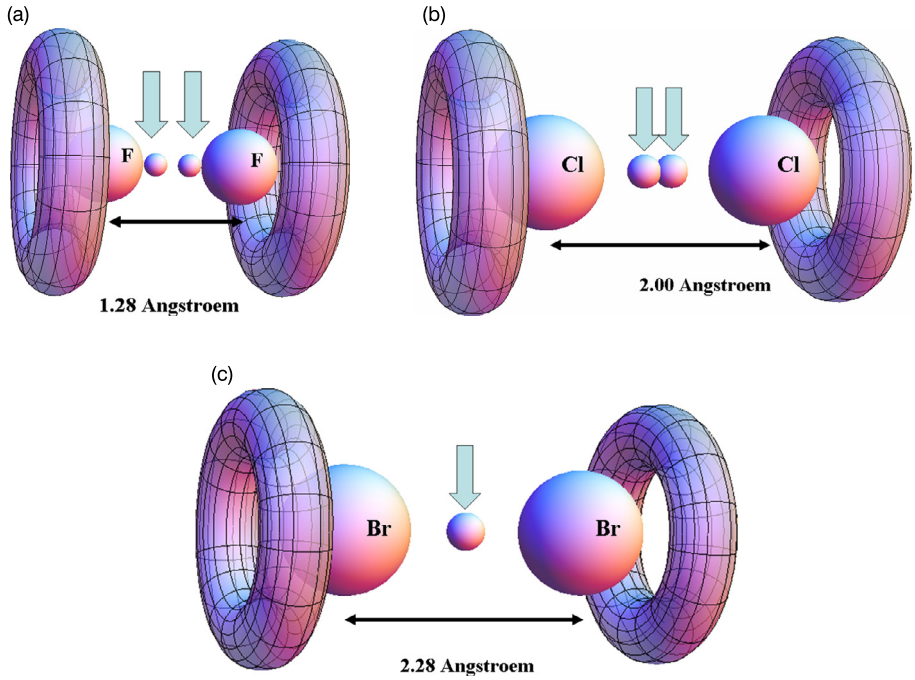


Fig. 3.9. A schematic picture showing isosurfaces of ELF(r) for (a) F_2, (b) Cl_2, and (c) Br_2. In all cases we see the peripheral tori, conserving the cylindrical symmetry of the system, and corresponding to the electronic lone pairs (three for each atom, together with the bond they form a nearly tetrahedral configuration). The iso-ELF islands shown by the arrows correspond to the regions with higher probability density for finding that electron pair which is responsible for the chemical bond. The cases (a) F_2 and (b) Cl_2 surprisingly show two such islands, while in the case of Br_2 we have a single island of the largest tendency to find an electron pair.

This gives the impression that ELF($\mathbf{r}$) tells us[57] what every freshman knows either from his teacher, or from Professor Gilbert Lewis, or from the VSEPR algorithm. What could the student and ourselves expect from Cl_2 and F_2? Obviously, the same! And yet the ELF($\mathbf{r}$) makes a surprise for us! It turns out that the ELF procedure shows *two* regions for the bond electron pair. Why? Well, there is an indication. Let us recall (say, from the valence bond method, Chapter 2), that among important valence bond structures are the covalent one and two ionic structures.

[57] Using strange shapes, colors, shading, and even reflexes of light on them, which shamelessly play the role of making us naively believe in all these things to be real!

In the case of Cl_2 they would be the Heitler–London function describing the covalent bond Cl–Cl and the two ionic structures corresponding to Cl^+Cl^- and Cl^-Cl^+, respectively. Such ionic structures are necessary for a reliable description of the molecule at finite internuclear distances.[58] In a particular ionic structure one electron is shifted towards one of the atoms, i.e., such a structure breaks the symmetry. However, the presence of two such structures (of equal weight, for a homopolar molecule) restores this symmetry. One may say that for Cl_2 and F_2 there is a large fluctuation of the bond electron pair position that strengthens the bond – a *charge-shift (CS) bond*). Therefore,

> besides the covalent bonds (like in Br–Br), the ionic bonds (as in NaCl, see Chapter V1-6), and the polar bonds (like in C–H), there are the CS bonds, the bonds with fluctuating position of the bonding electron pair.

The concept of the CS bonds as some distinct kind of chemical bonds comes from independent theoretical considerations,[59] but also seems to find its confirmation in a specialized visualization tool, which is in fact what the ELF idea really provides.

3.8 The DFT excited states

Ground states for a given symmetry

DFT is usually considered as a ground-state theory. One should, however, remember that the exact ground-state electron density ρ_0 contains also information about all the excited states (see p. 206). The problem is we do not know yet how to extract this information from ρ_0. Some of the excited states are the lowest-energy states belonging to a given irreducible representation of the symmetry group of the Hamiltonian. In such cases, forcing the proper symmetry of the Kohn–Sham orbitals leads to the solution for the corresponding excited state. Thus, these states are excited ones, but formally they can be treated as the ground states (in the corresponding irreducible representation).

Time-dependent DFT

Is it possible to detect excited states by exciting the ground state? Well, there is a promising path showing how to do it.[60] From Chapter V1-2 we know (see p. V1-114) that this requires

[58] For infinite distance they do not count.

[59] S. Shaik, D. Danovich, B. Silvi, D.L. Lauvergnat, Ph.C. Hiberty, *Chem. Eur. J.*, 11(2005)6358.

[60] E. Runge, E.K.U. Gross, *Phys. Rev. Letters*, 52(1984)997; M.E. Casida, "*Time Dependent Density Functional Response Theory for Molecules*," in *Recent Advances in Density Functional Methods, Part 1*, D.P. Chong (ed.), World Scientific, Singapore, 1995.

the time-dependent periodic perturbation $-\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} \exp(\pm i\omega t)$ of frequency ω, where $\hat{\boldsymbol{\mu}}$ denotes the dipole moment operator of the system and $\boldsymbol{\mathcal{E}}$ is the electric field amplitude. Such a theory is valid under the assumption that the perturbation is relatively small and the electronic states of the isolated molecule are still relevant. In view of that we consider only a linear response of the system to the perturbation. Let us focus on the dipole moment of the system as a function of ω. It turns out that at certain values $\omega = \omega_{0k} = \frac{E_k^{(0)} - E_0^{(0)}}{\hbar}$, for $E_k^{(0)}$ denoting the energy of the k-th excited state ($E_0^{(0)}$ is the ground-state energy), one has an abrupt change of the mean value of the dipole moment. In fact this means that for $\omega = \omega_{0k}$ the ω-dependent polarizability (the dipole moment change is proportional to the polarizability) goes to infinity, i.e., has a pole. By detecting these poles[61] we are able to calculate the excited states in the DFT within the accuracy of a few tenths of eV.

3.9 The hunted correlation dragon before our eyes

The DFT method has a long history behind it, which began with Thomas, Dirac, Fermi, etc. At the beginning the quantitative successes were quite modest (the electron gas theory, known as the Xα method). Real success came after a publication by Jan Andzelm and Erich Wimmer.[62] The DFT method, offering results at a correlated level for a wide spectrum of physical quantities, turned out to be roughly as inexpensive as the Hartree–Fock procedure – this is the most sensational feature of the method.

We have a beacon – exact electron density distribution of harmonium

Hohenberg and Kohn proved their famous theorem on the existence of the energy functional, but nobody was able to give the functional for any system. All the DFT efforts are directed towards elaborating such a potential, and the only criterion of whether a model is any good is comparison with experiment. However, it turned out that there is a system for which every detail of the DFT can be verified. Uniquely, the dragon may be driven out the hole and we may fearlessly and with impunity analyze all the details of its anatomy. The system is a bit

61 As a first guess may serve the orbital energy differences from the ground-state theory.

62 J. Andzelm, E. Wimmer, *J. Chem. Phys*., 96(1992)1280. Jan was my PhD student in the old days. In the paper by A. Scheiner, J. Baker, J. Andzelm, *J. Comp. Chem*., 18(1997)775, the reader will find an interesting comparison of the methods used. One of the advantages (or deficiencies) of the DFT methods is that they offer a wide variety of basis functions (in contrast to the *ab initio* methods, where Gaussian basis sets rule), recommended for some particular problems to be solved. For example, in electronics (Si, Ge) the plane wave $\exp(i\mathbf{kr})$ expansion is a preferred choice. However, these functions are not advised for catalysis phenomena with atoms that are rare on Earth. The Gaussian basis sets in the DFT had a temporary advantage (in the 1990s) over others, because the standard Gaussian programs offered analytically computed gradients (for optimization of the geometry). Now this is also offered by many DFT methodologies.

artificial; it is the harmonic helium atom (harmonium) discussed on p. V1-243, in which the two electrons attract the nucleus by a harmonic force, while repelling each other by Coulombic interaction. For some selected force constants k, e.g., for $k = \frac{1}{4}$, the Schrödinger equation *can be solved analytically*. The wave function is extremely simple (see p. 94). The electron density (normalized to 2) is computed as

$$\rho_0(\mathbf{r}) = 2N_0^2 \mathrm{e}^{-\frac{1}{2}r^2} \left\{ \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \left[\frac{7}{4} + \frac{1}{4}r^2 + \left(r + \frac{1}{r}\right) \mathrm{erf}\left(\frac{r}{\sqrt{2}}\right) \right] + \mathrm{e}^{-\frac{1}{2}r^2} \right\}, \tag{3.85}$$

where erf is the error function, $\mathrm{erf}\,(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp\left(-u^2\right) \mathrm{d}u$, and

$$N_0^2 = \frac{\pi^{\frac{3}{2}}}{(8 + 5\sqrt{\pi})}. \tag{3.86}$$

We should look at the $\rho_0(\mathbf{r})$ with great interest – it is a unique occasion, as it is probable you will never see again an *exact* result. The formula is not only exact, but on top of this it is simple. Kais et al. compare the exact results with two DFT methods: the BLYP (a version of B3LYP) and the Becke–Perdew (BP) method.[63]

Because of the factor $\exp(-1/2r^2)$ the density distribution ρ is concentrated on the nucleus.[64] The authors compare this density distribution with the corresponding Hartree–Fock density (appropriate for the potential used), and even with the density distribution related to the hydrogen-like atom (after neglecting $1/r_{12}$ in the Hamiltonian the wave function becomes an antisymmetrized product of the two hydrogen-like orbitals). In the latter case the electrons do not see each other[65] and the corresponding density distribution is too concentrated on the nucleus. As soon as the term $1/r_{12}$ is restored, the electrons immediately move apart and ρ on the nucleus decreases by about 30%. The second result is also interesting: the Hartree–Fock density is very close to ideal – it is almost the same curve.[66]

Total energy components

It turns out that in the case analyzed (and so far only in this case) we can calculate the *exact* total energy E (Eq. (3.19)), the "wonder " potential v_0 that in the Kohn–Sham model gives

63 The detailed references to these methods are given in S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, G.J. Laming, *J. Chem. Phys.*, 99(1993)417.

64 As it should be.

65 Even in the sense of the mean field (as it is in the Hartree–Fock method).

66 This is why the Hartree–Fock method is able to give 99.6% of the total energy. Nevertheless, in some cases this may not be a sufficient accuracy.

the exact density distribution ρ (Eq. (3.85)), and the exchange potential v_{x} and correlation v_{c} (Eqs. (3.72) and (3.73)).[67] Let us begin from the total energy.

In the second row of Table 3.1, labeled KS (for Kohn–Sham), the exact total energy is reported ($E[\rho_0] = 2.0000$ a.u.) and its components (bold figures) calculated according to Eqs. (3.12), (3.19), (3.20), (3.21), (3.70), and (3.74). The exact correlation energy E_{c} is calculated as the difference between the exact total energy and the listed components. Thus, $T_0[\rho_0]$ stands for the kinetic energy of the noninteracting electrons, $\int v\rho_0 \mathbf{dr}$ means the electron–nucleus attraction (positive, because the harmonic potential is positive), and $J[\rho_0]$ represents the self-interaction energy of ρ_0. According to Eq. (3.21) and taking into account ρ_0, i.e., twice a square of the orbital, we obtain $J[\rho_0] = 2\mathcal{J}_{11}$ with the Coulombic integral $\mathcal{J}_{11}$. However, the exchange energy is given by Eq. (3.74), $E_x = -\frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} K_{ij}$, and after summing over the spin coordinates we obtain the exchange energy $E_{\mathrm{x}} = -\mathcal{K}_{11} = -\mathcal{J}_{11}$. We see such a relation between J and E_{x} in the second row (KS[68]). The other rows report various approximations computed by HF, BLYP, and BP, each of them giving its own Kohn–Sham spin orbitals and its own approximation of the density distribution ρ_0. This density distribution was used for the calculation of the components of the total energy within each approximate method. Of course, the Hartree–Fock method (third row) gave 0 for the correlation energy, because there is no correlation in it except that which follows from the Pauli exclusion principle taken into account in the exchange energy (cf. Chapter 2).

Table 3.1. Harmonium (harmonic helium atom). Comparison of the components (a.u.) of the total energy $E[\rho_0]$ calculated by the HF, BLYP, and BP methods with the exact values (row KS, exact Kohn–Sham solution).

	$E[\rho_0]$	$T_0[\rho_0]$	$\int v\rho_0 \mathbf{dr}$	$J[\rho_0]$	$E_{\mathrm{x}}[\rho_0]$	$E_{\mathrm{c}}[\rho_0]$
KS	**2.0000**	**0.6352**	**0.8881**	**1.032**	**−0.5160**	**−0.0393**
HF	2.0392	0.6318	0.8925	1.030	−0.5150	0
BLYP	2.0172	0.6313	0.8933	1.029	−0.5016	−0.0351
BP	1.9985	0.6327	0.8926	1.028	−0.5012	−0.0538

It is remarkable that all the methods are doing quite well. The BLYP method gives the total energy with an error of 0.86% – twice as small as the Hartree–Fock method – while the BP functional missed by as little as 0.08%. The total energy components are a bit worse, which proves that a certain cancellation of errors occurs among the energy components, which improves the value of the total energy. The KS kinetic energy T_0 amounts to 0.6352, while that

[67] These potentials as functions of ρ or r.

[68] Only for spin-compensated two-electron systems we have $E_x[\rho] = -\frac{1}{2}J[\rho_0]$ and, therefore, $v_x = \frac{\delta E_x}{\delta \rho}$ can be calculated analytically. In all other cases, although E_x can be easily evaluated (knowing orbitals), the calculation of v_x is very difficult and costly (it can only be done numerically). In the present two-electron case, v_x^{HF} is a multiplicative operator rather than an integral operator.

calculated as the mean value of the kinetic energy operator (of two electrons) is 0.6644, a bit larger – the rest is in the exchange-correlation energy.[69]

Exact "wonder" v_0 potential – the correlation dragon is finally caught

Fig. 3.10 shows a unique thing, our "wonder" long-awaited, exact potential v_0 as a function of r, and alternatively as a function of $\rho^{\frac{1}{3}}$. We look at it with extreme curiosity. The exact $v_0(r)$ represents a monotonic function increasing with r and represents a modification (influence of the second electron) of the external potential v; we see that v_0 is shifted *upwards* with respect to v, because the electron repulsion is effectively included. As we can see, the best approximate potential is the Hartree–Fock one.

Exchange potential

As to the exchange potential v_x (Fig. 3.11, it has to be negative and indeed it is), how are the BLYP and BP exchange potentials doing? Their plots are very close to each other and go almost parallel to the exact exchange potential for most values of r, i.e., they are very good (any additive constant in any potential energy does not count). For small r, both DFT potentials undergo some strange vibration. This region (high density) is surely the most difficult to describe, and no wonder that simple formulae cannot accurately describe the exact electronic density distribution (Fig. 3.11).

Measuring the correlation dragon: it is a small beast!

The correlation potential v_c turns out (Fig. 3.12) to correspond to forces 10–20 *times smaller* than those typical for exchange potential v_x (just look at the corresponding slopes). This is an important message, because, as the reader may remember, at the very end we tried to push the dragon into the correlation hole v_c and, as we see now, we have succeeded. The dragon of the correlation energy turned out to be quite a small beast!

The exact correlation potential represents a smooth "hook-like" curve. The BLYP and BP correlation plots twine loosely like eels round about the exact curve,[70] and for small r exhibit some vibration similar to that for v_x. It is most impressive that the BLYP and BP curves twine as if they were in counterphase, which suggests that, if added, they might produce good[71] results.

69 As we have described before.

70 The deviations are very large.

71 Such temptations give birth to Babylon-type science.

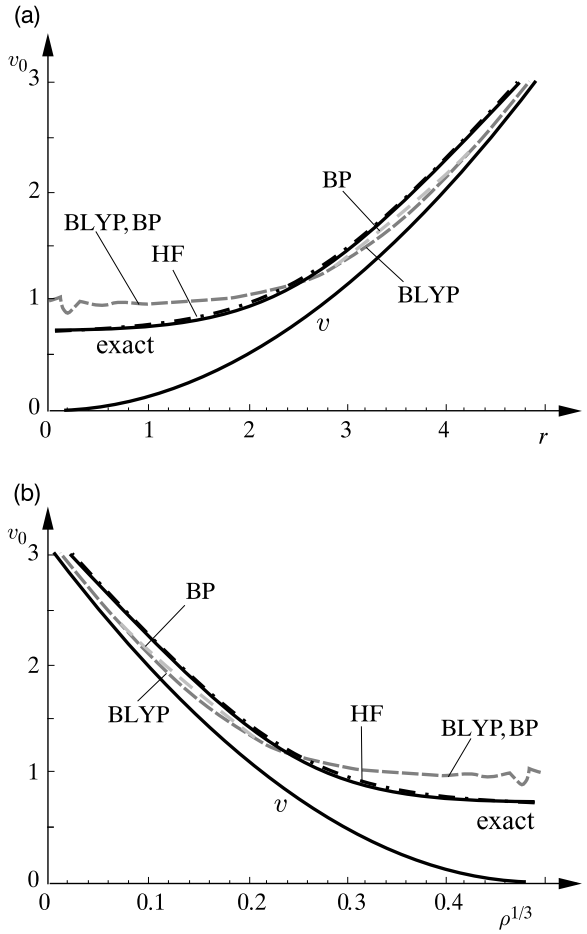


Fig. 3.10. Efficiency analysis of various DFT methods and comparison with the exact theory for the harmonium (with force constant $k = \frac{1}{4}$) according to Kais et al. (a) One-electron effective potential $v_0 = v + v_{\text{coul}} + v_{\text{xc}}$, with external potential $v = \frac{1}{2}kr^2$. (b) The same quantities as functions of $\rho^{\frac{1}{3}}$. The solid line corresponds to the exact results, the symbol HF pertains to the Fock potential (for the harmonic helium atom, $\cdot - \ \cdot -$), the symbols BLYP ($- - -$) and BP ($= = =$) stand for two popular DFT methods. With permission from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray and G.J. Lamming, *J. Chem. Phys.* 99(1993), American Institute of Physics.

Conclusion

The harmonic helium atom represents an instructive example that pertains to medium electronic densities. It seems that the dragon of the correlation energy does not have a hundred heads and is of quite good character, although a bit unpredictable.

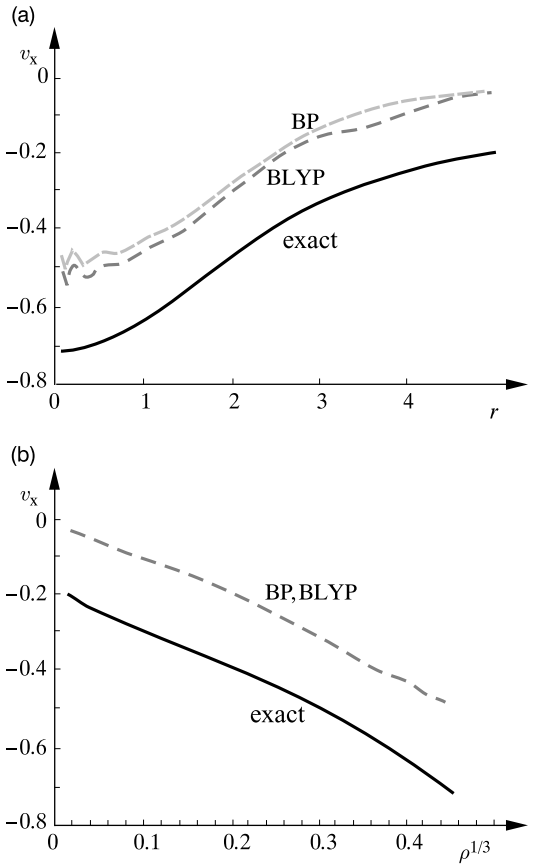


Fig. 3.11. Exchange potential. Efficiency analysis of various DFT methods and comparison with the exact theory for the harmonium (with the force constant $k = \frac{1}{4}$) according to Kais et al. (a) Exchange potential v_x as a function of the radius r. (b) A function of the density distribution ρ. The notation of Fig. 3.10 is used. It is seen that both DFT potentials produce plots that differ by nearly a constant from the exact potential (it is, therefore, an almost exact potential). The two DFT methods exhibit some nonphysical oscillations for small r. With permission from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray and G.J. Lamming, *J. Chem. Phys.* 99(1993), American Institute of Physics.

The results of various DFT versions are generally quite effective, although this comes from a cancellation of errors. Nevertheless, great progress has been made. At present many chemists prefer the DFT method (economy and accuracy) than getting stuck at the barrier of the configuration interaction excitations. And yet the method can hardly be called *ab initio*, since the exchange-correlation potential is tailored in a somewhat too practical manner.

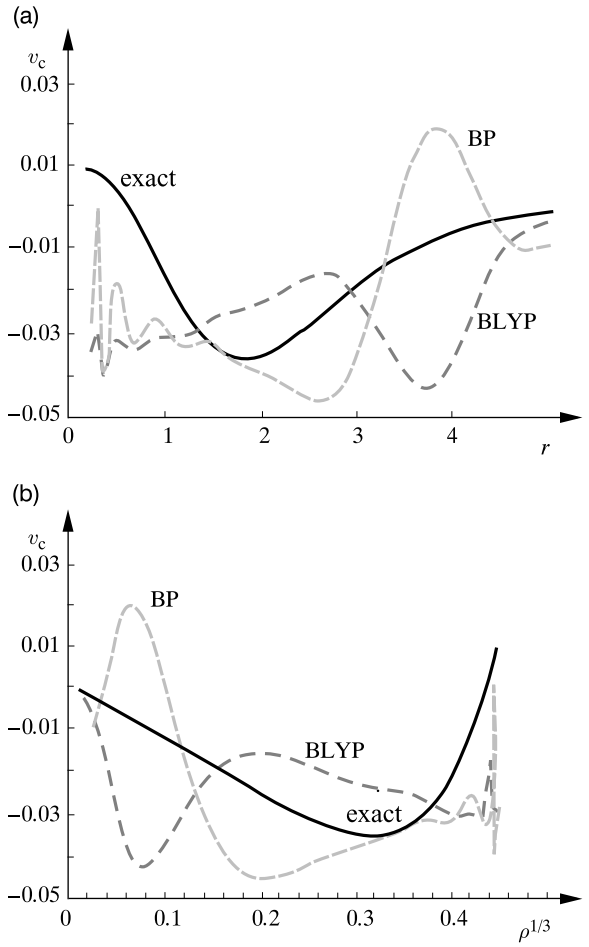


Fig. 3.12. The long-chased electron correlation dragon is finally found in its correlation hole and we have an exceptional occasion to see what it looks like. Correlation potential efficiency analysis of various DFT methods and comparison with the exact theory for the harmonic helium atom (with the force constant $k = \frac{1}{4}$) according to Kais et al. (a,b) Correlation potential v_c (less important than the exchange potential) as a function of (a) the radius r and (b) the density distribution ρ. Notation is as in Fig. 3.10. The DFT potentials produce plots that differ widely from the exact correlation potential. With permission from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray and G.J. Lamming, *J. Chem. Phys.* 99(1993), American Institute of Physics.

Summary

- The main theoretical concept of the DFT method is the electronic density distribution $\rho(\mathbf{r}) = N \sum_{\sigma_1=\frac{1}{2}}^{-\frac{1}{2}} \int \mathrm{d}\tau_2 \mathrm{d}\tau_3 \ldots \mathrm{d}\tau_N |\Psi(\mathbf{r}, \sigma_1, \mathbf{r}_2, \sigma_2, \ldots, \mathbf{r}_n, \sigma_N)|^2$, where $\mathbf{r}$ indicates a point in three-

dimensional space, and the summations are over all the spin coordinates of N electrons, while the integration is over the space coordinates of $(N-1)$ electrons. For example, within the molecular orbital (RHF) approximation $\rho = \sum_i 2\,|\varphi_i\,(\mathbf{r})|^2$ is the sum of the squares of all the molecular orbitals multiplied by their occupation number. The electronic density distribution ρ is a function of position in the three-dimensional space.

- Electronic density ρ carries a lot of information. It exhibits maxima at nuclei (with a discontinuity of the gradient, because of the cusp condition, p. 91). Bader analysis is based on identification of the critical (stationary) points of ρ (i.e., those for which $\nabla\rho = \mathbf{0}$), and for each of them the Hessian is computed (the second derivatives matrix). Diagonalization of the Hessian tells us whether the critical point corresponds to a maximum of ρ (nonnuclear attractor[72]), a minimum (e.g., cavities), a first-order saddle point (e.g., a ring center), or a second-order saddle point (chemical bond).
- The DFT relies on the two Hohenberg–Kohn theorems:
  - The *ground-state electronic density distribution* (ρ_0) *contains the same information as the ground-state wave function* (Ψ_0). Therefore, instead of a complex mathematical object (the ground-state wave function Ψ_0) depending on $4N$ variables we have a much simpler object (ρ_0) that depends on three variables (Cartesian coordinates) only.
  - A total energy functional of ρ *exists* that attains its minimum (the exact ground-state energy) at $\rho = \rho_0$. This mysterious functional is not yet known.
- Kohn and Sham presented the concept of a system with *noninteracting electrons,* subject however to some "wonder" external field $v_0\,(\mathbf{r})$ (instead of that of the nuclei), such that the resulting *density* ρ *remains identical to the exact ground-state density distribution* ρ_0. This *fictitious* system of electrons plays a very important role in DFT.
- Since the Kohn–Sham electrons do not interact, its wave function represents a single Slater determinant (called the Kohn–Sham determinant).
- We write the total energy expression as $E = T_0 + \int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r} + J[\rho] + E_{\mathrm{xc}}[\rho]$, which contains:
  - the kinetic energy of the *noninteracting* electrons (T_0),
  - the potential energy of the electron–nuclei interaction ($\int v(\mathbf{r})\rho(\mathbf{r})\mathrm{d}\mathbf{r}$),
  - the Coulombic electron–electron self-interaction energy ($J[\rho]$),
  - the remainder E_{xc}, i.e., the unknown exchange-correlation energy to be found.
- Using the single-determinant Kohn–Sham wave function (which gives the exact ρ_0) we vary the Kohn–Sham spin orbitals in order to find the minimum of the energy E.
- We are immediately confronted with the problem of how to find the unknown exchange-correlation energy E_{xc}, which is replaced by also an unknown exchange-correlation potential in the form of a functional derivative $v_{\mathrm{xc}} \equiv \frac{\delta E_{\mathrm{xc}}}{\delta\rho}$. We obtain the Kohn–Sham equation (resembling the Fock equation) $\{-\frac{1}{2}\Delta + v_0\}\phi_i = \varepsilon_i\phi_i$, with the "wonder potential" $v_0 = v + v_{\mathrm{coul}} + v_{\mathrm{xc}}$, where v_{coul} stands for the sum of the usual Coulombic operators (as in the Hartree–Fock method,[73] but built from the Kohn–Sham spin orbitals) and v_{xc} is an exchange-correlation potential to be found.
- The main problem now resides in the nature of E_{xc} (and v_{xc}). A variety of practical guesses we are forced to make begin here.
- The simplest guess is the local density approximation (LDA). We assume that E_{xc} can be summed up from the contributions of all the points in space, and that the individual contribution depends only on ρ computed at this point. Now, the key question is, *what does this dependence* $E_{\mathrm{xc}}[\rho]$ *look like*? LDA

[72] The maxima on the nuclei are excluded from the analysis, because of the discontinuity of $\nabla\rho$ mentioned above.

[73] It is, in fact, $\frac{\delta J[\rho]}{\delta\rho}$.

answers this question by using the following approximation: each point $\mathbf{r}$ in the three-dimensional space contributes to E_{xc} depending on the computed value of $\rho(\mathbf{r})$ *as if it were a homogeneous gas of uniform density* ρ, where the dependence $E_{\mathrm{xc}}[\rho]$ is *exactly* known.

- There are also more complex $E_{\mathrm{xc}}[\rho]$ functionals that go beyond the local approximation. They not only use the local value of ρ but sometimes also $\nabla\rho$ (gradient approximation).
- In each of these choices there is a lot of ambiguity. This, however, is restricted by some physical requirements.
- The requirements are related to the *electron pair distribution function* $\Pi(\mathbf{r}_1, \mathbf{r}_2) = N(N-1)\sum_{\mathrm{all}\ \sigma_i}\int|\Psi|^2 \mathrm{d}\mathbf{r}_3\mathrm{d}\mathbf{r}_4\ldots\mathrm{d}\mathbf{r}_N$, which takes into account the fact that the two electrons, shown by $\mathbf{r}_1$ and $\mathbf{r}_2$, avoid each other.
- First-order perturbation theory leads to the following exact expression for the total energy: $E = T_0 + \int\rho(\mathbf{r})v(\mathbf{r})\mathrm{d}\mathbf{r} + \frac{1}{2}\int\int\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\frac{\Pi\mathrm{aver}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}$, where $\Pi_{\mathrm{aver}}(\mathbf{r}_1, \mathbf{r}_2) = \int_0^1\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\lambda$, with the parameter $0 \le \lambda \le 1$ instrumental when transforming the system of *noninteracting* electrons ($\lambda = 0$, Kohn–Sham model) into the system of *fully interacting* ones ($\lambda = 1$) *all the time preserving the exact density distribution* ρ. Unfortunately, the function $\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)$ remains unknown.
- The function $\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)$ serves to define the electron hole functions, which will tell us where electron 2 prefers to be if electron 1 occupies the position $\mathbf{r}_1$. The exchange-correlation energy is related to the $\Pi^{\sigma\sigma'}_{\mathrm{aver}}$ function by $E_{\mathrm{xc}} = \frac{1}{2}\sum_{\sigma\sigma'}\int\int\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\frac{\Pi^{\sigma\sigma'}_{\mathrm{aver}}(\mathbf{r}_1,\mathbf{r}_2)-\rho_\sigma(\mathbf{r}_1)\rho_{\sigma'}(\mathbf{r}_2)}{r_{12}}$, where the sum is over the spin coordinate σ of electron 1 and spin coordinate σ' of electron 2, with the decomposition $\Pi_{\mathrm{aver}} = \Pi^{\alpha\alpha}_{\mathrm{aver}} + \Pi^{\alpha\beta}_{\mathrm{aver}} + \Pi^{\beta\alpha}_{\mathrm{aver}} + \Pi^{\beta\beta}_{\mathrm{aver}}$. For example, the number $\Pi^{\alpha\beta}_{\mathrm{aver}}\mathrm{d}V_1\mathrm{d}V_2$ is proportional to the probability of finding simultaneously an electron with the spin function α in the volume $\mathrm{d}V_1$ located at $\mathbf{r}_1$ and another electron with the spin function β in the volume $\mathrm{d}V_2$ located at $\mathbf{r}_2$, etc.
- The definition of the exchange-correlation hole function $h^{\sigma\sigma'}_{\mathrm{xc}}(\mathbf{r}_1, \mathbf{r}_2)$ is as follows: $E_{\mathrm{xc}} = \frac{1}{2}\sum_{\sigma\sigma'}\int\mathrm{d}\mathbf{r}_1\int\mathrm{d}\mathbf{r}_2\frac{\rho_\sigma(\mathbf{r}_1)}{r_{12}}h^{\sigma\sigma'}_{\mathrm{xc}}(\mathbf{r}_1, \mathbf{r}_2)$, which is equivalent to setting $h^{\sigma\sigma'}_{\mathrm{xc}}(\mathbf{r}_1,\ \mathbf{r}_2) = \frac{\Pi^{\sigma\sigma'}_{\mathrm{aver}}(\mathbf{r}_1,\mathbf{r}_2)}{\rho_\sigma(\mathbf{r}_1)} - \rho_{\sigma'}(\mathbf{r}_2)$. This means that the hole function is related to that part of the pair distribution function that indicates the *avoidance of the two electrons* (i.e., beyond their independent motion described by the *product* of the densities $\rho_\sigma(\mathbf{r}_1)\rho_{\sigma'}(\mathbf{r}_2)$).
- Due to the antisymmetry requirement for the wave function (Chapter V1-1) the holes have to satisfy some general (integral) conditions. The electrons with parallel spins have to avoid each other, i.e., $\int h^{\alpha\alpha}_{\mathrm{xc}}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = \int h^{\beta\beta}_{\mathrm{xc}}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = -1$ (one electron disappears from the neighborhood of the other), while the electrons with opposite spins are not influenced by the Pauli exclusion principle, i.e., $\int h^{\alpha\beta}_{\mathrm{xc}}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = \int h^{\beta\alpha}_{\mathrm{xc}}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = 0$.
- The exchange-correlation hole is a sum of the exchange hole and the correlation hole, $h^{\sigma\sigma'}_{\mathrm{xc}} = h^{\sigma\sigma'}_{\mathrm{x}} + h^{\sigma\sigma'}_{\mathrm{c}}$, where the exchange hole follows in a simple way from the Kohn–Sham determinant (and is therefore supposed to be known). Then, we have to guess the correlation holes. *All the correlation holes* have to satisfy the condition $\int h^{\sigma\sigma'}_{\mathrm{c}}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = 0$, which only means that the average has to be zero, but says nothing about the particular form of $h^{\sigma\sigma'}_{\mathrm{c}}(\mathbf{r}_1, \mathbf{r}_2)$. The only thing sure is, e.g., that close to the origin the function $h^{\sigma\sigma'}_{\mathrm{c}}$ has to be negative, and, therefore, for longer distances it has to be positive.
- Popular approximations, e.g., LDA, PW91, in general satisfy the integral conditions for the holes.
- Hybrid approximations (e.g., B3LYP), i.e., such linear combinations of the potentials that will ensure good agreement with experiment, become more and more popular.

- The DFT models can be tested when applied to exactly solvable problems with electronic correlation (like harmonium, Chapter V1-4). It turns out that despite the exchange, and especially correlation, DFT potentials deviate from the exact ones, the total energy is quite accurate.
- There is a possibility in DFT to calculate the excitation energies. This is possible within the time-dependent DFT. In this formulation one is looking at the frequency-dependent polarizabilities in the system subject to the electric field perturbation of frequency ω. The polarizabilities have poles at $\hbar\omega$ equal to an excitation energy.

Main concepts, new terms

adiabatic connection (p. 222)
attractor (p. 200)
Bader analysis (p. 197)
basin (p. 200)
B3LYP functional (p. 220)
catastrophe set (p. 201)
correlation hole (p. 229)
critical (stationary) points (p. 197)
charge-shift (CS) bond (p. 238)
density matrix (p. 194)
electron gas (p. 191)
electronic density (p. 194)
electron pair distribution (p. 221)
electron localization function (ELF) (p. 233)
exchange-correlation energy (p. 213)
exchange-correlation hole (p. 227)
exchange-correlation potential (p. 217)
exchange hole (p. 229)
Fermi hole (p. 231)
gradient approximation, NLDA (GEA) (p. 219)
Hohenberg–Kohn functional (p. 207)
Hohenberg–Kohn theorem (p. 204)
holes (p. 231)
hybrid approximations, NLDA (p. 220)
Kohn–Sham equation (p. 211)
Kohn–Sham system (p. 211)
Levy minimization (p. 207)
local density approximation (LDA) (p. 218)
nonnuclear attractor (p. 200)
one-particle density matrix (p. 249)
Perdew–Wang (p. 220)
self-interaction energy (p. 212)
spin polarization (p. 218)
time-dependent DFT (p. 238)
v-representability (p. 206)

From the research front

Computer technology has been revolutionary, not only because computers are fast. Much more important is that each programmer uses the full experience of his predecessors and easily "stands on the shoulders of giants." The computer era has made an unprecedented transfer of the most advanced theoretical tools from the finest scientists to practically everybody. Experimentalists often investigate large molecules. If there is a method like DFT, which gives answers to their vital questions in a shorter time than the *ab initio* methods, they will not hesitate and choose the DFT, even if the method is of semiempirical type. Something like this happens now. Nowadays the DFT procedure is applicable to systems with hundreds of atoms.

The DFT method is developing fast also in the conceptual sense,[74] e.g., the theory of reactivity ("charge sensitivity analysis"[75]) has been derived, which established a link between the intermolecular electron transfer and the charge density changes in atomic resolution. For systems in magnetic fields, current DFT was developed.[76] Relativistic effects[77] and time-dependent phenomena[78] are included in some versions of the theory.

Ad futurum

The DFT will be, of course, further elaborated. There are already investigations under way, which will allow us to calculate the dispersion energy.[79] The impetus will probably be directed towards such methods as the density matrix functional theory (DMFT) proposed by Levy,[80] and currently being developed by Jerzy Ciosłowski.[81] The idea is to abandon $\rho(\mathbf{r})$ as the central quantity, and instead use the one-particle density matrix $\rho(\mathbf{r}'; \mathbf{r})$ of Eqs. (3.1) and (3.2).

This method has the advantage that we are not forced to introduce the noninteracting Kohn–Sham electrons, because the mean value of the electron kinetic energy may be expressed directly by the new quantity (this follows from the definition)

$$T = -\frac{1}{2}\int \mathrm{d}\mathbf{r} \quad [\Delta_{\mathbf{r}}\rho(\mathbf{r}; \mathbf{r}')]|_{\mathbf{r}'=\mathbf{r}},$$

where the symbol $|_{\mathbf{r}'=\mathbf{r}}$ means replacing $\mathbf{r}'$ by $\mathbf{r}$ when the result $\Delta_{\mathbf{r}}\rho(\mathbf{r}; \mathbf{r}')$ is ready. Thus, in the DMFT exchange-correlation we have no kinetic energy left.

The success of the DFT approach will probably make the traditional *ab initio* procedures faster, up to the development of methods with linear scaling (with the number of electrons for long molecules). The massively parallel "computer farms" with 2000 processors currently, expected to reach millions soon, will saturate most demands of experimental chemistry. The results will be calculated fast, and it will be much more difficult to define an interesting target to compute. We will be efficient.

---

74 See, e.g., P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.*, 103(2003)1793.

75 R.F. Nalewajski, J. Korchowiec, "*Charge Sensitivity Approach to Electronic Structure and Chemical Reactivity*," World Scientific, Singapore, 1997; R.F. Nalewajski, J. Korchowiec, A. Michalak, "*Reactivity Criteria in Charge Sensitivity Analysis*," *Topics in Current Chemistry*, 183(1996)25; R.F. Nalewajski, "*Charge Sensitivities of Molecules and Their Fragments*, *Rev. Mod. Quant. Chem*., ed. K.D. Sen, World Scientific, Singapore (2002)1071; R.F. Nalewajski, R.G. Parr, *Proc. Natl. Acad. Sci. USA*, 97(2000)8879.

76 G. Vignale, M. Rasolt, *Phys. Rev. Letters*, 59(1987)2360, *Phys. Rev. B*, 37(1988)10685.

77 A.K. Rajagopal, J. Callaway, *Phys. Rev. B*, 7(1973)1912; A.H. MacDonald, S.H. Vosko, *J. Phys. C*, 12(1979)2977.

78 E. Runge, E.K.U. Gross, *Phys. Rev. Letters*, 52(1984)997; R. van Leeuwen, *Phys. Rev. Letters*, 82(1999)3863.

79 W. Kohn, Y. Meir, D. Makarov, *Phys. Rev. Letters*, 80(1998)4153; E. Hult, H. Rydberg, B.I. Lundqvist, D.C. Langreth, *Phys. Rev. B*, 59(1999)4708; J. Ciosłowski, K. Pernal, *J. Chem. Phys.*, 116(2002)4802.

80 M. Levy, *Proc. Nat. Acad. Sci. (USA)*, 76(1979)6062.

81 J. Ciosłowski, K. Pernal, *J. Chem. Phys.*, 111(1999)3396; J. Ciosłowski, K. Pernal, *Phys. Rev. A*, 61(2000)34503; J. Ciosłowski, P. Ziesche, K. Pernal, *Phys. Rev. B*, 63(2001)205105; J. Ciosłowski, K. Pernal, *J. Chem. Phys*., 115(2001)5784; J. Ciosłowski, P. Ziesche, K. Pernal, *J. Chem. Phys.*, 115(2001)8725.

We will have an efficient hybrid potential, say, of the B3LYP5PW2001/2002-type. There remains, however, a problem that already appears in laboratories. A colleague delivers a lecture and proposes a hybrid B3LYP6PW2003update,[82] more effective for aromatic molecules. What will these two scientists talk about? It is very good that the computer understands all this, but what about the scientists? In my opinion science will move into such areas as planning new materials and new molecular phenomena (cf. Chapter 7) with the programs mentioned above as tools.

Additional literature

W. Koch, M.C. Holthausen, "A Chemist's Guide to Density Functional Theory," New York, Wiley-VCH, 2000.

Very good and competent book, trying to be as clear as possible. It contains the theory and, in the second half, a description of the DFT reliability when calculating various physical and chemical properties.

R.H. Dreizler, E.K.U. Gross, "Density Functional Theory," Springer, Berlin, 1990.

A rigorous book on DFT for specialists in the field.

R.G. Parr, W. Yang, "Density Functional Theory of Atoms and Molecules," Oxford Univ. Press, Oxford, 1989.

The classic textbook on DFT for chemists.

A.D. Becke, in "Modern Electronic Structure Theory. Part II," D.R. Yarkony (ed.), World Scientific, p. 1022.

An excellent and comprehensible introduction into DFT written by a renowned expert in the field.

J. Andzelm, E. Wimmer, *J. Chem. Phys.*, 96(1992)1280.

A competent presentation of DFT technique introduced by the authors.

Richard F.W. Bader, "Atoms in Molecules. A Quantum Theory," Clarendon Press, Oxford, 1994.

An excellent book.

[82] The same pertains to the traditional methods. Somebody operating billions of the expansion functions meets a colleague using even more functions. It would be a pity if we changed into experts ("*this is what we are paid for...*") knowing which particular BLYP is good for calculating interatomic distances, which for charge distribution, etc.

Questions

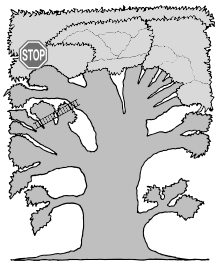
1. Consider Bader analysis. In the critical point of the charge density for a covalent bond:
   a. the value of the density is positive.
   b. the electronic density Hessian has precisely two negative eigenvalues.
   c. all three components of the electron density gradient are equal to zero.
   d. we are always in the middle of the distance between the two nuclei.
2. Consider Bader analysis. The electronic density Hessian calculated at the center of the benzene ring (of D_{6h} symmetry):
   a. has exactly one positive eigenvalue.
   b. has exactly two positive eigenvalues.
   c. all three components of the electron density gradient are equal to zero.
   d. the trace of the Hessian depends on the Cartesian coordinate system chosen.
3. Hohenberg and Kohn (ρ stands for the electron density, $\rho = \rho_0$ corresponds to the ideal ground-state electronic density, E_0 is the ground-state energy):
   a. gave the functional $E^{HK}[\rho]$ exhibiting a minimum that corresponds to the density ρ_0.
   b. have proved that having ρ one can obtain the ground-state wave function.
   c. have proved that from E_0 one can obtain ρ_0.
   d. have proved that there exists an energy functional $E^{HK}[\rho] \geq E^{HK}[\rho_0] = E_0$.
4. The Kohn–Sham system of electrons (ρ stands for the electron density, $\rho = \rho_0$ corresponds to the ideal ground-state electronic density, E_0 is the ground-state energy):
   a. represents N electrons leading to the Hartree–Fock electronic density.
   b. represents a system of N noninteracting electrons that give the same electronic density $\rho = \rho_0$ as the electronic density of the fully interacting system.
   c. is described by N spin orbitals, each being a solution of a one-electron equation.
   d. leads to the Slater determinant corresponding to the electronic density ρ_0.
5. In the LDA approximation (E_{xc} stands for the exchange-correlation energy):
   a. the uniform electron gas represents a system of N electrons in a box of volume V with the periodic boundary conditions.
   b. the uniform electron gas represents a system of N electrons in a box of volume V with the boundary condition of the wave function vanishing at the border of the box.
   c. the uniform electron gas represents a system of N electrons in a box of volume V with the periodic boundary conditions and the uniform distribution of the nuclear matter (to get the electrically neutral system).
   d. $E_{xc}[\rho]$ for molecules is such a functional of ρ that for a fixed ρ the value of E_{xc} is equal to the known value E_{xc}^{gas} for the uniform electron gas corresponding to the same density.
6. The DFT hybrid approximation (E_{xc} stands for the exchange-correlation energy):
   a. means using a linear combination of atomic hybrid orbitals in the expansion of the Kohn–Sham molecular orbitals.
   b. the B3LYP method belongs to the hybrid approximations.
   c. the hybrid approximations represent in fact some semiempirical methods.
   d. one uses as E_{xc} a linear combination of E_{xc} expressions stemming from several DFT functionals and from the Hartree–Fock method.
7. The exchange-correlation energy E_{xc} in the Kohn–Sham method:
   a. contains a part of the electron kinetic energy.
   b. effectively takes into account the Coulomb hole and the Fermi hole.

c. depends on a particular DFT functional.
d. is equal to zero.

8. The exchange-correlation hole functions satisfy:
   a. $\int h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = 0$ and $\int h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = -1$.
   b. $\int h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = -1$ and $\int h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = -1$.
   c. $\int h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\mathrm{d}\mathbf{r}_2 = 0$ and $h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}, \mathbf{r}) = -\rho_\beta(\mathbf{r})$.
   d. $h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2) = 0$ and $h_{\mathrm{xc}}^{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2) = -1$.
9. The DFT exchange energy E_x:
   a. is more important than the correlation energy.
   b. $E_x < 0$.
   c. is calculated using the exchange Hartree–Fock expression, but with the Kohn–Sham orbitals.
   d. must be a repulsion for He$\cdots$He and an attraction for H$\cdots$H.
10. The Kohn–Sham DFT method:
   a. is able to describe the interaction of the two argon atoms.
   b. is as time consuming as the Hartree–Fock method.
   c. does not take into account the electron correlation, because it uses a one-determinantal wave function.
   d. which would contain a correct exchange-correlation potential, would describe the dispersion interaction.

Answers

1a,b,c, 2b,c, 3b,d, 4b,c,d, 5c,d, 6b,c,d, 7a,b,c, 8a,c, 9a,b,c, 10b,d



CHAPTER 4

The Molecule Subject to Electric or Magnetic Fields

For the time being I was not aware of, but soon I have experienced by myself, how dangerous for our ship approaching the Magnetic Mountain was. (...) Not only the anchor, the iron trunks, the knives, spoons and other objects, but also the nails, that have been used to join the boards of the ship, jumped off just by themselves (...)

Bolesław Leśmian "Adventures of the Sailor Sindbad"

Where are we?

We are already in the crown of the TREE (left-hand side).

An example

How does a molecule react to an applied electric field? How to calculate the changes it undergoes? In some materials a strange phenomenon takes place: a monochromatic *red* laser light beam enters a transparent substance, and leaves the specimen as a *blue* beam. Why?

Another example, this time with a magnetic field. We apply some long-wave length electromagnetic field to a specimen. We do not see any absorption whatsoever. However, if, in addition, we apply a static magnetic field, gradually increasing in intensity, at some intensities we observe absorption. If we analyze the magnetic field values corresponding to the absorption, they cluster into some mysterious groups that depend on the chemical composition of the specimen. Why?

What is it all about?

The properties of a substance with and without an external electric field *differ*. The problem is how to compute the molecular properties in the electric field from the properties of the isolated molecule and the characteristics of the applied field. Molecules react also upon application of a magnetic field, which changes the internal electric currents and modifies the local magnetic field. A nucleus may be treated as a small magnet, which reacts to the local magnetic field it encounters. This local field depends not

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00012-0
Copyright © 2020 Elsevier B.V. All rights reserved.

only on the external magnetic field, but also on those from other nuclei, and on the electronic structure in the vicinity. This produces some energy levels in the spin system, with transitions leading to the nuclear magnetic resonance (NMR) phenomenon which has wide applications in chemistry, physics, and medicine.

The following topics will be described in the present chapter.

Why is this important?

There is no such a thing as an isolated molecule, since any molecule interacts with its neighborhood. In most cases this is the electric field of other molecules, which represents the only information about the external world the molecule has. The source of the electric field (another molecule or technical equipment) is of no importance. *Any molecule will respond to the electric field, but some will respond dramatically, while others may respond quite weakly.* The way they respond is of prime importance in technical and scientific applications.

The molecular electronic structure does not respond to a change in orientation of the nuclear magnetic moments, because the corresponding perturbation is too small. On the other hand, the molecular electronic structure influences the subtle energetics of interaction of the nuclear spin magnetic moments and these effects may be recorded in the NMR spectrum. This is of great practical importance, because it means *we have in the molecule under study a system of sounds (nuclear spins) which characterize the electronic structure practically without perturbing it*.

What is needed?

- Perturbation theory (Chapter V1-5, necessary),
- variational method (Chapter V1-5, advised),
- harmonic oscillator and rigid rotator (Chapter V1-4, advised),
- Breit Hamiltonian (Chapter V1-3, advised),
- Appendix D, p. 595 (advised),
- Appendix V1-G, p. V1-673 (necessary for magnetic properties),
- Appendix V1-N, p. V1-707 (advised),
- Appendix H, p. 627 (advised).

Classical works

John Hasbrouck Van Vleck (1899–1980), American physicist, professor at the University of Minnesota, received the Nobel Prize in 1977 for "*fundamental theoretical investigations of the electronic structure of magnetic and disordered systems*."

Peter Debye, as early as 1921, predicted in "*Molekularkräfte und ihre Elektrische Deutung*," *Physikalische Zeitschrift,* 22(1921)302, that a nonpolar gas or liquid of molecules with a nonzero quadrupole moment, when subject to a nonhomogeneous electric field, will exhibit the birefringence phenomenon due to the orientation of the quadrupoles in the electric field gradient. ★ The book by John Hasbrouck Van Vleck *Electric and Magnetic Susceptibilities*," Oxford University Press, 1932 represented enormous progress. ★ The theorem that forces acting on nuclei result from classical interactions with electron density (computed by a quantum mechanical method) was first proved by Hans Gustav Adolf Hellmann in the world's first textbook of quantum chemistry, "*Einführung in die Quantenchemie*," Deuticke, Leipzig and Vienna,[1] 1937, p. 285, and then, independently, by Richard Philips Feynman in "*Forces in Molecules*," published in *Physical Review*, 56(1939)340. ★ The first idea of nuclear magnetic resonance (NMR) came from a Dutch scholar, Cornelis Jacobus Gorter, in "*Negative Result in an Attempt to Detect Nuclear Spins*," in *Physica*, 3(1936)995. ★ The first *electron* paramagnetic resonance (EPR) measurement was carried out by Evgeniy Zavoiski from Kazan University (USSR) and reported in "*Spin-Magnetic Resonance in Paramagnetics*," published in *Journal of Physics (USSR)*, 9(1945)245, 447. ★ The first NMR absorption experiment was performed by Edward M. Purcell, Henry C. Torrey, and Robert V. Pound and published in "*Resonance Absorption by Nuclear Magnetic Moments in a Solid*," which appeared in *Physical Review*, 69(1946)37, while the first correct explanation of nuclear spin–spin coupling (through the chemical bond) was given by Norman F. Ramsey and Edward M. Purcell in "*Interactions between Nuclear Spins in Molecules*," published in *Physical Review*, 85(1952)143. ★ The first successful experiment in nonlinear optics with frequency doubling was reported by Peter A. Franken, Alan E. Hill, Wilbur C. Peters, and Gabriel Weinreich in "*Generation of Optical Harmonics*," published in *Physical Review Letters*, 7(1961)118. ★ Hendrik F. Hameka's book "*Advanced Quantum Chemistry. Theory of Interactions between Molecules and Electromagnetic Fields*" (1965) is also considered a classic work.

* * *

4.1 Hellmann–Feynman theorem

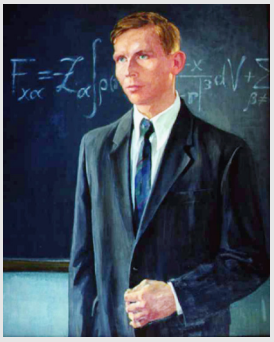
Let us assume that a system with Hamiltonian $\hat{H}$ is in a *stationary state* described by the (normalized) function ψ. Now let us begin to do a little "tinkering" with the Hamiltonian by introducing a parameter P. So we have $\hat{H}(P)$, and assume we may change the parameter

[1] A Russian edition had appeared a few months earlier, but it did not contain the theorem.

smoothly. For example, as the parameter P we may take the electric field intensity, or, if we assume the Born–Oppenheimer approximation, then as P we may take a nuclear coordinate.[2] If we change P in the Hamiltonian $\hat{H}(P)$, then its eigenfunctions and eigenvalues become functions of P.

The Hellmann–Feynman theorem pertains to the rate of the change[3] of $E(P)$:

Hans Gustav Adolf Hellmann (1903–1938), German physicist, one of the pioneers of quantum chemistry. He contributed to the theory of dielectric susceptibility, theory of spin, chemical bond theory (semiempirical calculations, also virial theorem and the role of kinetic energy), intermolecular interactions theory, electronic affinity, etc. Hellmann wrote the world's first textbook of quantum chemistry, "*Vviedieniye v kvantovuyu khimiyu*," a few months later edited in Leipzig as "*Einführung in die Quantenchemie*." In 1933 Hellmann presented his habilitation thesis at the Veterinary College of Hannover. As part of the paper work he filled out a form, in which according to the recent Nazi requirement he wrote that his wife was of Jewish origin. The Nazi ministry rejected the habilitation. The situation grew more and more dangerous (many students of the School were active Nazis) and the Hellmanns decided to emigrate. Since his wife originated from the Ukraine, they chose the Eastern route. Hellmann obtained a position at the Karpov Institute of Physical Chemistry in Moscow as a theoretical group leader. A leader of another group, the Communist Party First Secretary of the Institute (Hellmann's colleague and a coauthor of one of his papers) A.A. Zukhovitskiy, as well as the former First Secretary and leader of the Heterogenic Catalysis Group



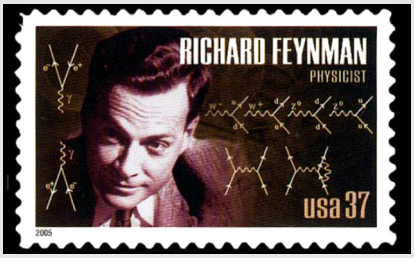
Mikhail Tiomkin, denounced Hellmann to the institution later called the KGB, which soon arrested him. Years later an investigation protocol was found in the KGB archives, with a text about Hellmann's spying written by somebody else but with Hellmann's signature. This was a common result of such "investigations." On May 16, 1938 Albert Einstein, and on May 18 three other Nobel Prize recipients, Irene Joliot-Curie, Frederick Joliot-Curie, and Jean-Baptiste Perrin, asked Stalin for mercy for Hellmann. Stalin ignored the eminent scholars' supplication, and on May 29, 1938 Hans Hellmann faced the firing squad and was executed. After W.H.E. Schwarz et al., *Bunsen-Magazin*, (1999)10, 60. Portrait reproduced from a painting by Tatiana Livschitz, courtesy of Professor Eugen Schwarz.

[2] Recall please that in the adiabatic approximation, the electronic Hamiltonian depends parametrically on the nuclear coordinates (Chapter V1-6). Then $E(P)$ corresponds to $E_k^0(R)$ from Eq. (V1-6.8).

[3] We may define $\left(\frac{\partial \hat{H}}{\partial P}\right)_{P=P_0}$ as an operator, being a limit when $P \to P_0$ of the operator sequence $\frac{\hat{H}(P)-\hat{H}(P_0)}{P-P_0}$.

Richard Philips Feynman (1919–1988), American physicist, for many years professor at the California Institute of Technology. His father was his first informal teacher of physics, who taught him the extremely important skill of independent thinking. Feynman studied at Massachusetts Institute of Technology, then in Princeton University, where he defended his PhD thesis under the supervision of John Archibald Wheeler. In 1945–1950 Feynman served as a professor at Cornell University. A paper plate thrown in the air by a student in the Cornell cafe was the first impulse for Feynman to think about creating a new version of quantum electrodynamics. For this achievement Feynman received the Nobel Prize in 1965 (cf. p. V1-17).

Feynman was a genius, who contributed to several branches of physics (superfluidity, weak interactions, quantum computers, nanotechnology). His textbook "*The Feynman Lectures on Physics*" is considered an unchallenged achievement in academic literature. Several of his books became best-sellers. Feynman was famous for his unconventional, straightforward, and crystal-clear thinking, as well as for his courage and humor. Curiosity and courage made possible his investigations of the ancient



Maya calendar and ant habits, as well as his activity in painting and music.

From John Slater's autobiography "*Solid State and Molecular Theory*," London, Wiley, (1975):

"*... The theorem known as the Hellmann–Feynman theorem, stating that the force on a nucleus can be rigorously calculated by electrostatics (...), remained, as far as I was concerned, only a surmise for several years. Somehow, I missed the fact that Hellmann, in Germany, proved it rigorously in* 1936*, and when a very bright undergraduate turned up in* 1938–1939 *wanting a topic for a bachelor's thesis, I suggested to him that he see if it could be proved. He came back very promptly with a proof. Since he was Richard Feynman (...), it is not surprising that he produced his proof without trouble*."

HELLMANN–FEYNMAN THEOREM

$$\frac{\partial E}{\partial P} = \langle\psi|\frac{\partial \hat{H}}{\partial P}|\psi\rangle. \tag{4.1}$$

The proof is simple. The differentiation with respect to P of the integrand in $E = \langle\psi|H|\psi\rangle$ gives

$$\begin{aligned}\frac{\partial E}{\partial P} &= \langle\frac{\partial \psi}{\partial P}|\hat{H}\psi\rangle + \langle\psi|\frac{\partial \hat{H}}{\partial P}\psi\rangle + \langle\psi|\hat{H}\frac{\partial \psi}{\partial P}\rangle \\ &= E\left(\langle\frac{\partial \psi}{\partial P}|\psi\rangle + \langle\psi|\frac{\partial \psi}{\partial P}\rangle\right) + \langle\psi|\frac{\partial \hat{H}}{\partial P}\psi\rangle = \langle\psi|\frac{\partial \hat{H}}{\partial P}\psi\rangle,\end{aligned} \tag{4.2}$$

because the expression in parentheses is equal to zero (we have profited from the fact that $\hat{H}$ is Hermitian and that ψ represents its eigenfunction[4]). Indeed, differentiating $\langle\psi|\psi\rangle = 1$ we have

$$0 = \langle\frac{\partial\psi}{\partial P}|\psi\rangle + \langle\psi|\frac{\partial\psi}{\partial P}\rangle, \tag{4.3}$$

which completes the proof.

Soon we will use the Hellmann–Feynman theorem to compute the molecular response to an electric field.[5]

---

[4] If, instead of the exact eigenfunction, we use an approximate function ψ, then the theorem would have to be modified. In such a case we have to take into account the terms $\langle\frac{\partial\psi}{\partial P}|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\frac{\partial\psi}{\partial P}\rangle$.

[5] In case P is a nuclear coordinate (say, x coordinate of nucleus C, denoted by X_C) and E stands for the potential energy for the motion of the nuclei (cf. Chapter V1-6, the quantity corresponds to E_0^0 of Eq. (V1-6.8)), the quantity $-\frac{\partial E}{\partial P} = F_{X_C}$ represents the x component of the force acting on the nucleus. The Helmann–Feynman theorem says that this component can be computed as the mean value of the derivative of the Hamiltonian with respect to the parameter P. Since the electronic Hamiltonian reads

$$\hat{H}_0 = -\frac{1}{2}\sum_i \Delta_i + V,$$

$$V = -\sum_A\sum_i \frac{Z_A}{r_{Ai}} + \sum_{i<j}\frac{1}{r_{ij}} + \sum_{A<B}\frac{Z_A Z_B}{R_{AB}},$$

after differentiating, we have

$$\frac{\partial\hat{H}_0}{\partial X_C} = \frac{\partial V}{\partial X_C} = \sum_i \frac{Z_C}{(r_{Ci})^3}(X_C - x_i) - \sum_{B(\neq C)}\frac{Z_C Z_B}{(R_{BC})^3}(X_C - X_B).$$

Therefore,

$$F_{X_C} = -\langle\psi|\frac{\partial\hat{H}}{\partial P}|\psi\rangle = Z_C\left[\int d\mathbf{r}_1\rho(1)\frac{x_1 - X_C}{(r_{C1})^3} - \sum_{B(\neq C)}\frac{Z_B}{(R_{BC})^3}(X_B - X_C)\right],$$

where $\rho(1)$ stands for the electronic density defined in Chapter 3 (Eqs. (3.1) and (3.2)).
The last term can be easily calculated from the positions of the nuclei, the first term requires calculation of the one-electron integrals. Note that the resulting formula says that the forces acting on the nuclei follow from the *classical* Coulomb interaction involving the electronic density ρ, even if the electronic density has been (and has to be) computed from quantum mechanics.

ELECTRIC PHENOMENA

4.2 The molecule immobilized in an electric field

The electric field intensity $\boldsymbol{\mathcal{E}}$ at a point represents the force acting on a unit positive point charge (probe charge), $\boldsymbol{\mathcal{E}} = -\nabla V$, where V stands for the electric field potential energy at this point.[6] Imagine the potential decreases linearly in space (Fig. 4.1a), the electric field intensity is constant (Fig. 4.1b,c). If at such a potential we shift the probe charge from a to $a + x$ $(x > 0)$, then the potential energy will *lower* by $V(a + x) - V(a) = -\mathcal{E}x < 0$, similarly as the potential energy of a stone will be lower after sliding it downhill.

If, instead of a unit charge, we shift the charge Q, then the energy will be lower by $-\mathcal{E}Qx$. It is seen that if we change the *direction* of the shift or the *sign* of the probe charge, then the energy will go *up* (in the case of the stone we may change only the direction). Therefore,

> the change of the potential energy in a homogeneous electric field $\boldsymbol{\mathcal{E}}$ when shifting charge Q by vector $\mathbf{r}$ is equal to $\Delta E = -Q\boldsymbol{\mathcal{E}} \cdot \mathbf{r}$.

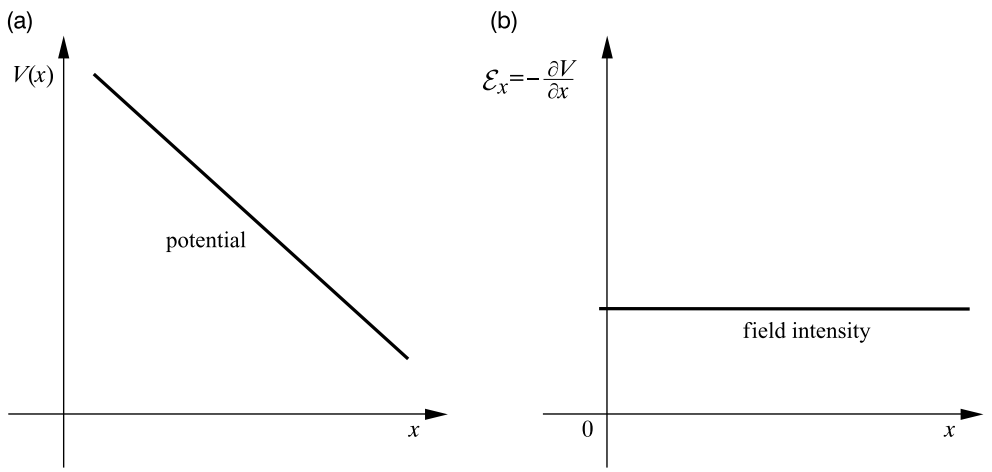


Fig. 4.1. Recalling the electric field properties. (a) The one-dimensional case: the potential V decreases with x. (b) This means that the electric field intensity $\boldsymbol{\mathcal{E}}$ is constant, i.e., the field is uniform. (c) The three-dimensional case. Uniform electric field $\boldsymbol{\mathcal{E}} = (\mathcal{E}, 0, 0)$. (d) Inhomogeneous electric field $\boldsymbol{\mathcal{E}} = (\mathcal{E}(x), 0, 0)$. (e) Inhomogeneous electric field $\boldsymbol{\mathcal{E}} = (\mathcal{E}_x(x, y), \mathcal{E}_y(x, y), 0)$.

[6] We see that two potential functions that differ by a constant will give the same forces, i.e., will describe identical physical phenomena (this is why this constant is arbitrary).

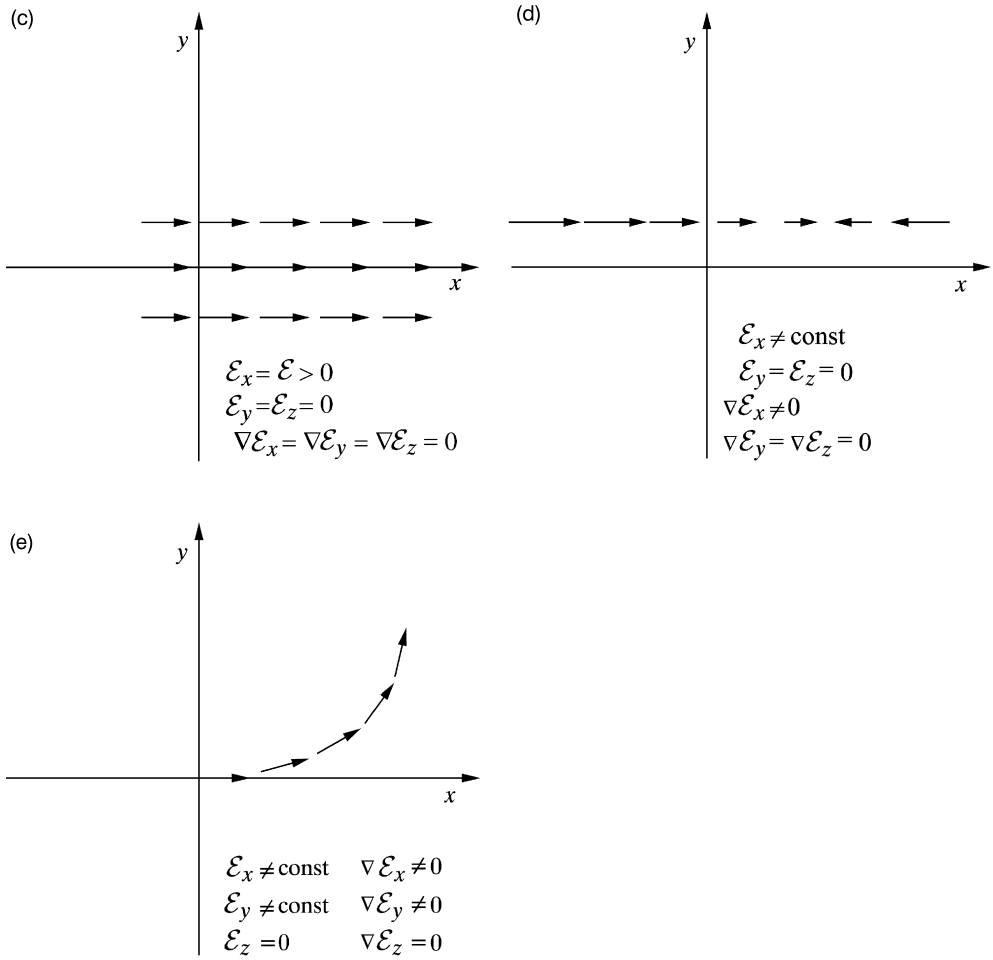


Fig. 4.1. (*continued*)

4.2.1 The electric field as a perturbation

The nonhomogeneous field at a slightly shifted point

Imagine a Cartesian coordinate system in three-dimensional space and a nonhomogeneous electric field (Fig. 4.1d,e) in it, $\boldsymbol{\mathcal{E}} = [\mathcal{E}_x(x, y, z), \mathcal{E}_y(x, y, z), \mathcal{E}_z(x, y, z)]$.

Assume the electric field vector $\boldsymbol{\mathcal{E}}(\mathbf{r}_0)$ is measured at a point indicated by the vector $\mathbf{r}_0$. What will we measure at a point shifted by a small vector $\mathbf{r} = (x, y, z)$ with respect to $\mathbf{r}_0$? The components of the electric field intensity represent smooth functions in space and this is why we may compute the electric field from the Taylor expansion (for each of the components $\mathcal{E}_x, \mathcal{E}_y, \mathcal{E}_z$

separately, all the derivatives are computed at point $\mathbf{r}_0$, see Fig. 4.2, indices $q, q', q'' = x, y, z$). We have

$$\mathcal{E}_x = + \left(\frac{\partial \mathcal{E}_x}{\partial x}\right)_0 x + \left(\frac{\partial \mathcal{E}_x}{\partial y}\right)_0 y + \left(\frac{\partial \mathcal{E}_x}{\partial z}\right)_0 z +$$
$$\frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial x^2}\right)_0 x^2 + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial x \partial y}\right)_0 xy + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial x \partial z}\right)_0 xz +$$
$$\frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial y \partial x}\right)_0 yx + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial y^2}\right)_0 y^2 + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial y \partial z}\right)_0 yz +$$
$$\frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial z \partial x}\right)_0 zx + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial z \partial y}\right)_0 zy + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial z^2}\right)_0 z^2 + \ldots =$$
$$\mathcal{E}_{x,0} + \sum_q \left(\frac{\partial \mathcal{E}_x}{\partial q}\right)_0 q + \frac{1}{2}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_x}{\partial q \partial q'}\right)_0 qq' + \ldots$$

and similarly

$$\mathcal{E}_y = \mathcal{E}_{y,0} + \sum_q \left(\frac{\partial \mathcal{E}_y}{\partial q}\right)_0 q + \frac{1}{2}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_y}{\partial q \partial q'}\right)_0 qq' + \ldots,$$
$$\mathcal{E}_z = \mathcal{E}_{z,0} + \sum_q \left(\frac{\partial \mathcal{E}_z}{\partial q}\right)_0 q + \frac{1}{2}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_z}{\partial q \partial q'}\right)_0 qq' + \ldots.$$

Energy gain due to a shift of the electric charge Q

These two electric field intensities (at points $\mathbf{r}_0$ and $\mathbf{r}_0 + \mathbf{r}$) have been calculated in order to consider the energy gain associated with the shift $\mathbf{r}$ of the electric point charge Q. Similar to the one-dimensional case just considered, we have the energy gain $\Delta E = -Q\boldsymbol{\mathcal{E}} \cdot \mathbf{r}$. There is only one problem: *which* of the two electric field intensities is to be inserted into the formula? Since the vector $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ is small ($\mathbf{i}, \mathbf{j}, \mathbf{k}$ stand for unit vectors corresponding to axes x, y, z, respectively), we may insert, e.g., the mean value of $\boldsymbol{\mathcal{E}}(\mathbf{r}_0)$ and $\boldsymbol{\mathcal{E}}(\mathbf{r}_0 + \mathbf{r})$. We quickly get the following:

$$\Delta E = -Q\boldsymbol{\mathcal{E}} \cdot \mathbf{r} = -Q\frac{1}{2}\left[\boldsymbol{\mathcal{E}}(\mathbf{r}_0) + \boldsymbol{\mathcal{E}}(\mathbf{r}_0 + \mathbf{r})\right]\mathbf{r} =$$
$$-\frac{1}{2}Q\left[\mathbf{i}\left(\mathcal{E}_{x,0} + \mathcal{E}_x\right) + \mathbf{j}\left(\mathcal{E}_{y,0} + \mathcal{E}_y\right) + \mathbf{k}\left(\mathcal{E}_{z,0} + \mathcal{E}_z\right)\right](\mathbf{i}x + \mathbf{j}y + \mathbf{k}z)$$
$$= -\mathcal{E}_{x,0}Qx - \mathcal{E}_{y,0}Qy - \mathcal{E}_{z,0}Qz +$$

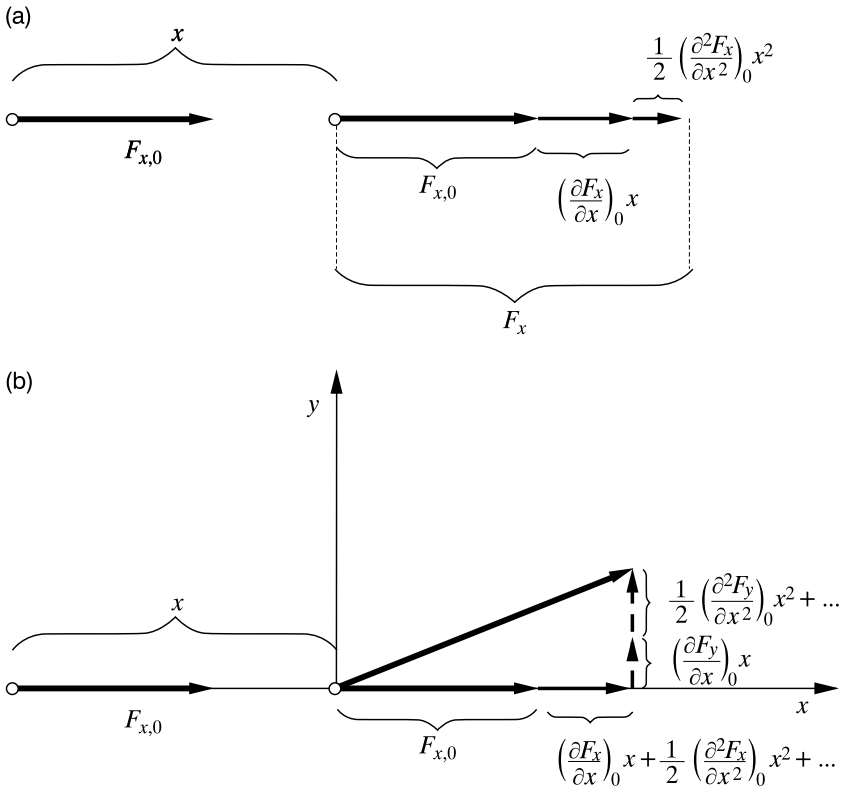


Fig. 4.2. The electric field computed at point $x \ll 1$ from its value (and the values of its derivatives) at point 0. (a) The one-dimensional case. (b) The two-dimensional case.

$$
\begin{aligned}
&- Q\frac{1}{2}\sum_q \left(\frac{\partial \mathcal{E}_x}{\partial q}\right)_0 qx - Q\frac{1}{4}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_x}{\partial q \partial q'}\right)_0 qq'x + \\
&- Q\frac{1}{2}\sum_q \left(\frac{\partial \mathcal{E}_y}{\partial q}\right)_0 qy - Q\frac{1}{4}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_y}{\partial q \partial q'}\right)_0 qq'y + \\
&- Q\frac{1}{2}\sum_q \left(\frac{\partial \mathcal{E}_z}{\partial q}\right)_0 qz - Q\frac{1}{4}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_z}{\partial q \partial q'}\right)_0 qq'z + \ldots \\
= &- \sum_q \mathcal{E}_{q,0}\tilde{\mu}_q - \frac{1}{2}\sum_{q,q'} \left(\frac{\partial \mathcal{E}_q}{\partial q'}\right)_0 \tilde{\Theta}_{qq'} - \frac{1}{4}\sum_{q,q',q''} \left(\frac{\partial^2 \mathcal{E}_q}{\partial q' \partial q''}\right)_0 \tilde{\Omega}_{qq'q''} + \ldots, \qquad (4.4)
\end{aligned}
$$

where "+..." denotes higher-order terms, while $\tilde{\mu}_q = Qq$, $\tilde{\Theta}_{qq'} = Qqq'$, $\tilde{\Omega}_{qq'q''} = Qqq'q''$, ... represent the components of the successive *moments* of a particle with electric charge Q pointed

by the vector $\mathbf{r}_0 + \mathbf{r}$ and calculated within the coordinate system located at $\mathbf{r}_0$. For example, $\tilde{\mu}_x = Qx$, $\tilde{\Theta}_{xy} = Qxy$, $\tilde{\Omega}_{xzz} = Qxz^2$, etc.

Traceless multipole moments

The components of such moments in general are not independent. The three components of the dipole moment are indeed independent, but among the quadrupole components we have the obvious relations $\tilde{\Theta}_{qq'} = \tilde{\Theta}_{q'q}$ from their definition, which reduces the number of independent components from 9 to 6. This however is not all. From the Maxwell equations (see Appendix V1-G, p. V1-673), we obtain the *Laplace equation* $\Delta V = 0$ (Δ means the Laplacian) valid for points without electric charges. Since $\boldsymbol{\mathcal{E}} = -\nabla V$ and therefore $-\nabla \boldsymbol{\mathcal{E}} = \Delta V$, we obtain

$$\nabla \boldsymbol{\mathcal{E}} = \sum_q \frac{\partial \mathcal{E}_q}{\partial q} = 0. \tag{4.5}$$

Thus, in the energy expression $-\frac{1}{2}\sum_{q,q'}\left(\frac{\partial \mathcal{E}_q}{\partial q'}\right)_0 \tilde{\Theta}_{qq'}$ of Eq. (4.4), the quantities $\tilde{\Theta}_{qq'}$ are not independent, since we have to satisfy condition (4.5).

We have therefore only five independent moments that are quadratic in coordinates. For the same reasons we have only seven (among 27) independent moments with the third power of coordinates. Indeed, ten original components $\Omega_{q,q',q''}$ with $(q, q', q'') = xxx, yxx, yyx, yyy, zxx, zxy, zzx, zyy, zzy, zzz$ correspond to all permutational nonequivalent moments. We have, however, three relations these components have to satisfy. They correspond to the three equations, each obtained from the differentiation of Eq. (4.5) over x, y, z, respectively. This results in only seven *independent* components[7] $\Omega_{q,q',q''}$.

These relations between moments can be taken into account (adding to the energy expression the zeros resulting from the Laplace equation (4.5)) and we may introduce what are known as the *traceless Cartesian multipole moments*[8] (the symbol with tilde), which may be chosen in the following way:

$$\mu_q \equiv \tilde{\mu}_q, \tag{4.6}$$

[7] In Appendix G on p. 613 the definition of the polar coordinates-based multipole moments is reported. The number of independent components of such moments is equal to the number of independent Cartesian components and equals $(2l + 1)$ for $l = 0, 1, 2, \ldots$ with the consecutive l pertaining, respectively, to the monopole (or charge) $(2l + 1 = 1)$, dipole (3), quadrupole (5), octupole (7), etc. (in agreement with what we have found a while before for the particular moments).

[8] The reader will find the corresponding formulae in the article by A.D. Buckingham, *Advan. Chem. Phys.*, 12(1967)107, or by A.J. Sadlej, "*Introduction to the Theory of Intermolecular Interactions*," Lund's Theoretical Chemistry Lecture Notes, Lund, 1990.

$$\Theta_{qq'} \equiv \frac{1}{2}\left[3\tilde{\Theta}_{qq'} - \delta_{qq'}\sum_{q}\tilde{\Theta}_{qq}\right]. \tag{4.7}$$

The adjective "traceless" results from relations of the type $\mathrm{Tr}\Theta = \sum_q \Theta_{qq} = 0$, etc.

Then, the expression for the energy contribution changes to (please check that both expressions are identical after using the Laplace formula)

$$\Delta E = -\sum_{q}\mathcal{E}_{q,0}\mu_q - \frac{1}{3}\sum_{q,q'}\left(\frac{\partial\mathcal{E}_q}{\partial q'}\right)_0\Theta_{qq'} - .... \tag{4.8}$$

Most often we first compute the moments and then use them to calculate the traceless multipole moments (cf. the table on p. 66).

System of charges in a nonhomogeneous electric field

Since we are interested in constructing the *perturbation operator* that is to be added to the Hamiltonian, from now on, according to the postulates of quantum mechanics (Chapter V1-1), we will treat the x, y, z coordinates in Eq. (4.8) as *operators* of multiplication by just x, y, z. In addition we would like to treat many charged particles, not just one, because we want to consider molecules. To this end we will sum up all the above expressions, computed for each charged particle, separately. As a result the Hamiltonian for the total system (nuclei and electrons) in the electric field $\boldsymbol{\mathcal{E}}$ represents the Hamiltonian of the system without field ($\hat{H}^{(0)}$) and the perturbation ($\hat{H}^{(1)}$), i.e.,

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)}, \tag{4.9}$$

where

$$\hat{H}^{(1)} = -\sum_{q}\hat{\mu}_q\mathcal{E}_q - \frac{1}{3}\sum_{qq'}\hat{\Theta}_{qq'}\mathcal{E}_{qq'}... \tag{4.10}$$

with the convention

$$\mathcal{E}_{qq'} \equiv \frac{\partial\mathcal{E}_q}{\partial q'},$$

where the field component and its derivatives are computed at a given point ($\mathbf{r}_0$), e.g., in the center of mass of the molecule, while $\hat{\mu}_q, \hat{\Theta}_{qq'}, \ldots$ denote the operators of the components of the traceless Cartesian multipole moments of the total system, i.e., of the molecule.[9] How can we imagine multipole moments? We may associate a given multipole moment with a simple object that exhibits a nonzero value for this particular moment, but all lower multipole moments equal zero.[10] Some of such objects are shown in Fig. 4.3, located between two charges Q and $-Q$, producing an "external field." Note that the multipole moment names (*dipole*, *quadrupole*, *octupole*) indicate the *number* of the point charges from which they are built.

Eq. (4.10) means that if the system exhibits nonzero multipole moments (before any interaction or due to the interaction), they will interact with the external electric field: the dipole with the electric field intensity, the quadrupole with its gradient, etc. Fig. 4.3 shows why this happens.

4.2.2 The homogeneous electric field

In the case of a *homogeneous* external electric field, the contribution to $\hat{H}^{(1)}$ comes from the first term in Eq. (4.10), i.e.,

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)} = \hat{H}^{(0)} - \hat{\mu}_x \mathcal{E}_x - \hat{\mu}_y \mathcal{E}_y - \hat{\mu}_z \mathcal{E}_z = \hat{H}^{(0)} - \hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}, \tag{4.11}$$

where the dipole moment operator $\hat{\boldsymbol{\mu}}$ has the form

$$\hat{\boldsymbol{\mu}} = \sum_i \mathbf{r}_i Q_i, \tag{4.12}$$

with the vector $\mathbf{r}_i$ indicating the particle i of charge Q_i.

Hence,

$$\frac{\partial \hat{H}}{\partial \mathcal{E}_q} = -\hat{\mu}_q. \tag{4.13}$$

From this it follows that

$$\langle \psi | \frac{\partial \hat{H}}{\partial \mathcal{E}_q} \psi \rangle = -\langle \psi | \hat{\mu}_q \psi \rangle = -\mu_q, \tag{4.14}$$

where μ_q is the expectation value of the q-th component of the dipole moment.

[9] Also calculated with respect to this point. This means that if the molecule is large, then $\mathbf{r}$ may become dangerously large. In such a case, as a consequence, the series (4.8) may converge slowly.

[10] *Higher* moments in general will be nonzero.

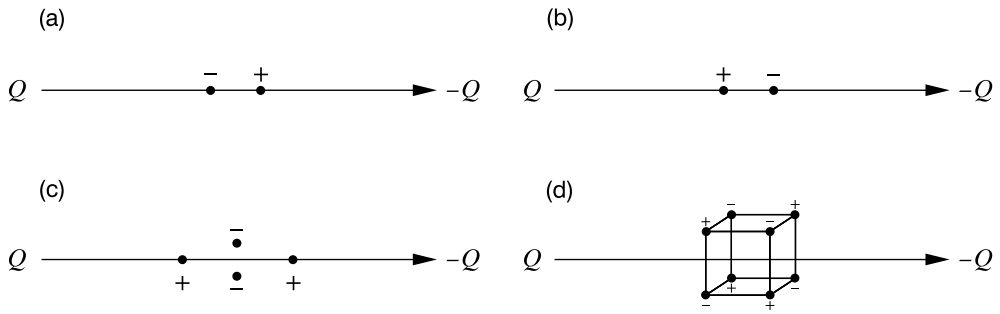


Fig. 4.3. Explanation of why a dipole moment interacts with the electric field intensity and a quadrupole moment with its gradient, while the octupole moment does not interact either with the first or with the second. The external electric field is produced by two *distant* electric charges $Q > 0$ and $-Q$ (for long distances between them the field in the central region between the charges resembles a homogeneous field) and interacts with an object (a dipole, a quadrupole, etc.) located in the central region. A favorable orientation of the object corresponds to the lowest interaction energy with Q and $-Q$. (a) Such a low-energy situation for a dipole: the charge "+" protrudes towards $-Q$, while the charge "−" protrudes towards Q. (b) The opposite situation, energetically nonfavorable. As we can see, the interaction energy of the dipole with the electric field *differentiates* these two situations. (c) Now, let us locate a quadrupole in the middle. Let us imagine that a neutral point object has just split into four point charges (of the same absolute value). The system lowers its energy by the "−" charges going off the axis, because they have increased their distance from the charge $-Q$, but at the same time the system energy has increased by the same amount, since the charges went off the symmetrically located charge $+Q$. What about the "+" charges? The splitting of the "++" charges leads to an energy gain for the right-hand side "+" charge, because it approached $-Q$ and went off the charge $+Q$, but the left-hand side "+" charge gives the opposite energy effect. Altogether the net result is zero. Conclusion: *the quadrupole does not interact with the homogeneous electric field*. Now, let us imagine an inhomogeneous field having a nonzero gradient along the axis (e.g., both Q charges differ by their absolute values). There will be no energy difference for the "−" charges, but one of the "+" charges will be attracted more strongly than the other. Therefore, the *quadrupole interacts with the field gradient*. We may foresee that the quadrupole will align with its longer axis along the field. (d) An octupole (all charges have the same absolute value). Indeed, the total charge, all the components of the dipole as well as of the quadrupole moment are equal to zero, but the octupole (eight charges in the vertices of a cube) is nonzero. Such an octupole does not interact with a homogeneous electric field (because the right and left sides of the cube do not gain anything when interacting), and it also does not interact with the field gradient (because each of the abovementioned sides of the cube is composed of two plus and two minus charges; what the first ones gain the second ones lose).

From the Hellmann–Feynman theorem we have

$$\langle \psi | \frac{\partial \hat{H}}{\partial \mathcal{E}_q} \psi \rangle = \frac{\partial E}{\partial \mathcal{E}_q}, \tag{4.15}$$

and therefore

$$\frac{\partial E}{\partial \mathcal{E}_q} = -\mu_q. \tag{4.16}$$

On the other hand, in the case of a *weak electric field* $\boldsymbol{\mathcal{E}}$ we may certainly write the Taylor expansion,

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} + \sum_q \left(\frac{\partial E}{\partial \mathcal{E}_q}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_q + \frac{1}{2!}\sum_{q,q'} \left(\frac{\partial^2 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_q \mathcal{E}_{q'} + \frac{1}{3!}\sum_{q,q',q''} \left(\frac{\partial^3 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'} \partial \mathcal{E}_{q''}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_q \mathcal{E}_{q'} \mathcal{E}_{q''} + \ldots, \tag{4.17}$$

where $E^{(0)}$ stands for the energy of the unperturbed molecule.

Linear and nonlinear responses to a homogeneous electric field

Comparing (4.16) and (4.17) we get

$$\frac{\partial E}{\partial \mathcal{E}_q} = -\mu_q = \left(\frac{\partial E}{\partial \mathcal{E}_q}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} + \sum_{q'} \left(\frac{\partial^2 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_{q'} + \frac{1}{2}\sum_{q'} \left(\frac{\partial^3 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'} \partial \mathcal{E}_{q''}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_{q'} \mathcal{E}_{q''} \ldots, \tag{4.18}$$

or replacing the derivatives by their equivalents (permanent dipole moment, molecular polarizability and hyperpolarizabilities)

$$\mu_q = \mu_{0q} + \sum_{q'} \alpha_{qq'} \mathcal{E}_{q'} + \frac{1}{2}\sum_{q'q''} \beta_{qq'q''} \mathcal{E}_{q'} \mathcal{E}_{q''} + \ldots. \tag{4.19}$$

The meaning of the formula for μ_q is clear: in addition to the permanent dipole moment $\boldsymbol{\mu}_0$ of the isolated molecule, we have its modification, i.e., an induced dipole moment, which consists of the *linear* part in the field ($\sum_{q'} \alpha_{qq'} \mathcal{E}_{q'}$) and of the *nonlinear* part ($\frac{1}{2}\sum_{q'q''} \beta_{qq'q''} \mathcal{E}_{q'} \mathcal{E}_{q''} + \ldots$). The quantities that characterize the molecule, vector $\boldsymbol{\mu}_0$ and tensors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ..., are of key importance. By comparing (4.18) with (4.19) we have the following relations:

for the permanent (field-independent) dipole moment of the molecule (component q),

$$\mu_{0q} = -\left(\frac{\partial E}{\partial \mathcal{E}_q}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}}; \tag{4.20}$$

for the total dipole moment (field-dependent),

$$\mu_q = -\left(\frac{\partial E}{\partial \mathcal{E}_q}\right); \tag{4.21}$$

for the component qq' of the *dipole polarizability* tensor,

$$\alpha_{qq'} = -\left(\frac{\partial^2 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} = \left(\frac{\partial \mu_q}{\partial \mathcal{E}_{q'}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}}; \tag{4.22}$$

for the component $qq'q''$ of the *dipole hyperpolarizability* tensor,

$$\beta_{qq'q''} = -\left(\frac{\partial^3 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'} \partial \mathcal{E}_{q''}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}}. \tag{4.23}$$

Next, we would obtain higher-order dipole hyperpolarizabilities (γ, ...), which will contribute to the characteristics of the way the molecule is polarized when subject to a weak electric field.

The homogeneous field: dipole polarizability and dipole hyperpolarizabilities

When using the definition of $\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ from Eq. (4.17) we have the following expression for the energy of the molecule in the electric field:

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} - \sum_q \mu_{0q}\mathcal{E}_q - \frac{1}{2}\sum_{qq'} \alpha_{qq'}\mathcal{E}_q\mathcal{E}_{q'} - \frac{1}{3!}\sum_{qq'q''} \beta_{qq'q''}\mathcal{E}_q\mathcal{E}_{q'}\mathcal{E}_{q''} - \frac{1}{4!}\sum_{qq'q''q'''} \gamma_{qq'q''q'''}\mathcal{E}_q\mathcal{E}_{q'}\mathcal{E}_{q''}\mathcal{E}_{q'''} \ldots. \tag{4.24}$$

Due to the homogeneous character of the electric field, this formula pertains exclusively to the interaction of the molecular *dipole* (the permanent dipole plus the induced linear and nonlinear response) with the electric field.

As seen from (4.19), the induced dipole moment with the components $\mu_q - \mu_{0q}$ may have a different direction from the applied electric field (due to the tensor character of the polarizability and hyperpolarizabilities). This is quite understandable, because the electrons will move in a direction which will represent a compromise between the direction the electric field forces them to move and the direction where the polarization of the molecule is easiest (Fig. 4.4).

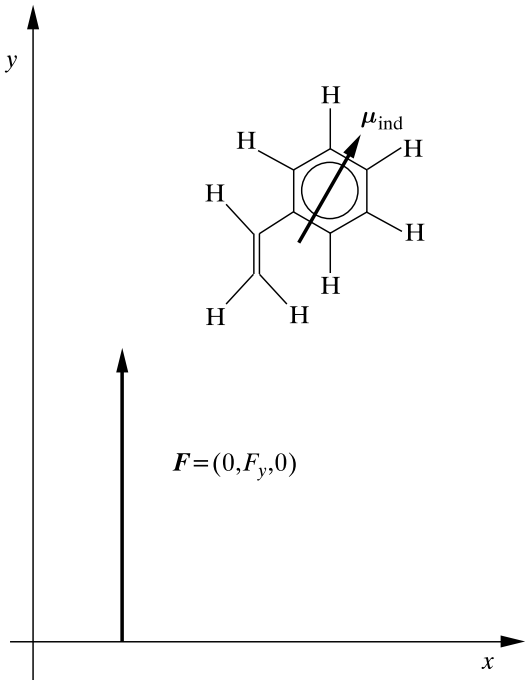


Fig. 4.4. The direction of the induced dipole moment may differ from the direction of the electric field applied (due to the tensor character of the polarizability and hyperpolarizabilities). Example: the vinyl molecule in a planar conformation. Assume the following Cartesian coordinate system: x (horizontal in the figure plane), y (vertical in the figure plane), and z (perpendicular to the figure plane), and the external electric field $\boldsymbol{\mathcal{E}} = (0, \mathcal{E}_y, 0)$. The x component of the induced dipole moment is equal to (within the accuracy of linear terms, Eq. (4.19)) $\mu_{ind,x} = \mu_x - \mu_{0x} \approx \alpha_{xy}\mathcal{E}_y$, $\mu_{ind,y} \approx \alpha_{yy}\mathcal{E}_y$, $\mu_{ind,z} \approx \alpha_{zy}\mathcal{E}_y$. Due to the symmetry plane $z = 0$ of the molecule (cf. p. 271) $\alpha_{zy} = \alpha_{zx} = 0$, and similarly for the hyperpolarizabilities, we have $\mu_{ind,z} = 0$. As we can see, despite the field having its x component equal to zero, the induced dipole moment x component is not equal to zero ($\mu_{ind,x} \neq 0$).

It is seen from Eqs. (4.19) and (4.22) that:

- As a second derivative of a continuous function E the polarizability represents a symmetric tensor ($\alpha_{qq'} = \alpha_{q'q}$).
- The polarizability characterizes *this part of the induced dipole moment, which is proportional to the field.*

- If nondiagonal components of the polarizability tensor are nonzero, then the charge flow direction within the molecule will differ from the direction of the field. This would happen when the electric field forced the electrons to flow into empty space, while they had a "highway" to travel along some chemical bonds (cf. Fig. 4.4).
- If a molecule is symmetric with respect to the plane $q = 0$, say, $z = 0$, then all the (hyper)polarizabilities with odd numbers of the indices q are equal to zero (cf. Fig. 4.4). It has to be like this, because otherwise a change of the electric field component from $\mathcal{E}_z$ to $-\mathcal{E}_z$ would cause a change in energy (see Eq. (4.24)), which is impossible, because the molecule is symmetric with respect to the plane $z = 0$.
- The dipole *hyperpolarizabilities* (β and higher-order) are very important, because if we limited ourselves to the first two terms of (4.19) containing only μ_{0q} and $\alpha_{qq'}$ (i.e., neglecting β and higher hyperpolarizabilities), the molecule would be equally easy to polarize in two opposite directions.[11] This is why, for a molecule with a center of inversion, all *odd* dipole hyperpolarizabilities (i.e., with an odd number of indices q) have to equal zero, because the invariance of the energy with respect to the inversion will be preserved that way. If the molecule does not exhibit an inversion center, the nonzero odd dipole hyperpolarizabilities ensure polarization of the molecule depends, in general, on whether we change the direction of the electric field vector to the opposite. This is how it should be. Why were the electrons to move to the same extent towards an electron donor (on one end of the molecule) and to an electron acceptor (on the other end)?

Does the dipole moment really exist?

Now, let us complicate things. What is $\boldsymbol{\mu}_0$? We used to say that it is the dipole moment of the molecule, $\boldsymbol{\mu}_0 = \langle\psi|\hat{\boldsymbol{\mu}}_0\psi\rangle$. Unfortunately, no molecule has a nonzero dipole moment in any of its stationary states. This follows from the invariance of the Hamiltonian with respect to the inversion operation and was described on p. V1-81. The mean value of the dipole moment operator is bound to be zero since $|\psi|^2$ is symmetric, while the dipole moment operator itself is antisymmetric with respect to the inversion. Thus for *any* molecule[12] $\mu_{0q} = 0$ for $q = x, y, z$.

Strange? No, not at all. The reason is the rotational part of the wave function (cf. p. V1-322). This is quite natural. Dear reader, have you ever thought about why the hydrogen atom does not exhibit a dipole moment despite having two opposite poles, that of the proton and that of the electron? The reason is the same. The electron in its ground state is described by the $1s$

[11] According to Eq. (4.19) the absolute value of the q component of the *induced dipole moment* $\boldsymbol{\mu}_{ind} = \boldsymbol{\mu} - \boldsymbol{\mu}_0$ would be identical for $\mathcal{E}_q$ as well as for $-\mathcal{E}_q$.

[12] "Everybody knows" that the HF molecule has a nonzero dipole moment. Common knowledge says that when an electric field is applied, the HF dipole gets aligned along the electric field vector. At any field, no matter how small? This would be an incredible scenario. No, the picture has to be more complex.

orbital, which does not prefer any direction, and the dipole moment integral for the hydrogen atom gives zero.[13] Evidently, we have got into trouble.

The trouble disappears after the Born–Oppenheimer approximation (the clamped nuclei approximation, cf. p. V1-318) is used, i.e., if we hold the *molecule fixed in space*. In such a case, the molecule *has* the dipole moment and this dipole moment is to be inserted into formulae as $\boldsymbol{\mu}_0$, and then we may calculate the polarizability, hyperpolarizabilities, etc. (see p. V1-82). But what do we do, when we do not apply the Born–Oppenheimer approximation? Yet, in experiments we do not use the Born–Oppenheimer approximation (or any other). We have to allow the molecule to rotate and then the dipole moment $\boldsymbol{\mu}_0$ disappears. Well, not quite, since the space now is no longer isotropic. There is a chance to measure a dipole moment. What do we measure then?

It is always good to see things working in a simple model, and simple models resulting in exact solutions of the Schrödinger equation have been described in Chapter V1-4. A good model for our rotating molecule may be the rigid rotator with a dipole moment (a charge Q on one mass and $-Q$ on the other).[14] The Hamiltonian remains, in principle, the same as for the rigid rotator, because we have to add a *constant* $-\frac{Q^2}{R}$ to the potential energy, which does not change anything. Thus the ground-state wave function is $Y_0^0 = \text{const}$ as before, which tells us that *every* orientation of the rigid dipolar rotor in space is *equally probable*.

In a homogeneous electric field such a wave function will not be a constant, but will have a single maximum for the electric field direction (and the minimum for the opposite direction). For any field intensity, this will correspond to the state of the lowest energy. This is natural, because a dipole should have a tendency to align along the orientation of the electric field.

What will it be if the rotator were in one of its excited states? Well, we can guess. A strong electric field will stop the rotation in order to align the rotator along the field and make from the rotator a kind of oscillator allowing only its vibrations about the field direction. In the first approximation this can be viewed as a harmonic oscillator, hence the equidistant distribution of the energy levels.[15]

---

[13] The same is true in any excited stationary state, because $\int \mathrm{d}\mathbf{r}x\,|\psi_{nlm}|^2 = 0$.

[14] This moment therefore has a constant length.

[15] The perturbation for small angle θ can be written as

$$\hat{H}^{(1)} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} = -RQ\mathcal{E}\cos\phi \simeq -\mathcal{E}RQ\left(1 - \frac{1}{2}\phi^2\right) = \text{const} + \frac{1}{2}\mathcal{E}RQ\phi^2 = \text{const} + \frac{1}{2}k\phi^2,$$

which corresponds to a harmonic dependence on ϕ with the force constant $k = \mathcal{E}RQ$.

The above reasoning suggests that the larger the field intensity the lower the energy of dipolar molecules. If this happened, the molecules would seek regions with the strongest electric field. We will see in a while that this is the case, but not always.[16]

A surprise at excited states

It has been shown in experiments that dipole molecules (*even the same* as those ones seeking high fields) may seek sometimes low field values, i.e., may be expelled from the electric field.[17] How is this possible?

Let us study these things in more detail. Imagine a free rigid rotator – a model of a dipolar molecule, where for the sake of simplicity we assume it now rotates within a plane. After separation of the center of mass (cf. Appendix V1-J, p. V1-691) one has to solve a problem for a single particle (with the reduced mass μ). It is like if one of the particles, having the negative charge $-Q$, resided in the origin, while the second particle, with charge Q, moved around in a circle of radius R, where R stands for the length of the rigid rotator. The only variable is the angle ϕ. This problem has been solved[18] in Chapter V1-4 (p. V1-193). The expression for the energy (after inserting $L = 2\pi R$) reads as[19]

$$E_J = -\frac{Q^2}{R} + J^2 \frac{\hbar^2}{2\mu R^2} = \text{const} + J^2 \frac{\hbar^2}{2\mu R^2}, \tag{4.25}$$

---

16 Such molecules are known as the "*high-field seekers.*" Basing on molecular dynamics it is possible to predict (H.J. Loesch, B. Scheel, *Phys. Rev. Letters*, 85(2000)2709) what happens in the following situation. Suppose we have a steel cylinder with a metal wire along its axis. A voltage difference is applied to the cylinder and the wire, resulting in an inhomogeneous electric field, the highest field being at the wire. A molecular beam of polar molecules (like NaCl, NaBr, NaI) when injected on one side of the cylinder begins to orbit in a helix-like motion about the wire. It is also possible to join the ends of the cylinder (making a torus) and forming a closed trajectory of the beam. Such devices might serve in the future as reservoirs of molecules in a given quantum state.

17 "*Low-field seekers.*"

18 The moving particle was an electron, but it does not matter. For the dipole there will be the electrostatic interaction of the two charges, but this interaction is constant (since R is a constant) and therefore irrelevant.

19 In a more formal derivation we write down first the Hamiltonian for a dipole rotator (two point masses with charges Q and $-Q$ and distance R): $H^{(0)} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 + \text{const}$, $\text{const} = -\frac{Q^2}{R}$. Next, we separate the center-of-mass motion (Appendix V1-J, p. V1-691) and get the Schrödinger equation with the Hamiltonian $\hat{H}^{(0)} = -\frac{\hbar^2}{2\mu}\Delta - \frac{Q^2}{R}$ which describes the relative motion of the two particles. After expressing Δ in spherical coordinates (Appendix V1-S, p. V1-735) and putting $R = \text{const}$ and $\theta = \pi/2$ (rotations about the z axis only), we get $\hat{H}^{(0)} = -\frac{\hbar^2}{2\mu R^2}\frac{\partial^2}{\partial\phi^2} + \text{const}$. The corresponding eigenfunctions are $\Phi_J(\phi) = \frac{1}{\sqrt{2\pi}}\exp(iJ\phi)$, $J = 0, \pm1, \pm2, ...$, with the eigenvalues given by Eq. (4.25).
The same result can be obtained equivalently by postulating that an integer number ($|J|$) of the de Broglie wave lengths (λ) should match the $2\pi R$ distance, i.e., $2\pi R = |J|\lambda$, where the quantum number $J = 0, \pm1, \pm2, ...$ From the de Broglie relation $\lambda = \frac{h}{p}$ (p stands for the momentum of the moving particle of mass μ) one gets $p = |J|\frac{\hbar}{R}$. Hence, the total energy (being the kinetic energy only) is $E = \frac{\mu v^2}{2} = \frac{p^2}{2\mu} = J^2\frac{\hbar^2}{2\mu R^2}$, as in Eq. (4.25).

for $J = 0, \pm 1, \pm 2, ...$ We have therefore the ground state corresponding to $J = 0$ and the doubly degenerate excited states with $J = \pm 1, \pm 2, ...$ corresponding to the wave functions $\Phi_J(\phi) = \frac{1}{\sqrt{2\pi}} \exp(iJ\phi)$. One of the excited-state wave functions, $\Phi_J(\phi)$, $J > 0$, corresponds to the rotation that increases ϕ, the second one, $\Phi_{-J}(\phi)$, describes the rotation in the opposite direction. Note that in each of these states the probability density of finding the rotating particle is uniform $|\Phi_J(\phi)|^2 = |\Phi_{-J}(\phi)|^2 = \frac{1}{2\pi} \neq f(\phi)$, i.e., the rotational motion is uniform (the wave functions are perfectly delocalized).

In the external electric field $\boldsymbol{\mathcal{E}}$ the Hamiltonian reads as (see (4.11))

$$\hat{H} = \hat{H}^{(0)} - \hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}. \tag{4.26}$$

Note that a uniform electric field along the $\phi = 0$ direction has to cause the degeneracy to be lifted. Indeed, for a very strong field the wave functions and the energies have to be similar to those of the harmonic oscillator (the dipole will oscillate about the $\phi = 0$ direction), which means nondegeneracy. How may such a transition from the rotator (the degenerate levels) to the oscillator (the nondegenerate levels) look?

There are two opposite effects manifesting themselves in every excited state: a free rotation of the particle ($E_J = J^2 \frac{\hbar^2}{2\mu R^2}$) is modified by the interaction of the dipole with the field that tends to stop the rotation and to orient the dipole along the field with the maximum energy gain (for a perfect alignment) $-\mu\mathcal{E} = -\mathcal{E}QR$. The ratio of these two tendencies can be characterized by parameter $\gamma = \frac{\mathcal{E}}{J^2}$.

Let us begin from an excited energy level with a negligible value of γ (a very weak field or a highly excited rotational level). We will have to do with a practically degenerate level with the two delocalized rotational states ($\pm J$) very similar to those of the free rotator.

Let us increase γ to some medium values (a stronger field, in our calculations $\frac{\hbar^2}{2\mu R^2} = 0.00005$ a.u. and $\mathcal{E}QR = 0.001$ a.u.) and consider the low energy levels. We obtain a symmetric (with respect to $\phi \to -\phi$) nodeless ground state wave function ψ_0, but unlike that for the free rotator, *showing a maximum amplitude for the orientation of the field* ($\phi = 0$). Thus, in this state the dipole has a propensity for orientation along the electric field. Next energy levels correspond to two wave functions (stemming from the nonperturbed functions with $J = \pm 1$), with a single nodal plane each. The lower level corresponds to the antisymmetric wave function ψ_1, with the nodal plane going through points $\phi = 0$ and $\phi = 180°$ and resembling very much the first excited state of the harmonic oscillator. Thus, in this state the dipolar rotator is oscillating about the direction of the field. The second one-node state, ψ_2, is symmetric and has maximum $|\psi_2|$ on the side *opposite* to the field! The reason is that all these functions have to orthogonal.[20]

[20] The wave functions represent the eigenfunctions of a Hermitian operator and correspond to different energies in a non-zero field. Such functions must be orthogonal (Appendix V1-B, p. V1-595).

The nodal plane of ψ_2 when orthogonal to that of ψ_1 keeps $\langle\psi_1|\psi_2\rangle = 0$. However, to fulfill $\langle\psi_0|\psi_2\rangle = 0$, one has to have $|\psi_2|$ small on the side of the field and *large on the side opposite to the field!*

We will come to similar conclusions for each pair of excited states with a given J.

For very large γ (very strong electric fields) the electric field will overcome the rotational kinetic energy for the states up to some large J: the ground state, as well as the excited states of both kinds described above, will lower their energies due to the overwhelming influence of the electric field. These states will be localized close to the direction ($\phi = 0$), all resembling the harmonic oscillator wave functions.

When in a nonhomogeneous field, the molecules in such states will seek the stronger field domains to lower their energies. The higher excited states will still be localized about the direction opposite to the field. The molecules in such states will be the low-field seekers, they will be expelled outside of the field.

The orientation of an electric dipole opposite to the electric field seems counterintuitive, but it is not. We will use here an analogy. Children like to use a swing, which is nothing else but a rotator in the gravitational field. If there were no such field (say, on a solid space ship) the swing would move around at a certain speed in one of two possible directions. The gravitational field (similarly as the electric field acting on the moving positive charge) forces the swing position to prefer the down direction. This means children like the high-field seeking states of the swing. However, besides the children there are acrobats, who manage not only to make almost free rotation (a delocalized state), but also to get a state of much higher energy, as shown in Fig. 4.5.

The acrobat, after exceeding some kinetic energy, is able to *spend more time being oriented opposite to the field* than along the field! In theoretician's reasoning there is a possibility to lower the energy of such a state by going off the gravitational field. This is therefore an analog of the low-field seeker dipole molecule.

4.2.3 The nonhomogeneous electric field: multipole polarizabilities and hyperpolarizabilities

Let us come back to the nonrotating (immobilized) molecules.

The formula $\mu_q = \mu_{0q} + \sum_{q'} \alpha_{qq'}\mathcal{E}_{q'} + \frac{1}{2}\sum_{q'q''} \beta_{qq'q''}\mathcal{E}_{q'}\mathcal{E}_{q''} + \ldots$ pertains to the polarizabilities and hyperpolarizabilities in a *homogeneous* electric field. The polarizability $\alpha_{qq'}$ characterizes a *linear* response of the molecular *dipole* moment to the electric field, the hyperpolarizability $\beta_{qq'q''}$ and the higher ones characterize the corresponding *nonlinear* response of

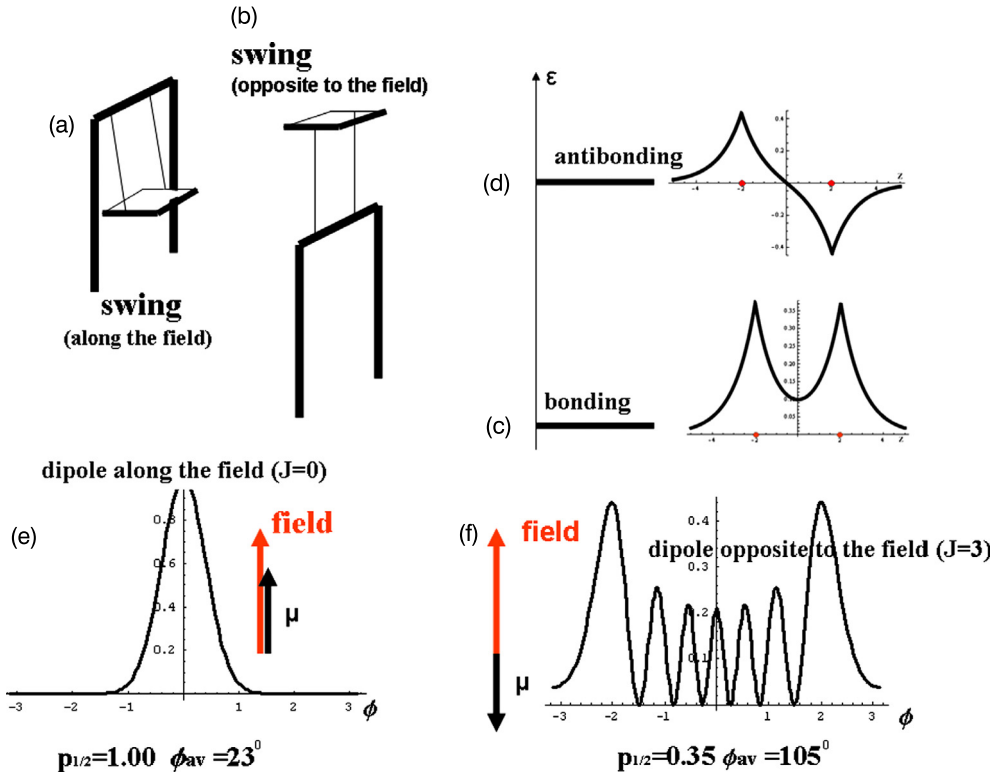


Fig. 4.5. Analogies of the high-field and low-field seeking states of dipolar molecules. (a,b) Two stable states of a swing: (a) downwards (a large stability) and (b) upwards (marginal stability). (c,d) Another analogy – two electronic states of the molecular ion H_2^+: (c) the ground state corresponding to the stable molecule (the bonding orbital) and (d) an excited state that corresponds to a dissociating molecule (the antibonding orbital). (e,f) The probability density of a given orientation ϕ of a dipole in the electric field (corresponding to $\phi = 0$) in two particular states. The quantity $p_{\frac{1}{2}}$ stands for the calculated probability that the dipole–field angle $\phi \in (-\pi/2, +\pi/2)$, ϕ_{av} means the mean value of $|\phi|$. (e) In the ground rotational state ($J = 0$) $\phi_{\mathrm{av}} = 23°$, i.e., the dipole prefers to be oriented along the field; in the excited rotational state ($J = 3$) the expected orientation is $\phi_{\mathrm{av}} = 105°$, i.e., mostly opposite to the field, the highest probability density is for the dipole direction almost opposite to the electric field. These data correspond to $\frac{\hbar^2}{2\mu R^2} = 0.00005$ a.u. and $\mathcal{E}QR = 0.001$ a.u.

the molecular *dipole* moment. However, a change of the molecular charge distribution contains more information than just that offered by the induced dipole moment. For a *nonhomogeneous* electric field the energy expression changes, because besides the dipole moment higher multipole moments (permanent as well as induced) come into play (see Fig. 4.3). Using the Hamiltonian (4.9) with the perturbation (4.10) (which corresponds to a molecule immersed in a nonhomogeneous electric field) we obtain the following energy expression from the Hellmann–

Feynman theorem (Eqs. (4.15) and (4.17)):

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} + E_\mu + E_\Theta + E_{\mu-\Theta} + ..., \tag{4.27}$$

where besides the unperturbed energy $E^{(0)}$ of the molecule, we have:

- the dipole–field interaction energy E_μ (including the permanent and induced dipole, these terms appeared earlier for the homogeneous field),

$$E_\mu = -\left[\sum_q \mu_{0q}\mathcal{E}_q + \frac{1}{2}\sum_{qq'}\alpha_{qq'}\mathcal{E}_q\mathcal{E}_{q'} + \frac{1}{6}\sum_{q,q',q''}\beta_{q,q',q''}\mathcal{E}_q\mathcal{E}_{q'}\mathcal{E}_{q''}...\right]; \tag{4.28}$$

- the terms that pertain to the *nonhomogeneity* of the electric field, i.e., the energy E_Θ of the interaction of the field gradient with the quadrupole moment (the permanent one Θ, the first term, and the induced one; C stands for the *quadrupole polarizability*, and then, in the terms denoted by "+..." there are the nonlinear responses with *quadrupole hyperpolarizabilities*),

$$E_\Theta = -\left[\frac{1}{3}\sum_{qq'}\Theta_{qq'}\mathcal{E}_{qq'} + \frac{1}{6}\sum_{qq'q''q'''}C_{qq'q''q'''}\mathcal{E}_{qq'}\mathcal{E}_{q''q'''} + ...\right]; \tag{4.29}$$

- the dipole–quadrupole cross-term $E_{\mu-\Theta}$,

$$E_{\mu-\Theta} = -\left[\frac{1}{3}\sum_{q,q',q''}A_{q,q'q''}\mathcal{E}_q\mathcal{E}_{q'q''} + \frac{1}{6}\sum_{q,q',q'',q'''}B_{qq',q''q'''}\mathcal{E}_q\mathcal{E}_{q'}\mathcal{E}_{q''q'''}\right]; \tag{4.30}$$

and

- the interaction of higher multipoles (permanent as well as induced: first, the octupole Ω with the corresponding *octupole polarizabilities and hyperpolarizabilities*, etc.) with the higher derivatives of the electric field together with the corresponding cross-terms denoted as +....

4.3 How to calculate the dipole moment

The dipole moment in normalized state $|n\rangle$ is to be calculated (according to the postulates of quantum mechanics, Chapter V1-1; the Born–Oppenheimer approximation is assumed) as the mean value $\boldsymbol{\mu} = \langle n|\hat{\boldsymbol{\mu}}|n\rangle$ of the dipole moment operator[21]

$$\hat{\boldsymbol{\mu}} = -\sum_i \mathbf{r}_i + \sum_A Z_A\mathbf{R}_A, \tag{4.31}$$

[21] As is seen, this is an operator having x, y, and z components in a chosen coordinate system and each of its components means a *multiplication by the corresponding coordinates and electric charges*.

where $\mathbf{r}_i$ are the vectors indicating the electrons and $\mathbf{R}_A$ shows nucleus A with the charge Z_A (in a.u.).

4.3.1 Coordinate system dependence

The dipole moment operator and the dipole moment itself do not depend on the choice of the origin of the coordinate system *only for a neutral molecule*. When two coordinate systems differ by translation $\mathbf{R}$, then, in general, we may obtain two different results:

$$\hat{\boldsymbol{\mu}} = \sum_i Q_i \mathbf{r}_i ,$$
$$\hat{\boldsymbol{\mu}}' = \sum_i Q_i \mathbf{r}_i' = \sum_i Q_i (\mathbf{r}_i + \mathbf{R}) = \hat{\boldsymbol{\mu}} + \sum_i Q_i \mathbf{R} = \hat{\boldsymbol{\mu}} + \mathbf{R} \sum_i Q_i . \tag{4.32}$$

It is seen that $\hat{\boldsymbol{\mu}}' = \hat{\boldsymbol{\mu}}$ only if $\sum_i Q_i = 0$, i.e., for a neutral system.[22]

This represents a special case of the theorem, saying that the lowest nonvanishing multipole moment does not depend on the choice of coordinate system; all others may depend on that choice.

4.3.2 Hartree–Fock approximation

In order to show the reader how we calculate the dipole moment in practice, let us use the Hartree–Fock approximation. Using the normalized Slater determinant $|\Phi_0\rangle$ we have as the Hartree–Fock approximation to the dipole moment

$$\boldsymbol{\mu} = \langle \Phi_0 | -\sum_i \mathbf{r}_i + \sum_A Z_A \mathbf{R}_A | \Phi_0 \rangle = \langle \Phi_0 | -\sum_i \mathbf{r}_i | \Phi_0 \rangle + \langle \Phi_0 | \sum_A Z_A \mathbf{R}_A | \Phi_0 \rangle = \boldsymbol{\mu}_{el} + \boldsymbol{\mu}_{nucl}, \tag{4.33}$$

where the integration goes over the electronic coordinates. The dipole moment of the nuclei $\boldsymbol{\mu}_{nucl} = \sum_A Z_A \mathbf{R}_A$ is very easy to compute, because in the Born–Oppenheimer approximation the nuclei occupy some fixed positions in space. The electronic component of the dipole moment $\boldsymbol{\mu}_{el} = \langle \Phi_0 | - \sum_i \mathbf{r}_i | \Phi_0 \rangle$, according to the Slater–Condon rules (rule I, Appendix V1-N on p. V1-707), amounts to $\boldsymbol{\mu}_{el} = -\sum_i n_i (\varphi_i | \mathbf{r}_i \varphi_i)$, where n_i stands for the occupation number of the orbital φ_i (let us assume double occupation, i.e., $n_i = 2$). After the LCAO expansion

[22] If you ever have to debug a computer program that calculates the dipole moment, then please remember there is a simple and elegant test at your disposal that is based on the above theorem. You just make two runs of the program for a neutral system each time using a different coordinate system (the two systems differing by a translation). The two results have to be identical.

is applied, $\varphi_i = \sum_j c_{ji}\chi_j$, and combining the coefficients c_{ji} into the bond-order matrix (see p. V1-506) **P**, we have

$$\boldsymbol{\mu}_{el} = -\sum_{kl} P_{lk}(\chi_k|\mathbf{r}|\chi_l). \tag{4.34}$$

This is in principle all we can say about calculation of the dipole moment in the Hartree–Fock approximation. The rest belongs to the technical side. We choose a coordinate system and calculate all the integrals of type $(\chi_k|\mathbf{r}\chi_l)$, i.e., $(\chi_k|x\chi_l)$, $(\chi_k|y\chi_l)$, $(\chi_k|z\chi_l)$. The bond-order matrix **P** is just a by-product of the Hartree–Fock procedure.

4.3.3 Atomic and bond dipoles

Within the Hartree–Fock model the total dipole moment can be decomposed into atomic and pairwise contributions, i.e.,

$$\boldsymbol{\mu}_{el} = -\sum_A \sum_{k\in A}\sum_{l\in A} P_{lk}(\chi_k|\mathbf{r}|\chi_l) - \sum_A \sum_{k\in A}\sum_{B\neq A}\sum_{l\in B} P_{lk}(\chi_k|\mathbf{r}|\chi_l), \tag{4.35}$$

where we assume that the atomic orbital centers (A, B) correspond to the nuclei.[23] To this end we construct the vectors, which indicate from the origin the nuclei and the centers of any pair of them. If the two atomic orbitals k and l belong to *the same* atom, then we insert $\mathbf{r} = \mathbf{R}_A + \mathbf{r}_A$, where $\mathbf{R}_A$ shows the atom (nucleus) A from the origin and $\mathbf{r}_A$ indicates the electron from the local origin centered on A. If k and l belong to *different* atoms, then $\mathbf{r} = \mathbf{R}_{AB} + \mathbf{r}_{AB}$, where $\mathbf{R}_{AB}$ indicates the center of the AB section and $\mathbf{r}_{AB}$ represents the position of the electron with respect to this center. Then

$$\begin{aligned}\boldsymbol{\mu}_{el} = &-\sum_A \mathbf{R}_A \sum_{k\in A}\sum_{l\in A} S_{kl}P_{lk} - \sum_A\sum_{k\in A}\sum_{l\in A} P_{lk}(\chi_k|\mathbf{r}_A|\chi_l)\\ &-\sum_A\sum_{B\neq A}\mathbf{R}_{AB}\sum_{k\in A}\sum_{l\in B} S_{kl}P_{lk} - \sum_A\sum_{k\in A}\sum_{B\neq A}\sum_{l\in B} P_{lk}(\chi_k|\mathbf{r}_{AB}|\chi_l).\end{aligned} \tag{4.36}$$

After adding the dipole moment of the nuclei we obtain

$$\boldsymbol{\mu} = \sum_A \boldsymbol{\mu}_A + \sum_A\sum_{B\neq A}\boldsymbol{\mu}_{AB}, \tag{4.37}$$

where

$$\boldsymbol{\mu}_A = \mathbf{R}_A(Z_A - \sum_{k\in A}\sum_{l\in A} S_{kl}P_{lk}) - \sum_{k\in A}\sum_{l\in A} P_{lk}(\chi_k|\mathbf{r}_A|\chi_l),$$

[23] We use the LCAO notation in the form $\varphi_i = \sum_A \sum_{k\in A} c_{ki}\chi_k$.

$$\boldsymbol{\mu}_{AB} = -\mathbf{R}_{AB} \sum_{k\in A}\sum_{l\in B} S_{kl} P_{lk} - \sum_{k\in A}\sum_{l\in B} P_{lk}(\chi_k|\mathbf{r}_{AB}|\chi_l).$$

We therefore have a quite interesting result.[24]

> The molecular dipole moment can be represented as the sum of the individual atomic dipole moments and the pairwise atomic dipole contributions.

The P_{lk} is large when k and l belong to the atoms forming a *chemical bond* (if compared to two nonbonded atoms, see Appendix D, p. 595); therefore the dipole moments related to *pairs* of atoms come practically uniquely from *chemical bonds*. The contribution of the lone pairs of atom A is hidden in the second term of $\boldsymbol{\mu}_A$ and may be quite large (cf. Appendix V1-T on p. V1-741).

4.3.4 Within the zero-differential overlap approximation

In several semiempirical methods of quantum chemistry (e.g., in the Hückel method) we assume the zero-differential overlap (ZDO) approximation, i.e., that $\chi_k \chi_l \approx (\chi_k)^2 \delta_{kl}$, and hence the second terms in $\boldsymbol{\mu}_A$ as well as in $\boldsymbol{\mu}_{AB}$ are equal to zero,[25] and therefore

$$\boldsymbol{\mu} = \sum_A \mathbf{R}_A (Z_A - \sum_{k\in A} P_{kk}) = \sum_A \mathbf{R}_A Q_A, \tag{4.38}$$

where $Q_A = (Z_A - \sum_{k\in A} P_{kk})$ represents the net electric charge of atom[26] A. This result is extremely simple: the dipole moment comes from the atomic charges only.

4.4 How to calculate the dipole polarizability

We have a formal expansion (4.24) for energy involving the dipole polarizability, but we need to calculate this expansion to be able to write the formula for computing $\alpha_{qq'}$.

---

[24] This does not represent a unique partitioning, only the total dipole moment should remain the same. For example, the individual atomic contributions include the lone pairs, which otherwise could be counted as a separate lone pair contribution.

[25] The second term in $\boldsymbol{\mu}_A$ is equal to zero, because the integrands $\chi_k^2 x$, $\chi_k^2 y$, $\chi_k^2 z$ are all antisymmetric with respect to transformation of the coordinate system, $x \to -x$, $y \to -y$, $z \to -z$.

[26] The molecule stays neutral. Indeed, $\sum_A \sum_{k\in A} P_{kk} = \sum_A \sum_{k\in A} \sum_i n_i c_{ki}^* c_{ki} = \sum_A \sum_{k\in A} \sum_i n_i |c_{ki}|^2 \approx \sum_i n_i = N$, where we have consequently used the ZDO approximation.

4.4.1 Sum over states (SOS) method

Perturbation theory gives the energy of the ground state $|0\rangle$ in a weak electric field as (the sum of the zero-, first-, and second-order energies,[27] see Chapter V1-5, the prime in the summation means excluding $n = 0$)

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} + \langle 0|\hat{H}^{(1)}|0\rangle + \sum_n{}' \frac{|\langle 0|\hat{H}^{(1)}|n\rangle|^2}{E_0^{(0)} - E_n^{(0)}} + \ldots. \tag{4.39}$$

If we assume a homogeneous electric field (see Eq. (4.11)), the perturbation is equal to $\hat{H}^{(1)} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}$, and we obtain

$$E = E^{(0)} - \langle 0|\hat{\boldsymbol{\mu}}|0\rangle \cdot \boldsymbol{\mathcal{E}} + \sum_n{}' \frac{[\langle 0|\hat{\boldsymbol{\mu}}|n\rangle \cdot \boldsymbol{\mathcal{E}}][\langle n|\hat{\boldsymbol{\mu}}|0\rangle \cdot \boldsymbol{\mathcal{E}}]}{E^{(0)} - E_n^{(0)}} + \ldots. \tag{4.40}$$

The first term represents the energy of the unperturbed molecule, the second term is a correction for the interaction of the permanent dipole moment with the field. The next term already takes into account that not only the permanent dipole moment but also the *induced moment* interacts with the electric field (Eq. (4.19)):

$$\sum_n{}' \frac{[\langle 0|\hat{\boldsymbol{\mu}}|n\rangle \cdot \boldsymbol{\mathcal{E}}][\langle n|\hat{\boldsymbol{\mu}}|0\rangle \cdot \boldsymbol{\mathcal{E}}]}{E^{(0)} - E_n^{(0)}} = -\frac{1}{2}\sum_{qq'} \alpha_{qq'}\mathcal{E}_q\mathcal{E}_{q'}, \tag{4.41}$$

where the component qq' of the polarizability is equal to

$$\alpha_{qq'} = 2\sum_n{}' \frac{\langle 0|\hat{\mu}_q|n\rangle\langle n|\hat{\mu}_{q'}|0\rangle}{\Delta_n}, \tag{4.42}$$

where $\Delta_n = E_n^{(0)} - E^{(0)}$. The polarizability has the dimension of a volume.[28]

Similarly, we may obtain the perturbational expressions for the dipole, quadrupole, octupole, etc., hyperpolarizabilities. For example, the ground-state dipole hyperpolarizability β_0 has the form (in the $qq'q''$component, the prime means that the ground state is omitted, we skip the derivation)

$$\beta_{qq'q''} = \sum_{n,m}{}' \frac{\langle 0|\hat{\mu}_q|n\rangle\langle n|\hat{\mu}_{q'}|m\rangle\langle m|\mu_{q''}|0\rangle}{\Delta_n\Delta_m} - \langle 0|\mu_q|0\rangle \sum_n{}' \frac{\langle 0|\hat{\mu}_{q'}|n\rangle\langle n|\hat{\mu}_{q''}|0\rangle}{(\Delta_n)^2}. \tag{4.43}$$

[27] Prime in the summation means that the 0-th state is excluded.

[28] Because μ^2 has the dimension of charge^2 × length^2, and Δ_n has the dimension of energy as for example in Coulombic energy, charge^2/length.

A problem with the SOS method is its slow convergence and the fact that whenever the expansion functions do not cover the energy continuum the result is incomplete.

Example 1 (The hydrogen atom in an electric field – perturbational approach). An atom or molecule, when located in an electric field, undergoes a *deformation*. We will show this in detail, taking the example of the hydrogen atom.

First, let us introduce a Cartesian coordinate system, within which the whole event will be described. Let the electric field be directed towards your right, i.e., has the form $\boldsymbol{\mathcal{E}} = (\mathcal{E}, 0, 0)$, with a constant $\mathcal{E} > 0$. The positive value of $\mathcal{E}$ means, according to the definition of electric field intensity, that a positive unit charge would move along $\boldsymbol{\mathcal{E}}$, i.e., from left to right. Thus, the anode is on your left and the cathode on your right.

We will consider a weak electric field; therefore, perturbation theory is applicable and this means just small corrections to the unperturbed situation. In our case the first-order correction to the wave function (Eq. (V1-5.26)) will be expanded in the series of hydrogen atomic orbitals (see Chapter V1-4, they form the complete set,[29] cf. Chapter V1-5), i.e.,

$$\psi_0^{(1)} = \sum_{k(\neq 0)} \frac{\left\langle \psi_k^{(0)} | \hat{H}^{(1)} | 1s \right\rangle}{E_0^{(0)} - E_k^{(0)}} \psi_k^{(0)}, \tag{4.44}$$

where $\psi_k^{(0)} \equiv |nlm\rangle$ and $E_k^{(0)} = -\frac{1}{2n^2}$ denote the orbitals and energies (in a.u.) of the isolated hydrogen atom, respectively (the unperturbed state $|nlm\rangle = |100\rangle = 1s$), and $\hat{H}^{(1)}$ is the perturbation, which for a homogeneous electric field has the form $\hat{H}^{(1)} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} = -\hat{\mu}_x \mathcal{E}$, with $\hat{\mu}_x$ standing for the dipole moment operator (its x component). The operator, according to Eq. (4.31), represents the following sum of products: charge (in our case of the electron or proton) times the x coordinate of the corresponding particle (let us denote them x and X, respectively), i.e., $\hat{\mu}_x = -x + X$, where the atomic units have been assumed. To keep the expression as simple as possible, let us locate the proton at the origin of the coordinate system,[30] i.e., $X = 0$. Finally, $\hat{H}^{(1)} = x\mathcal{E}$. Thus the perturbation $\hat{H}^{(1)}$ is simply proportional to the x coordinate of the electron.

In order not to work in vain, let us first check which unperturbed states k will contribute to the summation on the right-hand side of (4.44). The ground state ($k = 0$), i.e., the $1s$ orbital, is excluded (by the perturbation theory); next, $k = 1, 2, 3, 4$ denote the orbitals $2s, 2p_x, 2p_y, 2p_z$.

[29] Still they do not span the continuum.

[30] The proton might be located anywhere. The result does not depend on this choice, because the perturbation operators will differ by a constant. This, however, means that the nominator $\left\langle \psi_k^{(0)} \middle| \hat{H}^{(1)} | 1s \right\rangle$ in the formula will remain unchanged, because $\left\langle \psi_k^{(0)} | 1s \right\rangle = 0$.

The contribution of the $2s$ is equal to zero, because $\langle 2s|\hat{H}^{(1)}|1s\rangle = 0$ due to the antisymmetry of the integrand with respect to reflection $x \to -x$ ($\hat{H}^{(1)}$ changes its sign, while the orbitals $1s$ and $2s$ do not). A similar argument excludes the $2p_y$ and $2p_z$ orbitals. Hence, for the time being we have only a single candidate[31] $2p_x$. This time the integral is not zero and we will calculate it in a minute. If the candidates from the next shell ($n = 3$) are considered, similarly, the only nonzero contribution comes from $3p_x$. We will however stop our calculation at $n = 2$, because our goal is only to show how the whole machinery works. Thus, we need to calculate $\frac{\langle 2p_x|\hat{H}^{(1)}|1s\rangle}{E_0^{(0)}-E_1^{(0)}} = \frac{\langle 2p_x|x|1s\rangle}{E_0^{(0)}-E_1^{(0)}}\mathcal{E}$. The denominator is equal to $-1/2 + 1/8 = -3/8$ a.u. Calculation of the integral (fast exercise for students[32]) gives 0.7449 a.u. At $\mathcal{E} = 0.001$ a.u. we obtain the coefficient -0.001986 at the normalized orbital $2p_x$ in the first-order correction to the wave function. The negative value of the coefficient means that the orbital $-0.001986(2p_x)$ has its positive lobe oriented leftward.[33] The small absolute value of the coefficient results in such a tiny modification of the $1s$ orbital after the electric field is applied, that it will be practically invisible in Fig. 4.6. In order to make the deformation visible, let us use $\mathcal{E} = 0.1$ a.u. Then the admixture of $2p_x$ is equal to $-0.1986(2p_x)$, i.e., an approximate wave function of the hydrogen atom has the form $1s - 0.19862p_x$. Fig. 4.6 shows the unperturbed and the perturbed $1s$ orbital. As seen, the deformation makes an egg shape of the wave function (from a spherical one) – the electron is pulled towards the anode.[34] This is what we expected. Higher expansion functions ($3p_x, 4p_x, ...$) would change the shape of the wave function only a little.

Just *en passant* we may calculate a crude approximation to the dipole polarizability α_{xx}. From (4.42) we have

$$\alpha_{xx} \cong \frac{16}{3}\langle 2p_x|x(1s)\rangle^2 = \frac{16}{3}(0.7449)^2 = 2.96 \text{ a.u.}$$

The exact (nonrelativistic) result is $\alpha_{xx} = 4.5$ a.u. This shows that the number we have received is somewhat off, but after recalling that only a single expansion function has been used (instead of infinity of them) we should be quite happy with our result.[35]

---

[31] Note how fast our computation of the integrals proceeds. The main job (zero or not zero – that is the question) is done by the group theory.

[32] From p. V1-234 we have $\langle 2p_x\,|\,x\,|1s\rangle = \frac{1}{4\pi\sqrt{2}}\int_0^\infty \mathrm{d}r r^4 \exp\left(-\frac{3}{2}r\right)\int_0^\pi \mathrm{d}\theta \sin^3\theta \int_0^{2\pi}\mathrm{d}\phi \cos^2\phi = \frac{1}{4\pi\sqrt{2}}4!\left(\frac{3}{2}\right)^{-5}\frac{4}{3}\pi = 0.7449$, where we have used the formula $\int_0^\infty x^n \exp(-\alpha x)\,dx = n!\alpha^{-(n+1)}$ to calculate the integral over r.

[33] $2p_x \equiv x \times$ the positive spherically symmetric factor means the positive lobe of the $2p_x$ orbital is on your right (i.e., on the positive part of the x axis).

[34] This "pulling" results from adding together $1s$ and (with a negative coefficient) $2p_x$, i.e., we decrease the probability amplitude on the right-hand side of the nucleus, and increase it on the left-hand side.

[35] Such a situation is quite typical in the practice of quantum chemistry: the first terms of expansions give a lot, while the following ones give less and less, the total result approaching, with more and more pain, its limit. Note

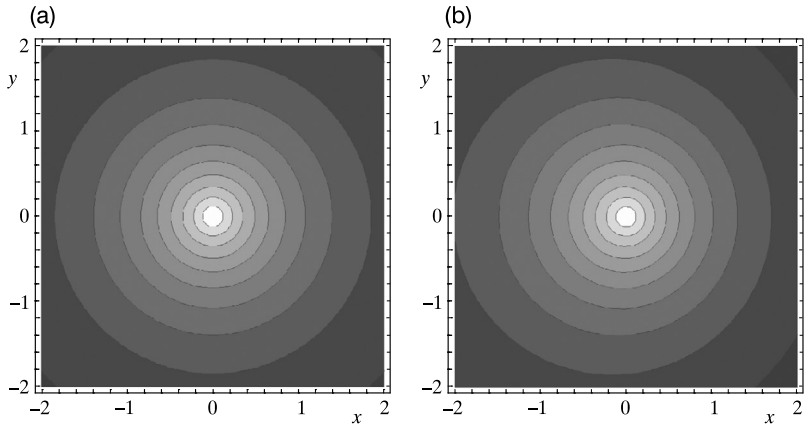


Fig. 4.6. Polarization of the hydrogen atom in an electric field. The wave function for (a) the unperturbed atom and (b) the atom in an electric field (a.u.) $\mathcal{E} = (0.1, 0, 0)$. As we can see, there are differences in the corresponding electronic density distributions: in the second case the wave function is deformed towards the anode (i.e., leftwards). Please note that the wave function is less deformed in the region close to the nucleus than in its left or right neighborhood. This is a consequence of the fact that the deformation is made by the $-0.1986(2p_x)$ function. Its main role is to subtract on the right and add on the left, and the smallest changes are at the nucleus, because $2p_x$ has its node there.

4.4.2 Finite field method

One may solve the Schrödinger equation including the term $-\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}$ in the Hamiltonian. The solution is valid then for this particular $\boldsymbol{\mathcal{E}}$. This procedure is known as the finite field method.

Example 2 (Hydrogen atom in electric field – variational approach). The polarizability of the hydrogen atom may be also computed by using a variational method (Chapter V1-5), in which the variational wave function $\psi = \chi_1 + c\chi_2$ is the $\chi_1 \equiv 1s$ plus an admixture (this is controlled by a variational parameter c) of the p-type orbital χ_2 with a certain exponential coefficient ζ (Ritz method of Chapter V1-5); see Appendix F, Eq. (F.1). As seen from Eq. (F.4), if χ_2 is taken as the $2p_x$ orbital (i.e., $\zeta = \frac{1}{2}$) we obtain $\alpha_{xx} = 2.96$ a.u., the same number we have already obtained by the perturbational method. However, if we take $\zeta = 1$, i.e., the same as in hydrogenic orbital $1s$, we will obtain $\alpha_{xx} = 4$ a.u. Well, a substantial improvement.

Is it possible to obtain even better results with the variational function ψ? Yes, it is. If we use the finite field method (with the electric field equal to $\mathcal{E}$ =0.01 a.u.), we will obtain[36] the minimum of E of Eq. (F.3) as corresponding to $\zeta_{opt} = 0.797224$. If we insert $\zeta = \zeta_{opt}$ into Eq. (F.4), we

that in the present case all terms are of the same sign, and we obtain better and better approximations when the expansion becomes longer and longer.

36 You may use *Mathematica* and the command FindMinimum[E,{ζ, 1}].

will obtain 4.475 a.u., only 0.5% off the exact result. This nearly perfect result is computed with a single correction function![37]

Sadlej relation – electric field-variant orbitals

In order to compute accurate values of $E(\boldsymbol{\mathcal{E}})$ extended LCAO expansions have to be used. Andrzej Sadlej[38] noticed that this huge numerical task in fact only takes into account a very simple effect: just a kind of *shift*[39] of the electronic charge distribution towards the anode. Since the atomic orbitals are usually centered on the nuclei and the electronic charge distribution shifts, to compensate for this still using the on-nuclei atomic orbitals requires monstrous and expensive LCAO expansions.

In LCAO calculations nowadays we most often use Gaussian-type orbitals (GTOs, see Chapter V1-8). They are rarely remembered as representing wave functions of the harmonic oscillator (cf. Chapter V1-4, here what oscillates is an electron), which they really do.[40] What would happen if an electron described by a GTO were subject to the electric field $\boldsymbol{\mathcal{E}}$?

> Sadlej noticed that the Gaussian-type orbital will change in a similar way to the wave functions of a charged harmonic oscillator in an electric field; these, however, simply shift.

Indeed, this can be shown as follows. The Schrödinger equation for the harmonic oscillator (here: an electron with $m = 1$ in a.u., its position is x) without any electric field is given on p. V1-217. According to the example of the hydrogen atom in an electric field, the Schrödinger equation for an electron oscillating in a homogeneous electric field $\mathcal{E} > 0$ has the form (k is a force constant)

$$\left(-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}kx^2 + \mathcal{E}x\right)\psi(x,\mathcal{E}) = E(\mathcal{E})\psi(x,\mathcal{E}). \tag{4.45}$$

Now, let us find such constants a and b that

---

[37] This success means that sometimes long expansions in the Ritz method may result from an unfortunate choice of expansion functions.

[38] A.J. Sadlej, *Chem. Phys. Letters*, 47(1977)50; A.J. Sadlej, *Acta Phys. Polon. A*, 53(1978)297.

[39] With a deformation.

[40] At least if they represent the 1s GTOs.

$$\frac{1}{2}kx^2 + \mathcal{E}x = \frac{1}{2}k(x-a)^2 + b. \tag{4.46}$$

We immediately get $a = -\mathcal{E}/k$, $b = -\frac{1}{2}ka^2$. The constant b is irrelevant, since it only shifts the zero on the energy scale. Thus,

> the solution to a charged harmonic oscillator (oscillating electron) in a homogeneous electric field represents the same function as without the field, but shifted by $-\frac{\mathcal{E}}{k}$.

Indeed, inserting $x' = x + \frac{\mathcal{E}}{k}$ leads to $\mathrm{d}/\mathrm{d}x = \mathrm{d}/\mathrm{d}x'$ and $\mathrm{d}^2/\mathrm{d}x^2 = \mathrm{d}^2/\mathrm{d}x'^2$, which gives a similar Schrödinger equation except that the harmonic potential is shifted. Therefore, the solution to the equation can be written as simply a zero-field solution shifted by $-\frac{\mathcal{E}}{k}$, i.e., $\psi(x') = \psi(x + \frac{\mathcal{E}}{k})$. This is quite understandable, because the operation means nothing but adding to the parabolic potential energy $kx^2/2$ a term proportional to x, i.e., a parabola potential again (Fig. 4.7b).

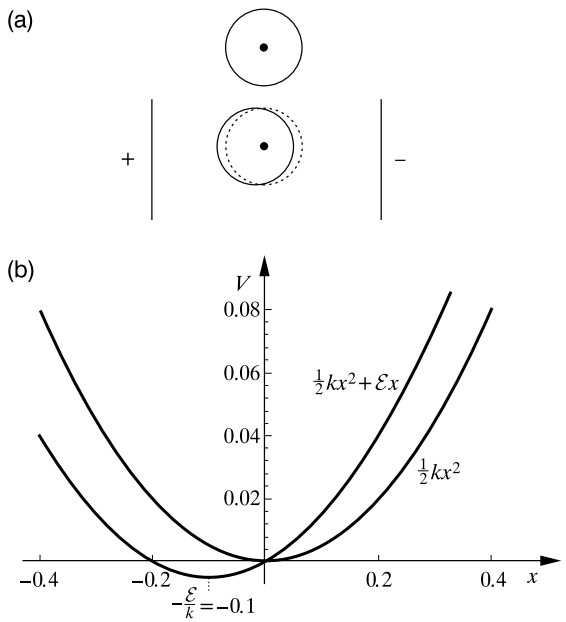
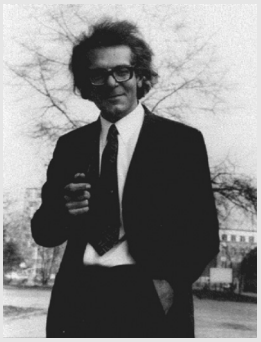


Fig. 4.7. Sadlej relation. (a) The electric field mainly causes a shift of the electronic charge distribution towards the anode. A Gaussian-type orbital represents the eigenfunction of a harmonic oscillator. (b) In a crude approximation let us suppose an electron oscillates in a parabolic potential energy well (with the force constant k). In this situation a homogeneous electric field $\mathcal{E}$ corresponds to the perturbation $\mathcal{E}x$, that *conserves the harmonicity with unchanged force constant* k.

Andrzej Jerzy Sadlej (1941–2010), Polish quantum chemist, the only person I know who wrote a scientific book while still being an undergraduate ("Elementary methods of quantum chemistry," PWN, Warsaw, 1966), the first book on quantum chemistry in Polish language. After half a century the book defends itself by its competence. The book you are keeping in your hands owes very much to friendship of the author with Andrzej. He did not expect any acknowledgement, his only concern was science.



To see how this displacement depends on the GTO *exponent*, let us recall its relation to the harmonic oscillator force constant k (cf. p. V1-217). The harmonic oscillator eigenfunction corresponds to a Gaussian orbital with an exponent equal to $\alpha/2$, where $\alpha^2 = k$ (in a.u.). Therefore, if we have a GTO with exponent equal to A, this means the corresponding harmonic oscillator has the force constant $k = 4A^2$. Now, if the homogeneous electric field $\mathcal{E}$ is switched on, the center of this atomic orbital has to move by $\Delta(A) = -\mathcal{E}/k = -\frac{1}{4}\mathcal{E}/A^2$. This means all the atomic orbitals have to move opposite to the applied electric field (as expected) and the displacement of the orbital is small if its exponent is large, and *vice versa*. Also, if the atomic electron charge distribution results from several GTOs (as in the LCAO expansion) it deforms in the electric field in such a way that the diffuse orbitals shift more, while the compact ones (with large exponents) shift only a little. Altogether this does not mean just a simple shift of the electronic charge density, but instead its shift accompanied by a deformation. On the other hand, we may simply optimize the GTO positions within the finite field Hartree–Fock method, and check whether the corresponding shifts $\Delta_{opt}(A)$ indeed follow the Sadlej relation.[41]

It turns out that the relation $\Delta_{opt}(A) \sim -E/A^2$ is satisfied to a good accuracy,[42] despite the fact that the potential energy in an atom does not represent that of a harmonic oscillator.

[41] We have tacitly assumed that in the unperturbed molecule the atomic orbitals occupy optimal positions. This assumption may sometimes cause trouble. If the centers of the atomic orbitals in an isolated molecule are nonoptimized, we may end up with a kind of antipolarizability: we apply the electric field and when the atomic orbital centers are optimized, the electron cloud moves opposite to that which we expect. This is possible only because in such a case the orbital centers mainly follow the strong intramolecular electric field, rather than the much weaker external field $\boldsymbol{\mathcal{E}}$ (J.M. André, J. Delhalle, J.G. Fripiat, G. Hennico, L. Piela, *Intern. J. Quantum Chem.*, 22S(1988)665).

[42] This is how the *electric field-variant orbitals* (EFVOs) were born: Andrzej's colleagues did not believe in this simple recipe for calculating polarizabilities, but they lost the bet (a bar of chocolate).

The electrostatic catastrophe of the theory

There is a serious problem in finite field theory. If even the *weakest* homogeneous electric field is applied and a very good basis set is used, we are *bound* to have some kind of catastrophe. A nasty word, but unfortunately reflecting quite adequately a mathematical horror we are going to be exposed to after adding to the Hamiltonian operator $\hat{H}^{(1)} = x\mathcal{E}$ with an electric field (here x symbolizes the component of the dipole moment).[43] The problem is that this operator is *unbound*, i.e., for a normalized trial function ϕ the integral $\langle\phi|\hat{H}^{(1)}\phi\rangle$ may attain ∞ or $-\infty$. Indeed, by gradually shifting the function towards the negative values of the x axis, we obtain more and more negative values of the integral, and for $x = -\infty$ we get $\langle\phi|\hat{H}^{(1)}\phi\rangle = -\infty$. In other words,

> when using atomic orbitals centered far from the nuclei in the region of the negative x (or allowing optimization of the orbital centers with the field switched on), we will lower the energy to $-\infty$, i.e., we obtain a catastrophe. This is quite understandable, because such a system (electrons separated from the nuclei and shifted far away along the x axis) has a huge dipole moment, and therefore *very low energy*, when interacting with the field.

Suppose calculations for a molecule in an electric field $\mathcal{E}$ are carried out. According to the Sadlej relation, we shift the corresponding atomic orbitals proportionally to $-\eta\mathcal{E}/A^2$, with $\eta > 0$, and the energy goes down. Around $\eta = \frac{1}{4}$, which according to Sadlej corresponds to optimal shifts,[44] we may expect the lowest energy. Then, for larger η, the energy has to go up. What if we continue to increase (Fig. 4.8) the shift parameter η?

The energy increase will continue only up to some critical value of η. Then, according to the discussion above, the energy will fall to $-\infty$, i.e., to a catastrophe. Thus the energy curve exhibits a barrier (Fig. 4.8) that is related to the basis set quality (its "saturation"): a poor basis means a high barrier, the ideal basis, i.e., the complete basis set, gives no barrier at all, just falling into the abyss with the polarizability going to infinity, etc. Therefore, rather paradoxically, reliable values of polarizability are to be obtained using a medium-quality basis set. An improvement of the basis will lead to worse results.[45]

This pertains to variational calculations. What about the perturbational method? In the first- and second-order corrections to the energy, the formulae contain the zero-order approximation

[43] The most dramatic form of the problem would appear if the finite field method were combined with the numerical solution of the Schrödinger or Fock equation.

[44] They are optimal for a parabolic potential.

[45] Once more in this book: wealth does not necessarily improve life.

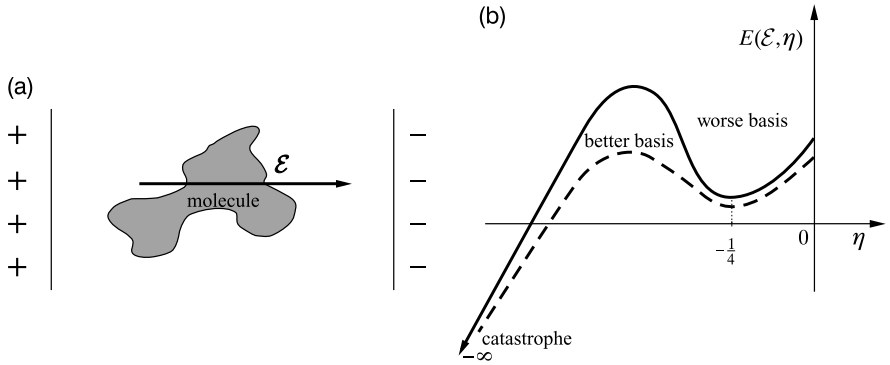


Fig. 4.8. (a) A molecule in a homogeneous electric field. (b) The term η is a parameter describing the shift of the Gaussian atomic orbitals along the electric field, with $\eta = 0$ showing the centering on the nuclei. The total energy $E(\mathcal{E}, x)$ is a function of the electric field intensity $\mathcal{E}$ and the basis set shift parameter η. Optimization of η gives a result close to the Sadlej value $\eta = \frac{1}{4}$; larger η values first lead to an increase of E, but then to a decrease towards a catastrophe: $lim_{x\to-\infty} E(\mathcal{E}, x) = -\infty$.

to the wave function $\psi_0^{(0)}$, e.g., $E^{(2)} = \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle$. If the origin of the coordinate system is located on the molecule, then the exponential decay of $\psi_0^{(0)}$ forces the first-order correction to the wave function $\psi_0^{(1)}$ to be also localized close to the origin, otherwise it would tend to zero through the shifting towards the negative values of x (this prevents the integral diverging to $-\infty$). However, the third-order correction to the energy contains the term $\langle\psi_0^{(1)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle$, which may already go to $-\infty$. Hence, the perturbation theory also carries the seed of future electrostatic catastrophe.

4.4.3 What is going on at higher electric fields

Polarization

The theory described so far is applicable only when the electric field intensity is small. Such a field can mainly polarize (a small deformation) the electronic charge distribution. More fascinating phenomena begin when the electric field gets stronger.

Deformation

Of course, the equilibrium configurations of the molecule with and without an electric field differ. In a simple case, say, the HCl molecule, the HCl distance increases. It has to increase, since the cathode pulls on the hydrogen atom and repels the chlorine atom, while the anode does the opposite. In more complex cases, like a flexible molecule, the field may change its conformation. This means that the polarizability results both from the electron cloud deformation and

the displacement of the nuclei. It turns out that the latter effect (called vibrational polarization) is of great importance.[46]

Dissociation

When the electric field gets stronger the molecule may dissociate into ions. To this end, the external electric field intensity has to become comparable to the electric field produced by the molecule itself in its neighborhood. The intramolecular electric fields are huge, the intermolecular ones are weaker but also very large (of the order of 10^8 V/m), much larger than those offered by current technical installations. No wonder then that the molecules may interact to such an extent that they may even undergo chemical reactions. When the interaction is weaker, the electric fields produced by molecules may lead to intermolecular complexes. Many beautiful examples may be found in biochemistry (see Chapters 5 and 7). A strong external electric field applied to a crystal may cause a cascade of processes, e.g., the so-called *displacive phase transitions*, when sudden displacements of atoms occur, and a new crystal structure appears.

Destruction

A sufficiently strong electric field will destroy the molecules through their ionization. The resulting ions accelerate in the field, collide with the molecules, and ionize them even more (these phenomena are accompanied by light emission as in vacuum tubes). Such processes may lead to the final decomposition of the system (plasma) with the electrons and the nuclei finally reaching the anode and cathode. We will have a vacuum.

Creation!

Let us keep increasing the electric field applied to the vacuum. Will anything interesting be going on? We know from Chapter V1-3 that when huge electric field intensities are applied (of the order of the electric field intensity in the vicinity of a proton – infeasible for the time being), then *the particles and antiparticles will leap off the vacuum*! The vacuum is not just nothing.

4.5 A molecule in an oscillating electric field

Constant and oscillating components

A nonzero hyperpolarizability indicates a *nonlinear* response (dipole moment proportional to the second and higher powers of the field intensity). This may mean an "inflated" reaction to

[46] J.-M. André, B. Champagne, in "Conjugated Oligomers, Polymers, and Dendrimers: From Polyacetylene to DNA," J.L. Brédas (ed.), Bibliothéque Scientifique Francqui, De Boeck Université, p. 349.

the applied field, a highly desired feature for contemporary optoelectronic materials. One such reaction is the second and third harmonic generation (SHG and THG, respectively), where light of frequency ω generates in a material light with frequencies 2ω and 3ω, respectively. A simple statement about why this may happen is shown below.[47]

Let us imagine a molecule immobilized in a laboratory coordinate system (like in an oriented crystal). Let us switch on a homogeneous electric field $\boldsymbol{\mathcal{E}}$, which has two components, i.e., a static component $\boldsymbol{\mathcal{E}}^0$ and an oscillating one $\boldsymbol{\mathcal{E}}^\omega$ with frequency ω. Then we have

$$\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{E}}^0 + \boldsymbol{\mathcal{E}}^\omega \cos(\omega t)\,. \tag{4.47}$$

We may imagine various experiments here: the steady field along x, y, or z and a light beam polarized along x, y, or z, we may also vary ω for each beam, etc. Such choices lead to a rich set of nonlinear optical phenomena.[48] What will the reaction of the molecule be in such an experiment? Let us see.[49]

Induced dipole moment

The total dipole moment of the molecule (i.e., the permanent moment $\boldsymbol{\mu}_0$ plus the induced moment $\boldsymbol{\mu}_{ind}$) will depend on time, because $\boldsymbol{\mu}_{ind}$ does. We have

$$\mu_q(t) = \mu_{0,q} + \mu_{ind,q}, \tag{4.48}$$

$$\mu_{ind,q}(t) = \sum_{q'} \alpha_{qq'}\mathcal{E}_{q'} + \frac{1}{2}\sum_{q'q''} \beta_{qq'q''}\mathcal{E}_{q'}\mathcal{E}_{q''} + \frac{1}{6}\sum_{q',q'',q'''} \gamma_{qq'q''q'''}\mathcal{E}_{q'}\mathcal{E}_{q''}\mathcal{E}_{q'''} + ... \tag{4.49}$$

Therefore, if we insert $\mathcal{E}_q = \mathcal{E}_q^0 + \mathcal{E}_q^\omega \cos(\omega t)$ as the electric field component for $q = x, y, z$, we obtain

$$\mu_q(t) = \mu_{0,q} + \sum_{q'} \alpha_{qq'}[\mathcal{E}_{q'}^0 + \mathcal{E}_{q'}^\omega \cos(\omega t)] + \frac{1}{2}\sum_{q'q''} \beta_{qq'q''}[\mathcal{E}_{q'}^0 + \mathcal{E}_{q'}^\omega \cos(\omega t)] \times [\mathcal{E}_{q''}^0 + \mathcal{E}_{q''}^\omega \cos(\omega t)] +$$

[47] The problem of how the polarizability changes as a function of inducing wave frequency is described in detail in J. Olsen, P. Jørgensen, *J. Chem. Phys.*, 82(1985)3235.

[48] S. Kielich, "*Molecular nonlinear optics*," Warszawa-Poznań, PWN, 1977.

[49] For the sake of simplicity we have used the same frequency and the same phases for the light polarized along x, y, and z.

$$\frac{1}{6}\sum_{q',q'',q'''}\gamma_{qq'q''q'''}[\mathcal{E}^0_{q'}+\mathcal{E}^\omega_{q'}\cos(\omega t)][\mathcal{E}^0_{q''}+\mathcal{E}^\omega_{q''}\cos(\omega t)]\times$$
$$[\mathcal{E}^0_{q'''}+\mathcal{E}^\omega_{q'''}\cos(\omega t)]+\ldots \tag{4.50}$$

Second (SHG) and third (THG) harmonic generation

After multiplication and simple trigonometry we have

$$\mu_q(t)=\mu_{\omega=0,q}+\mu_{\omega,q}\cos\omega t+\mu_{2\omega,q}\cos(2\omega t)+\mu_{3\omega,q}\cos(3\omega t)\,, \tag{4.51}$$

where the amplitudes μ corresponding to the coordinate $q\in x,y,z$ and to the particular resulting frequencies $0,\omega,2\omega,3\omega$ have the form given below. The polarizabilities and hyperpolarizabilities depend on the frequency ω and the direction of the incident light waves. According to a convention, a given (hyper)polarizability, e.g., $\gamma_{qq'q''q'''}(-3\omega;\omega,\omega,\omega)$, is characterized by the frequencies ω corresponding to the three directions x,y,z of the incident light polarization (preceded by minus the Fourier frequency of the term, -3ω, which symbolizes the photon energy conservation law). Some of the symbols, e.g., $\gamma_{qq'q''q'''}(-\omega;\omega,-\omega,\omega)$, after a semicolon have negative values, which means a partial (as in $\gamma_{qq'q''q'''}(-\omega;\omega,-\omega,\omega)$) or complete (as in $\beta_{q,q',q''}(0;-\omega,\omega)$) cancellation of the intensity of the oscillating electric field. The formulae for the amplitudes are

$$\mu_{\omega=0,q}=\mu_{0,q}+\sum_{q'}\alpha_{qq'}(0;0)\,\mathcal{E}^0_{q'}+\frac{1}{2}\sum_{q',q''}\beta_{qq'q''}(0;0,0)\mathcal{E}^0_{q'}\mathcal{E}^0_{q''}+$$
$$\frac{1}{6}\sum_{q',q'',q'''}\gamma_{qq'q''q'''}(0;0,0,0)\mathcal{E}^0_{q'}\mathcal{E}^0_{q''}\mathcal{E}^0_{q'''}+$$
$$\frac{1}{4}\sum_{q'q''}\beta_{q,q',q''}(0;-\omega,\omega)\mathcal{E}^\omega_{q'}\mathcal{E}^\omega_{q''}+\frac{1}{4}\sum_{q,q',q'',q'''}\gamma_{qq'q''q'''}(0;0,-\omega,\omega)\mathcal{E}^0_{q'}\mathcal{E}^\omega_{q''}\mathcal{E}^\omega_{q'''},$$

$$\mu_{\omega,q}=\sum_{q'}\alpha_{qq'}(-\omega;\omega)\,\mathcal{E}^\omega_{q'}+\sum_{q',q''}\beta_{qq'q''}(-\omega;\omega,0)\mathcal{E}^\omega_{q'}\mathcal{E}^0_{q''}+$$
$$\frac{1}{2}\sum_{q,q',q'',q'''}\gamma_{qq'q''q'''}(-\omega;\omega,0,0)\mathcal{E}^\omega_{q'}\mathcal{E}^0_{q''}\mathcal{E}^0_{q'''}+$$
$$\frac{1}{8}\sum\sum_{q,q',q'',q'''}\gamma_{qq'q''q'''}(-\omega;\omega,-\omega,\omega)\mathcal{E}^\omega_{q'}\mathcal{E}^\omega_{q''}\mathcal{E}^\omega_{q'''},$$

$$\mu_{2\omega,q} = \frac{1}{4}\sum_{q'q''}\beta_{q,q',q''}(-2\omega;\,\omega,\omega)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''} + \frac{1}{4}\sum\sum_{q,q',q'',q'''}\gamma_{qq'q''q'''}(-2\omega;\,\omega,\omega,0)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''}\mathcal{E}^{0}_{q'''}, \tag{4.52}$$

$$\mu_{3\omega,q} = \frac{1}{24}\sum_{q,q',q'',q'''}\gamma_{qq'q''q'''}(-3\omega;\,\omega,\omega,\omega)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''}\mathcal{E}^{\omega}_{q'''}. \tag{4.53}$$

We see that:

- An oscillating electric field may result in a nonoscillating dipole moment related to the hyperpolarizabilities $\beta_{q,q',q''}(0;\,-\omega,\omega)$ and $\gamma_{qq'q''q'''}(0;\,0,\,-\omega,\omega)$, which manifests as an electric potential difference on two opposite crystal faces.
- The dipole moment oscillates with the basic frequency ω of the incident light and in addition with two other frequencies, i.e., the second (2ω) and third (3ω) harmonics (SHG and THG, respectively). This is supported by experiment (mentioned in the example at the beginning of the chapter); applying incident light of frequency ω we obtain emitted light with frequencies[50] 2ω and 3ω.

Note that to generate a large second harmonic the material has to have large values of the hyperpolarizabilities β and γ. The THG needs a large γ. In both cases a strong laser electric field is a must. The SHG and THG therefore require support from the theoretical side: we are looking for high-hyperpolarizability materials and quantum mechanical calculations *before* an expensive organic synthesis is carried out.[51]

MAGNETIC PHENOMENA

The electric and magnetic fields (both of them are related by the Maxwell equations, Appendix V1-G) interact differently with matter, which is highlighted in Fig. 4.9, where the electron trajectories in both fields are shown. They are totally different; the trajectory in the magnetic field has a circle-like character, while in the electric field it is a parabola. This is why the description of magnetic properties differs so much from that of electric properties.

[50] This experiment was first carried out by P.A. Franken, A.E. Hill, C.W. Peters, G. Weinreich, *Phys. Rev. Letters*, 7(1961)118.

[51] In molecular crystals it is not sufficient that particular molecules have high values of hyperpolarizability. What counts is the hyperpolarizability of the crystal unit cell.

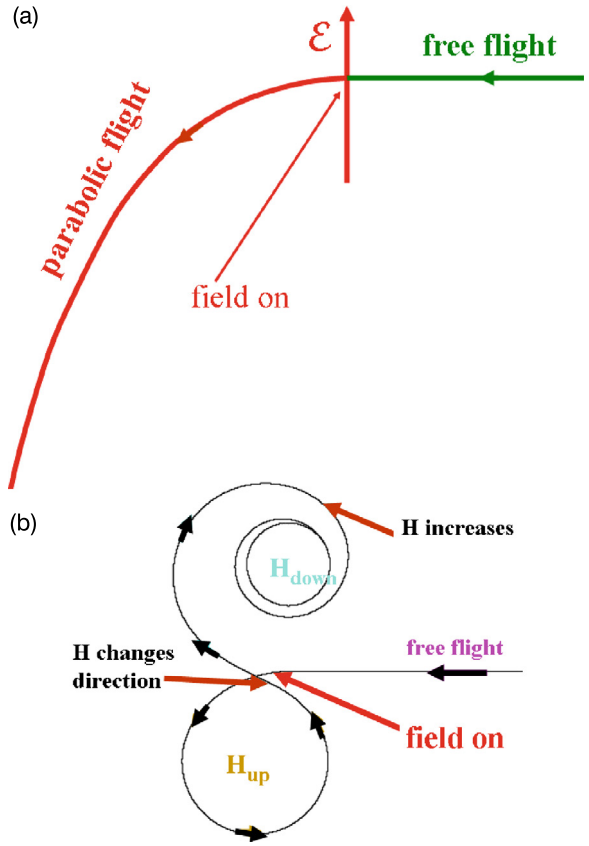


Fig. 4.9. Dramatic differences between electron trajectories in the electric and magnetic fields. (a) The uniform electric field $\mathcal{E}$ and the parabolic electron trajectory. (b) First the magnetic field is switched off, then it is set on (H_{up}) being perpendicular to the picture and oriented towards the reader. After the electron makes a single loop, the orientation of the magnetic field is inverted (H_{down}) and its intensity keeps increasing for a while. The latter causes the electron to make loops with smaller radius.

4.6 Magnetic dipole moments of elementary particles

4.6.1 Electron

An elementary particle, besides its orbital angular momentum, may also have internal angular momentum, or spin (cf. p. V1-33). In Chapter V1-3, the Dirac theory led to a relation between the *spin* angular momentum $\mathbf{s}$ of the electron and its *dipole magnetic moment* $\mathbf{M}_{spin,el}$ (Eq. (V1-3.63), p. V1-158):

$$\mathbf{M}_{spin,el} = \gamma_{el}\mathbf{s},$$

with the *gyromagnetic coefficient*[52]

$$\gamma_{el} = -2\frac{\mu_B}{\hbar},$$

where the Bohr magneton (m_0 is the electronic rest mass)

$$\mu_B = \frac{e\hbar}{2m_0 c}.$$

The gyromagnetic factor is twice as large as that appearing in the relation between the electron *orbital* angular momentum $\mathbf{L}$ and the associated magnetic dipole moment

$$\mathbf{M}_{orb,el} = -\frac{\mu_B}{\hbar}\mathbf{L}. \tag{4.54}$$

Quantum electrodynamics explains this effect much more precisely, predicting the factor very close to the experimental value[53] 2.0023193043737, known with the breathtaking accuracy ± 0.0000000000082.

4.6.2 Nucleus

Let us stay within the Dirac theory, as pertaining to a single elementary particle. If, instead of an electron, we take a *nucleus* of charge $+Ze$ and atomic mass[54] M, then we *would presume* (after insertion into the above formulae) the gyromagnetic factor should be $\gamma = -2\frac{\mu_{\text{nucl}}}{\hbar} = -2\frac{\frac{(-Ze)}{2Mm_H c}\hbar}{\hbar} = 2\frac{Z}{M}\frac{e\hbar}{2m_H c} = 2\frac{Z}{M}\frac{\mu_N}{\hbar}$, where $\mu_N = \frac{e\hbar}{2m_H c}$ (m_H denoting the proton mass) is known as the *nuclear magneton*.[55] For a proton ($Z = 1$, $M = 1$), we would have $\gamma_p = 2\mu_N/\hbar$, whereas the experimental value[56] is $\gamma_p = 5.59\mu_N/\hbar$. What is going on? In both cases we have a single elementary particle (electron or proton), both with the spin quantum number equal to $\frac{1}{2}$, so we might expect that nothing special will happen for the proton, and only the mass ratio and charge will make a difference. Instead we see that Dirac theory does pertain to the electron, but not to the nuclei. Visibly, the proton is more complex than the electron. We see that even the simplest nucleus has internal machinery, which results in the observed strange deviation.

---

52 From the Greek word *gyros,* or circle; it is believed that a circular motion of a charged particle is related to the resulting magnetic moment.

53 R.S. Van Dyck Jr., P.B. Schwinberg, H.G. Dehmelt, *Phys. Rev. Letters*, 59(1990)26.

54 Unitless quantity.

55 Ca. 1840 times smaller than the Bohr magneton (for the electron).

56 Also the gyromagnetic factor for an electron is expected to be ca. 1840 times larger than that for a proton. This means that a proton is expected to create a magnetic field ca. 1840 times weaker than the field created by an electron.

There are lots of quarks in the proton (three valence quarks and a sea of virtual quarks together with the gluons, etc.). The proton and electron polarize the vacuum differently and this results in different gyromagnetic factors. Other nuclei exhibit even stranger properties. Sometimes we even have negative gyromagnetic coefficients. In such a case their magnetic moment is opposite to the spin angular momentum. The complex nature of the internal machinery of the nuclei and vacuum polarization lead to the observed gyromagnetic coefficients.[57] Science has had some success here, e.g., for leptons,[58] but for nuclei the situation is worse. This is why we are simply forced to take this into account in the present book[59] and treat the *spin magnetic moments* of the nuclei as the experimental data:

$$\mathbf{M}_A = \gamma_A \mathbf{I}_A, \tag{4.55}$$

where $\mathbf{I}_A$ represents the spin angular momentum of nucleus A.

4.6.3 Dipole moment in the field

Electric field

The problem of an electric dipole μ rotating in an electric field was described on p. 272. When the field is switched off (cf. p. V1-229), the ground state is nondegenerate ($J = M = 0$) and represents a constant, while the excited states are all degenerate. After an electric field ($\boldsymbol{\mathcal{E}}$) is switched on, the ground-state wave function deforms in such a way as to prefer the alignment of the rotating dipole moment along the field, and – for the excited states – the degeneracy is lifted. Since we may always use a complete set of rigid rotator wave functions (at zero field), the deformed wave functions *have to be linear combinations of the wave functions corresponding to different* J.

Magnetic field

Imagine a spinning top such as children like to play with. If you make it spin (with angular momentum $\mathbf{I}$) and leave it in space *without any interaction*, then due to the fact that space is isotropic, its angular momentum will stay constant, i.e., the top will rotate about its axis with a constant speed and the axis will not move with respect to distant stars (Fig. 4.10a).

---

57 The relation between spin and magnetic moment is as mysterious as that between the magnetic moment and charge of a particle (the spin is associated with a rotation, while the magnetic moment is associated with a rotation of a charged object) or its mass. A neutron has spin equal to $\frac{1}{2}$ and magnetic moment similar to that of a proton, despite the zero electric charge. The neutrino has no charge, nearly zero mass and magnetic moment, and still spin equal to $\frac{1}{2}$.

58 And what about the "heavier brothers" of the electron, the muon and taon (cf. p. V1-386)? For the muon, the coefficient in the gyromagnetic factor (2.0023318920) is similar to that of the electron (2.0023193043737), just a bit larger, and agrees equally well with experiment. For the taon we have only a theoretical result, a little larger than the two other "brothers." Thus, the whole lepton family behaves hopefully in a similar way.

59 With a suitable respect of the Nature's complexity.

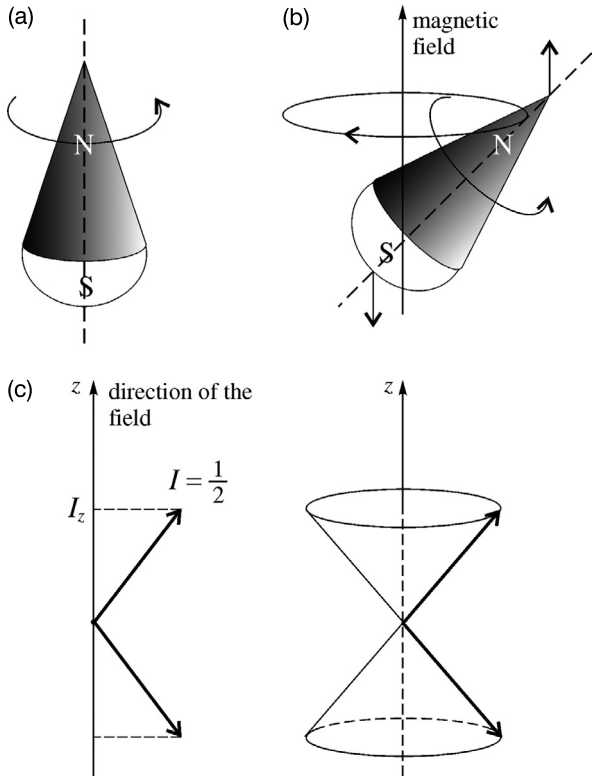


Fig. 4.10. Classical and quantum tops in space. (a) The space is isotropic and therefore the classical top preserves its angular momentum, i.e., its axis does not move with respect to distant stars and the top rotates about its axis with a constant speed. This behavior is used in gyroscopes that help to orient a space ship with respect to distant stars. (b) The same top in a homogeneous vector field. The space is no longer isotropic, and therefore the total angular momentum is no longer preserved. The projection of the total momentum on the field direction is still preserved. This is achieved by the precession of the top axis about the direction of the field. (c) A quantum top, i.e., an elementary particle with spin quantum number $I = \frac{1}{2}$ in the magnetic field. The projection I_z of its spin $\mathbf{I}$ is quantized, i.e., $I_z = m_I \hbar$ with $m_I = -\frac{1}{2}, +\frac{1}{2}$ and, therefore, we have two energy eigenstates that correspond to two precession cones, directed up and down.

The situation changes if a homogeneous vector field (e.g., a magnetic field) is switched on. Now, the space is no longer isotropic and the vector of the angular momentum is no longer conserved. However, the conservation law *for the projection of the angular momentum on the direction of the field* is still valid. This means that the top makes a precession about the field axis, because this is what keeps the projection constant (Fig. 4.10b). The magnetic dipole moment $\mathbf{M} = \gamma \mathbf{I}$ in the magnetic field $\mathbf{H} = (0, 0, H)$, $H > 0$, has as many stationary states as the number of possible projections of the spin angular momentum on the field direction. From Chapter V1-1,

we know that this number is $2I + 1$, where I is the spin quantum number of the particle (e.g., for a proton, $I = \frac{1}{2}$). The projections (Fig. 4.10c) are equal to $m_I\hbar$ with $m_I = -I, -I + 1, ...0, ... + I$. Therefore,

the energy levels in the magnetic field are equal to

$$E_{m_I} = -\gamma m_I \hbar H. \tag{4.56}$$

Note that the energy level splitting is proportional to the magnetic field intensity (Fig. 4.11).

If a nucleus has $I = \frac{1}{2}$, then the energy difference ΔE between the two states in a magnetic field H, one with $m_I = -\frac{1}{2}$ and the other one with $m_I = \frac{1}{2}$, equals $\Delta E = 2 \times \frac{1}{2}\gamma\hbar H = \gamma\hbar H$, and

$$\Delta E = h\nu_L, \tag{4.57}$$

where the Larmor[60] frequency is defined as

$$\nu_L = \frac{\gamma H}{2\pi}. \tag{4.58}$$

We see (Fig. 4.11) that for nuclei with $\gamma > 0$, lower energy corresponds to $m_I = \frac{1}{2}$, i.e., to the spin moment along the field (forming an angle $\theta = 54°44'$ with the magnetic field vector, see p. V1-33).

Note that

there is a difference between the energy levels of the electric dipole moment in an electric field and the levels of the magnetic dipole in a magnetic field. The difference is that for the magnetic dipole of an elementary particle the states do not have admixtures from the other I values (which are given by Nature), while for the electric dipole there are admixtures from states with other values of J.

This suggests that we may also expect such admixtures in the magnetic field. Anyway, it is at least true if the particle is complex. For example, the singlet state ($S = 0$) of the hydrogen molecule gets an admixture of the triplet state ($S = 1$) in the magnetic field, because the spin magnetic moments of both electrons tend to align parallel to the field.

60 Joseph Larmor (1857–1942), Irish physicist, professor at Cambridge University.

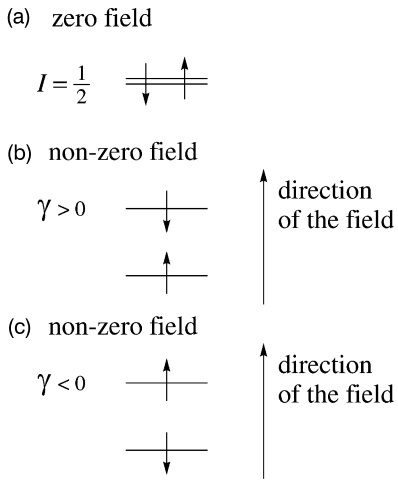


Fig. 4.11. Energy levels in magnetic field $\mathbf{H} = (0, 0, H)$ for a nucleus with spin angular momentum $\mathbf{I}$ corresponding to spin quantum number $I = \frac{1}{2}$. (a) The magnetic dipole moment is equal to $\mathbf{M} = \gamma\mathbf{I}$ and at the zero field the level is doubly degenerate. (b) For $\gamma > 0$ (e.g., a proton), $\mathbf{I}$ and $\mathbf{M}$ have the same direction. In a nonzero magnetic field the energy is equal to $E = -\mathbf{M}\cdot\mathbf{H} = -M_z H = -\gamma m_I \hbar H$, where $m_I = \pm\frac{1}{2}$. Thus, the degeneracy is lifted: the state with $m_I = \frac{1}{2}$, i.e., with the positive projection of $\mathbf{I}$ in the direction of the magnetic field has lower energy. (c) For $\gamma < 0$, $\mathbf{I}$ and $\mathbf{M}$ have the opposite direction. The state with $m_I = \frac{1}{2}$ has higher energy.

4.7 NMR spectra – transitions between the nuclear quantum states

Is there any possibility of making the nuclear spin flip from one quantum state to another? Yes. Evidently, we have to create distinct energy levels corresponding to different spin projections, i.e., to switch the magnetic field on (Figs. 4.11 and 4.12a). After the electromagnetic field is applied and its frequency matches the energy level difference, *the system* absorbs the energy. It looks as if *a nucleus* absorbs the energy and changes its quantum state. In a genuine NMR experiment, the electromagnetic frequency is fixed (radio wave lengths) and the specimen is scanned by a variable magnetic field. At some particular field values the energy difference matches the electromagnetic frequency and the transition (NMR) is observed.

The magnetic field a particular nucleus feels differs from the external magnetic field applied, because the electronic structure in which the nucleus is immersed makes its own contribution (Fig. 4.12b,c). Also the nuclear spins interact by creating their own magnetic fields.

We have not yet considered these effects in the nonrelativistic Hamiltonian (V1-2.1) on p. V1-76 (e.g., no spin–spin or spin–field interactions). The effects which we are now dealing with are so small (of the order of 10^{-11} kcal/mol) that they are of no importance for most

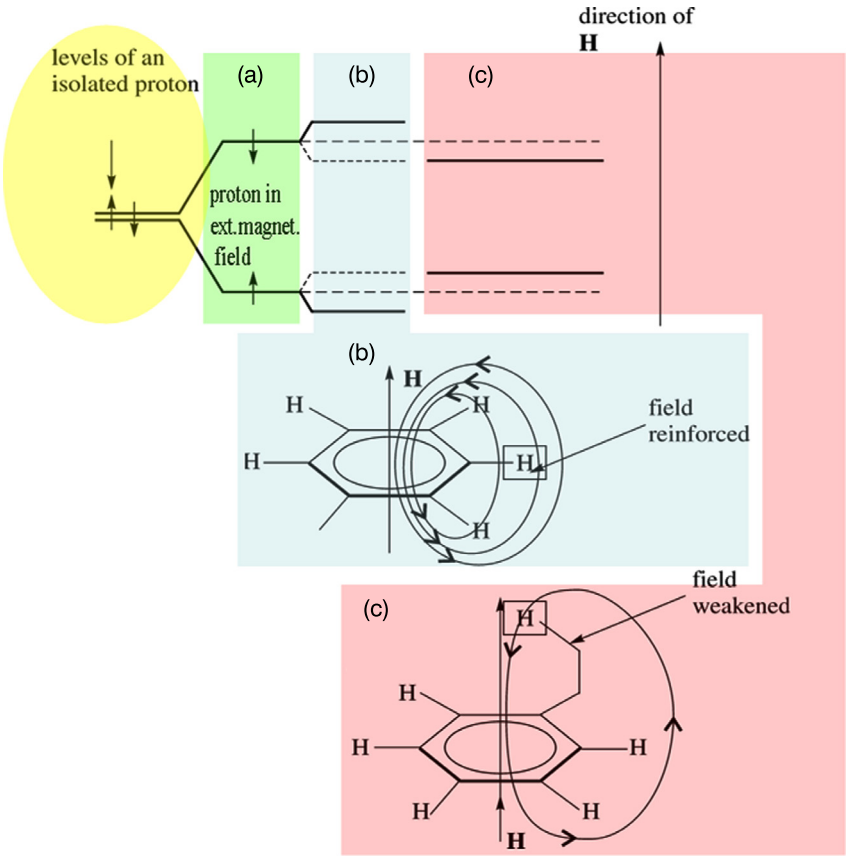


Fig. 4.12. Proton shielding by the electronic structure. (a) The energy levels of an isolated proton in a magnetic field. (b) The energy levels of the proton of the benzene ring (no nuclear spin interaction is assumed). The most mobile π electrons of benzene (which may be treated as a conducting circular wire) move around the benzene ring in response to the external magnetic field (perpendicular to the plane), thus producing an induced magnetic field. The latter one (when considered along the ring six-fold axis) opposes to the external magnetic field, but *at the position of the proton actually leads to an additional increase of the magnetic field felt by the proton*. This is why the figure shows the increase of the energy level difference due to the *electron shielding effect*. (c) The energy levels of another proton (located along the ring axis) in a similar molecule. This proton feels a local magnetic field that is decreased with respect to the external one (due to the induction effect).

applications, including UV-VIS, IR, Raman spectra, electronic structures, chemical reactions, and intermolecular interactions. This time, however, the situation is different: we are going to study very subtle interactions using the NMR technique which aims precisely at the energy levels that result from spin–spin and spin–magnetic field interactions. Even if these effects are very small, they can be observed. Therefore,

we have to consider more exact Hamiltonians.

- First, we have to introduce the interaction of our system with the electromagnetic field.
- Then we will consider the influence of the electronic structure on the magnetic field acting on the nuclei.
- Finally, the nuclear magnetic moment interaction ("coupling") will be considered.

4.8 Hamiltonian of the system in the electromagnetic field

The nonrelativistic Hamiltonian[61] $\hat{H}$ of the system of N particles (the j-th particle having mass m_j and charge q_j) moving in an external electromagnetic field with vector potential $\mathbf{A}$ and scalar potential ϕ may be written as[62]

$$\hat{H} = \sum_{j=1} [\frac{1}{2m_j} \left(\hat{\mathbf{p}}_j - \frac{q_j}{c} \mathbf{A}_j \right)^2 + q_j \phi_j] + \hat{V}, \tag{4.59}$$

where $\hat{V}$ stands for the "internal" potential coming from the mutual interactions of the particles, and $\mathbf{A}_j$ and ϕ_j denote the external vector[63] and scalar potentials $\mathbf{A}$ and ϕ calculated at the position of particle j, respectively.

4.8.1 Choice of the vector and scalar potentials

In Appendix V1-G on p. V1-673 it is shown that there is a certain arbitrariness in the choice of both potentials, which leaves the physics of the system unchanged. If for a homogeneous magnetic field $\mathbf{H}$ we choose the vector potential at the point indicated by $\mathbf{r} = (x, y, z)$ as

61 To describe the interactions of the spin magnetic moments, this Hamiltonian will soon be supplemented with the relativistic terms from the Breit Hamiltonian (p. V1-171).

62 To obtain this equation we may use Eq. (V1-3.33) as the starting point, which together with $E = mc^2$ gives with the accuracy of the first two terms the expression $E = m_0 c^2 + \frac{p^2}{2m_0}$. In the electromagnetic field, after introducing the vector and scalar potentials for a particle of charge q we have to replace E by $E - q\phi$, and $\mathbf{p}$ by $\left(\mathbf{p} - \frac{q}{c}\mathbf{A}\right)$. Then, after shifting the zero of the energy by $m_0 c^2$, the energy operator for a single particle reads as $\frac{1}{2m} \left(\hat{\mathbf{p}} - \frac{q}{c}\mathbf{A}\right)^2 + q\phi$, where $\mathbf{A}$ and ϕ are the values of the corresponding potentials at the position of the particle. For many particles we sum these contributions up and add the interparticle interaction potential (V). This is what we wanted to obtain (H. Hameka, "*Advanced Quantum Chemistry*," Addison-Wesley Publish. Co., Reading, Massachusetts (1965), p. 40).

63 Note that the presence of the magnetic field (and therefore of $\mathbf{A}$) makes it as if the charged particle moves faster on one side of the vector potential origin and slower on the opposite side.

(Eq. (V1-G.13)) $\mathbf{A}(\mathbf{r}) = \frac{1}{2}[\mathbf{H} \times \mathbf{r}]$, then, as shown in Appendix V1-G, we will satisfy the Maxwell equations, and in addition obtain the commonly used relation (Eq. (G.12)) $\mathrm{div}\mathbf{A} \equiv \nabla \mathbf{A} = 0$, known as the *Coulombic gauge*. In this way the origin of the coordinate system ($\mathbf{r} = \mathbf{0}$) was chosen as the origin of the vector potential (which need not be a rule).

Because $\boldsymbol{\mathcal{E}} = \mathbf{0}$ and $\mathbf{A}$ is time-independent, $\phi = const$ (p. V1-673), which of course means also that $\phi_j = const$; as an additive constant it may simply be eliminated from the Hamiltonian (4.59).

4.8.2 Refinement of the Hamiltonian

Let us assume the Born–Oppenheimer approximation (p. V1-320). Thus, the nuclei occupy some fixed positions in space, and in the electronic Hamiltonian (4.59) we have the electronic charges $q_j = -e$ and masses $m_j = m_0 = m$ (we skip the subscript 0 for the rest mass of the electron). Now, let us refine the Hamiltonian by adding the interaction of the particle magnetic moments (of the electrons and nuclei; the moments result from the orbital motion of the electrons as well as from the spin of each particle) with themselves and with the external magnetic field. Only these terms will count in the NMR experiment. We have, therefore, a "NMR refined Hamiltonian" of the system (the particular terms of the Hamiltonian correspond[64] to the relevant terms of the Breit Hamiltonian,[65] p. V1-171),

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_1 + \hat{\mathcal{H}}_2 + \hat{\mathcal{H}}_3 + \hat{\mathcal{H}}_4, \tag{4.60}$$

where (δ stands for the Dirac delta function, Appendix V1-E, and the spins have been replaced by the corresponding operators)

$$\hat{\mathcal{H}}_1 = \sum_{j=1}^{N} \frac{1}{2m}\left(\hat{\mathbf{p}}_j + \frac{e}{c}\mathbf{A}_j\right)^2 + \hat{V} + \hat{H}_{SH} + \hat{H}_{IH} + \hat{H}_{LS} + \hat{H}_{SS} + \hat{H}_{LL}, \tag{4.61}$$

$$\hat{\mathcal{H}}_2 = \gamma_{el} \sum_{j=1}^{N} \sum_{A} \gamma_A \left[\frac{\hat{\mathbf{s}}_j \cdot \hat{\mathbf{I}}_A}{r_{Aj}^3} - 3\frac{\left(\hat{\mathbf{s}}_j \cdot \mathbf{r}_{Aj}\right)\left(\hat{\mathbf{I}}_A \cdot \mathbf{r}_{Aj}\right)}{r_{Aj}^5} \right], \tag{4.62}$$

$$\hat{\mathcal{H}}_3 = -\gamma_{el}\frac{8\pi}{3} \sum_{j=1}^{N} \sum_{A} \gamma_A \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_j \cdot \hat{\mathbf{I}}_A, \tag{4.63}$$

64 All the terms used in the theory of magnetic susceptibilities and the Fermi contact term can be derived from classical electrodynamics.

65 Not to all of them. As we will see later, the NMR experimental spectra are described by using for each nucleus what is known as the shielding constant (related to the shielding of the nucleus by the electron cloud) and the internuclear coupling constants. The shielding and coupling constants enter in a specific way into the energy expression. Only those terms are included in the Hamiltonian that give nonzero contributions to these quantities.

$$\hat{\mathcal{H}}_4 = \sum_{A<B} \gamma_A \gamma_B \left[\frac{\hat{\mathbf{I}}_A \cdot \hat{\mathbf{I}}_B}{R_{AB}^3} - 3 \frac{\left(\hat{\mathbf{I}}_A \cdot \mathbf{R}_{AB}\right)\left(\hat{\mathbf{I}}_B \cdot \mathbf{R}_{AB}\right)}{R_{AB}^5} \right], \tag{4.64}$$

where in the global coordinate system the internuclear distance means the length of the vector $\mathbf{R}_{AB} = \mathbf{R}_B - \mathbf{R}_A$, while the electron–nucleus distance (of electron j with nucleus A) will be the length of $\mathbf{r}_{Aj} = \mathbf{r}_j - \mathbf{R}_A$. Then:

- In the term $\hat{\mathcal{H}}_1$, besides the kinetic energy operator in the external magnetic field (with vector potential $\mathbf{A}$, with the convention $\mathbf{A}_j \equiv \mathbf{A}\left(\mathbf{r}_j\right)$) given by $\sum_{j=1}^{N} \frac{1}{2m}\left(\hat{\mathbf{p}}_j + \frac{e}{c}\mathbf{A}_j\right)^2$, we have the Coulomb potential V of the interaction of all the charged particles.
  - Next, we have there the interaction of the spin magnetic moments of the electrons ($\hat{H}_{SH}$) and of the nuclei ($\hat{H}_{IH}$) with the field $\mathbf{H}$. These terms come from the first part of the term $\hat{H}_6$ of the Breit Hamiltonian, and represent the simple Zeeman terms:

Pieter Zeeman (1865–1943), Dutch physicist, professor at the University of Amsterdam. He became interested in the influence of a magnetic field on molecular spectra and discovered a field-induced splitting of the absorption lines in 1896. He shared the Nobel Prize with Hendrik Lorentz "*for their researches into the influence of magnetism upon radiation phenomena*" in 1902. The Zeeman splitting of star spectra allows us to determine the value of the magnetic field that was on the star at the moment the light was emitted!

  - $-\hat{\boldsymbol{\mu}} \cdot \mathbf{H}$, where $\hat{\boldsymbol{\mu}}$ is the magnetic moment operator of the corresponding particle. Why do we not have in $\hat{\mathcal{H}}_1$, together with $\hat{H}_{SH} + \hat{H}_{IH}$, the term $\hat{H}_{LH}$, i.e., the interaction of the electron orbital magnetic moment with the field? It would be so nice to have the full set of terms: the spin and the orbital magnetic moments interacting with the field. Everything is fine though; such a term is hidden in the mixed term resulting from $\frac{1}{2m}\left(\hat{\mathbf{p}}_j + \frac{e}{c}\mathbf{A}_j\right)^2$. Indeed, we get the corresponding Zeeman term from the transformation $\frac{e}{mc}\hat{\mathbf{p}}_j \cdot \mathbf{A}_j = \frac{e}{mc}\mathbf{A}_j \cdot \hat{\mathbf{p}}_j = \frac{e}{2mc}(\mathbf{H} \times \mathbf{r}_j) \cdot \hat{\mathbf{p}}_j = \frac{e}{2mc}\mathbf{H} \cdot (\mathbf{r}_j \times \hat{\mathbf{p}}_j) = \frac{e}{2mc}\mathbf{H} \cdot \hat{\mathbf{L}}_j = -\mathbf{H} \cdot (-\frac{e}{2mc}\hat{\mathbf{L}}_j) = -\mathbf{H} \cdot \mathbf{M}_{orb,el}(j)$, where $\mathbf{M}_{orb,el}(j)$ is, according to the definition of Eq. (4.54), the orbital magnetic moment of electron j.
  - The electronic spin–orbit terms ($\hat{H}_{LS}$), i.e., the corresponding magnetic dipole moment interactions, related to the term $\hat{H}_3$ in the Breit Hamiltonian.

  - The electronic spin–spin terms ($\hat{H}_{SS}$), i.e., the corresponding spin magnetic moment interactions, related to the term $\hat{H}_5$ in the Breit Hamiltonian.
  - The electronic orbit–orbit terms ($\hat{H}_{LL}$), i.e., the electronic orbital magnetic dipole interactions (corresponding to the term $\hat{H}_2$ in the Breit Hamiltonian).
- The (crucial for the NMR experiment) terms $\hat{\mathcal{H}}_2, \hat{\mathcal{H}}_3, \hat{\mathcal{H}}_4$ correspond to the magnetic "dipole–dipole" interaction involving *nuclear* spins (the term $\hat{H}_5$ of the Breit Hamiltonian): the classical electronic spin–nuclear spin interaction ($\hat{\mathcal{H}}_2$) plus the corresponding Fermi contact term[66] ($\hat{\mathcal{H}}_3$) and the classical interaction of the nuclear spin magnetic dipoles ($\hat{\mathcal{H}}_4$), this time without the contact term, because the nuclei are kept at long distances by the chemical bond framework.[67]

The magnetic dipole moment (of a nucleus or electron) "feels" the magnetic field acting on it through the vector potential $\mathbf{A}_j$ at the particle's position $\mathbf{r}_j$. This $\mathbf{A}_j$ is composed of the external field vector potential $\frac{1}{2}\left[\mathbf{H} \times \left(\mathbf{r}_j - \mathbf{R}\right)\right]$ (i.e., associated with the external magnetic field[68] $\mathbf{H}$), the individual vector potentials coming from the magnetic dipoles of the nuclei[69] $\sum_A \gamma_A \frac{\mathbf{I}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3}$ (and having their origins on the individual nuclei), and the vector potential $\mathbf{A}_{el}\left(\mathbf{r}_j\right)$ coming from the orbital and spin magnetic moments of all the electrons, so we have

$$\mathbf{A}_j \equiv \mathbf{A}\left(\mathbf{r}_j\right) = \frac{1}{2}\left[\mathbf{H} \times \mathbf{r}_{0j}\right] + \sum_A \gamma_A \frac{\mathbf{I}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3} + \mathbf{A}_{el}\left(\mathbf{r}_j\right), \tag{4.65}$$

where

$$\mathbf{r}_{0j} = \mathbf{r}_j - \mathbf{R}. \tag{4.66}$$

For closed shell systems (the majority of molecules) the vector potential $\mathbf{A}_{el}$ may be neglected, i.e., $\mathbf{A}_{el}\left(\mathbf{r}_j\right) \cong \mathbf{0}$, because the magnetic fields of the electrons are canceled out for a closed shell molecule (singlet state).

---

66 Let us take the example of the hydrogen atom in its ground state. Just note that the highest probability of finding the electron described by the orbital $1s$ is on the proton. The electron and the proton have spin magnetic moments that necessarily interact after they coincide. This effect is certainly something other than just the dipole–dipole interaction, which as usual describes the magnetic interaction for long distances. We have to have a correction for very short distances – this is the Fermi contact term.

67 And atomic electronic shell structure.

68 The vector $\mathbf{R}$ indicates the origin of the external magnetic field $\mathbf{H}$ vector potential from the global coordinate system (cf. Appendix V1-G and the commentary there related to the choice of origin).

69 Recalling the force lines of a magnet, we see that the magnetic field vector $\mathbf{H}$ produced by the nuclear magnetic moment $\gamma A \mathbf{I}_A$ should reside within the plane of $\mathbf{r}_{Aj}$ and $\gamma_A \mathbf{I}_A$. This means that $\mathbf{A}$ has to be orthogonal to the plane. This is ensured by $\mathbf{A}_j$ proportional to $\gamma_A \mathbf{I}_A \times \mathbf{r}_{Aj}$.

Rearranging terms

When such a vector potential $\mathbf{A}$ is inserted into $\hat{\mathcal{H}}_1$ (just patiently take the square of the content of the parentheses) we immediately get

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}^{(1)}, \tag{4.67}$$

where $\hat{\mathcal{H}}_0$ is the usual nonrelativistic Hamiltonian for the isolated system,

$$\hat{\mathcal{H}}_0 = -\sum_j \frac{\hbar^2}{2m}\Delta_j + \hat{V}, \tag{4.68}$$

$$\hat{\mathcal{H}}^{(1)} = \sum_k^{11} \hat{B}_k, \tag{4.69}$$

while *a few* minutes of careful calligraphy leads to the following result[70]:

$$\hat{B}_1 = \frac{e^2}{2mc^2}\sum_{A,B}\sum_j \gamma_A\gamma_B \frac{\hat{\mathbf{I}}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3}\frac{\hat{\mathbf{I}}_B \times \mathbf{r}_{Bj}}{r_{Bj}^3}, \tag{4.70}$$

$$\hat{B}_2 = \frac{e^2}{8mc^2}\sum_j \left(\mathbf{H}\times\mathbf{r}_{0j}\right)\cdot\left(\mathbf{H}\times\mathbf{r}_{0j}\right), \tag{4.71}$$

$$\hat{B}_3 = -\frac{i\hbar e}{mc}\sum_A\sum_j \gamma_A \nabla_j \cdot \frac{\hat{\mathbf{I}}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3}, \tag{4.72}$$

$$\hat{B}_4 = -\frac{i\hbar e}{2mc}\sum_j \nabla_j \cdot \left(\mathbf{H}\times\mathbf{r}_{0j}\right), \tag{4.73}$$

$$\hat{B}_5 = \frac{e^2}{2mc^2}\sum_A\sum_j \gamma_A \left(\mathbf{H}\times\mathbf{r}_{0j}\right)\cdot\frac{\hat{\mathbf{I}}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3}, \tag{4.74}$$

$$\hat{B}_6 = \hat{\mathcal{H}}_2 = \gamma_{el}\sum_{j=1}^{N}\sum_A \gamma_A\left[\frac{\hat{\mathbf{s}}_j\cdot\hat{\mathbf{I}}_A}{r_{Aj}^3} - 3\frac{\left(\hat{\mathbf{s}}_j\cdot\mathbf{r}_{Aj}\right)\left(\hat{\mathbf{I}}_A\cdot\mathbf{r}_{Aj}\right)}{r_{Aj}^5}\right], \tag{4.75}$$

$$\hat{B}_7 = \hat{\mathcal{H}}_3 = -\gamma_{el}\frac{8\pi}{3}\sum_{j=1}\sum_A \gamma_A\delta\left(\mathbf{r}_{Aj}\right)\hat{\mathbf{s}}_j\cdot\hat{\mathbf{I}}_A, \tag{4.76}$$

[70] The operators $\hat{B}_3$ and $\hat{B}_4$ contain the nabla (differentiation) operators. It is worth noting that this differentiation pertains to *everything on the right-hand side of the nabla, including any function on which* $\hat{B}_3$ *and* $\hat{B}_4$ *operators will act*.

$$\hat{B}_8 = \hat{H}_{SH} = -\gamma_{el} \sum_j \hat{\mathbf{s}}_j \cdot \mathbf{H}, \tag{4.77}$$

$$\hat{B}_9 = \hat{\mathcal{H}}_4 = \sum_{A<B} \gamma_A \gamma_B \left[\frac{\hat{\mathbf{I}}_A \cdot \hat{\mathbf{I}}_B}{R_{AB}^3} - 3 \frac{\left(\hat{\mathbf{I}}_A \cdot \mathbf{R}_{AB}\right)\left(\hat{\mathbf{I}}_B \cdot \mathbf{R}_{AB}\right)}{R_{AB}^5} \right], \tag{4.78}$$

$$\hat{B}_{10} = \hat{H}_{IH} = -\sum_A \gamma_A \hat{\mathbf{I}}_A \cdot \mathbf{H}, \tag{4.79}$$

$$\hat{B}_{11} = \hat{H}_{LS} + \hat{H}_{SS} + \hat{H}_{LL}. \tag{4.80}$$

We are just approaching the coupling of our theory with the NMR experiment. To this end, let us first define an empirical Hamiltonian, which serves in the NMR experiment to find what are known as the nuclear shielding constants and the spin–spin coupling constants. Then we will come back to the perturbation $\hat{\mathcal{H}}^{(1)}$.

4.9 Effective NMR Hamiltonian

NMR spectroscopy[71] means recording the electromagnetic wave absorption by a system of interacting nuclear magnetic dipole moments.[72] It is important to note that the energy differences detectable by contemporary NMR equipment are of the order of 10^{-13} a.u., while the breaking of a chemical bond corresponds to about 10^{-1} a.u. This is why

> all possible changes of the spin state of a system of nuclei do not change the chemical properties of the molecule. This is really what we could only dream of: we have something like observatory stations (the nuclear spins) that are able to detect tiny chemical bond details.

As will be seen in a moment, to reproduce NMR spectra we need an effective and rotation-averaged Hamiltonian that describes the interaction of the nuclear magnetic moments with the magnetic field and with themselves.

---

71 The first successful experiment of this kind was described by E.M. Purcell, H.C. Torrey, R.V. Pound, *Phys. Rev.*, 69(1946)37.

72 The wave lengths used in the NMR technique are of the order of meters (radio frequencies).

4.9.1 Signal averaging

NMR experiments usually pertain to long (many hours) recording of the radio wave radiation coming from a liquid specimen. Therefore, we obtain a static (time-averaged) record, which involves various kinds of averaging:

- over the rotations of any single molecule that contributes to the signal (we assume that each dipole keeps the same orientation in space when the molecule is rotating); these rotations can be free or restrained;
- over all the molecules present in the specimen;
- over the vibrations of the molecule (including internal rotations).

4.9.2 Empirical Hamiltonian

The empirical NMR Hamiltonian contains some parameters that take into account the electronic cloud structure in which the nuclei are immersed. *These NMR parameters will represent our target*.

Now, let us proceed in this direction.

To interpret the NMR data, it is sufficient to consider an *effective* Hamiltonian (containing explicitly only the nuclear magnetic moments; the electron coordinates are absent and the electronic structure enters only implicitly through some interaction parameters). In matrix notation we have

$$\hat{\mathcal{H}} = -\sum_A \gamma_A \mathbf{H}^T (\mathbf{1} - \boldsymbol{\sigma}_A) \mathbf{I}_A + \sum_{A<B} \gamma_A \gamma_B \{\mathbf{I}_A^T (\mathbf{D}_{AB} + \mathbf{K}_{AB}) \mathbf{I}_B\}, \tag{4.81}$$

where $\mathbf{I}_C \equiv \left(I_{C,x}, I_{C,y}, I_{C,z}\right)^T$ stands for the spin angular momentum of nucleus C, while $\boldsymbol{\sigma}_A$, $\mathbf{D}_{AB}, \mathbf{K}_{AB}$ denote the symmetric square matrices of dimension three (*tensors*):

- $\boldsymbol{\sigma}_A$ is a *shielding constant* tensor of nucleus A. Due to this shielding, nucleus A feels a *local field* $\mathbf{H}_{loc} = (\mathbf{1} - \boldsymbol{\sigma}_A)\mathbf{H} = \mathbf{H} - \boldsymbol{\sigma}_A \mathbf{H}$ instead of the external field $\mathbf{H}$ applied (due to the tensor character of $\boldsymbol{\sigma}_A$ the vectors $\mathbf{H}_{loc}$ and $\mathbf{H}$ may differ by their length and direction). The formula assumes that the shielding is proportional to the external magnetic field intensity that causes the shielding. Thus, the first term in the Hamiltonian $\hat{\mathcal{H}}$ may also be written as $-\sum_A \gamma_A \mathbf{H}_{loc}^T \mathbf{I}_A$.
- $\mathbf{D}_{AB}$ is the 3×3 matrix describing the (direct) *dipole–dipole interaction through space* defined above.
- $\mathbf{K}_{AB}$ is also a 3×3 matrix that takes into account that two magnetic dipoles interact additionally through the framework of the chemical bonds or hydrogen bonds that separate them. This is known as the *reduced spin–spin intermediate coupling tensor*.

Without electrons

Let us imagine, just for fun, removing all the electrons from the molecule (and keep them safely in a drawer), while the nuclei still reside in their fixed positions in space. The Hamiltonian would consist of two types of terms:

- *the Zeeman term*: interaction of the nuclear magnetic moments with the external electric field (the nuclear analog of the first term in $\hat{H}_6$ of the Breit Hamiltonian, p. V1-171), $-\sum_A \mathbf{H}\cdot\hat{\mathbf{M}}_A = -\sum_A \gamma_A \mathbf{H}\cdot\hat{\mathbf{I}}_A$;
- the "through space" dipole–dipole nuclear magnetic moment interaction (the nuclear analog of the $\hat{H}_5$ term in the Breit Hamiltonian), $\sum_{A<B}\gamma_A\gamma_B\{\hat{\mathbf{I}}_A\cdot(\mathbf{D}_{AB}\hat{\mathbf{I}}_B)$: $\mathbf{D}_{AB} = \frac{\mathbf{i}\cdot\mathbf{j}}{R_{AB}^3} - 3\frac{(\mathbf{i}\cdot\mathbf{R}_{AB})(\mathbf{j}\cdot\mathbf{R}_{AB})}{R_{AB}^5}$, where $\mathbf{i},\mathbf{j}$ denote the unit vectors along the x, y, z axes, e.g., $(\mathbf{D}_{AB})_{xx} = \frac{1}{R_{AB}^3} - 3\frac{(R_{AB,x})^2}{R_{AB}^5}$, $(\mathbf{D}_{AB})_{xy} = -3\frac{R_{AB,x}R_{AB,y}}{R_{AB}^5}$, etc., with $\mathbf{R}_{AB}$ denoting the vector pointing to nucleus B from nucleus A (of length R_{AB}).

Rotations average out the dipole–dipole interaction

What would happen if we rotated the molecule? In the theory of NMR, there are a lot of notions stemming from classical electrodynamics. In the isolated molecule the total angular momentum has to be conserved (this follows from the isotropic properties of space). The total angular momentum comes not only from the particles' orbital motion, but also from their spin contributions. The empirical (nonfundamental) conservation law pertains to the total spin angular momentum alone (cf. p. V1-86), as well as the individual spins separately. The spin magnetic moments are oriented in space, and this orientation results from the history of the molecule and may be different in each molecule of the substance. These spin states are nonstationary. The *stationary states correspond to some definite values of the square of the total spin of the nuclei and of the spin projection on a chosen axis*. According to quantum mechanics (Chapter V1-1), only these values are to be measured. For example, in the hydrogen molecule there are two stationary nuclear spin states: one with parallel spins (orthohydrogen) and the other with antiparallel spins (parahydrogen). Then we may assume that the hydrogen molecule has two "nuclear gyroscopes" that keep pointing them in the same direction in space when the molecule rotates (Fig. 4.13).

Let us see what will happen if we average the interaction of two magnetic dipole moments (the formula for the interaction of two dipoles will be derived in Chapter 5, p. 363). We have $E_{dip-dip} = \frac{\mathbf{M}_A\cdot\mathbf{M}_B}{R_{AB}^3} - 3\frac{(\mathbf{M}_A\cdot\mathbf{R}_{AB})(\mathbf{M}_B\cdot\mathbf{R}_{AB})}{R_{AB}^5}$. Assume (without losing the generality of the problem) that $\mathbf{M}_A$ resides at the origin of a polar coordinate system and has a constant direction along the z axis, while the dipole $\mathbf{M}_B$ just moves on the sphere of radius R_{AB} around $\mathbf{M}_A$ (all orientations are equally probable), the $\mathbf{M}_B$ vector preserving the same direction in space,

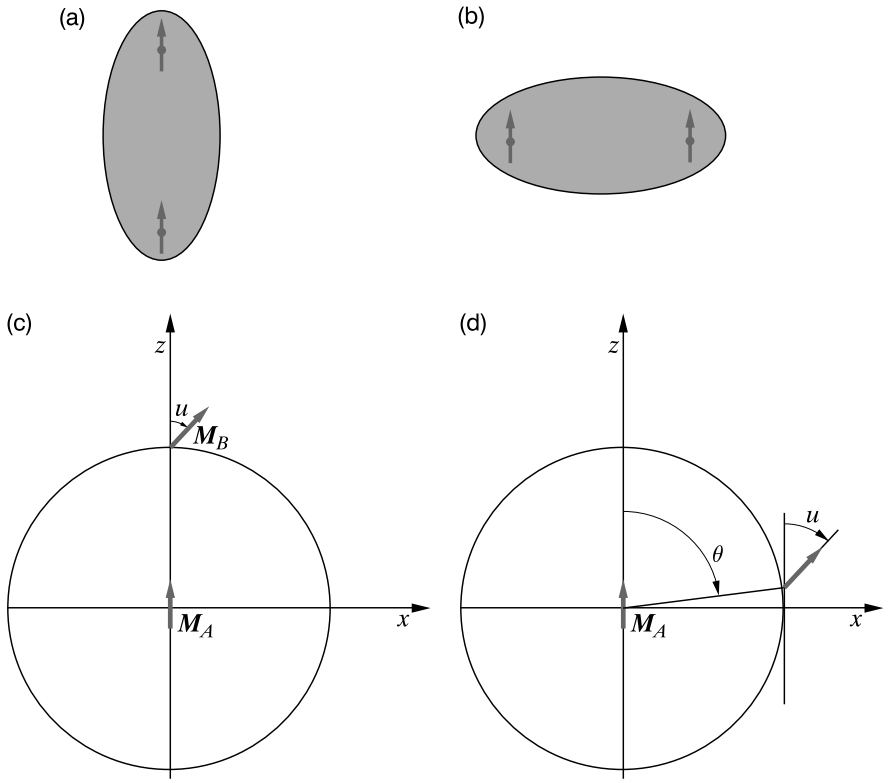


Fig. 4.13. Rotation of a molecule and the nuclear magnetic moments. (a) The orientation of the nuclear magnetic moments in the orthohydrogen at the perpendicular configuration of the nuclei. (b) The same, but the molecule is oriented horizontally. (c) In the theory of NMR, we assume (in a classical way) that the motion of the molecule does not influence the orientation of both nuclear magnetic moments averaging the dipole–dipole interaction over all possible orientations. Let us immobilize the magnetic moment $\mathbf{M}_A$ along the z axis. The magnetic moment $\mathbf{M}_B$ will move on the sphere of radius 1, both moments still keeping the same direction in space, $(\theta, \phi) = (u, 0)$. (d) One of such configurations. Averaging over all possible orientations gives zero (see the text).

$(\theta, \phi) = (u, 0)$, all the time. Now, let us calculate the average value of $E_{dip-dip}$ with respect to all possible positions of $\mathbf{M}_B$ on the sphere. We have

$$\bar{E}_{dip-dip} = \frac{1}{4\pi}\int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\phi E_{dip-dip}$$
$$= \frac{1}{4\pi}\int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\phi \left[\frac{1}{R_{AB}^3}\mathbf{M}_A \cdot \mathbf{M}_B - \frac{3}{R_{AB}^5}(\mathbf{M}_A \cdot \mathbf{R}_{AB})(\mathbf{M}_B \cdot \mathbf{R}_{AB})\right]$$

$$
\begin{aligned}
&= \frac{M_A M_B}{4\pi R_{AB}^3} \int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\phi \, [\cos u - 3\cos\theta \cos(\theta - u)] \\
&= \frac{M_A M_B}{2R_{AB}^3} \int_0^{\pi} d\theta \sin\theta \, [\cos u - 3\cos\theta \cos(\theta - u)] \\
&= \frac{M_A M_B}{R_{AB}^3} \{\cos u - \frac{3}{2} \int_0^{\pi} d\theta \sin\theta \cos\theta \, [\cos\theta \cos u + \sin\theta \sin u]\} \\
&= \frac{M_A M_B}{R_{AB}^3} \{\cos u - \frac{3}{2} \left[\cos u \cdot \frac{2}{3} + \sin u \cdot 0\right]\} = 0.
\end{aligned} \tag{4.82}
$$

Thus, the averaging gives 0 *irrespective of the radius* R_{AB} *and of the angle u between the two dipoles. This result was obtained when assuming the orientations of both dipoles do not change (the abovementioned "gyroscopes") and that all angles* θ *and* ϕ *are equally probable.*

Averaging over molecular rotations

An NMR experiment requires long recording times. This means that each molecule, when rotating freely (gas or liquid[73]) with respect to the NMR apparatus, acquires all possible orientations with equal probability. The equipment will detect an averaged signal. This is why the proposed effective Hamiltonian has to be averaged over the rotations. As we have shown, such an averaging causes the mean dipole–dipole interaction (containing $\mathbf{D}_{AB}$) to be equal to zero. If we assume that the external magnetic field is along the z axis, then the averaged Hamiltonian reads as (cf. Appendix H)

$$
\hat{\mathcal{H}}_{av} = -\sum_A \gamma_A (1 - \sigma_A) H_z \hat{I}_{A,z} + \sum_{A<B} \gamma_A \gamma_B K_{AB} (\hat{\mathbf{I}}_A \cdot \hat{\mathbf{I}}_B), \tag{4.83}
$$

where $\sigma_A = \frac{1}{3}\left(\sigma_{A,xx} + \sigma_{A,yy} + \sigma_{A,zz}\right) = \frac{1}{3} Tr\boldsymbol{\sigma}_A$, and $K_{AB} = \frac{1}{3} Tr\mathbf{K}_{AB}$.

This Hamiltonian is at the basis of NMR spectra interpretation. An experimentalist adjusts σ_A for all the magnetic nuclei and K_{AB} for all their interactions, in order to reproduce the observed spectrum. Any theory of NMR spectra should explain the values of these parameters.

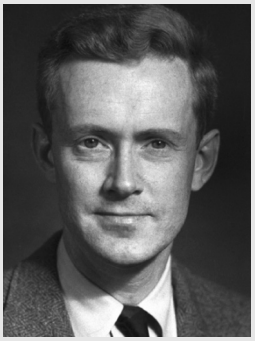
[73] This is not the case in the solid state.

Adding the electrons – why the nuclear spin interaction does not average out to zero

We know already why $\mathbf{D}_{AB}$ averages out to zero, but why this is not true for $\mathbf{K}_{AB}$?

Norman F. Ramsey (1915–2011), American physicist, professor at the University of Illinois and Columbia University, then from 1947 at Harvard University. He was first of all an outstanding experimentalist in the domain of NMR measurements in molecular jets, but his "hobby" was theoretical physics. Ramsey carried out the first accurate measurement of the neutron magnetic moment and gave a lower bound theoretical estimation to its dipole moment. In 1989 he received the Nobel Prize "*for the invention of the separated oscillatory fields method and its use in the hydrogen maser and other atomic clocks*."

Edwards Mills Purcell (1912–1997), American physicist, professor at the Massachusetts Institute of Technology and Harvard University. His main domains were relaxation phenomena and magnetic properties in low temperatures. He received in 1952 the Nobel Prize together with Felix Bloch "*for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith*."



Ramsey and Purcell[74] explained this by what is known as the spin induction mechanism, described in Fig. 4.14. Spin induction leads to the result that, in the averaging of $\mathbf{K}_{AB}$, the spin–spin configurations have different weights than in the averaging of $\mathbf{D}_{AB}$. This effect is due to the chemical bonds, because it makes a difference if the correlating electrons have their spins oriented parallel with or perpendicular to the bond line.

Where does such an effect appear in quantum chemistry? One of the main candidates may be the term $\hat{H}_3$ (the Fermi contact term in the Breit Hamiltonian, p. V1-171) which couples the

[74] N.F. Ramsey, E.M. Purcell, *Phys. Rev.*, 85(1952)143.

orbital motion of the electrons with their spin magnetic moments. This is a relativistic and hence very small effect, and therefore the rotational averaging leaves only a small value of K_{AB}.

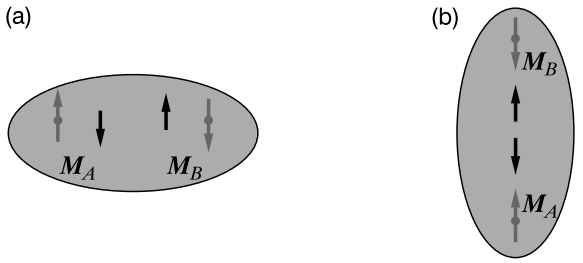


Fig. 4.14. The nuclear spin–spin coupling (Fermi contact) mechanism through chemical bond AB. The electrons repel each other and therefore correlate their motion (cf. p. 104). This is why, when one of them is close to nucleus A, the second prefers to run off to nucleus B. An electron close to A will exhibit a tendency (i.e., the corresponding energy will be lower than in the opposite case) to have a spin antiparallel to the spin of A. The second electron, close to B, must have opposite spin to its partner, and therefore will exhibit a tendency to have its spin *the same as that of nucleus* A. We may say that the second electron exposes the spin of nucleus A right at the position of nucleus B. Such a mechanism gives a much stronger magnetic dipole interaction than that through empty space. (a) A favorable configuration of nuclear and electron spins, all perpendicular to the bond. (b) The same situation after the molecule is rotated by 90°. The electronic correlation energy will obviously differ in these two orientations of the molecule, and this results in different averaging than in the case of the interaction through space.

4.9.3 Nuclear spin energy levels

Calculating the mean value of Hamiltonian (4.83), we obtain the energy of the nuclear spins in the magnetic field, i.e.,

$$E = -\sum_A (1-\sigma_A)\,\gamma_A H m_{I,A}\hbar + \sum_{A<B} \gamma_A \gamma_B K_{AB} \left\langle \hat{\mathbf{I}}_A \cdot \hat{\mathbf{I}}_B \right\rangle,$$

where $\left\langle \hat{\mathbf{I}}_A \cdot \hat{\mathbf{I}}_B \right\rangle$ is the mean value of the scalar product of the two spins calculated by using their spin functions. This expression can be simplified by the following transformation:

$$\begin{aligned} E &= -\sum_A (1-\sigma_A)\,\gamma_A H m_{I,A}\hbar + \sum_{A<B} \gamma_A \gamma_B K_{AB} \left\langle \hat{I}_{A,x}\hat{I}_{B,x} + \hat{I}_{A,y}\hat{I}_{B,y} + \hat{I}_{A,z}\hat{I}_{B,z} \right\rangle \\ &= -\sum_A (1-\sigma_A)\,\gamma_A H m_{I,A}\hbar + \sum_{A<B} \gamma_A \gamma_B K_{AB} \left(0\cdot 0 + 0\cdot 0 + \hbar^2 m_{I,A} m_{I,B} \right), \end{aligned}$$

because the mean values of $\hat{I}_{C,x}$ and $\hat{I}_{C,y}$ calculated for the spin functions of nucleus C both equal 0 (for the α or β functions describing a nucleus with $I_C = \frac{1}{2}$, see Chapter V1-1). Therefore,

the energy becomes a function of the magnetic spin quantum numbers $m_{I,C}$ for all the nuclei with a nonzero spin $\mathbf{I}_C$,

$$E\left(m_{I,A}, m_{I,B}, ...\right) = -\hbar H \sum_A \left(1 - \sigma_A\right) \gamma_A m_{I,A} + \sum_{A<B} h J_{AB} m_{I,A} m_{I,B}, \tag{4.84}$$

where the commonly used nuclear *spin–spin coupling constant* is defined as

$$J_{AB} \equiv \frac{\hbar}{2\pi} \gamma_A \gamma_B K_{AB}. \tag{4.85}$$

Note that since hJ_{AB} has the dimension of the energy, J_{AB} itself is a frequency and may be expressed in Hz.

Due to the presence of the rest of the molecule (electron shielding) the Larmor frequency $\nu_A = \frac{H\gamma_A}{2\pi}\left(1 - \sigma_A\right)$ is changed by $-\sigma_A \frac{H\gamma_A}{2\pi}$ with respect to the Larmor frequency $\frac{H\gamma_A}{2\pi}$ for the isolated proton. Such changes are usually expressed (as "ppm," i.e., "parts per million[75]") by the *chemical shift* δ_A,

$$\delta_A = \frac{\nu_A - \nu_{ref}}{\nu_{ref}} \cdot 10^6 = \frac{\sigma_{ref} - \sigma_A}{\sigma_{ref}} \cdot 10^6, \tag{4.86}$$

where ν_{ref} is the Larmor frequency for a reference nucleus (for protons this means by convention the proton Larmor frequency in tetramethylsilane, $Si(CH_3)_4$). The chemical shifts (unlike the Larmor frequencies) are independent of the magnetic field applied.

Example 3 (The carbon nucleus in an external magnetic field). A single carbon ^{13}C nucleus (spin quantum number $I_C = \frac{1}{2}$) in a molecule with nonmagnetic other nuclei.

As seen from Eq. (4.84) such a nucleus in magnetic field H has two energy levels (for $m_{I,C} = \pm\frac{1}{2}$, Fig. 4.15a),

$$E\left(m_{I,C}\right) = -\hbar H \left(1 - \sigma_C\right) \gamma_C m_{I,C},$$

where the shielding constant σ_C characterizes the vicinity of the nucleus. For the isolated nucleus $\sigma_C = 0$.

75 This means the chemical shift (unitless quantity) has to be multiplied by 10^{-6} to obtain $\frac{\nu_A - \nu_{ref}}{\nu_{ref}}$.

Example 4 (The methane molecule $^{13}CH_4$ in magnetic field H). This time there is an additional magnetic field coming from four equivalent protons, each having $I_H = \frac{1}{2}$. The energy levels of the carbon magnetic spin result from the magnetic field and from the $m_{I,H}$'s of the protons according to Eq. (4.84) (Fig. 4.15). The resonance of the ^{13}C nucleus means transition between energy levels that correspond to $m_{I,C} = \pm\frac{1}{2}$ and all the m_{I,H_i} being constant.[76] Thus, the lower level corresponds to

$$E_+\left(m_{I,H_1}, m_{I,H_2}, m_{I,H_3}, m_{I,H_4}\right) = -\frac{\hbar}{2}H\left(1-\sigma_C\right)\gamma_C + \frac{h}{2}^1 J_{CH}\left(m_{I,H_1} + m_{I,H_2} + m_{I,H_3} + m_{I,H_4}\right),$$

and at the higher level we have the energy

$$E_-\left(m_{I,H_1}, m_{I,H_2}, m_{I,H_3}, m_{I,H_4}\right) = \frac{\hbar}{2}H\left(1-\sigma_C\right)\gamma_C - \frac{h}{2}^1 J_{CH}\left(m_{I,H_1} + m_{I,H_2} + m_{I,H_3} + m_{I,H_4}\right).$$

Since $m_{I,H_i} = \pm\frac{1}{2}$, each of the levels $E_\pm$ will be split into five levels (Fig. 4.15):

- a nondegenerate level arising from all $m_{I,H_i} = \frac{1}{2}$,
- a quadruply degenerate level that comes from all $m_{I,H_i} = \frac{1}{2}$, except one equal to $-\frac{1}{2}$ (there are four positions of this one),
- a sextuple degenerate level that results from two $m_{I,H_i} = \frac{1}{2}$ and two $m_{I,H_i} = -\frac{1}{2}$ (six ways of achieving this),
- a quadruply degenerate level that comes from all $m_{I,H_i} = -\frac{1}{2}$, except one that equals $\frac{1}{2}$ (there are four positions of this one),
- a nondegenerate level arising from all $m_{I,H_i} = -\frac{1}{2}$.

Example 5 (Nuclear resonances in C_2H_5OH). Our result may be generalized for n equivalent protons, which interact with a nuclear spin in an external magnetic field. The proton magnetic moment may be aligned either along or opposite to the external magnetic field. The number of ways for k moments aligned along and $(n-k)$ moments aligned opposite is $\binom{n}{k}$. For the previous case (methane) with $n = 4$ and for $k = 0, 1, 2, 3, 4$ the numbers of equivalent positions are: 1, 4, 6, 4, 1, which leads to the degeneracy of the nuclear energy levels of the carbon nuclear momentum, as shown in Fig. 4.15.

What would happen if in a single molecule one had several groups of equivalent nuclei? By the way, what does "equivalent" mean? For example, how many equivalent protons do we have in

[76] The NMR selection rule for a given nucleus says that the single nucleus undergoes a flip.

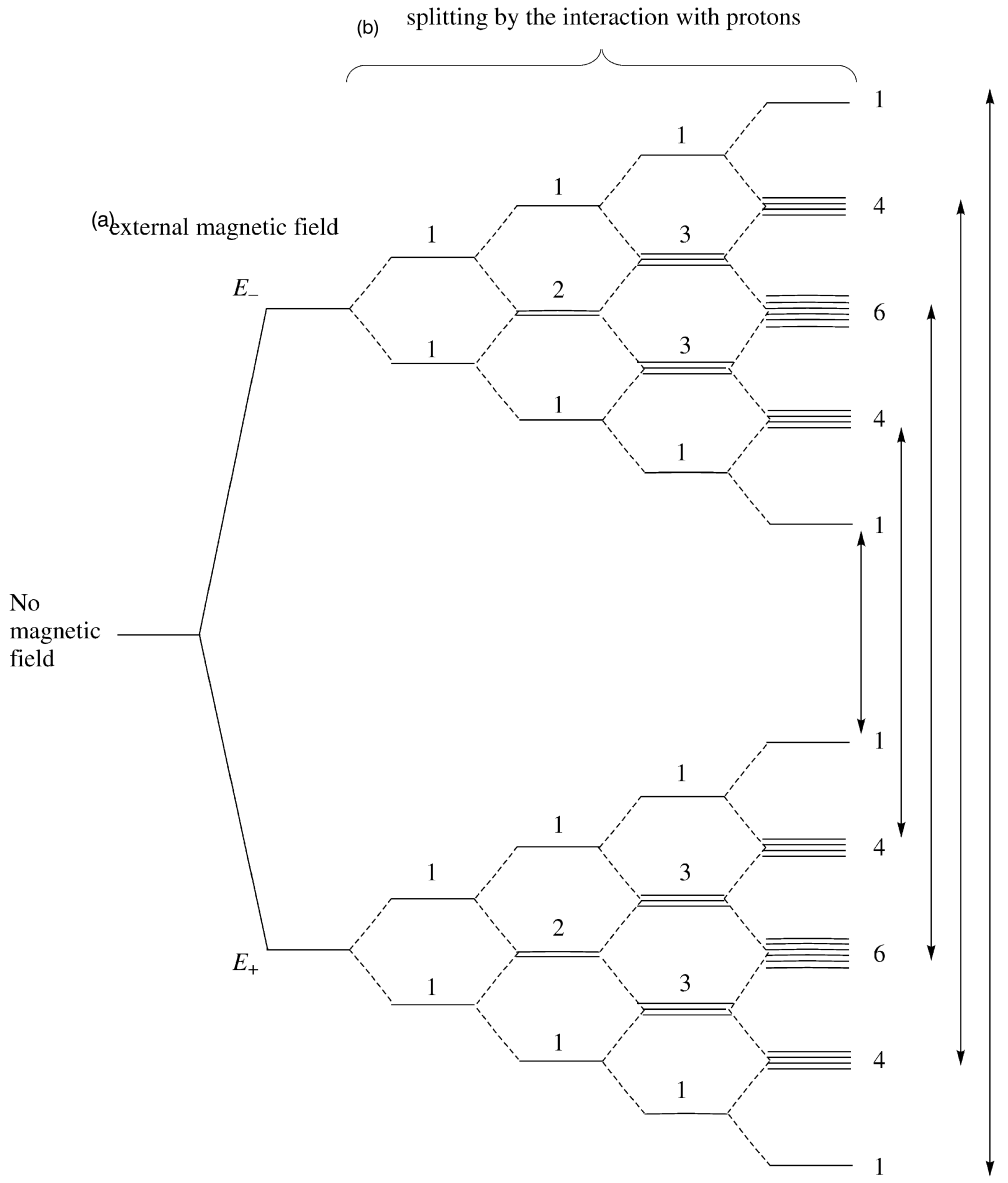


Fig. 4.15. The energy levels of the ^{13}C magnetic moment in an external magnetic field and in the methane molecule. (a) The spin energy levels of the ^{13}C atom in an external magnetic field. (b) Additional interaction of the ^{13}C spin with the four equivalent proton magnetic moments switched on. As we can see the energy levels in each branch follow the Pascal triangle rule. The splits within the branch come from the coupling of the nuclei and are field-independent. The E_+ and E_- energies are field-dependent; increasing the field means a tuning of the separation between the energy levels. The resonance takes place when the field-dependent energy difference matches the energy of the electromagnetic field quanta. The NMR selection rule means that only the indicated transitions take place. Since the energy split due to the coupling of the nuclei is very small, the levels E_+ are equally occupied and therefore the NMR intensities satisfy the ratio $1:4:6:4:1$.

a molecule such as C_2H_5OH? Already the chemists' way of writing this formula (a result of long debates in the distillery room) suggests that we know something special about one of these protons. It turned out that this peculiarity comes from binding to the oxygen atom, while other hydrogen atoms are bound to carbon atoms. These other protons form two groups, however, which is reflected in a more detailed formula: CH_3–CH_2–OH. We were discussing until now the nonequivalence, but what about equivalence? Are the three protons in the methyl group –CH_3 equivalent for a chemist? If we take into account the conformational states, we see several conformers possible, but in none of these conformations the three protons are equivalent, although *the roles played by the three protons in the whole set of the conformations are identical*! The situation gets better if one recalls that NMR experiments take a long time and pertain to many molecules in solution. Therefore, every molecule is able to visit all the conformations (for sufficiently high temperature), including those resulting from rotations of the –CH_3 group about the C–C bond. There is no good reason to think that the three protons of the methyl group are nonequivalent in the NMR experiment; the same pertains to the two protons of the –CH_2– group.[77]

The group of n_1 equivalent protons modifies the magnetic field felt by a nucleus (not belonging to the group), which will undergo a resonance transition. Due to the same coupling constants of this particular nucleus with all the nuclei of the group, a splitting of the NMR signal to $(n_1 + 1)$ signals occurs. For example, in the case of the ^{13}C resonance in methane we got $n_1 + 1 = 4 + 1 = 5$ signals.

Possibly, another group of n_2 equivalent protons interacts with the nucleus differently (another value of the coupling constant) and causes its own splitting of each of the previously described signals into $(n_2 + 1)$ signals.

There is one more problem. What will be if the nucleus, the resonance of which we are considering, is equivalent to some other nuclei? This may happen in two cases. In the first case the equivalent protons belong to the same molecule, in the second one they belong to different molecules. This second case happens always in macroscopic samples used in NMR, and the

[77] In the case of rigid molecules such an averaging does not take place. In such a case a subtle difference appears between the concepts of the chemical and magnetic equivalence of two nuclei, a and a'. One has chemical equivalence of a and a' if the two nuclei have the same neighborhood. If a and a' are chemically nonequivalent, they are also nonequivalent magnetically. However, *two chemically equivalent nuclei may turn out to be magnetically nonequivalent.* It will happen, when one finds at least one nucleus (excluding a and a'), that its nuclear coupling constant with a differs from that for a'. As an example may serve two protons in $H_2C{=}CF_2$. Indeed, let us take one of the fluorine nuclei. One of the protons corresponds to the *cis*, the other to the *trans* coupling with this particular nucleus.

result is that the NMR signal is stronger if the concentration is higher.[78] The NMR installation does not know anything about our concept of molecule. Therefore,

> in the case of equivalent nuclei from the same molecule one may treat all these nuclei as a "collective" single nucleus undergoing the resonance with the intensity multiplied by the number of the equivalent nuclei.

Such an effect looks natural if one recalls that the NMR electromagnetic waves correspond to radio frequencies. This means a long wave, which is unable to distinguish the equivalent nuclei distributed in a tiny section of space.

The NMR spectra may be quite complicated, especially when the chemical shifts turn out to be less important than the spin–spin coupling. In the example considered the opposite is true (such spectra are known as first-order NMR spectra), and the situation is easier. What therefore should one expect as the NMR spectrum of the CH_3–CH_2–OH molecule? Well, roughly speaking we expect three signals shifted with respect to what one gets for the commonly used tetramethylsilane reference (internal standard) signal: one for the protons of the –CH_3 group, one for those of the CH_2 group, and one for the proton of the OH group. We also expect that the shift of the methyl group signal should be the smallest one (because of the similarity to the reference), the shift for the methylene group should be larger (less similar to the reference), and the OH signal should differ very much from the reference (resulting in the largest shift). The experiment confirms this rough estimation (Fig. 4.16a). In addition, it is reasonable to expect the corresponding intensities to satisfy the proportion 3 : 2 : 1 (proportionality to the number of the protons). Again the experiment (Fig. 4.16a) seems to show something like that (when considering the area under the peaks), besides the fact that we see some quite complex splitting of each of the three signals.

We are not satisfied by the above rule of thumb and we want to discover why each signal has such a complex structure. We suspect that this has to do with the interaction with the neighbors of the resonating nucleus. Let us start from the collective nucleus H_3 from the methyl group. It interacts with a group of two equivalent protons (methylene group) separated from it by three bonds and with a single proton of the OH group separated by the four bonds. Therefore, first of all one may expect the resonance of the collective nucleus H_3 from the CH_3 group split into $n_{CH_2} + 1 = 3$ peaks, and their intensity ratio (caused by the degeneracy of the energy levels) 1 : 2 : 1. Additionally, each of the resulting signals should be split into two lines due to two possible spin orientations of the proton nucleus from the hydroxyl group. In the experiment we see the first splitting (Fig. 4.16b1), but not the second one. This is because of too weak

[78] It is also broader, because every molecule, even if chemically identical, has a slightly different geometry.

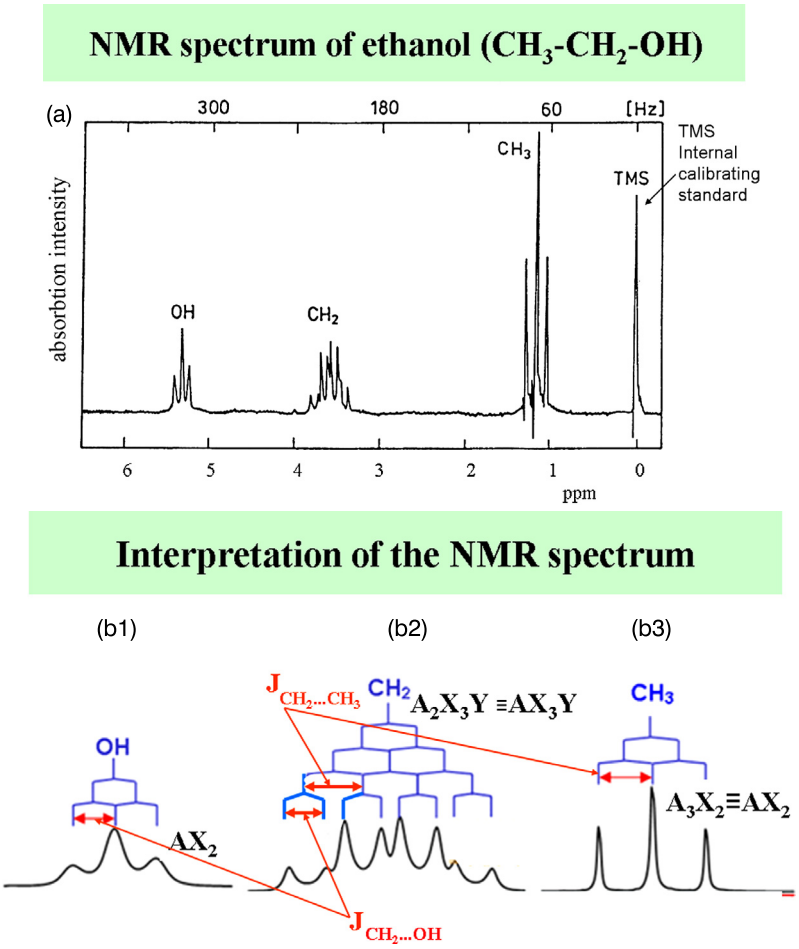


Fig. 4.16. Analysis of the proton magnetic resonance spectrum for liquid ethanol. (a) The NMR experimental spectrum. From the right-hand side: the tetramethylsilane (TMS) signal – the reference chemical shift taken as equal to $\delta = 0$, at $\delta \approx 1$ ppm the triplet signal from the protons of the CH_3 group, at about 3.5 ppm a multiplet from the CH_2 group, at $\delta \approx 5.3$ ppm there is a triplet coming from the resonance of the proton from the hydroxyl group. (b) Rationalization of the experimental spectrum. (b1) The hydroxyl group proton interacts magnetically with the protons of the CH_2 group, splitting the hydroxyl proton signal into three signals with the intensity ratio $1 : 2 : 1$ (there is no splitting visible from the CH_3 group, these protons are too far away). (b2) The explanation of the CH_2 multiplet signal goes in two steps. First, the splitting from the interaction with the CH_3 group is taken into account (the coupling constant $J_{CH_2-CH_3}$), resulting in a quartet with the intensity ratio $1 : 3 : 3 : 1$. Next, each of the resulting signals is split into two lines due to the interaction with the hydroxyl proton (with a smaller spin–spin coupling constant J_{CH_2-OH} than before, the intensity ratio is $1 : 1$). (b3) The signal from the collective nucleus of the CH_3 group is split by the interaction with the methylene group protons into three signals with the spin–spin coupling constant $J_{CH_2-CH_3}$ (the interaction with the hydroxyl group proton is too weak to be visible). Nomenclature: $A_nX_mY_p$ denotes the signal of the collective nucleus A (n equivalent nuclei) split by m equivalent protons X and p equivalent protons Y (nonequivalent to X).

sensitivity of the NMR equipment used; the hydroxyl proton is simply too far away and the splitting is small.

Now, what about –CH_2–? The signal from this "collective proton" is split by the methyl group protons into $n_{CH_3} + 1 = 4$ lines with the intensity ratio $1 : 3 : 3 : 1$ (the separation between them should be identical to that found for CH_3, because the two kinds of protons interact with the same coupling constant $J_1 = 7.2$ Hz). Each of these signals will be additionally split into two lines due to the interaction with the hydroxyl proton (the corresponding coupling constant is $J_2 = 5.1$ Hz). As a result one gets a quite complex multiplet of signals, its structure fully rationalized by the above coupling constants (see Fig. 4.16b2).

Finally, we will consider the proton from the hydroxyl group. Its signal will be split into $n_{CH_2} + 1 = 3$ lines with the intensity ratio $1 : 2 : 1$ (Fig. 4.16b3), because of two protons from the methylene group. The influence of the distant methyl group is not visible in the spectrum of this accuracy. The chemical shift dominates and is the largest for the protons of this molecule. It turns out that the chemical shift of this proton is sensitive to the details of the intermolecular interactions for this particular proton participates in the hydrogen bonds and other interactions, including possible chemical reactions.[79]

We therefore have important information: the NMR spectrum may serve for identification of chemical interactions within the molecules, as well as of intermolecular interactions.

4.10 The Ramsey theory of the NMR chemical shift

An external magnetic field $\mathbf{H}$ and/or the magnetic field produced by the nuclear magnetic dipole moments $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$... certainly represent an extremely weak perturbation to the molecule, and therefore the perturbational methods described in Chapter V1-5 seem to be a perfect choice.

We decide to apply the theory through the second order. Such an effect is composed of two parts:

[79] My NMR colleagues told me that the chemical shift of this proton may change very much, even changing the order of the multiplets observed. The chemical shift depends on such things as whether the ethanol is fresh made or not, how long it has been kept in the bottle, whether there are some traces of water in it or not, etc. This proton looks so crazy that one may think it is responsible for the widely known extraordinary properties of the ethanol molecule. Is the rest of the molecule innocent? Certainly not. We have protons in organic and inorganic acids and nothing comes out of it. Maybe we have to have a functional OH group? No, because water itself would act much stronger than the ethanol. Maybe the proton is innocent and this is the –C_2H_5 group that should be blamed? No, because H–C_2H_5 does not act like that. It seems one should have a hydroxyl group very close to a bulky hydrophobic group. Indeed, the –CH_3 group has an effect similar to that of –C_2H_5, except that in addition it hurts the eyes and even may kill the drinker.

- the first-order correction (*the diamagnetic contribution*),
- the second-order correction (*the paramagnetic contribution*).

The corresponding energy change due to the perturbation $\hat{\mathcal{H}}^{(1)}$ from Eq. (4.69) (the prime means that $k = 0$, i.e., the ground state is excluded from the summation)

$$\Delta E = E_0^{(1)} + E_0^{(2)} = \left\langle \psi_0^{(0)} | \hat{\mathcal{H}}^{(1)} \psi_0^{(0)} \right\rangle + \sum_k{}' \frac{\left\langle \psi_0^{(0)} | \hat{\mathcal{H}}^{(1)} \psi_k^{(0)} \right\rangle \left\langle \psi_k^{(0)} | \hat{\mathcal{H}}^{(1)} \psi_0^{(0)} \right\rangle}{E_0^{(0)} - E_k^{(0)}}. \tag{4.87}$$

4.10.1 Shielding constants

In the Hamiltonian (4.83) the shielding constants occur in the term $\mathbf{I}_A \cdot \mathbf{H}$. The perturbation operator $\hat{H}^{(1)}$ contains a lot of terms, but most of them, when inserted into the above formula, are unable to produce terms that behave like $\mathbf{I}_A \cdot \mathbf{H}$. Only some very particular terms could produce such a dot product dependence. A minute of reflection leads directly to $\hat{B}_3$, $\hat{B}_4$, $\hat{B}_5$, and $\hat{B}_{10}$ as the only terms of the Hamiltonian that have any chance of producing the dot product form.[80] Therefore, using the definition of the reduced resolvent $\hat{R}_0$ of Eq. (2.88) we have[81]

$$\Delta E = E_0^{(1)} + E_0^{(2)} = \left\langle \psi_0^{(0)} | (\hat{B}_{10} + \hat{B}_5) \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} | \left(\hat{B}_3 \hat{R}_0 \hat{B}_4 + \hat{B}_4 \hat{R}_0 \hat{B}_3 \right) \psi_0^{(0)} \right\rangle. \tag{4.88}$$

After averaging the formula over rotations and extracting the proper term proportional to $\mathbf{I}_A$ and $\mathbf{H}$ (details given in Appendix H, p. 627) we obtain as the shielding constant of nucleus A

---

80 There is an elegant way to single out the only necessary B_i's that give a contribution to the energy proportional to the product $x_i x_j$ (no higher terms included), where x_i and x_j stand for some components of the magnetic field intensity $\mathbf{H}$ and/or of the nuclear spin $\mathbf{I}_A$'s (that cause perturbation of the molecule). As to the first-order correction ("diamagnetic") we calculate the second derivative $\left(\frac{\partial^2 \hat{\mathcal{H}}^{(1)}}{\partial x_i \partial x_i} \right)_{\mathbf{H}=\mathbf{0}, \mathbf{I}_i = \mathbf{0}}$ of the Hamiltonian $\hat{\mathcal{H}}^{(1)}$ with respect to the components of $\mathbf{H}$ and/or $\mathbf{I}_A$, afterwards inserting $\mathbf{H} = \mathbf{0}$ and $\mathbf{I}_A = \mathbf{0}$ (i.e., calculating the derivative at zero perturbation). Then the diamagnetic correction to the energy is $\left\langle \psi_0^{(0)} | \left(\frac{\partial^2 \hat{\mathcal{H}}^{(1)}}{\partial x_i \partial x_i} \right)_{\mathbf{H}=\mathbf{0}, \mathbf{I}_i = \mathbf{0}} \psi_0^{(0)} \right\rangle$. As to the second-order correction ("paramagnetic"), we calculate the first derivatives, $\left(\frac{\partial \hat{\mathcal{H}}^{(1)}}{\partial x_i} \right)_{\mathbf{H}=\mathbf{0}, \mathbf{I}_i = \mathbf{0}}$ and $\left(\frac{\partial \hat{\mathcal{H}}^{(1)}}{\partial x_j} \right)_{\mathbf{H}=\mathbf{0}, \mathbf{I}_i = \mathbf{0}}$, and therefore the contribution to the energy is $\sum_k{}' \frac{\left\langle \psi_0^{(0)} | \left(\frac{\partial \hat{\mathcal{H}}^{(1)}}{\partial x_i} \right)_{\mathbf{H}=\mathbf{0}, \mathbf{I}_i = \mathbf{0}} \psi_k^{(0)} \right\rangle \left\langle \psi_k^{(0)} | \left(\frac{\partial \hat{\mathcal{H}}^{(1)}}{\partial x_j} \right)_{\mathbf{H}=\mathbf{0}, \mathbf{I}_i = \mathbf{0}} \psi_0^{(0)} \right\rangle}{E_0^{(0)} - E_k^{(0)}}$.

81 Note that whenever the reduced resolvent appears in a formula infinite summation over unperturbed states is involved.

$$\sigma_A = \frac{e^2}{3mc^2}\left\langle \psi_0^{(0)} | \sum_j (\mathbf{r}_{0j}\cdot\mathbf{r}_{Aj})\frac{1}{r_{Aj}^3}\psi_0^{(0)}\right\rangle + \tag{4.89}$$

$$-\frac{e^2}{6m^2c^2}\left\langle \psi_0^{(0)} | \left[\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\hat{R}_0\left(\sum_j \hat{\mathbf{L}}_{0j}\right) + \left(\sum_j \hat{\mathbf{L}}_{0j}\right)\hat{R}_0\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\right]\psi_0^{(0)}\right\rangle,$$

where

$$\hat{\mathbf{L}}_{Aj} = -i\hbar\left(\mathbf{r}_{Aj}\times\nabla_j\right) \tag{4.90}$$

and

$$\hat{\mathbf{L}}_{0j} = -i\hbar\left(\mathbf{r}_{0j}\times\nabla_j\right) \tag{4.91}$$

stand for the angular momenta operators for electron j calculated with respect to the position of nucleus A and with respect to the origin of vector potential $\mathbf{A}$, respectively.

4.10.2 Diamagnetic and paramagnetic contributions

The result (4.89) has been obtained in two parts,

$$\sigma_A = \sigma_A^{\text{dia}} + \sigma_A^{\text{para}}, \tag{4.92}$$

called the *diamagnetic contribution*,

$$\sigma_A^{\text{dia}} = \frac{e^2}{3mc^2}\left\langle \psi_0^{(0)} | \sum_j (\mathbf{r}_{0j}\cdot\mathbf{r}_{Aj})\frac{1}{r_{Aj}^3}\psi_0^{(0)}\right\rangle,$$

and the *paramagnetic contribution,*

$$\sigma_A^{\text{para}} = -\frac{e^2}{6m^2c^2}\left\langle \psi_0^{(0)} | \left[\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\hat{R}_0\left(\sum_j \hat{\mathbf{L}}_{0j}\right) + \left(\sum_j \hat{\mathbf{L}}_{0j}\right)\hat{R}_0\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\right]\psi_0^{(0)}\right\rangle.$$

Each of these contributions looks suspicious. Indeed, the diamagnetic contribution explicitly depends on the choice of origin $\mathbf{R}$ of vector potential $\mathbf{A}$ through $\mathbf{r}_{0j} = \mathbf{r}_j - \mathrm{R}$ (see Eq. (4.66)). Similarly, the paramagnetic contribution also depends on choice through $\hat{\mathbf{L}}_{0j}$ and Eq. (4.66). The *practical* importance of the choice of $\mathbf{R}$ is stressed in Appendix V1-G. Since both contributions depend on the choice, they cannot have any physical significance separately.

Is it possible that the *sum* of the two contributions is invariant with respect to choice of $\mathbf{R}$? Yes, it is! The invariance has fortunately been proved.[82] This is good, because any measurable quantity cannot depend on arbitrary choice of the origin of the coordinate system.

[82] A. Abragam, "*The Principles of Nuclear Magnetism*", Clarendon Press, Oxford (1961).

4.11 The Ramsey theory of NMR spin–spin coupling constants

We will apply the same philosophy to calculate the nuclear coupling constant. Taking into account the Hamiltonian $\hat{\mathcal{H}}^{(1)}$ from Eq. (4.69) (p. 305), we note that the only terms in $\hat{\mathcal{H}}^{(1)}$ that have the chance to contribute to the NMR coupling constants (see Eq. (4.84)) are

$$\begin{aligned}\Delta E = E_0^{(1)} + E_0^{(2)} &= \left\langle \psi_0^{(0)} | (\hat{B}_1 + \hat{B}_9)\psi_0^{(0)} \right\rangle + \\ &+ \left\langle \psi_0^{(0)} | \left(\hat{B}_3 + \hat{B}_6 + \hat{B}_7\right) \hat{R}_0 | \left(\hat{B}_3 + \hat{B}_6 + \hat{B}_7\right) \psi_0^{(0)} \right\rangle \qquad (4.93) \\ &= E_{\text{dia}} + E_{\text{para}}, \qquad (4.94)\end{aligned}$$

because we are looking for terms that could result in the scalar product of the nuclear magnetic moments. The first term is the diamagnetic contribution (E_{dia}), the second one is the paramagnetic contribution (E_{para}).

4.11.1 Diamagnetic contribution

There are two diamagnetic contributions in the total diamagnetic effect $\left\langle \psi_0^{(0)} | (\hat{B}_1 + \hat{B}_9)\psi_0^{(0)} \right\rangle$:

- The $\left\langle \psi_0^{(0)} | \hat{B}_9 \psi_0^{(0)} \right\rangle$ term simply represents the $\sum_{A<B} \gamma_A \gamma_B \mathbf{I}_A^T \mathbf{D}_{AB} \mathbf{I}_B$ contribution of Eq. (4.81), i.e., the *direct ("through space") nuclear spin–spin interaction.*This calculation does not require anything except summation over spin–spin terms. However, *as has been shown, averaging over free rotations of the molecule in the specimen renders this term equal to zero*.
- The $\left\langle \psi_0^{(0)} | \hat{B}_1 \psi_0^{(0)} \right\rangle$ term can be transformed in the following way:

$$\left\langle \psi_0^{(0)} | \hat{B}_1 \psi_0^{(0)} \right\rangle = \left\langle \psi_0^{(0)} | \left(\frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \frac{\mathbf{I}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3} \frac{\mathbf{I}_B \times \mathbf{r}_{Bj}}{r_{Bj}^3} \right) \psi_0^{(0)} \right\rangle =$$
$$\frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \left\langle \psi_0^{(0)} | \frac{\left(\mathbf{I}_A \times \mathbf{r}_{Aj}\right) \cdot \left(\mathbf{I}_B \times \mathbf{r}_{Bj}\right)}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle.$$

Now, note that $(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$. Taking $\mathbf{A} = \mathbf{I}_A, \mathbf{B} = \mathbf{r}_{Aj}, \mathbf{C} = \mathbf{I}_B \times \mathbf{r}_{Bj}$ we first have the following:

$$\left\langle \psi_0^{(0)} | \hat{B}_1 \psi_0^{(0)} \right\rangle = \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \mathbf{I}_A \cdot \left\langle \psi_0^{(0)} | \frac{\mathbf{r}_{Aj} \times \left(\mathbf{I}_B \times \mathbf{r}_{Bj}\right)}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle.$$

Recalling that $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}\,(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}\,(\mathbf{A}\cdot\mathbf{B})$ this term (called the *diamagnetic spin–orbit contribution [DSO]*[83]) reads as

$$E_{\mathrm{DSO}}=\frac{e^2}{2mc^2}\sum_{A,B}\sum_{j}\gamma_A\gamma_B\mathbf{I}_A\cdot\left[\mathbf{I}_B\left\langle\psi_0^{(0)}|\frac{\mathbf{r}_{Aj}\cdot\mathbf{r}_{Bj}}{r_{Aj}^3r_{Bj}^3}\psi_0^{(0)}\right\rangle-\left\langle\psi_0^{(0)}|\mathbf{r}_{Bj}\frac{\mathbf{r}_{Aj}\cdot\mathbf{I}_B}{r_{Aj}^3r_{Bj}^3}\psi_0^{(0)}\right\rangle\right].$$

We see that we need to calculate some integrals with monoelectronic operators, which is an easy task.

4.11.2 Paramagnetic contribution

The paramagnetic contribution E_{para} to the energy can be split into several terms, i.e.,

$$\begin{aligned}E_{\mathrm{para}}&=\left\langle\psi_0^{(0)}|\left(\hat{B}_3+\hat{B}_6+\hat{B}_7\right)\hat{R}_0\left(\hat{B}_3+\hat{B}_6+\hat{B}_7\right)\psi_0^{(0)}\right\rangle\\&=E_{\mathrm{PSO}}+E_{\mathrm{SD}}+E_{\mathrm{FC}}+\text{mixed terms},\end{aligned}$$

where we have:

- the *paramagnetic spin–orbit contribution*,

$$E_{\mathrm{PSO}}=\left\langle\psi_0^{(0)}|\hat{B}_3\hat{R}_0\hat{B}_3\psi_0^{(0)}\right\rangle,$$

with $\hat{B}_3$ meaning the interaction between the *nuclear spin* magnetic moment and the magnetic moment resulting from the *electronic angular momenta* of the individual electrons in an atom,

- the *spin–dipole contribution*,

$$E_{\mathrm{SD}}=\left\langle\psi_0^{(0)}|\hat{B}_6\hat{R}_0\hat{B}_6\psi_0^{(0)}\right\rangle,$$

which describes the interaction energy of the magnetic spin dipoles, the nuclear with the electronic dipole,

- the *Fermi contact interaction*,

$$E_{\mathrm{FC}}=\left\langle\psi_0^{(0)}|\hat{B}_7\hat{R}_0\hat{B}_7\psi_0^{(0)}\right\rangle,$$

which is related to the electronic spin–nuclear spin interaction with zero distance between them,

[83] The name comes, of course, from the nuclear spin–electronic orbit interaction.

- the mixed terms, which contain $\left\langle \psi_0^{(0)} | \hat{B}_i \hat{R}_0 \hat{B}_j \psi_0^{(0)} \right\rangle$ for $i, j = 3, 6, 7$ and $i \neq j$; these terms are either exactly zero or (in most cases, not always) small.[84]

4.11.3 Coupling constants

The energy contributions have to be averaged over rotations of the molecule and the coupling constants are to be extracted from the resulting formulae. How this is performed is shown in Appendix H on p. 627.

Using this result, the nuclear spin–spin coupling constant is calculated as the sum of the diamagnetic and paramagnetic contributions,

$$J_{AB} = J_{AB}^{\text{dia}} + J_{AB}^{\text{para}}, \tag{4.95}$$

$$J_{AB}^{\text{dia}} \equiv J_{AB}^{\text{DSO}}, \tag{4.96}$$

$$J_{AB}^{\text{para}} = J_{AB}^{\text{PSO}} + J_{AB}^{\text{SD}} + J_{AB}^{\text{FC}} + J_{AB}^{\text{mixed}}, \tag{4.97}$$

where the particular contributions to the coupling constant are[85]

$$J_{AB}^{\text{DSO}} = \frac{e^2\hbar}{3\pi mc^2} \gamma_A \gamma_B \sum_j \left\langle \psi_0^{(0)} | \frac{\mathbf{r}_{Aj} \cdot \mathbf{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle,$$

$$J_{AB}^{\text{PSO}} = \frac{1}{3\pi} \hbar \left(\frac{e}{mc}\right)^2 \gamma_A \gamma_B \sum_{j,lAj} \left\langle \psi_0^{(0)} | \hat{\mathbf{L}}_{Aj} \hat{R}_0 \hat{\mathbf{L}}_{Bl} \psi_0^{(0)} \right\rangle,$$

[84] Let us consider all cross terms. First, let us check that $\left\langle \psi_0^{(0)} | \hat{B}_3 \hat{R}_0 \hat{B}_6 \psi_0^{(0)} \right\rangle = \left\langle \psi_0^{(0)} | \hat{B}_6 \hat{R}_0 \hat{B}_3 \psi_0^{(0)} \right\rangle = \left\langle \psi_0^{(0)} | \hat{B}_3 \hat{R}_0 \hat{B}_7 \psi_0^{(0)} \right\rangle = \left\langle \psi_0^{(0)} | \hat{B}_7 \hat{R}_0 \hat{B}_3 \psi_0^{(0)} \right\rangle = 0$. Note that $\hat{B}_6$ and $\hat{B}_7$ both contain electron spin operators, while $\hat{B}_3$ does not. Let us assume that, as is usually the case, $\psi_0^{(0)}$ is a singlet function. Recalling Eq. (2.88) this implies that, for $\left\langle \psi_0^{(0)} | \hat{B}_3 \psi_k^{(0)} \right\rangle$ to survive, the function $\psi_k^{(0)}$ has to have the same multiplicity as $\psi_0^{(0)}$. This however kills the other factors, $\left\langle \psi_0^{(0)} | \hat{B}_6 \psi_k^{(0)} \right\rangle$ and $\left\langle \psi_0^{(0)} | \hat{B}_7 \psi_k^{(0)} \right\rangle$ terms describing the magnetic interaction of nuclei with exactly the same role played by electrons with α and β spins. Thus, the products $\left\langle \psi_0^{(0)} | \hat{B}_3 \psi_k^{(0)} \right\rangle \left\langle \psi_k^{(0)} | \hat{B}_6 \psi_0^{(0)} \right\rangle$ and $\left\langle \psi_0^{(0)} | \hat{B}_3 \psi_k^{(0)} \right\rangle \left\langle \psi_k^{(0)} | \hat{B}_7 \psi_0^{(0)} \right\rangle$ are zero.

The mixed term $\left\langle \psi_0^{(0)} | \hat{B}_6 \hat{R}_0 \hat{B}_7 \psi_0^{(0)} \right\rangle$ vanishes for the isotropic electron cloud around the nucleus, because in the product $\left\langle \psi_0^{(0)} | \hat{B}_6 \psi_k^{(0)} \right\rangle \left\langle \psi_k^{(0)} | \hat{B}_7 \psi_0^{(0)} \right\rangle$ the Fermi term $\left\langle \psi_k^{(0)} | \hat{B}_7 \psi_0^{(0)} \right\rangle$ survives if $\psi_0^{(0)} \psi_k^{(0)}$ calculated at the nucleus is nonzero. This kills, however, $\left\langle \psi_0^{(0)} | \hat{B}_6 \psi_k^{(0)} \right\rangle$, because for $\psi_0^{(0)} \psi_k^{(0)} \neq 0$ (which as a rule comes from a $1s$ orbital, isotropic situation) the electron–nucleus dipole–dipole magnetic interaction averages to zero when different positions of the electron are considered. For nonisotropic cases this mixed contribution can be of importance.

[85] The empirical Hamiltonian (4.84) contains only the $A > B$ contributions; therefore, the factor 2 appears in J.

$$J_{AB}^{\mathrm{SD}} = \frac{1}{3\pi}\hbar\gamma_{el}^2\gamma_A\gamma_B \sum_{j,l=1}^{N} \left\langle \psi_0^{(0)} \middle| \left[\frac{\hat{\mathbf{s}}_j}{r_{Aj}^3} - 3\frac{\left(\hat{\mathbf{s}}_j \cdot \mathbf{r}_{Aj}\right)\mathbf{r}_{Aj}}{r_{Aj}^5} \right] \hat{R}_0 \left[\frac{\hat{\mathbf{s}}_l}{r_{Bl}^3} - 3\frac{\left(\hat{\mathbf{s}}_l \cdot \mathbf{r}_{Bl}\right)(\mathbf{r}_{Bl})}{r_{Bl}^5} \right] \psi_0^{(0)} \right\rangle,$$

$$J_{AB}^{\mathrm{FC}} = \frac{1}{3\pi}\hbar\left(\frac{8\pi}{3}\right)^2 \gamma_{el}^2\gamma_A\gamma_B \sum_{j,l=1} \left\langle \psi_0^{(0)} | \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_j \hat{R}_0 \delta\left(\mathbf{r}_{Bl}\right) \hat{\mathbf{s}}_l \psi_0^{(0)} \right\rangle.$$

Thus,

the nuclear spin magnetic moments are coupled via their magnetic interaction with the *electronic* magnetic moments:

- $J_{AB}^{\mathrm{DSO}} + J_{AB}^{\mathrm{PSO}}$ results from the electronic *orbital* magnetic dipole moments,
- $J_{AB}^{\mathrm{SD}} + J_{AB}^{\mathrm{FC}}$ corresponds to such interactions with the electronic *spin* magnetic dipole moments.

As to the integrals involved, the Fermi contact contribution J_{AB}^{FC} (just the value of the wave function at the nucleus position) is the easiest to compute. Assuming that $\psi_k^{(0)}$ states are Slater determinants, the diamagnetic spin–orbit contribution J_{AB}^{DSO} requires some (easy) one-electron integrals of the type $\left\langle \psi_1 \middle| \frac{x_{Aj}}{r_{Aj}^3} \psi_2 \right\rangle$, the paramagnetic spin–orbit contribution J_{AB}^{PSO} needs some one-electron integrals involving $\hat{\mathbf{L}}_{Aj}$ operators, which require differentiation of the orbitals, the spin–dipole contribution J_{AB}^{SD} leads also to some simple one-electron integrals, but handling the spin operators is needed (see p. V1-33), as for J_{AB}^{FC}. All the formulae require an infinite summation over states (due to the presence of $\hat{R}_0$), which is very tedious. This is why, in contemporary computational techniques, some other approaches, mainly what is called propagator theory, are used.[86]

4.11.4 The Fermi contact coupling mechanism

There are no simple rules, but usually the most important contribution to J_{AB} comes from the Fermi contact term (J_{AB}^{FC}); the next most important is the paramagnetic spin–orbit term J_{AB}^{PSO}, and other terms, including the mixed contributions J_{AB}^{mixed}, are of small importance. Let us consider the Fermi contact coupling mechanism between two protons through a single bond (the coupling constant J_{AB} denoted as ${}^1J_{HH}$). The proton and the electron close to it prefer to have opposite spins (Fig. 4.14). Then the other electron of the bond (being closer to the other nucleus) shows the other nucleus the spin of first nucleus, and therefore the second nucleus

[86] J. Linderberg, Y. Öhrn, "*Propagators in Quantum Chemistry*," Second edition, John Wiley and Sons, Ltd., 2004.

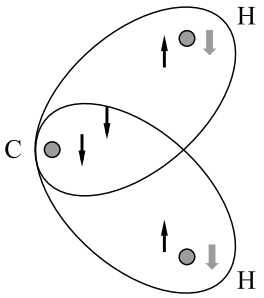


Fig. 4.17. Is the proton–proton coupling constant through two bonds (H–C–H), i.e., $^2J_{HH}$, positive or negative? Please recall that $^1J_{HH} > 0$ (see Fig. 4.14a), where the induction mechanism is described. The interaction through two bonds depends on what happens at the central carbon atom: are the spins of the two electrons there (one from each bond C–H) parallel or antiparallel? Hund's rule suggests they prefer to be parallel. This means that the situation with the two proton spins parallel is more energetically favorable, and this means $^2J_{HH} < 0$. This rule of thumb may fail when the carbon atom participates in multiple bonds, as in ethylene (see Section "From the research front").

prefers to have the opposite spin with respect to the first nucleus. According to Eq. (4.84), since $m_{I,A}m_{I,B} < 0$, we have $J_{AB} \equiv^1 J_{HH} > 0$. What about $^2J_{HH}$? This time, to have a through bond interaction we have to have a central atom, like carbon ^{12}C (i.e., with zero magnetic moment) (Fig. 4.17). The key point now is what happens at the central atom; is it preferable to have on it two parallel or two antiparallel electron spins? We do not know, but we may have a suggestion. Hund's rule says that in the case of orbital degeneracy (in our case: this corresponds to two *equivalent* C–H bonds), the electrons prefer to have parallel spins. This suggests that the two distant proton spins have a negative coupling constant, i.e., $^2J_{HH} < 0$, which is indeed the case. The same argument suggests that $^3J_{HH} > 0$, etc.[87]

4.12 Gauge-invariant atomic orbitals (GIAOs)

The coupling constants in practical applications may depend on the choice of vector and scalar potentials. The arbitrariness in the choice of the potentials $\mathbf{A}$ and ϕ ("gauge choice") does not represent any problem for an *atom*, because it is natural to locate the origin on the nucleus. The

[87] Thus, although calculation of the coupling constants is certainly complex, we have in mind a simple model of the nuclear spin–spin interaction that seems to work. We love such models, because they enable us to predict numbers knowing other numbers, or to predict new phenomena. This gives the impression that we understand what happens. This is by no means true. What the electrons are doing and how the spin magnetic moments interact is too complicated, but nevertheless we may suspect the main principles of the game. Such models help us to discuss things with others, to communicate some conjectures, to verify them, and to get more and more confidence in ourselves. Until one day something goes wrong. Then we try to understand why it happened. This *may* require a revision of our model, i.e., a new model, etc.

same reasoning however means a serious problem for a molecule, because even though any choice is equally justified, this justification is only theoretical, not practical. Should the origin be chosen at the center of mass, at the center of the electron cloud, halfway between them, or at another point? *An unfortunate (although mathematically fully justified) choice of the vector potential origin would lead to correct results, but only after calculating and summing up all the contributions to infinity, including application of the complete set of atomic orbitals.* These requirements are too demanding.

4.12.1 London orbitals

Atomic orbitals are used in quantum chemistry as the building blocks of many-electron functions (cf. p. V1-495). The question where to center the orbitals sometimes represents a serious problem. Additionally, in the case of a magnetic field, there is the abovementioned arbitrariness of choice of the vector potential origin. A remedy to the second problem was found by Fritz London[88] in the form of atomic orbitals that depend explicitly on the field applied. Each atomic orbital $\chi\left(\mathbf{r}-\mathbf{R}_C\right)$ centered on nucleus C (with position shown by vector $\mathbf{R}_C$) and describing an electron pointed by vector $\mathbf{r}$ is replaced by the *London orbital* in the following way:

LONDON ATOMIC ORBITALS

$$\chi_L\left(\mathbf{r}-\mathbf{R}_C;\mathbf{A}_C\right)=\exp\left(-i\mathbf{A}_C\cdot\mathbf{r}\right)\chi\left(\mathbf{r}-\mathbf{R}_C\right), \tag{4.98}$$

where $\mathbf{A}_C$ stands for the value of vector field $\mathbf{A}$ at nucleus C, and $\mathbf{A}$ corresponds to the origin O according to (V1-G.13) on p. V1-675, where $\mathbf{H}$ denotes the intensity of a homogeneous magnetic field (no contribution from the magnetic field created by the nuclei, etc.).

As seen, the London orbitals are not invariant with respect to the choice of vector potential origin,

e.g., with respect to shifting the origin of the coordinate system in Eq. (V1-G.13) by vector $\mathbf{R}$,

$$\mathbf{A}'\left(\mathbf{r}\right)=\frac{1}{2}\left[\mathbf{H}\times\left(\mathbf{r}-\mathbf{R}\right)\right]=\mathbf{A}(\mathbf{r})-\frac{1}{2}\left[\mathbf{H}\times\mathbf{R}\right]. \tag{4.99}$$

[88] F. London, *J. Phys. Radium*, 8(1937)397.

Indeed,

$$\chi_L\left(\mathbf{r}-\mathbf{R}_C;\mathbf{A}'_C\right)=\exp\left(-i\mathbf{A}'_C\cdot\mathbf{r}\right)\chi\left(\mathbf{r}-\mathbf{R}_C\right)$$
$$=\exp\left(-i\mathbf{A}_C\cdot\mathbf{r}\right)\exp(i\frac{1}{2}\left[\mathbf{H}\times\mathbf{R}\right]\cdot\mathbf{r})\chi\left(\mathbf{r}-\mathbf{R}_C\right)$$
$$=\exp(i\frac{1}{2}\left[\mathbf{H}\times\mathbf{R}\right]\cdot\mathbf{r})\chi_L\left(\mathbf{r}-\mathbf{R}_C;\mathbf{A}_C\right)\neq\chi_L\left(\mathbf{r}-\mathbf{R}_C;\mathbf{A}_C\right).$$

Despite this property the London orbitals are also known as gauge-invariant atomic orbitals (GIAOs).

4.12.2 Integrals are invariant

Let us calculate the overlap integral S between two London orbitals centered at points C and D. After shifting the origin of the coordinate system in Eq. (V1-G.13) by vector $\mathbf{R}$ we get

$$S=\left\langle\chi_{L,1}\left(\mathbf{r}-\mathbf{R}_C;\mathbf{A}'_C\right)|\chi_{L,2}\left(\mathbf{r}-\mathbf{R}_D;\mathbf{A}'_D\right)\right\rangle=\left\langle\exp\left(-i\mathbf{A}'_C\cdot\mathbf{r}\right)\chi_1|\exp\left(-i\mathbf{A}'_D\cdot\mathbf{r}\right)\chi_2\right\rangle=$$
$$\left\langle\chi_1|\exp\left(-i(\mathbf{A}'_D-\mathbf{A}'_C)\cdot\mathbf{r}\right)\chi_2\right\rangle,$$

i.e., *the result is independent of* $\mathbf{R}$. It turns out[89] that all the integrals needed, i.e., those of kinetic energy, nuclear attraction, and electron repulsion (cf. Appendix V1-Q on p. V1-729), are invariant with respect to an arbitrary shift of the origin of vector potential $\mathbf{A}$.

This means that when we use the London orbitals the results do not depend on the choice of vector potential origin.

Summary

- The Hellmann–Feynman theorem tells us about the rate at which the energy changes when we change parameter P in the Hamiltonian (e.g., the intensity of the electric field). This rate is $\frac{\partial E}{\partial P}=\langle\psi(P)|\frac{\partial\hat{H}}{\partial P}|\psi(P)\rangle$, where $\psi(P)$ means the *exact* solution to the Schrödinger equation (with energy $E(P)$) at value P of the parameter.

[89] T. Helgaker, P. Jørgensen, *J. Chem. Phys.*, 95(1991)2595.

Electric phenomena

- When a molecule is located in a nonhomogeneous electric field the perturbation operator has the form $\hat{H}^{(1)} = -\sum_q \hat{\mu}_q \mathcal{E}_q - \frac{1}{3}\sum_{qq'} \hat{\Theta}_{qq'} \mathcal{E}_{qq'} \ldots$, where $\mathcal{E}_q$ for $q = x, y, z$ denotes the electric field components along the corresponding axes of a Cartesian coordinate system, $\mathcal{E}_{qq'}$ stands for the q' component of the gradient of $\mathcal{E}_q$, while $\hat{\mu}_q, \hat{\Theta}_{qq'}$ stand for the operators of the corresponding components of the dipole and quadrupole moments. In a homogeneous electric field ($\mathcal{E}_{qq'} = 0$) this reduces to $\hat{H}^{(1)} = -\sum_q \hat{\mu}_q \mathcal{E}_q$.
- After using the last expression in the Hellmann–Feynman theorem we obtain the dependence of the dipole moment components on the (weak) field intensity, $\mu_q = \mu_{0q} + \sum_{q'} \alpha_{qq'} \mathcal{E}_{q'} + \frac{1}{2}\sum_{qq'} \beta_{qq'q''} \mathcal{E}_q \mathcal{E}_{q'} \mathcal{E}_{q''} + \ldots$, where μ_{0q} stands for the component corresponding to the isolated molecule, $\alpha_{qq'}$ denotes the q, q' component of the (dipole) polarizability tensor, $\beta_{qq'q''}$ is the corresponding component of the (dipole) first hyperpolarizability tensor, etc. The quantities $\mu_{0q}, \alpha_{qq'}, \beta_{qq'q''}$ in a given Cartesian coordinate system characterize the *isolated* molecule (no electric field) and represent the target of the calculation methods.
- Reversing the electric field direction may in general give different absolute values of the induced dipole moment only because of nonzero hyperpolarizability $\beta_{qq'q''}$ and higher-order hyperpolarizabilities.
- In a nonhomogeneous field we have the following interactions:
  - of the *permanent dipole moment* of the molecule with the electric field, $-\boldsymbol{\mu}_0 \boldsymbol{\mathcal{E}}$,
  - of the *induced dipole moment* proportional to the field ($\sum_{q'} \alpha_{qq'} \mathcal{E}_{q'}$) with the field plus higher-order terms proportional to higher powers of the field intensity involving *dipole hyperpolarizabilities*,
  - of the *permanent quadrupole moment* $\Theta_{qq'}$ of the molecule with the field gradient, $-\frac{1}{3}\sum_{qq'} \Theta_{qq'} \mathcal{E}_{qq'}$,
  - of the *induced quadrupole moment* proportional to the field gradient with the field gradient ($-\frac{1}{4}\sum_{qq'q''q'''} C_{qq'q''q'''} \mathcal{E}_{qq'} \mathcal{E}_{q''q'''}$, the quantity C is called the *quadrupole polarizability*) plus higher-order terms containing *quadrupole hyperpolarizabilities*,
  - higher multipole interactions.
- In the LCAO MO approximation, the dipole moment of the molecule can be decomposed into the sum of the atomic dipole moments and the dipole moments of the atomic pairs.
- The dipole polarizability may be computed by:
  - the *sum over states (SOS) method*, which is based on second-order correction to the energy in the perturbational approach,
  - the *finite field method*, e.g., a variational approach in which the interaction with a weak homogeneous electric field is included in the Hamiltonian. The components of the polarizability are computed as the second derivatives of the energy with respect to the corresponding field components (the derivatives are calculated at zero field). In practical calculations within the LCAO MO approximation we often use the Sadlej relation that connects the shift of a Gaussian atomic orbital with its exponent and the electric field intensity.
- In laser fields we may obtain a series of nonlinear effects (proportional to higher powers of field intensity), including the doubling and tripling of the incident light frequency.

Magnetic phenomena

- An elementary particle has a magnetic dipole moment $\mathbf{M}$ proportional to its spin angular momentum $\mathbf{I}$, i.e., $\mathbf{M} = \gamma \mathbf{I}$, where γ stands for what is called the gyromagnetic factor (characteristic for the kind of particle).
- The magnetic dipole of a particle with spin $\mathbf{I}$ (corresponding to spin quantum number I) in homogeneous magnetic field H has $(2I + 1)$ energy states $E_{m_I} = -\gamma m_I \hbar H$, where $m_I = -I, -I + 1, ..., +I$. Thus, the energy is proportional to H.
- The Hamiltonian of a system in an electromagnetic field has the form $\hat{H} = \sum_{j=1}[\frac{1}{2m_j}\left(\hat{\mathbf{p}}_j - \frac{q_i}{c}\mathbf{A}_j\right)^2 + q_j\phi_j] + \hat{V}$, where $\mathbf{A}$ and ϕ denote the vector and scalar fields (both are functions of position in the three-dimensional space, here they are calculated at particle j) that characterize the external electromagnetic field.
- Potentials $\mathbf{A}$ and ϕ contain, in principle (see Appendix V1-G), the same information as the magnetic and electric fields $\mathbf{H}$ and $\boldsymbol{\mathcal{E}}$. There is an arbitrariness in the choice of $\mathbf{A}$ and ϕ.
- In order to calculate the energy states of a system of nuclei (detectable in NMR spectroscopy) we have to use the Hamiltonian $\hat{H}$ given above supplemented with the interaction of all magnetic moments, related to the orbital and spin of the electrons and the nuclei.
- The refinement is based on classical electrodynamics and the usual quantum mechanical rules for forming operators (Chapter V1-1) or, alternatively, on the relativistic Breit Hamiltonian (p. V1-171). This is how we get the Hamiltonian (4.67) which contains the usual nonrelativistic Hamiltonian plus the perturbation (4.69) with a number of terms (p. 305).
- NMR experimentalists use an empirical Hamiltonian (Eq. (4.83)), in which they have the interaction of the nuclear spin magnetic moments with the magnetic field (the Zeeman effect), the latter weakened by the shielding of the nuclei by the electrons plus the dot products of the nuclear magnetic moments weighted by the coupling constants. The experiment gives both the shielding (σ_A) and the coupling (J_{AB}) constant.
- Nuclear spin coupling takes place through the induction mechanism in the chemical bond (cf. Figs. 4.14 and 4.17). Of key importance for this induction is high electron density at the position of the nuclei (the so-called Fermi contact term, Fig. 4.14).
- The perturbational theory of shielding and coupling constants was given by Ramsey. According to the theory, each quantity consists of diamagnetic and paramagnetic contributions. The diamagnetic term is easy to calculate, the paramagnetic one is more demanding.
- Each of the contributions individually depends on the choice of the origin of the vector potential $\mathbf{A}$, while their sum, i.e., the shielding constant, is invariant with respect to this choice.
- The London atomic orbitals $\chi_L = \exp(-i\mathbf{A}_C \cdot \mathbf{r})\,\chi(\mathbf{r} - \mathbf{R}_C)$ used in calculations for a molecule in a magnetic field depend explicitly on that field, through the value $\mathbf{A}_C$ of the vector potential $\mathbf{A}$ calculated at the center $\mathbf{R}_C$ of the usual atomic orbital $\chi(\mathbf{r} - \mathbf{R}_C)$.
- The most important feature of London orbitals is that all the integrals appearing in calculations are invariant with respect to the origin of the vector potential. This is why results obtained using London orbitals are also independent of that choice.

Main concepts, new terms

atomic dipoles (p. 279)
Bohr magneton (p. 295)
bond dipoles (p. 279)
Cartesian multipole moments (p. 263)
chemical shift (p. 313)
coupling constant (p. 307, 322)
coupling mechanism (p. 325)
diamagnetic effect (p. 321)
diamagnetic spin–orbit contribution (p. 323)
dipole hyperpolarizability (p. 269)
dipole, quadrupole, octupole moments (p. 263)
dipole polarizability (p. 269)
direct spin–spin interaction (p. 322)
empirical NMR Hamiltonian (p. 307)
Fermi contact contribution (p. 323)
finite field method (p. 284)
Gauge-invariant atomic orbital (GIAO) (p. 327)
gyromagnetic factor (p. 295)
Hellmann–Feynman theorem (p. 256)
homogeneous electric field (p. 266)
nonhomogeneous electric field (p. 263)
intermediate spin–spin coupling (p. 307)
linear response (p. 268)
local field (p. 307)
London orbitals (p. 327)
magnetic dipole (p. 295)
Maxwell equations (p. V1-673)
multipole hyperpolarizability (p. 268)
multipole moments (p. 264)
multipole polarizability (p. 268)
NMR (p. 307)
NMR Hamiltonian (p. 307)
nonlinear response (p. 268)
nuclear magneton (p. 295)
oscillating electric field (p. 290)
paramagnetic effect (p. 321)
paramagnetic spin–orbit (p. 323)
Ramsey theory (p. 319)
Sadlej relation (p. 285)
second/third harmonic generation (p. 290)
shielding constants (p. 307)
spin–dipole contribution (p. 323)
spin magnetic moment (p. 295)
sum over states method (p. 281)
traceless multipole moments (p. 264)
zero-differential overlap (ZDO) (p. 280)

From the research front

The electric dipole (hyper)polarizabilities are not easy to calculate, because:

- the sum over states (SOS) method converges slowly, i.e., a huge number of states have to be taken into account, including those belonging to a continuum;
- the finite field method requires a large quantity of atomic orbitals with small exponents (they describe the lion's share of the electron cloud deformation), although, being diffuse, they do not contribute much to the minimized energy (and lowering the energy is the only indicator that tells us whether a particular function is important or not).

More and more often in their experiments chemists investigate *large* molecules. Such large objects cannot be described by "global" polarizabilities and hyperpolarizabilities (except perhaps optical properties, where the wave length is often much larger than the size of the molecule). How such large molecules function (interacting with other molecules) depends first of all on their local properties. We have to

Table 4.1. Comparison of theoretical and experimental shielding constants. The shielding constant σ_A (unitless quantity) is as usual expressed in ppm, i.e., the number given has to be multiplied by 10^{-6} to obtain σ_A of Eq. (4.84). The Hartree–Fock, MP2, and MP4 results are calculated in J. Gauss, *Chem. Phys. Letters*, 229(1994)198; the CCSD(T) in J. Gauss, J.F. Stanton, *J. Chem. Phys.*, 104(1996)2574; and the CASSCF in K. Ruud, T. Helgaker, R. Kobayashi, P. Jørgensen, K.L. Bak, H.J. Jensen, *J. Chem. Phys.*, 100(1994)8178. For the Hartree–Fock method, see Chapter V1-8; for the other methods mentioned here, see Chapter 2. The references to the corresponding experimental papers are given in T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev.*, 99(1999)293. The experimental error is estimated for σ_H in ammonia as ± 1.0, for σ_O as ± 17.2, for σ_H in water as ± 0.015, for σ_F as ± 6, and for σ_H in hydrogen fluoride as ± 0.2.

Method	CH_4		NH_3		H_2O		HF	
	σ_C	σ_H	σ_N	σ_H	σ_O	σ_H	σ_F	σ_H
Hartree–Fock	194.8	31.7	262.3	31.7	328.1	30.7	413.6	28.4
MP2	201.0	31.4	276.5	31.4	346.1	30.7	424.2	28.9
MP4	198.6	31.5	269.9	31.6	337.5	30.9	418.7	29.1
CCSD(T)	198.9	31.6	270.7	31.6	337.9	30.9	418.6	29.2
CASSCF	200.4	31.19	269.6	31.02	335.3	30.21	419.6	28.49
Experiment	**198.7**	**30.61**	**264.54**	**31.2**	**344.0**	**30.052**	**410**	**28.5**

Table 4.2. Comparison of theoretical and experimental spin–spin coupling constants ${}^nJ_{AB}$ for ethylene (n denotes the number of separating bonds) in Hz. For the methods used, see Chapter 2. All references to experimental and theoretical results can be found in T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev.*, 99(1999)293.

Method	Spin–spin coupling constants J_{AB} for ethylene in Hz					
	${}^1J_{CC}$	${}^1J_{CH}$	${}^2J_{CH}$	${}^2J_{HH}$	${}^3J_{HH-cis}$	${}^3J_{HH-trans}$
MC SCF	71.9	146.6	−3.0	−2.7	10.9	18.1
EOM-CCSD	70.1	153.23	−2.95	0.44	11.57	17.80
Experiment	**67.457**	**156.302**	**−2.403**	**2.394**	**11.657**	**19.015**

replace such characteristics by new ones offering atomic resolution, similar to those proposed in the techniques of Stone or Sokalski (p. 598), where individual atoms are characterized by their multipole moments, polarizabilities, etc.

Not long ago, the shielding and especially spin–spin coupling constants were very hard to calculate with reasonable accuracy. Nowadays these quantities are computed routinely using commercial software with atomic London orbitals (or other than GIAO basis sets).

The current possibilities of the theory in predicting the nuclear shielding constants and the nuclear spin–spin coupling constants are shown in Tables 4.1 and 4.2.

Note that the accuracy of the theoretical results for shielding constants is nearly the same as that of experiment. As to the spin–spin coupling constants, the theoretical results are only slightly off experimental values.

Ad futurum

It seems that the SOS method will be gradually sent out of business. The finite field method (in the electric field responses) will become more and more important, due to its simplicity. It remains however to solve the problem of processing the information we get from such computations and translating it into the abovementioned local characteristics of the molecule.

Contemporary numerical methods allow routine calculation of polarizability. It is difficult with the hyperpolarizabilities that are much more sensitive to the quality of the atomic basis set used. The hyperpolarizabilities relate to nonlinear properties, which are in high demand in new materials for technological applications.

Problems such as the dependence of the molecular spectra and of the molecular conformations and structure on the external electric field (created by our equipment or by a neighboring molecule) will probably become more and more important. This may pertain especially the femtosecond spectroscopy, where the laser electric fields are very strong.

The theory of the molecular response to an electric field and the theory of the molecular response to a magnetic field look, despite some similarities, as if they were "from another story." One of the reasons is that the electric field response can be described by solving the Schrödinger equation, while that corresponding to the magnetic field is based inherently on relativistic effects, much less investigated beside some quite simple examples. Another reason may be the scale difference; the electric effects are much larger than the magnetic ones.

However, the theory for the interaction of matter with the electromagnetic field has to be coherent. The finite field method, so gloriously successful in electric field effects, is in the "stone age" stage for magnetic field effects. The propagator methods[90] look the most promising; these allow for easier calculation of NMR parameters than the SOS methods.

Additional literature

A.D. Buckingham, *Advan. Chem. Phys*., 12(1967)107.

A classical paper on molecules in a static or periodic electric field.

H.F. Hameka, "Advanced Quantum Chemistry. Theory of Interactions between Molecules and Electromagnetic Fields," Addison-Wesley Publish. Co., Reading, Massachusetts, USA (1965).

This is a first-class book, although it presents the state of the art before the *ab initio* methods for calculating the magnetic properties of molecules.

T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev*., 99(1999)293.

A review article on the magnetic properties of molecules (NMR) with a presentation of suitable contemporary theoretical methods.

[90] J. Linderberg, Y. Öhrn, "*Propagators in Quantum Chemistry*," Second edition, John Wiley and Sons, Ltd., 2004.

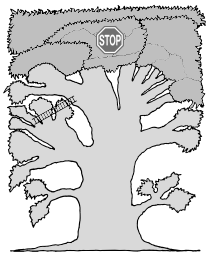
Questions

1. The Hellmann–Feynman theorem says that $\frac{\partial E}{\partial P} = \langle\psi|\frac{\partial \hat{H}}{\partial P}|\psi\rangle$ (where $\hat{H}$ is the Hamiltonian that depends on parameter P). This is true, when:
   a. ψ stands for the ground-state wave function.
   b. ψ is of class Q.
   c. ψ is the exact Hartree–Fock function.
   d. ψ is an eigenfunction of $\hat{H}$.
2. In the expression for the energy of a molecule in an electric field:
   a. a component of the quadrupole moment is multiplied by the gradient of the field.
   b. a component of the dipole moment is multiplied by the intensity of the field.
   c. the dipole hyperpolarizability represents a coefficient at the third power of the electric field intensity.
   d. the dipole polarizability represents a coefficient at the square of the electric field intensity.
3. A nonpolar molecule (but having a nonzero quadrupole moment) in the field with a nonzero gradient:
   a. will interact with the field.
   b. will orient in such a way as to have the dipole moment along the electric field.
   c. will orient in such a way as to have the longer axis of the quadrupole along the field.
   d. will orient in such a way as to have the longer axis of the quadrupole along the gradient of the field.
4. The second harmonic generation in a uniform electric field depends on the following molecular property:
   a. dipole hyperpolarizability β.
   b. quadrupole and octupole polarizability.
   c. octupole hyperpolarizability.
   d. dipole hyperpolarizability γ.
5. In some variational calculations for a molecule without electric field the positions of the atomic orbitals have been optimized. In the finite field variational method a small shift of a certain Gaussian atomic orbital off a nucleus:
   a. will always increase the energy.
   b. will always decrease energy, while at a larger shift the energy will increase.
   c. may decrease the energy if the shift is opposite to the electric field.
   d. will decrease the energy if the sufficiently small shift pertains to all atomic orbitals and is opposite to the electric field.
6. The magnetic moment $\mathbf{M}$ (of an elementary particle):
   a. is the same as its spin angular momentum.
   b. has the same length for the electron and for the proton.
   c. interacts with uniform magnetic field $\mathbf{H}$, and the interaction energy is equal to $\frac{1}{2}\mathbf{H}\cdot\mathbf{M}$.
   d. interacts with uniform magnetic field $\mathbf{H}$, and the interaction energy is equal to $-\mathbf{H}\cdot\mathbf{M}$.
7. The hydrogen molecule in its electronic singlet state:
   a. when in magnetic field acquires an admixture of the triplet electronic state.
   b. may be of two kinds depending on the singlet or triplet state of its nuclei.
   c. has electron spins forming an angle of 180°.
   d. must have opposite nuclear spins.

8. The vector potential **A** of the electromagnetic field corresponds to the uniform magnetic field **H**. Then:
   a. $\mathbf{A} = \mathrm{rot}\mathbf{H}$.
   b. **A** may be chosen in such a way as to satisfy the Maxwell equation $\mathbf{H} = \mathrm{rot}\mathbf{A}$.
   c. **A** also represents a uniform vector field.
   d. **A** and $\mathbf{A} - \nabla \exp(-x^2)$ give the same magnetic field **H**.
9. The shielding constant for a nucleus consists of the diamagnetic and paramagnetic parts:
   a. the diamagnetic part depends on the choice of the origin of the vector potential **A**.
   b. the paramagnetic part depends on the choice of the origin of the vector potential **A**.
   c. as physical effects none of them can depend on the choice of the origin of the vector potential **A**.
   d. the sum of these parts does not depend on the choice of the origin of the vector potential **A**.
10. The value of the London orbital χ_L $(\mathbf{r} - \mathbf{R}; \mathbf{A})$ calculated at the point indicated by vector **r** depends on:
   a. the magnetic field at point **R**.
   b. the vector potential at point **r**.
   c. the vector potential at point $\mathbf{r} - \mathbf{R}$.
   d. the vector potential **A** at the point indicated by **R**.

Answers

1a,d, 2a,b,c,d, 3a,d, 4a,d, 5c,d, 6d, 7a,b,c, 8b,d, 9a,b,d, 10d



CHAPTER 5

Intermolecular Interactions

Remember when discoursing about water, to induce first experience, then reason.
from notes by Leonardo da Vinci *(*1452–1519*)*

Where are we?

We are already in the crown of the TREE.

An example

Look at a snow flake on your hand. Why does such a fascinating, regular structure exist? And why does it then change to a few drops of water? Why do water molecules stick together? Visibly they attract each other *for some reason*. The interaction must not be too strong, however, since the snow flake transforms so easily and then the water drops evaporate (without destroying the water molecules!).

What is it all about?

THEORY OF INTERMOLECULAR INTERACTIONS

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00013-2
Copyright © 2020 Elsevier B.V. All rights reserved.

- Accuracy should be the same
- Basis set superposition error (BSSE)
- Good and bad news about the supermolecular method

Perturbational approach (△) **p. 351**

- Intermolecular distance – what does it mean?
- Polarization approximation (two molecules)
- Intermolecular interactions: physical interpretation
- Electrostatic energy in the multipole representation plus penetration energy
- Induction energy in the multipole representation
- Dispersion energy in the multipole representation
- Resonance interaction – excimers

Symmetry-adapted perturbation theory (SAPT) (♦ ℧ Ⓢ) **p. 377**

- Polarization approximation is illegal
- Constructing a symmetry-adapted function
- The perturbation is always large in polarization approximation
- Iterative scheme of SAPT
- Symmetry forcing
- A link to the variational method – the Heitler–London interaction energy
- Summary: the main contributions to the interaction energy

Convergence problems (♦ ℧ Ⓢ) **p. 392**

- Padé approximants may improve convergence

Nonadditivity of intermolecular interactions (△) **p. 398**

- Interaction energy represents the nonadditivity of the total energy
- Many-body expansion of the rigid interaction energy
- What is additive, what is not?
- Additivity of the electrostatic interaction
- Exchange nonadditivity
- Induction nonadditivity
- Additivity of the second-order dispersion energy
- Nonadditivity of the third-order dispersion energy

ENGINEERING OF INTERMOLECULAR INTERACTIONS

Idea of the molecular surface (△) **p. 411**

- Van der Waals atomic radii
- A concept of molecular surface
- Confining molecular space – the nanovessels
- Molecular surface under high pressure

Decisive forces (△) **p. 414**

- Distinguished role of the valence repulsion and electrostatic interaction

- Hydrogen bond
- Coordination interaction
- Electrostatic character of molecular surface – the maps of the molecular potential
- Hydrophobic effect

Construction principles (△) **p. 424**

- Molecular recognition – synthons
- "Key-and-lock," template, and "hand-and-glove" synthon interactions
- Convex and concave – the basics of strategy in the nanoscale

Chapter V1-8 dealt with the question of why atoms form molecules. Electrons and nuclei attract each other, and this results in almost exact electrical neutralization of matter. Despite this neutrality, atoms and molecules interact, because:

- two atoms or molecules cannot occupy the same space,
- electrons and nuclei in an atom or molecule may still interact with those in other atoms or molecules. This chapter will tell us about the very reason for and details of the interaction.

Why is this important?

What is the most important message humanity ever learned about matter? According to Richard Feynman the message would be: "*The world is built of atoms, which repel each other at short distances and attract at longer ones*." If the intermolecular interactions were suddenly switched off, the world would disintegrate in about a femtosecond, that is, in a single period of atomic vibration (the atoms simply would not come back when shifted from their equilibrium positions). Soon after, everything would evaporate and a sphere of gas, the remainder of the Earth, would be held together by gravitational forces. Is that not enough?

What is needed?

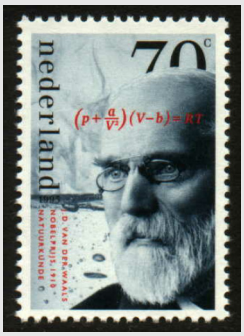
- Perturbation theory (Chapter V1-5, absolutely necessary),
- variational method (Chapter V1-5, recommended),
- Appendix G, p. 613 (necessary),
- Many-body perturbation theory (MBPT) (Chapter 2, p. 165, necessary),
- reduced resolvent (Chapter 2, p. 165, necessary),
- Appendix E, p. 601 (recommended),
- Appendix V1-T (mentioned).

Classical works

Democritus of Abdera (ca. 460 BCE–ca. 370 BCE), Greek philosopher, formulated the first atomic theory. Its traces go further back in time, but it is Democritus who produced a much more elaborated picture. According to him, Nature represents a constant motion of indivisible and permanent particles (atoms), whose interactions result in various materials. *It turned out after almost 25 centuries that this hypothesis was basically correct!* All the written works of Democritus have been lost, but his ideas continued to have an important impact on science for centuries.

Rudjer Josip Bosković (1711–1787), Croatian physicist, mathematician, astronomer, and philosopher from the beautiful Dubrovnik. His expertise in statics and mechanics decided that it was him who was chosen for reparation of masterpieces such as the St. Peter's basilica dome (of 42 meters diameter).

Johannes Diderik van der Waals (1837–1923), Dutch physicist, professor at the University of Amsterdam. His research topic was the influence of intermolecular forces on the properties of gases (equation of state of the real gas, 1873) and liquids. In 1910 van der Waals received the Nobel Prize "*for his work on the equation of state for gases and liquids*."



Such an important subject was recognized very early. The idea that the cohesion of matter stems from the interaction of small indivisible particles ("atoms") comes from Democritus. ★ An idea similar to that cited by Feynman was first stated clearly by the Croatian scientist Rudjer Bosković in "*Theoria Philosophiae naturalis*," Venice, 1763. ★ Padé approximants were first proposed in the PhD thesis of Henri Padé entitled "*Sur la représentation approchée d'une fonction pour des fractions rationnelles*," which was published in *Annales Scientifique de l'École Normale Supérieure, Suppl. [3]*, 9(1892)1. ★ The role of intermolecular interactions was highlighted in the work of Johannes Diderik van der Waals, especially in "*Die Kontinuität des gasformigen und flüssigen Zustandes*," Barth, Leipzig (1899, 1900). From that time on, intermolecular interactions are often called van der Waals interactions. ★ Determination of parenthood of the hydrogen bond idea is

not an easy task. There are reasonable indications that the idea comes from the work "*Über Haupt- und Nebenvalenzen und die Constitution der Ammoniumverbindungen*" in "*Liebig's Annalen der Chemie*," 322(1902)261, by the Swiss organic chemist Alfred Werner, the father of coordination chemistry and 1913 Nobel Prize winner. ★ The concept of ionic radii was first proposed by Linus Pauling in "*The Sizes of Ions and the Structure of Ionic Crystals*," *Journal of the American Chemical Society*, 49(1927)765. ★ The quantum mechanical explanation of intermolecular forces, including the ubiquitous dispersion interactions, was given by Fritz London in "*Zur Theorie und Systematik der Molekularkräfte*," *Zeitschrift für Physik*, 63(1930)245 and in "*Über einige Eigenschaften und Anwendungen der Molekularkräfte*" from *Zeitschrift für Physikalische Chemie (B)*, 11(1930)222. ★ The hydrophobic effect was first highlighted by Walter Kauzmann in the paper "*Some Factors in the Interpretation of Protein Denaturation*," in *Advances in Protein Chemistry*, 14(1959)1. The effect was further elaborated by George Nemethy, Harold Scheraga, Frank Stillinger, and David Chandler among others. ★ Resonance interactions were first described by Robert S. Mulliken in the article "*The Interaction of Differently Excited Like Atoms at Large Distances*," in *Physical Reviews*, 120(1960)1674. ★ Bogumił Jeziorski and Włodzimierz Kołos extended the existing theory of intermolecular forces to intermediate distances ("*On the Symmetry Forcing in the Perturbation Theory of Weak Intermolecular Interactions*," *International Journal of Quantum Chemistry*, 12 Suppl. 1, (1977)91).

* * *

THEORY OF INTERMOLECULAR INTERACTIONS

There are two principal methods of calculating the intermolecular interactions: the supermolecular method and the perturbational method. Both assume the Born–Oppenheimer approximation.

We may all agree about what the total system under consideration should be.[1] Any idea of interaction poses one fundamental question: what kind of objects interact? The answer represents our arbitrary decision and has profound consequences. Even if we may be able to describe perfectly the total system for a large spectrum of our choices, the work needed may depend critically on the choice of the interacting objects.

5.1 Idea of the rigid interaction energy

The configuration of the nuclei of the total system can be defined by a set of coordinates given by vector $\mathbf{R}$. We divide the whole system into the interacting subsystems ("molecules"): $A, B, C, ...$, with their *internal* geometries (configurations of the nuclei) defined by $\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, ...$ and the fixed numbers of electrons $N_A, N_B, N_C, ...$ The rest of the

[1] This is always a compromise. We (quite arbitrarily) cut the system out from the Universe and say that it does not interact with the rest of the Universe...

coordinates ("external") that determine the intermolecular distances and the orientations of the molecules in a global coordinate system will be denoted as $\mathbf{R}_{ex}$. We have

$$\text{all coordinates} = \text{external} + \text{internal}.$$

Let us define the rigid interaction energy at the configuration $\mathbf{R}$ of the nuclei as

$$E_{int}\left(\mathbf{R}_{ex}; \mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, ...\right) = E_{ABC...}\left(\mathbf{R}\right) - \left[E_A\left(\mathbf{R}_A\right) + E_B\left(\mathbf{R}_B\right) + E_C\left(\mathbf{R}_C\right) + ...\right], \quad (5.1)$$

where the notation $E_{int}\left(\mathbf{R}_{ex}; \mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, ...\right)$ means that E_{int} is a function of $\mathbf{R}_{ex}$ and depends parametrically on $\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, ...$, i.e., $\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, ...$ *are fixed*. $E_{ABC...}$ is the ground-state electronic energy (corresponding to $E_0^{(0)}$ from Eq. (V1-6.22)) of the total system, and $E_A\left(\mathbf{R}_A\right), E_B\left(\mathbf{R}_B\right), E_C\left(\mathbf{R}_C\right), ...$ are the electronic energies of the n subsystems (molecules), calculated *at the same positions of the nuclei* as those in the total system.

This definition implies that the interaction energy represents just a theoretical concept, not a measurable quantity. Indeed, we calculate $E_{int}(\mathbf{R})$ for any geometry $\mathbf{R}$ we wish to consider, which may have nothing to do with the optimized geometry of the total system or the geometry of the isolated molecules. We will see in a while that there are also other reasons why E_{int} cannot be measured.

5.2 Idea of the internal relaxation

One may modify the concept of the rigid interaction energy (Eq. (5.1)) by allowing relaxation of the individual molecules at a fixed $\mathbf{R}_{ex}$. This seems to be physically appealing since chemists often think of molecules as approaching each other, and during such an approach the molecules *reorient and change* due to the interaction. This is not shown explicitly when calculating the rigid interaction energy, because we see the relaxation of the electronic structures of the interacting molecules, but we do keep the nuclear frameworks of the individual molecules unrelaxed. The geometry of the individual molecules should change when they approach.

The internally relaxed interaction energy $E_{int}^{relax}\left(\mathbf{R}_{ex}\right)$ may be defined by

$$E_{int}^{relax}\left(\mathbf{R}_{ex}\right) = \min_{\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C ...} E_{ABC...}\left(\mathbf{R}\right) - \left[E_A\left(\mathbf{R}_A^0\right) + E_B\left(\mathbf{R}_B^0\right) + E_C\left(\mathbf{R}_C^0\right) + ...\right], \quad (5.2)$$

where $\min_{\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C ...} E_{ABC...}\left(\mathbf{R}\right)$ means that the minimum of $E_{ABC...}$ is achieved by simultaneous optimization of the internal coordinates (i.e., internal geometries) $\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C ...$ at all the other coordinates fixed ($\mathbf{R}_{ex}$). In the second term are the electronic energies of the isolated molecules with the optimized geometries $\mathbf{R}_A^0, \mathbf{R}_B^0, \mathbf{R}_C^0, ...$ Thus, in general both terms in Eq. (5.2) differ from those of Eq. (5.1).

After minimization $\min_{\mathbf{R}_A,\mathbf{R}_B,\mathbf{R}_C...} E_{ABC...}(\mathbf{R})$ we get the total electronic energy as a function of $\mathbf{R}_{ex}$, with the molecules $A, B, C, ...$ distorted by the interaction (their geometry differing from $\mathbf{R}_A^0, \mathbf{R}_B^0, \mathbf{R}_C^0, ...$).[2]

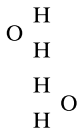
The relaxed interaction energy idea may be extended still further by allowing also energy optimization within a subset of the $\mathbf{R}_{ex}$ coordinates. For example, one of the choices may allow optimization of the rotational degrees of freedom of each molecule (still keeping other degrees of freedom in $\mathbf{R}_{ex}$ fixed).

5.3 Interacting subsystems

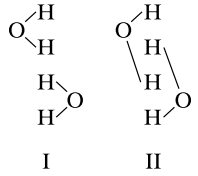
5.3.1 Natural division

Although the notion of interaction energy is of great practical value, its theoretical meaning is a little bit fuzzy. Right at the beginning we have a question: interaction *of what*? We view the system *as composed of particular subsystems*, which, once isolated, have to be put together.

For instance, the supersystem

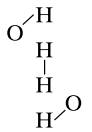


may be considered as two interacting water molecules, but even then we still have an uncertainty, i.e., whether the two molecules correspond to (I) or to (II):



In addition the system might be considered as composed of a hydrogen molecule interacting with two OH radicals:

[2] The minimization performed may be the subject of the multiple minima problem (see Chapter V1-7). In such a case one may get a nonunique, though still physically meaningful, $E_{int}^{relax}(\mathbf{R}_{ex})$ function, depending on which of the multiple minima has been achieved. This physical meaning is not complete since one has to take into account the zero vibration energy (we will do it in a while).



etc.

Choice of subsystems is of no importance from the point of view of mathematics, but is of crucial importance from the point of view of calculations in theoretical chemistry.

The particular choice of subsystem should depend on the kind of experiment with which we wish to compare our calculations, e.g.:

- we are interested in the interaction of water molecules when studying water evaporation or freezing;
- we are interested in the interaction of atoms and ions that exist in the system when heating water to 1000°C.

Let us stress that in any case when choosing subsystems we are forced to single out *particular atoms* belonging to subsystems[3] A and B. It is not sufficient to define the kind of molecules participating in the interaction (see our two first examples for $n = 2$).

If, when dividing a system into n subsystems in two ways (I and II), we obtain $|E_{int}|_I < |E_{int}|_{II}$, then division I will be called more *natural* than division II.

5.3.2 What is most natural?

Which division is *most* natural? We do not have any experience in answering such questions. What? Why should we have any difficulties? It is sufficient to consider all possible divisions and to choose the one which requires the lowest separation energy. Unfortunately, this is not so obvious. Let us consider two widely separated water molecules (Fig. 5.1a).

Right between the molecules, e.g., in the middle of the O–O separation, we place two point charges $q > 0$ and $-q$, i.e., we place nothing, since the charges cancel each other out (please compare with a similar trick on p. 74). We therefore just have two water molecules. Now we start our game. We say that the charges are *real*: one belongs to one of the molecules and the other to the second (Fig. 5.1b). The charge q could be anything, but we want to use it for a very special goal: to construct the two subsystems in a *more natural* way than just two water molecules. It is interesting that after the choice is made, any of the subsystems has *lower* energy than that of isolated water, since the molecules are oriented in such a way as to attract each other. This means that the value of q can be chosen from an interval making the choice of subsystems more natural. For a certain $q = q_{opt}$ we would obtain as the interaction energy

[3] This means that interaction energy idea belongs to classical concepts. In a quantum system particles of the same kind are indistinguishable. A quantum system does not allow us to separate a part from the system. Despite this, the interaction energy idea is important and useful.

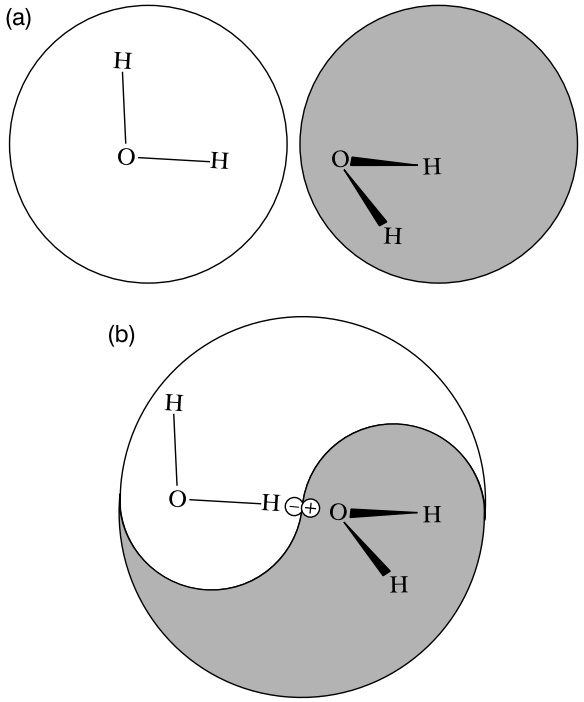


Fig. 5.1. Part–entity relationship. Two distinct ways of dividing the $(H_2O)_2$ system into subsystems. Division (a) is traditional. The interacting objects are two isolated water molecules and the interaction energy is equal to about -5 kcal/mol (attraction). Division (b) is more subtle. We wish to treat a certain point in space of charge zero as composed of two fictitious charges $q > 0$ and $-q$, and one of the charges is ascribed to one, and the other to the other molecule. In this way two *new subsystems* are defined, each of them composed of the water molecule *and* the corresponding point charge. The value of q may be chosen in such a way as to produce the interaction energy of the new subsystems close to 0. Therefore this is a more natural choice of subsystems than the traditional one. The total interaction energy of the two water molecules is now absorbed within the interactions of the fictitious point charges with "their" water molecules. Each of the point charges takes over the interaction of "its" water molecule with the rest of the Universe. Hence, I have permitted myself (with the necessary *licentia poetica*) to use the *yin* and *yang* symbols – the two basic elements of ancient Chinese philosophy.

of the new subsystems $E'_{int} = 0$. This certainly would be the most natural choice,[4] with the "dressed" water molecules, not seeing each other.[5]

[4] Although not unique, since the charges could be chosen at different points in space, and we could also use point multipoles, etc.

[5] Allusion to the elementary particles "dressed" by interactions (see section "*What is it all about?*", Chapter V1-8). It is worth noting that we have to superpose the subsystems *first* (then the fictitious charges disappear), and *then* calculate the interaction energy of the water molecules deformed by the charges.

In the remainder of this chapter we will not use any fictitious charges.

5.4 Binding energy

Interaction energy can be calculated at any configuration **R** of the nuclei. We may ask whether any "privileged" configuration exists with respect to the interaction energy. This was our subject in Chapter V1-7, and it turned out that the electronic energy may have many minima (equilibria) as a function of **R**. For each of such configurations we may define the *binding energy with respect to a particular dissociation channel* as the difference of the corresponding interaction energies (all subsystems at the optimal positions $\mathbf{R}_{opt(j)}$ of the nuclei with respect to the electronic energy E_j), i.e.,

$$E_{bind} = E_{ABC...}\left(\mathbf{R}_{opt(tot)}\right) - \sum_{j=A,B,C,...} E_j\left(\mathbf{R}_{opt(j)}\right) = \min_{\mathbf{R}_{ex}} E_{int}^{relax}\left(\mathbf{R}_{ex}\right). \tag{5.3}$$

At a given configuration $\mathbf{R}_{opt(tot)}$ we usually have many dissociation channels differing by the possible products and their quantum mechanical states.

5.5 Dissociation energy

The calculated interaction energy of Eq. (5.1) and the binding energies are only theoretical quantities and cannot be measured. The measurable quantity is the *dissociation energy*

$$E_{diss} = E_{bind} + [\Delta E_{0,tot} - \sum_{j=A,B,C,...} \Delta E_{0,j}], \tag{5.4}$$

where $\Delta E_{0,tot}$ stands for what is known as the *zero vibration energy* of the total system (cf. p. V1-427) at the equilibrium geometry $\mathbf{R}_{opt(tot)}$, and $\Delta E_{0,j}$ for $j = A, B, C, ...$ represents the zero-vibration energies for the subsystems. In the harmonic approximation $\Delta E_{0,tot} = \frac{1}{2}\sum_i h\nu_{i,tot}$ and $\Delta E_{0,j} = \frac{1}{2}\sum_i h\nu_{i,j}$, ... at their equilibrium geometries $\mathbf{R}_{opt(j)}$.

Let us stress that the formula $\frac{1}{2}h\nu$ for the zero-vibration energy is related to the harmonic approximation.[6] Generally,

> the zero-vibration energy has to be determined as the difference between the lowest vibrational energy level and the energy of the bottom of the potential well.

[6] In which ν is well defined.

5.6 Dissociation barrier

If a molecule receives dissociation energy it is most often a sufficient condition for its dissociation (Fig. 5.2a). Sometimes however the dissociation energy is too low to get dissociation, and the reason is that there is an energy barrier to be overcome (Fig. 5.2b). The barrier may be high and the system is stable even if the dissociation products have (much) lower energy. Catenans, rotaxans, and endohedral complexes shown in Fig. 5.2c may serve as examples.

The energy necessary to overcome the barrier from the trap side is equal to

$$E_{bar} = E_{\#} - \left(E_{\min} + \frac{1}{2} \sum_j h\nu_j \right),$$

where $E_{\min}$ is the energy of the bottom of the well, $E_{\#}$ represents the barrier top energy, and $\frac{1}{2}\sum_j h\nu_j$ is the zero-vibration energy of the well.

5.7 Supermolecular approach

In the supermolecular method the interaction energy is calculated from its definition (5.1) using any reliable method of the electronic energy calculation. For the sake of brevity we will consider the interaction of only two subsystems: A and B.

5.7.1 Accuracy should be the same

There is a problem though. We are unable to solve the Schrödinger equation exactly either for AB or for A or for B. We have to use approximations. If so, we have to worry about *the same accuracy* of calculation for AB, A, and B. From this we may expect that

> in determining E_{AB} as well as E_A and E_B the same theoretical method should be preferred, because any method introduces its own systematic error and we may hope these errors will be canceled out at least partially in the above formula.

This problem is already encountered at the stage of basis set choice. For example, suppose we have decided to carry out the calculations within the Hartree–Fock method in the LCAO MO approximation. The same method has to be used for AB, A, and B. But what does this really mean? Which of the two protocols should be used?

1. Consider the atomic basis set Ω that consists of the atomic orbitals centered on the nuclei of A (set Ω_A) and on the nuclei of B (set Ω_B), i.e., $\Omega = \Omega_A \cup \Omega_B$.
2. Calculate E_{AB} using Ω, E_A using Ω_A, and E_B using Ω_B (Fig. 5.3a).

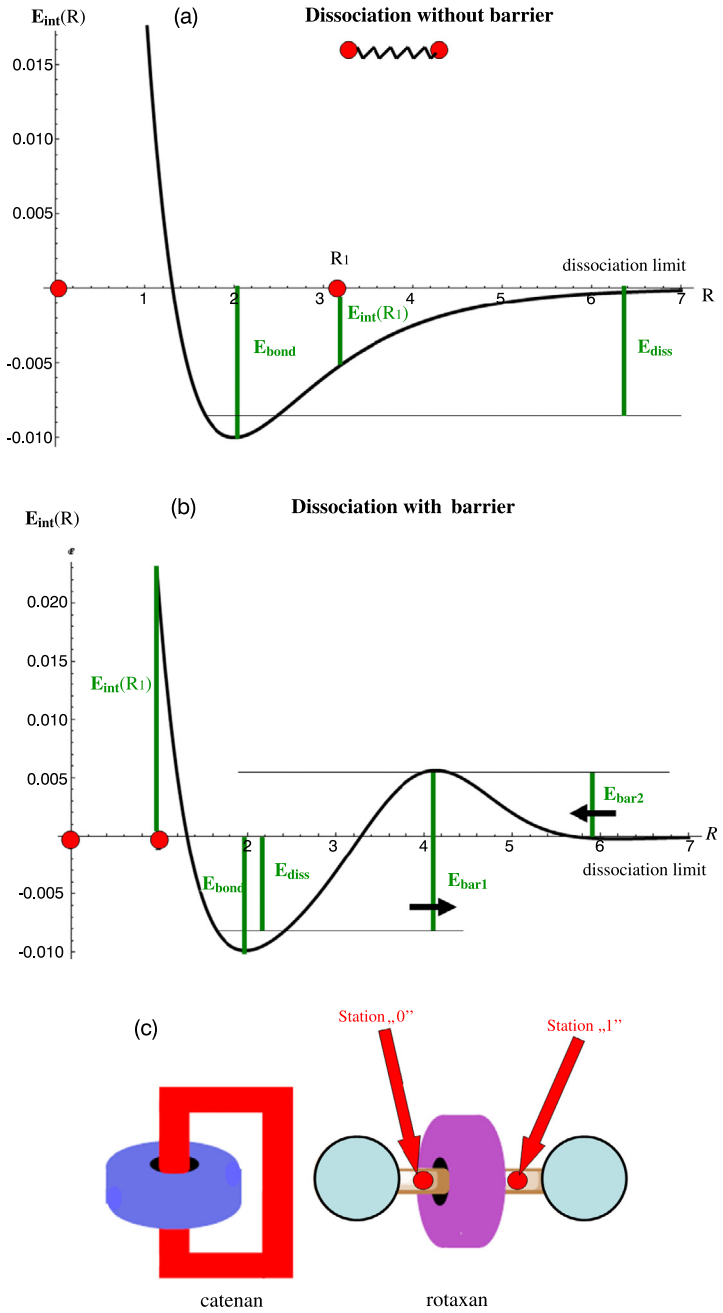


Fig. 5.2. Interaction energy E_{int}, binding energy E_{bind}, dissociation energy E_{diss}, and barrier energy E_{bar}. (a) Common situation: the interaction energy of two atoms or molecules (circles) represents a simple function of their distance R. (b) A more complex situation: there is a barrier for dissociation of the complex (E_{bar1}) and a barrier (of height E_{bar2}) for approaching the atoms. (c) Two parts of the molecule are interlocked. In such a situation like a catenan (two interlocked rings) and a rotaxan (a ring moving along a wire) the E_{bar1} energy is very large.

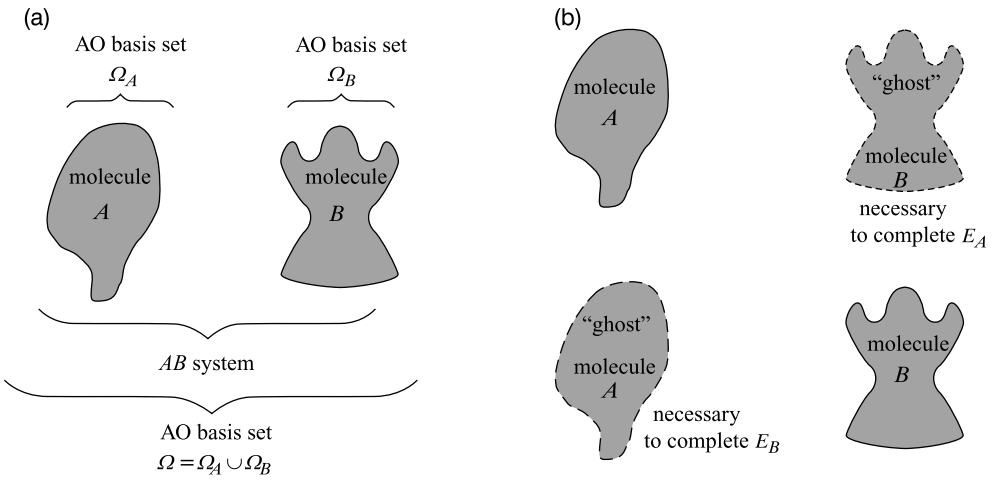


Fig. 5.3. (a) Basis set superposition error (BSSE) problem. Each of the molecules offers its own atomic orbitals to the total basis set $\Omega = \Omega_A \cup \Omega_B$. (b) The counterpoise method, in which the calculations for a single subsystem are performed within the full atomic basis set Ω: the orbitals centered on it and what are called ghost orbitals centered on the partner.

Apparently everything looks logical, but *we did not use the same method* when calculating the energies of AB, A, and B. The basis set used has been different depending on what we wanted to calculate. Thus, it seems more appropriate to calculate all three quantities using the same basis set Ω.

5.7.2 Basis set superposition error (BSSE)

Such an approach is supported by the following reasoning. When the calculations for E_{AB} are performed within the basis set Ω we calculate implicitly not only the interaction energy, but we also allow the individual subsystems to lower their energy – with no physical reason whatsoever – and this lowering corresponds to the $\Omega - \Omega_A$ for A or $\Omega - \Omega_B$ for B. Conclusion: by subtracting from E_{AB} the energies E_A calculated with Ω_A and E_B calculated with Ω_B we are left not only with the interaction energy (as should be), but also with an unwanted and nonphysical extra term (an error) connected with the artificial lowering of the subsystems' energies when calculating E_{AB}. This error is called the *basis set superposition error* (BSSE).

...and the remedy

To remove the BSSE we may consider the use of the basis set Ω not only for E_{AB} but also for E_A and E_B. This procedure, called the *counterpoise method*, was first introduced by Boys and

Bernardi.[7] Application of the full basis set Ω when calculating E_A results in the wave function of A containing not only its own atomic orbitals, but also the atomic orbitals of the ("absent") partner B, the "*ghost orbitals*" (Fig. 5.3b). As a by-product, the charge density of A would exhibit broken symmetry with respect to the symmetry of A itself (if any), e.g., the helium atom would have a small dipole moment, etc.

5.7.3 Good and bad news about the supermolecular method

Two deficiencies

When performing the subtraction in Eq. (5.1), we obtain a number representing the interaction energy at a certain distance and orientation of the two subsystems.

> The resulting E_{int} has two disadvantages: it is less precise than E_A and E_B, and it does not tell us anything about why the particular value is obtained.

The first disadvantage could be compared (following Coulson[8]) to weighing the captain's hat by first weighing the ship with the captain wearing his hat and then the ship with the captain without his hat (Fig. 5.4).

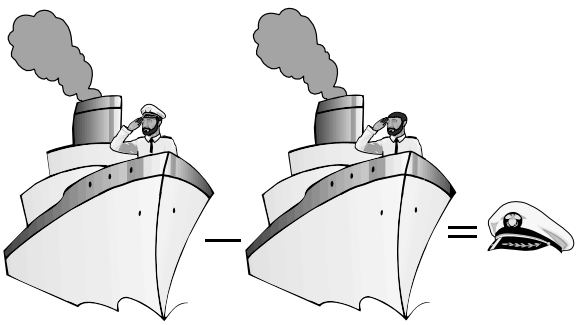


Fig. 5.4. In the supermolecular method we subtract two large numbers that differ only slightly and lose accuracy in this way. It resembles determining the weight of the captain's hat by weighing first the ship with the captain wearing his hat, repeating the procedure with the captain without his hat, and subtracting the two results. In order not to obtain a result like 240 kg or so, we have to have at our disposal a very accurate method of weighing things.

[7] S.F. Boys, F. Bernardi, *Mol. Phys.*, 19(1970)553.
[8] C.A. Coulson "*Valence*," Oxford University Press, 1952.

Formally everything is perfect, but there is a cancellation of significative digits in E_{AB} and $(E_A + E_B)$, which may lead to a very poor interaction energy.

The second deficiency deals with the fact that the interaction energy obtained is just a number and we will have no idea why the number is of such magnitude.[9]

Both deficiencies will be removed in the perturbational approach to intermolecular interaction. Then the interaction energy will be calculated directly and we will be able to tell of which physical contributions it consists.

Important advantage

A big advantage of the supermolecular method is its applicability at any intermolecular distance, i.e., independently of how strong the interaction is.

5.8 Perturbational approach

5.8.1 Intermolecular distance – what does it mean?

What is the distance (in kilometers) between the German and Turkish populations, or what does the distance between two buses mean? Because of the nonzero dimensions of both objects, it is difficult to tell what the distance could be and any measure of it will be arbitrary. It is the same story with molecules. Up till now we did not need a notion for the intermolecular distance; the positions of the nuclei were sufficient. At the beginning we need only an infinite distance and therefore, in principle, any definition will be acceptable. Later, however, we will be forced to specify the intermolecular distance (cf. p. 359 and Appendix G on p. 613). The final numerical values should not depend on this choice, but intermediary results could depend on it. It will turn out that despite the existing arbitrariness, we will prefer those definitions which are based upon the center-of-charge distance or something similar.

[9] The severity of this can be diminished by analyzing the supramolecular interaction energy expression (using molecular orbitals of A and B) and identifying the physically distinguishable terms by the kind of molecular integrals of which they are composed (K. Kitaura, K. Morokuma, *Intern. J. Quantum Chem.*, 10(1976)325).

5.8.2 Polarization approximation (two molecules)

According to the Rayleigh–Schrödinger perturbation theory (Chapter V1-5) the unperturbed Hamiltonian $\hat{H}^{(0)}$ is a sum of the isolated molecules' Hamiltonians, i.e.,

$$\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B.$$

Following quantum theory tradition in the present chapter the symbol for the perturbation operator will be changed (when compared to Chapter V1-5): $\hat{H}^{(1)} \equiv V$.

The perturbation theory may be applied to any state. Hereafter however we highlight the most important case: the perturbation of the ground state (denoted by the subscript "0"). In what is called the *polarization approximation*, the zero-order wave function will be taken as a product,[10]

$$\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}, \tag{5.5}$$

where $\psi_{A,0}$ and $\psi_{B,0}$ are the exact ground-state wave functions for the isolated molecules A and B, respectively, i.e.,

$$\hat{H}_A\psi_{A,0} = E_{A,0}\psi_{A,0},$$
$$\hat{H}_B\psi_{B,0} = E_{B,0}\psi_{B,0}.$$

We will assume that, because of the large separation of the two molecules, the electrons of molecule A are *distinguishable* from the electrons of molecule B. We have to stress the classical (therefore, "illegal") flavor of this approximation. Secondly, we assume that the exact wave functions of both isolated molecules, $\psi_{A,0}$ and $\psi_{B,0}$, are at our disposal (which never happens in chemical practice).

Of course, function $\psi_0^{(0)}$ is only an approximation to the exact wave function of the total system. Intuition tells us that this approximation is probably very good, because we assume the perturbation is small and as usual the product function $\psi_0^{(0)} = \psi_{A,0}\ \psi_{B,0}$ is an exact wave function for the *noninteracting* system.

[10] In this way we eliminate a complication which sometimes may occur. The n-th state of the two noninteracting molecules comes, of course, from *some* states of the isolated molecules A and B. It may happen (most often when the two molecules are identical) that two different sets of the states give the same energy $E_n^{(0)}$; typically, this may happen upon the exchange of excitations of both molecules. Then $\psi_n^{(0)}$ has to be taken as a linear combination of these two possibilities, which leads to profound changes of the formulae with respect to the usual cases, see p. 376.

The chosen $\psi_0^{(0)}$ has a wonderful feature, namely, it represents an eigenfunction of the $\hat{H}^{(0)}$ operator, as is required by the Rayleigh–Schrödinger perturbation theory (Chapter V1-5).

The function has also an unpleasant feature: it differs from the exact wave function by symmetry. For example, it is easy to see that

the function $\psi_0^{(0)}$ *is not* antisymmetric with respect to the electron exchanges between molecules, while the exact function has to be antisymmetric with respect to any exchange of electron labels.

This deficiency exists for any intermolecular distance.[11] We will soon pay a high price for this.

First-order effect: electrostatic energy

The first-order correction (Chapter V1-5, p. V1-277)

$$E_0^{(1)} \equiv E_{\text{elst}} \equiv E_{pol}^{(1)} = \left\langle \psi_0^{(0)} | V \psi_0^{(0)} \right\rangle \tag{5.6}$$

represents what is called the electrostatic interaction energy (E_{elst}). To stress that E_{elst} is the first-order correction to the energy in the polarization approximation, the quantity will be alternatively denoted by $E_{pol}^{(1)}$. The electrostatic energy represents the Coulombic interaction of two "frozen" charge distributions corresponding to the isolated molecules A and B, because it is the mean value of the Coulombic interaction energy operator V calculated with the wave function $\psi_0^{(0)}$ being the product of the wave functions of the isolated molecules $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$.

Second-order energy: induction and dispersion energies

The second-order energy in the polarization approximation approach can be expressed in a slightly different way than it was done in Chapter V1-5.

[11] We may say that the range of the Pauli principle is infinity. If somebody paints some electrons green and others red (we do this in the perturbational method), they are in "no man's land," between the classical and quantum worlds. Since the wave function $\psi_0^{(0)}$ does not have the proper symmetry, the corresponding operator $\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B$ is just a mathematical object not quite appropriate to the total system under study.

The n-th state of the total system at long intermolecular distances corresponds to some states n_A and n_B of the individual molecules, i.e.,

$$\psi_n^{(0)} = \psi_{A,n_A}\psi_{B,n_B} \tag{5.7}$$

and[12]

$$E_n^{(0)} = E_{A,n_A} + E_{B,n_B}. \tag{5.8}$$

Using this assumption, the second-order correction to the ground-state energy (we assume $n = 0$ and $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$) can be expressed as (see Chapter V1-5, p. V1-279)

$$E_0^{(2)} = \sum_{n_A}\sum_{n_B}{}' \frac{|\langle\psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})}, \tag{5.9}$$

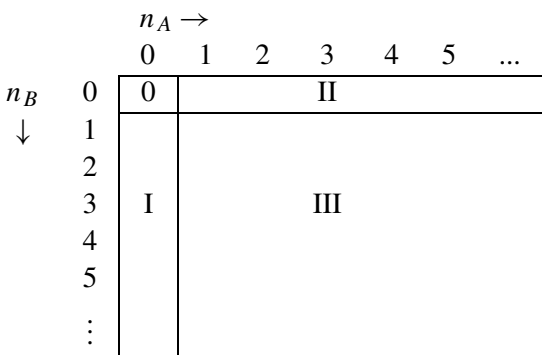
where "prime" in the summation means excluding the ground state i.e. $(n_A, n_B) = (0, 0)$. The summation in $E_0^{(2)}$ can be divided in the following way:

$$E_0^{(2)} = \sum_{n_A}\sum_{n_B}{}' \ldots = \sum_{(n_A=0,n_B\neq 0)} \ldots + \sum_{(n_A\neq 0,n_B=0)} \ldots + \sum_{(n_A\neq 0,n_B\neq 0)} \ldots . \tag{5.10}$$

Let us construct a matrix $\mathbf{A}$ (of infinite dimension) composed of the element $A_{00} = 0$ and the other elements calculated from the formula

$$A_{n_A,n_B} = \frac{|\langle\psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})} \tag{5.11}$$

and divide it into the following parts I, II and III:



[12] Also in this case we exclude the resonance interaction.

The quantity $E_0^{(2)}$ is a sum of all the elements of **A**. This summation will be carried out in three steps. First, the sum of all the elements of column 0 (part I, $n_A = 0$) represents the *induction energy* associated with forcing a change in the charge distribution of molecule B by the charge distribution of the isolated ("frozen") molecule A. Second, the sum of all the elements of row 0 (part II, $n_B = 0$) has a similar meaning, but the roles of the molecules are interchanged. Finally, the sum of all the elements of the "interior" of the matrix (part III, n_A and n_B not equal to zero) represents the *dispersion energy*. Therefore,

$$E_0^{(2)} = \underbrace{E_{\text{ind}}(A \to B)}_{\text{I}} + \underbrace{E_{\text{ind}}(B \to A)}_{\text{II}} + \underbrace{E_{\text{disp}}}_{\text{III}}, \tag{5.12}$$

where

$$\begin{aligned}
E_{\text{ind}}(A \to B) &= \sum_{n_B}{}' \frac{|\langle \psi_{A,0}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{B,0} - E_{B,n_B})}, \\
E_{\text{ind}}(B \to A) &= \sum_{n_A}{}' \frac{|\langle \psi_{A,n_A}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A})}, \\
E_{\text{ind}} &= E_{\text{ind}}(A \to B) + E_{\text{ind}}(B \to A), \\
E_{\text{disp}} &= \sum_{n_A}{}' \sum_{n_B}{}' \frac{|\langle \psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})}.
\end{aligned} \tag{5.13}$$

The electrostatic and induction interactions are visualized in Fig. 5.5.

What do these formulae tell us?

One thing has to be made clear. In Eq. (5.13) we sometimes see arguments for the interacting molecules undergoing excitations. We have to recall however that all the time we are interested in the ground state of the total system, and we calculate its energy and wave function. The excited-state wave functions appearing in the formulae are only a consequence of the fact that the first-order correction to the wave function is expanded in a complete basis set chosen deliberately as $\{\psi_n^{(0)}\}$. If we took another basis set, e.g., the wave functions of another isoelectronic molecule, we would obtain the same numerical results (although Eq. (5.13) will not hold), but the argument would be removed. From the mathematical point of view, the very essence of the perturbation theory means a small deformation of the starting $\psi_0^{(0)}$ function. This tiny deformation is the target of the expansion in the basis set $\{\psi_n^{(0)}\}$. In other words, the perturbation theory is just a cosmetic of the $\psi_0^{(0)}$: add a small hump here (Fig. 5.6), subtract a small function there,

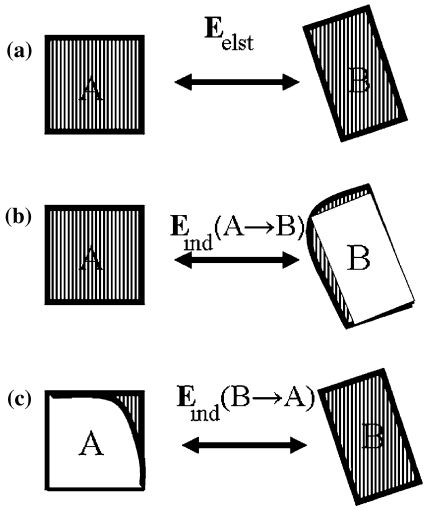


Fig. 5.5. The essence of the electrostatic and induction interactions (a schematic visualization). (a) The electrostatic energy ($E_{\rm elst} \equiv E_0^{(1)} \equiv E_{\rm pol}^{(1)}$) represents the classical Coulombic interaction of the "frozen" charge distributions of molecule A and molecule B, the same as those of the isolated molecules. (b) The induction energy consists of two contributions. The first one, $E_{\rm ind}(A \to B)$, means a modification of the electrostatic energy by allowing a polarization of molecule B by the frozen (i.e., unperturbed) molecule A symbolized as the electrostatic interaction of the shadowed charges. (c) The second contribution to the induction energy, $E_{\rm ind}(B \to A)$, corresponds to the exchange of the roles of the molecules.

etc. Therefore, the presence of the excited wave functions in the formulae is not an argument for observing some physical excitations. We may say that the mathematical procedure took what we have prepared for it, and we have prepared excited states.

This does not mean that the energy eigenvalues of the molecule have no influence on its induction or dispersion interactions with other molecules.[13] However, this is a different story. It has to do with whether the small deformation we have been discussing depends on the energy eigenvalues spectrum of the individual molecules.

The denominators in the expressions for the induction and dispersion energies suggest that the lower excitation energies of the molecules, the larger their deformation, induction, and dispersion energy.

[13] The smaller the gap between the ground and excited states of the molecule, the larger the polarizability (see Chapter 4).

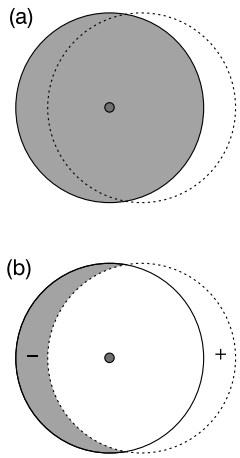


Fig. 5.6. A perturbation of the wave function is a *small* correction. (a) How a wave function (gray) spherically symmetric with respect to the nucleus (black) can be transformed into a function that is shifted off the nucleus (dotted circle)? (b) The function representing the correction (subtract on the left and add on the right) shown schematically. Please note the function has symmetry of a p orbital.

5.8.3 Intermolecular interactions: physical interpretation

Now we would like to recommend the reader to study the multipole expansion concept (Appendix G on p. 613, also Chapter 4, p. 264).

> The very essence of the multipole expansion is a replacement of the trouble-making Coulombic interaction of two point-like particles (one from molecule A, the other from molecule B) by an infinite sum of easily calculable interactions of what are called multipoles. Each interaction term has in the denominator (instead of r_{12}) an integer power of the intermolecular distance (R) between the origins of the two coordinate systems localized in the individual molecules.

In other words, multipole expansion describes the intermolecular interaction of two nonspherically symmetric, distant objects by the "interaction" of deviations (multipoles) from spherical symmetry.

To prepare ourselves for the application of the multipole expansion, let us introduce two Cartesian coordinate systems with x and y axes in one system parallel to the corresponding axes in the other system, and with the z axes colinear (see Fig. G.1 on p. 614). One of the systems is

connected to molecule A, the other one to molecule B, and the distance between the origins is R ("intermolecular distance").[14]

The operator V of the interaction energy of two molecules may be written as

$$V = -\sum_j \sum_a \frac{Z_a}{r_{aj}} - \sum_i \sum_b \frac{Z_b}{r_{bi}} + \sum_{ij} \frac{1}{r_{ij}} + \sum_a \sum_b \frac{Z_a Z_b}{R_{ab}}, \tag{5.14}$$

where we have used the convention that the summations over i and a correspond to all electrons and nuclei of molecule A, respectively, and those over j and b to molecule B. Since the molecules are assumed to be distant, we have a *practical* guarantee that the interacting particles are distant too. In V many terms with inverse interparticle distance are present. For any such term we may write the corresponding multipole expansion (Appendix G, p. 615, $s = \min(k, l)$),

$$\begin{aligned}
-\frac{Z_a}{r_{aj}} &= \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(a)^* \hat{M}_B^{(l,m)}(j), \\
-\frac{Z_b}{r_{bi}} &= \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(i)^* \hat{M}_B^{(l,m)}(b), \\
\frac{1}{r_{ij}} &= \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(i)^* \hat{M}_B^{(l,m)}(j), \\
\frac{Z_a Z_b}{R_{ab}} &= \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(a)^* \hat{M}_B^{(l,m)}(b),
\end{aligned}$$

where

$$A_{kl|m|} = (-1)^{l+m} \frac{(k+l)!}{(k+|m|)!(l+|m|)!}, \tag{5.15}$$

[14] A sufficient condition for the multipole expansion convergence is such a separation of the charge distributions of both molecules that they could be enclosed in two nonpenetrating spheres located at the origins of the two coordinate systems. This condition cannot be fulfilled with molecules, because their electronic charge density distribution extends to infinity. The consequences of this are described in Appendix G. However, the better the sphere condition is fulfilled (by a proper choice of the origins), the more effective in describing the interaction energy are the first terms of the multipole expansion.
The very fact that we use closed sets (like the spheres) in the theory witnesses that in the polarization approximation we are in "no man's land," between the quantum and classical worlds.

and the multipole moment $M_C^{(k,m)}(n)$ pertains to particle n and is calculated in "its" coordinate system $C \in \{A, B\}$. For example,

$$\hat{M}_A^{(k,m)}(a) = Z_a R_a^k P_k^{|m|}(\cos\theta_a)\exp(im\phi_a), \tag{5.16}$$

where R_a, θ_a, ϕ_a are the polar coordinates of nucleus a (with charge Z_a) of molecule A taken in the coordinate system of molecule A. When all such expansions are inserted into the formula for V, we may perform the following chain of transformations:

$$\begin{aligned}
V = & -\sum_j \sum_a \frac{Z_a}{r_{aj}} - \sum_i \sum_b \frac{Z_b}{r_{bi}} + \sum_{ij} \frac{1}{r_{ij}} + \sum_a \sum_b \frac{Z_a Z_b}{R_{ab}} \\
\cong & \sum_j \sum_a \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(a)^* \hat{M}_B^{(l,m)}(j) \\
& + \sum_i \sum_b \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(i)^* \hat{M}_B^{(l,m)}(b) \\
& + \sum_{ij} \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(i)^* \hat{M}_B^{(l,m)}(j) \\
& + \sum_a \sum_b \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(a)^* \hat{M}_B^{(l,m)}(b) \\
= & \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \{[\sum_a \hat{M}_A^{(k,m)}(a)]^* [\sum_j \hat{M}_B^{(l,m)}(j)] \\
& + [\sum_i \hat{M}_A^{(k,m)}(i)]^* [\sum_b \hat{M}_B^{(l,m)}(b)] + [\sum_i \hat{M}_A^{(k,m)}(i)]^* [\sum_j \hat{M}_B^{(l,m)}(j)] \\
& + [\sum_a \hat{M}_A^{(k,m)}(a)]^* [\sum_b \hat{M}_B^{(l,m)}(b)]\} \\
= & \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} [\sum_a \hat{M}_A^{(k,m)}(a) \\
& + \sum_i \hat{M}_A^{(k,m)}(i)]^* [\sum_b \hat{M}_B^{(l,m)}(b) + \sum_j \hat{M}_B^{(l,m)}(j)] \\
= & \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)*} \hat{M}_B^{(l,m)}.
\end{aligned} \tag{5.17}$$

In the square brackets we can recognize the multipole moment operators for the entire molecules calculated in "their" coordinate systems,

$$\hat{M}_A^{(k,m)} = \sum_a \hat{M}_A^{(k,m)}(a) + \sum_i \hat{M}_A^{(k,m)}(i),$$
$$\hat{M}_B^{(l,m)} = \sum_b \hat{M}_B^{(l,m)}(b) + \sum_j \hat{M}_B^{(l,m)}(j).$$

This has the form of a single multipole expansion, but this time the multipole moment operators correspond to *entire molecules*.

Using the table of multipoles (p. 618), we may easily write down the multipole operators for the individual molecules. The lowest moment is the net charge (monopole) of the molecules,

$$\hat{M}_A^{(0,0)} = q_A = (Z_A - N_A),$$
$$\hat{M}_B^{(0,0)} = q_B = (Z_B - N_B),$$

where Z_A is the sum of all the nuclear charges of molecule A and N_A is its number of electrons (similarly for B). The next moment is $\hat{M}_A^{(1,0)}$, which is a component of the dipole operator equal to

$$\hat{M}_A^{(1,0)} = -\sum_i z_i + \sum_a Z_a z_a, \tag{5.18}$$

where the small letters z denote the z coordinates of the corresponding particles measured in coordinate system A (the capital Z denotes the nuclear charge). Similarly, we could very easily write other multipole moments and the operator V takes the form (see, Appendix G)

$$V = \frac{q_A q_B}{R} - R^{-2}(q_A \hat{\mu}_{Bz} - q_B \hat{\mu}_{Az}) + R^{-3}(\hat{\mu}_{Ax}\hat{\mu}_{Bx} + \hat{\mu}_{Ay}\hat{\mu}_{By} - 2\hat{\mu}_{Az}\hat{\mu}_{Bz})$$
$$+R^{-3}(q_A \hat{Q}_{B,z^2} + q_B \hat{Q}_{A,z^2}) + \ldots,$$

where the monopole q_A is the net charge of molecule A,

$$\hat{\mu}_{Ax} = -\sum_i x_i + \sum_a Z_a x_a \ ,$$
$$\hat{Q}_{A,z^2} = -\sum_i \frac{1}{2}(3z_i^2 - r_i^2) + \sum_a Z_a \frac{1}{2}(3z_a^2 - R_a^2),$$

and symbol A means that all these moments are measured in coordinate system A. The other quantities have similar definitions and are easy to derive.[15]

5.8.4 Electrostatic energy in the multipole representation plus the penetration energy

Electrostatic energy (p. 353) represents the first-order correction in polarization perturbational theory and is the mean value of V with the product wave function $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$. Since now we have the multipole representation of V, we may insert it into Eq. (5.6).

Let us stress, for the sake of clarity, that V is an *operator* that contains the *operators* of the molecular multipole moments and that the integration is, as usual, carried out over the x, y, z, σ coordinates of all electrons (the nuclei have positions fixed in space according to the Born–Oppenheimer approximation), i.e., over the coordinates of electrons 1, 2, 3, etc. In the polarization approximation we know which electrons belong to molecule A ("we have painted them green") and which belong to molecule B ("red"). We have a comfortable situation, because every term in V represents a *product* of an operator depending on the coordinates of the electrons belonging to A and of an operator depending on the coordinates of the electrons of molecule B. This (together with the fact that in the integral we have a *product* of $|\psi_{A,0}|^2$ and $|\psi_{B,0}|^2$) results in a product of two integrals: one over the electronic coordinates of A and the other over the electronic coordinates of B. This is the reason why we like multipoles so much.

Therefore,

> the expression for $E_0^{(1)} = E_{elst}$ *formally* has to be of exactly the same form as the multipole representation of V, the only difference being that in V we had the molecular multipole *operators*, whereas in E_{elst} we have the molecular multipoles themselves as the *mean values* of the corresponding molecular multipole operators in the ground state (the index "0" has been omitted on the right-hand side).

However, the operator V from Eq. (5.14) and the operator in the multipole form (5.17) are equivalent only when the multipole series converges. It does so when the interacting objects are nonoverlapping, which is not the case. The electronic charge distributions penetrate and this causes a small difference (*penetration energy* E_{penetr}) between the E_{elst} calculated with and without the multipole expansion. The penetration energy vanishes very fast with intermolecular distance R (cf. Appendix V1-S, p. V1-735). We have

[15] There is one thing that may bother us, namely, that $\hat{\mu}_{Bz}$ and $\hat{\mu}_{Az}$ appear in the charge–dipole interaction terms with opposite signs, so they are not on equal footing. The reason is that the two coordinate systems are also not on equal footing, because the z coordinate of coordinate system A points to B, whereas the opposite is not true (see Appendix G).

$$E_{\text{elst}} = E_{multipol} + E_{penetr}, \tag{5.19}$$

where $E_{multipol}$ contains all the terms of the multipole expansion

$$E_{multipol} = \frac{q_A q_B}{R} - R^{-2}(q_A \mu_{Bz} - q_B \mu_{Az}) + R^{-3}(\mu_{Ax}\mu_{Bx} + \mu_{Ay}\mu_{By} - 2\mu_{Az}\mu_{Bz}) + R^{-3}(q_A Q_{B,z^2} + q_B Q_{A,z^2}) + \ldots$$

The molecular multipoles are

$$q_A = \langle \psi_{A,0}| - \sum_i 1 + \sum_a Z_a |\psi_{A,0}\rangle = (-\sum_i 1 + \sum_a Z_a)\langle \psi_{A,0}|\psi_{A,0}\rangle = \sum_a Z_a - N_A$$

$=$ the same as operator q_A,

$$\mu_{Ax} = \langle \psi_{A,0}|\hat{\mu}_{Ax}\psi_{A,0}\rangle = \langle \psi_{A,0}| - \sum_i x_i + \sum_a Z_a x_a |\psi_{A,0}\rangle \tag{5.20}$$

and similarly for the other multipoles.

Since the multipoles in the formula for $E_{multipol}$ pertain to the isolated molecules, we may say that the electrostatic interaction represents the interaction of the permanent multipoles of both molecules.

Dipole–dipole

The above multipole expansion also represents a useful source for the expressions for particular multipole–multipole interactions.

Let us take as an example the important case of the dipole–dipole interaction.

From the above formulae the dipole–dipole interaction $E_{dip-dip} = \frac{1}{R^3}(\mu_{Ax}\mu_{Bx} + \mu_{Ay}\mu_{By} - 2\mu_{Az}\mu_{Bz})$ reads also as

$$E_{dip-dip} = \frac{1}{R^3}\left(\boldsymbol{\mu}_{A\perp} \cdot \boldsymbol{\mu}_{B\perp} - 2\boldsymbol{\mu}_{A\parallel} \cdot \boldsymbol{\mu}_{B\parallel}\right), \tag{5.21}$$

where $\perp$ and $\parallel$ mean that we have to do with the vector components perpendicular to the axis z connecting the two point dipoles and parallel to this axis, respectively.

This is a short and easy to memorize formula, and we might be completely satisfied in using it *provided we always remember the particular coordinate system used for its derivation.*

Taking into account our coordinate system, the vector pointing to the coordinate system origin a from b is $\mathbf{R} = (0, 0, R)$. Then we can express $E_{dip-dip}$ in a very useful form *independent of any choice of coordinate system* (cf., e.g., pp. V1-171, 302), i.e.,

DIPOLE–DIPOLE INTERACTION

$$E_{dip-dip} = \frac{\boldsymbol{\mu}_A \cdot \boldsymbol{\mu}_B}{R^3} - 3\frac{(\boldsymbol{\mu}_A \cdot \mathbf{R})\,(\boldsymbol{\mu}_B \cdot \mathbf{R})}{R^5}. \tag{5.22}$$

This form of the dipole–dipole interaction has been used in Chapters V1-3 and 4.

It is important to understand the energetics of the ubiquitous dipole–dipole interaction. It is sufficient to consider the two dipoles having orientation within a plane formed by $\boldsymbol{\mu}_A$ and the vector $\mathbf{R}$, connecting both point dipoles. If the dipole moments and the vector $\mathbf{R}$ do not form the common plane, we represent one of the dipole moments (say, $\boldsymbol{\mu}_B$) as a sum of two components: one forming the common plane with $\boldsymbol{\mu}_A$ and $\mathbf{R}$ and the other orthogonal to this plane, i.e., $\boldsymbol{\mu}_B = \boldsymbol{\mu}_{B\parallel} + \boldsymbol{\mu}_{B\perp}$. The dipole–dipole interaction is a sum of the dipole–dipole interactions with these two components, but the interaction of $\boldsymbol{\mu}_A$ with $\boldsymbol{\mu}_{B\perp}$ is zero. Indeed, we have $E_{dip-dip} = \frac{\boldsymbol{\mu}_A\cdot(\boldsymbol{\mu}_{B\parallel}+\boldsymbol{\mu}_{B\perp})}{R^3} - 3\frac{(\boldsymbol{\mu}_A\cdot\mathbf{R})(\boldsymbol{\mu}_{B\parallel}+\boldsymbol{\mu}_{B\perp})\cdot\mathbf{R}}{R^5} = \frac{\boldsymbol{\mu}_A\cdot\boldsymbol{\mu}_{B\perp}}{R^3} + \frac{\boldsymbol{\mu}_A\cdot\boldsymbol{\mu}_{B\parallel}}{R^3} - 3\frac{(\boldsymbol{\mu}_A\cdot\mathbf{R})(\boldsymbol{\mu}_{B\perp}\cdot\mathbf{R})}{R^5} - 3\frac{(\boldsymbol{\mu}_A\cdot\mathbf{R})(\boldsymbol{\mu}_{B\parallel}\cdot\mathbf{R})}{R^5} = 0 + \frac{\boldsymbol{\mu}_A\cdot\boldsymbol{\mu}_{B\parallel}}{R^3} + 0 - 3\frac{(\boldsymbol{\mu}_A\cdot\mathbf{R})(\boldsymbol{\mu}_{B\parallel}\cdot\mathbf{R})}{R^5} = \frac{\boldsymbol{\mu}_A\cdot\boldsymbol{\mu}_{B\parallel}}{R^3} - 3\frac{(\boldsymbol{\mu}_A\cdot\mathbf{R})(\boldsymbol{\mu}_{B\parallel}\cdot\mathbf{R})}{R^5}$.

Therefore,

as to the dipole–dipole interaction whatever is nonzero pertains to the situation where the two dipole moments and $\mathbf{R}$ form the common plane.

Hence, the orientation of these dipole moments with respect to an axis (chosen as $\mathbf{R}$) can be characterized by the angles θ_A and θ_B, which the vectors form with the axis. The dipole–dipole formula (5.21) gives in this case

$$E_{dip-dip} = \frac{\mu_A \mu_B}{R^3}\left(\sin\theta_A \sin\theta_B - 2\cos\theta_A \cos\theta_B\right). \tag{5.23}$$

Fig. 5.7 shows function $E_{dip-dip}\,(\theta_A, \theta_B)$. As one can see, the strongest dipole–dipole attraction corresponds to the two dipoles aligned as $\rightarrow\rightarrow$ ($\theta_A = \theta_B = 0$) or as $\leftarrow\leftarrow$ ($\theta_A =$

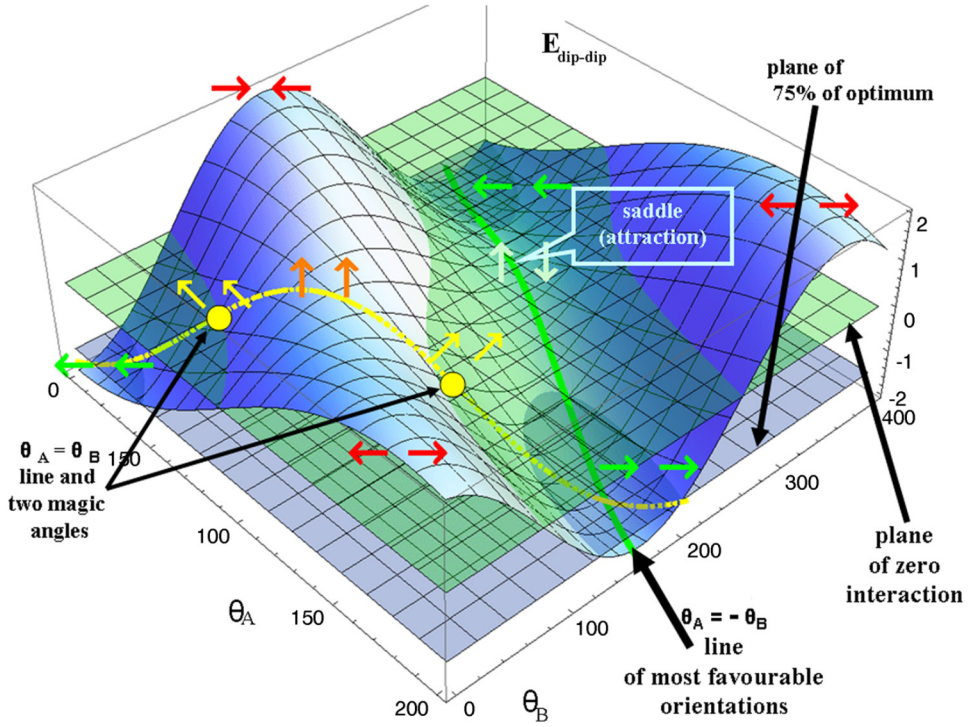


Fig. 5.7. The dipole–dipole electrostatic interaction as a function of their orientation within a plane (for convenience in $\frac{\mu_A \mu_B}{R^3}$ energy units, R stands for the fixed distance between these two point-like dipoles). The angles θ_A and θ_B measure the angular deviation of the dipoles from the connecting axis. The upper plane corresponds to the interaction energy equal to 0. The lowest energy is -2 and corresponds to two colinear parallel orientations: $\longrightarrow \quad \longrightarrow$ and $\longleftarrow \quad \longleftarrow$ (shown by arrows). When moving along the energetic valley (the central part of the figure, a line from bottom up) one can pass from one to the other of these configurations (via the configuration $\uparrow\downarrow$ of energy -1), all the time having the dipole–dipole attraction! Therefore, if some other interactions (like, e.g., a steric hindrance) tried to destabilize the optimum dipole–dipole colinear configuration, an interesting low-energy compromise would be possible – just rotate the two dipoles starting from the configuration $\longrightarrow \quad \longrightarrow$ towards the antiparallel configuration $\uparrow\downarrow$ (i.e., putting $\theta_A = -\theta_B \neq 0$ instead of $\theta_A = \theta_B = 0$ and arriving at configuration $\nearrow \quad \searrow$). In the two points labeled by small circles (the line shown corresponds to $\theta_A = \theta_B$) both dipoles form with the axis what is known as the magic angle θ_{mag} (the corresponding configurations are $\nwarrow\nwarrow$ and $\nearrow\nearrow$), and their interaction vanishes. The energy maxima (equal to $+2$) correspond to the configurations $\longrightarrow \quad \longleftarrow$ and $\longleftarrow \quad \longrightarrow$.

$\theta_B = 180°$). The worst energetically are the orientations $\rightarrow\leftarrow$ ($\theta_A = 0, \theta_B = 180°$) and $\leftarrow\rightarrow$ $\theta_A = 180°, \theta_B = 0$). The lateral interaction corresponding to the antiparallel configurations $\uparrow\downarrow$ ($\theta_A = -\theta_B = 90°$) or $\downarrow\uparrow$ ($\theta_A = -\theta_B = -90°$) also corresponds to attraction, but it does not

represent any minimum; these are two saddle points. The lateral parallel configurations ↑↑ and ↓↓ mean repulsion of the absolute value opposite to the ↑↓ attraction.

There are many configurations giving zero dipole–dipole interaction (shown in Fig. 5.7 as intersection of the $E_{dip-dip}(\theta_A, \theta_B)$ function with the upper plane). However, if one forces the noninteracting dipoles to be parallel ($\theta_A = \theta_B = \theta$), the zero energy will correspond only to the particular θ angle known as magic angle[16] θ_{mag} corresponding to the ↗↗ (or ↖↖) configuration.

In the central part of Fig. 5.7 one sees a deep valley connecting the most important configurations of two dipoles – two global minima with energy equal to −2 (for the configurations →→ and ←←) separated by the saddle point corresponding to ↑↓ or ↓↑.

If some other (than dipole–dipole) interactions tried to distort the parallel colinear configuration of the interacting dipoles (optimal one), it could be done without increasing too much the energy provided the distortion would correspond to the opposite angular deviations of both dipoles with respect to the **R** axis (i.e., the dipoles would be tilted in opposite directions, the dipole orientations moved towards the saddle point configuration ↑↓, Fig. 5.7). In such a way one gets an easy-to-achieve compromise – common in experimental dipole–dipole configurations: ↗ ↘ (see Example 1 on p. 366).

Is the electrostatic interaction important?

Electrostatic interaction can be attractive or repulsive. For example, in the electrostatic interaction between Na^+ and Cl^- the main role will be played by the charge–charge interaction, which is negative and therefore represents attraction, while for $Na^+ \cdots Na^+$ the electrostatic energy will be positive (repulsion). For neutral molecules the electrostatic interaction may depend on their *orientation* to such an extent that the sign may change. This is an exceptional feature of great importance peculiar only to electrostatic interaction.

When the distance R is small if compared to size of the interacting subsystems, multipole expansion gives bad results. To overcome this the total charge distribution may be divided into *atomic* segments (Appendix D). Each atom would carry its charge and other multipoles, and the electrostatic energy would be the sum of the atom–atom contributions, any of which would represent a series similar[17] to $E_0^{(1)}$.

[16] This angle plays an important role in the solid state nuclear magnetic resonance measurements. We have $E_{dip-dip} = \frac{\mu_A \mu_B}{R^3}\left(\sin^2\theta_{\text{mag}} - 2\cos^2\theta_{\text{mag}}\right) = \frac{\mu_A \mu_B}{R^3}\left(1 - 3\cos^2\theta_{\text{mag}}\right) = 0$, and from this we get $\theta_{\text{mag}} = \arccos\frac{1}{\sqrt{3}} = 54.74°$.

[17] A.J. Stone, *Chem. Phys. Lett.*, 83(1981)233; A.J. Stone, M. Alderton, *Mol. Phys.*, 56(1985)1047; W.A. Sokalski, R. Poirier, *Chem. Phys. Lett.* 98, (1983)86; W.A. Sokalski, A. Sawaryn, *J. Chem. Phys.*, 87(1987)526.

Example 1. Let us calculate (by using the Hartree–Fock method) the rigid interaction energy of two hydrogen fluoride molecules at the fixed F···F distance equal to 5 Å (each molecule has the optimum length equal to 0.911 Å). We obtain the following results[18] (in the last column the dipole–dipole interaction energy computed from Eq. (5.23) is displayed, all energies in a.u.):

Configuration	E_{int}	$E_{dip-dip}$
$\rightarrow\rightarrow$ colinear parallel	−0.00143	−0.00143
$\rightarrow\leftarrow$ colinear antiparallel	+0.00320	+0.00143
$\uparrow\uparrow$ parallel dipoles	+0.00071	+0.00071
$\uparrow\downarrow$ antiparallel dipoles	−0.00071	−0.00071

It is seen that, except the colinear antiparallel case, the electrostatic dipole–dipole interaction dominates the other interactions contributing to the Hartree–Fock result. A lot of molecular integrals, so many terms to calculate, the SCF iterative method included, but the situation is correctly described by a primitive dipole–dipole term. The colinear antiparallel configuration turned out to be a too difficult case here; the discrepancy is large. This is because even if the intermolecular distance F···F is kept constant, in reality the molecules are too close in this particular case, e.g., the H···H distance is about 3 Å only. In such a situation not only the interaction of the point dipoles, but also other interactions, like quadrupole–dipole, quadrupole–quadrupole, other electrostatic contributions, and nonelectrostatic interactions (e.g., valence repulsion, induction), enter the game, so one may expect trouble.

Example 2 (Dipole–internal field interactions in proteins). In the above example any of the dipoles corresponded to the entire molecule. One may consider, however, the molecular dipole moment as a sum of the dipole moments distributed on all atoms and bonds of the molecule (see Eq. (4.37) on p. 279) and then sum up the *intramolecular* electrostatic dipole–dipole interactions.

Here we give an example of this idea. The native conformation is important, because many proteins exert their biological activity only in such conformation. It is currently assumed that the native conformation corresponds to the lowest-energy conformational state, i.e., to the global minimum of energy (cf. p. V1-413). A hypothesis has been presented[19] that at the energy minimum, each peptide bond (–HN–CO–) of the molecule is oriented quite well along the electric field created by the rest of the molecule.[20] The hypothesis has been tested considering a polyalanine oligomer in the α-helical conformation. It turned out that indeed, except of the very ends of the α-helix, the angular deviation from the alignment was of the order of a few degrees only.

[18] We take the standard basis set 6-31G(d), for which the H–F bond length of the individual molecule is optimal. The results are not corrected for the BSSE.

[19] L. Piela, H.A. Scheraga, *Biopolymers*, 26(1987)*S*33.

[20] Each atom of the molecule had an electric charge assigned according to a force field used.

Moreover, for other conformations considered (breaking the α-helix), the deviations from the dipole–field colinearity had the local character, thus being able to serve for computing a diagnosis how to change the conformation in order to lower the electrostatic energy. This hypothesis has then been confirmed for a large class of other proteins in their native conformations, not only in their α-helical and β-type motifs, but also in other structural elements of proteins.[21]

Reality or fantasy?

In principle, this part (about electrostatic interactions) may be considered as completed. I am tempted, however, to include some "obvious" subjects, which will turn out to lead us far away from the usual track of intermolecular interactions.

Let us consider the Coulomb interaction of two point charges q_1 on molecule A and q_2 on molecule B, both charges separated by distance r,

$$E_{\text{elst}} = \frac{q_1 q_2}{r}. \tag{5.24}$$

This is an outstanding formula.

- first of all we have the amazing power of r with the *exact* value -1;
- second, change of the charge *sign* does not make any profound changes in the formula, except the change of *sign* of the interaction energy;
- third, the formula is bound to be false (it has to be only an approximation), since instantaneous interaction is assumed, whereas the interaction has to have time to travel between the interacting objects and during that time the objects change their distance (see Chapter V1-3, p. V1-171).

From these remarks some apparently obvious observations follow, i.e., that E_{elst} is invariant with respect to the following operations:

- $q_1' = -q_1, \quad q_2' = -q_2$ (charge conjugation, Chapter V1-2, p. V1-85),
- $q_1' = q_2, \quad q_2' = q_1$ (*exchange* of charge positions),
- $q_1' = -q_2, \quad q_2' = -q_1$ (charge conjugation *and* exchange of charge positions).

These invariance relations, when treated literally and rigorously, are not of particular usefulness in theoretical chemistry. They may, however, open new possibilities when considered as some limiting cases. Chemical reaction mechanisms very often involve the interaction of molecular ions. Suppose we have a particular reaction mechanism. Now, let us make the charge conjugation of all the objects involved in the reaction (this would require the change of matter to antimatter). This will preserve the reaction mechanism. We cannot make such changes in chemistry. However, we may think of some *other molecular systems*, which have similar geometry

[21] D. Ripoll, J.A. Vila, H.A. Scheraga, *Proc. Natl. Acad. Sciences (USA)*, 102(2005)7559.

but opposite overall charge pattern ("counterpattern"). The new reaction has a chance to run in a similar direction as before. This concept is parallel to the idea of *Umpolung* functioning in organic chemistry. It seems that nobody has looked from that point of view at all known reaction mechanisms.

5.8.5 Induction energy in the multipole representation

The induction energy contribution consists of two parts: $E_{\mathrm{ind}}(A \to B)$ and $E_{\mathrm{ind}}(B \to A)$, or the polarization energy of molecule B in the electric field of the unperturbed molecule A and *vice versa*, respectively.

The goal of the present section is to take apart the induction mechanism by showing its multipole components. If we insert the multipole representation of V into the induction energy $E_{\mathrm{ind}}(A \to B)$, then

$$
\begin{aligned}
E_{\mathrm{ind}}(A \to B) = \sum_{n_B}{}' \frac{|\langle \psi_{A,0}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{E_{B,0}-E_{B,n_B}} &= \sum_{n_B}{}' \frac{1}{E_{B,0}-E_{B,n_B}} \\
&\{|R^{-1}q_A \cdot 0 - R^{-2}q_A\langle\psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle + R^{-2}\cdot 0 \\
&+ R^{-3}[\mu_{Ax}\langle\psi_{B,n_B}|\hat{\mu}_{Bx}\psi_{B,0}\rangle + \mu_{Ay}\langle\psi_{B,n_B}|\hat{\mu}_{By}\psi_{B,0}\rangle \\
&- 2\mu_{Az}\langle\psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle] + \ldots|\}^2 \\
= \sum_{n_B}{}' \frac{1}{E_{B,0}-E_{B,n_B}} &\{|-R^{-2}q_A\langle\psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle + \\
&R^{-3}[\mu_{Ax}\langle\psi_{B,n_B}|\hat{\mu}_{Bx}\psi_{B,0}\rangle + \mu_{Ay}\langle\psi_{B,n_B}|\hat{\mu}_{By}\psi_{B,0}\rangle \\
&- 2\mu_{Az}\langle\psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle] + \ldots|\}^2 \\
= -\frac{1}{2}\frac{1}{R^4}q_A^2\alpha_{B,zz} &+ \ldots,
\end{aligned}
$$

where

- the zeros appearing in the first part of the derivation come from the orthogonality of the eigenstates of the isolated molecule B,
- symbol "+..." stands for higher powers of R^{-1},
- $\alpha_{B,zz}$ represents the zz component of the dipole polarizability tensor of molecule B, which absorbed the summation over the excited states of B according to definition (4.42).

A molecule in the electric field of another molecule

Note that $\frac{1}{R^4}q_A^2$ represents the square of the electric field intensity $\mathcal{E}_z(A \to B) = \frac{q_A}{R^2}$ measured on molecule B and created by the net charge of molecule A. Therefore, we have $E_{\text{ind}}(A \to B) = -\frac{1}{2}\alpha_{A,zz}\ \mathcal{E}_z^2(A \to B) + \ldots$ according to formula (4.24) describing the molecule in an electric field. For molecule B its partner molecule A represents an external world creating the electric field, and molecule B has to behave as described in Chapter 4 (and *vice versa*). The net charge of A created the electric field $\mathcal{E}_z(A \to B)$ on molecule B, which as a consequence induced on B a dipole moment $\mu_{B,\text{ind}} = \alpha_{B,zz}\ \mathcal{E}_z(A \to B)$ according to formula (4.19). This is associated with the interaction energy term $\frac{1}{2}\alpha_{B,zz}\ \mathcal{E}_z^2(A \to B)$ (see Eq. (4.24)).

There is however a small problem. Why is the induced moment proportional only to the net charge of molecule A? This would be absurd. Molecule B does not know anything about multipoles of molecule A; it only knows about the *local* electric field that acts on it and has to react to that field by a suitable polarization. Everything is all right, though. The rest of the problem is in the formula for $E_{\text{ind}}(A \to B)$. So far we have analyzed the electric field on B coming from the net charge of A, but the other terms of the formula will give contributions to the electric field coming from *all other* multipole moments of A. Then, the response of B will pertain to the total electric field created by "frozen" A on B, as should be. A similar story can be given for $E_{int}(B \to A)$.

5.8.6 Dispersion energy in the multipole representation

After inserting V in the multipole representation into the expression for the dispersion energy we obtain

$$\begin{aligned} E_{\text{disp}} = &\sum_{n_A}{}' \sum_{n_B}{}' \frac{1}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})} |R^{-1} q_A q_B \cdot 0 \cdot 0 - R^{-2} q_A \cdot 0 \cdot (\mu_{Bz})_{n_B,0} \\ &- R^{-2} q_B \cdot 0 \cdot (\mu_{Az})_{n_A,0} + R^{-3}[(\mu_{Ax})_{n_A,0}\,(\mu_{Bx})_{n_B,0} + \left(\mu_{Ay}\right)_{n_A,0}\left(\mu_{By}\right)_{n_B,0} \\ &- 2\,(\mu_{Az})_{n_A,0}\,(\mu_{Bz})_{n_B,0}] + \ldots|^2 \\ = &\sum_{n_A}{}' \sum_{n_B}{}' \frac{|R^{-3}[(\mu_{Ax})_{n_A,0}\,(\mu_{Bx})_{n_B,0} + \left(\mu_{Ay}\right)_{n_A,0}\left(\mu_{By}\right)_{n_B,0} - 2\,(\mu_{Az})_{n_A,0}\,(\mu_{Bz})_{n_B,0}] + \ldots|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})}, \end{aligned}$$

where $(\mu_{Ax})_{n_A,0} = \langle \psi_{A,n_A} | \hat{\mu}_{Ax} \psi_{A,0} \rangle$, $(\mu_{Bx})_{n_B,0} = \langle \psi_{B,n_B} | \hat{\mu}_{Bx} \psi_{B,0} \rangle$ and similarly the other quantities. The zeros in the first part of the equality chain come from the orthogonality of the eigenstates of each of the molecules.

The square in the formula pertains to all terms. The other terms, not shown in the formula, have powers of R^{-1} higher than R^{-3}.

Hence, if we squared the total expression, the most important term would be the dipole–dipole contribution with the asymptotic R^{-6} distance dependence.

As we can see (Eq. (5.13)), its calculation requires *double* electronic excitations (one on the first, the other on the second interacting molecules), and these already belong to the correlation effect (cf. p. 174).

The dispersion interaction is a pure correlation effect and therefore the methods used in a supermolecular approach, which do not take into account the electronic correlation (as for example the Hartree–Fock method), are unable to produce any nonzero dispersion contribution.

Dispersion energy model – calculation on fingers

We will try to catch the very essence of this intermolecular interaction. As always we will try to be as simple as possible in order to construct the simplest model, which still contains the very reason for dispersion interaction to happen.

- Well, why to complicate things and consider molecules? Do atoms interact via dispersion interaction? Yes. Let us take atoms then.
- Is it essential to have more than two electrons? No.
- Let us take, therefore, two hydrogen atoms, each in its ground state $1s$, and being at a long internuclear distance R.
- Is it of importance for seeing the phenomenon to have the "full-size" three-dimensional hydrogen atoms? No, because the effect comes from the electron correlation and such correlation may happen even for the toy one-dimensional hydrogen atoms, with the electrons moving along the internuclear axis only.
- Is it essential indeed to have such motion? No, we may still simplify things and give only the possibility of correlating *two positions* for each of the two electrons (along the internuclear axis, so our toy will be not only one-dimensional, but also "granular"): on the left-hand side and on the right-hand side (Fig. 5.8), the fixed electron–proton distance being $a \ll R$.

Now, let us calculate the interaction energy of the two "toy hydrogen atoms" at large distances R by using the dipole–dipole interactions for all four possible situations from Eq. (5.22) assuming the local coordinate systems on the protons. In the total potential energy, there is a common contribution, identical in all four situations: the interaction within the individual atoms, $-2/a$ (the remainder is the interaction energy $E_{int}(i)$). The potential energy resulting from the Coulomb interactions of the particles is therefore

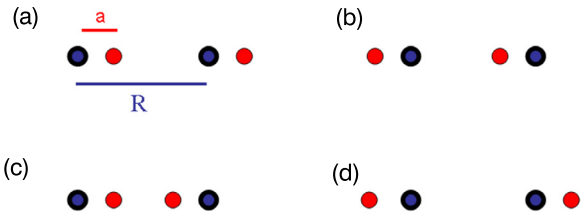


Fig. 5.8. Dispersion energy origin shown using a primitive model of two interacting hydrogen atoms. The nuclei (larger circles) occupy some fixed positions in space (at distance R), while each of the electrons (smaller circles) can occupy only two positions: on the left and on the right of the corresponding nucleus (at distance a from it). A popular explanation for the dispersion interaction is that, due to electron repulsion, the situations (a) and (b) occur more often than situation (c), and this is why the dispersion interaction represents a net attraction of dipoles. The positions of the electrons that correspond to (a) and (b) represent two favorable instantaneous dipole–instantaneous dipole interactions, while (c) corresponds to two nonfavorable instantaneous dipole–instantaneous dipole interactions. A trouble with this explanation is that there is also the possibility of having electrons far apart as in (d). This most favorable situation (the longest distance between the electrons) means, however, repulsion of the resulting dipoles. It may be shown, though, that the net result (dispersion interaction) is still an attraction (see the text) as it should be.

$$V(i) = -2/a + E_{int}(i),$$

where the first term on the right-hand side is twice (two "hydrogen atoms") the electron–nucleus interaction and the second one corresponds to four possible situations

Situation, i	Fig. 5.8	Interaction energy $E_{int}(i)$
1	a	$-2\frac{\mu^2}{R^3}$
2	b	$-2\frac{\mu^2}{R^3}$
3	c	$+2\frac{\mu^2}{R^3}$
4	d	$+2\frac{\mu^2}{R^3}$

with $\boldsymbol{\mu} = (0, 0, \pm a)$ for each of the interacting atoms according to definition (5.18), and $\mu \equiv a$ in a.u. Note that if we assume the same probability for each situation, the interatomic interaction energy per situation would be zero, i.e., $\frac{1}{4}\sum_i E_{int}(i) = 0$. These situations have, however, different probabilities (p_i), because the electrons repel each other, and the total potential energy depends on where they actually are. *Note that the probabilities should be different only because of the electron correlation.* If we could guess somehow these probabilities $p_i, i = 1, 2, 3, 4$, then we could calculate the mean interaction energy of our model one-dimensional atoms

as $\bar{E}_{int} = \sum_i p_i E_{int}(i)$. In this way we could see whether it corresponds to net attraction ($\bar{E}_{int} < 0$) or repulsion ($\bar{E}_{int} > 0$), which is most interesting for us. Well, but how to calculate them?[22]

We may suspect that for the ground state (we are interested in the ground state of our system), the lower the potential energy $V(i)$, the higher the probability density p_i. This is what happens for the harmonic oscillator, for the Morse oscillator, for the hydrogen-like atom, etc. It looks as a general rule. Is there any tip that could help us work out what such a dependence might be? If you do not know where to begin, then think of the harmonic oscillator model as a starting point! This is what people usually do as a first guess. As seen from Eq. (V1-4.22), the ground-state wave function for the harmonic oscillator may be written as $\psi_0 = A\exp[-BV(x)]$, where $B > 0$, and $V(x)$ stands for the potential energy for the harmonic oscillator. Therefore the probability density changes as $\rho = A^2\exp[-2BV(x)]$. Interesting... Let us assume that a similar thing happens[23] for the probabilities p_i of finding electrons 1 and 2 and they may be reasonably estimated as $p_i = N'A^2\exp[-2BV(i)]$, where $V(i) = -2/a + E_{int}(i)$ plays the role of potential energy. Finally, $p_i = N\exp[-2BE_{int}(i)]$, and $N = 1/\sum_i \exp[-2BE_{int}(i)]$ is the normalization constant ensuring that in our model $\sum_i p_i = 1$. For long distances R (small $E_{int}(i)$) we may expand this expression in a Taylor series and obtain $p_i = \frac{[1-2BE_{int}(i)]}{\sum_j \exp[-2BE_{int}(j)]} \approx \frac{1-2BE_{int}(i)}{\sum_j (1-2BE_{int}(j)+\ldots)} = \frac{1-2BE_{int}(i)}{4-2B\cdot 0+\sum_j \frac{1}{2}[2BE_{int}(j)]^2+\ldots} \approx \frac{1}{4} - \frac{B}{2}E_{int}(i)$, where the Taylor series has been truncated to the accuracy of the linear terms in the interaction. Then for the mean interaction energy we have

$$\bar{E}_{int} = \sum_i p_i E_{int}(i) \approx \sum_i \left[\frac{1}{4} - \frac{B}{2}E_{int}(i)\right] E_{int}(i) =$$

$$\frac{1}{4}\sum_i E_{int}(i) - \frac{B}{2}\sum_i [E_{int}(i)]^2 = 0 - \frac{B}{2}\frac{16\mu^4}{R^6} = -8B\frac{\mu^4}{R^6} < 0.$$

[22] In principle we could look at what people have calculated in the most sophisticated calculations for the hydrogen molecule at a large R, and assign the p_i's as the squares of the wave function value for the corresponding four positions of both electrons. Since these wave functions are awfully complex and do not contribute anything qualitatively different, we leave this path without regret.

[23] This is like having the electron attached to the nucleus by a harmonic spring (instead of Coulombic attraction).

The approximations we have made were extremely crude, but in spite of this we were able to grasp four important features of the correct dispersion energy:

- it comes from the Coulombic correlation of electrons,
- it corresponds to attractive interaction,
- it vanishes with distance as R^{-6},
- it is proportional to the fourth power of the instantaneous dipole moment.

What really contributed to the success of the above reasoning is the dipole–dipole character of the interaction,[24] not the fact that we have used a harmonic model.

Examples

The electrostatic interaction energy of two molecules can be calculated from Eq. (5.6). However, it is very important for a chemist to be able to predict the main features of the electrostatic interaction *without any calculation* at all, based on some general rules. This will create chemical intuition or chemical common sense so important in planning, performing, and understanding experiments.

How to recognize that a particular multipole–multipole interaction represents attraction or repulsion? First we replace the molecules by their lowest nonzero (nonpoint-like) multipoles represented by point charges, e.g., ions by $+$ or $-$, dipolar molecules by $+-$, quadrupoles by $+^{-}_{-}+$, etc. In order to do this we have to know which atoms are electronegative and which are electropositive.[25] After doing this we replace the two molecules by the multipoles. If the nearest neighbor charges in the two interacting multipoles are of opposite sign, the multipoles attract, otherwise they repel each other (Fig. 5.9).

The data of Table 5.1 were obtained assuming a long intermolecular distance and the molecular orientations as shown in the table.

In composing Table 5.1 some helpful information has been used:

- *Induction and dispersion energies always represent attraction, except in some special cases when they are zero*. These special cases are obvious, e.g., it is impossible to induce some changes on molecule B if molecule A does not have any nonzero permanent multipoles. Also, the dispersion energy is zero if an interacting subsystem has no electrons on it.

[24] And that the lower the potential energy $V(x)$, the higher the value of the ground-state wave function $\psi(x)$ and the higher the corresponding probability density.

[25] This is common knowledge in chemistry and is derived from experiments as well as from quantum mechanical calculations. The latter provide the partial atomic charges from what is called population analysis (see Appendix D). Despite its nonuniqueness it would satisfy our needs. A unique and elegant method of calculation of atomic partial charges is related to the Bader analysis described on p. 197.

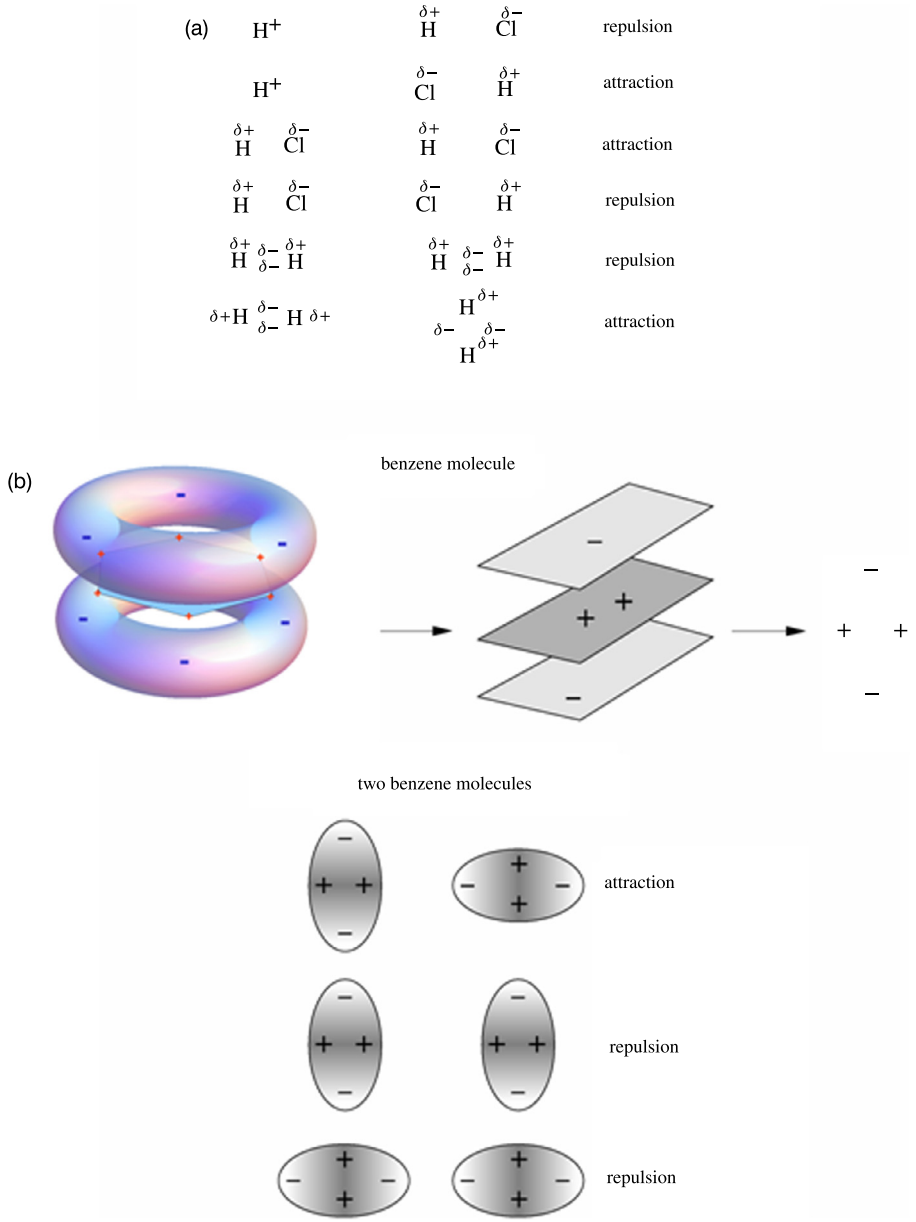
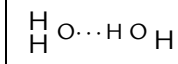
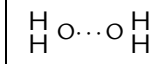


Fig. 5.9. For sufficiently large intermolecular separations the interaction of the lowest nonvanishing multipoles dominates. Whether this is an attraction or repulsion can be recognized by representing the molecular charge distributions by nonpoint-like multipoles (clusters of point charges). If such multipoles point to each other by point charges of the opposite (same) sign, then the electrostatic interaction of the molecules is attraction (repulsion). (a) A few examples of simple molecules and the atomic partial charges. (b) Even the interaction of the two benzene molecules obeys this rule: in the face-to-face configuration they repel, while in the perpendicular configuration they attract each other.

Table 5.1. The asymptotic interaction energy (proportional to R^{-n}, the table gives the exponent n) of two molecules in their electronic ground states. For each pair of molecules or atoms a short characteristic of their electrostatic, induction, and dispersion interactions is given. It consists of n and the sign (in parentheses) of the corresponding interaction type, where the "$-$" sign means attraction, the "$+$" sign means repulsion, and 0 corresponds to the absence of such an interaction.

System	Electrostatic	Induction	Dispersion
He···He	0	0	6(−)
He···H^+	0	4(−)	0
He···HCl	0	6(−)	6(−)
H^+···HCl	2(+)	4(−)	0
HCl···ClH	3(+)	6(−)	6(−)
HCl···HCl	3(−)	6(−)	6(−)
H–H···He	0	8(−)	6(−)
H–H···H–H	5(+)	8(−)	6(−)
	5(−)	8(−)	6(−)
	3(−)	6(−)	6(−)
	3(+)	6(−)	6(−)

- *Electrostatic energy is nonzero if both interacting molecules have some nonzero permanent multipoles.*
- *Electrostatic energy is negative (positive) if the lowest nonvanishing multipoles of the interacting partners attract (repel) themselves.*[26]
- The dispersion energy always decays as R^{-6}.
- The electrostatic energy vanishes as $R^{-(k+l+1)}$, where the 2^k-pole and 2^l-pole represent the lowest nonvanishing multipoles of the interacting subsystems.

[26] This statement is true for sufficiently long distances.

- The induction energy vanishes as $R^{-2(k+2)}$, where the 2^k-pole is the lower of the two lowest nonzero permanent multipoles of molecules A and B. The formula is easy to understand if we take into account that the lowest *induced* multipole is always a dipole ($l = 1$) and that the induction effect is of the second order (hence a factor 2 in the exponent).

5.8.7 Resonance interaction – excimers

We have analyzed so far interactions of the ground-state molecules that produce the ground electronic state of the total system. In excited states one may find an important modification of the interaction energy.

Suppose we have a system of noninteracting identical molecules A and B, with A in the ground and B in an excited state. An acceptable zero-order wave function might be taken as $\psi_1^{(0)} = \psi_{A,0}\psi_{B,n}$. Note, however, that transferring the excitation to the other molecule leads to another function, $\psi_2^{(0)} = \psi_{A,n}\psi_{B,0}$, and this new function corresponds to the same energy. Thus, we have a degeneracy in the perturbation theory (see vol. 1, p. V1-282). We have to solve (with $H_{ij}^{(1)} \equiv V_{ij} \equiv \left\langle \psi_i^{(0)} | V \psi_j^{(0)} \right\rangle$)

$$\det \left\{ H_{ij}^{(1)} - E^{(1)} \delta_{ij} \right\} = 0, \tag{5.25}$$

$i, j = 1, 2$, for the unknown $E^{(1)}$. We have two solutions, *gerade g* and *ungerade u* (two possible unperturbed starting points, $|V_{12}|$ is known as the *resonance energy*),

$$E_g^{(1)} = V_{11} - |V_{12}|,$$
$$E_u^{(1)} = V_{11} + |V_{12}|,$$

corresponding to two different zero-order wave functions

$$\psi_g^{(0)} = \frac{1}{\sqrt{2}}(\psi_{A,0}\psi_{B,n} + \psi_{A,n}\psi_{B,0}) \qquad \text{“gerade”},$$
$$\psi_u^{(0)} = \frac{1}{\sqrt{2}}(\psi_{A,0}\psi_{B,n} - \psi_{A,n}\psi_{B,0}) \qquad \text{“ungerade”}.$$

The first order of the perturbation theory gives therefore the additional resonance energy gain $-|V_{12}| = -|\langle \psi_{A,0}\psi_{B,n} | V \psi_{A,n}\psi_{B,0} \rangle|$, when compared to the usual electrostatic interaction V_{11}. The gain comes from the exchange of excitations (photons) between A and B, which results in the creation of the excimer complex in solutions and molecular crystals; their energy may reach the order of 10^{-3} a.u. Application of the multipole expansion to V gives for electrically neutral molecules

$$V_{12} = \frac{(\mu_{Ax})_{0,n}\,(\mu_{Bx})_{n,0} + \left(\mu_{Ay}\right)_{0,n}\left(\mu_{By}\right)_{n,0} - 2\,(\mu_{Az})_{0,n}\,(\mu_{Bz})_{n,0}}{R^3} + \ldots,$$

where $(\mu_{Ax})_{0,n} = \langle \psi_{A,0} | \hat{\mu}_{Ax} \psi_{A,n} \rangle$, $(\mu_{Bx})_{n,0} = \langle \psi_{B,n} | \hat{\mu}_{Bx} \psi_{B,0} \rangle$, etc. This expression is similar to what has been derived for the dispersion interaction, except that then it was squared (giving rise to R^{-6} dependence), whereas for the resonance interaction we have the R^{-3} decay, which is much slower. This is why the resonance interactions are important – they are very long-range.

5.9 Symmetry-adapted perturbation theory (SAPT)

The SAPT approach, as opposed to the polarization approximation valid for long distances only, is applicable not only for large but also for intermediate intermolecular separations when the electron clouds of both molecules overlap to such an extent that:

- the polarization approximation, i.e., ignoring the Pauli principle (p. 351), becomes a very poor approximation;
- the multipole expansion becomes invalid.

5.9.1 Polarization approximation is illegal

The polarization approximation zero-order wave function $\psi_{A,0}\psi_{B,0}$ will be deprived of the privilege of being the unperturbed function $\psi_0^{(0)}$ in a perturbation theory. Since it will still play an important role in the theory, let us denote it by $\varphi^{(0)} = \psi_{A,0}\psi_{B,0}$.
The polarization approximation seems to have (at first glimpse) a very strong foundation, because at long intermolecular distances R, the zero-order *energy* is close to the exact one. The trouble is, however, that a similar statement is not true for the zero-order *wave function* $\varphi^{(0)}$ and the exact wave function at *any* intermolecular distance (even at infinity).

Let us take the example of two ground-state hydrogen atoms. The polarization approximation zero-order wave function (using the normalized atomic orbitals)

$$\varphi^{(0)}(1,2) = 1s_a(1)\alpha(1)\ 1s_b(2)\beta(2), \tag{5.26}$$

where the spin functions have been introduced (the Pauli principle is ignored[27]).

This function is neither symmetric (since $\varphi^{(0)}(1,2) \neq \varphi^{(0)}(2,1)$), nor antisymmetric (since $\varphi^{(0)}(1,2) \neq -\varphi^{(0)}(2,1)$), and therefore is "illegal" and in principle not acceptable.

[27] This is the essence of the polarization approximation.

5.9.2 Constructing a symmetry-adapted function

In the Born–Oppenheimer approximation the electronic ground-state wave function of H_2 has to be the eigenfunction of the nuclear inversion symmetry operator $\hat{I}$ interchanging nuclei a and b (cf. Appendix V1-C). Since $\hat{I}^2 = 1$, the eigenvalues can be either -1 (called u symmetry) or $+1$ (g symmetry).[28] The ground state is of g symmetry; therefore, the projection operator $\frac{1}{2}(1 + \hat{I})$ will take care of that (it says: make a fifty-fifty combination of a function and its counterpart coming from the exchange of nuclei a and b).[29] On top of this, the wave function has to fulfill the Pauli exclusion principle, which we will ensure by the antisymmetrizer $\hat{A}$ (cf. p. V1-707), which, when acting, gives either an antisymmetric function or zero. Altogether the proper symmetry will be ensured by projecting $\varphi^{(0)}$ using the idempotent projection operator

$$\hat{\mathcal{A}} = \frac{1}{2}(1 + \hat{I})\hat{A}. \tag{5.27}$$

As a projection $\hat{\mathcal{A}}\varphi^{(0)}$ of $\varphi^{(0)}$ we obtain

$$\begin{aligned}
\hat{A}\frac{1}{2}(1 + \hat{I})\varphi^{(0)} &= \frac{1}{2!}\frac{1}{2}(1 + \hat{I})\sum_P (-1)^p \hat{P}[1s_a(1)\alpha(1)1s_b(2)\beta(2)] \\
&= \frac{1}{4}(1 + \hat{I})[1s_a(1)\alpha(1)1s_b(2)\beta(2) - 1s_a(2)\alpha(2)1s_b(1)\beta(1)] \\
&= \frac{1}{4}[1s_a(1)\alpha(1)1s_b(2)\beta(2) - 1s_a(2)\alpha(2)1s_b(1)\beta(1) + 1s_b(1)\alpha(1)1s_a(2)\beta(2) \\
&\quad - 1s_b(2)\alpha(2)1s_a(1)\beta(1)] \\
&= \frac{1}{2\sqrt{2}}[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)]\{\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]\}.
\end{aligned}$$

This result is proportional to the Heitler–London wave function (from p. 127, where its important role in chemistry is highlighted)

$$\psi_{HL} \equiv \psi_0^{(0)} = N[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)]\{\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]\}, \tag{5.28}$$

where N is the normalization constant.

The function is of the same symmetry as the exact solution to the Schrödinger equation (antisymmetric with respect to the exchange of electrons and symmetric with respect to the exchange

[28] The symbols come from German: g or *gerade* (even) and u or *ungerade* (odd).

[29] We ignore the proton spins.

of protons). It is easy to calculate[30] that $N = 2[(1 + S^2)]^{-1/2}$, where S stands for the overlap integral of the atomic orbitals $1s_a$ and $1s_b$.

5.9.3 The perturbation is always large in polarization approximation

Let us check (Appendix V1-B) how distant functions $\varphi^{(0)}$ and $\psi^{(0)}$ are in the Hilbert space (they are both normalized, i.e., they are unit vectors in the Hilbert space). We will calculate the norm of difference $\varphi^{(0)} - \psi_0^{(0)}$. If the norm were small, then the two functions would be close in the Hilbert space. Let us see. We have

$$
\begin{aligned}
||\varphi^{(0)} - \psi_0^{(0)}|| &\equiv \left[\int (\varphi^{(0)} - \psi_0^{(0)})^*(\varphi^{(0)} - \psi_0^{(0)})\mathrm{d}\tau\right]^{\frac{1}{2}} \\
&= \left[1 + 1 - 2\int \varphi^{(0)}\psi_0^{(0)}\mathrm{d}\tau\right]^{\frac{1}{2}} = \left\{2 - 2\int \left[1s_a(1)\alpha(1)1s_b(2)\beta(2)\right]\right. \\
&\quad \left. N\left[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)\right]\left\{\frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right]\right\}\mathrm{d}\tau\right\}^{\frac{1}{2}} \\
&= \left\{2 - N\sqrt{2}\int \left[1s_a(1)1s_b(2)\right]\left[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)\right]\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\right\}^{\frac{1}{2}} \\
&= \left\{2 - \frac{1}{\sqrt{1+S^2}}(1 + S^2)\right\}^{\frac{1}{2}} = \{2 - \sqrt{1+S^2}\}^{1/2},
\end{aligned}
$$

where we have assumed that the functions are real. When $R \to \infty$, then $S \to 0$ and

$$
\lim_{R\to\infty} ||\varphi^{(0)} - \psi_0^{(0)}|| = 1 \neq 0. \tag{5.29}
$$

It is therefore clear that the Heitler–London wave function differs considerably and that this huge difference *does not vanish* when $R \to \infty$.

The two normalized functions $\varphi^{(0)}$ and $\psi_0^{(0)}$ represent two unit vectors in the Hilbert space (see Fig. 5.10). The scalar product of the two unit vectors $\left\langle \varphi^{(0)} | \psi_0^{(0)} \right\rangle$ is equal to

[30] $\int |\psi_{HL}|^2 \mathrm{d}\tau_1 \mathrm{d}\tau_2 = |N|^2(2 + 2S^2)\{\sum_{\sigma_1}\sum_{\sigma_2}\frac{1}{2}\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right]^2\} = |N|^2 2(1 + S^2)\frac{1}{2}(1 + 1 - 2\cdot 0) = 1$. Hence, $N = \frac{1}{\sqrt{2(1+S^2)}}$. In a moment we will need function $\psi_0^{(0)}$ with the intermediate normalization with respect to $\varphi^{(0)}$, i.e., satisfying $\langle \psi_0^{(0)} | \varphi^{(0)} \rangle = 1$, instead of $\langle \psi_0^{(0)} | \psi_0^{(0)} \rangle = 1$. Then N will be different and equal to $\langle \varphi^{(0)} | \hat{A}\varphi^{(0)} \rangle^{-1}$.

$\cos\theta$. Let us calculate this angle θ_{lim} which corresponds to R tending to ∞. The quantity $\lim_{R\to\infty} ||\varphi^{(0)} - \psi_0^{(0)}||^2 = \lim_{R\to\infty} \int(\varphi^{(0)} - \psi_0^{(0)})^*(\varphi^{(0)} - \psi_0^{(0)})\mathrm{d}\tau = \lim_{R\to\infty}[2 - 2\cos\theta] =$ 1. Hence, $\cos\theta_{\text{lim}} = \frac{1}{2}$, and therefore $\theta_{\text{lim}} = 60°$. This means that the three unit vectors $\varphi^{(0)}$, $\psi_0^{(0)}$, and $\varphi^{(0)} - \psi_0^{(0)}$ for $R \to \infty$ form an equilateral triangle, and therefore $\varphi^{(0)}$ represents a highly "handicapped" function, which lacks about a half with respect to a function of the proper symmetry.[31] This is certainly bad news.

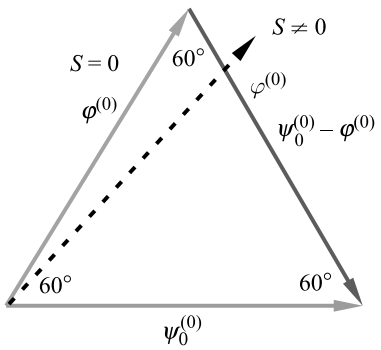


Fig. 5.10. A view from the Hilbert space showing that the polarization approximation is very bad even for the infinite intersystem distance. The normalized functions $\varphi^{(0)}$ and $\psi_0^{(0)}$ for the hydrogen molecule as unit vectors belonging to the Hilbert space. The functions differ widely at any intermolecular distance R. For $S = 0$, i.e., for long internuclear distances, the difference $\psi_0^{(0)} - \varphi^{(0)}$ represents a vector of the Hilbert space having length 1. Therefore, for $R = \infty$ the three vectors $\varphi^{(0)}$, $\psi_0^{(0)}$, and $\psi_0^{(0)} - \varphi^{(0)}$ form an equilateral angle. For shorter distances the angle between $\varphi^{(0)}$ and $\psi_0^{(0)}$ becomes smaller than 60°.

Therefore, the perturbation V has to be treated as *always large*, because it is responsible for a huge wave function change: from the unperturbed one of bad symmetry to the exact one of the correct symmetry.

In contrast to this, there would be no problem at all with the vanishing of the $||\psi_0^{(0)} - \psi_0||$ as $R \to \infty$, *where* ψ_0 *represents an exact ground-state solution of the Schrödinger equation. Indeed,* $\psi_0^{(0)}$ *correctly describes the dissociation of the molecule into two hydrogen atoms (both in the* 1*s state), as well as both functions having the same symmetry for all interatomic distances.* Therefore,

[31] In Appendix E (p. 601) we show how the charge distribution changes when the Pauli exclusion principle is forced by a proper projection of the $\varphi^{(0)}$ wave function.

the Heitler–London wave function represents a good approximation to the exact function for long (and we hope medium) intermolecular distances. *Unfortunately, it is not the eigenfunction of* $\hat{H}^{(0)}$ and therefore we cannot construct the usual Rayleigh–Schrödinger perturbation theory.

And this is the second item of bad news today...

5.9.4 Iterative scheme of SAPT

We now have two issues: either to construct another zero-order Hamiltonian, for which the $\psi_0^{(0)}$ function would be an eigenfunction (then the perturbation would be small and the Rayleigh–Schrödinger perturbation theory might be applied), or to abandon any Rayleigh–Schrödinger perturbation scheme and replace it by something else. The first of these possibilities was developed intensively in many laboratories, including ours in Warsaw. The approach had the deficiency that the operators appearing in the theories depended explicitly on the basis set used, and therefore there was no guarantee a basis-independent theory exists.

The second possibility relies on an iterative solution of the Schrödinger equation, forcing the proper symmetry of the intermediate functions. The method was proposed mainly by Bogumił Jeziorski and Włodzimierz Kołos.

Claude Bloch was probably the first to write the Schrödinger equation in the form shown in formulae[32] (2.94) and (2.83). Let us recall them in a notation adapted to the present situation. We write

$$\psi_0 = \varphi^{(0)} + \hat{R}_0(E_0^{(0)} - E_0 + V)\psi_0,$$

$$E_0 = E_0^{(0)} + \langle \varphi^{(0)} | V \psi_0 \rangle,$$

where we assume that $\varphi^{(0)}$ satisfies

$$\hat{H}^{(0)}\varphi^{(0)} = E_0^{(0)}\varphi^{(0)}$$

with the eigenvalues of the unperturbed Hamiltonian $\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B$ given as the sum of the energies of the isolated molecules A and B, i.e.,

$$E_0^{(0)} = E_{A,0} + E_{B,0},$$

[32] C. Bloch, *Nucl. Phys.*, 6(1958)329.

and ψ_0 is the exact ground-state solution to the Schrödinger equation with the total nonrelativistic Hamiltonian $\hat{H}$ of the system,

$$\hat{H}\psi_0 = E_0\psi_0.$$

We focus our attention on the difference $\mathcal{E}_0$ between E_0, which is our target, and $E_0^{(0)}$, which is at our disposal as the unperturbed energy. We may write the Bloch equations in a form exposing the interaction energy $\mathcal{E}_0{=}E_0 - E_0^{(0)}$, i.e.,

$$\psi_0 = \varphi^{(0)} + \hat{R}_0(-\mathcal{E}_0 + V)\psi_0,$$
$$\mathcal{E}_0 = \langle\varphi^{(0)}|V\psi_0\rangle.$$

The equations are valid for intermediate normalization $\langle\varphi^{(0)}|\psi_0\rangle = 1$. This system of equations for $\mathcal{E}_0$ and ψ_0 might be solved by an iterative method:

ITERATIVE SCHEME:

$$\psi_0(n) = \varphi^{(0)} + \hat{R}_0[-\mathcal{E}(n) + V]\psi_0(n-1), \tag{5.30}$$
$$\mathcal{E}_0(n) = \langle\varphi^{(0)}|V\psi_0(n-1)\rangle, \tag{5.31}$$

$n = 1, 2, 3, ...$, where n is the *iteration* number.

In practice everything depends on the starting point chosen, how many iterations have to be done, and whether convergence will be achieved. One of the most beautiful features of iterative schemes is that despite the freedom is usually very large, we are able to reach the limit, i.e., to get $||\psi_0(n+1) - \psi_0(n)|| < \varepsilon$ and $|\mathcal{E}_0(n+1) - \mathcal{E}_0(n)| < \varepsilon$ for $n > n_0$ and for an arbitrarily small $\varepsilon > 0$.

Polarization scheme replaced

We start in the zeroth iteration with $\psi_0\,(0) = \varphi^{(0)}$. From the second of the iterative equations we get right away $\mathcal{E}_0(1) = \langle\varphi^{(0)}|V\varphi^{(0)}\rangle \equiv E_{\text{pol}}^{(1)}$, which is the interaction energy accurate up to the first order of the polarization perturbational scheme. After inserting this result and $\psi_0(0) = \varphi^{(0)}$ to the first equation, one obtains (note that $\hat{R}_0\varphi^{(0)} = 0$) $\psi_0(1) = \varphi^{(0)} + \hat{R}_0(V - \langle\varphi^{(0)}|V\varphi^{(0)}\rangle)\varphi^{(0)} = \varphi^{(0)} + \hat{R}_0V\varphi^{(0)}$. This is a sum of the unperturbed wave function and of the first-order correction (Eq. (V1-5.26) on p. V1-279), exactly what we have obtained in the polarization approximation. The second iteration gives $\mathcal{E}_0(2) = \langle\varphi^{(0)}|V\psi_0(1)\rangle = \langle\varphi^{(0)}|V\left(\varphi^{(0)} + \hat{R}_0V\varphi^{(0)}\right)\rangle = E_{\text{pol}}^{(1)} + \langle\varphi^{(0)}|V\hat{R}_0V\varphi^{(0)}\rangle = E_{\text{pol}}^{(1)} + E_{\text{pol}}^{(2)}$, because the second term

represents nothing else but the second-order energy correction in the polarization approximation.

When repeating the above iterative scheme and grouping the individual terms according to the powers of V, at each turn we obtain the exact expression appearing in the Rayleigh–Schrödinger polarization approximation (Chapter V1-5) plus some higher-order terms.

It is worth noting that $\mathcal{E}_0(n)$ is the *sum* of all corrections of the Rayleigh–Schrödinger *up to the* n-th order with respect to V plus some higher-order terms. For large R, the quantity $\mathcal{E}_0(n)$ is an arbitrarily good approximation to the exact interaction energy.

Of course, the rate at which the iterative procedure converges depends very much on the starting point chosen. From this point of view, the start from $\psi_0(0) = \varphi^{(0)}$ is particularly unfortunate, because the remaining (roughly) 50% of the *wave function* has to be restored by the hard work of the perturbational series (high-order corrections are needed). This will be especially pronounced for long intermolecular distances, where the exchange interaction energy will not be reproduced in any finite order.

Murrell–Shaw and Musher–Amos (MS-MA) perturbation theory

A much more promising starting point in Eq. (5.30) seems to be $\psi_0(0) = \psi_0^{(0)}$, because the symmetry of the wave function is already correct and the function itself represents an exact solution at $R = \infty$. For convenience the intermediate normalization is used (see p. 275), $\left\langle \varphi^{(0)} | \psi_0^{(0)} \right\rangle = 1$, i.e., $\psi_0^{(0)} = N\hat{\mathcal{A}}\varphi^{(0)}$ with $N = \langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle^{-1}$. The first iteration of Eqs. (5.30) and (5.31) gives the first-order correction to the energy (which we split into the polarization part and the rest)

$$\mathcal{E}_0(1) = N\langle \varphi^{(0)} | V\hat{\mathcal{A}}\varphi^{(0)} \rangle = E_{\mathrm{pol}}^{(1)} + E_{\mathrm{exch}}^{(1)}, \qquad (5.32)$$

$$E_{\mathrm{pol}}^{(1)} \equiv E_{\mathrm{elst}} = \langle \varphi^{(0)} | V\varphi^{(0)} \rangle.$$

We have obtained the electrostatic energy already known from the polarization approximation plus an important correction $E_{\mathrm{exch}}^{(1)}$ which we will discuss in a minute.

The first-iteration wave function will be obtained in the following way. First, we will use the commutation relation $\hat{\mathcal{A}}\hat{H} = \hat{H}\hat{\mathcal{A}}$ or

$$\hat{\mathcal{A}}(\hat{H}^{(0)} + V) = (\hat{H}_0 + V)\hat{\mathcal{A}}. \qquad (5.33)$$

Of course,

$$\hat{\mathcal{A}}(\hat{H}^{(0)} - E_0^{(0)} + V) = (\hat{H}^{(0)} - E_0^{(0)} + V)\hat{\mathcal{A}}, \tag{5.34}$$

which gives[33] $V\hat{\mathcal{A}} - \hat{\mathcal{A}}V = [\hat{\mathcal{A}}, \hat{H}^{(0)} - E_0^{(0)}]$, as well as $(V - \mathcal{E}_1)\hat{\mathcal{A}} = \hat{\mathcal{A}}(V - \mathcal{E}_1) + [\hat{\mathcal{A}}, \hat{H}^0 - E_0^{(0)}]$. Now we are ready to use Eq. (5.31) with $n = 1$. We have

$$\begin{aligned}
\psi_0(1) &= \varphi^{(0)} + \hat{R}_0(V - \mathcal{E}_0(1))\psi_0^{(0)} = \varphi^{(0)} + N\hat{R}_0(V - \mathcal{E}_0(1))\hat{\mathcal{A}}\varphi^{(0)} \\
&= \varphi^{(0)} + N\hat{R}_0\{\hat{\mathcal{A}}(V - \mathcal{E}_0(1)) + \\
&\quad \hat{\mathcal{A}}(\hat{H}^{(0)} - E_0^{(0)}) - (\hat{H}^{(0)} - E_0^{(0)})\hat{\mathcal{A}}\}\varphi^{(0)} \\
&= \varphi^{(0)} + N\hat{R}_0\hat{\mathcal{A}}(V - \mathcal{E}_0(1))\varphi^{(0)} + N\hat{R}_0\hat{\mathcal{A}}(\hat{H}^{(0)} - E_0^{(0)})\varphi^{(0)} \\
&\quad - N\hat{R}_0(\hat{H}^{(0)} - E_0^{(0)})\hat{\mathcal{A}}\varphi^{(0)}.
\end{aligned}$$

The third term is equal to 0, because $\varphi^{(0)}$ is an eigenfunction of $\hat{H}^{(0)}$ with an eigenvalue $E_0^{(0)}$. The fourth term may be transformed by decomposing $\hat{\mathcal{A}}\varphi^{(0)}$ into the vector (in the Hilbert space) parallel to $\varphi^{(0)}$ or $\langle\hat{\mathcal{A}}\varphi^{(0)}|\varphi^{(0)}\rangle\varphi^{(0)}$ and the vector orthogonal to $\varphi^{(0)}$, or $(1 - |\varphi^{(0)}\rangle\langle\varphi^{(0)}|)\mathcal{A}\varphi^{(0)}$. The result of $\hat{R}_0(\hat{H}^{(0)} - E_0^{(0)})$ acting on the first vector is zero (p. 166), while the second vector gives $(1 - |\varphi^{(0)}\rangle\langle\varphi^{(0)}|)\mathcal{A}\varphi^{(0)}$. This gives as the first-iteration ground-state wave function $\psi_0(1)$

$$\begin{aligned}
\psi_0(1) &= \varphi^{(0)} + N\hat{R}_0\hat{\mathcal{A}}(V - \mathcal{E}_0(1))\varphi^{(0)} + N\hat{\mathcal{A}}\varphi^{(0)} - N\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle\varphi^{(0)} \\
&= \frac{\hat{\mathcal{A}}\varphi^{(0)}}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} + N\hat{R}_0\hat{\mathcal{A}}(V - \mathcal{E}_0(1))\varphi^{(0)} = \hat{\mathcal{B}}\varphi^{(0)} - N\hat{R}_0\hat{\mathcal{A}}(\mathcal{E}_0(1) - V)\varphi^{(0)},
\end{aligned}$$

where

$$\hat{\mathcal{B}}\varphi^{(0)} = \frac{\hat{\mathcal{A}}\varphi^{(0)}}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle}. \tag{5.35}$$

After inserting $\psi_0(1)$ into the iterative scheme (5.30) with $n = 2$ we obtain the second-iteration energy

[33] Let us stress *en passant* that the left-hand side is of the first order in V, while the right-hand side is of the zeroth order. Therefore, in SAPT, the order is not a well-defined quantity, its role is taken over by the iteration number.

$$\mathcal{E}_0\,(2) = \langle \varphi^{(0)} | V\,\psi_0\,(1)\rangle = \frac{\langle \varphi^{(0)} | V\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)}\rangle} - N\langle \varphi^{(0)} | V\,\hat{R}_0\hat{\mathcal{A}}[\mathcal{E}_0\,(1) - V]\varphi^{(0)}\rangle. \tag{5.36}$$

These equations are identical to the corresponding corrections in perturbation theory derived by Murrell and Shaw[34] and by Musher and Amos[35] (MS–MA).

5.9.5 Symmetry forcing

Finally, there is good news. It turns out that we may formulate a general iterative scheme which is able to produce various procedures, known and unknown in the literature. In addition the scheme has been designed by my nearest neighbor colleagues (Jeziorski and Kołos). This scheme reads as

$$\psi_0(n) = \varphi^{(0)} + \hat{R}_0[-\mathcal{E}_0(n) + V]\hat{\mathcal{F}}\psi_0(n-1),$$
$$\mathcal{E}_0(n) = \langle \varphi^{(0)} | V\hat{\mathcal{G}}\psi_0(n-1)\rangle,$$

where in Eqs. (5.30) and (5.31) we have inserted operators $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$ which have to fulfill the obvious condition

$$\hat{\mathcal{F}}\psi_0 = \hat{\mathcal{G}}\psi_0 = \psi_0, \tag{5.37}$$

where ψ_0 is the solution to the Schrödinger equation.

WHY FORCE THE SYMMETRY?

At the end of the iterative scheme (convergence) the insertion of the operators $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$ has no effect at all, but *before that* their presence may be crucial for the numerical convergence. This is the goal of symmetry forcing.

This method of generating perturbation theories has been called by the authors the *symmetry forcing* method in SAPT.

[34] J.N. Murrell, G. Shaw, *J. Chem. Phys.*, 46(1867)1768.
[35] J.I. Musher, A.T. Amos, *Phys. Rev.*, 164(1967)31.

Table 5.2. Symmetry forcing in various perturbation schemes of the SAPT compared with the polarization approximation. The operator $\hat{\mathcal{B}}$ is defined by $\hat{\mathcal{B}}\chi = \hat{\mathcal{A}}\chi / \langle\varphi^{(0)}|\hat{\mathcal{A}}\chi\rangle$.

Perturbation scheme	$\psi_0(0)$	$\hat{\mathcal{F}}$	$\hat{\mathcal{G}}$
polarization	$\varphi^{(0)}$	1	1
symmetrized polarization	$\varphi^{(0)}$	1	$\hat{\mathcal{B}}$
MS–MA	$\hat{\mathcal{B}}\varphi^{(0)}$	1	1
Jeziorski–Kołos scheme[a]	$\hat{\mathcal{B}}\varphi^{(0)}$	$\hat{\mathcal{A}}$	1
EL-HAV[b]	$\hat{\mathcal{B}}\varphi^{(0)}$	$\hat{\mathcal{A}}$	$\hat{\mathcal{B}}$

[a] B. Jeziorski, W. Kołos, *Int. J. Quantum Chem.*, 12(1977)91.
[b] Eisenschitz–London and Hirschfelder–van der Avoird perturbation theory: R. Eisenschitz, F. London, *Zeit. Phys.*, 60(1930)491; J.O. Hirschfelder, *Chem. Phys. Letters*, 1(1967)363; A. van der Avoird, *J. Chem. Phys.*, 47(1967)3649.

Polarization collapse removed

The corrections obtained in SAPT differ from those of the polarization perturbational method.

To show the relation between the results of the two approaches, let us first introduce some new quantities. The first is an idempotent antisymmetrizer

$$\hat{\mathcal{A}} = C\hat{\mathcal{A}}^A\hat{\mathcal{A}}^B(1+\hat{P})$$

with $C = \frac{N_A!N_B!}{(N_A+N_B)!}$, where $\hat{\mathcal{A}}^A$, $\hat{\mathcal{A}}^B$ are idempotent antisymmetrizers for molecules A and B, each molecule contributing N_A and N_B electrons. Permutation operator $\hat{P}$ contains all the electron exchanges between molecules A and B, i.e.,

$$\hat{P} = \hat{P}^{AB} + \hat{P}',$$
$$\hat{P}^{AB} = -\sum_{i\in A}\sum_{j\in B}\hat{P}_{ij},$$

with $\hat{P}^{AB}$ denoting the single exchanges only, and $\hat{P}'$ the rest of the permutations, i.e., the double, triple, etc., exchanges. Let us stress that $\varphi^{(0)} = \psi_{A,0}\psi_{B,0}$ represents a product of two antisymmetric functions[36] and therefore $\hat{\mathcal{A}}\varphi^{(0)} = C(1+\hat{P}^{AB}+\hat{P}')\psi_{A,0}\psi_{B,0}$. Taking into account the operator $\hat{P}$ in $\langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle$ and $\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle$ produces

$$E^{(1)} = \frac{\langle\psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle + \langle\psi_{A,0}\psi_{B,0}|V\hat{P}^{ab}\psi_{A,0}\psi_{B,0}\rangle + O(S^4)}{1+\langle\psi_{A,0}\psi_{B,0}|\hat{P}^{ab}\psi_{A,0}\psi_{B,0}\rangle + O(S^4)}, \tag{5.38}$$

and therefore

[36] The product itself does not have this symmetry.

in the polarization approximation

$$E_{\text{pol}}^{(1)} \equiv E_{\text{elst}} = \langle \varphi^{(0)} | V \varphi^{(0)} \rangle \tag{5.39}$$

while in the SAPT

$$E^{(1)} = \frac{\langle \varphi^{(0)} | V \hat{\mathcal{A}} \varphi^{(0)} \rangle}{\langle \varphi^{(0)} | \hat{\mathcal{A}} \varphi^{(0)} \rangle}, \tag{5.40}$$

$$E^{(1)} = E_{\text{pol}}^{(1)} + E_{\text{exch}}^{(1)}, \tag{5.41}$$

where *the exchange interaction* in the first-order perturbation theory

$$\begin{aligned} E_{\text{exch}}^{(1)} = {} & \langle \psi_{A,0} \psi_{B,0} | V P^{AB} \psi_{A,0} \psi_{B,0} \rangle \\ & - \langle \psi_{A,0} \psi_{B,0} | V \psi_{A,0} \psi_{B,0} \rangle \langle \psi_{A,0} \psi_{B,0} | P^{AB} \psi_{A,0} \psi_{B,0} \rangle + O(S^4). \end{aligned} \tag{5.42}$$

In the most commonly encountered interaction of closed shell molecules the $E_{\text{exch}}^{(1)}$ term represents the *valence repulsion.*

The symbol $O(S^4)$ stands for all the terms that vanish with the fourth power of the overlap integrals or faster.[37] The valence repulsion already appears (besides the electrostatic energy $E_{\text{pol}}^{(1)}$) in the first order of the perturbation theory as a result of the Pauli exclusion principle.[38]

We have gained a remarkable thing, which may be seen by taking the example of two interacting subsystems: Na^+ and Cl^-. In the polarization approximation the electrostatic, induction, and dispersion contributions to the interaction energy are negative, the total energy will go down, and we would soon have a catastrophe: both subsystems would occupy the same place in space and according to the energy calculated (that could attain even $-\infty$, Fig. 5.11) the system would be extremely happy. This is absurd.

If this were true, we could not exist. Indeed, sitting safely on a chair we have an equilibrium of the gravitational force and..., well, and what? First of all, the force coming from the valence

[37] This means that we also take into account such a decay in other than overlap integrals, e.g., $\langle 1s_a 1s_b | 1s_b 1s_a \rangle$ is of the order S^2, where $S = \langle 1s_a | 1s_b \rangle$. Thus the criterion is the differential overlap rather than the overlap integral.

[38] An intriguing idea: the polarization approximation should be an extremely good approximation for the interaction of a molecule with an antimolecule (built from antimatter). Indeed, in the molecule we have electrons, in the antimolecule we have positrons, and no antisymmetrization (between the systems) is needed. Therefore a product wave function should be a very good starting point. No valence repulsion will appear, the two molecules will penetrate like ghosts. Soon after, "a tremendous lightning will be seen and the terrible thunder of annihilation will be heard." The system will disappear.

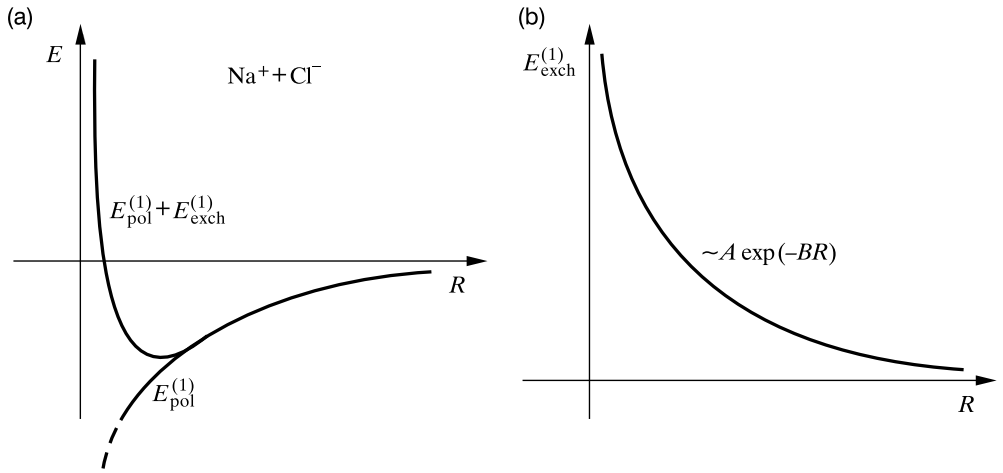


Fig. 5.11. Interaction energy of Na^+ and Cl^-. The polarization approximation gives an absurdity for small separations: the subsystems attract very strongly (mainly because of the electrostatic interaction), while they should repel very strongly. (a) The absurdity is removed when the valence repulsion is taken into account. (b) The valence repulsion alone modeled by the term $A\exp(-BR)$, where A and B are positive constants.

repulsion we are talking about. It is claimed sometimes that quantum effects are peculiar to small objects (electrons, nuclei, atoms, molecules) and are visible when dealing with such particles. We see, however, that we owe even sitting on a chair to the Pauli exclusion principle (a quantum effect).

The valence repulsion removes the absurdity of the polarization approximation, which made the collapse of the two subsystems possible.

5.9.6 A link to the variational method – the Heitler–London interaction energy

Since the $\hat{\mathcal{A}}\varphi^{(0)}$ wave function is a good approximation to the exact ground-state wave function at high values of R, we may calculate what is called the Heitler–London interaction energy (E_{int}^{HL}) as the mean value of the total (electronic) Hamiltonian minus the energies of the isolated subsystems, i.e.,

$$E_{int}^{HL} = \frac{\langle \hat{\mathcal{A}}\varphi^{(0)} | \hat{H} \hat{\mathcal{A}}\varphi^{(0)} \rangle}{\langle \hat{\mathcal{A}}\varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle} - (E_{A,0} + E_{B,0}).$$

This expression may be transformed in the following way:

$$\begin{aligned} E_{int}^{HL} &= \frac{\langle\varphi^{(0)}|\hat{H}\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} - (E_{A,0}+E_{B,0}) \\ &= \frac{\langle\varphi^{(0)}|\hat{H}^{(0)}\hat{\mathcal{A}}\varphi^{(0)}\rangle + \langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} - (E_{A,0}+E_{B,0}) \\ &= \frac{(E_{A,0}+E_{B,0})\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle + \langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} - (E_{A,0}+E_{B,0}) \\ &= \frac{\langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle}. \end{aligned}$$

Therefore, the Heitler–London interaction energy is equal to the first-order SAPT energy, so we have

$$E_{int}^{HL} = E^{(1)}.$$

5.9.7 Summary: the main contributions to the interaction energy

From the first two iterations ($n = 2$) of the SAPT scheme (Eq. (5.30), p. 382) we got the energy $\mathcal{E}_0\,(2)$, which now will be written in a slightly different form, i.e.,

$$\mathcal{E}_0\,(2) = E_{\text{elst}} + E_{\text{exch}}^{(1)} - \frac{\langle\varphi^{(0)}|V\hat{R}_0\hat{\mathcal{A}}[\mathcal{E}_0\,(1) - V]\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle}. \tag{5.43}$$

We recognize on the right-hand side (the first two terms) the complete first-order contribution, i.e., the electrostatic energy (E_{elst}) and the valence repulsion energy ($E_{\text{exch}}^{(1)}$). From the definition of the induction and dispersion energies (Eqs. (5.13) on p. 355), as well as from the reduced resolvent $\hat{R}_0$ of Eq. (2.88) on p. 165 (applied here to the individual molecules), one may write

$$\hat{R}_0 = \hat{R}_{0,\text{ind}} + \hat{R}_{0,\text{disp}}, \tag{5.44}$$

where the induction part of the resolvent

$$\hat{R}_{0,\text{ind}} = \hat{R}_{0,\text{ind}(A\to B)} + \hat{R}_{0,\text{ind}(B\to A)} \tag{5.45}$$

corresponds to deforming B by A and *vice versa*, i.e.,

$$\hat{R}_{0,\text{ind}(A\to B)} = \sum_{n_B(\neq 0)} \frac{|\psi_{A,0}\psi_{B,n_B}\rangle\langle\psi_{A,0}\psi_{B,n_B}|}{E_{B,0} - E_{B,n_B}}, \tag{5.46}$$

$$\hat{R}_{0,\mathrm{ind}(B\to A)} = \sum_{n_A(\neq 0)} \frac{\left|\psi_{A,n_A}\psi_{B,0}\right\rangle\left\langle\psi_{A,n_A}\psi_{B,0}\right|}{E_{A,0} - E_{A,n_A}}, \tag{5.47}$$

while the dispersion part of the resolvent reads as

$$\hat{R}_{0,\mathrm{disp}} = \sum_{(n_A,n_B)\neq(0,0))} \frac{\left|\psi_{A,n_A}\psi_{B,n_B}\right\rangle\left\langle\psi_{A,n_A}\psi_{B,n_B}\right|}{\left(E_{A,0} - E_{A,n_A}\right) + \left(E_{B,0} - E_{B,n_B}\right)}. \tag{5.48}$$

Let us insert these formulae into Eq. (5.43). We get (because $\hat{R}_{0,\mathrm{ind}}\varphi^{(0)} = \hat{R}_{0,\mathrm{ind}}\psi_{A,0}\psi_{B,0} = 0$ and, for similar reasons, also $\hat{R}_{0,\mathrm{disp}}\varphi^{(0)} = 0$)

$$\mathcal{E}_0\,(2) - (E_{\mathrm{elst}} + E^{(1)}_{\mathrm{exch}}) = \tag{5.49}$$

$$\frac{\langle\varphi^{(0)}|V\hat{R}_{0,\mathrm{ind}}\hat{\mathcal{A}}V\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} + \frac{\langle\varphi^{(0)}|V\hat{R}_{0,\mathrm{disp}}\hat{\mathcal{A}}V\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} = \tag{5.50}$$

$$(E_{\mathrm{ind}} + E^{(2)}_{\mathrm{ind-exch}}) + (E_{\mathrm{disp}} + E^{(2)}_{\mathrm{disp-exch}}). \tag{5.51}$$

In this way we have introduced the following:

- *The induction-exchange energy* $E^{(2)}_{\mathrm{ind-exch}}$ representing a modification of the induction energy, known from the polarization approximation (E_{ind}),

$$E^{(2)}_{\mathrm{ind-exch}} = \frac{\langle\psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{ind}}\hat{\mathcal{A}}V\psi_{A,0}\psi_{B,0}\rangle}{\langle\psi_{A,0}\psi_{B,0}|\hat{\mathcal{A}}\left(\psi_{A,0}\psi_{B,0}\right)\rangle} - E_{\mathrm{ind}}.$$

Note that after using the definition of the antisymmetrization operator $\hat{\mathcal{A}}$ (Eq. (V1-N.1) on p. V1-707) as well as applying the Taylor expansion, one gets as the first term, i.e., $\langle\psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{ind}}V\psi_{A,0}\psi_{B,0}\rangle \equiv E_{\mathrm{ind}}$, which cancels out the second term of the last equation. The other terms,

$$E^{(2)}_{\mathrm{ind-exch}} = \langle\psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{ind}}\hat{P}_{AB}V\psi_{A,0}\psi_{B,0}\rangle \tag{5.52}$$

$$- \langle\psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{ind}}V\psi_{A,0}\psi_{B,0}\rangle\langle\psi_{A,0}\psi_{B,0}|\hat{P}_{AB}\left(\psi_{A,0}\psi_{B,0}\right)\rangle \tag{5.53}$$

$$+ O(S^4), \tag{5.54}$$

represent, therefore, the corrections for the induction energy due to the forcing of the Pauli exclusion principle (for they appear as a result of the antisymmetrization) and are said to take care of the electron exchanges between the two molecules (due to the single electron exchanges from the permutation operator $\hat{P}_{AB}$, the double exchanges, etc., see p. 387).[39]

[39] One more thing. Because $\hat{R}_{0,\mathrm{ind}} = \hat{R}_{0,\mathrm{ind}(A\to B)} + \hat{R}_{0,\mathrm{ind}(B\to A)}$ one may split $E^{(2)}_{\mathrm{ind-exch}}$ into a part associated with a modification of polarization (due to the valence repulsion) of B by A and a similar part for polarization modification of A by B.

- *The dispersion-exchange energy* $E^{(2)}_{\mathrm{disp-exch}}$, which is a modification of the dispersion energy defined in the polarization perturbation theory (E_{disp}),

$$E^{(2)}_{\mathrm{disp-exch}} = \frac{\langle \psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{disp}}\hat{\mathcal{A}}V\psi_{A,0}\psi_{B,0}\rangle}{\langle \psi_{A,0}\psi_{B,0}|\hat{\mathcal{A}}\left(\psi_{A,0}\psi_{B,0}\right)\rangle} - E_{\mathrm{disp}}. \tag{5.55}$$

After representing $\hat{\mathcal{A}}$ by permutation operators and application of the Taylor expansion to the inverse of the denominator, the first term $\langle \psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{disp}}V\psi_{A,0}\psi_{B,0}\rangle \equiv E_{\mathrm{disp}}$ cancels out with the second term on the right-hand side. The other terms, as in the case of $E^{(2)}_{\mathrm{ind-exch}}$, result from the Pauli exclusion principle, i.e.,

$$E^{(2)}_{\mathrm{disp-exch}} = \langle \psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{disp}}\hat{P}_{AB}V\psi_{A,0}\psi_{B,0}\rangle \tag{5.56}$$

$$- \langle \psi_{A,0}\psi_{B,0}|V\hat{R}_{0,\mathrm{disp}}V\psi_{A,0}\psi_{B,0}\rangle\langle \psi_{A,0}\psi_{B,0}|\hat{P}_{AB}\left(\psi_{A,0}\psi_{B,0}\right)\rangle \tag{5.57}$$

$$+ O(S^4). \tag{5.58}$$

In the SAPT scheme, which takes into account the overlapping of the electron clouds of both interacting molecules, one obtains the terms known from the polarization approximation (no antisymmetrization) plus some important modifications.

- In the first order, due to the Pauli exclusion principle (i.e., to the antisymmetrization) the electrostatic energy E_{elst} known from the polarization approximation is supplemented by the valence repulsion energy $E^{(1)}_{\mathrm{exch}}$.
- In the second order, the Pauli deformation (cf. Appendix E on p. 601) of the electronic density in the AB complex results in exchange-based modifications ($E^{(2)}_{\mathrm{ind-exch}}$ and $E^{(2)}_{\mathrm{disp-exch}}$) of the induction and dispersion interactions (E_{ind} and E_{disp}) known from the polarization perturbation theory.

One may therefore suspect that an exchange-stemming restriction of polarization as well as of the electron correlation (the essence of the dispersion energy) due to the Pauli exclusion principle should result in a smaller energy gain, and therefore $E^{(2)}_{\mathrm{ind-exch}}$ and $E^{(2)}_{\mathrm{disp-exch}}$ should represent positive (i.e., destabilizing) energy effects. There is no theoretical evidence for this to be true, but the numerical results seem to confirm the above conjecture (Table 5.3).

The electron permutations lead to the integrals that decay with R similarly as powers of the overlap integrals, i.e., the above modifications decay very fast with increasing R.

Table 5.3. The SAPT contributions to the interaction energy calculated by using the DFT electronic density for Ne_2, $(H_2O)_2$, and $(C_6H_6)_2$. Numerical experience: the values for $E^{(2)}_{\mathrm{ind-exch}}$ and $E^{(2)}_{\mathrm{disp-exch}}$ are both positive!

Contribution	Ne_2 (cm^{-1})[a]	$(H_2O)_2$ ($\frac{kcal}{mol}$)[a]	$(C_6H_6)_2$ ($\frac{kcal}{mol}$)[b]
E_{elst}	−6.24	−7.17	−0.01
$E^{(1)}_{exch}$	26.40	5.12	3.82
E_{ind}	−4.76	−2.22	−1.30
$E^{(2)}_{ind-exch}$	4.88	1.13	1.07
E_{disp}	−46.23	−2.15	−5.77
$E^{(2)}_{disp-exch}$	1.85	0.36	0.51
E_{int}	−22.65	−4.50	−1.67

[a] A.J. Misquitta, K. Szalewicz, *J. Chem. Phys.*, 122(2005)214109.
[b] S. Tsuzuki, H.P. Lüthi, *J. Chem. Phys.*, 114(2001)3949.

5.10 Convergence problems

A very slow convergence (of "pathological" character) is bound to appear. To see why, let us recall the Heitler–London function (p. 378), which for very large R approximates very well the exact wave function. Since the function is so accurate, and therefore indispensable for a good description, an effective iterative process is expected to reproduce it in the first few iterations. Of course, the Heitler–London wave function can be represented as a linear combination of the functions belonging to a complete basis set, because any complete basis set can do it. In particular, a complete basis set can be formed from products of two functions belonging to complete one-electron basis sets[40]; the first one, describing electron 1, represents an orbital centered on nucleus a, while the second function (of electron 2) is an orbital centered on nucleus b. There is, however, a difficulty of a purely numerical character: the second half of the Heitler–London wave function corresponds to electron 1 at b, and electron 2 at a, *just opposite to what the functions belonging to the complete set describe*. This would result in prohibitively long expansions (related to a large number of iterations and a high order of the perturbation theory) necessary for the theory to reproduce the second half of the Heitler–London function. This effort is similar to expanding the function $1s_b$ (i.e., localized far away of the center a) as a linear combination of the functions $1s_a, 2s_a, 2p_{xa}, 2p_{xa}, 2p_{ya}, 2p_{za}$, ... It can be done in principle, but in practice it requires prohibitively long expansions, which should ensure a perfect destructive interference (of the wave functions) at a and, at the same time, a perfect constructive interference at b.

40 The hydrogen orbitals form a complete set, because they are the eigenfunctions of a Hermitian operator (Hamiltonian).

In perturbation theories all calculated corrections are simply *added* together. This may lead to partial sums that converge slowly or do not converge. This pertains also to the SAPTs. Why? Let us see Table 5.2. One of the perturbational schemes given there, namely, that called the symmetrized polarization approximation, is based on calculation of the wave function exactly as in the polarization approximation scheme, but just before calculation of the corrections to the energy, the polarization wave function is projected on the antisymmetrized space. This procedure is bound to cause trouble. The system excessively changes its charge distribution without paying any attention to the Pauli exclusion principle (thus allowing it to polarize itself in a nonphysical way; this may be described as "*overpolarization*"), and then the result has to be modified in order to fulfill a principle[41] (the Pauli principle).

This became evident after a study called the *Pauli blockade*.[42] It was shown that if the Pauli exclusion principle is not obeyed, the electrons of subsystem A can flow, without any penalty and totally nonphysically, to the low-energy orbitals of B. This may lead to occupation of that orbital by, e.g., four electrons (!), whereas the Pauli principle admits only a maximum of a double occupation.

Thus, any realistic deformation of the electron clouds has to take into account the exchange interaction (valence repulsion) or the Pauli principle. Because of this, we have introduced what is called the *deformation-exchange interaction energy*:

$$E_{\text{def}-\text{exch}} = \mathcal{E}_0(2) - (E_{\text{elst}} + E_{\text{disp}}). \tag{5.59}$$

5.10.1 Padé approximants may improve convergence

Any perturbational correction carries information. Summing up (this is the way we calculate the total effect) these corrections mean a certain processing of the information. We may ask an amazing question: *is there any possibility of taking the same corrections and squeezing out more information*[43] *than just making the sum*?

In 1892 Henri Padé[44] wrote his doctoral dissertation in mathematics and presented some fascinating results.

---

41 This is similar to allowing all plants to grow as they want and, just after harvesting everything, select the wheat alone. We cannot expect much from such an agriculture.

42 M. Gutowski, L. Piela, *Mol. Phys*., 64(1988)337.

43 That is, a more accurate result.

44 H. Padé, *Ann. Sci. École Norm. Sup. Suppl. [3]*, 9(1892)1.

For a power series

$$f(x) = \sum_{j=0}^{\infty} a_j x^j \tag{5.60}$$

we may define a *Padé approximant* $[L/M]$ as the ratio of two polynomials,

$$[L/M] = \frac{P_L(x)}{Q_M(x)}, \tag{5.61}$$

where $P_L(x)$ is a polynomial of at most the L-th order, while $Q_M(x)$ is a polynomial of the M-th order. The coefficients of the polynomials P_L and Q_M will be determined by the following condition:

$$f(x) - [L/M] = \text{terms of higher order than } x^{L+M}. \tag{5.62}$$

In other words,

the first $(L + M)$ terms of the Taylor expansion for a function $f(x)$ and for its Padé approximant are identical.

Since the nominator and denominator of the approximant can be harmlessly multiplied by any nonzero number, we may set, without losing anything, the following normalization condition:

$$Q_M(0) = 1. \tag{5.63}$$

Let us assume also that $P_L(x)$ and $Q_M(x)$ do not have any common factor (if they did, we should redefine Eq. (5.61)).

If we now write the polynomials as

$$P_L(x) = p_0 + p_1 x + p_2 x^2 + \ldots p_L x^L,$$
$$Q_M(x) = 1 + q_1 x + q_2 x^2 + \ldots q_M x^M,$$

then multiplying Eq. (5.61) by Q and forcing the coefficients at the same powers of x to be equal we obtain the following system of equations for the unknowns p_i and q_i (there are $L + M + 1$

of them, the number of equations is the same):

$$\begin{aligned}
&a_0 && = p_0,\\
&a_1 + a_0 q_1 && = p_1,\\
&a_2 + a_1 q_1 + a_0 q_2 && = p_2,\\
&a_L + a_{L-1} q_1 + \ldots a_0 q_L && = p_L,\\
&a_{L+1} + a_L q_1 + \ldots a_{L-M+1} q_M && = 0,\\
&\vdots\\
&a_{L+M} + a_{L+M-1} q_1 + \ldots a_L q_M && = 0.
\end{aligned} \tag{5.64}$$

Note, please, that the sum of the subscripts in either term is a constant integer from the range $[0, L + M]$, which is connected to the abovementioned equal powers of x.

Example 3 (Mathematical). The method is best illustrated in action. Let us take a function

$$f(x) = \frac{1}{\sqrt{1-x}}. \tag{5.65}$$

Let us expand f in a Taylor series, i.e.,

$$f(x) = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \frac{5}{16}x^3 + \frac{35}{128}x^4 + \ldots. \tag{5.66}$$

Therefore, $a_0 = 1$, $a_1 = \frac{1}{2}$, $a_2 = \frac{3}{8}$, $a_3 = \frac{5}{16}$, $a_4 = \frac{35}{128}$. Now let us forget that these coefficients came from the Taylor expansion of $f(x)$. Many other functions may have the same *beginning* of the Taylor series. Let us calculate some partial sums of the right-hand side of Eq. (5.66). We have the following results:

Approx. $f(\frac{1}{2})$	Sum up to the n-th term
$n = 1$	1.00000
$n = 2$	1.25000
$n = 3$	1.34375
$n = 4$	1.38281
$n = 5$	1.39990

We see that the Taylor series "works very hard" approaching $f(\frac{1}{2}) = \sqrt{2} = 1.414213562$; it succeeds, but not without pain and effort.

Now let us check how one of the simplest Padé approximants, namely, $[1/1]$, performs the same job. By definition

$$\frac{(p_0 + p_1 x)}{(1 + q_1 x)}. \tag{5.67}$$

Solving (5.64) gives as the approximant[45]

$$\frac{(1-\frac{1}{4}x)}{(1-\frac{3}{4}x)}. \tag{5.68}$$

Let us stress that information contained in the power series (5.60) has been limited to a_0, a_1, a_2 (all other coefficients *have not been used*). For $x=\frac{1}{2}$ the Padé approximant $[1/1]$ has the value

$$\frac{(1-\frac{1}{4}\frac{1}{2})}{(1-\frac{3}{4}\frac{1}{2})}=\frac{7}{5}=1.4, \tag{5.69}$$

which is *more effective* than the painful efforts of the Taylor series that used a coefficients up to a_4 (this gave 1.39990). To be fair, we have to compare the Taylor series result that used only a_0, a_1, a_2, and this gives only 1.34375! Therefore, the approximant failed by 0.01, while the Taylor series failed by 0.07. The Padé approximant $[2/2]$ has the form

$$[2,2]=\frac{(1-\frac{3}{4}x+\frac{1}{16}x^2)}{(1-\frac{5}{4}x+\frac{5}{16}x^2)}. \tag{5.70}$$

For $x=\frac{1}{2}$ its value is equal to $\frac{41}{29}=1.414$, which means an accuracy of 10^{-4}, while *without Padé approximants, but using the same information contained in the coefficients, we get an accuracy two orders of magnitude worse.*

Our procedure did not contain information that the function expanded is $(1-x)^{-\frac{1}{2}}$, for we gave the first five terms of the Taylor expansion only. Despite this, the procedure determined with high accuracy what will give higher terms of the expansion.

Example 4 (Quantum mechanical). This is not the end of the story yet. The reader will see in a minute some things which will be even more strange. Perturbation theory also represents a power series (with respect to λ) with coefficients that are energy corrections. If perturbation is small, the corrections are small as well; in general, the smaller, the higher the perturbation order. As a result, a partial sum of a few low-order corrections usually gives sufficient accuracy. However, the higher the order, the more difficult are the corrections to calculate. Therefore, we may ask if there is any possibility of obtaining good results and at a low price by using the Padé approximants. In Table 5.4 the results of a perturbational study of H_2^+ by Jeziorski et al. are presented.[46]

[45] Indeed, $L=M=1$, and therefore the equations for the coefficients p and q are the following: $p_0=1$, $\frac{1}{2}+q_1=p_1$, $\frac{3}{8}+\frac{1}{2}q_1=0$. This gives the solution $p_0=1$, $q_1=-\frac{3}{4}$, $p_1=-\frac{1}{4}$.

[46] B. Jeziorski, K. Szalewicz, M. Jaszuński, *Chem. Phys. Letters*, 61(1979)391.

Table 5.4. Convergence of the MS–MA perturbational series for the hydrogen atom in the field of a proton ($2p\sigma_u$ state) for internuclear distance R (a.u.). The error (in %) is given for the sum of the original perturbational series and for the Padé $[L+1, L]$ approximant, and is calculated with respect to the variational method (i.e., the best for the basis set used).

$2L+1$	$R = 12.5$		$R = 3.0$	
	pert. series	$[L+1, L]$	*pert. series*	$[L+1, L]$
3	0.287968	0.321460	0.265189	0.265736
5	0.080973	−0.303293	0.552202	−1.768582
7	0.012785	−0.003388	0.948070	0.184829
9	−0.000596	−0.004147	1.597343	0.003259
11	−0.003351	−0.004090	2.686945	0.002699
13	−0.003932	−0.004088	4.520280	0.000464
15	−0.004056	−0.004210	7.606607	0.000009
17	−0.004084	−0.001779	12.803908	0.000007
19	−0.004090	0.000337	21.558604	−0.000002
21	−0.004092	−0.000003	36.309897	0.000001

For $R = 12.5$ a.u., we see that the approximants had a very difficult task. First of all they "recognized" the series limit, already about $2L + 1 = 17$. Before that, they were less effective than the original series. It has to be stressed, however, that they "recognized" it *extremely* well (see $2L + 1 = 21$). In contrast to this, the (traditional) partial sums ceased to improve when L increased. This means that either the partial sum series converges[47] to a false limit or it converges to the correct limit, but does so extremely slowly. We see from the variational result (the error is calculated with respect to this) that the convergence *is* false.[48] If the variational result had not been known, we would say that the series has already converged. However, the Padé approximants said: "*no, this is a false convergence*," and they were right.

For $R = 3.0$ a.u. (see table) the original series represents a real tragedy. For this distance, the perturbation is too large and the perturbational series just evidently *diverges*. The greater our effort, the greater the error of our result. The error is equal to 13% for $2L + 1 = 17$, then to 22% for $2L + 1 = 19$, and attains 36% for $2L + 1 = 21$. Despite of these hopeless results, it turns out that the problem represent "peanuts" for the Padé approximants.[49] They were already much better for $L = 3$.

[47] We have only numerics as argument though... The described case represents an example of vicious behavior when summing up a series: starting from a certain order of the perturbational series the improvement becomes extremely slow (we see the "false" convergence), although we are far from the true limit.

[48] At least as we see it numerically.

[49] Similar findings are reported in T.M. Perrine, R.K. Chaudhuri, K.F. Freed, *Intern. J. Quantum Chem.*, 105(2005)18.

Why are the Padé approximants so effective?

The apparent garbage produced by the perturbational series for $R = 3.0$ a.u. represented for the Padé approximants precise information that the absurd perturbational corrections pertain the energy of the $2p\sigma_u$ state of the hydrogen atom in the electric field of the proton. How come? Visibly low-order perturbational corrections, even if absolutely crazy, somehow carry information about the physics of the problem. The convergence properties of the Rayleigh–Schrödinger perturbation theory depend critically on the poles of the function approximated (see the discussion on p. V1-284). A pole cannot be described by any power series (as happens in perturbation theories), whereas the Padé approximants have poles built in the very essence of their construction (the denominator as a polynomial). This is why they may fit so well to the nature of the problems under study.[50]

5.11 Nonadditivity of intermolecular interactions

5.11.1 Interaction energy represents the nonadditivity of the total energy

The *total* energy of interacting molecules is not an additive quantity, i.e., does not represent the sum of the energies of the isolated molecules. The reason for this nonadditivity is the interaction energy.

Let us see whether the interaction energy itself has some additive properties. First of all the interaction energy requires the declaration of which fragments of the total system we treat as (interacting) molecules (see the beginning of this chapter). The only real system is the total system, not these subsystems. Therefore, the subsystems can be chosen in many different ways.

If the theory is exact, the total system can be described at any such choice (cf. p. 75). Only the supermolecular theory that is invariant with respect to such choices is.[51] The perturbation theory so far has no such powerful feature (this problem is not even raised in the literature), because it requires the intra- and intermolecular interactions to be treated on the same footing.

5.11.2 Many-body expansion of the rigid interaction energy

A next question could be: *is the interaction energy pairwise additive*, i.e.,

> is the interaction energy a sum of *pairwise* interactions?

[50] There are cases, however, where Padé approximants may fail in a spectacular way.

[51] However, for rather trivial reasons, i.e., interaction energy represents a by-product of the method. The main goal is the total energy, which by definition is independent of the choice of subsystems.

If this were true, it would be sufficient to calculate all possible interactions of pairs of molecules in the configuration identical to that they have in the total system[52] and our problem would be solved.

For the time being let us take the example of a stone, a shoe, and a foot. The owner of the foot will certainly remember the nonzero value of the three-body interaction, while nothing special happens when you put a stone into the shoe, or your foot into the shoe, or a small stone on your foot (much smaller two-body interactions). The molecules behave similarly –

their molecular interactions are not pairwise additive.

In the case of three interacting molecules (as well as the shoes, etc.), there is an effect of a strictly three-body character, *which cannot be reduced to ("explained") two-body interactions*. Similarly for larger numbers of molecules, there is a nonzero four-body effect, because all cannot be calculated as two- and three-body interactions, etc.

In what is called the many-body expansion for N molecules A_1, A_2, ..., A_N the interaction energy $E_{int}(A_1A_2...A_N)$, i.e., the difference between the total energy $E_{A_1A_2...A_N}$ and the sum of the energies of the isolated molecules[53] E_{A_i}, can be represented as a series of m-body interaction terms $\Delta E\,(m, N)$, $m = 2, 3, ..., N$, i.e.,

$$E_{int} = E_{A_1A_2...A_N} - \sum_{i=1}^{N} E_{A_i} = \sum_{i>j}^{N} \Delta E_{A_iA_j}\,(2, N) + \sum_{i>j>k}^{N} \Delta E_{A_iA_jA_k}\,(3, N) + ...\Delta E_{A_1A_2...A_N}\,(N, N)\,. \tag{5.71}$$

The $\Delta E\,(m, N)$ contribution to the interaction energy of N molecules ($m \leq N$) represents the sum of the interactions of m molecules (all possible combinations of m molecules among N molecules keeping their configurations fixed as in the total system) inexplicable by the interactions of smaller number of molecules.

One more question. Should we really stay with the idea of the rigid interaction energy? For instance, we may be interested in how the conformation of the AB complex changes in the presence of molecule C. This is also a three-body interaction. Such dilemmas have not yet been solved in the literature.

52 This would be much less expensive than the calculation for the total system.

53 In the same internal configuration as that in the total system and with the fixed number of electrons.

Example 5 (Three helium atoms). Let us take the helium atoms in the equilateral triangle configuration with the side length $R = 2.52$ Å. This value corresponds to 90% of twice the helium atom van der Waals radius[54] (1.40 Å). The MP2 calculation[55] gives the energy for the three atoms equal to $E_{\mathrm{He}_3} = -8.5986551$ a.u., for a pair of atoms $E_{\mathrm{He}_2} = -5.7325785$ a.u., and for a single atom $E_{\mathrm{He}} = -2.8663605$ a.u. Therefore, the interaction energy of a single atomic pair is equal to 0.0001425 a.u.; three such interactions give what is known as the two-body contribution (or the two-body effect) $\triangle E_{\mathrm{two-body}} = \sum_{i<j}^{3} \Delta E_{A_i A_j}(2,3) = 0.0001425 \cdot 3 = 0.0004275$ a.u. The three-body effect represents the rest, i.e., is calculated as the difference between E_{He_3} and the sum of the one-body and two-body effects $\triangle E_{\mathrm{three-body}} = E_{\mathrm{He}_3} - E_{\mathrm{He}} \cdot 3 - \triangle E_{\mathrm{two-body}} = -0.000001$ a.u. Therefore, we get the ratio $\left|\frac{\triangle E_{\mathrm{three-body}}}{\triangle E_{\mathrm{two-body}}}\right| = 0.26\%$. This means the three-body effect is quite small.

Example 6 (Locality or communication). Do two helium atoms feel each other when separated by a bridge molecule? In other words, do the intermolecular interactions have the local character or not? By the local character we mean that their influence is limited to, say, the nearest neighbor atoms.[56] Let us check this by considering a bridge molecule (two cases: butane and butadiene) between two helium atoms. The first helium atom pushes the terminal carbon atom of the butane (the distance $R_{C\,He} = 1.5$ Å, the attack is perpendicular to the C–C bond). The question is whether the second helium atom in the equivalent position but at the other end of the molecule will feel that pushing or not (*the positions of all nuclei do not change*, we are within the Born–Oppenheimer approximation). This means we are interested in a three-body effect.

Let us calculate the three-body interaction for He$\cdots$butane$\cdots$He and for He$\cdots$butadiene$\cdots$He. A large three-body effect would mean a communication between the two helium atoms through the electronic structure of the bridge. For the butane bridge we get $\left|\frac{\triangle E_{\mathrm{three-body}}}{\triangle E_{\mathrm{two-body}}}\right| = 0.02\%$, and for the butadiene bridge we obtain $\left|\frac{\triangle E_{\mathrm{three-body}}}{\triangle E_{\mathrm{two-body}}}\right| = 0.53\%$. Thus, we see the locality, but it is quite remarkable that there is an order of magnitude larger helium–helium communication for the case of butadiene (the distance between the terminal carbon atoms is similar in both cases: 3.7 Å for butadiene and 3.9 Å for butane), which is known for having conjugated double and single bonds, C=C–C=C, while butane has only single C–C–C–C bonds. It seems we touch here one of the aspects of the all important difference in chemistry between the labile π electron structure (butadiene) and the stiff σ electron structure (butane).

---

54 The notion of the van der Waals radius of an atom (it will be introduced a bit later), even if very useful, is a bit arbitrary. The abovementioned 90% is set for didactic purpose in order to get the effect clearly visible.

55 The description of the MP2 method can be found on p. 169. In the calculations reported here the 6-31G(d) basis set is used, without BSSE correction.

56 The intermolecular interactions in general are certainly nonlocal in the above sense. As an example may serve the allosteric effect: when one ligand binds to a protein, it may help (or inhibit) binding of another ligand far away. Crucial there are conformational transformations, which are absent in our example.

5.11.3 What is additive, what is not?

Already vast experience has been accumulated and some generalizations are possible.[57] For three argon atoms in an equilibrium configuration, the three-body term is of the order of 1%. It should be noted, however, that in the argon *crystal* there is a lot of three-body interactions and the three-body effect increases to about 7%. On the other hand, for liquid water the three-body effect is of the order of 20%, and the higher contributions are about 5%. Three-body effects are sometimes able to determine the crystal structure and have a significant influence on the physical properties of the system close to a phase transition ("critical region").[58]

In the case of the interaction of metal atoms, the nonadditivity is much larger than that for the noble gases, and the three-body effects may attain several dozen percent. This is important information since the force fields widely used in molecular mechanics (see p. V1-406) are based almost exclusively on effective pairwise interactions (neglecting the three- and more-body contributions).[59]

Although the intermolecular interactions are nonadditive, we may ask whether individual contributions to the interaction energy (electrostatic, induction, dispersion, valence repulsion) are additive.

Let us begin from the electrostatic interaction.

5.11.4 Additivity of the electrostatic interaction

Suppose we have three molecules, A, B, C, intermolecular distances are long, and therefore it is possible to use the polarization perturbation theory, in a very similar way to that presented in the case of two molecules (p. 351). In this approach, the unperturbed Hamiltonian $\hat{H}^{(0)}$ represents the sum of the Hamiltonians for the isolated molecules A, B, C. Let us change the abbreviations a little bit to be more concise for the case of three molecules. A *product* function $\psi_{A,n_A}\psi_{B,n_B}\psi_{C,n_C}$ will be denoted by $|n_A n_B n_C\rangle = |n_A\rangle\,|n_B\rangle\,|n_C\rangle$, where n_A, n_B, n_C $(= 0, 1, 2, ...)$ stand for the quantum numbers corresponding to the orthonormal wave functions for molecules A, B, C, respectively. The functions $|n_A n_B n_C\rangle = |n_A\rangle\,|n_B\rangle\,|n_C\rangle$ are the eigenfunctions of $\hat{H}^{(0)}$, i.e.,

$$\hat{H}^{(0)}\,|n_A n_B n_C\rangle = [E_A(n_A) + E_B(n_B) + E_C(n_C)]\,|n_A n_B n_C\rangle\,,$$

[57] V. Lotrich, K. Szalewicz, *Phys. Rev. Letters*, 79(1997)1301.

[58] R. Bukowski, K. Szalewicz, *J. Chem. Phys.*, 114(2001)9518.

[59] That is, the effectivity of a force field relies on such a choice of interaction parameters that the experimental data are reproduced (in such a way the parameters implicitly contain part of the higher-order terms).

which is analogous to Eqs. (5.7) and (5.8) on p. 354.

The perturbation is equal to $\hat{H} - \hat{H}^{(0)} = V = V_{AB} + V_{BC} + V_{AC}$, where the operators V_{XY} contain all the Coulomb interaction operators involving the nuclei and electrons of molecule X and those of molecule Y.

Let us recall that the electrostatic interaction energy $E_{\text{elst}}(ABC)$ of the ground-state ($n_A = 0, n_B = 0, n_C = 0$) molecules is defined as the first-order correction to the wave function in the polarization approximation perturbation theory,[60]

$$E^{(1)}_{\text{pol}} \equiv E_{\text{elst}}(ABC) = \langle 0_A 0_B 0_C | V | 0_A 0_B 0_C \rangle = \langle 0_A 0_B 0_C | V_{AB} + V_{BC} + V_{AC} | 0_A 0_B 0_C \rangle \,,$$

where the quantum numbers 000 have been supplied (maybe because of my excessive caution) by the redundant and self-explanatory indices $(0_A, 0_B, 0_C)$.

The integration in the last formula goes over the coordinates of all electrons. In the polarization approximation, the electrons can be unambiguously divided into three groups: those belonging to A, B, and C. Because the zero-order wave function $|0_A 0_B 0_C\rangle$ represents a product $|0_A\rangle\,|0_B\rangle\,|0_C\rangle$, the integration over the electron coordinates of one molecule can be easily performed and yields

$$E_{\text{elst}}(A, B, C) = \left\langle 0_A 0_B | \hat{V}_{AB} | 0_A 0_B \right\rangle + \left\langle 0_B 0_C | \hat{V}_{BC} | 0_B 0_C \right\rangle + \left\langle 0_A 0_C | \hat{V}_{AC} | 0_A 0_C \right\rangle ,$$

where to get the first term, integration was performed over the electrons of C, to get the second over the electrons of A, and to get the third over those of B.

Now, let us look at the last formula. We easily see that the individual terms simply represent the electrostatic interaction energies of pairs of molecules: AB, BC, and AC, which we would obtain in the perturbational theory (within the polarization approximation) for the interactions of AB, BC, and AC, respectively. Conclusion:

> the electrostatic interaction is pairwise additive.

5.11.5 Exchange nonadditivity

What about the exchange contribution? This contribution *does not exist in the polarization approximation*. It appears only in SAPT in pure form in the first-order energy correction and coupled to other effects in higher-order energy corrections.[61] The exchange interaction is dif-

[60] The $E_{\text{elst}}(ABC)$ term in SAPT represents only part of the first-order correction to the energy (the rest being the valence repulsion).

[61] Such terms are bound to appear. For example, the induction effect is connected to deformation of the electron density distribution. The interaction (electrostatic, exchange, dispersive, etc.) of such a deformed object will change with respect to that of the isolated object. The coupling terms take care of this change.

ficult to interpret, because it appears as a result of the antisymmetry of the wave function (Pauli exclusion principle). The antisymmetry is forced by the postulates of quantum mechanics (p. V1-40) and its immediate consequence is that the probability density of finding two electrons with the same spin and space coordinates is equal to zero.

PAULI DEFORMATION

The Pauli exclusion principle leads to a *deformation* of the wave functions describing the two molecules (by projecting the product-like wave function by the antisymmetrizer $\hat{A}$) with respect to the product-like wave function. The Pauli deformation (cf. Appendix E) appears already in the zeroth order of perturbation theory, whereas in the polarization approximation, the deformation of the wave function appears in the first order and is not related to the Pauli exclusion principle.

The antisymmetrizer pertains to the permutation symmetry of the wave function with respect to the coordinates of all electrons and therefore is different for a pair of molecules and for a system of three molecules.

The expression for the three-body nonadditivity of the valence repulsion[62] is given by Eq. (5.43), based on definition (5.40) of the first-order correction in SAPT[63] and definition (5.71) of the three-body contribution. We have

$$E^{(1)}_{\mathrm{exch},ABC} = N_{ABC}\,\langle 0_A 0_B 0_C|\,\hat{V}_{AB}+\hat{V}_{BC}+\hat{V}_{AC}\,\left|\hat{A}_{ABC}(0_A 0_B 0_C)\right\rangle - \sum_{(XY)=(AB),(AC),(BC)} N_{XY}\,\langle 0_X 0_Y|\,\hat{V}_{XY}|\hat{A}_{XY}(0_X 0_Y), \tag{5.72}$$

where $N_{ABC}\hat{A}_{ABC}\,|0_A 0_B 0_C\rangle$, $N_{AB}\hat{A}_{AB}\,|0_A 0_B\rangle$, etc., represent the normalized (N_{ABC} etc. are the normalization coefficients) antisymmetrized product-like wave functions of the systems ABC, AB, etc. The antisymmetrizer $\hat{A}_{ABC}$ pertains to subsystems A, B, C, similarly $\hat{A}_{AB}$ pertains to A and B, etc., all antisymmetrizers containing only the intersystem electron exchanges, and the summation goes over all pairs of molecules.

There is no chance of proving that the exchange interaction is additive, i.e., that Eq. (5.72) gives 0. Let us consider the simplest possible example, where each molecule has only a single electron: $|0_A(1)\,0_B(2)\,0_C(3)\rangle$. The operator $\hat{A}_{ABC}$ makes (besides other terms) the following permutation: $\hat{A}_{ABC}\,|0_A(1)\,0_B(2)\,0_C(3)\rangle = \ldots - \frac{1}{(N_A+N_B+N_C)!}\,|0_A(3)\,0_B(2)\,0_C(1)\rangle + \ldots$,

62 B. Jeziorski, M. Bulski, L. Piela, *Intern. J. Quantum Chem.*, 10(1976)281.

63 Because, as we have already proved, the rest, i.e., the electrostatic energy, is an additive quantity.

which according to Eq. (5.72) leads to the integral

$$-\frac{1}{(N_A+N_B+N_C)!}\,N_{ABC}\,\langle 0_A(1)\,0_B(2)\,0_C(3)|\,\frac{1}{r_{12}}\,|0_A(3)\,0_B(2)\,0_C(1)\rangle =$$
$$-\frac{1}{(N_A+N_B+N_C)!}\,N_{ABC}\,\langle 0_A(1)\,0_B(2)|\,\frac{1}{r_{12}}\,|0_B(2)\,0_C(1)\rangle$$

involving the wave functions centered on A, B, and C. This means that the term belongs to the *three-body* effect.

The permutation operators of which the $\hat{A}_{ABC}$ operator is composed correspond to the identity permutation[64] as well as to the exchange of one or more electrons between the interacting subsystems: $\hat{A}_{ABC} = 1+$ single exchanges $+$ double exchanges$+$...

It is easy to demonstrate[65] that

> *the larger the number of electrons exchanged, the less important such exchanges are*, because the resulting contributions would be proportional to higher and higher powers of the (as a rule small) overlap integrals (S).

SE Mechanism[66]

The smallest nonzero number of electron exchanges in $\hat{A}_{ABC}$ is equal to 1. Such an exchange may only take place between two molecules, say, AB.[67] This results in terms proportional to S^2 in the three-body expression, where S means the overlap integral between the molecular orbitals of A and B. The third molecule *does not participate in the electron exchanges, but is not just a spectator in the interaction*. If it were, the interaction would not be three-body (Fig. 5.12).

> SE MECHANISM
>
> Molecule C interacts electrostatically with the *Pauli deformation of molecules A and B (i.e., with the multipoles that represent the deformation)*. Such a mixed interaction is called the SE mechanism.

64 The operator reproduces the polarization approximation expressions in SAPT.

65
- First, we write down the exact expression for the first-order exchange nonadditivity.
- Then we expand the expression in the Taylor series with respect to those terms that arise from all electron exchanges except the identity permutation.
- Next, we see that the exchange nonadditivity expression contains terms of the order of S^2 and higher, where S stands for the overlap integrals between the orbitals of the interacting molecules.

66 *Single exchange.*

67 After that we have to consider AC and BC.

When the double electron exchanges are switched on, we would obtain part of the three-body effect of the order of S^4. Since S is usually of the order of 10^{-2}, this contribution is expected to be small, although caution is advised, because the *number* of such terms is much more important.

TE Mechanism[68]

Is there any contribution of the order of S^3? Yes. The antisymmetrizer $\hat{A}_{ABC}$ is able to make the single electron exchange between, e.g., A and B, but by mediation of C. The situation is schematically depicted in Fig. 5.12.

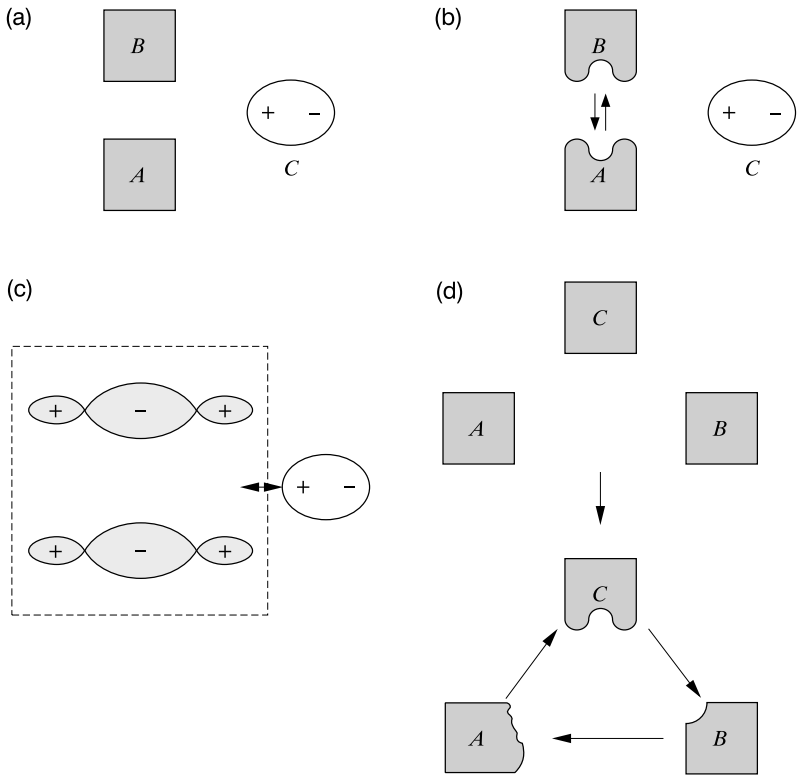


Fig. 5.12. A scheme of the SE and TE exchange nonadditivities. (a)–(c) The single exchange (SE) mechanism. (a) Three noninteracting molecules (schematic representation of electron densities). (b) Pauli deformation of molecules A and B. (c) Electrostatic interaction of the Pauli deformation resulting from single electron exchanges between A and B with the dipole moment of C. (d) The triple exchange (TE) mechanism: molecules A and B exchange an electron with the mediation of molecule C. All molecular electron density distributions undergo Pauli deformation.

[68] *Triple exchange.*

TE MECHANISM
A molecule is involved in a single exchange with another molecule by mediation of a third one.

Short-rangeness

The Pauli deformation has a local (i.e., short-range) character (see Appendix E). Let us imagine that molecule B is very long and the configuration corresponds to $A\ B\ C$. When C is far from A, the three-body effect is extremely small, because almost everything in the interaction is of the two-body character. If molecule C approaches A and has some nonzero low-order multipoles, e.g., a charge, then it may interact by the SE mechanism even from afar. Both mechanisms (SE and TE) operate only at short $C\,A$ distances.

The exchange interaction is nonadditive, but the effects pertain to the contact region of both molecules. The Pauli exclusion principle does not have any finite range in space, i.e., after being introduced it has serious implications for the wave function even at infinite intermolecular distance (cf. p. 379). Despite this, it always leads to the differential overlap of atomic orbitals (as in overlap or exchange integrals), which decays exponentially with increasing intermolecular distance. This may not be true for the SE mechanism, which has a partly long-range character, and some caution is needed when considering its spatial range.

5.11.6 Induction nonadditivity

The nonadditivity of the intermolecular interaction results mainly from the nonadditivity of the induction contribution.

How do we convince ourselves about the nonadditivity? This is very easy. It will be sufficient to write the expression for the induction energy for the case of three molecules and to see whether it gives the sum of the pairwise induction interactions. Before we do this, let us write the formula for the total second-order energy correction (similar to the case of two molecules on p. 354). We have

$$E^{(2)}(ABC) = {\sum_{n_A,n_B,n_C}}' \frac{|\langle n_A n_B n_C|\, V\, |0_A 0_B 0_C\rangle|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)] + [E_C(0_C) - E_C(n_C)]}. \tag{5.73}$$

According to perturbation theory, the term with all the indices equal to zero has to be omitted in the above expression (symbolized by $\sum'$). It is much better like this, because otherwise

the denominator would be zero. Further, the terms with all nonzero indices are equal to zero. Indeed, let us recall that V is the sum of the Coulomb potentials corresponding to all three *pairs* of the three molecules. This is the reason why it is easy to perform the integration over the electron coordinates of the third molecule (not involved in the pair). A similar operation has been performed for the electrostatic interaction. This time, however, the integration makes the term equal to zero, because of the orthogonality of the ground and excited states of the third molecule. All this leads to the conclusion that to have a nonzero term in the summation, among the three indices *it has to be one index or two indices of zero value*. Let us perform the summation in two stages: all the terms with only two zero indices (or a single nonzero index) will make a contribution to $E_{\text{ind}}(ABC)$, while all the terms with only one zero index (or two nonzero indices) will sum to $E_{\text{disp}}(ABC)$, i.e.,

$$E^{(2)}(ABC) = E_{\text{ind}}(ABC) + E_{\text{disp}}(ABC), \tag{5.74}$$

where the first term represents the *induction energy*

$$E_{\text{ind}}(ABC) = E_{\text{ind}}(AB \to C) + E_{\text{ind}}(AC \to B) + E_{\text{ind}}(BC \to A),$$

where

$E_{\text{ind}}(BC \to A) \equiv \sum_{n_A \neq 0} \frac{|\langle n_A 0_B 0_C|V|0_A 0_B 0_C\rangle|^2}{E_A(0_A) - E_A(n_A)}$ means that the "frozen" molecules B and C *acting together* polarize molecule A, etc. The second term in (5.74) represents the *dispersion energy* (this will be considered later on).

For the time being let us consider the induction energy $E_{\text{ind}}(ABC)$. Writing V as the sum of the Coulomb interactions of the pairs of molecules we have

$$E_{\text{ind}}(BC \to A) = \sum_{n_A \neq 0} \frac{|\langle n_A 0_B 0_C|V_{AB} + V_{BC} + V_{AC}|0_A 0_B 0_C\rangle|^2}{E_A(0_A) - E_A(n_A)} =$$
$$\sum_{n_A \neq 0} \frac{|\langle n_A 0_B|V_{AB}|0_A 0_B\rangle + \langle n_A 0_C|V_{AC}|0_A 0_C\rangle|^2}{E_A(0_A) - E_A(n_A)}.$$

Look at the square in the nominator. The induction nonadditivity arises just because of this. If the square were equal to the sum of squares of the two components, the total expression shown explicitly would be equal to the induction energy corresponding to the polarization of A by the frozen charge distribution of B plus a similar term corresponding to the polarization of A by C, i.e., the polarization occurring *separately*. Together with the other terms in $E_{\text{ind}}(AB \to C) + E_{\text{ind}}(AC \to B)$ *we would obtain the additivity* of the induction energy $E_{\text{ind}}(ABC)$. However, besides the sum of squares we also have the mixed terms. They will produce the nonadditivity of the induction energy, i.e.,

$$E_{\text{ind}}(ABC) = E_{\text{ind}}(AB) + E_{\text{ind}}(BC) + E_{\text{ind}}(AC) + \Delta_{\text{ind}}(ABC)\,. \tag{5.75}$$

Thus, we obtain the following expression for the *induction nonadditivity* $\Delta_{\text{ind}}(ABC)$:

$$\Delta_{\text{ind}}(ABC) = 2\text{Re}\sum_{n_A\neq 0}\frac{\langle n_A 0_B|\, V_{AB}\,|0_A 0_B\rangle\,\langle n_A 0_C|\, V_{AC}\,|0_A 0_C\rangle}{[E_A(0_A) - E_A(n_A)]} + ..., \tag{5.76}$$

where the term displayed explicitly is the nonadditivity of $E_{\text{ind}}(BC \to A)$ and "+..." stands for the nonadditivities of $E_{\text{ind}}(AB \to C) + E_{\text{ind}}(AC \to B)$.

Example 7 (Induction nonadditivity). To show that the induction interaction of two molecules depends on the presence of the third molecule, let us consider the system shown in Fig. 5.13.

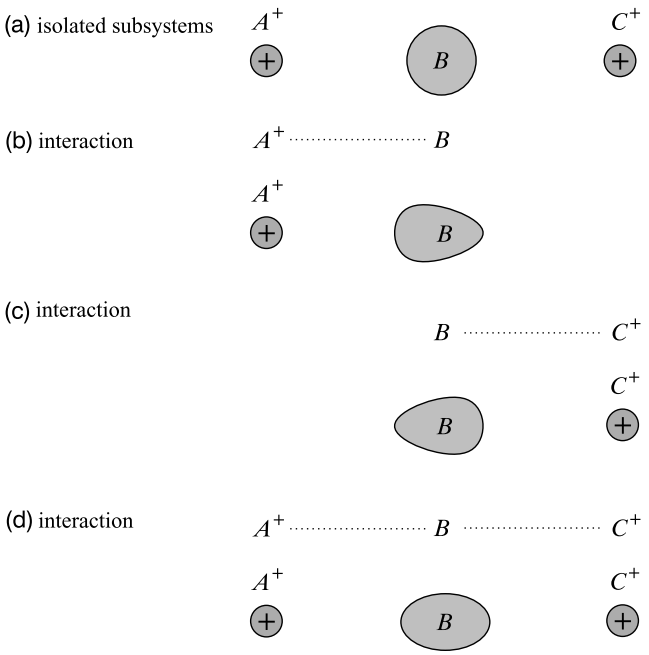


Fig. 5.13. The induction interaction may produce a large nonadditivity. (a) Two distant nonpolarizable cations, A^+, C^+, and a small, polarizable neutral molecule B placed exactly in the middle between A and C. (b) The two-body induction interaction A^+B, a strong polarization. (c) The two-body induction interaction BC^+, a strong polarization. (d) The two cations polarize molecule B. Their electric field vectors cancel each other out in the middle of B and give a small electric field intensity within B (a weak polarization).

Let molecule B be placed half-way between A^+ and C^+, so the configuration of the system is $A^+ \cdots\cdots B \cdots\cdots C^+$, with long distances between the subsystems. In such a situation, the total interaction energy is practically represented by the induction contribution plus the constant electrostatic repulsion of A^+ and C^+. Is the three-body term (induction nonadditivity) large? We will show in a minute that this term is large and positive (destabilizing). Since the electric field intensities nearly cancel out within molecule B, despite the high polarizability of the latter, the induction energy will be small. On the contrary, the opposite is true when considering two-body interaction energies. Indeed, A^+ polarizes B very strongly, C^+ does the same, resulting in high stabilization due to high two-body induction energy. Since the total effect is nearly zero, the induction nonadditivity is bound to be a large positive number.[69]

Self-consistency and polarization catastrophe

The second-order induction effects pertain to polarization by the charge distributions corresponding to the isolated molecules. However, the induced multipoles introduce a change in the electric field and in this way contribute to further changes in charge distribution. These effects already belong to the third[70] and higher orders of perturbation theory.

It is therefore evident that a two-body interaction model cannot manage the induction interaction energy. This is because we have to ensure that any subsystem, e.g., A, has to experience polarization in an electric field which is the *vector sum* of the electric fields from all its partner subsystems (B, C, ...) and calculated at the position of A. The calculated induced dipole moment of A (we focus on the lowest multipole) creates the electric field that produces some changes in the dipole moments of B, C, ..., which in turn change the electric field acting on all the molecules, including A. The circle must close and the polarization procedure has to be performed until self-consistency is reached. This can often be done, although such a simplified interaction model does not allow for geometry optimization, which may lead to a *polarization catastrophe* ending up with induction energy equal to $-\infty$ (due to an excessive approach and the lack of the Pauli blockade described on p. 392).

5.11.7 Additivity of the second-order dispersion energy

The dispersion energy is a second-order correction; Eq. (5.13) gives the formula for interaction of two molecules. For *three* molecules we obtain the following formula for the dispersion part

[69] If the intermolecular distances were small, B would not be in the middle of AC or molecule B would be of large spatial dimension, and the strength of our conclusion would diminish.

[70] Each of the induced multipoles is proportional to V, their interaction introduces another V; altogether this gives a term proportional to VVV, i.e., indeed of the third order.

of the second-order effect (see the discussion on p. 407):

$$E_{\text{disp}}(ABC) = \sum_{(n_A,n_B)\neq(0_A,0_B)} \frac{|\langle n_A n_B 0_C|\, V_{AB} + V_{BC} + V_{AC}\,|0_A 0_B 0_C\rangle|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)]} + \ldots,$$

where $+\ldots$ denotes analogous terms with summations over n_A, n_C as well as n_B, n_C. Among three integrals in the nominator only the first one will survive, since the others vanish due to the integration over the coordinates of the electrons of molecule Z not involved in the interaction V_{XY}; we have

$$E_{\text{disp}}(ABC) = \sum_{(n_A,n_B)\neq(0_A,0_B)} \frac{|\langle n_A n_B 0_C|\, V_{AB}\,|0_A 0_B 0_C\rangle + 0 + 0|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)]} + \sum_{(n_A,n_C)\neq(0_A,0_C)} \ldots + \sum_{(n_B,n_C)\neq(0_B,0_C)} \ldots.$$

In the first term we can integrate over the coordinates of C. Then the first term displayed in the above formula turns out to be the dispersion interaction of A and B,

$$E_{\text{disp}}(ABC) = \sum_{(n_A,n_B)\neq(0,0)} \frac{|\langle n_A n_B|\, V_{AB}\,|0_A 0_B\rangle|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)]} + \sum_{(n_A,n_C)\neq(0_A,0_C)} \ldots + \sum_{(n_B,n_C)\neq(0_B,0_C)} \ldots = E_{\text{disp}}(AB) + E_{\text{disp}}(AC) + E_{\text{disp}}(BC).$$

Thus, we have proved that

> the dispersion interaction (second order of the perturbation theory) *is additive*, i.e.,
>
> $$E_{\text{disp}}(ABC) = E_{\text{disp}}(AB) + E_{\text{disp}}(AC) + E_{\text{disp}}(BC).$$

5.11.8 Nonadditivity of the third-order dispersion interaction

One of the third-order energy terms represents a correction to the dispersion energy. The correction as shown by Axilrod and Teller[71] has a three-body character. The part connected to the interaction of three distant instantaneous dipoles on A, B, and C reads as

$$E^{(3)}_{\text{disp}} = 3C^{(3)}_{ddd} \frac{1 + 3\cos\theta_A \cos\theta_B \cos\theta_C}{R^3_{AB} R^3_{AC} R^3_{BC}}, \tag{5.77}$$

[71] B.M. Axilrod, E. Teller, *J. Chem. Phys.*, 11(1943)299.

where R_{XY} and θ_X denote the sides and the angles of the ABC triangle and $C^{(3)}_{ddd} > 0$ represents a constant. The formula shows that

> when the ABC system is in a linear configuration, the dispersion contribution is negative, i.e., *stabilizing*, while the equilateral triangle configuration corresponds to a *destabilization*.

ENGINEERING OF INTERMOLECULAR INTERACTIONS

5.12 Idea of molecular surface

5.12.1 van der Waals atomic radii

It would be of practical importance to know how close two molecules can approach each other. We will not enter this question too seriously, because this problem cannot have an elegant solution: it depends on the direction of approach, and the atoms involved, as well as how strongly the two molecules collide. Searching for the effective radii of atoms would be nonsense if the valence repulsion were not a sort of "soft wall" or if the atom sizes were very sensitive to molecular details. Fortunately, it turns out that an atom, despite different roles played in molecules, can be characterized by its approximate radius, called the *van der Waals radius*. The radius may be determined in a naive, but quite effective, way. For example, we may make two HF molecules approach axially with the fluorine atoms head-on and then find the distance[72] R_{FF} at which the interaction energy is equal to, say, +5 kcal/mol (repulsion). The proposed fluorine atom radius would be $r_F = \frac{R_{FF}}{2}$. A similar procedure may be repeated with two HCl molecules with the resulting r_{Cl}. Now, let us consider an axial complex H–F$\cdots$Cl–H with the intermolecular distance also corresponding to +5 kcal/mol. What F$\cdots$Cl distance are we expecting? Of course, something close to $r_F + r_{Cl}$. It turns out that we are about right. This is why the atomic van der Waals radius concept is so attractive from the practical point of view.

5.12.2 A concept of molecular surface

We may define a superposition of atomic van der Waals spheres. This defines what is called the van der Waals surface of the molecule,[73] or a molecular shape – a concept of great importance

[72] Using a reliable quantum mechanical method.

[73] The van der Waals surface of a molecule may sometimes be very complex, e.g., a molecule may have two or more surfaces (like fullerenes).

and of the same arbitrariness (as the radii themselves). In principle three factors decide about what happens in chemistry: the molecular shape (surface), its chemical characteristics, and the solvent. The first and third features result in at least two possible spatial shapes of molecules (as shown in Fig. 5.14): the solvent-excluded surface (SES) and the solvent-accessible surface (SAS). Despite the fact that any notion of molecular surface is to a large extent arbitrary, the molecular surface remains one of the most important factors.

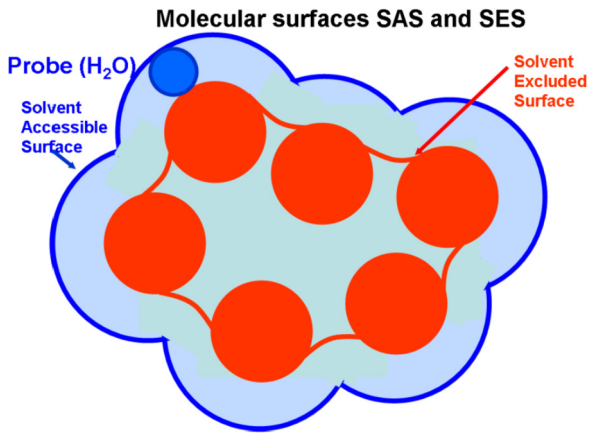


Fig. 5.14. After achieving a consensus about a van der Waals contact distance, one is able to construct the solvent-excluded surface (SES) and the solvent-accessible surface (SAS). A solvent (probe) molecule taken as a sphere is rolling on the SES, thus defining the SAS. Both surfaces define the all important overall shape of the molecule, which decides about the interaction with other molecules.

In a similar way we may define *ionic radii*,[74] to reproduce the ion packing in ionic crystals, as well as *covalent radii*, to foresee chemical bond lengths.

5.12.3 Confining molecular space – the nanovessels

Molecules at long distances interact through the mediation of the electric fields created by them. The valence repulsion is of a different character, since it results from the Pauli exclusion principle, and may be interpreted as an energy penalty for an attempt by electrons of the same spin coordinate to occupy the same space (cf. Chapter V1-1 and p. 105).

Luty and Eckhardt[75] have highlighted the role of pushing one molecule by another. Let us imagine an atomic probe, e.g., a helium atom. The pushing by the probe deforms the molec-

[74] This concept was introduced by Pauling, based on crystallographic data (L. Pauling, *J. Amer. Chem. Soc.*, 49(1927)765).

[75] T. Luty, C.J. Eckhardt, in "*Reactivity of Molecular Solids*," eds. E. Boldyreva, V. Boldyrev, Wiley, 1999, p. 51.

ular electronic wave function (Pauli deformation), but motion of the electrons is accompanied by the motion of the nuclei. Both motions may lead to dramatic events. For example, we may wonder how an explosive reaction takes place. The spire hitting the material in its metastable chemical state is similar to the helium atom probe pushing a molecule. Due to the pushing the molecule distorts to such an extent that the HOMO-LUMO separation vanishes and the system rolls down (see Chapter 6) to a deep potential energy minimum on the corresponding potential energy hypersurface. Closing of the HOMO-LUMO gap takes place within the reaction barrier. Since the total energy is conserved, the large reaction net energy gain goes to highly excited vibrational states (in the classical approximation corresponding to large amplitude vibrations). The amplitude may be sufficiently large to ensure the pushing of the next molecules in the neighborhood and a chain reaction starts with exponential growth.

Now imagine a lot of atomic probes confining the space (like a cage or template) available to a molecule under study. In such a case the molecule will behave differently from a free one. For example,

- a protein molecule, when confined, *will fold to another conformation*;
- some photochemical reactions that require a space for the rearrangement of molecular fragments *will not occur* if the space is not accessible;
- in a restricted space some other chemical reactions will take place (new chemistry – *chemistry in "nanovessels"*);
- some unstable molecules may *become stable* when enclosed in a nanovessel.

These are fascinating and little explored topics.

5.12.4 Molecular surface under high pressure

Chemical reactions in nanovessels, which may run in another way than without them, suggest that the atomic structure of materials may change when under increasing pressure. Suppose we have to do with a molecular crystal and we are studying what happens after applying isotropic pressure.

The transformations after increasing pressure may be strange, but quite a lot can be said using the following rules of thumb:

- The weaker the interatomic bond (chemical or of the van der Waals character), the larger the change of its length. As a manifestation of this, the pressure after passing a threshold value may change the packing of the molecules, leaving the molecules themselves virtually unchanged (*displacive phase transitions*).

- At higher pressure, the chemical bonds, often of the mixed character (ionic and covalent [polarized bonds]), become more and more covalent. This is because the overlap integrals increase very much due to shorter interatomic distances (exponential increase) and as a consequence the covalent bond character becomes stronger (cf. p. V1-511).
- For even larger pressures the atomic coordination numbers begin to increase. This corresponds to the formation of new chemical bonds, which make the atoms to be closer in space. In order to make the new bonds, the atoms, when squeezed, open their closed core electronic shells (e.g., under sufficiently high pressures even the noble gas atoms open their electronic shells, cf. Chapter 6). Some new phase transitions are accompanying these changes, known as *reconstructive phase transitions*.
- At even higher pressures the crystal structure becomes of the close-packed type, which means a further minimization of the crystal's empty spaces.
- Empirical observation: the atoms when under high pressure behave similarly as the atoms belonging to the same group and the next period of the Mendeleev table do at lower pressures. For example, the silicon under high pressure behaves as germanium under lower pressure, etc.

5.13 Decisive forces

5.13.1 Distinguished role of the valence repulsion and electrostatic interaction

- The valence repulsion plays the role of a soft wall that forbids the closed shell molecules to approach too closely. This represents a very important factor, since those molecules that do not fit together receive an energy penalty.
- The electrostatic contribution plays a prominent role in the intermolecular interaction, because of its long-range character and the all-important dependence on the mutual orientation of the molecules.

The induction and dispersion contributions, even if sometimes larger than the electrostatic interaction, usually play a less important role. This is because only the electrostatics may change the sign of the energy contribution when the molecules reorient, thus playing the pivotal role in the interaction energy.

The induction and dispersion contributions are negative (nearly independent of the mutual orientation of the molecules), and we may say, as a rule of thumb, that their role is to stabilize the structure already determined by the valence repulsion and the electrostatics.

5.13.2 Hydrogen bond

Among the electrostatic interactions, the most important are those having an especially strong dependence on orientation. The most representative of them are the *hydrogen bonds (also known as hydrogen bridges) X–H$\cdots$Y, where an electronegative atom X plays the role of a proton donor, while an electronegative atom Y plays the role of a proton acceptor (Fig. 5.15b). Most often the hydrogen bond X–H$\cdots$Y deviates only a little from linearity.* Additionally, the XY separation usually falls into a narrow range: 2.7–3.1 Å, i.e., it is to a large extent independent of X and Y. The hydrogen bond features are also unique, because of the extraordinary properties of the hydrogen atom in the hydrogen bridge. This is the only atom which occasionally may attain the partial charge equal to 0.45 a.u., which means it represents a nucleus devoid to a large extent its electron density. This is one of the reasons why the hydrogen bond is strong when compared with other types of intermolecular interactions.

Example 8 (Hydrogen bond in the water dimer). This is the most important hydrogen bond on Earth, because water and hydrogen bonds in water mean Life. Moreover, since it may be thought as being representative for other hydrogen bonds, we treat the present example as highlighting the main consequences of formation of a generic hydrogen bond. We will perform the quantum mechanical calculations (of medium quality) within the Born–Oppenheimer approximation, including the electron correlation and optimizing the system geometry.[76]

In Fig. 5.15a the optimized geometry of the isolated water molecule is shown ($r_{\mathrm{OH}} = 0.959$ Å, $\measuredangle\mathrm{HOH} = 103.4°$) as well as the net Mulliken charges on the individual atoms[77] ($Q_{\mathrm{O}} = -0.511$, $Q_{\mathrm{H}} = 0.256$). All these numbers change when two water molecules form a hydrogen bond (Fig. 5.15b). We see the following changes:

- The complex has the $C_{2\mathrm{v}}$ symmetry with the proton donor molecule and the oxygen atom of the proton acceptor molecule within the symmetry plane (the book's page).
- The optimized oxygen–oxygen distance turns out to be $R_{\mathrm{OO}} = 2.91$ Å, the hydrogen bond is *almost linear* ($\measuredangle\mathrm{O–H}\cdots\mathrm{O} = 176.7°$), and the optimized O–H bond length (of the hydrogen bridge) is $r_{\mathrm{OH},donor} = 0.965$, with a tiny upwards deviation of the proton from the linearity of the O–H$\cdots$O bridge (the effect of repulsion with the protons of the acceptor). *This quite common quasilinearity suggests the electrostatics is important in forming hydrogen bonds*. If it were true, we should see this also from the relative position of the molecular dipoles. At first sight we might be disappointed though, because the dipole moments are almost perpendicular instead of being colinear. However, the two dipole moments deviate in the

[76] The Møller–Plesset method accurate up to the second order is used (the MP2 method described in Chapter 2), the basis set is 6-311G(d,p). This ensures results of a decent quality.

[77] See Appendix D on p. 595. The charges are calculated within the Hartree–Fock method that precedes the MP2 perturbational calculation.

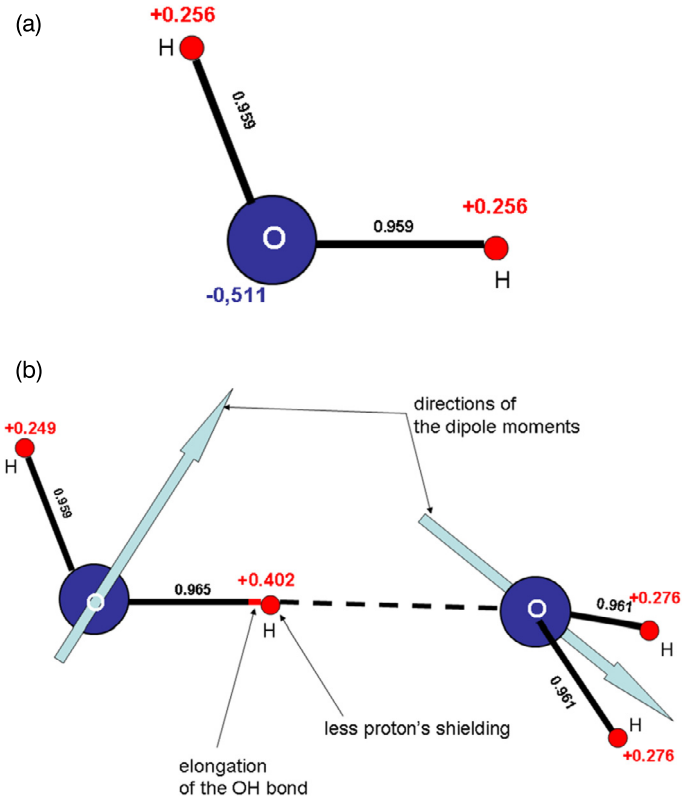


Fig. 5.15. The hydrogen bond in the water dimer described by the MP2 method. (a) The optimal configuration of an isolated water molecule. (b) The optimal configuration of the complex of two water molecules. The numbers displayed at the nuclei represent the atomic Mulliken net charges (in a.u.), those at the bonds represent their lengths in Å. Conclusion: the hydrogen bridge O–H···O is almost linear and the bridge hydrogen atom is devoid of some 40% of its electronic cloud (exceptional case in chemistry).

opposite directions from the colinear configuration (similarly as for the HF···HF complex) with the deviation angle about ±45° (this means that they are nearly orthogonal). As seen from Fig. 5.7 on p. 364, despite such a large deviation, the electrostatics is still favorable.[78]

[78] Let us calculate from Eq. (5.23) on p. 363. We have $E_{dip-dip} = \frac{\mu_A^2}{R^3}(\sin\theta_A \sin\theta_B - 2\cos\theta_A \cos\theta_B) = \frac{\mu_A^2}{R^3}\left(\frac{1}{\sqrt{2}}(-\frac{1}{\sqrt{2}}) - 2\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\right) = \frac{\mu_A^2}{R^3}\left(-\frac{3}{2}\right)$, as compared with the optimal value $-2\frac{\mu_A^2}{R^3}$. The loss of the dipole–dipole electrostatics is of the order of 25%. The energy loss must be more than compensated by other attractive interactions.

- We do not see any significant charge transfer between the molecules. Indeed, the Mulliken population analysis shows only a tiny $0.02e$ electron transfer from the proton donor to the proton acceptor.
- The donor molecule is perturbed more than the acceptor one. This is seen first of all from the Mulliken net charges on hydrogen atoms and from the deformation of geometry (see Fig. 5.15). One sees *a remarkable positive charge on the bridge hydrogen atom* $Q_{\mathrm{H},donor} = 0.402$; no other nucleus in the whole chemistry is so poor in electrons. This hydrogen bond effect is confirmed in the NMR recordings (low shielding constant).
- The calculated binding energy[79] is $E_{bind} = E^{MP2} - 2E^{MP2}_{\mathrm{H_2O}} = -6.0\frac{\mathrm{kcal}}{\mathrm{mol}}$. *This order of magnitude of the binding energy, about* 20 *times smaller than the energy of chemical bonds, is typical for hydrogen bridges.*

5.13.3 Coordination interaction

Coordination interactions appear if an electronic pair of one subsystem (electron donor) lowers its energy by interacting[80] with an electron acceptor offering an empty orbital, e.g., a cation (acceptor) interacts with an atom or atoms (donor) offering lone electronic pairs. This may be also seen as a special kind of electrostatic interaction.[81] Fig. 5.16 shows a ring-shaped molecule (crown ether) offering lone pairs for the interaction with a cation.

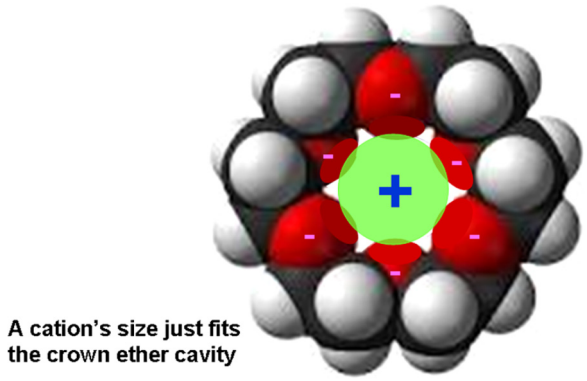


Fig. 5.16. A cation fits the hole of a ring molecule (crown ether). The hole surface has partial negative charges, a result of the presence of oxygen atoms.

79 No BSSE correction included.
80 Forming a molecular orbital.
81 A lone pair has a large dipole moment, which interacts with the positive charge of the acceptor.

When concentrating on the ligands we can see that in principle they represent a negatively charged cavity (lone pairs) waiting for a monoatomic cation with *dimensions of a certain range only*. The interaction of such a cation with the ligand would be exceptionally large and therefore "specific" for such a pair of interacting moieties, which is related to the selectivity of the interaction.

Let us consider a water solution containing Li^+, Na^+, K^+, Rb^+, and Cs^+ ions. After adding the abovementioned crown ether and after the equilibrium state is attained, only the cation of a special size that fits the hole perfectly will bind with the crown ether. For the other ions the equilibrium will be shifted towards their association with water molecules, not the cryptand. This is remarkable information. The crown ethers represent an example of generic molecular recognition (Fig. 5.16).

We are able to selectively extract objects of some particular shape and dimensions.

5.13.4 Electrostatic character of molecular surface – the maps of the molecular potential

Chemical reaction dynamics is possible only for very simple systems. Chemists, however, have most often to do with medium-size or large molecules. Would it be possible to tell anything about the barriers for chemical reactions in such systems? Most chemical reactions start from a situation when the molecules are far away, but already interact. The main contribution is the electrostatic interaction energy, which is of the most long-range character. Electrostatic interaction depends strongly on the mutual orientation of the two molecules (*steric effect*). Therefore, the orientations are biased towards the privileged ones (energetically favorable). There is quite a lot of experimental data suggesting that privileged orientations lead, at smaller distances, to low reaction barriers. There is no guarantee of this, but it often happens for electrophilic and nucleophilic reactions, because the attacking molecule prefers those parts of the partner that correspond to high electron density (for electrophilic attack) or low electron density (for nucleophilic attack).

We may use an electrostatic probe (e.g., a unit positive charge) to detect which parts of the molecule "like" the approaching charge (energy lowering) and which do not (energy increasing).

The electrostatic interaction energy of the point-like unit charge (probe) in position $\mathbf{r}$ with molecule A is described by the following formula (the definition of the electrostatic potential produced by molecule A):

$$V_A(\mathbf{r}) = +\sum_a \frac{Z_a}{|\mathbf{r}_a - \mathbf{r}|} - \int \frac{\rho_A(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \mathrm{d}\mathbf{r}', \tag{5.78}$$

where the first term describes the interaction of the probe with the nuclei denoted by index a, and the second term means the interaction of the probe with the electron density distribution of molecule A denoted by ρ_A (according to Chapter 3).

In the Hartree–Fock or Kohn–Sham approximation (Chapter 3, p. 196; we assume the n_i-tuple occupation of the molecular orbital $\varphi_{A,i}$, $n_i = 0, 1, 2$),

$$\rho_A(\mathbf{r}) = \sum_i n_i |\varphi_{A,i}(\mathbf{r})|^2. \tag{5.79}$$

These molecular orbitals $\varphi_{A,i}$ are calculated in the literature for the isolated molecule, but one can imagine them to be computed *in the presence of the probe charge*. The first possibility has one pro: we characterize the isolated molecule under study; however, we do not see how the charge flows in our molecule under the influence of the attacking object and this should count during the attack. Taking the second possibility removes this deficiency, but only in part since the attacking object usually does not have a unit charge on it.

Therefore, in order to obtain $V_A(\mathbf{r})$ at point $\mathbf{r}$ it is sufficient to calculate the distances of the point from any of the nuclei (trivial) as well as the one-electron integrals, which appear after inserting into (5.78) $\rho_A(\mathbf{r}') = 2\sum_i |\varphi_{A,i}(\mathbf{r}')|^2$. Within the LCAO MO approximation the electron density distribution ρ_A represents the sum of products of two atomic orbitals (in general centered at two different points). As a result the task reduces to calculating typical one-electron three-center integrals of the nuclear attraction type (cf. Chapter V1-8 and Appendix V1-Q), because the third center corresponds to the point $\mathbf{r}$ (Fig. 5.17). There is no computational problem with this for contemporary quantum chemistry.

In order to represent $V_A(\mathbf{r})$ graphically we usually choose to show an equipotential surface corresponding to a given absolute value of the potential, while additionally indicating its sign (Fig. 5.17). The sign tells us which parts of the molecule are preferred for the probe-like object to attack and which not. In this way we obtain basic information about the reactivity of different parts of the molecule.[82]

[82] Having the potential calculated according to (5.78) we may ask about the set of atomic charges that reproduce it. Such charges are known as electrostatic potential charges (ESP).

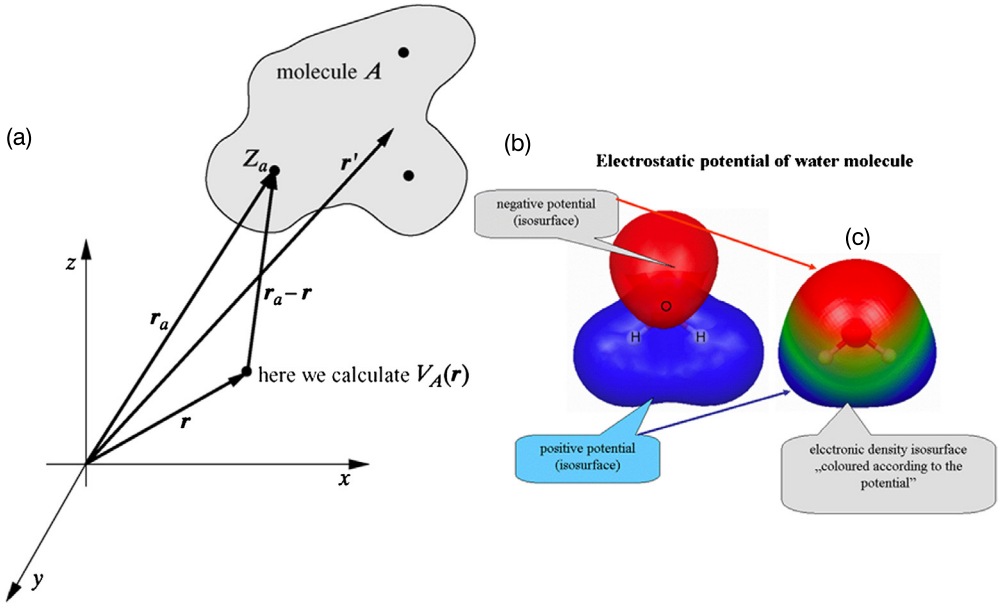


Fig. 5.17. The molecular electrostatic potential (MEP) represents the electrostatic interaction of the positive unit charge (probe) with molecule A. (a) The coordinate system and the vectors used in Eq. (5.78). (b) The equipotential surfaces $|V_A(\mathbf{r})|$ for the water molecule. (c) Another way of showing the MEP: one computes $V_A(\mathbf{r})$ on the molecular surface (defined somehow). MEP in a.u. means the interaction of a proton with the molecule. In more expensive books this is shown by coloring the surface using a certain convention: color↔MEP. Such information is more useful, because the role of the MEP is to predict the site of attack of another molecule, which is able to approach the surface. It is seen that the proton will attack the region of the oxygen atom, while an attack of Cl^- would happen from the side of the hydrogens.

5.13.5 Hydrophobic effect

This is quite a peculiar type of interaction, which appears mainly (not only) in water solutions.[83] The hydrophobic interaction does not represent any particular new interaction (beyond those we have already considered), because at least potentially they could be explained by the electrostatic, induction, dispersion, valence repulsion, and other interactions already discussed.

The problem, however, may be seen from a different point of view. The basic interactions have been derived as if operating in vacuum. However, in a medium the molecules interact with one another through the mediation of other molecules, including those of the solvent. In particular,

83 W. Kauzmann, *Advan. Protein Chem.*, 14(1959)1. A contemporary theory is given in K. Lum, D. Chandler, J.D. Weeks, *J. Phys. Chem.*, 103(1999)4570.

a water medium creates an elastic network of hydrogen bonds[84] that surround the hydrophobic moieties, trying to expel them from the solvent and, therefore, pushing them together (i.e., minimizing the hole in the hydrogen bond network), which has the effect of imitating their mutual attraction, resulting in the formation of a sort of "oil drop."

> We may say in a rather simplistic way that hydrophobic molecules aggregate not because they attract particularly strongly, but because water strongly prefers them to be out of its hydrogen bond net structure.

Hydrophobic interactions have a complex character and are not yet fully understood. They have an important entropic component (Fig. 5.18a).

The interaction depends strongly on the size of the matching hydrophobic moieties. For small sizes, e.g., such as two methane molecules in water, the hydrophobic interaction is small; it increases considerably for larger moieties (Fig. 5.18). The hydrophobic effects become especially structure determining in what is called *amphiphilic* macromolecules, with their van der Waals surfaces differing in hydrophobic character (hydrophobic/hydrophilic). Amphiphilic molecules (Fig. 5.19) are able to self-organize, forming structures up to the nanometer scale ("nanostructures").

> When the hydrophobic moieties in water are free to move, they usually form a separate phase, like, e.g., the oil phase above the water phase or an oil film on the surface of water. If this freedom is limited by other interactions, like those in amphiphilic molecules, a compromise is achieved in which the hydrophobic moieties keep together, while the hydrophilic ones are exposed to water (e.g., lipid bilayer or micelle). Finally, when the positions of the hydrophobic amino acids are constrained by the polypeptide backbone, they have the tendency to be buried as close to the center of the globular protein as possible, while the hydrophilic amino acid residues are exposed to the surrounding water.

Who attacks whom?

In chemistry a probe will not be a point charge, but rather a neutral molecule or an ion. Nevertheless our new tool (electrostatic potential) will still be useful:

---

84 M.N. Rodnikova, *J. Phys. Chem. (Russ.)*, 67(1993)275.

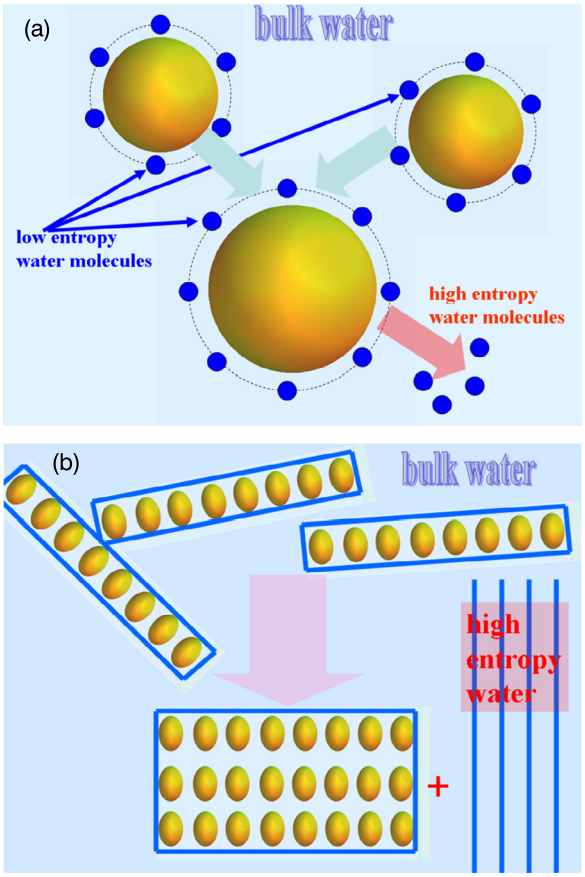


Fig. 5.18. The hydrophobic effect has an important entropic stabilization. (a) Two oil drops, when unified, leave some high-entropy making water molecules. (b) This effect gets much stronger for contacting hydrophobic surfaces.

- If the probe represents a cation, it will attack those parts of molecule A which are electron-rich (electrophilic reaction).
- If the probe represents an anion, it will attack the electron-deficient parts (nucleophilic reaction).
- If the probe represents a neutral molecule (B) with partial charges on its atoms, *its* electrostatic potential V_B is the most interesting. Those AB configurations that correspond to the contacts of the associated sections of V_A and V_B with the opposite signs are the most (electrostatically) stable.

The site of the molecular probe (B) which attacks an electron-rich site of A itself has to be of electron-deficient character (and *vice versa*). Therefore, from the point of view of the attacked

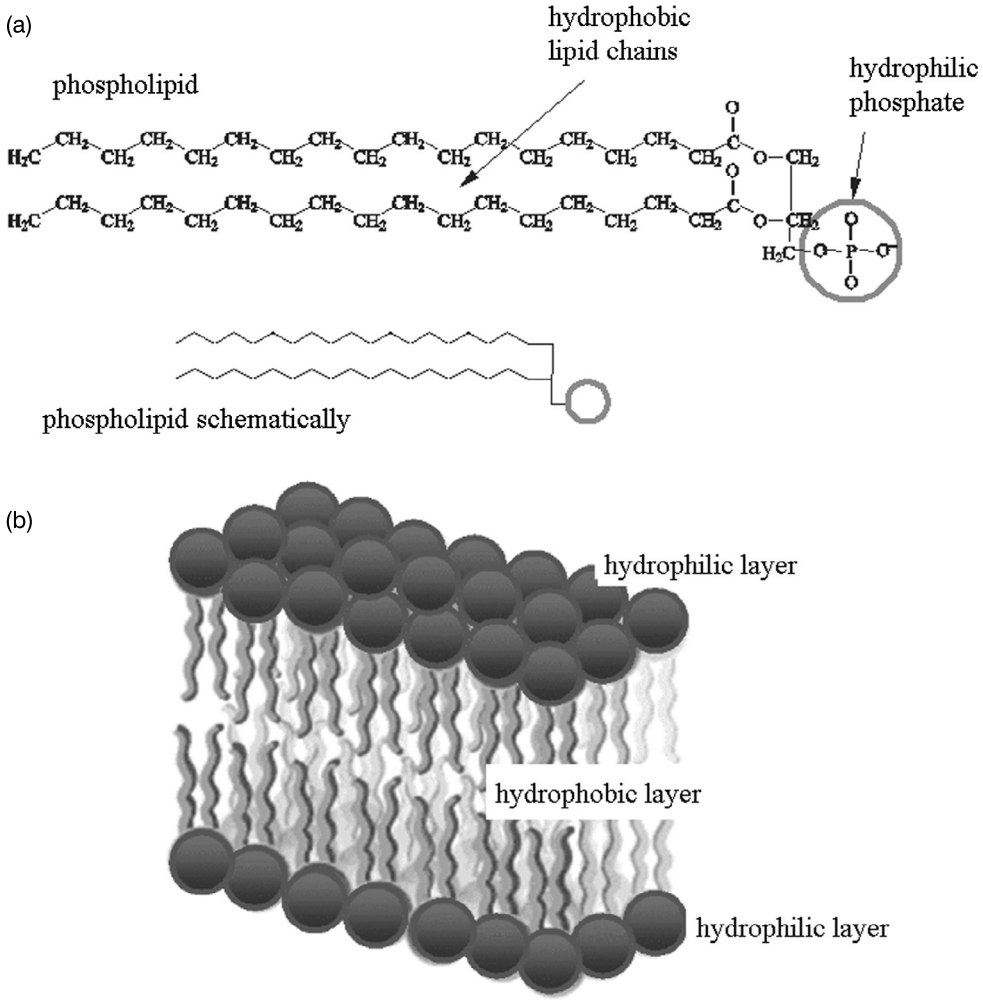


Fig. 5.19. An amphiphilic molecule, i.e., two contradictions (hydrophobicity and hydrophilicity) side by side and their consequences in water. (a) A phospholipid with two hydrophobic aliphatic "tails" and a hydrophilic phosphate "head" (its chemical structure given), shown schematically. (b) The hydrophobic effect in water leads to formation of a lipid bilayer structure, while the hydrophilic heads are exposed towards the bulk water (due to a strong hydration effect). The lipid bilayer plays in biology the role of the cell membrane.

molecule (*A*), everything looks "upside down": an electrophilic reaction becomes nucleophilic and *vice versa*. When two objects exhibit an affinity to each other, who attacks whom represents a delicate and ambiguous problem, and let it be that way. Therefore, where does such nomenclature in chemistry come from? Well, it comes from the concept of didactics...

5.14 Construction principles

5.14.1 Molecular recognition – synthons

Organic molecules often have quite a few donor and acceptor substituents. These names may pertain to donating/accepting electrons or protons (cf. the charge conjugation described on p. 367). Sometimes a particular side of a molecule displays a system of donors and acceptors. Such a system "awaiting" interaction with a *complementary object* is called a *synthon* (the notion introduced by the Indian scholar Desiraju[85]). Crown ethers therefore contain the synthons able to recognize a narrow class of cations (with sizes within a certain range). In Fig. 5.20 we show another example of synthons based on hydrogen bonds. Due to the particular geometry of the molecules as well as to the abovementioned weak dependence of the XY distance on X and Y, both synthons are complementary.

The example just reported is of immense importance, because it pertains to guanine (G), cytosine (C), adenine (A), and thymine (T). Thanks to these two pairs of synthons (GC and AT) we exist, because G, C, A, and T represent the four letters which are sufficient to write the Book of Life word by word in a single molecule of DNA. The words, the sentences, and the chapters of this book decide the majority of the very essence of your (and my) personality. The whole DNA strand may be considered as a large single synthon. The synthon has its important counterpart, which fits the DNA perfectly because of the complementarity. The molecular machine which synthesizes this counterpart molecule (a "negative") is the polymerase, a wonderful molecule you will read about in Chapter 7. Any error in this complementarity results in a mutation.[86]

Fig. 5.21 shows the far-reaching consequences of an apparently minor change in the molecular structure (replacement of 16 hydrogen atoms by fluorines). In case R=H the molecule has a cone-like shape, while for R=F (larger atoms) the cone has to open. This means that the molecule has a concave part and resembles a plate. What happens next represents a direct consequence of this:

- In the first case the cone-like molecules associate laterally, which has to end up as a spherical structure similar to a micelle. The spheres pack into a cubic structure (like in NaCl).
- In the second case the "plates" match with each other and one gets stacks of them. The most effective packing is of the stack-to-stack type, leading to hexagonal liquid crystal packing.

5.14.2 "Key-lock," template-like, and "hand-glove" synthon interactions

The spectrum of the energy levels of a molecule represents something like its finger print. An energy level corresponds to certain electronic, vibrational, and rotational states (Chapter V1-6).

[85] G.R. Desiraju, "*Crystal Engineering, The Design of Organic Solids*," Elsevier, Amsterdam, 1989.

[86] Representing a potential or real danger, as well as a chance for evolution.

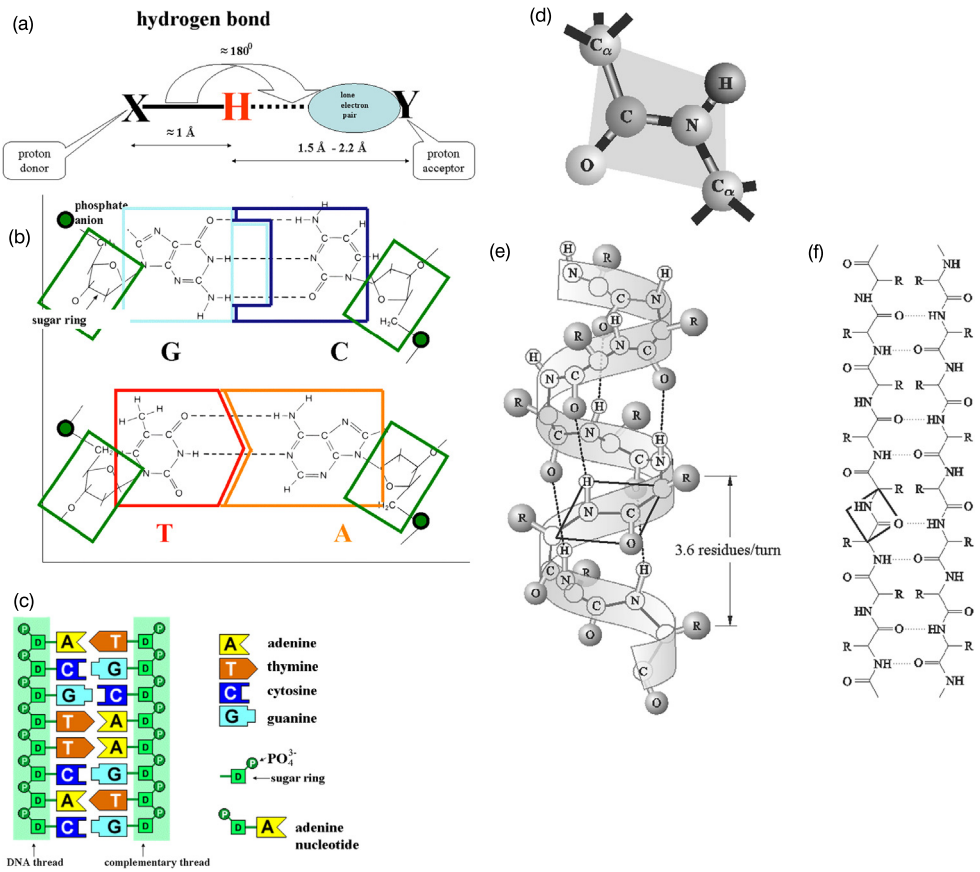


Fig. 5.20. The hydrogen bond and Nature's all important hydrogen bond-based synthons. (a) A hydrogen bond is almost linear. (b) The synthon of adenine (A) fits the synthon of thymine (T), forming a complementary AT pair (two hydrogen bonds involved) of these nucleobases. The guanine synthon (G) fits the cytosine synthon (C) forming the complementary GC pair (three hydrogen bonds involved). (c) A section of the DNA double helix shown in the form of a synthon–synthon interaction scheme. A single DNA thread is an alternant polymer of the sugar (five-carbon) rings (deoxyribose) and of the phosphate groups PO_4^{3-}, thus being a polyanion (the negative charge is compensated in the solution by the corresponding number of cations). Each deoxyribose offers an important substituent from the set of the nucleobases A, T, G, and C. In order to make the double helix, the second DNA thread must have the complementary bases to those of the first one. In this way the two threads represent two complementary polysynthons bound by the hydrogen bonds. (d) A protein represents a chain-like polymer, the monomers being the peptide bonds –NH–CO–, one of them highlighted in the figure. The peptide bonds are bound together through the bridging carbon atoms (known as the C_α carbons): $-(\mathrm{NH{-}CO{-}C_\alpha HR}_i)_N-$, where N is usually of the order of hundreds, and the functional groups R_i, $i = 1, 2, ..., 20$, stand for the amino acid side chains (Nature has 20 kinds of these side chains in genetic code). (e,f) Two structural motifs dominate in protein three-dimensional structures, both created by the hydrogen bond-based self-recognition of the synthons within the protein molecule (one peptide bond is highlighted). (e) The α-helical motif, in which a system of intramolecular hydrogen bonds NH$\cdots$OC forms a helical structure. (f) The β-sheet motif, based as well on the intramolecular hydrogen bonds NH$\cdots$OC, but formed by a lateral approaching of the peptide β-strands.

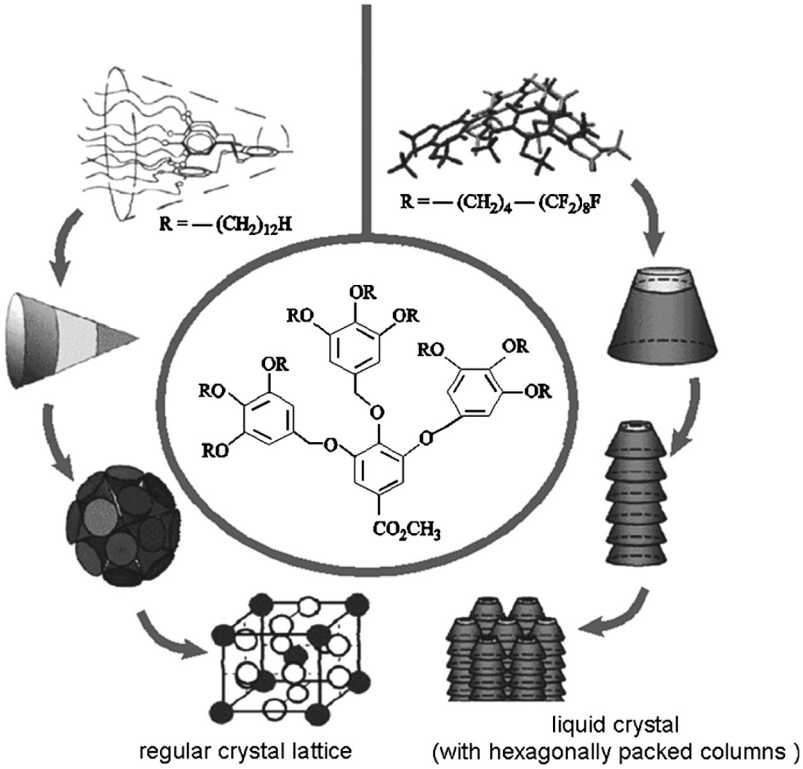


Fig. 5.21. The key-lock interaction in two marginally different situations: the molecules involved differ by replacing some of the hydrogens by fluorine atoms. The figure shows how profound consequences this seemingly small detail has. In one case we obtain a cone-like molecule (left) and then a crystal of cubic symmetry, in the other one the molecule has a plate-like shape (right), and because of that, we get finally a liquid crystal with hexagonal packing of columns of these molecular plates (Donald A. Tomalia, *Nature Materials*, 2(2003)711).

Different *electronic* states[87] may be viewed as representing different chemical bond patterns. Different *vibrational* states[88] form series, each series for an energy well on the PES. The energy level pattern is completed by the rotational states of the molecule as a whole. Since the electronic excitations are of the highest energy, the PES of the ground electronic state is most important. For larger molecules such a PES is characterized by a lot of potential energy wells corresponding to the conformational states. If the bottoms of the excited conformational wells are of high energy (with respect to the lowest-energy well, Fig. 5.22a), then the molecule in

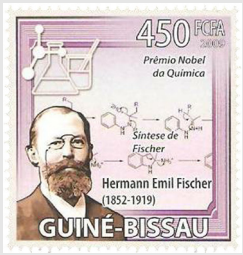
[87] In the Born–Oppenheimer approximation, each corresponding to a potential energy hypersurface.
[88] Including internal rotations, such as those of the methyl group.

its ground state may be called "rigid," because high energy is needed to change the molecular conformation.

If such rigid molecules *A* and *B* match perfectly each other, this corresponds to the *key-lock* type of molecular recognition. To match, the interacting molecules sometimes have only to orient properly in space when approaching one another and then dock (the AT or GC pairs may serve as an example). The key-lock concept of Fischer from 100 years ago (concerning enzyme–substrate interaction) is considered as the foundation of supramolecular chemistry – the chemistry that deals with the complementarity and matching of molecules.

One of the molecules, if rigid enough, may serve as a *template* for another molecule, which is flexible (Fig. 5.22b). Finally two flexible molecules (Fig. 5.22c) may pay an energy penalty for acquiring higher-energy conformations, but such ones which lead to a very strong interaction of the molecules in the *hand-glove* type of molecular recognition.

Hermann Emil Fischer (1852–1919), German chemist, professor at the universities in Strasbourg, Munich, Erlangen, Würzburg, and Berlin. Known mainly for his excellent works on the structure of sugar compounds. His (recognized decades later) correct determination of the absolute conformation of sugars was based solely on the analysis of their chemical properties. Even today this would require advanced physicochemical investigations. In 1902 Fischer



received the Nobel Prize "*for his work on sugar and purine syntheses*."

5.14.3 Convex and concave – the basics of strategy in the nanoscale

Chemistry can be seen from various points of view. In particular, one may think of it (liberating oneself for a while from the overwhelming influence of the Mendeleev table) as of a general strategy to tailor matter in the nanoscale,[89] in order to achieve a suitable physicochemical goal. Then we see an interesting feature; until recently the concepts of chemistry have been based on the intermolecular interaction of *essentially convex molecules*.[90]

89 Let us stress that chemists always had to do with so fashionable nowadays "nanostructures." The molecules have dimensions of the order of tens to hundreds of Å or more, i.e., 1–10 nm.

90 We do not count here some small concave details of molecular surfaces that were usually of the size of a single atom and did not play any particular role in theoretical considerations.

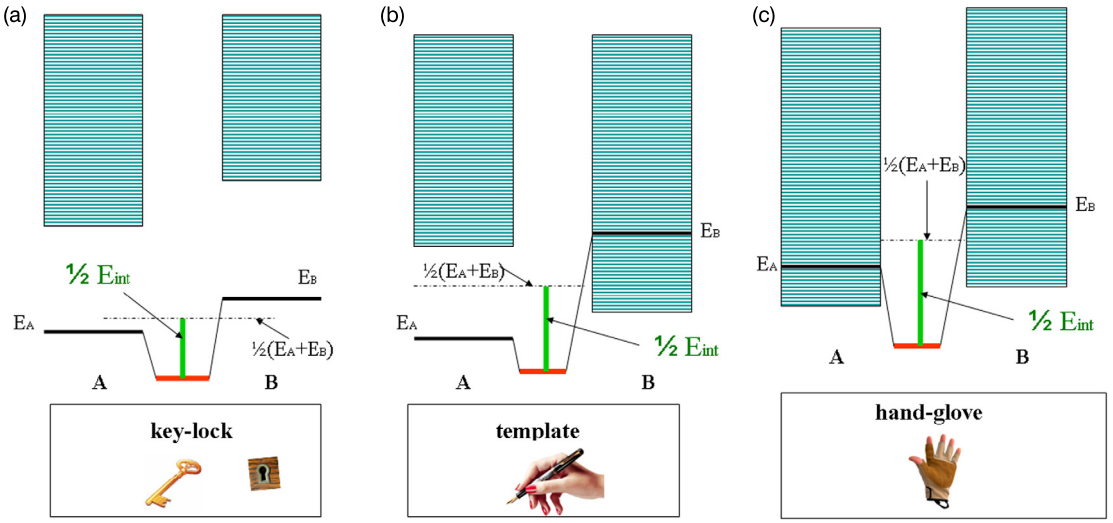


Fig. 5.22. The key-lock, template, and hand-glove interactions. Any molecule may be characterized by a spectrum of its energy levels. (a) In the key-lock-type interaction of two rigid molecules A and B their low-energy conformational states are separated from the quasicontinuum high-energy conformational states (including possibly those of some excited electronic states) by an energy gap, in general different for A and B. Due to the effective synthon interactions the energy per molecule decreases substantially with respect to that of the isolated molecules, leading to molecular recognition without significant changes in molecular shape. (b) In the template-like interaction one of the molecules is rigid (large energy gap), while the other one offers a quasicontinuum of conformational states. Among the latter, there is one that (in spite of being a conformational excited state), due to the perfect matching of synthons, results in considerable energy lowering, much below the energy of isolated molecules. Thus, one of the molecules has to distort in order to get perfect matching. (c) In the hand-glove type of interaction the two interacting molecules offer quasicontinua of their conformational states. Two of the excited conformational states correspond to such molecular shapes that match each other perfectly (molecular recognition) and lower the total energy considerably. This lowering is so large that it is able to overcome the conformational excitation energy (an energy cost of molecular recognition).

Practical use of molecular fitting in chemistry, taking place when *a convex molecule interacts with a concave one*, turned out to be a real breakthrough.[91] To reach this stage the chemists had to learn how to synthesize concave molecules, necessarily large ones, to host the convex ones. The convex–concave interaction has some peculiar features, which make precise operation in the nanoscale possible:

91 Strangely enough chemistry reached this stage as late as in 1960s (Charles Pedersen and his crown ethers).

- a convex molecule of a particular shape may fit a specific concave one, much better and with much more important energy gain and entropy loss than other molecules can (*molecular recognition*);
- elimination of some potential convex reactants ("guest molecules") to enter a reaction center inside a concave pocket of the "host molecule," just because their shape makes the contact impossible (due to steric hindrance, i.e., an excessive valence repulsion, Fig. 5.23a1);
- one may isolate the reaction center from some unwanted neighborhood (Fig. 5.23a2), which in extreme cases leads to molecular reaction vessels, even as large as a biological cell.

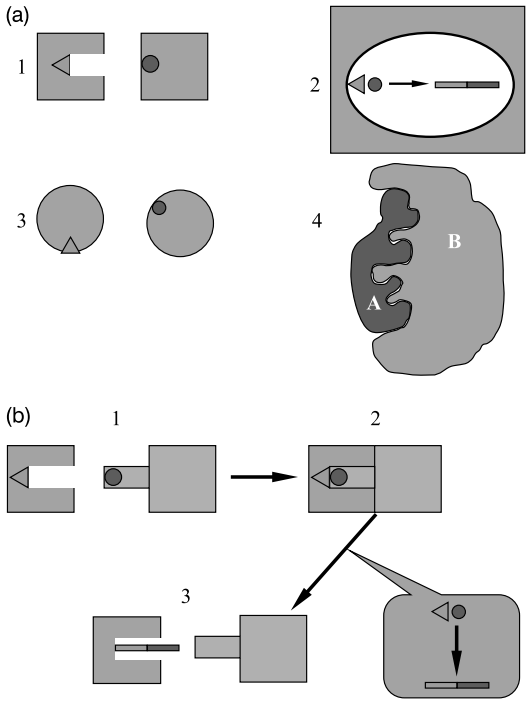


Fig. 5.23. How to make a particular chemical reaction of two molecules happen with high yield (scheme)? The figure shows several ways to get two reaction centers (one in each molecule, symbolized by triangle and circle) close in space. (a1) This architecture does not allow the reaction to proceed. (a2) The reaction will proceed with higher probability when in a reaction cavity (the cavity protects also the molecules from influence of other molecules of the neighborhood). (a3) For the two centers to meet, there must be a lot of unsuccessful attempts (which is known as entropic barrier). (a4) Molecules A and B attract effectively in one configuration only, at which they have a large contact surface. (b1,b2) A model of catalytic center (reaction cavity); the molecules fit best exactly at a configuration for which the reaction centers meet. (b3) The reaction products do not fit the cavity, which results in leaving the cavity.

- overcoming the entropy barrier (many nonreactive orientations) shown in Fig. 5.23a3 and positioning the reactants in space in some particular unique reactive orientation (similar to that shown in Fig. 5.23a4);
- a strong binding (many interatomic contacts distributed on *large contact surface*), in order to make a molecular complex (Fig. 5.23a4);
- at the same time sufficiently weak intermolecular binding in order to allow for reorganization of the complex;
- assuring the reaction centers (*predefined* by chemists) to be close in space and therefore forcing the reaction to proceed with a high yield (Fig. 5.23b1,b2);
- *leaving the reaction pocket by the products*, because they do not have enough space in the pocket (Fig. 5.23b3).

Therefore,

> we imagine an idea of a *molecular machine*, which offers a specific reaction space (pocket), selects from the neighborhood the right objects that match the pocket, fixes their positions as to have the key reaction centers close in space, makes the reaction proceed, and then removes the products from the reaction space, leaving the pocket ready for the next reaction.

Summary

- Interaction energy of two molecules (at a given geometry) may be calculated within any reliable quantum mechanical method by subtracting from the total system energy the sum of the energies of the subsystems. This is called a *supermolecular method*.
- The supermolecular method has at least one important advantage: *it works independently of the interaction strength and of the intermolecular distance*. The method has the disadvantage that due to the subtraction, a loss of accuracy may occur (especially for long distances) and no information is obtained about the structure of the interaction energy.
- In the supermolecular method there is a need to compensate for what is called the basis set superposition error (BSSE). The error appears because due to the incompleteness of the atomic basis set (Ω_A, Ω_B), the isolated subsystem A profits from the Ω_A basis set only, while when interacting lowers its energy also due to the total $\Omega_A \cup \Omega_B$ basis set (the same pertains to any of the subsystems). As a result a part of the calculated interaction energy does not come from the interaction, but from the problem of the basis set used (BSSE) described above. A remedy is called the counterpoise method, in which all quantities (including the energies of the individual subsystems) are calculated within the $\Omega_A \cup \Omega_B$ basis set.
- The perturbational method has limited applicability:
  - at long intermolecular separations what is called the polarization approximation may be used;
  - at medium distances a more general formalism called SAPT may be applied;

  - at short distances (of the order of chemical bond lengths) perturbational approaches are inapplicable.
- The main advantage of the perturbational approach is its ability to directly calculate the interaction energy.
- Another advantage of a low-order perturbational approach is the possibility of dividing the interaction energy into well-defined physically distinct energy contributions.
- In a polarization approximation approach, the unperturbed wave function is assumed as a *product* of the exact wave functions of the individual subsystems, i.e., $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$. The corresponding zero-order energy is $E_0^{(0)} = E_{A,0} + E_{B,0}$.
- Then, the first-order correction to the energy represents what is called the *electrostatic* interaction energy, $E_0^{(1)} = E_{\text{elst}} = \langle \psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle$, which is the Coulombic interaction (at a given intermolecular distance) of the frozen charge density distributions of the individual, noninteracting molecules. After using the multipole expansion E_{elst} can be divided into the sum of the multipole–multipole interactions plus a remainder, called the penetration energy. A multipole–multipole interaction corresponds to the permanent multipoles of the isolated molecules. An individual multipole–multipole interaction term (2^k-pole with 2^l-pole) vanishes asymptotically as $R^{-(k+l+1)}$, e.g., the dipole–dipole term decreases as $R^{-(1+1+1)} = R^{-3}$.
- In the second order we obtain the sum of the induction and dispersion terms, i.e., $E^{(2)} = E_{\text{ind}} + E_{\text{disp}}$.
- The induction energy splits into $E_{\text{ind}}(A \to B) = \sum'_{n_B} \frac{|\langle \psi_{A,0}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{E_{B,0}-E_{B,n_B}}$, which pertains to polarization of molecule B by the unperturbed molecule A, and $E_{\text{ind}}(B \to A) = \sum'_{n_A} \frac{|\langle \psi_{A,n_A}\psi_{B,0}|V|\psi_{A,0}\psi_{B,0}\rangle|^2}{E_{A,0}-E_{A,n_A}}$, with the roles of the molecules reversed. The induction energy can be represented as the permanent multipole–induced multipole interaction, where the interaction of the 2^k-pole with the 2^l-pole vanishes as $R^{-2(k+l+1)}$.
- The dispersion energy is defined as $E_{\text{disp}} = \sum'_{n_A}\sum'_{n_B} \frac{|\langle \psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0}-E_{A,n_A})+(E_{B,0}-E_{B,n_B})}$ and represents a result of the electronic correlation. After applying the multipole expansion, the effect can be described as a series of instantaneous multipole–instantaneous multipole interactions, with the individual terms decaying asymptotically as $R^{-2(k+l+1)}$. The most important contribution is the dipole–dipole ($k = l = 1$), which vanishes as R^{-6}.
- The polarization approximation fails for medium and short distances. For medium separations we may use *symmetry-adapted perturbation theory* (SAPT). The unperturbed wave function is symmetry-adapted, i.e., has the same symmetry as the exact function. This is not true for the polarization approximation, where the product-like $\varphi^{(0)}$ does not exhibit the proper symmetry with respect to electron exchanges between the interacting molecules. The symmetry adaptation is achieved by a projection of $\varphi^{(0)}$.
- SAPT *reproduces all the energy corrections that appear in the polarization approximation* (E_{elst}, E_{ind}, E_{disp}, ...) *and provides some exchange-type terms* (in each order of the perturbation).
- The most important exchange term is the valence repulsion appearing in the first-order correction to the energy,
  $E_{\text{exch}}^{(1)} = \langle \psi_{A,0}\psi_{B,0}|V\hat{P}^{AB}\psi_{A,0}\psi_{B,0}\rangle - \langle \psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle\langle \psi_{A,0}\psi_{B,0}|\hat{P}^{AB}\psi_{A,0}\psi_{B,0}\rangle + O(S^4)$,
  where $\hat{P}^{AB}$ stands for the single exchanges' permutation operator and $O(S^4)$ represents all the terms decaying as the fourth power of the overlap integral or faster.
- The interaction energy of N molecules *is not pairwise additive*, i.e., is not the sum of the interactions of all possible pairs of molecules. Among the energy corrections up to the second order, the exchange

and, first of all, the induction terms contribute to the nonadditivity. The electrostatic and dispersion (in the second order) contributions are pairwise additive.

- The nonadditivity is highlighted in what is called the *many-body expansion of the interaction energy*, where the interaction energy is expressed as the sum of two-body, three-body, etc., energy contributions. The m-body contribution, in a system of $N \geq m$ molecules, is defined as that part of the interaction energy that is nonexplicable by any interactions of smaller number of molecules, but explicable by the interactions among m molecules.
- The dispersion interaction in the third-order perturbation theory contributes to the three-body nonadditivity and is called the Axilrod–Teller energy. The term represents a correlation effect. Note that the effect is negative for three bodies in a linear configuration.
- The most important contributions: electrostatic, valence repulsion, induction, and dispersion, lead to a richness of supramolecular structures.
- The molecular surface (although not having an unambiguous definition) is one of the most important features of the molecules involved in the molecular recognition.
- The electrostatic interaction plays a particularly important role, because it is of a long-range character and is very sensible to relative orientation of the subsystems. The hydrogen bond X–H$\cdots$Y represents an example of the domination of the electrostatic interaction. This results in its directionality, linearity, and a small (as compared to typical chemical bonds) interaction energy, of the order of 5 kcal/mol.
- The valence repulsion is one of the most important energy contributions, because it controls how the interacting molecules fit together in space.
- The induction and dispersion interactions for polar systems, although contributing significantly to the binding energy, in most cases do not have a decisive role in forming structure and only slightly modify the geometry of the resulting structures.
- In aqueous solutions the solvent structure contributes very strongly to the intermolecular interaction, leading in particular to what is called the hydrophobic effect. The effect expels the nonpolar subsystems from the lattice of hydrogen bonds in water, thus causing the aliphatic (nonpolar) moieties to approach, which *looks* like an attraction.
- A molecule may have such a shape that it fits that of another molecule (synthons, small valence repulsion and a large number of attractive atom–atom interactions).
- In this way molecular recognition may be achieved by the key-lock-type fit (the molecules nondistorted), the template fit (one molecule distorted), or the hand-glove-type fit (both molecules distorted).
- Molecular recognition may be planned by chemists and used to build complex molecular architectures, in a way similar to that in which living matter operates.

Main concepts, new terms

amphiphilicity (p. 420)
Axilrod–Teller dispersion energy (p. 410)
basis set superposition error (BSSE) (p. 349)
binding energy (p. 346)
catenans (p. 347)
dipole–dipole (p. 362)
dispersion energy (p. 353)
dissociation barrier (p. 347)
dissociation energy (p. 346)
electrostatic energy (p. 353)

From the research front

Intermolecular interactions influence any liquid and solid state measurements. Physicochemical measurement techniques give only some indications of the shape of a molecule, except NMR, X-ray, and neutron analyses, which provide the atomic positions in space, but are very expensive. This is why there is a need for theoretical tools which may offer such information in a less expensive way. For very large molecules, such an analysis uses the force fields described in Chapter V1-7. This is currently the most powerful theoretical tool for determining the approximate shape of molecules with numbers of atoms even of the order of thousands. To obtain more reliable information about intermolecular interactions we may perform quantum-mechanical calculations within a supermolecular approach, necessarily of an *ab initio* type, because other methods give rather low-quality results. If the particular method chosen is the Hartree–Fock approach (currently limited to hundreds atoms), we have to remember that it cannot take into account any dispersion contribution to the interaction energy by *definition*.[92] *Ab initio* calculations of the correlation energy still represent a challenge. High-quality calculations for a molecule with a hundred atoms may be carried out using the MP2 method. Still more time consuming are the CCSD(T) and SAPT calculations, which are feasible only for systems with a dozen atoms, but offer an accuracy of 1 kcal/mol, required for chemical applications.

[92] The dispersion energy represents an electronic correlation effect, absent in the Hartree–Fock energy.

Ad futurum

No doubt the computational techniques will continue to push the limits mentioned above. The more coarse the method used, the more spectacular this pushing will be. The most difficult to imagine would be a great progress in methods using explicitly correlated wave functions. It seems that pushing the experimental demands and calculation time required will cause experimentalists (they will perform the calculations[93]) to prefer a rough estimation using primitive methods rather than waiting too long for a precise result (still not very appropriate, because obtained without taking the influence of solvent, etc., into account). It seems that in the near future we may expect theoretical methods exploiting the synthon concept. It is evident that a theoretician has to treat the synthons on an equal footing with other parts of molecule, but a practice-oriented theoretician cannot do that; otherwise he would wait virtually forever for something to happen in the computer, while in reality the reaction takes only a picosecond or so. Still further in the future we will see the planning of hierarchic multilevel supramolecular systems, taking into account the kinetics and competitiveness among such structures. In the still more distant future, functions performed by such supramolecular structures, as well as their sensitivity to changing external conditions, will be investigated.

Additional literature

J.O. Hirschfelder, C.F. Curtiss, R.B. Bird, "Molecular Theory of Gases and Liquids," Wiley, New York, 1964.

A thick "bible" (1249 pages) of intermolecular interactions. We find there everything before the advent of computers and of SAPT.

H. Margenau, N.R. Kestner, "Theory of Intermolecular Forces," Pergamon, Oxford, 1969.

A lot of detailed derivations. There is also a chapter devoted to the nonadditivity of the interaction energy – a *rara avis* in textbooks.

"Molecular Interactions," eds. H. Ratajczak, W.J. Orville-Thomas, Wiley, Chichester, 1980.

A three-volume edition containing a selection of articles by experts in the field.

A.J. Stone, "The Theory of Intermolecular Forces," Oxford Univ. Press, Oxford, 1996.

The book contains the basic knowledge in the field of intermolecular interactions given in the language of perturbation theory as well as the multipole expansion (a lot of useful formulae for the electrostatic, induction, and dispersion contributions). This very well-written book presents many important problems in a clear and comprehensive way.

"Molecular Interactions", ed. S. Scheiner, Wiley, Chichester, 1997.

A selection of articles written by experts.

93 What will theoreticians do? My answer is given in Chapter 7.

P. Hobza, R. Zahradnik, "Intermolecular Complexes," Elsevier, Amsterdam, 1988.

Many useful details.

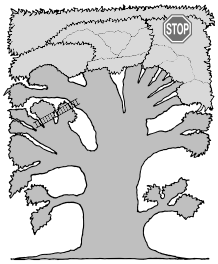
Questions

1. The rigid interaction energy of subsystems A and B in the AB system:
   a. does not represent a measurable quantity.
   b. requires calculation of the electronic energy of AB, A, and B, where the configurations of the nuclei in A and B are the same as those in AB.
   c. requires calculation of the electronic energy of AB, A, and B, where the configurations of the nuclei in A and B correspond to the minimum energy of these subsystems.
   d. the interaction energy of two water molecules in the H_4O_2 system is an unambiguously defined quantity.
2. The Boys–Bernardi method of removing the basis set superposition error (BSSE) means:
   a. a high-quality calculation: all calculations within the most extended basis set available, feasible separately for A, B, and AB.
   b. the energy of the total system should be calculated within the joint sets of the atomic orbitals centered on the individual subsystems, while the energy of the subsystems should be calculated within their individual basis sets.
   c. the atomic basis set for AB represents the sum of the basis sets for A and B.
   d. all quantities are calculated within a common basis set being the sum of the sets for the individual molecules.
3. The zero-order wave function in the polarization perturbation theory (for a finite intermolecular distance):
   a. does not satisfy the Pauli exclusion principle.
   b. represents a product of the wave functions for the polarized molecules.
   c. admits that two electrons with the same spin coordinates occupy the same point in space.
   d. represents a product of the wave functions for the isolated molecules.
4. Induction energy (R denotes the intermolecular distance):
   a. decays as R^{-7} for the interaction of two hydrogen molecules.
   b. decays as R^{-6} for the immobilized water and ammonia molecules.
   c. represents an attraction.
   d. is an electronic correlation effect.
5. Dispersion energy (R denotes the intermolecular distance):
   a. is not equal to zero for two polar molecules.
   b. the Hartree–Fock method overestimates its value.
   c. represents an electronic correlation effect.
   d. for R sufficiently large decays as R^{-6}.
6. The multipole moments of a point particle of charge q and the coordinates x, y, z:
   a. the only nonzero multipole moment of a point-like particle represents its charge, i.e., its monopole.
   b. the z component of the dipole moment is equal to qz.
   c. a point-like particle cannot have a dipole moment, therefore we always have $\mu_z = 0$.
   d. the values of the multipole moments of a particle depend on the coordinate system chosen.

7. In the SAPT for $H_2O\cdots H_2O$:
   a. the dispersion energy appears in the second order.
   b. one obtains a minimum of the electronic energy.
   c. the valence repulsion appears in the first order.
   d. the zero-order wave function is asymmetric with respect to exchange of the coordinates of any two electrons.
8. In the SAPT for $Ne\cdots Ne$:
   a. the wave function of the zeroth order represents an antisymmetrized product of the wave function for the isolated neon atoms.
   b. the electrostatic energy appears in the first order and is equal to zero within the multipole approximation.
   c. the exchange corrections appear in every order.
   d. the penetration part of the electrostatic energy is equal to zero.
9. Additivity and nonadditivity of the interaction:
   a. the electrostatic energy is always additive, while the induction energy is always nonadditive.
   b. the three-body contribution represents this part of the interaction energy of N subsystems, which cannot be explained by the pairwise interactions of the subsystems.
   c. the dispersion energy (in the second order) contains only the pairwise interactions of the subsystems.
   d. additivity means that the interaction energy is a sum of the pairwise interaction energies.
10. In supramolecular chemistry:
   a. molecular recognition means a strong attraction of the molecules at their unique mutual configuration only.
   b. the angle $O–H\cdots O$ in the hydrogen bond $HO–H\cdots OH_2$ is equal to 180.0°.
   c. the term hydrogen bond pertains to the hydrogen–hydrogen interaction, where the hydrogens belong to different molecules.
   d. the "hand-glove" interaction is also known as the "key-lock" interaction.

Answers

1a,b, 2d, 3a,c,d, 4b,c, 5a,c,d, 6b,d, 7a,b,c, 8a,b,c, 9a,c,d, 10a



CHAPTER 6

Chemical Reactions

Enter through the narrow gate, for the gate is wide and the way is broad that leads to destruction...
Saint Matthew (7.13)

Where are we?

We are already picking fruit in the crown of the TREE.

Example

Why do two substances react and another two do not? Why does a reaction start when increasing the temperature? Why does a reaction mixture change color? As we know from Chapter V1-6, this tells us about some important electronic structure changes. On the other hand the products (that differ so much from the reactants) tell us about profound changes in the positions of the nuclei that take place simultaneously. Something dramatic is going on. But what and why?

What is it all about?

How to describe a chemical reaction in terms of quantum mechanics?

The structure of the chapter is the following.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00014-4
Copyright © 2020 Elsevier B.V. All rights reserved.

AB INITIO APPROACH

Accurate solutions (three atoms) (℧ ⧫✠) **p. 453**

- Coordinate system and Hamiltonian
- Solution to the Schrödinger equation
- Berry phase (✠)

APPROXIMATE METHODS

Intrinsic reaction coordinate (IRC) (△) **p. 460**

Reaction path Hamiltonian method (℧ ⧫) **p. 463**

- Energy close to IRC
- Vibrational adiabatic approximation
- Vibrational nonadiabatic model
- Application of the reaction path Hamiltonian method to the reaction $H_2+OH \rightarrow H_2O+H$

Acceptor–donor (AD) theory of chemical reactions (℧ ⧫) **p. 479**

- A simple model of nucleophilic substitution – MO, AD, and VB formalisms
- MO picture → AD picture
- Reaction stages
- Contributions of structures as the reaction proceeds
- Nucleophilic attack – the model is more general: $H^- +$ ethylene → ethylene $+ H^-$
- The model looks even more general: the electrophilic attack $H^+ + H_2 \rightarrow H_2 + H^+$
- The model works also for the nucleophilic attack on a polarized bond

Symmetry-allowed and symmetry-forbidden reactions (△) **p. 501**

- Woodward–Hoffmann symmetry rules
- AD formalism
- Electrocyclic reactions
- Cycloaddition reaction
- Barrier means a cost of opening the closed shells

Barrier for the electron transfer (℧ Ⓢ) **p. 509**

- Diabatic and adiabatic potential
- Marcus theory
- Solvent-controlled electron transfer

We are already acquainted with the toolbox for describing the electronic structure at *any* position of the nuclei. It is time now to look at possible *large* changes of the electronic structure at *large* changes of nuclear positions. The two motions, that of electrons and that of nuclei, will be coupled together (especially in a small region of the configurational space).

Our plan consists of four parts:

- In the first part (after using the Born–Oppenheimer approximation fundamental to this chapter), we assume that we have calculated the ground-state electronic energy, i.e., the potential energy for the nuclear motion. It will turn out that the hypersurface has a *characteristic "drain-pipe" shape, and its bottom in the central section, in many cases, exhibits a barrier*. Taking a three-atom example, we will show how the problem *could* be solved *if we were capable* of calculating the quantum dynamics of the system accurately.
- In the second part we will concentrate on a specific representation of the system's energy that takes *explicitly* into account the abovementioned reaction drain-pipe ("*reaction path Hamiltonian*"). Then we will focus on the description of how a chemical reaction proceeds. Just to be more specific, an example will be shown in detail.
- In the third part (acceptor–donor theory of chemical reactions) we will find the answer to the question of *where the reaction barrier comes from and what happens to the electronic structure when the reaction proceeds*.
- The fourth part will pertain to the *reaction barrier height* in electron transfer (a subject closely related to the third part).

Why is this important?

Chemical reactions are at the heart of chemistry, making possible the achievement of its ultimate goals, which include synthesizing materials with desired properties. What happens in the chemist's flask is a complex phenomenon which consists of an astronomical number of elementary reactions of individual molecules. In order to control the reactions in the flask, it would be good to *understand first the rules which govern these elementary reaction acts*. This is the subject of the present chapter.

What is needed?

- Hartree–Fock method (Chapter V1-8, necessary),
- conical intersection (Chapter V1-6, necessary),
- normal modes (Chapter V1-7, necessary),
- Appendices V1-N (recommended), V1-E (just mentioned), I (necessary), V1-J (recommended), and V1-G (just mentioned),
- elementary statistical thermodynamics or even phenomenological thermodynamics: entropy, free energy (necessary).

Classical works

Everything in chemistry began in the 1920s.

The first publications that considered conical intersection – a key concept for chemical reactions – were two articles from the Budapest schoolmates Janos (John) von Neumann and Jenó Pál (Eugene) Wigner

John Charles Polanyi (born 1929), Canadian chemist of Hungarian origin, son of Michael Polanyi (one of the pioneers in the field of chemical reaction dynamics), professor at the University of Toronto. John was attracted to chemistry by Meredith G. Evans, who was a student of his father. Three scholars, John Polanyi, Yuan Lee, and Dudley Herschbach, shared the 1986 Nobel Prize "*for their contributions concerning the dynamics of chemical elementary processes*."

Yuan T. Lee is a native of Taiwan, called by his colleagues "a Mozart of physical chemistry." He wrote that he was deeply impressed by a biography of Marie Curie and that her idealism decided his own path. Yuan Lee constructed a "supermachine" for studying crossing molecular beams and the reactions in them. One of the topics was the alkali metal atom–iodine collisions.

"*Über merkwürdige diskrete Eigenwerte*," published in *Physikalische Zeitschrift,* 30(1929)465, and "*Über das Verhalten von Eigenwerten bei adiabatischen Prozessen*," which also appeared in *Physikalische Zeitschrift*, 30(1929)467. ★ Then a paper "*The Crossing of Potential Surfaces*" by their younger schoolmate Edward Teller was published in the *Journal of Chemical Physics*, 41(1937)109. ★ A classical theory of the "reaction drain-pipe" with entrance and exit channels was first proposed by Henry Eyring, Harold Gershinowitz, and Cheng E. Sun in "*Potential Energy Surface for Linear* H_3," *Journal of Chemical Physics*, 3(1935)786, and then by Joseph O. Hirschfelder, Henry Eyring, and Brian Topley in the article "*Reactions Involving Hydrogen Molecules and Atoms*," in *Journal of Chemical Physics*, 4(1936)170, and by Meredith G. Evans and Michael Polanyi in "*Inertia and Driving Force of Chemical Reactions*," which appeared in *Transactions of the Faraday Society*, 34(1938)11. ★ Hugh Christopher Longuet-Higgins, Uno Öpik, Maurice H.L. Pryce, and Robert A. Sack, in a splendid paper "*Studies of the Jahn–Teller Effect*," *Proceedings of the Royal Society of London A*, 244(1958)1, noted for the first time that the wave function changes its phase close to a conical intersection, which later became known as the Berry phase. ★ The acceptor–donor description of chemical reactions was first proposed by Robert S.J. Mulliken in "*Molecular Compounds and their Spectra*," *Journal of the American Chemical Society*, 74(1952)811. ★ The idea of the intrinsic reaction coordinate (IRC) was first given by Isaiah Shavitt in "*The Tunnel Effect Corrections in the Rates of Reactions with Parabolic and Eckart Barriers*," Report WIS-AEC-23, Theoretical Chemistry Lab., University of Wisconsin, (1959), as well as by Morton A. Eliason and Joseph O. Hirschfelder in the *Journal of the Chemical Physics*, 30(1959)1426 in the article "*General Collision Theory Treatment for the Rate of Bimolecular, Gas Phase Reactions.*" ★ The symmetry rules allowing some reactions and forbidding others were first proposed by Robert B. Woodward and Roald Hoffmann in two letters to the editor:

"*Stereochemistry of Electrocyclic Reactions*" and "*Selection Rules for Sigmatropic Reactions*," *Journal of American Chemical Society*, 87(1965)395, 2511, as well as by Kenichi Fukui and Hiroshi Fujimoto

Dudley Herschbach writes in his CV that he spent his childhood in a village close to San Jose, picking fruit, milking cows, etc. Thanks to his wonderful teacher he became interested in chemistry. He graduated from Harvard University (physical chemistry), where as he says, he has found "an exhilarating academic environment." In 1959 he became professor at the University of California at Berkeley.

in an article published in the *Bulletin of the Chemical Society of Japan*, 41(1968)1989. ★ The concept of the steepest descent method was formulated by Kenichi Fukui in "*A Formulation of the Reaction Coordinate*," which appeared in the *Journal of Physical Chemistry*, 74(1970)4161, although the idea seems to have a longer history. ★ Other classical papers include a seminal article by Sason S. Shaik, "*What Happens to Molecules as They React? Valence Bond Approach to Reactivity*," in *Journal of the American Chemical Society*, 103(1981)3692. ★ The Hamiltonian path method was formulated by William H. Miller, Nicolas C. Handy, and John E. Adams in the article "*Reaction Path Hamiltonian for Polyatomic Molecules*," in the *Journal of the Chemical Physics*, 72(1980)99. ★ The first quantum dynamics simulation was performed by a PhD student, George C. Schatz (under the supervision of Aron Kupperman), for the reaction $H_2+H\rightarrow H+H_2$, reported in "*Role of Direct and Resonant Processes and of their Interferences in the Quantum Dynamics of the Collinear* $H+H_2$ *Exchange Reaction*," in *Journal of Chemical Physics*, 59(1973)964. ★ John Polanyi, Dudley Herschbach, and Yuan Lee proved that the lion's share of the reaction energy is delivered through the rotational degrees of freedom of the products, e.g., J.D. Barnwell, J.G. Loeser, D.R. Herschbach, "*Angular Correlations in Chemical Reactions. Statistical Theory for Four-Vector Correlations*," published in the *Journal of Physical Chemistry*, 87(1983)2781. ★ Ahmed Zewail (Egypt/USA) developed an amazing experimental technique known as femtosecond spectroscopy, which for the first time allowed the study of the reacting molecules at different stages of an ongoing reaction ("*Femtochemistry – Ultrafast Dynamics of The Chemical Bond*," vols. I and II, A.H. Zewail, World Scientific, New Jersey, Singapore (1994)). ★ Among others, Josef Michl, Lionel Salem, Donald G. Truhlar, Robert E. Wyatt, and W. Ronald Gentry also contributed to the theory of chemical reactions.

* * *

6.1 Hypersurface of the potential energy for nuclear motion

Theoretical chemistry is currently in a stage which experts in the field characterize as "the primitive beginnings of chemical *ab initio* dynamics.[1]" The majority of the systems studied so far are *three-atomic*.[2]

The Born–Oppenheimer approximation works wonders, as it is possible to consider the (classical or quantum) dynamics of the *nuclei*, while the electrons disappear from the scene (their role became, after determining the potential energy for the motion of the nuclei, described in the electronic energy, the quantity corresponding to $E_0^0(R)$ from Eq. (V1-6.8) on p. V1-316).

Even with this approximation our job is not simple:

- The reactants as well as the products may be quite large systems and the many-dimensional ground-state potential energy hypersurface $E_0^0(\mathbf{R})$ may have a very complex shape, whereas we are most often interested in the small fragment of the hypersurface that pertains to a particular one of many possible chemical reactions.
- We have many such hypersurfaces $E_k^0(\mathbf{R})$, $k = 0, 1, 2, ...$, each corresponding to an electronic state: $k = 0$ means the ground state, $k = 1, 2, ...$ correspond to the excited states. There are processes which take place on a single hypersurface without changing the chemical bond pattern,[3] but the very essence of chemical reaction is to change the bond pattern, and therefore excited states come into play.

It is quite easy to see where the fundamental difficulty lies. Each of the hypersurfaces $E_k^0(\mathbf{R})$ for the motion of $N > 2$ nuclei depends on $(3N - 6)$ atomic coordinates (the number of translational and rotational degrees of freedom was subtracted).

Determining the hypersurface is not an easy matter:

- A high accuracy of 1 kcal/mol is required, which is (for a fixed configuration) very difficult to achieve for *ab initio* methods,[4] and even more difficult for the semiempirical or empirical methods.

---

1 R.D. Levine and R.B. Bernstein, "*Molecular Reaction Dynamics and Chemical Reactivity*," Oxford University Press, 1987.

2 John Polanyi recalls that the reaction dynamics specialists used to write as the first equation on blackboard A + BC → AB + C, which made any audience burst out laughing. However, one of the outstanding specialists (Zare) said about the simplest of such reactions (H_3) (*Chem. Engin. News*, June 4 (1990)32): "*I am smiling, when somebody calls this reaction the simplest one. Experiments are extremely difficult, because one does not have atomic hydrogen in the stockroom, especially the high speed hydrogen atoms (only these react). Then, we have to detect the product, i.e., the hydrogen, which is a transparent gas. On top of that it is not sufficient to detect the product in a definite spot, but we have to know which quantum state it is in.*"

3 Strictly speaking a change of conformation or formation of an intermolecular complex represents a chemical reaction. Chemists, however, reserve this notion for more profound changes of electronic structure.

4 We have seen in Chapter 2 that the correlation energy is very difficult to calculate.

- The *number of points* on the hypersurface which have to be calculated is extremely large and increases exponentially with the system size.[5]
- There is no general methodology telling us what to do with the calculated points. There is a consensus that we should approximate the hypersurface by a smooth analytical function, but no general solution has yet been offered.[6]

6.1.1 Potential energy minima and saddle points

Let us denote $E_0^0(\mathbf{R}) \equiv V$. The most interesting points of the hypersurface V are its *critical points*, i.e., the points for which the gradient ∇V is equal to zero:

$$G_i = \frac{\partial V}{\partial X_i} = 0 \quad \text{for} \quad i = 1, 2, \ldots, 3N, \tag{6.1}$$

where X_i denote the Cartesian coordinates that describe the configurations of N nuclei. Since $-G_i$ represents the force acting along the axis X_i, no forces act on the atoms in the configuration of a critical point.

There are several types of critical points. Each type can be identified after considering the *Hessian*, i.e., the matrix with elements

$$V_{ij} = \frac{\partial^2 V}{\partial X_i \partial X_j} \tag{6.2}$$

calculated for the critical point. There are three types of critical points: maxima, minima, and saddle points (cf. Chapter V1-7 and Fig. V1-7.11, as well as the Bader analysis, p. 197). The saddle points, as will be shown in a while, are of several classes depending on the signs of the Hessian eigenvalues. Six of the eigenvalues are equal to zero (rotations and translations of the total system, see p. V1-416), because this type of motion proceeds without any change of the potential energy V.

We will concentrate on the remaining $(3N - 6)$ eigenvalues:

---

[5] Indeed, if we assume that ten values for each coordinate axis are sufficient (and this looks like a rather poor representation), then for N atoms we have 10^{3N-6} quantum mechanical calculations of good quality to perform. This means that for $N = 3$ we may still pull it off, but for larger N everybody has to give up. For example, for the reaction $HCl + NH_3 \rightarrow NH_4Cl$ we would have to calculate 10^{12} points in the configurational space, while even a single point is a computational problem.

[6] Such an approximation is attractive for two reasons: first, we dispose the (approximate) values of the potential energy for *all* points in the configuration space (not only those for which the calculations were performed), and second, the analytical formula may be differentiated and the derivatives give the forces acting on the atoms.
It is advisable to construct the abovementioned analytical functions following some theoretical arguments. These are supplied by intermolecular interaction theory (see Chapter 5).

- In the minimum the $(3N-6)$ Hessian eigenvalues $\lambda_k \equiv \omega_k^2$ (ω is the angular frequency of the corresponding normal modes) are all positive.
- In the maximum all are negative.
- For a saddle point of the n-th order, $n = 1, 2, \ldots, 3N-7$, the n eigenvalues are negative, the rest positive. Thus, a first-order saddle point corresponds to all but one of the positive Hessian eigenvalues, i.e., one of the angular frequencies ω is therefore imaginary.

The eigenvalues were obtained by diagonalization of the Hessian. Such diagonalization corresponds to a rotation of the local coordinate system (cf. p. V1-421). Imagine a two-dimensional surface that at the minimum could be locally approximated by an ellipsoidal valley. The diagonalization means such a rotation of the coordinate system x, y that both axes of the ellipse coincide with the new axes x', y' (Chapter V1-7). On the other hand, if our surface locally resembled a cavalry saddle, diagonalization would lead to such a rotation of the coordinate system that one axis would be directed along the horse, and the other across.[7]

IR and Raman spectroscopy providing the vibration frequencies and force constants tell us a lot about how the energy hypersurface close to minima looks, both for the reactants and the products. On the other hand, theory and recently also the femtosecond spectroscopy[8] are the only sources of information about the first-order saddle points. However, the latter are extremely important for determining reaction rates since any saddle point is a kind of pivot point – it is as important for the reaction as the Rubicon was for Caesar.[9]

The simplest chemical reactions are those which do not require crossing any reaction barrier. For example, the reaction $Na^+ + Cl^- \rightarrow NaCl$ or recombination of two radicals are not accompanied by bond breaking or bond formation and take place *without any barrier*.[10]

After the barrierless reactions, there is a group of reactions in which the reactants and the products are separated on the hypersurface V by a single saddle point (no intermediate products). *How do we describe such a reaction in a continuous way?*

---

7 A cavalry saddle represents a good example of the first-order saddle of a two-dimensional surface.

8 In this spectroscopy we hit a molecule with a laser pulse of a few femtoseconds. The pulse perturbs the system, and when relaxing it is probed by a series of new pulses, each giving a spectroscopic fingerprint of the system. A femtosecond is an incredibly short time, during which light is able to run only about $3 \cdot 10^{-5}$cm. Ahmed Zewail, the discoverer of this spectroscopy, received the Nobel Prize in 1999.

9 In 49 BCE Julius Caesar heading his Roman legions crossed the Rubicon river (the border of his province of Gaul), which initiated civil war with the central power in Rome. His words "*alea iacta est*" (the die is cast) became a symbol of a final and irreversible decision.

10 As a matter of fact, the formation of van der Waals complexes may also belong to this group. However, in large systems, when precise docking of molecules take place, the final docking may occur with a steric barrier.

A potential energy surface (PES) (and therefore all possible configurations of the nuclei) can be divided into separate valleys with the saddle points at their borders. Let us focus on a particular energy valley of the ground-state PES. One may ask about the stability of the system occupying this valley. A suitable kinetic energy may be sufficient to allow the system pass the saddle points at the border of the valley (Fig. 6.1). The distances from the bottom of the valley to these saddle points may be different (depending on how large a deformation of nuclear configuration is needed to reach a given saddle point) and, more importantly, may correspond to different energies to be overcome ("barriers for a given reaction"). The destabilization is the easiest when moving towards the *lowest-energy saddle point*, which is often associated with the motion described by the lowest-frequency vibrational mode. The larger the kinetic energy of the system, the more numerous possibilities to crossing over the saddle points, each characterized by some particular products.[11]

A possibility to pass to next and next valleys (through the saddle points) means a chain of chemical reactions.

Fig. 6.1. An analogy between the electronic ground-state PES and a mountain landscape, as well as between the stability of a molecular system and the stability of a ball in the landscape. On the right-hand side is seen the lake; its bottom's lowest point represents the most stable position of the ball. If the ball had a small kinetic energy, it would only circulate around the bottom of the valley, which is analogous to molecular vibrations (preserving the pattern of the chemical bonds). If the kinetic energy were larger, this would mean a larger amplitude of such oscillations, in particular such one that corresponds to passing over the barrier (see the arrow towards the lowest saddle point) and reaching the next valley, which means a change of the chemical bond pattern; one reaction channel is open. One sees that when the kinetic energy is getting larger, the number of possible open reaction channels increases quickly (other arrows).

[11] In other words, opening new reaction channels.

6.1.2 Distinguished reaction coordinate (DRC)

We often define a reaction path in the following way:

- First, we make a choice of a particular distance (s) between the reacting molecules (e.g., an interatomic distance, one of the atoms belongs to molecule A, the other to B).
- Then we minimize the potential energy by optimization of all atomic positions, while keeping the s distance fixed.[12]
- We change s by small increments from its reactant value until the product value is obtained (for each s optimizing all other distances).
- This defines a path (DRC) in the configurational space; the progress along the path is measured by s.

A deficiency of the DRC is an arbitrary choice of the distance. The energy profile obtained (the potential energy versus s) depends on the choice. Often the DRC is reasonably close to the reactant geometry and becomes misleading when close to the product value (or *vice versa*). There is no guarantee that such a reaction path passes through the saddle point. On top of this other coordinates may undergo discontinuities, which feels a little catastrophic.

6.1.3 Steepest descent path (SDP)

Because of the Boltzmann distribution the potential energy *minima* are most important, mainly low-energy ones.[13]

The saddle points of the first order are also important, because it is believed that any two minima may be connected by a chain of first-order saddle points. Several first-order saddle points to pass mean a multistage reaction that consists of several steps, each one representing a pass through a single first-order saddle point (elementary reaction). Thus, *the least energy demanding path from the reactants to products goes via a saddle point of the first order*. This steepest descent path (SDP) is determined by the direction $-\nabla V$. First, we choose a first-order saddle point $\mathbf{R}_0$, then we diagonalize the Hessian matrix calculated at this point and the eigenvector $\mathbf{L}$ corresponding to the single negative eigenvalue of the Hessian (cf. p. V1-419). Now, let us move all atoms *a little* from position $\mathbf{R}_0$ in the direction indicated in the configurational space by $\mathbf{L}$, and then let us follow vector $-\nabla V$ until it reduces to zero (then we are at the minimum).

[12] This is equivalent to calculating the relaxed interaction energy of Eq. (5.2).

[13] Putting aside some subtleties (e.g., does the minimum support a vibrational level), the minima correspond to stable structures, since a small deviation from the minimum position causes a gradient of the potential to become nonzero, and this means a force pushing the system back towards the minimum position.

In this way we have traced half the SDP. The other half will be determined starting down from the other side of the saddle point and following the $-\mathbf{L}$ vector first.

In a moment we will note a certain disadvantage of the SDP, which causes us to prefer another definition of the reaction path (see p. 461).

6.1.4 Higher-order saddles

The first-order saddle points play a prominent role in the theory of chemical reactions. How can a chemist imagine a higher-order saddle point? Are they feasible at all in chemistry? The first-order saddle point may be modeled by a bond A–X with atom X in a position with a repulsion with atom B, from which a departure of X means an energy relief. A multiple-order saddle point may correspond to a geometry with several such atoms stuck in a high-energy position.

6.1.5 Our goal

We would like to present a theory of elementary chemical reactions within the Born–Oppenheimer approximation, i.e., which describes nuclear motion on the potential energy hypersurface.

We have the following alternatives:

1. To perform *molecular dynamics*[14] on the hypersurface V (a point on the hypersurface represents the system under consideration).
2. To solve the *time-independent Schrödinger equation* $\hat{H}\psi = E\psi$ for the motion of the nuclei with potential energy V.
3. To solve the time-dependent *Schrödinger equation* with the boundary condition for $\psi(x, t = 0)$ in the form of a wave packet.[15] The wave packet may be directed into the entrance channel towards the reaction barrier (from various starting conditions). In the barrier range, the wave packet splits into a wave packet crossing the barrier and a wave packet reflected from the barrier (cf. p. V1-203).
4. To perform a semiclassical analysis that highlights the existence of the SDP, or a similar path, leading from the reactant to the product configuration.

Before going to more advanced approaches, let us consider possibility 1.

[14] A classical approach. We have to ensure that the bonds may break; this is a very nontypical molecular dynamics.

[15] For example, a Gaussian function (in the nuclear coordinate space) moving from a position in this space with a starting velocity.

6.2 Chemical reaction dynamics (a pioneers' approach)

The SDP does not represent the only possible reaction path. It is only the *least energy-expensive path* from reactants to products. In real systems, the point representing the system will attempt to get through the pass in many different ways. Many such attempts are unsuccessful (non-reactive trajectories). If the system leaves the entrance channel (reactive trajectories), it will not necessarily pass through the saddle point, because it may have some extra kinetic energy, which may allow it to go with a higher energy than that of the barrier. Everything depends on the starting position and velocity of the point running through the entrance channel.

In the simplest case of a three-atom reaction,

$$\mathrm{A} + \mathrm{BC} \rightarrow \mathrm{AB} + \mathrm{C},$$

the potential energy hypersurface represents a function of $3N - 6 = 3$ coordinates (the translations and rotations of the total system were separated). Therefore, even in such a simple case, it is difficult to draw this dependence. To make this possible we may simplify the problem by considering only a limited set of geometries, e.g., the three atoms in a linear configuration. In such a case we have only two independent variables[16] R_{AB} and R_{BC} and the function $V(R_{AB}, R_{BC})$ may be visualized by a map quite similar to those used in geography. The map has a characteristic shape, shown in Fig. 6.2:

[16] After separating the center-of-mass motion. The separation may be done in the following way. The kinetic energy operator has the form

$$\hat{T} = -\frac{\hbar^2}{2M_A}\frac{\partial^2}{\partial X_A^2} - \frac{\hbar^2}{2M_B}\frac{\partial^2}{\partial X_B^2} - \frac{\hbar^2}{2M_C}\frac{\partial^2}{\partial X_C^2}.$$

We introduce some new coordinates:

- the center-of-mass coordinate $X_{CM} = \frac{M_A X_A + M_B X_B + M_C X_C}{M}$ with the total mass $M = M_A + M_B + M_C$,
- $R_{AB} = X_B - X_A$,
- $R_{BC} = X_C - X_B$.

To write the kinetic energy operator in the new coordinates we start with relations

$$\frac{\partial}{\partial X_A} = \frac{\partial R_{AB}}{\partial X_A}\frac{\partial}{\partial R_{AB}} + \frac{\partial X_{CM}}{\partial X_A}\frac{\partial}{\partial X_{CM}} = -\frac{\partial}{\partial R_{AB}} + \frac{M_A}{M}\frac{\partial}{\partial X_{CM}},$$

$$\frac{\partial}{\partial X_B} = \frac{\partial R_{AB}}{\partial X_B}\frac{\partial}{\partial R_{AB}} + \frac{\partial R_{BC}}{\partial X_B}\frac{\partial}{\partial R_{BC}} + \frac{\partial X_{CM}}{\partial X_B}\frac{\partial}{\partial X_{CM}} = \frac{\partial}{\partial R_{AB}} - \frac{\partial}{\partial R_{BC}} + \frac{M_B}{M}\frac{\partial}{\partial X_{CM}},$$

$$\frac{\partial}{\partial X_C} = \frac{\partial R_{BC}}{\partial X_C}\frac{\partial}{\partial R_{BC}} + \frac{\partial X_{CM}}{\partial X_C}\frac{\partial}{\partial X_{CM}} = \frac{\partial}{\partial R_{BC}} + \frac{M_C}{M}\frac{\partial}{\partial X_{CM}}.$$

After squaring these operators and substituting them into $\hat{T}$ we obtain, after a brief derivation,

$$\hat{T} = -\frac{\hbar^2}{2M}\frac{\partial^2}{\partial X_{CM}^2} - \frac{\hbar^2}{2\mu_{AB}}\frac{\partial^2}{\partial R_{AB}^2} - \frac{\hbar^2}{2\mu_{BC}}\frac{\partial^2}{\partial R_{BC}^2} + \hat{T}_{ABC},$$

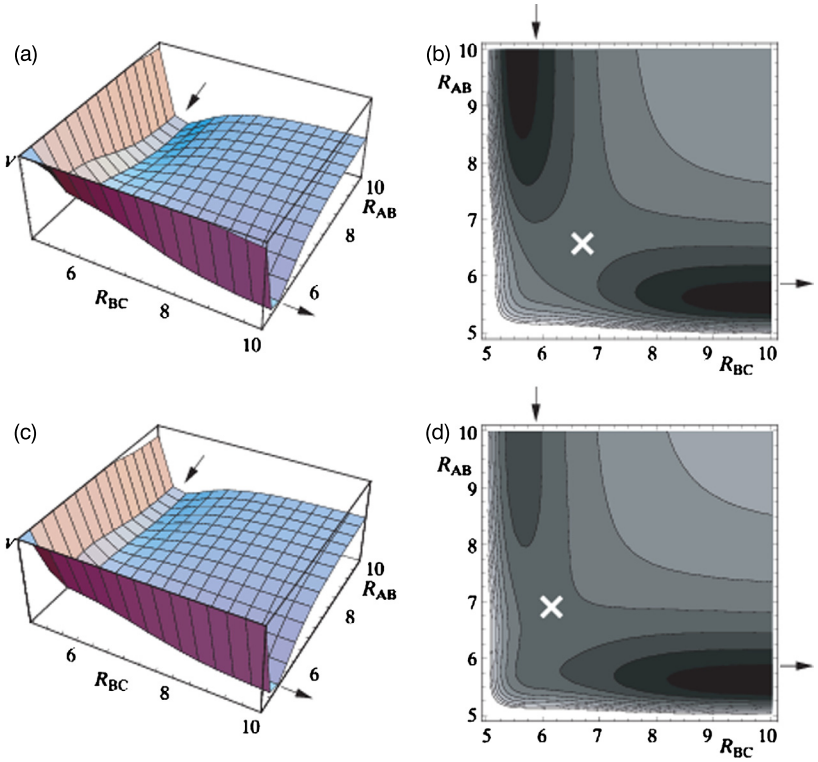


Fig. 6.2. The "drain-pipe" A + BC → AB + C (for a fictitious colinear system). The surface of the potential energy for the motion of the nuclei is a function of distances R_{AB} and R_{BC}. On the left-hand side there is the view of the surface, while on the right-hand side the corresponding maps are shown. The barrier positions are given by the crosses on the right-hand figures. (a,b) The symmetric entrance and exit channels with the separating barrier. (c,d) An exothermic reaction with the barrier in the entrance channel ("an early barrier"). (e,f) An endothermic reaction with the barrier in the exit channel ("a late barrier"). This endothermic reaction will not proceed spontaneously, because due to the equal width of the two channels, the reactant free energy is lower than the product free energy. (g,h) A spontaneous endothermic reaction, because due to the much wider exit channel (as compared to the entrance channel) the free energy is lower for the products. Note that in this particular case there is a van der Waals complex well in the entrance channel just before the barrier. There is no such well in the exit channel.

where the reduced masses $\frac{1}{\mu_{AB}} = \frac{1}{M_A} + \frac{1}{M_B}$, $\frac{1}{\mu_{BC}} = \frac{1}{M_B} + \frac{1}{M_C}$, whereas $\hat{T}_{ABC}$ stands for the mixed term $\hat{T}_{ABC} = \frac{\hbar^2}{M_B} \frac{\partial^2}{\partial R_{AB} \partial R_{BC}}$.

In this way we obtain the center-of-mass motion separation (the first term). The next two terms represent the kinetic energy operators for the independent pairs AB and BC, while the last one is the mixed term $\hat{T}_{ABC}$, whose presence is understandable: atom B participates in two motions, those associated with R_{AB} and R_{BC}. We may eventually get rid of $\hat{T}_{ABC}$ after introducing a skew coordinate system (the coordinates are determined

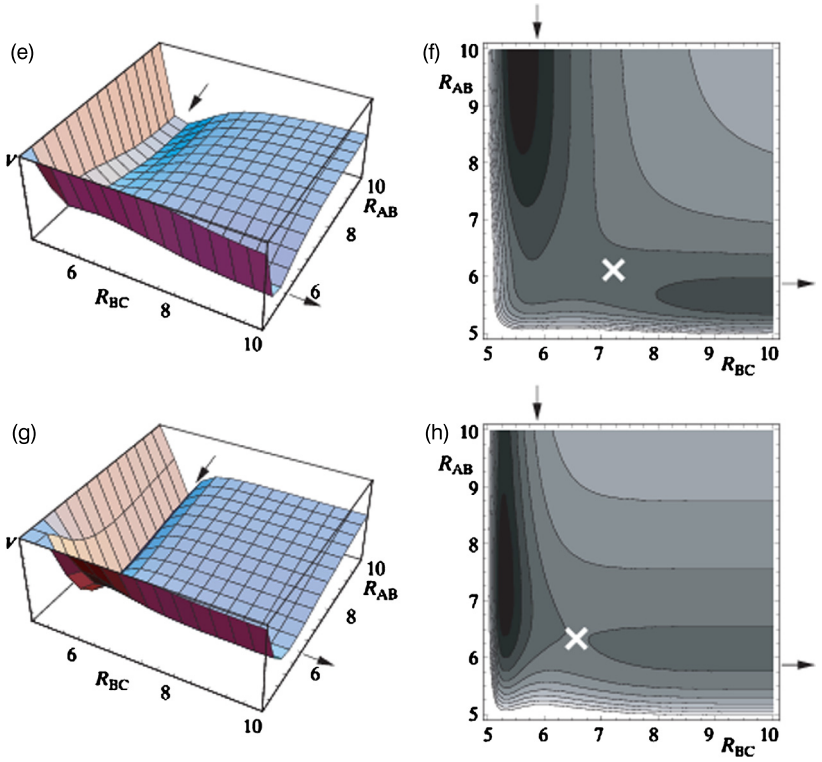


Fig. 6.2. (*continued*)

- "*Reaction map*." First of all we can see the characteristic "drain-pipe" shape of the potential energy V for the motion of the nuclei, i.e., the function $V(R_{AB}, R_{BC}) \to \infty$ for $R_{AB} \to 0$ or for $R_{BC} \to 0$; therefore, we have a high energy wall along the axes. When R_{AB} and R_{BC} are both large we have a kind of plateau that goes gently downhill towards the bottom of the curved drain-pipe, that extends nearly parallel to the axes. The chemical reaction $\mathrm{A} + \mathrm{BC} \to \mathrm{AB} + \mathrm{C}$ means a motion (shown by arrows) close to the bottom of the "drain-pipe" from a point corresponding to a large R_{AB}, while R_{BC} has a value corresponding to the equilibrium BC length to a point, corresponding to a large R_{BC} and R_{AB} with a value corresponding to the length of the isolated molecule AB.

by projections parallel to the axes). After rewriting $\hat{T}$, we obtain the following condition for the angle θ between the two axes, which ensures the mixed terms vanish: $\cos\theta_{opt} = \frac{2}{M_B}\frac{\mu_{AB}\mu_{BC}}{\mu_{AB}+\mu_{BC}}$. If all atoms have equal masses, we obtain $\theta_{opt} = 60°$.

- *Barrier*. A projection of the "drain-pipe" bottom on the R_{AB} R_{BC} plane gives the SDP. Therefore, the SDP represents one of the important features of the "landscape topography." Travel on the PES along the SDP is not a flat trip, because the drain-pipe consists of two valleys: the reactant valley (*entrance channel*) and the product valley (*exit channel*) separated by a pass (*saddle point*), which causes the reaction barrier. The saddle point corresponds to the situation in which the old chemical bond is already weakened (but still exists), while the new bond is just emerging. This explains (as has been shown by Henry Eyring, Michael Polanyi, and Meredith Evans) why the energy required to go from the entrance to the exit barrier is much smaller than the dissociation energy of BC, e.g., for the reaction $H + H_2 \rightarrow H_2 + H$ the activation energy (to overcome the reaction barrier) amounts only to about 10% of the hydrogen molecule binding energy. Simply, when the B–C bond breaks, a new bond A–B forms at the same time compensating for the energy cost needed to break the B–C bond.

The barrier may have different positions in the reaction "drain-pipe," e.g., it may be in the entrance channel (early barrier, Fig. 6.2c,d) or in the exit channel (late barrier, Fig. 6.2e,f), or it may be in between (symmetric case, Fig. 6.2a,b). The barrier position influences the course of the reaction.

When determining the SDP, kinetic energy was neglected, i.e., the motion of the point representing the system resembles a "crawling." A chemical reaction does not, however, represent any crawling over the energy hypersurface, but rather a dynamics that begins in the entrance channel and ends in the exit channel, including motion "uphill" the potential energy V. Overcoming the barrier thus is only possible when the system has an excess of kinetic energy.

What will happen if we have an early barrier? A possible reactive trajectory for such a case is shown in Figs. 6.2c and 6.3a.

It is seen that the most effective way to pass the barrier is to set the point (representing the system) in fast motion along the entrance channel. This means that atom A has to have high kinetic energy when attacking the molecule BC. After passing the barrier the point slides downhill, entering the exit channel. Since, after sliding down, it has high kinetic energy, a *bobsleigh effect* is taking place, i.e., the point climbs up the potential wall (as a result of the repulsion of atoms A and B) and then moves by making zigzags similar to a bobsleigh team. This zigzag means, of course, that strong oscillations of AB take place (and atom C leaves the rest of the system). Thus,

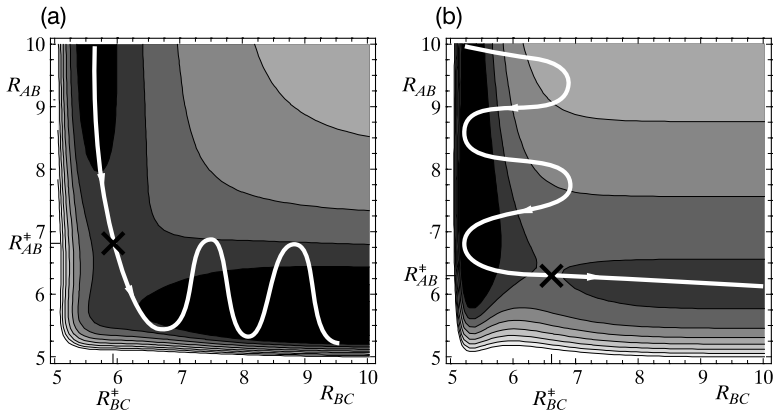


Fig. 6.3. A potential energy map for the colinear reaction A + BC → AB + C as a function of R_{AB} and R_{BC}. The distances $R^{\#}_{AB}$ and $R^{\#}_{BC}$ determine the saddle point position. (a) A reactive trajectory. If the point that represents the system runs sufficiently fast along the entrance channel towards the barrier, it will overcome the barrier by a "*charge ahead*." Then, in the exit channel the point has to oscillate, which means product vibrations. (b) A reaction with a late barrier. In the entrance channel a promising reactive trajectory is shown as the wavy line. This means the system oscillates in the entrance channel in order to be able to attack the barrier directly after passing the corner area (bobsleigh effect).

early location of a reaction barrier may result in a vibrationally excited product.

A different thing happens when the barrier is *late*. A possible reactive (i.e., successful) trajectory is shown in Fig. 6.3b. For the point to overcome the barrier it has to have a high momentum along the BC axis, because otherwise it would climb up the potential energy wall in vain as the energy cost was too large. This may happen if the point moves along a zigzag-like way *in the entrance channel* (as shown in Fig. 6.3b). This means that

to overcome a late barrier, the *vibrational excitation* of the reactant BC is effective,

because an increase in the kinetic energy of A will not produce much. Of course, the conditions for the reaction to occur matter less for high collision energies of the reactants. On the other hand, a too fast a collision may lead to the occurrence of unwanted reactions, e.g., dissociation of the system into A + B + C. Thus there is an energy window for any given reaction.

AB INITIO APPROACH

6.3 Accurate solutions (three atoms[17])

6.3.1 Coordinate system and Hamiltonian

This approach to the chemical reaction problem corresponds to point 2 on p. 447.

Jacobi coordinates

For three atoms of masses M_1, M_2, M_3, with total mass $M = M_1 + M_2 + M_3$, we may introduce the Jacobi coordinates (see p. V1-402) in three different ways (Fig. 6.4).

Each of the coordinate systems (let us label them $i, j, k = 1, 2, 3$) highlights two atoms "close" to each other and a third "distant." Now, let us choose a pair of vectors $\mathbf{r}_k, \mathbf{R}_k$ for each of the choices of the Jacobi coordinates by the following procedure ($\mathbf{X}_i$ represents the vector identifying nucleus i in a space-fixed coordinate system [SFCS], cf. Appendix V1-J). First, let us define $\mathbf{r}_k$. We write

$$\mathbf{r}_k = \frac{1}{d_k}(\mathbf{X}_j - \mathbf{X}_i), \tag{6.3}$$

where the square of the mass scaling parameter equals

$$d_k^2 = \left(1 - \frac{M_k}{M}\right)\frac{M_k}{\mu}, \tag{6.4}$$

while μ represents the reduced mass (for three masses)

$$\mu = \sqrt{\frac{M_1 M_2 M_3}{M}}. \tag{6.5}$$

Now the second vector needed for the Jacobi coordinates is chosen as

$$\mathbf{R}_k = d_k\left[\mathbf{X}_k - \frac{M_i\mathbf{X}_i + M_j\mathbf{X}_j}{M_i + M_j}\right]. \tag{6.6}$$

The three Jacobi coordinate systems are related by the following formulae (cf. Fig. 6.4):

$$\begin{pmatrix} \mathbf{r}_i \\ \mathbf{R}_i \end{pmatrix} = \begin{pmatrix} \cos\beta_{ij} & \sin\beta_{ij} \\ -\sin\beta_{ij} & \cos\beta_{ij} \end{pmatrix}\begin{pmatrix} \mathbf{r}_j \\ \mathbf{R}_j \end{pmatrix}, \tag{6.7}$$

[17] The method was generalized for an arbitrary number of atoms (D. Blume, C.H. Greene, "*Monte Carlo Hyperspherical Description of Helium Cluster Excited States*," 2000).

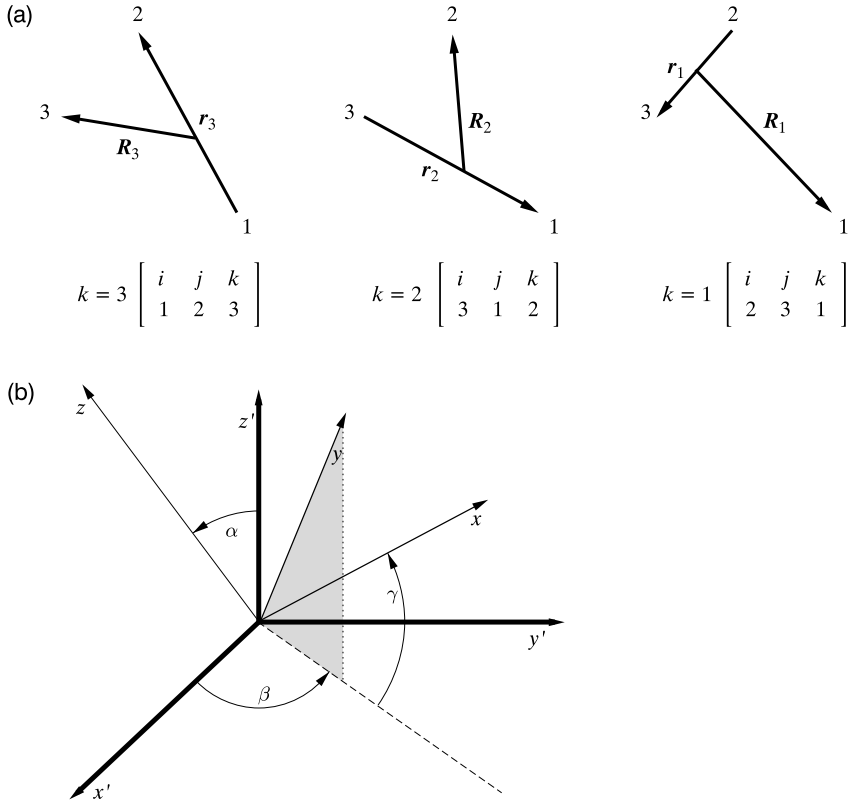


Fig. 6.4. (a) The three equivalent Jacobi coordinate systems. (b) The Euler angles show the mutual orientation of the two co-centered Cartesian coordinate systems. First, we project the y axis on the x', y' plane (the result is the dashed line). The first angle α is the angle between axes z' and z, the two other (β and γ) use the projection line described above. The relations among the coordinates are given by H. Eyring, J. Walter, G.E. Kimball, "*Quantum Chemistry*," John Wiley, New York, 1967.

$$\tan \beta_{ij} = -\frac{M_k}{\mu}, \qquad \beta_{ij} = -\beta_{ji}. \tag{6.8}$$

The Jacobi coordinates will now be used to define what is called the (more convenient) hyperspherical democratic coordinates.

Hyperspherical democratic coordinates

When a chemical reaction proceeds, the role of the atoms changes and using the same Jacobi coordinate system all the time leads to technical problems. In order not to favor any of the three

atoms despite possible differences in their masses, we introduce *hyperspherical democratic coordinates*.

First, let us define the z axis of a Cartesian coordinate system, which is perpendicular to the molecular plane at the center of mass, i.e., parallel to $\mathbf{A} = \frac{1}{2}\mathbf{r} \times \mathbf{R}$, where $\mathbf{r}$ and $\mathbf{R}$ are *any* (just as in democracy, the result is the same) of the vectors $\mathbf{r}_k, \mathbf{R}_k$. Note that by definition $|\mathbf{A}|$ represents the surface of the triangle built of the atoms. Now, let us construct the axes x and y of the rotating with molecule coordinate system (RMCS, cf. p. V1-348) in the plane of the molecule taking care that:

- the Cartesian coordinate system is right-handed,
- the axes are oriented along the main axes of the moments of inertia,[18] with $I_{yy} = \mu\left(r_y^2 + R_y^2\right) \geq I_{xx} = \mu\left(r_x^2 + R_x^2\right)$.

Finally, we introduce hyperspherical democratic coordinates equivalent to the RMCS:

- the first coordinate measures the *size of the system*, or its "radius,"

$$\rho = \sqrt{R_k^2 + r_k^2}, \tag{6.9}$$

where ρ has no subscript, because the result is independent of k (to check this use Eq. (6.7)),

- the second coordinate describes the system's *shape*,

$$\cos\theta = \frac{2\,|\mathbf{A}|}{\rho^2} \equiv u. \tag{6.10}$$

Since $|\mathbf{A}|$ is the triangular surface, $2\,|\mathbf{A}|$ means the surface of the corresponding parallelogram. The last surface (in the nominator) is compared to the surface of a square with side ρ (in the denominator; if u is small, the system is elongated like an ellipse with three atoms on its circumference)

- the third coordinate represents the angle ϕ_k for any of the atoms (in this way we determine where the k-th atom is on the ellipse), where

$$\cos\phi_k = \frac{2(\mathbf{R}_k \cdot \mathbf{r}_k)}{\rho^2 \sin\theta} \equiv \cos\phi. \tag{6.11}$$

As chosen, the hyperspherical democratic coordinates which cover all possible atomic positions within the plane $z = 0$ have the following ranges: $0 \leq \rho < \infty, 0 \leq \theta \leq \frac{\pi}{2}, 0 \leq \phi \leq 4\pi$.

[18] These directions are determined by diagonalization of the inertia moment matrix (cf. Appendix V1-L).

Hamiltonian in these coordinates

The hyperspherical democratic coordinates represent a useful alternative for the RMCS from Appendix V1-J (they themselves form another RMCS), and therefore do not depend on the orientation with respect to the body-fixed coordinate system (BFCS). However, the molecule has somehow to "be informed" that it rotates (preserving the length and the direction of the total angular momentum), because a centrifugal force acts on its parts and the Hamiltonian expressed in the BFCS (cf. Appendix V1-J) has to contain information about this rotation.

The exact kinetic energy expression for a polyatomic molecule in a space-fixed coordinate system (SFCS, cf. Appendix V1-J) has been given on p. V1-345. After separation of the center-of-mass motion, the Hamiltonian is equal to $\hat{H} = \hat{T} + V$, where V represents the electronic energy playing the role of the potential energy for the motion of the nuclei (an analog of $E_0^0(R)$ from Eq. (V1-6.8), we assume the Born–Oppenheimer approximation). In the hyperspherical democratic coordinates we obtain[19]

$$\hat{H} = -\frac{\hbar^2}{2\mu\rho^5}\frac{\partial}{\partial\rho}\rho^5\frac{\partial}{\partial\rho} + \hat{\mathcal{H}} + \hat{\mathcal{C}} + V(\rho,\theta,\phi), \tag{6.12}$$

with

$$\hat{\mathcal{H}} = \frac{\hbar^2}{2\mu\rho^2}\left[-\frac{4}{u}\frac{\partial}{\partial u}u\left(1-u^2\right)\frac{\partial}{\partial u} - \frac{1}{1-u^2}\left(4\frac{\partial^2}{\partial\phi^2} - \hat{J}_z^2\right)\right], \tag{6.13}$$

$$\hat{\mathcal{C}} = \frac{\hbar^2}{2\mu\rho^2}\left[\frac{1}{1-u^2}4i\,\hat{J}_z u\frac{\partial}{\partial\phi} + \frac{2}{u^2}\left[\hat{J}_x^2 + \hat{J}_y^2 + \sqrt{1-u^2}\left(\hat{J}_x^2 - \hat{J}_y^2\right)\right]\right], \tag{6.14}$$

where the first part and the term with $\frac{\partial^2}{\partial\phi^2}$ in $\hat{\mathcal{H}}$ represent what are called deformation terms, next there is a term with $\hat{J}_z^2$ describing the rotations about the z axis, and the terms in $\hat{\mathcal{C}}$ contain the Coriolis term (with $i\,\hat{J}_z u\frac{\partial}{\partial\phi}$).

6.3.2 Solution to the Schrödinger equation

We will need in a while some basis functions that depend on the angles θ and ϕ, preferentially each of them somehow adapted to the problem we are solving. These basis functions will be generated as the eigenfunctions of $\hat{\mathcal{H}}$ obtained at a fixed value $\rho = \rho_p$, i.e.,

$$\hat{\mathcal{H}}(\rho_p)\,\Phi_{k\Omega}(\theta,\phi;\rho_p) = \varepsilon_{k\Omega}(\rho_p)\,\Phi_{k\Omega}(\theta,\phi;\rho_p), \tag{6.15}$$

where, because of two variables θ, ϕ, we have two quantum numbers k and Ω (numbering the solutions of the equations).

[19] J.G. Frey, B.J. Howard, *Chem. Phys.*, 99(1985)415.

The total wave function that also takes into account rotational degrees of freedom (θ, ϕ) is constructed as (the quantum number $J = 0, 1, 2, ...$ determines the length of the angular momentum of the system, while the quantum number $M = -J, -J+1, ..., 0, ...J$ gives the z component of the angular momentum) a linear combination of the basis functions $U_{k\Omega} = D^{JM}_{\Omega}(\alpha, \beta, \gamma)\, \Phi_{k\Omega}(\theta, \phi; \rho_p)$, i.e.,

$$\psi^{JM} = \rho^{-\frac{5}{2}} \sum_{k\Omega} F^{J}_{k\Omega}(\rho; \rho_p)\, U_{k\Omega}(\alpha, \beta, \gamma, \theta, \phi; \rho_p), \tag{6.16}$$

where α, β, γ are the three Euler angles that define the orientation of the molecule with respect to the distant stars, $D^{JM}_{\Omega}(\alpha, \beta, \gamma)$ represent the eigenfunctions of the symmetric top,[20] $\Phi_{k\Omega}$ are the solutions to Eq. (6.15), and $F^{J}_{k\Omega}(\rho; \rho_p)$ stand for the ρ-dependent expansion coefficients, i.e., functions of ρ (centered at point ρ_p). Thanks to $D^{JM}_{\Omega}(\alpha, \beta, \gamma)$ the function ψ^{JM} is the eigenfunction of the operators $\hat{J}^2$ and $\hat{J}_z$.

In what is known as the *close coupling method* the function from Eq. (6.16) is inserted into the Schrödinger equation $\hat{H}\psi^{JM} = E_J \psi^{JM}$. Then the resulting equation is multiplied by a function $U_{k'\Omega'} = D^{JM}_{\Omega'}(\alpha, \beta, \gamma)\, \Phi_{k'\Omega'}(\theta, \phi; \rho_p)$ and integrated over angles $\alpha, \beta, \gamma, \theta, \phi$, which means taking into account all possible orientations of the molecule in space (α, β, γ) and all possible shapes of the molecule (θ, ϕ) which are allowed for a given size ρ. We obtain a set of linear equations for the unknowns $F^{J}_{k\Omega}(\rho_p; \rho)$, i.e.,

$$\rho^{-\frac{5}{2}} \sum_{k\Omega} F^{J}_{k\Omega}(\rho; \rho_p) \left\langle U_{k'\Omega'} \middle| \left(\hat{H} - E_J\right) U_{k\Omega} \right\rangle_{\omega} = 0. \tag{6.17}$$

The summation extends over some assumed set of k, Ω (the number of k, Ω pairs is equal to the number of equations). The symbol $\omega \equiv (\alpha, \beta, \gamma, \theta, \phi)$ means integration over the angles. The system of equations is solved numerically.

If, when solving the equations, we apply the boundary conditions suitable for a discrete spectrum (vanishing for $\rho = \infty$), we obtain the stationary states of the three-atomic molecule. We are interested in chemical reactions in which one atom joins a diatomic molecule, and after a while another atom flies out, leaving (after reaction) the remaining diatomic molecule. Therefore, we have to apply suitable boundary conditions. As a matter of fact we are not interested in details of the collision; we are positively interested in what comes to our detector from the spot where the reaction takes place. What may happen at a certain energy E to a given reactant state (i.e., what the product state is; such a reaction is called "*state-to-state*") is determined by the corresponding *cross-section*[21] $\sigma(E)$. The cross-section can be calculated from what is

[20] D.M. Brink, G.R. Satchler, "*Angular Momentum*," Clarendon Press, Oxford, 1975.

[21] After summing up the experimental results over all the angles, this is already to be compared with the result of the abovementioned integration over angles.

called the **S** matrix, whose elements are constructed from the coefficients $F^J_{k\Omega}\left(\rho;\rho_p\right)$ found from Eq. (6.17). The **S** matrix plays a role of an energy-dependent dispatcher: such a reactant state changes to such a product state with such and such probability.

We calculate the *reaction rate* k assuming all possible energies E of the system (satisfying the Boltzmann distribution) and taking into account that fast products arrive more often at the detector when counting per unit time as follows:

$$k = \text{const} \int dE\ E\sigma(E) \exp(-\frac{E}{k_B T}), \tag{6.18}$$

where k_B is the Boltzmann constant.

The calculated reaction rate constant k may be compared with the result of the corresponding "state-to-state" experiment.

6.3.3 Berry phase

When considering accurate quantum dynamics calculations (point 3 on p. 447) researchers met the problem of what is called the Berry phase.

In Chapter V1-6 the wave function (V1-6.9) corresponding to the adiabatic approximation was assumed. In this approximation the electronic wave function depends parametrically on the positions of the nuclei. Let us imagine we take one (or more) of the nuclei on an excursion. We set off, going slowly (in order to allow the electrons to adjust i.e. solving the Schrödinger equation), the wave function deforms, and then we are back home and put the nucleus exactly in place. Did the wave function come back exactly too? Not necessarily. By definition (cf. Chapter V1-2) a class Q function has to be a unique function of coordinates. This, however, does not pertain to a parameter we were changing. What certainly came back is the probability density $\psi_k(\mathbf{r};\mathbf{R})^*\psi_k(\mathbf{r};\mathbf{R})$, because it decides that we cannot distinguish the starting and the final situations. *The wave function itself might undergo a phase change, i.e., the starting function is equal to* $\psi_k(\mathbf{r};\mathbf{R}_0)$*, while the final function is* $\psi_k(\mathbf{r};\mathbf{R}_0)\exp(i\phi)$ *and* $\phi \neq 0$. This phase shift is called the Berry phase.[22] Did it happen or not? Sometimes we can tell.

In Fig. 6.5 the absence (a and b) and the acquiring (c and d) of the geometric (or Berry) phase is explained. The phase appears when the system makes a "trip" in configurational space (shown as a trajectory of the oriented hand print). We may make the problem of the Berry phase more familiar by taking an example from everyday life. Please put your arm down against your body

[22] The discoverers of this effect were H.C. Longuet-Higgins, U. Öpik, M.H.L. Pryce, and R.A. Sack, *Proc. Roy. Soc. London A*, 244(1958)1. The problem of this geometric phase diffused into the consciousness of physicists much later, after an article by M.V. Berry, *Proc. Roy. Soc. London A*, 392(1984)45.

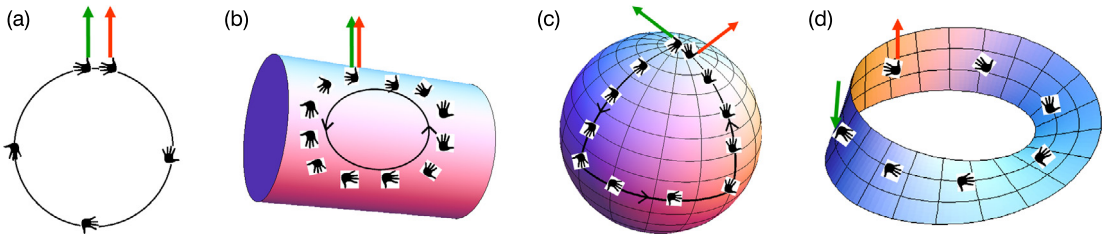


Fig. 6.5. The phase of the wave function (shown as a vector in the Hilbert space) may change when traveling (see the hand prints, the phase acquired represents the actual direction of the tumb) in the space of parameters the wave function depends on. The Berry phase is that part of the phase which is related to the topology of the parameter space. In cases (a) and (b) the Berry phase is absent (zero angle between the vectors before and after the turn), while (c) and (d) show the two vectors differing by the Berry phase after a full turn.

with the thumb directed forward. During the operations described below, please do not move the thumb with respect to the arm. Now stretch your arm horizontally sideways, rotate it to your front and then put down along your body. Note that now your thumb is not directed towards your front anymore, but towards your body. When your arm came back, the thumb had made a rotation of 90°.

Your thumb corresponds to $\psi_k(\mathbf{r};\mathbf{R})$, i.e., a vector in the Hilbert space, which is coupled with a slowly varying neighborhood ($\mathbf{R}$ corresponds to the hand positions). When the neighborhood returns, the vector may have been rotated in the Hilbert space (i.e., multiplied by a phase $\exp(i\phi)$).

Let us consider a quantum dynamics description of a chemical reaction according to point 3 on p. 447. For example, let us imagine a molecule BC fixed in space, with atom B directed to us. Now, atom A, represented by a wave packet, rushes towards atom B. We may imagine that atom A approaches the molecule and makes a bond with atom B (atom C leaves the diatomic molecule) *or* atom A may first approach atom C, then turn back, and make a bond with atom B (as before). The two possibilities correspond to two waves, which finally meet and interfere. If the phases of the two waves differed, we would see this in the results of the interference. The scientific community was surprised that some details of the reaction $H + H_2 \rightarrow H_2 + H$ at higher energies are impossible to explain without taking the Berry phase[23] into account. One of the waves described above made a turn around the conical intersection point (because it had to by-pass the equilateral triangle configuration, cf. Chapter V1-6). As shown in the work of Longuet-Higgins et al. mentioned above, this is precisely the reason why the function acquires a phase shift. We have shown in Chapter V1-6 (p. V1-371) that such a trip around a conical intersection point results in changing the phase of the function by π.

[23] Y.-S.M. Wu, A. Kupperman, *Chem. Phys. Letters*, 201(1993)178.

APPROXIMATE METHODS

6.4 Intrinsic reaction coordinate (IRC)

This section addresses point 4 of our plan from p. 447.

On p. 446 two reaction coordinates were proposed: DRC and SDP. Use of the first of them may lead to some serious difficulties (like energy discontinuities). The second reaction coordinate will undergo in a moment a useful modification and will be replaced by what is known as the intrinsic reaction coordinate (IRC).

What is the IRC?

Let us use the Cartesian coordinate system once more with $3N$ coordinates for the N nuclei; we write $X_i, i = 1, ...3N$, where X_1, X_2, X_3 denote the x, y, z coordinates of atom 1 of mass M_1, etc. The i-th coordinate is therefore associated with mass M_i of the *corresponding* atom. The classical Newtonian equation of motion for the system of N nuclei is[24]

$$M_i \ddot{X}_i = -\frac{\partial V}{\partial X_i} \quad \text{for} \quad i = 1, \ldots, 3N. \tag{6.19}$$

Let us introduce what are called *mass-weighted coordinates* (or, more precisely, weighted by the square root of mass)

$$x_i = \sqrt{M_i} X_i . \tag{6.20}$$

Our goal now will be to solve the Newton equation, but not allowing the system to acquire the kinetic energy. Instead we want the system crawling as if moving with large friction. First, we will rewrite Eq. (6.19) in the mass-weighted coordinates

$$\sqrt{M_i}\sqrt{M_i}\ddot{X}_i = -\frac{\partial V}{\partial x_i}\frac{\partial x_i}{\partial X_i} = \sqrt{M_i}\left(-\frac{\partial V}{\partial x_i}\right) \tag{6.21}$$

and obtain

$$\ddot{x}_i = -\frac{\partial V}{\partial x_i} \equiv -g_i , \tag{6.22}$$

[24] Mass × acceleration equals force; a dot over the symbol means a time derivative.

where g_i stands for the i-th component of the gradient of potential energy V calculated in the mass-weighted coordinates. To show the full motion this equation will be repeatedly integrated over t (finite but sufficiently short time to assume the gradient stays unchanged) and in such a case we get

$$\dot{x}_i = -g_i t + v_{0,i}, \tag{6.23}$$

where the initial speed is set[25] $v_{0,i} = 0$, which finally gives

$$\frac{\mathrm{d}x_i}{-g_i} = t\,\mathrm{d}t = \textit{independent of } i. \tag{6.24}$$

Thus,

> in the coordinates weighted by the square roots of the masses, a displacement of atom number i is proportional to the potential gradient (and does not depend on the atom mass).

If mass-weighted coordinates were not introduced, a displacement of the point representing the system on the potential energy map *would not follow the direction of the negative gradient or the steepest descent* (on a geographic map such a motion would look natural, because slow rivers flow this way). Indeed, the formula analogous to (6.24) would have the form $\frac{\mathrm{d}X_i}{-G_i} = \frac{t}{M_i}\mathrm{d}t$, where $G_i = \frac{\partial V}{\partial X_i}$ and therefore, during a single watch tick $\mathrm{d}t$, light atoms would travel long distances while heavy atoms would travel short distances. Thus, the motion would depend on mass, contrary to the discovery of Galileo Galilei for free fall. This means a counter-intuitive crawling across the gradient direction.

Thus, after introducing mass-weighted coordinates, the masses disappear (in formulas we may treat them as 1*).* The atomic displacements in this space will be measured in units of $\sqrt{\text{mass}}\times$length, usually in $\sqrt{u}a_0$, where $12u = {}^{12}\text{C}$ atomic mass, $u = 1822.887\ m$ (m is the electron mass), and sometimes also in units of $\sqrt{u}$Å.

Eq. (6.24) takes into account our assumption about the zero initial speed of the atom in any of the integration steps (also called *"trajectory-in-molasses"*), because otherwise we would have an additional term in $\mathrm{d}x_i$: the initial velocity times time. Shortly speaking, when the watch ticks,

[25] To keep the kinetic energy reduced to zero (crawling).

the system, represented by a point in $3N$-dimensional space, crawls over the potential energy hypersurface along the negative gradient of the hypersurface (in mass-weighted coordinates). When the system starts from a saddle point of the first order, a small deviation of the position makes the system crawl down on one or the other side of the saddle. The trajectory of the nuclei during such a motion is called the *intrinsic reaction coordinate* (IRC).

The point that represents the system slides down with infinitesimal speed along the IRC.

Measuring the travel along the IRC

In the space of the mass-weighted coordinates, trajectory IRC represents a certain curve $\mathbf{x}_{\mathrm{IRC}}$ that depends on a parameter s, i.e., $\mathbf{x}_{\mathrm{IRC}}(s)$.

The parameter s measures the length along the reaction path IRC

(e.g., in $\sqrt{u}a_0$ or $\sqrt{u}$Å). Let us take two close points on the IRC and construct the vector $\boldsymbol{\xi}(s) = \mathbf{x}_{\mathrm{IRC}}(s + \mathrm{d}s) - \mathbf{x}_{\mathrm{IRC}}(s)$. Then

$$(\mathrm{d}s)^2 = \sum_i [\xi_i(s)]^2. \tag{6.25}$$

We assume that $s = 0$ corresponds to the saddle point, $s = -\infty$ to the reactants, and $s = \infty$ to the products (Fig. 6.6).

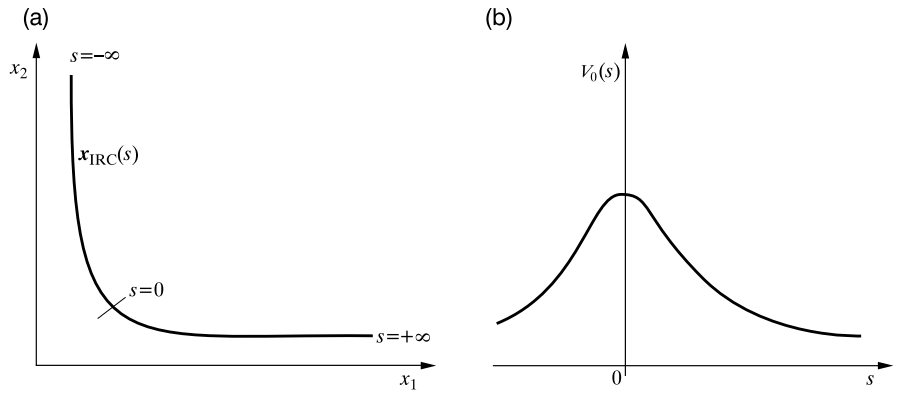


Fig. 6.6. A schematic representation of the IRC. (a) Curve $\mathbf{x}_{\mathrm{IRC}}(s)$ and (b) energy profile when moving along the IRC (i.e., curve $V_0(\mathbf{x}_{\mathrm{IRC}}(s))$) in the case of two mass-weighted coordinates x_1, x_2.

For each point on the IRC, i.e., on the curve $\mathbf{x}_{\mathrm{IRC}}(s)$, we may read the mass-weighted coordinates, and use them to calculate the coordinates of each atom. Therefore, each point on the IRC corresponds to a certain spatial and electronic structure of the system.

6.5 Reaction path Hamiltonian method

6.5.1 Energy close to IRC

A hypersurface of the potential energy represents an expensive product. We first have to calculate the potential energy for a grid of points. If we assume that ten points per coordinate is a sufficient number, then we have to perform 10^{3N-6} advanced quantum mechanical calculations; for $N = 10$ atoms this gives 10^{24} calculations, which is an impossible task. Now you see why specialists like three-atomic systems so much.

Are all the points necessary? For example, if we assume low energies, the system will in practice stay close to the IRC. Why, therefore, worry about other points? This idea was exploited by Miller, Handy, and Adams.[26] They decided to introduce the coordinates that are natural for the problem of motion in the reaction "drain-pipe." The approach corresponds to point 4 on p. 447.

The authors derived the

REACTION PATH HAMILTONIAN
an approximate expression for the energy of the reacting system in the form that stresses the existence of the IRC and of deviations from it.

This formula (*Hamilton function of the reaction path*) has the following form:

$$H(s, p_s, \{Q_k, P_k\}) = T(s, p_s, \{Q_k, P_k\}) + V(s, \{Q_k\}), \tag{6.26}$$

where T is the kinetic energy, V stands for the potential energy, s denotes the reaction coordinate along the IRC, $p_s = \frac{ds}{dt}$ represents the momentum coupled with s (mass = 1), and $\{Q_k\}$, $k = 1, 2, ...3N - 7$, stand for other coordinates orthogonal to the reaction path $\mathbf{x}_{IRC}(s)$ (this is why Q_k will depend on s) and the momenta $\{P_k\}$ conjugated with them.

We obtain the coordinates Q_k in the following way. At point s of the reaction path we diagonalize the Hessian, i.e., the matrix of the second derivatives of the potential energy, and consider all the resulting normal modes ($\omega_k(s)$ are the corresponding frequencies, cf. Chapter V1-7)

[26] W.H. Miller, N.C. Handy, J.E. Adams, *J. Chem. Phys.,* 72(1980)99.

other than that which corresponds to the reaction coordinate s (the latter corresponds to the "imaginary"[27] frequency ω_k). The diagonalization also gives the orthonormal vectors $\mathbf{L}_k(s)$, each having a direction in the $(3N-6)$-dimensional configurational space (the mass-weighted coordinate system). The coordinate $Q_k \in (-\infty, +\infty)$ measures the displacement along the direction of $\mathbf{L}_k(s)$. The coordinates s and $\{Q_k\}$ are called the *natural coordinates*. To stress that Q_k is related to $\mathbf{L}_k(s)$, we will write it as $Q_k(s)$.

The potential energy close to the IRC can be approximated (*harmonic approximation*, cf. p. V1-218) by

$$V(s, \{Q_k\}) \cong V_0(s) + \frac{1}{2}\sum_{k=1}^{3N-7} \omega_k(s)^2 Q_k(s)^2, \tag{6.27}$$

where the term $V_0(s)$ represents the potential energy that corresponds to the bottom of the reaction "drain-pipe" at a point s along the IRC, while the second term tells us what will happen to the potential energy if we displace the point (i.e., the system) perpendicular to $\mathbf{x}_{IRC}(s)$ along all the normal oscillator coordinates. In the *harmonic approximation* for the oscillator k, the energy goes up by half the force constant $\times$ the square of the normal coordinate Q_k^2. The force constant is equal to ω_k^2, because the mass is equal to 1.

The kinetic energy turns out to be more complicated. We have

$$T(s, p_s, \{Q_k, P_k\}) = \frac{1}{2}\frac{[p_s - \sum_{k=1}^{3N-7}\sum_{k'=1}^{3N-7} B_{kk'} Q_{k'} P_k]^2}{[1 + \sum_{k=1}^{3N-7} B_{ks} Q_k]^2} + \sum_{k=1}^{3N-7} \frac{P_k^2}{2}. \tag{6.28}$$

The last term is recognized as the vibrational kinetic energy for the independent oscillations perpendicular to the reaction path (recall that the mass is treated as equal to 1). If in the first term we insert $B_{kk'} = 0$ and $B_{ks} = 0$, the term would be equal to $\frac{1}{2}p_s^2$ and, therefore, would describe the kinetic energy of a point moving as if the reaction coordinate were a straight line.

[27] For large $|s|$ the corresponding ω^2 is close to zero. When $|s|$ decreases (we approach the saddle point), ω^2 becomes negative (i.e., ω is imaginary). For simplicity we will call this the "imaginary frequency" for any s.

CORIOLIS AND CURVATURE COUPLINGS

- $B_{kk'}$ are called the Coriolis coupling constants. They couple the normal modes perpendicular to the IRC.
- The B_{ks} are called the *curvature coupling constants*, because they would be equal to zero if the IRC was a straight line. They couple the translational motion along the reaction coordinate with the vibrational modes orthogonal to it.
- All the above coupling constants B depend on s.

Therefore, in the reaction path Hamiltonian we have the following quantities that characterize the reaction "drain-pipe":

- the reaction coordinate s that measures the progress of the reaction along the "drain-pipe,"
- the value $V_0(s) \equiv V_0(\mathbf{x}_{IRC}(s))$ represents the energy that corresponds to the bottom of the "drain-pipe"[28] at the reaction coordinate s,
- the width of the "drain-pipe" is characterized[29] by a set $\{\omega_k(s)\}$,
- the curvature of the "drain-pipe" is hidden in constants B; their definition will be given later in this chapter. Coefficient $B_{kk'}(s)$ tells us how normal modes k and k' are coupled together, while $B_{ks}(s)$ is responsible for a similar coupling between reaction path $\mathbf{x}_{\text{IRC}}(s)$ and vibration k perpendicular to it.

6.5.2 Vibrational adiabatic approximation

Most often when moving along the bottom of the "drain-pipe," potential energy $V_0(s)$ only changes moderately when compared to the potential energy changes the molecule undergoes when oscillating perpendicularly[30] to $\mathbf{x}_{IRC}(s)$. Simply, the valley bottom profile results from the fact that the molecule hardly holds together when moving along the reaction coordinate s, *a chemical bond breaks*, while *other bonds remain strong* and it is not so easy to stretch their strings. This suggests that there is *slow* motion along s and *fast* oscillatory motion along the coordinates Q_k.

Since we are mostly interested in the slow motion along s, we may average over the fast motion.

[28] That is, the classical potential energy corresponding to the point of the IRC given by s (this gives an idea of how the potential energy changes when walking along the IRC).

[29] A small ω corresponds to a wide valley when measured along a given normal mode coordinate ("soft" vibration), a large ω means a narrow valley ("hard" vibration).

[30] That is, when moving along the coordinates $Q_k, k = 1, 2, ...3N - 7$.

The philosophy behind the idea is that while the system moves slowly along s, it undergoes a large number of oscillations along Q_k. After such vibrational averaging the only information that remains about the oscillations are the vibrational quantum energy levels for each of the oscillators (of course, the levels will depend on s).

VIBRATIONAL ADIABATIC APPROXIMATION

The fast vibrational motions will be treated quantum mechanically and their total energy will enter the potential energy for the classical motion along s.

This approximation parallels the adiabatic approximation made in Chapter V1-6, where fast motion of electrons was separated from the slow motion of nuclei in a similar way. There after averaging over fast electronic motion, the total electronic energy became the potential energy for motion of nuclei, here after averaging over fast vibrational motion perpendicular to IRC the total vibrational energy (the energy of the corresponding harmonic oscillators in their quantum states) becomes the potential energy for the slow motion along s. This concept is called the *vibrational adiabatic approximation*.

In this approximation, to determine the stage of the reaction we give two *classical* quantities, i.e., *where* the system is on the reaction path (s) and *how fast* the system moves along the reaction path (p_s), as well as what the *quantum states* of the oscillators vibrating perpendicularly to the reaction path are (vibrational quantum number $v_k = 0, 1, 2, ...$ for each of the oscillators). Therefore, the potential energy for the (slow) motion along the reaction coordinate s is[31]

$$V_{adiab}(s; v_1, v_2, \ldots, v_{3N-7}) = V_0(s) + \sum_{k=1}^{3N-7} (v_k + \frac{1}{2})\hbar[\omega_k(s) - \omega_k(-\infty)], \qquad (6.29)$$

where we have chosen an arbitrary additive constant in the potential as equal to the vibrational energy of the reactants (with minus sign), $-\sum_{k=1}^{3N-7}(v_k + \frac{1}{2})\hbar\omega_k(-\infty)$. Note that even though $v_k = 0$ for each of the oscillators, there is a nonzero vibrational correction to the classical potential energy $V_0(s)$, because the zero-vibrational energy changes if s changes.

[31] Even if (according to the vibrational adiabatic approximation) the vibrational quantum numbers were kept constant during the reaction, their energies as depending on s through ω would change.

The vibrational adiabatic potential V_{adiab} was created for a given set of the vibrational quantum numbers v_k, fixed during the reaction process. Therefore, it is impossible to exchange energy between the vibrational modes (we assume therefore the Coriolis coupling constants $B_{kk'} = 0$), as well as between the vibrational modes and the reaction path (we assume the curvature coupling constants $B_{ks} = 0$). This would mean a change of v_k's.

From Eq. (6.29) we may draw the following conclusion.

When during the reaction the frequency of a normal mode decreases dramatically (which corresponds to breaking of a chemical bond), the square bracket becomes negative. This means that an excitation of the bond before the reaction decreases the (vibrational adiabatic) reaction barrier and the reaction rate will increase.

As a matter of fact, this is quite obvious; a vibrational excitation that engages the chemical bond to be broken already weakens the bond before the reaction.

Why do chemical reactions proceed?

Exothermic reactions

When the reactants (products) have kinetic energy higher than the barrier and the corresponding momentum p_s is large enough, with a high probability the barrier will be overcome (cf. p. V1-205). Even if the energy is lower than the barrier there is still a nonzero probability of passing to the other side because of the tunneling effect. In both cases (passing over and under the barrier) it is easier when the kinetic energy is large.

The barrier *height* is usually different for the reactions reactants → products and products → reactants (Fig. 6.1). If the barrier height is smaller for the reactants, this *may* result in an excess of the product concentration over the reactant concentration.[32] Since the reactants have higher energy than the products, the potential energy excess will change into kinetic energy[33] of the products (which is observed as a temperature increase – the reaction is exothermic). This may happen if the system has the possibility to pump the potential (i.e., electronic) energy excess into the translational and rotational degrees of freedom or to a "third body or bodies" (through collisions, e.g., with the solvent molecules) or has the possibility to emit light quanta. If the system has no such possibilities the reaction will not take place.

[32] Because a lower barrier is easier to overcome.
[33] Most often rotational energy.

Endothermic reactions

The barrier height does not always decide.

Besides the barrier height the *widths* of the entrance and exit channels also count.

For the time being let us take an example with $V_0(-\infty) = V_0(\infty)$, i.e., the barrier calculated from IRC is *the same* in both directions. Imagine a narrow entrance channel, i.e., large force constants for the oscillators, and a wide exit channel, i.e., low force constants (Figs. 6.2g,h and 6.7).

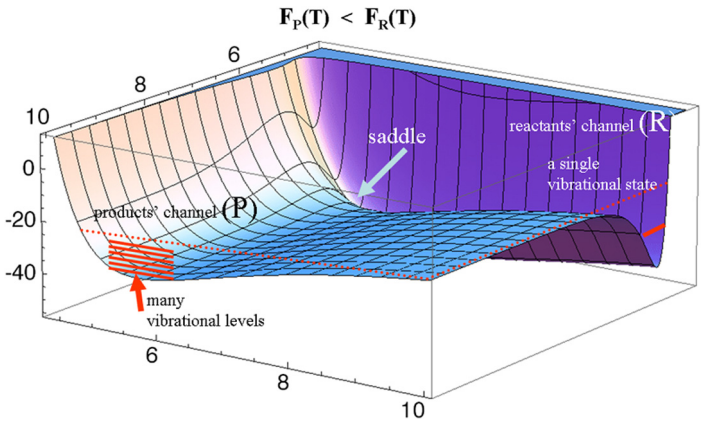


Fig. 6.7. Why do some endothermic reactions proceed spontaneously? The figure shows the reaction "drain-pipe" V. As we can see, the reactants have lower energy than the products. Yet it is not V that decides the reaction to proceed, but the free energy $F = E - TS$, where T is the temperature and S the entropy. The free energy depends on the density of the vibrational states of the reactants and products. The more numerous the low-energy vibrational levels, the larger the entropy and the lower the free energy, if $T > 0$. As we can see, the reactant vibrational levels are scarce (only one is shown), while on the product side they are densely distributed. When the energy gain related to the entropy overcomes the potential energy loss, the (endothermic) reaction will proceed spontaneously.

Note that, if the entrance channel is narrow while the exit channel is wide, the density of the vibrational states will be small in the entrance channel and large in the exit channel. Therefore, for $T > 0$ there will be a lot of possibilities to occupy the low-energy vibrational levels for the products, and only a few possibilities for the reactants. This means a high entropy of the products and a small entropy of the reactants, i.e., the products will be more stabilized by the entropy than the reactants.[34] Once again we can see that

[34] It pertains to the term $-TS$ in the free energy. In such a chemistry masterpiece as biology the endothermic reactions which are not spontaneous, are often coupled with exothermic ones. In such a way the whole process may go spontaneously.

while the energy in a spontaneous endothermic reaction increases, the decisive factor is the free energy which *decreases*. The reactants → products reaction occurs "uphill" the potential energy, but "downhill" the free energy.

Kinetic and thermodynamic pictures

- In a macroscopic reaction carried out in a chemist's flask we have a statistical ensemble of the systems that are in different microscopic stages of the reaction.
- The ensemble may be modeled by a reaction "drain-pipe" (we assume the barrier) with a lot of points, each representing one of the reacting systems.
- When the macroscopic reaction begins (e.g., we mix two reactants) a huge number of points appear in the entrance channel, i.e., we have the reactants only. As the reactant molecules assemble or dissociate the points appear or disappear.
- If the barrier were high (no tunneling) and temperature $T = 0$, all the reactants would be in their zero-vibrational states[35] and in their ground rotational states. This state would be stable even when the product valley corresponded to a lower energy (this would be a metastable state).
- Raising the temperature causes some of the ensemble elements in the entrance channel to acquire energy comparable to the barrier height. Those elements might have a chance to pass the barrier either by tunneling (for energy smaller than the barrier) or by going over the barrier. Not all of the elements with sufficient energies would pass the barrier, but only those with reactive trajectories (corresponding to a proper orientation and kinetic energy).
- After passing the barrier the energy is conserved, but in case of exothermic reactions the excess of energy changes into the translational, vibrational, and rotational energy of products or may be transferred to a "third body" (e.g., the solvent molecules) or changed into electromagnetic radiation. The latter if in the infrared region means a hot testtube – typical for the exotermic reactions. On the other hand, a spontaneous endotermic reaction takes the necessary energy from the neighborhood, which means cooling or even freezing (like when producing icecream).
- The probability of the reactive trajectories might be calculated in a way similar to that described in Chapter V1-4 (tunneling[36]), with additionally taking into account the initial vibrational and rotational states of the reactants as well as averaging over the energy level populations.
- The products would also have a chance to pass the barrier back to the reactant side (Nature does not know what we call reactants), but at the beginning the number of the elements

[35] Please note that even in this case ($T = 0$) the energy of these states would not only depend on the bottom of the valley, $V_0(s)$, but also on the valley's width through $\omega_k(s)$, according to Eq. (6.29).

[36] See also H. Eyring, J. Walter, G.E. Kimball, "*Quantum Chemistry*," John Wiley, New York, 1967.

passing the barrier in the reactant-to-product direction would be larger (nonequilibrium state).

- However, the higher the product concentration, the more often the products transform into the reactants. As an outcome we arrive at the *thermodynamic equilibrium state*, in which the average numbers of the elements passing the barrier per unit time in either direction are equal.
- If the barrier is high and the energies considered are low, then the stationary states of the system could be divided into those (of energy $E_{i,R}$) which have high amplitudes in the entrance channel (the reactant states) and those (of energy $E_{i,P}$) with high amplitudes in the exit channel (product states). In such a case we may calculate the partition function for the reactants as (we choose the same reference energy)

$$Z_R(T) = \sum_i g_i \exp\left(-\frac{E_{i,R} - E_{0,R}}{k_B T}\right)$$

and for the products as

$$Z_P(T) = \sum_i g_i \exp\left(-\frac{E_{i,P} - E_{0,R}}{k_B T}\right) = \sum_i g_i \exp\left(-\frac{E_{i,P} - E_{0,P} + \Delta E}{k_B T}\right),$$

where g_i stands for the degeneracy of the i-th energy level and the difference of the ground-state levels is $\Delta E = E_{0,P} - E_{0,R}$.

- Having the partition functions, we may calculate (at a given temperature and volume and a fixed number of particles[37]) the free or Helmholtz energy (F) corresponding to the entrance and exit channels (in thermodynamic equilibrium) as follows:

$$F_R(T) = -k_B T \frac{\partial}{\partial T} \ln Z_R(T), \tag{6.30}$$

$$F_P(T) = -k_B T \frac{\partial}{\partial T} \ln Z_P(T). \tag{6.31}$$

- The reaction goes in such a direction as to attain the minimum free energy F.
- The higher the density of states in a given channel (this corresponds to higher entropy), the lower F. The density of the vibrational states is higher for wider channels (see Fig. V1-4.16).

[37] Similar considerations may be performed for a constant pressure (instead of volume). The quantity that then attains the minimum at the equilibrium state is the Gibbs potential G.

6.5.3 Vibrational nonadiabatic model

Coriolis coupling

The vibrational adiabatic approximation is hardly justified, because the reaction channel is curved. This means that motion along s couples with some vibrational modes, and also the vibrational modes couple among themselves. We therefore have to use the nonadiabatic theory and this means we need nonzero coupling coefficients B. The Miller–Handy–Adams reaction path Hamiltonian theory gives the following expression for $B_{kk'}$ (Fig. 6.8):

$$B_{kk'}(s) = \frac{\partial \mathbf{L}_k(s)}{\partial s} \cdot \mathbf{L}_{k'}(s), \tag{6.32}$$

where $\mathbf{L}_k, k = 1, 2, ...3N - 7$ represent the orthonormal eigenvectors ($3N$-dimensional, cf. Chapter V1-7, p. V1-421) of the normal modes Q_k of frequency ω_k corresponding to s.

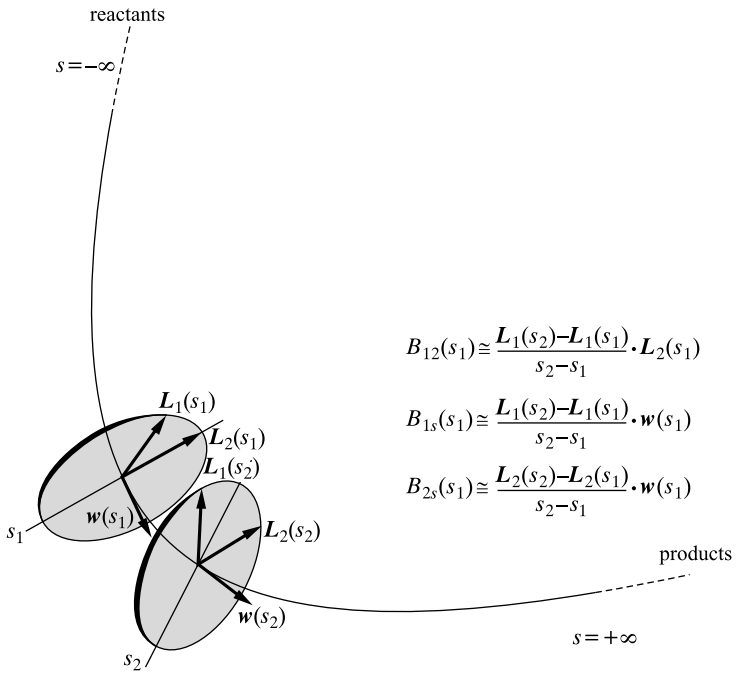


Fig. 6.8. A scheme: the Coriolis coupling coefficient (B_{12}) and the curvature coefficients (B_{1s} and B_{2s}) related to the normal modes 1 and 2 and reaction coordinate s. Diagonalization of the two Hessians calculated at points $s = s_1$ and $s = s_2$ gives two corresponding normal mode eigenvectors $\mathbf{L}_1(s_1)$ and $\mathbf{L}_2(s_1)$ as well as $\mathbf{L}_1(s_2)$ and $\mathbf{L}_2(s_2)$. At both points s_1 and s_2 we also calculate the versors $\mathbf{w}(s_1)$ and $\mathbf{w}(s_1)$ that are tangent to the IRC. The calculated vectors inserted into the formulae (on the right-hand side of the Figure) give the approximations to B_{1s}, B_{2s}, and B_{12}.

If the derivative in the above formula is multiplied by an increment of the reaction path Δs, we obtain $\frac{\partial \mathbf{L}_k(s)}{\partial s}\Delta s$, which represents a change of normal mode vector $\mathbf{L}_k$ when the system moved along the IRC by the increment Δs. This change might be similar to normal mode eigenvector $\mathbf{L}_{k'}$. This means that $B_{kk'}$ measures how much eigenvector $\mathbf{L}_{k'}(s)$ resembles the *change* of eigenvector $\mathbf{L}_k(s)$ (when the system moves a bit along the reaction path).[38] Coupling coefficient $B_{kk'}$ is usually especially large close to those values of s for which $\omega_k \cong \omega_{k'}$, i.e., for the crossing points of the vibrational frequency (or energy) curves $\omega_k(s)$. These are the points where we may expect an important energy flow from one normal mode to another, because the energy quanta match ($\hbar\omega_k(s) \cong \hbar\omega_{k'}(s)$). Coriolis coupling means that the directions of $\mathbf{L}_k$ and $\mathbf{L}_{k'}$ change when the reaction proceeds, and this resembles a rotation in the configurational space about the IRC.

Curvature couplings

Curvature coupling constant B_{ks} links the motion along the reaction valley with the normal modes orthogonal to the IRC (Fig. 6.8), i.e.,

$$B_{ks}(s) = \frac{\partial \mathbf{L}_k(s)}{\partial s} \cdot \mathbf{w}(s), \tag{6.33}$$

where $\mathbf{w}(s)$ represents the unit vector tangent to the intrinsic reaction path $\mathbf{x}_{IRC}(s)$ at point s. Coefficient $B_{ks}(s)$ therefore represents a measure of how the *change* in the normal mode eigenvector $\mathbf{L}_k(s)$ resembles a motion along the IRC. Large $B_{ks}(s)$ makes energy flow from the normal mode to the reaction path (or *vice versa*) much easier.

DONATING MODES

The modes with large $B_{ks}(s)$ in the *entrance* channel are called the *donating modes*, because an excitation of such modes makes an energy transfer to the reaction coordinate degree of freedom (an increase of the kinetic energy along the reaction path) possible. This will make the reaction rate increase.

[38] Differentiating the orthonormality condition $\mathbf{L}_k(s) \cdot \mathbf{L}_{k'}(s) = \delta_{kk'}$ we obtain

$$\frac{\partial}{\partial s}\left[\mathbf{L}_k(s) \cdot \mathbf{L}_{k'}(s)\right] =$$
$$\left[\frac{\partial \mathbf{L}_k(s)}{\partial s} \cdot \mathbf{L}_{k'}(s) + \mathbf{L}_k(s) \cdot \frac{\partial \mathbf{L}_{k'}(s)}{\partial s}\right] =$$
$$\left[\frac{\partial \mathbf{L}_k(s)}{\partial s} \cdot \mathbf{L}_{k'}(s) + \frac{\partial \mathbf{L}_{k'}(s)}{\partial s} \cdot \mathbf{L}_k(s)\right] =$$
$$B_{kk'} + B_{k'k} = 0.$$

Hence, $B_{kk'} = -B_{k'k}$.

In the vibrational adiabatic approximation, coefficients B_{ks} are equal to zero. This means that in such an approximation an exothermic reaction would not transform the net reaction energy (hidden in vibrations of the reactants' chemical bonds) into the kinetic energy of translational motion of products and vibrations of the products, what is typical for such reactions (heat production). However, as was shown by John Polanyi and Dudley Herschbach, the reactions do not go this way – a majority of the reaction energy goes into the rotational degrees of freedom (excited states of some modes). The rotations are hidden in those vibrations at $s = 0$ which are similar to the internal rotations and in the limit of $s \to +\infty$ transform into product rotations. Next, the excited products emit infrared quanta in the process of infrared fluorescence. This means that in order to have a realistic description of the reaction we have to abandon the vibrational adiabatic approximation.

6.5.4 Application of the reaction path Hamiltonian method to the reaction $H_2 + OH \to H_2O + H$

The reaction represents one of a few polyatomic systems for which precise calculations were performed.[39] It may be instructive to see how a practical implementation of the reaction path Hamiltonian method looks.

Potential energy hypersurface

The *ab initio* configuration interaction calculations (p. 134) of the potential energy hypersurface for the system under study were performed by Walsh and Dunning[40] within the Born–Oppenheimer ("clamped nuclei", p. V1-320). The electronic energy obtained as a function of the nuclear configuration plays the role of the potential energy for the motion of the nuclei. The calculation gave the electronic energy for a relatively scarce set of the configurations of the nuclei, but then the numerical results were fitted by an analytical function.[41] The IRC energy profile obtained is shown in Fig. 6.9.

It is seen from Fig. 6.9 that the barrier height for the reactants is equal to about 6.2 kcal/mol, while the reaction energy calculated as the difference of the products minus the energy of the reactants is equal to about −15.2 kcal/mol (an exothermic reaction). What happens to the atoms when the system moves along the reaction path? This is deduced from the calculated PES and shown in Fig. 6.10.

---

[39] G.C.J. Schatz, *J. Chem. Phys.*, 74(1981)113; D.G. Truhlar, A.D. Isaacson, *J. Chem. Phys.*, 77(1982)3516; A.D. Isaacson, D.G. Truhlar, *J. Chem. Phys.*, 76(1982)380; and above all the paper by Thom Dunning Jr. and Elfi Kraka in "*Advances in Molecular Electronic Structure Theory: The Calculation and Characterization of Molecular Potential Energy Surfaces*," ed. T.H. Dunning Jr., JAI Press, Inc., Greenwich, CN, (1989)1.

[40] S.P. Walsh, T.H. Dunning Jr., *J. Chem. Phys.*, 72(1980)1303.

[41] G.C. Schatz, H. Elgersma, *Chem. Phys. Letters*, 73(1980)21.

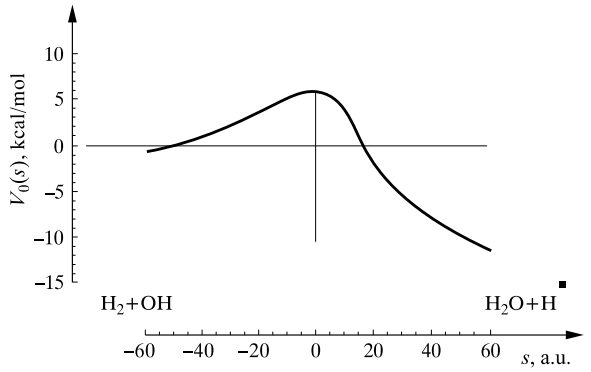


Fig. 6.9. The reaction $H_2 + OH \rightarrow H_2O + H$ energy profile $V_0(s)$ for $-\infty \leq s \leq \infty$. The value of the reaction coordinate $s = -\infty$ corresponds to the reactants, while $s = \infty$ corresponds to the products. It turns out that the product energy is lower than the energy of the reactants (i.e., the reaction is exothermic). The barrier height in the entrance channel calculated as the difference of the top of the barrier and the lowest point of the entrance channel amounts to 6.2 kcal/mol. According to T. Dunning Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*," ed. T. Dunning Jr., JAI Press, Greenwich, CN, 1989, courtesy of the authors.

The saddle point configuration of the nuclei when compared to those corresponding to the reactants and to products tells us whether the barrier is early or late. The difference of the O–H distances for the saddle point and for the product (H_2O) amounts to 0.26 Å, which represents $\frac{0.26}{0.97} = 27\%$, while the H–H distance difference for the saddle point and of the reactant (H_2) is equal to 0.11 Å, which corresponds to $\frac{0.11}{0.74} = 15\%$. In conclusion, the saddle point resembles the reactants more than the products, i.e., the barrier is early.

Normal mode analysis

Let us see what the normal mode analysis gives when performed for some selected points along the IRC. The calculated frequencies are shown in Fig. 6.11 as wave numbers $\bar{\nu} = \omega/(2\pi c)$.

As we can see, before the reaction takes place we have two normal mode frequencies ω_{HH} and ω_{OH}. When the two reacting subsystems approach one another we have to treat them as entity. The corresponding number of vibrations is $3N - 6 = 3 \times 4 - 6 = 6$ normal modes. Fig. 6.11 shows five (real) frequencies. Let us begin from the left-hand side, two of them have frequencies close to those of HH and OH, three others have frequencies close to zero and correspond to the vibrational and rotational motions of the loosely bound reactants,[42] the last "vibrational mode" is connected with a motion along the reaction path and has imaginary frequency. Such

[42] Van der Waals interactions (see Chapter 5).

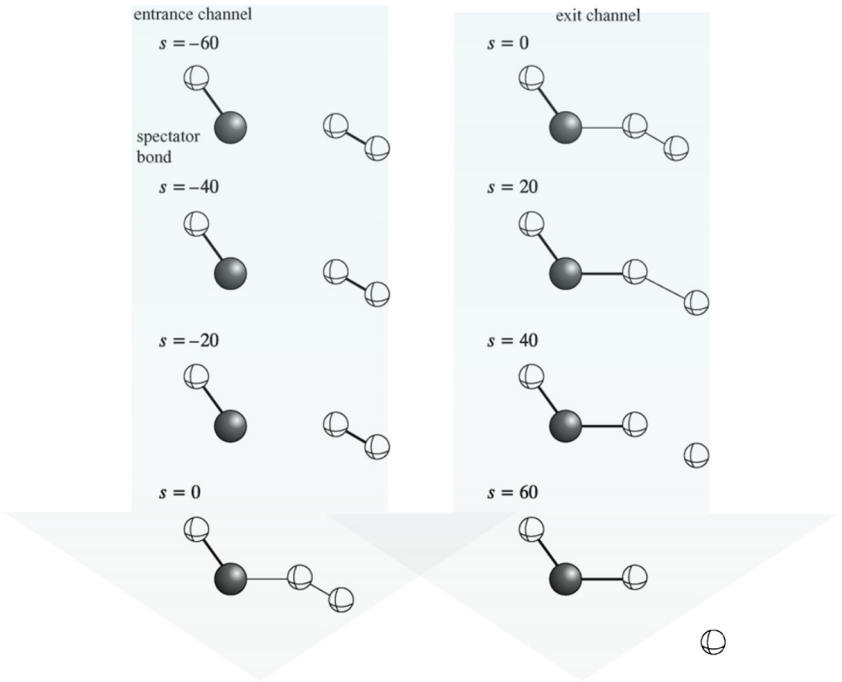


Fig. 6.10. The optimum atomic positions in the reacting system $H_2 + OH \rightarrow H_2O + H$ as functions of the reaction coordinate s. According to T. Dunning Jr. and E. Kraka, from "Advances in Molecular Electronic Structure Theory," ed. T. Dunning Jr., JAI Press, Greenwich, CN, 1989, courtesy of the authors.

a frequency means that the corresponding curvature of the potential energy is negative.[43] For example, at the saddle point, when moving along the reaction path, we have a potential energy maximum instead of minimum as would be the case for a regular oscillator.

The frequency ω_{HH} drops down close to the saddle point. This is precisely the bond to be broken. Interestingly, the frequency minimum is attained at 9 a.u. beyond the saddle point. Afterwards the frequency *increases* fast and when the reaction is completed it turns out to be the O–H symmetric stretching frequency of the water molecule. Knowing only this, we can tell what has happened: the H–H bond was broken and a new O–H bond was formed. At the end of the reaction path we have, in addition, the antisymmetric stretching mode of the H_2O, which evolved from the starting value of ω_{OH} while changing only a little (this reflects that one O–H bond exists all the time and in fact represents a "spectator" to the reaction) as well as the H–O–H bending mode, which appeared as a result of the strengthening of an intermolecular

[43] Note that $\omega = \sqrt{k/m}$, where the force constant k stands for the second derivative of the potential energy, i.e., its curvature.

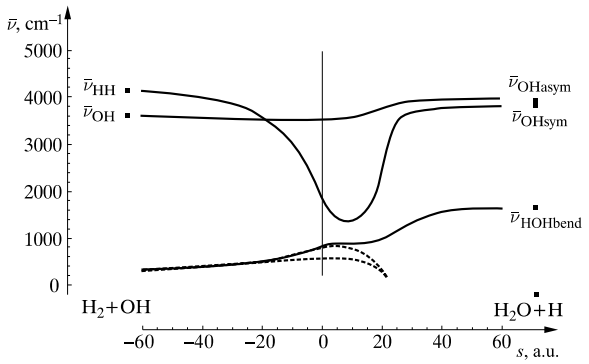


Fig. 6.11. The reaction $H_2 + OH \rightarrow H_2O + H$. The vibrational frequencies (in wave numbers $\bar{\nu} = \omega/(2\pi c)$) for the normal modes along the coordinate s. Only the real wave numbers are given (the "vibration" along s is imaginary and not given). The dots indicate the asymptotic values. According to T. Dunning Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*," ed. T. Dunning Jr., JAI Press, Greenwich, CN, 1989, courtesy of the authors.

interaction in $H_2 + OH$ when the reaction proceeded. The calculations have shown that this vibration corresponds to the *symmetric* stretching mode[44] of the H_2O. The two other modes at the beginning of the reaction have almost negligible frequencies, and after an occasional increasing of their frequencies near the saddle point end up with zero frequencies for large s. Of course, at the end we have to have $3 \times 3 - 6 = 3$ vibrational modes of the H_2O, and we do.

Example 1 (Vibrational adiabatic approximation). Let us consider several versions of the reaction that differ by assuming various vibrational states of the reactants.[45] Using Eq. (6.29), for each set of the vibrational quantum numbers we obtain the vibrational adiabatic potential V_{adiab} as a function of s (Fig. 6.12).

The adiabatic potentials obtained are instructive. It turns out that:

[44] At first sight this looks like contradicting chemical intuition since the antisymmetric mode is apparently compatible to the reaction path (one hydrogen atom being far away while the other is close to the oxygen atom). However, everything is all right. The SCF LCAO MO calculations for the water molecule within a medium-size basis set give the O–H bond length equal to 0.95 Å, whereas the O–H radical bond length is equal to 1.01 Å. This means that when the hydrogen atom approaches the O–H radical (making the water molecule), the hydrogen atom of the radical has to get *closer* to the oxygen atom. The resulting motion of both hydrogen atoms is similar to the symmetric (not antisymmetric) mode.

[45] We need the described frequencies of the modes which are orthogonal to the reaction path.

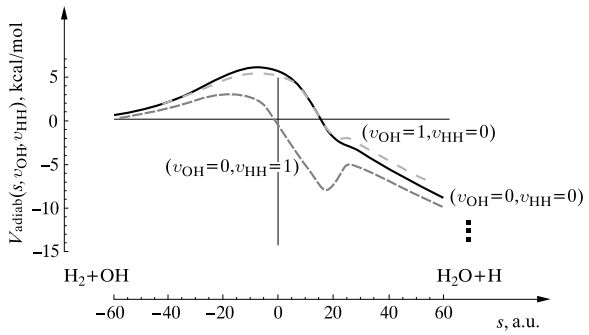


Fig. 6.12. The reaction $H_2 + OH \rightarrow H_2O + H$ (within the vibrational adiabatic approximation). Three sets of the vibrational numbers $(v_{OH}, v_{HH}) = (0, 0), (1, 0), (0, 1)$ were chosen. Note that the height and position of the barrier depend on the vibrational quantum numbers assumed. An excitation of H_2 decreases considerably the barrier height. According to T. Dunning Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*," ed. T. Dunning Jr., JAI Press, Greenwich, CN, 1989, courtesy of the authors.

- The adiabatic potential corresponding to the vibrational ground state $(v_{OH}, v_{HH}) = (0, 0)$ gives lower barrier height than the classical potential $V_0(s)$ (5.9 versus 6.1 kcal/mol). The reason for this is the lower zero-vibration energy for the saddle point configuration than for the reactants.[46]
- The adiabatic potential for the vibrational ground state has its maximum at $s = -5$ a.u., not at the saddle point $s = 0$.
- Excitation of the O–H stretching vibration does not significantly change the energy profile; in particular the barrier is lowered by only about 0.3 kcal/mol. Thus, the O–H is definitely a spectator bond.
- This contrasts with what happens when the H_2 molecule is excited. In such a case the barrier is lowered by as much as about 3 kcal/mol. This suggests that the H–H stretching vibration is a "donating mode."

Example 2 (Nonadiabatic theory). Now let us consider the vibrational nonadiabatic procedure. To do this we have to include the coupling constants B. This is done in the following way. Moving along the reaction coordinate s we perform the normal mode analysis resulting in the vibrational eigenvectors $\mathbf{L}_k(s)$. This enables us to calculate how these vectors change and to determine the derivatives $\partial \mathbf{L}_k / \partial s$. Now we may calculate the corresponding dot prod-

[46] This stands to reason, because when the Rubicon is crossed, all the bonds are weakened with respect to the reactants.

ucts (see Eqs. (6.32) and (6.33)) and obtain the coupling constants $B_{kk'}(s)$ and $B_{ks}(s)$ at each selected point s. A role of the coupling constants B in the reaction rate can be determined after dynamic studies assuming various starting conditions (the theory behind this approach will not be presented in this book). Yet some important information may be extracted just by inspecting functions $B(s)$. The functions $B_{ks}(s)$ are shown in Fig. 6.13.

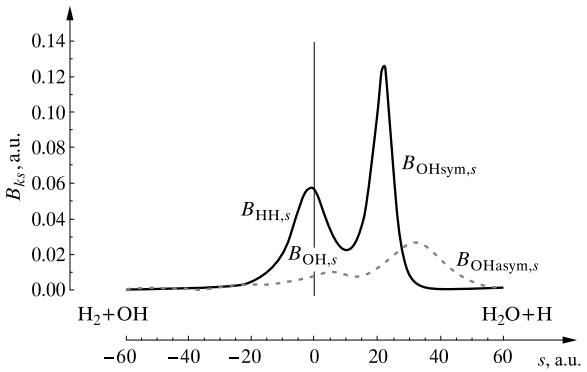


Fig. 6.13. The reaction $H_2 + OH \rightarrow H_2O + H$. The curvature coupling constants $B_{ks}(s)$ as functions of s. The $B_{ks}(s)$ characterize the coupling of the k-th normal mode with the reaction coordinate s. According to T. Dunning Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*," ed. T. Dunning Jr., JAI Press, Greenwich, CN, 1989, courtesy of the authors.

As we can see:

- In the entrance channel the value of $B_{\mathrm{OH},s}$ is close to zero; therefore, there is practically no coupling between the O–H stretching vibrations and the reaction path and hence there will be practically no energy flow between those degrees of freedom. This might be expected from a weak dependence of ω_{OH} as a function of s. Once more we see that the O–H bond plays only the role of a reaction spectator.
- This is not the case for $B_{\mathrm{HH},s}$. This quantity attains its maximum just before the saddle point (let us recall that the barrier is early). The energy may, therefore, flow from the vibrational mode of H_2 to the reaction path (and *vice versa*) and a vibrational excitation of H_2 may have an important impact on the reaction rate (recall please the lowering of the adiabatic barrier when this mode is excited).

The Coriolis coupling constants $B_{kk'}$ as functions of s are plotted in Fig. 6.14 (only for the O–H and H–H stretching and H–O–H bending modes).

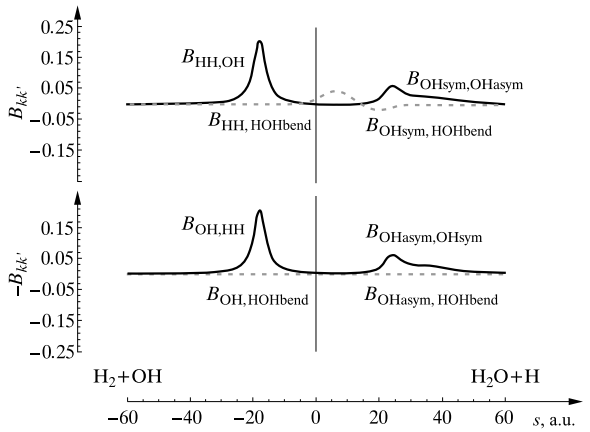


Fig. 6.14. The reaction $H_2 + OH \rightarrow H_2O + H$. The Coriolis coupling constants $B_{kk'}(s)$ as functions of s. A high value of $B_{kk'}(s)$ means that close to reaction coordinate s the changes of the k-th normal mode eigenvector resemble eigenvector k'. According to T. Dunning Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*," ed. T. Dunning Jr., JAI Press, Greenwich, CN, 1989, courtesy of the authors.

As we can see from Fig. 6.14:

- The maximum coupling for the H–H and O–H modes occurs long before the saddle point (close to $s = -18$ a.u.), enabling the system to exchange energy between the two vibrational modes.
- In the exit channel we have quite significant couplings between the symmetric and antisymmetric O–H modes and the H–O–H bending mode.

This means that formation of H–O$\cdots$H obviously influences both the O–H stretching and H–O–H bending.

6.6 Acceptor–donor (AD) theory of chemical reactions

6.6.1 A simple model of nucleophilic substitution – MO, AD, and VB formalisms

Let us take an example of a simple substitution reaction,

$$H^- + H_2 \rightarrow H_2 + H^-, \tag{6.34}$$

and consider what is known as the acceptor–donor (AD) formalism. The formalism may be treated as intermediate between the configuration interaction (CI) and the valence bond (VB) approaches (Chapter 2). Any of the three formalisms is equivalent to the two others, provided they differ only by a linear transformation of many-electron basis functions.

In the CI formalism the Slater determinants are built of the molecular spin orbitals.
In the VB formalism the Slater determinants are built of the atomic spin orbitals.
In the AD formalism the Slater determinants are built of the acceptor and donor spin orbitals.

6.6.2 MO picture → AD picture

As usual molecular orbitals for the total system φ_1, φ_2, φ_3 in a minimal basis set may be expressed (Fig. 6.15a,b) using the molecular orbital of the donor (n, in our case the $1s$ atomic orbital of H^-) and acceptor molecular orbitals (bonding χ and antibonding[47] χ^*), i.e.,

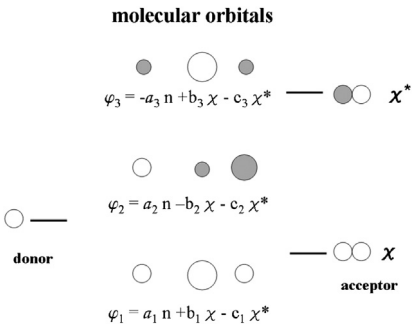


Fig. 6.15. A schematic representation of the molecular orbitals and their energies: of the donor (H^-, n representing the hydrogen atom $1s$ orbital), of the acceptor (H_2, bonding χ and antibonding χ^* of the hydrogen molecule), and of the total system H_3 in a linear configuration (center of the figure). In all cases we use the approximation that the molecular orbitals are built of the three $1s$ hydrogen atomic orbitals only.

$$\begin{aligned} \varphi_1 &= a_1 n + b_1 \chi - c_1 \chi^*, \\ \varphi_2 &= a_2 n - b_2 \chi - c_2 \chi^*, \\ \varphi_3 &= -a_3 n + b_3 \chi - c_3 \chi^*, \end{aligned} \tag{6.35}$$

where $a_i, b_i, c_i > 0$, for $i = 1, 2, 3$. This convention comes from the fact that φ_1 is of the lowest energy and therefore exhibits no node, φ_2 has to play the role of the orbital second in energy

[47] The starlet does not mean any complex conjugation. It simply says: "antibonding".

scale and therefore has a single node, while φ_3 is the highest in energy and therefore has two nodes.[48]

Any N-electron Slater determinant Ψ^{MO} composed of the molecular spin orbitals $\{\phi_i\}, i = 1, 2, ...$ (cf. Eq. (V1-N.1) on p. V1-707), may be written as a linear combination of the Slater determinants Ψ_i^{AD} composed of the spin orbitals $u_i, i = 1, 2, ...$, of the *acceptor and donor*[49] (AD) picture[50]

$$\Psi_k^{MO} = \sum_i C_k\,(i)\,\Psi_i^{AD}. \tag{6.36}$$

In a moment we will be interested in some of the coefficients $C_k\,(i)$. For example, the expansion for the ground-state Slater determinant (in the MO picture)

$$\Psi_0 = N_0\,|\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2| \tag{6.37}$$

gives

$$\Psi_0 = C_0(\mathrm{DA})\Psi_{DA} + C_0(\mathrm{D^+A^-})\Psi_{D^+A^-} + \ldots, \tag{6.38}$$

where $\bar{\varphi}_i$ denotes the spin orbital with spin function β, φ_i denotes the spin orbital with spin function α, N_0 stands for the normalization coefficient, and Ψ_{DA}, $\Psi_{D^+A^-}$ represent the normalized Slater determinants with the following electronic configurations: in Ψ_{DA}, $n^2\chi^2$; in $\Psi_{D^+A^-}$, $n^1\chi^2(\chi^*)^1$; etc.

48 Positive values of a, b, c make the node structure described above possible.

49 We start from the Slater determinant built of N molecular spin orbitals. Any of these is a linear combination of the spin orbitals of the donor and acceptor. We insert these combinations into the Slater determinant and expand the determinant according to the first row (Laplace expansion, see Appendix V1-A on p. V1-589). As a result we obtain a linear combination of the Slater determinants all having the *donor or acceptor* spin orbitals in the first row. For each of the resulting Slater determinants we repeat the procedure, but focusing on the second row, then the third row, etc. We end up with a linear combination of the Slater determinants that contain *only* the donor or acceptor spin orbitals. We concentrate on one of them, which contains some particular donor and acceptor orbitals. We are interested in the coefficient $C_k(i)$ that multiplies this Slater determinant number "i."

50 A similar expansion can also be written for the *atomic* spin orbitals (VB picture) instead of the donors and acceptors (AD picture).

Roald Hoffmann, American chemist, born in 1937 in Złoczów (then Poland) to a Jewish family, professor at Cornell University in Ithaca, USA. Hoffmann discovered the symmetry rules that pertain to some reactions of organic compounds. In 1981 he shared the Nobel Prize with Kenichi Fukui "*for their theories, developed independently, concerning the course of chemical reactions*." Roald Hoffmann is also a poet and playwright. His poetry is influenced by chemistry, in which, as he wrote, he was inspired by Marie Curie.
His CV reads as a film script. When in 1941 the Germans entered Złoczów, the four-year-old Roald was taken with his mother to a labor camp. One of the Jewish detainees betrayed a camp conspiration network to the Germans. They massacred the camp, but Roald and his mother had earlier been smuggled out of the camp by his father and hidden in a Ukrainian teacher's house. Soon after, his father was killed. The Red Army pushed the Germans out in 1944 and Roald and his mother went via Przemyśl to Cracow. In 1949 they finally reached America. Roald Hoffmann

graduated from Stuyvesant High School, Columbia University, and Harvard University. In Harvard Roald met the excellent chemist Professor Robert Burns Woodward (syntheses of chlorophyl, quinine, strychnine, and cholesterol, penicilline structure, vitamins), a Nobel Prize winner in 1965. Woodward alerted Hoffmann to a mysterious behavior of polyens in substitution reactions. Roald Hoffmann clarified the problem using the symmetry properties of the molecular orbitals (now called the Woodward–Hoffmann symmetry rules, cf. p. 501). Photo by the author.

We are mainly interested in the coefficient $C_0(\mathrm{DA})$. As shown by Fukui, Fujimoto, and Hoffmann (cf. Appendix I, p. 635),[49]

$$C_0(\mathrm{DA}) \approx \langle \Psi_{DA} | \Psi_0 \rangle = \begin{vmatrix} a_1 & b_1 \\ a_2 & -b_2 \end{vmatrix}^2 = (a_1 b_2 + a_2 b_1)^2, \tag{6.39}$$

where in the determinant, the coefficients of the donor and acceptor orbitals appear in those molecular orbitals φ_i of the total system that are occupied in the ground-state Slater determinant Ψ_0 (the coefficients of n and χ in φ_1 are a_1 and b_1, respectively, while those in φ_2 are a_2 and $-b_2$, respectively, see Eqs. (6.35)).

[49] We assume that the orbitals n, χ, and χ^* are orthogonal (approximation).

Kenichi Fukui (1918–1998), Japanese chemist, professor at Kyoto University. One of the first scholars who stressed the importance of the IRC, and introduced what is called the frontier orbitals (mainly HOMO and LUMO), which govern practically all chemical processes. Fukui received the Nobel Prize in chemistry in 1981.

Function Ψ_0 will play a prominent role in our story, but let us take now two excited states, the doubly excited configurations of the total system,[50] i.e.,

$$\Psi_{2d} = N_2|\varphi_1\bar{\varphi}_1\varphi_3\bar{\varphi}_3| \tag{6.40}$$

and

$$\Psi_{3d} = N_3|\varphi_2\bar{\varphi}_2\varphi_3\bar{\varphi}_3|, \tag{6.41}$$

where N_i stand for the normalization coefficients. Let us ask about the coefficients that *they* produce for the DA configuration (let us call these coefficients $C_2(\mathrm{DA})$ for Ψ_{2d} and $C_3(\mathrm{DA})$ for Ψ_{3d}), i.e.,

$$\Psi_{2d} = C_2(\mathrm{DA})\Psi_{DA} + C_2(\mathrm{D}^+\mathrm{A}^-)\Psi_{D^+A^-} + \ldots, \tag{6.42}$$

$$\Psi_{3d} = C_3(\mathrm{DA})\Psi_{DA} + C_3(\mathrm{D}^+\mathrm{A}^-)\Psi_{D^+A^-} + \ldots. \tag{6.43}$$

According to the result described above (see p. 635) we obtain

$$C_2(\mathrm{DA}) = \begin{vmatrix} a_1 & b_1 \\ -a_3 & b_3 \end{vmatrix}^2 = (a_1b_3 + a_3b_1)^2, \tag{6.44}$$

$$C_3(\mathrm{DA}) = \begin{vmatrix} a_2 & -b_2 \\ -a_3 & b_3 \end{vmatrix}^2 = (a_2b_3 - a_3b_2)^2. \tag{6.45}$$

[50] We will need this information later to estimate the configuration interaction role in calculating the CI ground state.

Please recall that at every reaction stage the main object of interest will be the ground state of the system. The ground state will be dominated[51] by various resonance structures. As usual the resonance structures are associated with the corresponding chemical structural formulae with the proper chemical bond pattern.

Such formulae enable us to calculate the contributions of the particular donor–acceptor resonance structures (e.g., DA, D^+A^-, etc., cf. p. 126) in the Slater determinants built of the molecular orbitals (6.35) of the total system. If one of these structures prevailed at a given stage of the reaction, this would represent important information about what has happened in the course of the reaction.

6.6.3 Reaction stages

We would like to know the a, b, c values at *various reaction stages*, because we could then calculate the coefficients C_0, C_2, and C_3 (Eqs. (6.39), (6.44), and (6.45)) for the DA as well as for other donor–acceptor structures (e.g., D^+A^-, see below) and deduce what really happens during the reaction.

Reactant stage (R)

The simplest situation is at the starting point. When H^- is far away from H–H, then of course (Fig. 6.15) $\varphi_1 = \chi$, $\varphi_2 = n$, $\varphi_3 = -\chi^*$ (the sign does not matter, we follow the convention $a, b, c > 0$). Hence, we have $b_1 = a_2 = c_3 = 1$, while the other a, b, $c = 0$; therefore,

i	a_i	b_i	c_i
1	0	1	0
2	1	0	0
3	0	0	1

Using Eqs. (6.39), (6.44), and (6.45) we get (the superscript R recalls that the results correspond to reactants)

$$C_0^R(\mathrm{DA}) = (0 \cdot 1 + 1 \cdot 1)^2 = 1, \tag{6.46}$$

$$C_2^R(\mathrm{DA}) = 0, \tag{6.47}$$

$$C_3^R(\mathrm{DA}) = (1 \cdot 0 - 0 \cdot 0)^2 = 0. \tag{6.48}$$

[51] That is, these structures will correspond to the highest expansion coefficients.

When the reaction begins, the reactants are correctly described as a Slater determinant with doubly occupied n and χ orbitals, which corresponds to the DA structure.

This is, of course, what we expected to obtain for the electronic configuration of the noninteracting reactants.

Intermediate stage (I)

What happens at the intermediate stage (I)?

It will be useful to express first the atomic orbitals $1s_a$, $1s_b$, $1s_c$ through orbitals n, χ, χ^* (they span the same space). From Chapter V1-8 (p. V1-511) we obtain

$$1s_a = n, \tag{6.49}$$

$$1s_b = \frac{1}{\sqrt{2}}\left(\chi - \chi^*\right), \tag{6.50}$$

$$1s_c = \frac{1}{\sqrt{2}}\left(\chi + \chi^*\right), \tag{6.51}$$

where we have assumed that the overlap integrals between different atomic orbitals are equal to zero.[52]

The intermediate stage corresponds to the situation in which the hydrogen atom in the middle (b) is at the same distance from a as from c, and therefore the two atoms are equivalent. This implies that the nodeless, one-node, and two-node orbitals have the following form (where ◯ stands for the $1s$ orbital and ● for the $-1s$ orbital):

$$\begin{aligned}
\varphi_1 &= ◯\,◯\,◯ = \tfrac{1}{\sqrt{3}}(1s_a + 1s_b + 1s_c),\\
\varphi_2 &= ◯\,\cdot\,● = \tfrac{1}{\sqrt{2}}(1s_a - 1s_c),\\
\varphi_3 &= ●\,◯\,● = \tfrac{1}{\sqrt{3}}(-1s_a + 1s_b - 1s_c).
\end{aligned} \tag{6.52}$$

Inserting formulae (6.50) we obtain

$$\begin{aligned}
\varphi_1 &= \tfrac{1}{\sqrt{3}}(n + \sqrt{2}\chi + 0\cdot\chi^*),\\
\varphi_2 &= \tfrac{1}{\sqrt{2}}\big(n - \tfrac{1}{\sqrt{2}}(\chi + \chi^*)\big),\\
\varphi_3 &= \tfrac{1}{\sqrt{3}}\big(-n + 0\cdot\chi - \sqrt{2}\chi^*\big)
\end{aligned} \tag{6.53}$$

[52] In reality they are small, but nonzero. Putting them zero makes it possible to write down the molecular orbitals without any calculation.

or

	a_i	b_i	c_i
$i=1$	$\frac{1}{\sqrt{3}}$	$\sqrt{\frac{2}{3}}$	0
$i=2$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$
$i=3$	$\frac{1}{\sqrt{3}}$	0	$\sqrt{\frac{2}{3}}$

(6.54)

From Eqs. (6.39), (6.44), and (6.45) we have

$$C_0^I(\mathrm{DA}) = (\frac{1}{\sqrt{3}}\frac{1}{2} + \frac{1}{\sqrt{2}}\sqrt{\frac{2}{3}})^2 = \frac{3}{4} = 0.75, \tag{6.55}$$

$$C_2^I(\mathrm{DA}) = (\frac{1}{\sqrt{3}} \cdot 0 + \sqrt{\frac{2}{3}}\frac{1}{\sqrt{3}})^2 = \frac{2}{9} = 0.22, \tag{6.56}$$

$$C_3^I(\mathrm{DA}) = (\frac{1}{\sqrt{2}} \cdot 0 - \frac{1}{2}\frac{1}{\sqrt{3}})^2 = \frac{1}{12} = 0.08. \tag{6.57}$$

The first is the most important of these three numbers. Something happens to the electronic ground state of the system. At the starting point, the ground-state wave function had a DA contribution equal to $C_0^R(\mathrm{DA}) = 1$, while now this contribution has decreased to $C_0^I(\mathrm{DA}) = 0.75$. Let us see what will happen next.

Product stage (P)

How does the reaction end up?

Let us see how molecular orbitals φ corresponding to the products are expressed by n, χ, and χ^* (they were defined for the starting point). At the end we have the molecule H–H (made of the middle and left-hand side hydrogen atoms) and the outgoing ion H^- (made of the right hydrogen atom).

Therefore, the lowest-energy orbital at the end of the reaction has the form

$$\varphi_1 = \frac{1}{\sqrt{2}}(1s_a + 1s_b) = \frac{1}{\sqrt{2}}n + \frac{1}{2}\chi - \frac{1}{2}\chi^*, \tag{6.58}$$

which corresponds to $a_1 = \frac{1}{\sqrt{2}}$, $b_1 = \frac{1}{2}$, $c_1 = \frac{1}{2}$.

Since the φ_2 orbital is identified with $1s_c$, from Eqs. (6.39), (6.44), and (6.45) we obtain $a_2 = 0$, $b_2 = c_2 = \frac{1}{\sqrt{2}}$ (never mind that all the coefficients are multiplied by -1) and finally as φ_3 we obtain the antibonding orbital

$$\varphi_3 = \frac{1}{\sqrt{2}}(1s_a - 1s_b) = \frac{1}{\sqrt{2}}n - \frac{1}{2}\chi + \frac{1}{2}\chi^*, \tag{6.59}$$

i.e., $a_3 = \frac{1}{\sqrt{2}}$, $b_3 = \frac{1}{2}$, $c_3 = \frac{1}{2}$ (the sign is reversed as well). This leads to

i	a_i	b_i	c_i
1	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$
2	0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
3	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$

(6.60)

Having a_i, b_i, c_i for the end of reaction, we may easily calculate $C_0^P(\mathrm{DA})$ of Eq. (6.39) as well as $C_2^P(\mathrm{DA})$ and $C_3^P(\mathrm{DA})$ from Eqs. (6.44) and (6.45), respectively, for the reaction products

$$C_0^P(\mathrm{DA}) = (\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + 0 \cdot \frac{1}{2})^2 = \frac{1}{4}, \tag{6.61}$$

$$C_2^P(\mathrm{DA}) = (\frac{1}{\sqrt{2}} \cdot \frac{1}{2} + \frac{1}{\sqrt{2}} \cdot \frac{1}{2})^2 = \frac{1}{2}, \tag{6.62}$$

$$C_3^P(\mathrm{DA}) = (0 \cdot \frac{1}{2} - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}})^2 = \frac{1}{4}. \tag{6.63}$$

Now we can reflect for a while. It is seen that during the reaction some important changes occur, namely,

> when the reaction begins, the system is 100% described by the structure DA, while after the reaction it resembles this structure only by 25%.

Role of the configuration interaction

We may object that our conclusions look quite naive. Indeed, there is something to worry about. We have assumed that, independently of the reaction stage, the ground-state wave function represents a single Slater determinant Ψ_0, whereas we should rather use a configuration interaction expansion. In such an expansion, besides the dominant contribution of Ψ_0, double excitations would be the most important (p. 174), which in our simple approximation of the three φ orbitals means a leading role for Ψ_{2d} and Ψ_{3d}, i.e.,

$$\Psi_{CI} = \Psi_0 + \kappa_1 \Psi_{2d} + \kappa_2 \Psi_{3d} + \ldots.$$

The two configurations are multiplied by some *small* coefficients κ (because all the time we deal with the electronic ground state dominated by Ψ_0). It will be shown that the κ coefficients in the CI expansion $\Psi = \Psi_0 + \kappa_1 \Psi_{2d} + \kappa_2 \Psi_{3d}$ are *negative*. This will serve us to make a more

detailed analysis (than that performed so far) of the role of the DA structure at the beginning and end of the reaction.

The coefficients κ_1 and κ_2 may be estimated using perturbation theory with Ψ_0 as unperturbed wave function. The first-order correction to the wave function is given by Eq. (V1-5.26) on p. V1-279, where we may safely insert the total Hamiltonian $\hat{H}$ instead of the operator[53] $\hat{H}^{(1)}$ (this frees us from saying what $\hat{H}^{(1)}$ looks like). Then we obtain

$$\kappa_1 \cong \frac{\langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle}{E_0 - E_{2d}} < 0, \tag{6.64}$$

$$\kappa_2 \cong \frac{\langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle}{E_0 - E_{3d}} < 0, \tag{6.65}$$

because from the Slater–Condon rules (Appendix V1-N) we have $\left\langle \Psi_0 | \hat{H} \Psi_{2d} \right\rangle = \langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle - \langle \varphi_2 \bar{\varphi}_2 | \bar{\varphi}_3 \varphi_3 \rangle = \langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle - 0 = \langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle$ and, similarly, $\left\langle \Psi_0 | \hat{H} \Psi_{3d} \right\rangle = \langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle$, where E_0, E_{2d}, E_{3d} represent the energies of the corresponding states. The integrals $\langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle = (\varphi_2 \varphi_2 | \varphi_3 \varphi_3)$ and $\langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle = (\varphi_1 \varphi_1 | \varphi_3 \varphi_3)$ are Coulombic *repulsions* of a certain electron density distribution with *the same* charge distribution[54]; *therefore*, $\langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle > 0$ *and* $\langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle > 0$. Therefore, indeed $\kappa_1 < 0$ and $\kappa_2 < 0$.

Thus, the contribution of the DA structure to the ground-state CI function results mainly from its contribution to the single Slater determinant Ψ_0 *(coefficient* $C_0(\mathrm{DA})$*), but is modified by a small correction* $\kappa_1 C_2(\mathrm{DA}) + \kappa_2 C_3(\mathrm{DA})$*, where* $\kappa < 0$.

What are the values of $C_2(\mathrm{DA})$ and $C_3(\mathrm{DA})$ at the beginning and at the end of the reaction? At the beginning our calculations gave $C_2^R(\mathrm{DA}) = 0$ and $C_3^R(\mathrm{DA}) = 0$. Note that $C_0^R(\mathrm{DA}) = 1$. Thus the electronic ground state at the start of the reaction mainly represents the DA structure.

And what about the end of the reaction? We have calculated that $C_2^P(\mathrm{DA}) = \frac{1}{2} > 0$ and $C_3^P(\mathrm{DA}) = \frac{1}{4} > 0$. This means that at the end of the reaction the coefficient corresponding to the DA structure will be *certainly smaller* than $C_0^P(\mathrm{DA}) = 0.25$, the value obtained for the single-determinant approximation for the ground-state wave function.

Thus,

> taking the CI expansion into account *makes our conclusion* based on the single Slater determinant even *sharper*. When the reaction starts, the wave function means the DA structure, while when it ends, this contribution is very strongly reduced.

[53] Because the unperturbed wave function Ψ_0 is an eigenfunction of the $\hat{H}^{(0)}$ Hamiltonian and is orthogonal to any of the expansion functions.

[54] $\varphi_2^* \varphi_3$ in the first case and $\varphi_1^* \varphi_3$ in the second one.

What therefore represents the ground-state wave function at the end of the reaction?

6.6.4 Contributions of the structures as the reaction proceeds

To answer this question let us consider first all possible occupations of the three energy levels (corresponding to n,χ,χ^*) by four electrons. As before we assume for the orbital energy levels $\varepsilon_\chi < \varepsilon_n < \varepsilon_{\chi^*}$. In our simple model the number of such singlet-type occupations is equal only to six (Table 6.1 and Fig. 6.16).

Table 6.1. All possible singlet-type occupations of the orbitals n, χ, and χ^* by four electrons. These occupations correspond to the resonance structures DA, D^+A^-, DA*, D^+A^{-*}, $D^{+2}A^{-2}$, and DA**.

ground state	DA	$(n)^2(\chi)^2$
singly excited state	D^+A^-	$(n)^1(\chi)^2(\chi^*)^1$
singly excited state	DA*	$(n)^2(\chi)^1(\chi^*)^1$
doubly excited state	D^+A^{-*}	$(n)^1(\chi)^1(\chi^*)^2$
doubly excited state	$D^{+2}A^{-2}$	$(\chi)^2(\chi^*)^2$
doubly excited state	DA**	$(n)^2(\chi^*)^2$

Now, let us ask what is the contribution of each of these structures[55] in Ψ_0, Ψ_{2d}, and Ψ_{3d} in the three stages of the reaction. This question is especially important for Ψ_0, because this Slater determinant is dominant for the ground-state wave function. The corresponding contributions in Ψ_{2d} and Ψ_{3d} are less important, because these configurations enter the ground-state CI wave function multiplied by the tiny coefficients κ. We have already calculated these contributions for the DA structure. The contributions of all the structures are given[56] in Table 6.2 (the largest contributions in boldface).

First, let us focus on which structures contribute mainly to Ψ_0 at the three stages of the reaction. As has been determined,

[55] We have already calculated some of these contributions.

[56] Our calculations gave $C_0^I(\mathrm{DA}) = 0.75$, $C_2^I(\mathrm{DA}) = 0.22$, $C_3^I(\mathrm{DA}) = 0.08$. In Table 6.2 these quantities are 0.729, 0.250, 0.020. The only reason for the discrepancy may be the nonzero overlap integrals, which were neglected in our calculations and were taken into account in those given in Table 6.2.

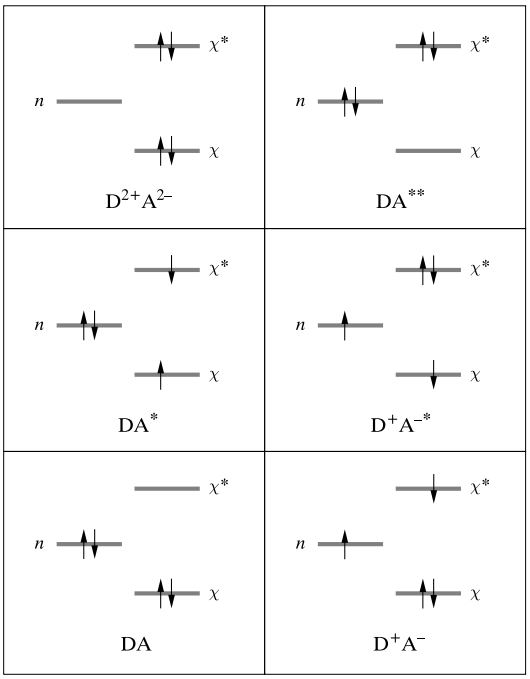


Fig. 6.16. The complete set of the six singlet wave functions ("structures") that arise from occupation of the donor orbital n and of the two acceptor orbitals (χ and χ^*).

- at point R we have only the contribution of the DA structure;
- at point I the contribution of DA decreases to 0.729, other structures come into play with the dominant D^+A^- structure (the coefficient equal to -0.604);
- at point P there are three dominant structures, i.e., D^+A^-, D^+A^{-*}, and $D^{+2}A^{-2}$.

Now we may think of going beyond the single determinant approximation by performing the CI. At the R stage the DA structure dominates as before, but has some small admixtures of DA** (because of Ψ_{3d}) and $D^{+2}A^{-2}$ (because of Ψ_{2d}), while at the product stage the contribution of the DA structure almost vanishes. Instead, some important contributions of the excited states appear, mainly of the D^+A^-, D^+A^{-*}, and $D^{+2}A^{-2}$ structures, but also other structures of lower importance.

Table 6.2. Contributions $C_0(i), C_2(i), C_3(i)$ of the six donor–acceptor structures (i) in the three Slater determinants Ψ_0, Ψ_{2d}, and Ψ_{3d} built of molecular orbitals at the three reaction stages reactant (R), intermediate (I), and product (P) (S. Shaik, *J. Am. Chem. Soc.*, 103(1981)3692). The $|C_0(i)| > 0.5$ are in boldface.

Structure	MO determinant		R	I	P
DA	Ψ_0	$C_0(\mathrm{DA})$	1	**0.729**	0.250
	Ψ_{2d}	$C_2(\mathrm{DA})$	0	0.250	0.500
	Ψ_{3d}	$C_3(\mathrm{DA})$	0	0.020	0.250
D^+A^-	Ψ_0	$c_0(D^+A^-)$	0	**−0.604**	**−0.500**
	Ψ_{2d}	$C_2(D^+A^-)$	0	0.500	0.000
	Ψ_{3d}	$C_3(D^+A^-)$	0	0.103	0.500
DA^*	Ψ_0	$C_0(DA^*)$	0	0.177	0.354
	Ψ_{2d}	$C_2(DA^*)$	0	0.354	−0.707
	Ψ_{3d}	$C_3(DA^*)$	0	0.177	0.354
D^+A^{-*}	Ψ_0	$C_0(D^+A^{-*})$	0	0.103	**0.500**
	Ψ_{2d}	$C_2(D^+A^{-*})$	0	0.500	0.000
	Ψ_{3d}	$C_3(D^+A^{-*})$	0	−0.604	−0.500
DA^{**}	Ψ_0	$C_0(DA^{**})$	0	0.021	0.250
	Ψ_{2d}	$C_2(DA^{**})$	0	0.250	0.500
	Ψ_{3d}	$C_3(DA^{**})$	1	0.729	0.250
$D^{+2}A^{-2}$	Ψ_0	$C_0(D^{+2}A^{-2})$	0	0.250	**0.500**
	Ψ_{2d}	$C_2(D^{+2}A^{-2})$	1	0.500	0.000
	Ψ_{3d}	$C_3(D^{+2}A^{-2})$	0	0.250	0.500

The value of such qualitative conclusions comes from the fact that they do not depend on the approximation used, e.g., on the atomic basis set, neglecting the overlap integrals, etc.

For example, the contributions of the six structures in Ψ_0 calculated using the Gaussian atomic basis set STO-3G and within the extended Hückel method are given in Table 6.3. Despite the fact that even the geometries used for the R, I, and P stages are slightly different, the qualitative results are the same. It is rewarding to learn things that do not depend on detail.

Table 6.3. More advanced methods find the same! Contributions $C_0(i)$ of the six donor–acceptor structures (i) in the Ψ_0 Slater determinant at three different stages (R, I, and P) of the reaction (S. Shaik, *J. Am. Chem. Soc.*, 103(1981)3692). Bold figures show the most important contributions.

Structure	STO-3G			Extended Hückel		
	R	***I***	***P***	***R***	***I***	***P***
DA	**1.000**	**0.620**	0.122	**1.000**	**0.669**	0.130
D^+A^-	0.000	**−0.410**	**−0.304**	−0.012	**−0.492**	**−0.316**
DA*	0.000	0.203	0.177	0.000	0.137	0.179
D^+A^{-*}	0.000	0.125	**0.300**	0.000	0.072	**0.298**
DA**	0.000	0.117	**0.302**	0.000	0.176	**0.301**
$D^{+2}A^{-2}$	0.000	0.035	0.120	0.000	0.014	0.166

Where do the final structures D^+A^-, D^+A^{-*}, and $D^{+2}A^{-2}$ come from?

As seen from Table 6.2, the main contributions at the end of the reaction come from the D^+A^-, D^+A^{-*}, and $D^{+2}A^{-2}$ structures. What do they correspond to *when the reaction starts*? From Table 6.2 it follows that the $D^{+2}A^{-2}$ structure simply represents Slater determinant Ψ_{2d} (Fig. 6.17). But where do the D^+A^- and D^+A^{-*} structures come from? There are no such contributions either in Ψ_0, in Ψ_{2d}, or in Ψ_{3d}. It turns out however that a similar analysis applied to the normalized configuration[57] $N|\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_3|$ at stage R gives *exclusively* the D^+A^- structure, while applied to the $N|\varphi_1\bar{\varphi}_2\varphi_3\bar{\varphi}_3|$ determinant, it gives *exclusively* the D^+A^{-*} structure (Fig. 6.17). So we have traced them back. The first of these configurations corresponds to a single-electron excitation from HOMO to LUMO – this is, therefore, the lowest excited state of the reactants. Our picture is clarified:

> the reaction starts from DA; at the intermediate stage (transition state) we have a large contribution of the first excited state that at the starting point was the D^+A^- structure related to the excitation of an electron from HOMO to LUMO.

6.6.5 *Nucleophilic attack – the model is more general: H^- + ethylene → ethylene + H^-*

Maybe the acceptor–donor theory described above pertains only to the H^- + H–H reaction? Fortunately, its applicability goes far beyond. Let us consider a nucleophilic attack S_N2 of the H^- ion on the ethylene molecule (Fig. 6.18), perpendicular to the ethylene plane towards the

[57] N stands for the normalization coefficient.

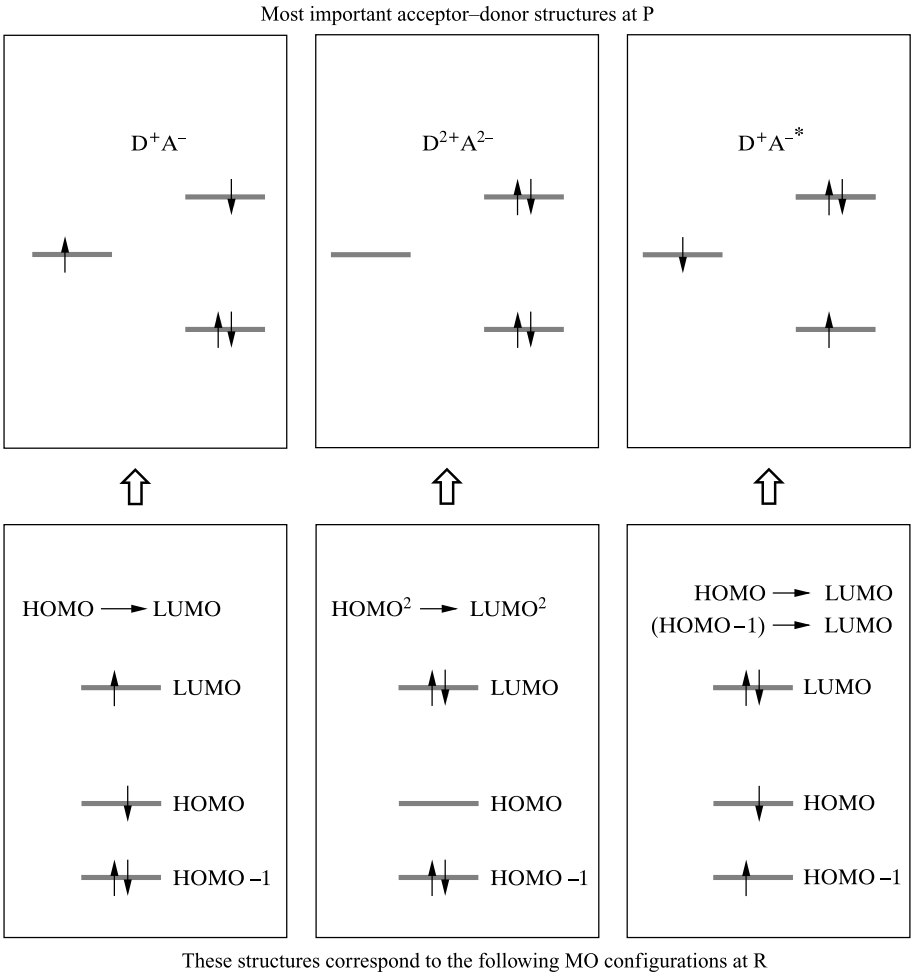


Fig. 6.17. What final structures are represented at the starting point?

position of one of the carbon atoms. The arriving ion binds to the carbon atom, forming the C–H bond, while another proton with two electrons (i.e., H^- ion) leaves the system. Such a reaction looks like having only academic interest (except that some isotopic molecules are involved, e.g., when one of the protons is replaced by a deuteron), but comprehension comes from the simplest examples possible, when the fewest things change.

The LCAO MO calculations for the ethylene molecule give the following result. The HOMO orbital is of the π bonding character, while the LUMO represents the antibonding π^* orbital (both are linear combinations of mainly carbon $2p_z$ atomic orbitals, z being the axis perpendicular to the ethylene plane). On the energy scale the H^- $1s$ orbital goes between the π and π^*

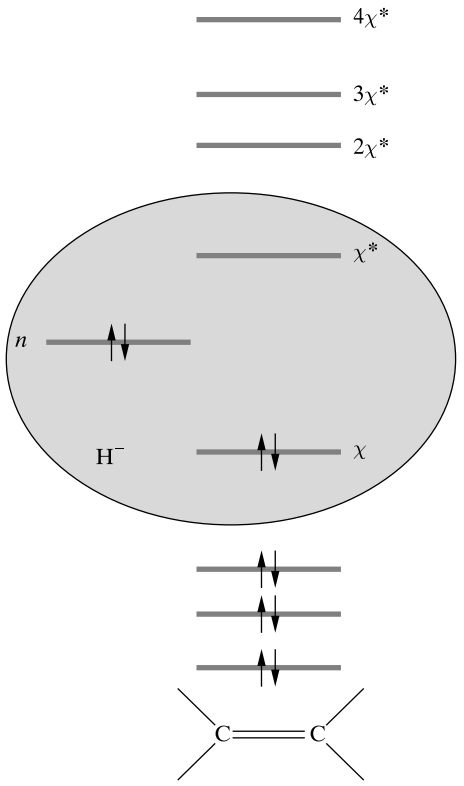


Fig. 6.18. The AD approach turns out to be more general. Nucleophilic substitution of ethylene by H^-. The figure aims to demonstrate that, despite the consideration of a more complex system than the $H^- + H_2 \rightarrow H_2 + H^-$ reaction discussed so far, the machinery behind the scene works in the same way. The attack of H^- goes perpendicularly to the ethylene plane, onto one of the carbon atoms. The figure shows the (orbital) energy levels of the donor (H^-, left-hand side) and of the acceptor (ethylene, right-hand side). Similarly as for $H^- + H_2$ the orbital energy of the donor orbital n is between the acceptor orbital energies χ and χ^* corresponding to the bonding π and antibonding π^* orbitals, respectively. Other molecular orbitals of the ethylene (occupied of low energy and virtual: $2\chi^*, 3\chi^*, ...$) play a marginal role, due to high energetic separation from the energy level of n.

energies, similarly as what happened with the χ and χ^* orbitals in the $H^- + H–H$ reaction. The virtual orbitals (let us call them $2\chi^*$, $3\chi^*$, $4\chi^*$) are far away up in the energetic scale, while the doubly occupied σ-type orbitals are far down in the energetic scale. Thus, the H^- $n = 1s$ orbital energy is close to that of χ and χ^*, while other orbitals are well separated from them.

This energy level scheme allows for many possible excitations, far more numerous than considered before. *Nevertheless*, because of the effective mixing of only those donor and acceptor orbitals that are of comparable energies, *the key partners are, as before*, n, χ, *and* χ^*. The role

of the other orbitals is only marginal: their admixtures will only slightly deform the shape of the main actors of the drama n, χ, and χ^*, which are known as the *frontier orbitals*. The coefficients for various acceptor–donor structures in the expansion of Ψ_0 are shown in Table 6.4. The calculations were performed using the extended Hückel method[58] at three stages of the reaction (R, in which the H^- ion is at a distance of 3 Å from the attacked carbon atom; I, with a distance of 1.5 Å, and P, with a distance equal to 1.08 Å; in all cases the planar geometry of the ethylene was preserved). It is seen (Table 6.4) that despite the fact that a more complex method was used, the emerging picture is basically the same (see bold digits): at the beginning the DA structure prevails, at the intermediate stage we have a "hybrid" of the DA and D^+A^- structures, while at the end we have a major role for the D^+A^- and D^+A^{-*} structures. We can also see that even if some higher excitations were taken into account (to the orbitals $2\sigma^*, 3\sigma^*$) they play only a marginal role. The corresponding population analysis (not reported here) indicates a basically identical mechanism. This resemblance extends also to the S_{N2} nucleophilic substitutions in aromatic compounds.

Table 6.4. Expansion coefficients $C_0(i)$ at the acceptor–donor structures in the ground-state wave function at various stages of the reaction: reactant (R), intermediate (I), and product (P) (S. Shaik, *J. Am. Chem. Soc.*, 103(1981)3692). The most important contributions are shown in boldface.

Structure	Coefficients		
	R	***I***	***P***
DA	**1.000**	**0.432**	0.140
D^+A^- $(n \to \pi^*)$	0.080	**0.454**	**0.380**
DA* $(\pi \to \pi^*)$	−0.110	−0.272	−0.191
D^+A^{-*} $(n \to \pi^*, \pi \to \pi^*)$	−0.006	−0.126	**−0.278**
D^+A^- $(n \to 2\sigma^*)$	$<10^{-4}$	0.006	0.004
D^+A^- $(n \to 3\sigma^*)$	$<10^{-4}$	−0.070	−0.040

6.6.6 *The model looks even more general: the electrophilic attack $H^+ + H_2 \to H_2 + H^+$*

Let us see whether this mechanism is even more general and consider the *electrophilic* substitution in the model reaction $H^+ + H\text{–}H \to H\text{–}H + H^+$. This time the role of the donor is played by the hydrogen molecule, while that of the acceptor is taken over by the proton.

[58] Introduced to chemistry by Roald Hoffmann. He uses to say that he cultivates chemistry with an old, primitive tool – a method with respect to which all other methods are better... This method, however, ensures access to the wealth of the complete Mendeleev periodic table.

The total number of electrons is only two. The DA structure corresponds to $(\chi)^2(n)^0(\chi^*)^0$. Other structures are defined by full analogy with the previous case of the H_3^- system: structure D^+A^- means $(\chi)^1(n)^1(\chi^*)^0$, structure $D^{+*}A^-$ obviously corresponds to $(\chi)^0(n)^1(\chi^*)^1$, structure D^*A to $(\chi)^1(n)^0(\chi^*)^1$, $D^{**}A$ to $(\chi)^0(n)^0(\chi^*)^2$, and $D^{+2}A^{-2}$ to $(\chi)^0(n)^2(\chi^*)^0$. As before, the ground-state Slater determinant may be expanded into the contributions of these structures. The results (the overlap neglected) are collected in Table 6.5.

Table 6.5. Expansion coefficients at the acceptor–donor structures for the reaction of proton with the hydrogen molecule at three different stages of the reaction: reactant (R), intermediate (I), and product (P) (S. Shaik, *J. Am. Chem. Soc.*, 103(1981)3692).

Structure	Coefficients		
	R	***I***	***P***
DA	**1.000**	**0.729**	0.250
D^+A^-	0	**0.604**	**0.500**
$D^{+*}A^-$	0	−0.104	**−0.500**
D^*A	0	−0.177	−0.354
$D^{**}A$	0	0.021	0.250
$D^{+2}A^{-2}$	0	0.250	0.500

It is worth stressing that we obtain the same reaction machinery as before. First, at stage R the DA structure prevails, at intermediate stage I we have a mixture of the DA and D^+A^- structures, and we end up (stage P) with D^+A^- and $D^{+*}A^-$ (the energy levels for the donor are the same as the energy levels were previously for the acceptor, hence we have $D^{+*}A^-$, and not D^+A^{-*} as before). This picture would not change qualitatively if we considered electrophilic substitution of the ethylene or benzene.

6.6.7 The model works also for the nucleophilic attack on the polarized bond

Let us consider a nucleophilic attack of the species X^- on the polarized double bond >C=Y (planarity assumed), where Y represents an atom more electronegative than carbon (say, oxygen), i.e.,

$$X^- + > C{=}Y \rightarrow > C{=}X + Y^-.$$

The problem is similar to the ethylene case except the double bond is polarized. The arguments of the kind already used for ethylene make it possible to limit ourselves exclusively to the frontier orbitals n, $\pi \equiv \chi$, and $\pi^* \equiv \chi^*$ (Figs. 6.19a and 6.20a).

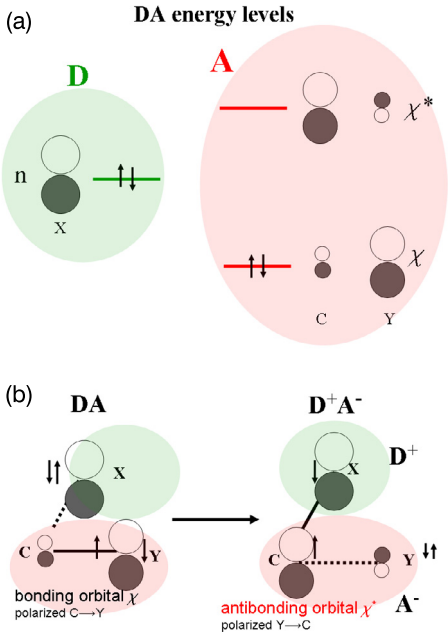


Fig. 6.19. Nucleophilic attack $X^- + >C{=}Y \rightarrow >C{=}X + Y^-$, planar configuration. The orbitals $\chi\,(=\pi)$ and $\chi^*(=\pi^*)$ are polarized. Their polarizations are opposite, the darker parts of the atomic orbitals correspond to negative values of the orbitals. (a) The orbital energy levels with the starting electronic configuration (DA structure) occupied by four electrons. (b) The nucleophilic substitution reaction S_N2 (X^- attacks within the system's plane, all orbitals shown are perpendicular to this plane). The starting DA structure has an electron pair on the π orbital (polarized bond $C \rightarrow Y$). An electron from the donor goes to the acceptor's π^* (with the opposite polarization $Y \rightarrow C$), creating the crucial intermediate D^+A^- structure. This means a large amplitude for the electron sitting on C. With the unpaired electron left on X they are ready to form the (new) C–X bond. The old bond C–Y is becoming weaker due to the antibonding character of π^*.

The bonding π orbital may be approximated as a linear combination of the $2p_z$ atomic orbitals (the z axis is perpendicular to the $>C{=}Y$ plane) of Y and C, i.e.,

$$\pi = a \cdot (2p_z)_C + b \cdot (2p_z)_Y, \tag{6.66}$$

where we assume (by convention) that the coefficients satisfy $a, b > 0$. Note that the orbital π is *polarized* this time, and due to a higher electronegativity of Y we have $b > a$ with the normalization condition $a^2 + b^2 = 1$ (we neglect the overlap integrals between the atomic orbitals). In this situation the antibonding orbital π^* may be obtained from the orthogonality condition of the orbital π as

$$\pi^* = b \cdot (2p_z)_C - a \cdot (2p_z)_Y. \tag{6.67}$$

The role of the donor orbital n will be played by $(2p_z)_X$. Note that π^* has the opposite polarization to that of π, i.e., the electron described by π^* prefers to be close to the less electronegative carbon atom.

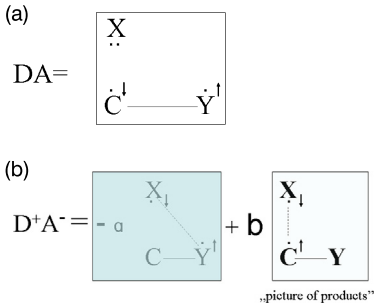


Fig. 6.20. A pictorial description of the DA and D^+A^- structures. (a) For a large donor–acceptor distance the electronic ground state is described by the DA structure. (b) Structure D^+A^- becomes very important already for the intermediate stage (I). Since $|b| \gg |a|$ the structures show the formation of the new covalent bond C–X (opposite spins of the electrons). The more important structure in (b) is visible better.

What happens at the intermediate stage, when the D^+A^- structure enters into play? In this structure one electron occupying n goes to χ^*. The Slater determinant that corresponds to this structure (one of its rows contains a spin orbital corresponding to orbital $\pi^* = b(2p_z)_C - a(2p_z)_Y$) may be expanded as a linear combination of the two Slater determinants, which instead of the spin orbital mentioned above have the corresponding atomic spin orbital ($(2p_z)_C$ or $(2p_z)_Y$). According to Eq. (6.67) we have

$$D^+A^- \equiv \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix} =$$

$$\begin{vmatrix} \pi^*(1)\alpha(1) & \pi^*(2)\alpha(2) & \dots & \pi^*(N)\alpha(N) \\ n(1)\beta(1) & n(2)\beta(2) & \dots & n(N)\beta(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix} =$$

$$\begin{vmatrix} \left[b(2p_z(1))_C - a(2p_z(1))_Y\right]\alpha(1) & \left[b(2p_z(2))_C - a(2p_z(2))_Y\right]\alpha(2) & \dots \\ n(1)\beta(1) & n(2)\beta(2) & \dots \\ \dots & \dots & \dots \end{vmatrix} =$$

$$b \begin{vmatrix} [2p_z(1)_C]\alpha(1) & [2p_z(2)]_C\alpha(2) & \dots \\ n(1)\beta(1) & n(2)\beta(2) & \dots \\ \dots & \dots & \dots \end{vmatrix}$$

$$-a \begin{vmatrix} \left[2p_z(1)_Y\right]\alpha(1) & \left[2p_z(2)_Y\right]\alpha(2) & \dots \\ n(1)\beta(1) & n(1)\beta(1) & \dots \\ \dots & \dots & \dots \end{vmatrix} =$$

$$-a\left\{\dot{\mathrm{X}}^{\downarrow}\mathrm{C} - \dot{\mathrm{Y}}^{\uparrow} \text{ structure}\right\} + b\left\{\dot{\mathrm{X}}^{\downarrow}\dot{\mathrm{C}}^{\uparrow} - \mathrm{Y} \text{ structure}\right\} \approx b\left\{\dot{\mathrm{X}}^{\downarrow}\dot{\mathrm{C}}^{\uparrow} - \mathrm{Y} \text{ structure}\right\}.$$

This is what Figs. 6.19b and 6.20b show using a pictorial representation.

Summarizing Sections 6.6.4–6.6.7 one may conclude:

- The reason for chemical reaction to occur is crossing over (in energy scale) of two energy hypersurfaces, each corresponding to a diabatic state (VB structure), from which one corresponds to the old bond pattern and the other to the new bond pattern.
- The crossing results in energy barriers when the system goes from the reactants to products or the other way.
- These two barriers have usually different heights. They represent the energy expense needed for opening the reaction channel.
- At the crossing region an electron jumps from the donor D to the acceptor A, thus creating the D^+A^- structure. We see from the VB results why the variational method has chosen the D^+A^- structure among the six possible ones: it corresponds to the new bond getting stronger and a weakening of the old bond.
- The mechanism given is general and applies wherever at least one of the reactants has a closed shell. When both the reacting molecules are of the open shell type, there will be no avoided crossing and no reaction barrier: the reactants are already prepared for the reaction.

What is going on in the chemist's flask?

Let us imagine the molecular dynamics on the energy hypersurface calculated using a quantum mechanical method (classical force fields are not appropriate since they offer nonbreakable

chemical bonds). The system is represented by a point that slides downhill (with an increasing velocity) and climbs uphill (with deceasing velocity). The system has a certain kinetic energy, because chemists usually heat their flasks.

Let us assume that first the system wanders through those regions of the ground-state hypersurface which are far in the energy scale from other electronic states. In such a situation, the adiabatic approximation (Chapter V1-6) is justified and the electronic energy (corresponding to the hypersurface) represents potential energy for the motion of the nuclei. The system corresponds to a given chemical bond pattern (we may work out a structural formula). The point representing the system in the configurational space "orbits" at the bottom of an energy well, which means the bond lengths vibrate, as do bond angles and torsional angles, but a single bond remains single, double remains double, etc.

Due to, e.g., heating the flask accidentally the system *climbs up* the wall of the potential energy well. This may mean, however, that it approaches a region of the configurational space in which another diabatic hypersurface (corresponding to another electronic state) *lowers* its energy to such an extent that the two hypersurfaces tend to intersect. In this region the adiabatic approximation fails, since we have *two* electronic states of comparable energies (both have to be taken into account), and the wave function cannot be taken as the product of an electronic function and a function describing the nuclear motion (as is required by the adiabatic approximation). As a result of mixing the crossing is avoided and two *adiabatic* hypersurfaces (upper and lower) appear. Each is composed of two parts. One part corresponds to a molecule looking as if it had one bond pattern, while the other part pertains to a different bond pattern. The bond pattern on each of the adiabatic hypersurfaces changes and the Rubicon for this change is represented by the boundary, i.e., the region of the avoided crossing that separates the two diabatic parts of the adiabatic hypersurface. Therefore, when the system in its dynamics goes uphill and enters the boundary region, the corresponding bond pattern becomes fuzzy and changes to another pattern after crossing the boundary. The reaction is completed.

What will happen next? The point representing the system in the configurational space continues to move and it may happen to arrive at another avoided-crossing region[59] and its energy is sufficient to overcome the corresponding barrier. This is the way multistep chemical reactions happen. It is important to realize that, in experiments, we have to do with an ensemble of such points rather than with a single one. The points differ by their positions (configurations of the nuclei) and nuclear momenta. Only a fraction of them has sufficiently high

[59] This new avoided crossing may turn out to be the old one. In such a case the system will cross the barrier in the opposite direction. Any chemical reaction is reversible (to different extents).

kinetic energy, proper momenta and molecules' orientation to cross the reaction barrier. The rest wander through a superbasin (composed of numerous basins) of the initial region thus undergoing vibrations, rotations (including internal rotations), etc. Of those which cross a barrier, only a fraction crosses the same barrier (i.e., the barrier of the same reaction). Others, depending on the initial conditions (nuclear positions and momenta) may cross other barriers. The art of chemistry means that in such a complicated situation it is still possible to perform reactions with nearly 100% yield and obtain a pure chemical compound – the chemist's target.

6.7 Symmetry-allowed and symmetry-forbidden reactions

6.7.1 Woodward–Hoffmann symmetry rules

The rules pertain to such an approach of two molecules that all the time some symmetry elements of the nuclear framework are preserved (there is a symmetry group associated with the reaction, see Appendix V1-C). Then:

- the molecular orbitals belong to the irreducible representations of the group;
- we assume that during the approach the orbital energies change, *but their electron occupancies do not*;
- the reaction is symmetry-allowed when the total energy (often taken as the sum of the orbital energies) lowers when going from reactants to products, otherwise it is symmetry-forbidden.

6.7.2 AD formalism

A VB point of view is simple and beautiful, but sometimes the machinery gets stuck. For example, this may happen when the described mechanism has to be rejected, because it does not meet some symmetry requirements. E.g. imagine that instead of a linear approach of H^- to H_2, we consider a T-shape configuration. In such a case the all important D^+A^- structure becomes *useless* for us, because the resonance integral which is proportional to the overlap integral between the 1*s* orbital of H^- (HOMO of the donor) and χ^* (LUMO of the acceptor) is equal to zero for symmetry reasons. If the reaction were to proceed, we would have had to form molecular orbitals from the above orbitals and this is impossible.

Yet there is an emergency exit from this situation. Let us turn our attention to the D^+A^{-*} structure, which corresponds to a doubly occupied χ^*, but a singly occupied χ. *This* structure would lead to the reaction, because the overlap integral of $1s$ H^- and χ H–H has a nonzero value. In this way,

> a forbidden symmetry will simply cause the system to choose as the lead another structure, such that it allows the formation of new bonds in *this situation*.

The above example shows that symmetry can play an important role in chemical reactions. In what follows the role of symmetry will be highlighted.

6.7.3 Electrocyclic reactions

Professor Woodward had an intriguing puzzle for young Roald Hoffmann. The puzzle had to do with cyclization of alkens (molecules with alternant double and single carbon-carbon bonds). All these molecules had at the ends the planar $H_2C=$ group and in the cyclization reaction these terminal groups had to *rotate* somehow in order to form a C–C single bond, $–H_2C–CH_2–$, closing the cycle. In the case of $H_2C=$ the direction of the rotation (conrotation or disrotation, Fig. 6.21a) is irrelevant for the final product. However, when a similar cyclization were carried out with HRC = groups ($R \neq H$), the two products possible *would differ*; these are two distinct isomers.

One might, however, presume that both products will appear, since which rotation is carried out is a matter of a stochastic process.

> The crux of the puzzle was, however, that only one of the two possible transformations took place, as if one of the rotations was forbidden, but another one allowed. It was even worse, since the allowed rotation was sometimes conrotation and sometimes disrotation, depending which alken was considered.

Roald Hoffmann solved the problem introducing the notion of symmetry-allowed and symmetry-forbidden chemical reactions. What really mattered was the symmetry of the π

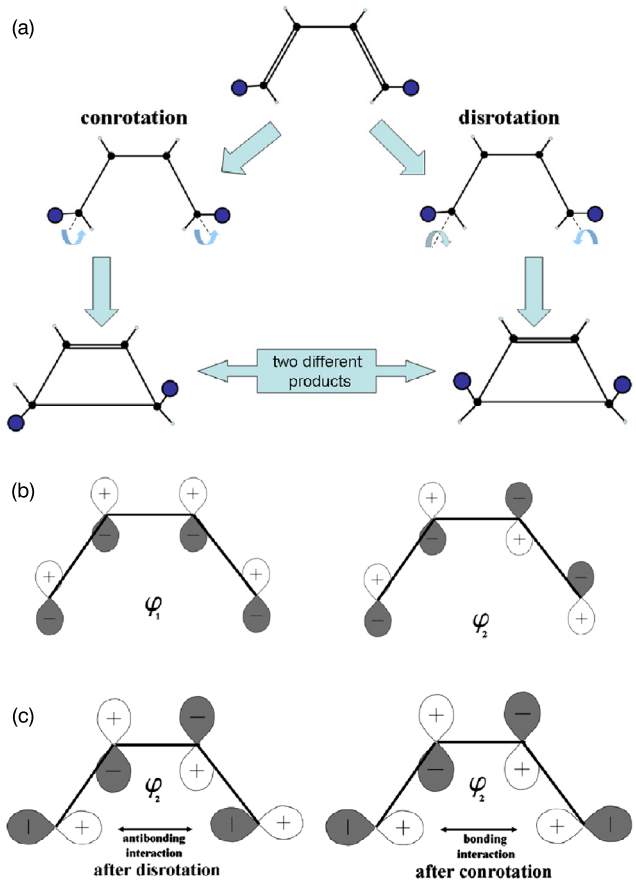


Fig. 6.21. Cyclization of cis-butadiene. (a) Start from the center: conrotation and disrotation lead in general to different products, but only one of these transformations is symmetry-allowed! (b) The doubly occupied π orbitals of the butadiene: φ_1 (HOMO-1) and φ_2 (HOMO). (c) The transformation of φ_2 under conrotation and disrotation. Conrotation is symmetry-allowed, disrotation is symmetry-forbidden.

molecular orbitals involved. These are the orbitals which certainly have to change their character when the reaction proceeds; one of them changes from π to σ. In the case of cis-butadiene we have two π orbitals (Fig. 6.21b): φ_1 and φ_2 (which are HOMO-1 and HOMO, respectively). The two electrons occupying φ_1 do not change their π character, they just go on a two-center C–C π orbital in the cyclobutene.

The two electrons occupying φ_2 change their character from π to σ. Note that they transform differently under conrotation or under disrotation (Fig. 6.21c). Conrotation preserves the C_2 axis in the carbon atoms' plane, while during disrotation the mirror plane perpendicular to the carbon atoms' plane is preserved. Conrotation results in the in-phase (i.e., *bonding*, or low-energy) overlapping of the terminal $2p$ atomic orbitals, while disrotation would lead to the out-of-phase high-energy σ^* orbital. Thus, conrotation is symmetry-allowed (because there is an important energy gain), while disrotation is forbidden (because it would correspond to an increase of the energy). For other alkens the same reasoning leads to the conclusion that $4n$ π electrons gives that conrotation is allowed (the butadiene means $n = 1$), while disrotation is forbidden, while $4n + 2$ π electrons gives that disrotation is allowed, while conrotation is forbidden. This is closely related to the type of isomer – the product of the cyclization reaction. The very reason of such a selection rule is that for the $(4n + 2)$ case the HOMO π orbital has the two terminal $2p$ orbitals in-phase, while for the $4n$ case the HOMO π orbital has the two terminal $2p$ orbitals out-of-phase (as in Fig. 6.21b).

6.7.4 Cycloaddition reaction

Let us take the example of the cycloaddition of two ethylene molecules when they bind together, forming cyclobutane. The frontier orbitals of the ground-state ethylene molecule are the doubly occupied π (HOMO) and the empty π^* (LUMO) molecular orbitals.

Fig. 6.22a shows that the reaction would go towards the products if we prepared the reacting ethylene molecules in their triplet states. Such a triplet state has to be stabilized during the reaction, while the state corresponding to the old bond pattern should lose its importance. Is it reasonable to expect the triplet state to be of low energy in order to have the chance to be pulled sufficiently down the energy scale? Yes, it is, because the triplet state arises by exciting an electron from the HOMO (i.e., π) to the LUMO (i.e., π^*), and this energy cost is the lowest possible (in the orbital picture). Within the π electron approximation the Slater determinant corresponding to the triplet state (and representing the corresponding molecular orbitals as a linear combination of the carbon $2p_z$ atomic orbitals denoted simply as a and b) has the form

$$N \det\Big(\pi(1)\alpha(1)\pi^*(2)\alpha(2)\Big) = \tag{6.68}$$

$$= N\Big(\pi(1)\alpha(1)\pi^*(2)\alpha(2) - \pi(2)\alpha(2)\pi^*(1)\alpha(1)\Big) \tag{6.69}$$

$$= N\alpha(1)\alpha(2)\Big(\big(a(1)+b(1)\big)\big(a(2)-b(2)\big) - \big(a(2)+b(2)\big)\big(a(1)-b(1)\big)\Big) \tag{6.70}$$

$$= -2N\alpha(1)\alpha(2)\Big(a(1)b(2) - a(2)b(1)\Big). \tag{6.71}$$

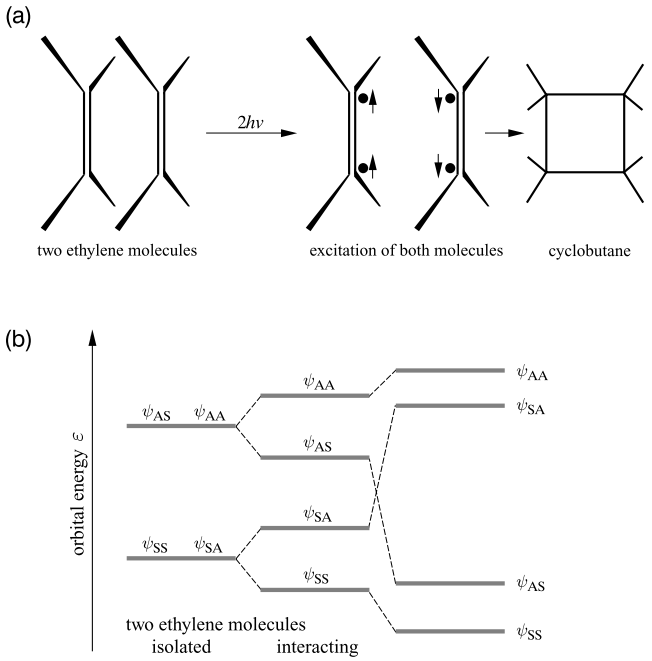


Fig. 6.22. Two equivalent schemes for the cycloaddition reaction of ethylene. (a) Two ethylene molecules, each after excitation to the triplet state, should dimerize, forming cyclobutane, because everything is prepared for electron pairing and formation of the new bonds (see text). (b)–(d) We obtain this picture from the Woodward–Hoffmann rules. According to these rules we assume that the ethylene molecules preserve two planes of symmetry, P_1 and P_2, during all stages of the reaction. We concentrate on four π electrons – the main actors in the drama. At the beginning, i.e. when the intermolecular distance is large, see (c), the lowest-energy molecular orbital of the total system (b,c) is of the SS type (i.e., symmetric with respect to P_1 and P_2). The other three orbitals (not shown in (c)) are of higher energies that increases in the following order (b): SA, AS, AA. Hence, the four electrons occupy SS and SA, (b). (d) The situation after the reaction. The four electrons are no longer of the π type, we now call them the σ type, and they occupy the hybrid orbitals shown in the figure. Once more, the lowest energy (b) corresponds to the SS symmetry orbital (d). The three others (not shown in (d)) have higher energy, but their order *is different than before* (b): AS, SA, AA. The four electrons should occupy, therefore, the SS- and AS-type orbitals, whereas (according to the Woodward–Hoffmann rule) they still occupy SS and SA. This is energetically unfavorable and such a thermic reaction does not proceed. Yet, if before the reaction the ethylene molecules were excited to the triplet state $(\pi)^1(\pi^*)^1$ (then the two molecules have AS symmetry), then at the end of the reaction they would correspond to the configuration $(SS)^2(AS)^2$, of very low energy, and the photochemical reaction proceeds.

Such a function means that when one electron is on the first carbon atom, the other is on the second carbon atom (no ionic structures). The "parallel" electron spins of one molecule may

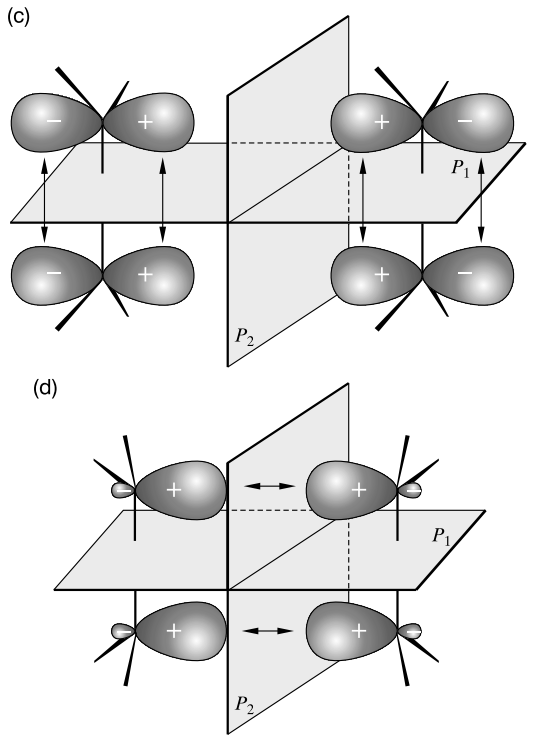


Fig. 6.22. (*continued*)

be in the opposite direction to the similar electron spins of the second molecule. Everything is prepared for the cycloaddition, i.e., formation of the new chemical bonds.

Similar conclusions can be drawn from the Woodward–Hoffmann symmetry rules.

Example 3 (Two ethylene molecules – Diels–Alder reaction). The two ethylene molecules are oriented as shown in Fig. 6.22c. Let us focus on the frontier (HOMO and LUMO) orbitals at long intermolecular distances. All are built of the symmetry orbitals composed of the four 2p carbon atomic orbitals (perpendicular to the planes corresponding to the individual molecules) and can be classified as symmetric (S) or antisymmetric (A) with respect to the symmetry planes P_1 and P_2. Moreover, by recognizing the bonding or antibonding interactions, without performing any calculations, we can tell that the SS symmetry orbital is of the lowest energy (because of the bonding character of the intra- as well as intermolecular interactions), then the SA symmetry (the bonding intramolecular – the π orbitals and the antibonding intermolecular, the intramolecular being more important) follows, next the AS symmetry (the antibonding intra- and bonding intermolecular orbitals π^*), and the highest-energy orbital AA (the antibonding

intra- and intermolecular). Since the number of electrons involved is four, they occupy the SS and SA orbitals.[60] This is what we have at the beginning of the reaction.

What do we have at the end of the reaction? At the end there are no π electrons whatsoever; instead we have two new σ chemical bonds, each built from the two sp^3 hybrids (Fig. 6.22d shows only the SS symmetry orbitals, the three other ones are not shown) oriented from the first former ethylene molecule to the other.[61] Therefore, we may form new symmetry orbitals once again, and recognize their bonding and antibonding character and hence the energetic order of their orbital energies without any calculations, just by inspection (Fig. 6.22b). The lowest energy corresponds, of course, to SS (because the newly formed σ chemical bonds correspond to the bonding combination and the lateral bonding overlap of the hybrids is also of the bonding character, Fig. 6.22d, the three other symmetry orbitals not shown), *the next in energy however is the* AS (because of the bonding interactions in the newly formed σ bonds, while the lateral orbital interaction is weakly antibonding), then follows the SA symmetry orbital (antibonding interaction along the bonds that is only slightly compensated by the lateral overlap of the hybrids), and finally, the highest energy corresponds to the totally antibonding orbital of the AA symmetry.

According to the Woodward–Hoffmann rules, the four π electrons on which we focus occupy the SS and SA orbitals from the beginning to the end of the reaction, Fig. 6.22d. This corresponds to low energy at the beginning of the reaction (R), but is very unfavorable at its end (P), because the unoccupied AS orbital is lower in the energy scale.

And what if we were smart and excited the reactants by laser? This would allow double occupation of the AS symmetry orbital right at the beginning of the reaction, leading to a low-energy configuration. To excite one electron per molecule means to put one on orbital π^*, while the second electron stays on orbital π. Of two possible spin states (singlet and triplet) the triplet state is lower in energy (see Chapter V1-8, p. V1-539). This situation is described by Eq. (6.71), and the result is that *when one electron sits on nucleus a, the other sits on b. These electrons have parallel spins – everything is prepared for the reaction*.

Therefore, the two ethylene molecules, when excited to the triplet state, open their closed shells in such a way that favors cycloaddition.

> The cycloaddition is, therefore, forbidden in the ground state (no thermally activated reaction) and allowed via an excited state (photochemistry).

[60] What a nasty historical association.

[61] We have to do with a four-member ring; therefore the sp^3 hybrids match the bond pattern only roughly.

6.7.5 Barrier means a cost of opening the closed shells

Now we can answer more precisely the question of what happens when two molecules react. When the molecules are of the closed shell character, first a change of their electronic structure has to take place. For that to happen, the kinetic energy of molecular collisions (temperature plays an important role) has to be sufficiently high in order to push and distort[62] the nuclear framework, together with the electron cloud of each of the partners (kinetic energy contra valence repulsion described in Chapter 5), to such an extent that the new configuration already corresponds to that behind the reaction barrier. For example, in the case of an electrophilic or nucleophilic attack, these changes correspond to the transformation $D \rightarrow D^+$ and $A \rightarrow A^-$, while in the case of cycloaddition they correspond to the excitation of the reacting molecules to their triplet states. *These changes make the unpaired electrons move to the proper reaction centers. As long as this state is not achieved, the changes within the molecules are small and, at most, a molecular complex forms*, in which the partners preserve their integrity and their main properties. The profound changes follow from a quasiavoided crossing of the DA diabatic hypersurface with an excited-state diabatic hypersurface, the excited state being to a large extent a "picture of the product".[63]

> Reaction barriers appear because the reactants have to open their valence shells and prepare the electronic structure to be able to form new bonds. This means their energy goes up until the "right" excited structure lowers its energy so much that the system slides down the new diabatic hypersurface to the product configuration.

The right structure means the electronic distribution in which, for each to-be-formed chemical bond, there is a set of two properly localized unpaired electrons. *The barrier height depends on the energetic gap between the starting structure and the excited state which is the "picture" of the products.* By proper distortion of the geometry (due to the valence repulsion with neighbors) we achieve a "pulling down" of the excited state mentioned, but the same distortion causes the ground state to go up. The larger the initial energy gap, the harder it is to make the two states interchange their order. The reasoning is supported by the observation that the barrier height

62 Two molecules cannot occupy the same volume due to the Pauli exclusion principle.

63 Even the noble gases open their electronic shells when subject to extreme conditions. For example, xenon atoms under a pressure of about 150 GPa change their electronic structure so much, that their famous closed shell electronic structure ceases to be the ground state. Please recall the Pauli blockade (Chapter 5). Space restrictions for an atom or molecule by the excluded volume of other atoms, i.e., mechanical pushing, lead to changes in its electronic structure. These changes may be very large under high pressure. The energy of some excited states decreases so much that the xenon atoms begin to exist in the *metallic state* (see, e.g., M.I. Eremetz, E.A. Gregoryantz, V.V. Struzhkin, H. Mao, R.J. Hemley, N. Muldero, N.M. Zimmerman, *Phys. Rev. Letters,* 85(2000)2797). The pioneers of these investigations were R. Reichlin, K.E. Brister, A.K. McMahan, M. Ross, S. Martin, Y.K. Vohra, A.L. Ruoff, *Phys. Rev. Letters*, 62(1989)669.

for electrophilic or nucleophilic attacks is roughly proportional to the *difference between the donor ionization energy and the acceptor electronic affinity, while the barrier for cycloaddition increases with the excitation energies of the donor and acceptor to their lowest triplet states*. Such relations show the great interpretative power of the AD formalism. We would not see this in the VB picture, because it would be difficult to correlate the VB structures based on the atomic orbitals with the ionization potentials or the electron affinities of the molecules involved. The best choice is to look at all three pictures (MO, AD, and VB) simultaneously. This is what we have done.

6.8 Barrier for the electron transfer reaction

In AD theory, a chemical reaction of two closed shell entities means opening their electronic shells (accompanied by an energy cost) and then forming the new bonds (accompanied by an energy gain). The electronic shell opening might have been achieved in two ways: either electron transfer from the donor to the acceptor or an excitation of each molecule to the triplet state and subsequent electron pairing between the molecules.

Now we will be interested in the barrier *height* when the first of these possibilities occurs.

6.8.1 Diabatic and adiabatic potential

Example 4 (Electron transfer in $H_2^+ + H_2$). Let us imagine two molecules, H_2^+ and H_2, in a parallel configuration[64] at distance R from one another and having identical length 1.75 a.u. (Fig. 6.23a). The value chosen is the arithmetic mean of the two equilibrium separations (2.1 a.u. for H_2^+, 1.4 a.u. for H_2).

There are two geometry parameters to change (Fig. 6.23): the length q_L of the left molecule and the length q_R of the right molecule. Instead of these two variables we may consider the other two: their sum and their difference. Since our goal is to be as simple as possible, we will assume[65] that $q_L + q_R = \text{const}$, and therefore the geometry of the total nuclear framework may be described by a single variable, i.e., $q = q_R - q_L$, with $q \in (-\infty, \infty)$.

First we assume that the extra electron is on the left-hand side molecule, so we have the $H_2 \cdots H_2^+$ system. It is quite easy to imagine what happens when q changes from 0 (i.e., from both bonds of equal length) to a somewhat larger value. Variable $q = q_R - q_L > 0$ means that $q_R > q_L$; therefore, when q increases a bit, the energy of the system will decrease, because the H_2^+ molecule elongates, while the H_2 molecule shortens, i.e., both approach their equilibrium

64 We freeze all the translations and rotations.

65 The assumption stands to reason, because a shortening of one molecule will be accompanied by an almost identical lengthening of the other when they exchange an electron.

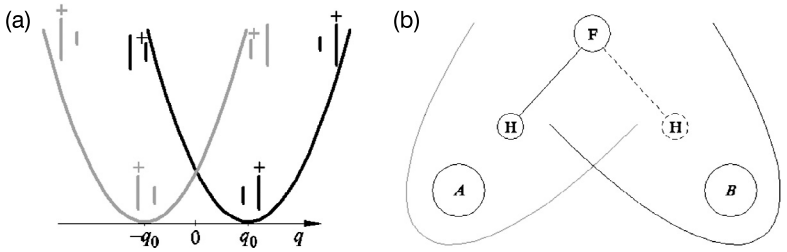


Fig. 6.23. An electron transfer is accompanied by a geometry change. (a) The black curve corresponds to the system $H_2 \cdots H_2^+$ (i.e., the extra electron is on the left-hand side molecule), the gray curve pertains to the system $H_2^+ \cdots H_2$ (the electron is on the right-hand side molecule). Variable q equals the difference of the bond lengths of the right-hand side molecule and the left-hand side molecule. At $q = \pm q_0$ both molecules have their optimum bond lengths. (b) The HF pendulum oscillates between two sites, A and B, which accommodate an extra electron becoming either A^-B or AB^-. The curves similar to parabolas, the black one for AB^- and the gray one for A^-B, denote the energies of the diabatic states as functions of the pendulum angle.

geometries. If q increases further, it will soon reach the value $q = q_0 = 2.1 - 1.4 = 0.7$ a.u., the optimum value for both molecules. A further increase of q will mean, however, a kind of discomfort for each of the molecules, and the energy will go up; for large q it will go up very much. And what will happen for $q < 0$? This means an elongation of an already too long H_2 and a shortening of an already too short H_2^+. The potential energy goes up and the total plot is similar to a parabola with the minimum at $q = q_0 > 0$ (see the black curve on Fig. 6.23a).

If, however, we assume that the extra electron resides all the time on the right-hand side molecule, so we have to do with $H_2^+ \cdots H_2$, then we will obtain the identical parabola-like curve as before, but with the minimum position at $q = -q_0 < 0$ (see the gray curve on Fig. 6.23a).

DIABATIC AND ADIABATIC POTENTIALS

Each of these curves with a single minimum and with the extra electron residing all the time on a given molecule represents the *diabatic* potential energy curve for the motion of the nuclei. If, when the donor–acceptor distance changes, the electron *keeps pace* with it and jumps on the acceptor, then increasing or decreasing q from 0 gives a similar result: we obtain a single electronic ground-state potential energy curve with *two* minima in positions $\pm q_0$. This is the *adiabatic* curve.

Asking whether the adiabatic or diabatic potential has to be applied is equivalent to asking *whether the nuclei are slow enough that the electron keeps pace (adiabatic) or not*

(diabatic) with their motion.[66] This is within the spirit of the adiabatic approximation (cf. Chapter V1-6, p. V1-321). Also, a diabatic curve corresponding to the same electronic structure (the extra electron sitting on one of the molecules all the time) is a special case of the diabatic hypersurface that preserved the same chemical bond pattern.

Example 5 (The "HF pendulum."). Similar conclusions come from another ideal system, namely, the hydrogen fluoride molecule treated as the pendulum of a grandfather clock (the hydrogen atom down, the clock axis going through the fluorine atom) moving over two molecules, A and A^-; one of them, A^-, accommodates an extra electron (Fig. 6.23b).

The electron is negatively charged, the hydrogen atom in the HF molecule carries a partial positive charge, and both objects attract each other. If the electron sits on the left-hand molecule and during the pendulum motion does not keep pace and does not jump over to the right-hand side molecule, the potential energy has a single minimum for the angle $-\theta_0$ (the diabatic potential might be approximated by a parabola-like curve with the minimum at $-\theta_0$). A diabatic analogous curve with the minimum at θ_0 arises when the electron resides on B all the time. When the electron keeps pace with any position of the pendulum, we have a single adiabatic potential energy curve with two minima, i.e., at $-\theta_0$ and at θ_0.

6.8.2 Marcus theory

The contemporary theory of the electron transfer reaction was proposed by Rudolph Marcus.[67] The theory is based to a large extent on the harmonic approximation for the diabatic potentials involved, i.e., the diabatic curves represent parabolas. One of

Rudolph Arthur Marcus (born 1923), American chemist, professor at the University of Illinois in Urbana and at the California Institute of Technology in Pasadena. In 1992 Marcus received the Nobel Prize "*for his contribution to the theory of electron transfer reactions in chemical systems*."

[66] In the electron transfer reaction $H_2^+ + H_2 \rightarrow H_2 + H_2^+$ the energy of the reactants is equal to the energy of the products, because the reactants and the products represent the same system. Is it therefore a kind of fiction? Is there any reaction at all taking place? From the point of view of a bookkeeper (thermodynamics) no reaction took place, but from the point of view of a molecular observer (kinetics) such a reaction may take place. It is especially visible when instead of one of the hydrogen atoms we use deuterium; then the reaction $HD^+ + H_2 \rightarrow HD + H_2^+$ becomes real even for the bookkeeper (mainly because of the difference in the zero-vibration energies of the reactants and products).

[67] The reader may find a good description of the theory in a review article by P.F. Barbara, T.J. Meyer, M.A. Ratner, *J. Phys. Chem.*, 100(1996)13148.

the parabolas corresponds to the reactants $V_R(q)$, the other to the products $V_P(q)$ of the electron transfer reaction (Fig. 6.24).[68]

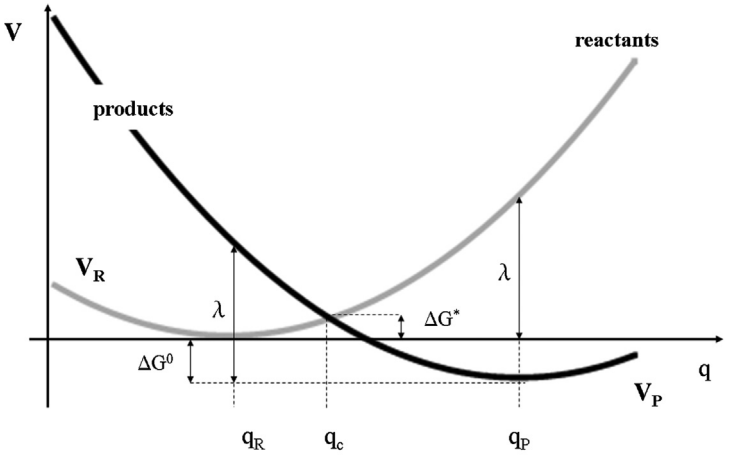


Fig. 6.24. The Marcus theory is based on two parabolic diabatic potentials $V_R(q)$ and $V_P(q)$ for the reactants (gray curve) and products (black curve), having minima at q_R and q_P, respectively. The quantity $\Delta G^0 \equiv V_P(q_P) - V_R(q_R) < 0$ represents the energy difference between the products and the reactants at their equilibrium geometries, the reaction barrier $\Delta G^* \equiv V_R(q_c) - V_R(q_R) = V_R(q_c) > 0$, where q_c corresponds to the intersection of the parabolas. The reorganization energy $\lambda \equiv V_R(q_P) - V_R(q_R) = V_R(q_P) > 0$ represents the energy expense for making the geometry of the reactants identical with that of the products (and *vice versa*).

Now, let us assume that both parabolas have the same curvature (force constant f).[69] The reactants correspond to the parabola with the minimum at q_R (without losing generality we adopt a convention that at $q = q_R$ the energy is equal zero), i.e.,

$$V_R(q) = \frac{1}{2} f (q - q_R)^2,$$

while the parabola with the minimum at $q = q_P$ is shifted in the energy scale by ΔG^0 ($\Delta G^0 < 0$ corresponds to an exothermic reaction[70]), i.e.,

$$V_P(q) = \frac{1}{2} f (q - q_P)^2 + \Delta G^0.$$

[68] Let the mysterious q be a single variable for a while, whose deeper meaning will be given later. In order to make the story more concrete let us think about two reactant molecules (R) that transform into the product molecules (P): $A^- + B \rightarrow A + B^-$.

[69] This widely used assumption is better fulfilled for large molecules when one electron more or less does not change much.

[70] That is, the energy of the reactants is higher than the energy of the products (as in Fig. 6.24).

So far we just treat the quantity ΔG^0 as a potential energy difference $V_P(q_P) - V_R(q_R)$ of the model system under consideration ($H_2^+ + H_2$ or the "HF pendulum"), even though the symbol suggests that this interpretation will be generalized in the future.

Such parabolas represent a simple situation. The parabolas' intersection point q_c satisfies by definition $V_R(q_c) = V_P(q_c)$. This gives[71]

$$q_c = \frac{\Delta G^0}{f}\frac{1}{q_P - q_R} + \frac{q_P + q_R}{2}.$$

Of course on the parabola diagram, the two minima are the most important, the intersection point q_c and the corresponding energy, which represents the reaction barrier reactants → products.

MARCUS EQUATION

The electron transfer reaction barrier is calculated as

$$\Delta G^* = V_R(q_c) = \frac{1}{4\lambda}\left(\lambda + \Delta G^0\right)^2, \tag{6.72}$$

where the energy λ (*reorganization energy*) represents the energy difference between the energies of the products in the equilibrium configuration of the reactants $V_P(q_R)$ and the energy in the equilibrium configuration of the products $V_P(q_P)$, i.e.,

$$\lambda = V_P(q_R) - V_P(q_P) = \frac{1}{2}f(q_R - q_P)^2 + \Delta G^0 - \Delta G^0 = \frac{1}{2}f(q_R - q_P)^2.$$

The reorganization energy is therefore always positive (energy expense).

REORGANIZATION ENERGY

Reorganization energy is the energy cost needed for making products in the nuclear configuration of the reactants.

If we ask about the energy needed to transform the optimal geometry of the reactants into the optimal geometry of the products, we obtain the same number. Indeed, we immediately obtain $V_R(q_P) - V_R(q_R) = \frac{1}{2}f(q_R - q_P)^2$, which is the same as before. Such a result is a consequence of the harmonic approximation and the same force constant assumed for V_R and V_P. It is seen

[71] If the curves did not represent parabolas, we might have serious difficulties. This is why we need harmonicity.

that the barrier for the thermic electron transfer reaction is higher if the geometry change is wider upon the electron transfer (large $(q_R - q_P)^2$) and if the system is stiffer (large f).

Svante August Arrhenius (1859–1927), Swedish physical chemist and astrophysicist, professor at the Stockholm University, originator of the electrolytic theory of ionic dissociation, measurements of the temperature of planets and of the solar corona, and of the theory deriving life on Earth from outer space. In 1903 he received the Nobel Prize in chemistry "*for the services he has rendered to the advancement of chemistry by his electrolytic theory of dissociation*."

From the Arrhenius theory the electron transfer (ET) reaction rate constant reads as

$$k_{ET} = Ae^{-\frac{\left(\lambda+\Delta G^0\right)^2}{4\lambda k_B T}}. \tag{6.73}$$

How would the reaction rate change if parabola $V_R(q)$ stays in place, while parabola $V_P(q)$ moves in the energy scale with respect to it? In experimental chemistry this may correspond to a *class* of the chemical reactions $A^- + B \rightarrow A + B^-$, with A (or B) from a homological series of compounds. The homology suggests that the parabolas are similar, because the mechanism is the same (the reactions proceed similarly), and the situations considered differ only by a lowering of the second parabola with respect to the first, i.e., the reactions in the series become more and more exothermic. We may have four qualitatively different cases:

Case 1. If the lowering is zero, i.e., $\Delta G^0 = 0$, the reaction barrier is equal to $\lambda/4$ (Fig. 6.25a).

Case 2. If $\left|\Delta G^0\right| < \lambda$, the reaction barrier is lower, because of the subtraction in the exponent, and the reaction rate *increases* (Fig. 6.25b). Hence, the ΔG^0 is the "driving force" in such reactions.

Case 3. When $\left|\Delta G^0\right|$ keeps increasing, at $\left|\Delta G^0\right| = \lambda$ the reorganization energy cancels out the driving force, and the barrier vanishes to zero. Note that this represents the highest reaction rate possible (Fig. 6.25c).

Case 4. Inverse Marcus region (Fig. 6.25d). Let us imagine now that we keep increasing the driving force. We have a reaction for which $\Delta G^0 < 0$ and $\left|\Delta G^0\right| > \lambda$. Compared to the previous case, the *driving force has increased, whereas the reaction rate decreases*. This might look like a possible surprise for experimentalists. A case like this is called the *inverse Marcus region*,

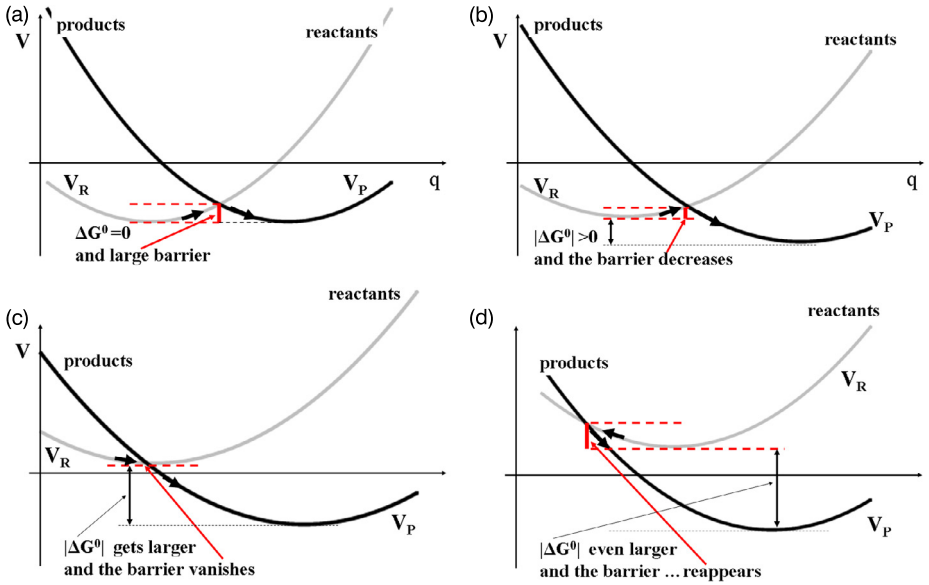


Fig. 6.25. Four qualitatively different cases in the Marcus theory. The reactant (product) parabola is gray (black). We assume $\Delta G^0 \leq 0$. (a) $\Delta G^0 = 0$, hence $\Delta G^* = \frac{\lambda}{4}$. (b) $\left|\Delta G^0\right| < \lambda$. (c) $\left|\Delta G^0\right| = \lambda$. (d) Inverse Marcus region $\left|\Delta G^0\right| > \lambda$.

foreseen by Marcus in the 1960s, using the two-parabola model. People could not believe this prediction until experimental proof[72] in 1984.

New meaning of the variable q

Let us make a subtraction. We write

$$\begin{aligned} V_R(q) - V_P(q) &= f(q-q_R)^2/2 - f(q-q_P)^2/2 - \Delta G^0 \\ &= \frac{f}{2}[2q - q_R - q_P][q_P - q_R] - \Delta G^0 = Aq + B, \end{aligned} \tag{6.74}$$

where A and B represent constants. This means that in parabolic approximation

> the diabatic potential energy difference depends *linearly* on coordinate q. In other words, measuring the stage of a given electron transfer reaction we may use either q or $V_R(q) - V_P(q)$.

72 J.R. Miller, L.T. Calcaterra, G.L. Closs, *J. Am. Chem. Soc.*, 97(1984)3047.

6.8.3 Solvent-controlled electron transfer

The above examples and derivations pertain to a one-dimensional model of electron transfer (a single variable q), while in reality (imagine a solution) the problem pertains to a huge number of variables. What happens here? Let us take the example of electron transfer between Fe^{2+} and Fe^{3+} ions in an aqueous solution, $Fe^{2+} + Fe^{3+} \rightarrow Fe^{3+} + Fe^{2+}$ (Fig. 6.26).[73]

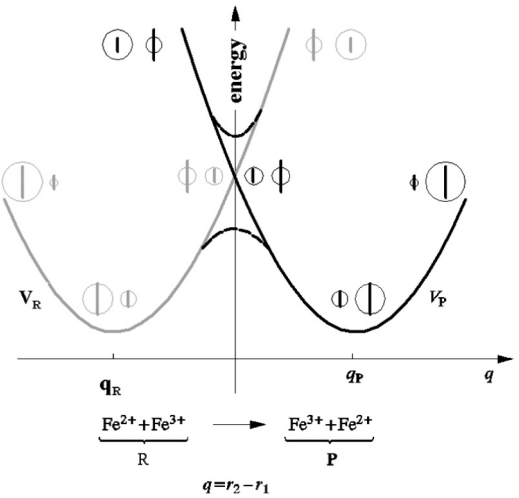


Fig. 6.26. The diabatic potential energy curves, with V_R for the reactants (gray) and V_P for the products (black) pertaining to the electron transfer reaction $Fe^{2+} + Fe^{3+} \rightarrow Fe^{3+} + Fe^{2+}$ in aqueous solution. The curves depend on the variable $q = r_2 - r_1$ that describes the *solvent*, which is characterized by the radius r_1 of the cavity for the first (say, left) ion and by the radius r_2 of the cavity for the second ion. For the sake of simplicity we assume $r_1 + r_2 = const$ and equal to the sum of the ionic radii of Fe^{2+} and Fe^{3+}. For several points q the cavities are drawn (circles with radii varying with q) as well as the vertical sections that symbolize the diameters of the left and right ions (the diameters of the ions stay constant). In this situation, the plots V_R and V_P have to differ widely. The dashed lines represent the adiabatic curves (in the peripherial sections they coincide with the diabatic curves).

It turns out that

> solvent behavior is of key importance for the electron transfer process.

The ions Fe^{2+} and Fe^{3+} are hydrated. For the reaction to proceed, the solvent has to *reorganize* itself next to both ions. The hydration shell of the Fe^{2+} ion is of larger radius than the hydration shell of the Fe^{3+} ion, because Fe^{3+} is smaller than Fe^{2+} and, in addition, creates a stronger

[73] In this example $\Delta G^0 = 0$, i.e., case 1 considered above.

electric field due to its higher charge, thus attracting the hydration shell stronger (Fig. 6.26). Both factors add to a more compact association of the water molecules with the Fe^{3+} ion than with Fe^{2+}. In a crude approximation, the state of the solvent may be characterized by the radii of the spherical cavities, that have to be created in water structure to accommodate the ions. Let us assume possible cavities have variable radii r_1 and r_2, whereas the fixed ionic radii are r_{Fe2+} and r_{Fe3+}, with $r_{Fe2+} > r_{Fe3+}$. Let us assume, for the sake of simplicity, that $r_1 + r_2 = r_{Fe2+} + r_{Fe3+} = const$ and introduce a single variable $q = r_2 - r_1$ that in this situation characterizes the state of the solvent. Thus, instead of billions of parameters we have only one. Let us see what happens when q changes.

We first consider that the extra electron sits on the left ion all the time (gray reactant curve V_R, Fig. 6.26) and the variable q is a negative number (with a high absolute value, i.e., $r_1 \gg r_2$). As seen from Fig. 6.26, the energy is very high, because the solvent squeezes the Fe^{3+} ion out (the second cavity is too small). It does not help that the Fe^{2+} ion has a lot of space in its cavity. Now we begin to move towards higher values of q. The first cavity begins to shrink, for a while without any resistance from the Fe^{2+} ion, and the second cavity begins to lose its pressure, thus making the Fe^{3+} ion more and more happy. The energy decreases. Finally we reach the minimum of V_R at $q = q_R$ and the radii of the cavities match the ions perfectly. Meanwhile variable q continues to increase. Now the solvent squeezes the Fe^{2+} ion out, while the cavity for Fe^{3+} becomes too large. The energy increases again, because of the first effect. We arrive at $q = 0$. The cavities are of equal size, but match neither of the ions well. This time the Fe^{2+} ion experiences some discomfort, and after passing the point $q = 0$ the pain increases more and more, and the energy continues to increase. The whole story pertains to the extra electron sitting on the left ion all the time (no jump, i.e., the reactant situation). A similar dramatic story can be told when the electron is sitting all the time on the right ion (product situation). In this case we obtain the V_P plot (black).

The V_R and V_P plots just described represent the diabatic potential energy curves for the motion of the nuclei, each valid for the extra electron residing on the same ion all the time. Fig. 6.26 also shows the adiabatic curve when the extra electron has enough time to adjust to the motion of the approaching nuclei and the solvent, and jumps at the partner ion.

Taking a single parameter q to describe the electron transfer process in a solvent is certainly a crude simplification. Actually there are billions of variables in the game, describing the degrees of freedom of the water molecules in the first and further hydration shells. One of the important steps towards successful description of the electron transfer reaction was the Marcus postulate,[74] which says that

[74] Such collective variables are used very often in everyday life. Who cares about all the atomic positions when studying a ball rolling down an inclined plane? Instead, we use a single variable (the position of the center of the ball) which gives us a perfect description of the system in a certain energy range.

despite the multidimensionality of the problem, Eq. (6.74) is still valid, i.e., the separation $V_R - V_P$ is a single variable describing the position of the system on the electron transfer reaction path (it is therefore a collective coordinate that describes the positions of the solvent molecules).

There is no doubt a low potential energy value is important, but it is also important how often this value can be reached by the system. This is connected to the *width* of the low-energy basin associated with the *entropy*[75] and to the *free energy*. In statistical thermodynamics we introduce the idea of the mean force, related to the free energy. Imagine a system in which we have two motions on different time scales: fast (e.g., of small solvent molecules) and slow (e.g., which change the shape of a macromolecule). To focus on the slow motion, we average the energy over the fast motions (the Boltzmann factor needed will introduce a temperature dependence in the resulting energy). In this way, from the potential energy, we obtain the *mean force potential* depending only on the slow variables,[76] sometimes called the free energy (which is a function of the geometry of the macromolecule, cf. p. V1-415).

The second Marcus assumption is that the ordinate axis should be treated as the mean force potential, or the free energy G, rather than just potential energy V.

It is very rare in theoretical chemistry that a many-dimensional problem can be transformed to a single variable problem. This is why the Marcus idea described above of a collective coordinate provokes the reaction: "no way." However, as it turned out later, this simple postulate lead to a solution that grasps the essential features of electron transfer.

What do the Marcus parabolas mean?

The example just considered of the electron transfer reaction $Fe^{2+} + Fe^{3+} \rightarrow Fe^{3+} + Fe^{2+}$ reveals that in this case the reaction barrier is controlled by the solvent, i.e., by billions of

[75] A wide potential energy well can accommodate a lot of closely lying vibrational levels and therefore the number of possible states of the system in a given energy range may be huge (large entropy). Please recall the particle-in-a-box problem (Chapter V1-4): the longer the box, the tighter the energy levels.

[76] The free energy is defined as $F(T) = -kT\frac{\partial}{\partial T}\ln Z$, where $Z = \sum_i \exp\left(-\frac{E_i}{kT}\right)$ represents the partition function (also known as the sum of states), where E_i stands for the i-th energy level. In the classical approach this energy level corresponds to the potential energy $V(x)$, where x represents a point in configurational space, and the sum corresponds to an integral over the total configurational space $Z = \int dx \exp\left(-\frac{V}{kT}\right)$. Note that the free energy is a function of the temperature only, not of the spatial coordinates x. If however, the integration were only carried out over part of the variables, say, only the fast variables, then Z, and therefore also F, would become a function of the slow variables and of the temperature (mean force potential). Despite the incomplete integration, we sometimes use the name "free energy" for this mean force potential by saying that "the free energy is a function of coordinates..."

coordinates. As shown by Marcus, this plethora can be effectively replaced by a single collective variable. Only after this approximation may we draw the diabatic parabola-like curves. The intersection point of the two diabatic curves can be found easily only after assuming their parabolic character. And yet any collective variable means motion along a line in an extremely complex configurational space (solvent molecules plus reactants). Moving along this line means that, according to Marcus, we encounter the intersection of the ground and excited electronic states. As shown in Chapter V1-6, such a crossing occurs at the conical intersection. Is it therefore that during the electron transfer reaction the system goes through the conical intersection point? How to put together such notions as reaction barrier, reaction path, entrance and exit channels, not to speak of acceptor–donor theory? Fig. 6.27, pertaining to the reaction DA $\rightarrow$ D^+A^-, will be of some help to us.

- The *diabatic hypersurfaces*, one corresponding to DA (i.e., the extra electron is on the donor all the time) and the second to D^+A^- (i.e., the extra electron resides on the acceptor), undergo the conical intersection (Fig. 6.27a, note the two colors used). For conical intersection to happen at least three atoms are required. Imagine a simple model, with a diatomic acceptor A and an atom D as donor. Atom D has a dilemma: does it transfer the electron to the first or the second atom of A? This dilemma means conical intersection. The variables ξ_1 and ξ_2 described in Chapter V1-6 were chosen (they lead to splitting of the adiabatic hypersurfaces, Fig. 6.27b,c), which measure the deviation of the donor D with respect to the corner of the equilateral triangle of side equal to the length of the diatomic molecule A. The conical intersection point, i.e., (0, 0), corresponds to the equilateral triangle. The figure also shows the upper and lower cones touching at (0, 0).
- The conical intersection led to two *adiabatic* hypersurfaces: lower (electronic ground state) and upper (electronic excited state). Each of the adiabatic hypersurfaces shown in Fig. 6.27 consists of the "dark" half (the diabatic states of the reactants, DA) and the "light" half (the diabatic state of the products, D^+A^-). The border between them reveals the intersection of the two diabatic states and represents the line of change of the electronic structure reactants/products. Crossing the line means the chemical reaction happens.
- The "avoided crossing" occurs everywhere along the border except at the conical intersection. It is rather improbable that the reactive trajectory passes through the conical intersection, because it usually corresponds to higher energy. It will usually pass by (this resembles an avoided crossing) and the electronic state DA changes to electronic state D^+A^- or *vice versa*. This is why we speak of the avoided crossing in a polyatomic molecule, whereas the concept pertains to diatomics only (see V1-6.13).
- Passing the border is easiest at *two* points (Fig. 6.27c). These are the two saddle points (barriers I and II). A thermic electron transfer reaction goes through one of them. In each case we obtain different products. Both saddle points differ in that D, when attacking A, has the choice of joining either of the two ends of A, usually forming *two different* products.

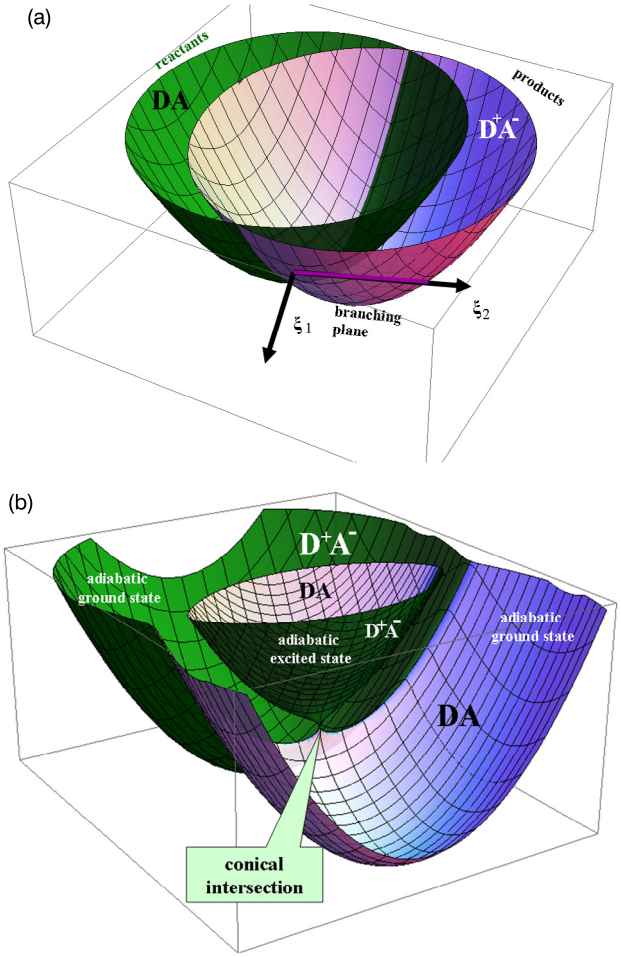


Fig. 6.27. Electron transfer in the reaction DA → D^+A^- as well as the relation of the Marcus parabolas to the concepts of the conical intersection, diabatic and adiabatic states, entrance and exit channels, and reaction barrier. (a) Two diabatic surfaces of the electronic energy as functions of the ξ_1 and ξ_2 variables that describe the deviation from the conical intersection point (within the bifurcation plane, cf. p.V1-369). Both surfaces are shown schematically in the form of the two intersecting paraboloids: one for the reactants (DA), the second for products (D^+A^-). (b) The same, but the hypersurfaces are presented more realistically. The upper and lower adiabatic surfaces touch at the conical intersection point. Note that each of the adiabatic surfaces is two-color meaning two different chemical structures. (c) Another view of the same surfaces. On the ground-state adiabatic surface (the lower one) we can see two reaction channels I and II, each with its reaction barrier (see the text). On the upper adiabatic surface an energy valley is visible that symbolizes a bound state that is separated from the conical intersection by a reaction barrier. (d) The Marcus parabolas represent the sections of the diabatic surfaces along the corresponding reaction channel, hence at a certain distance from the conical intersection. Hence, the parabolas in reality cannot intersect (undergo an avoided crossing).

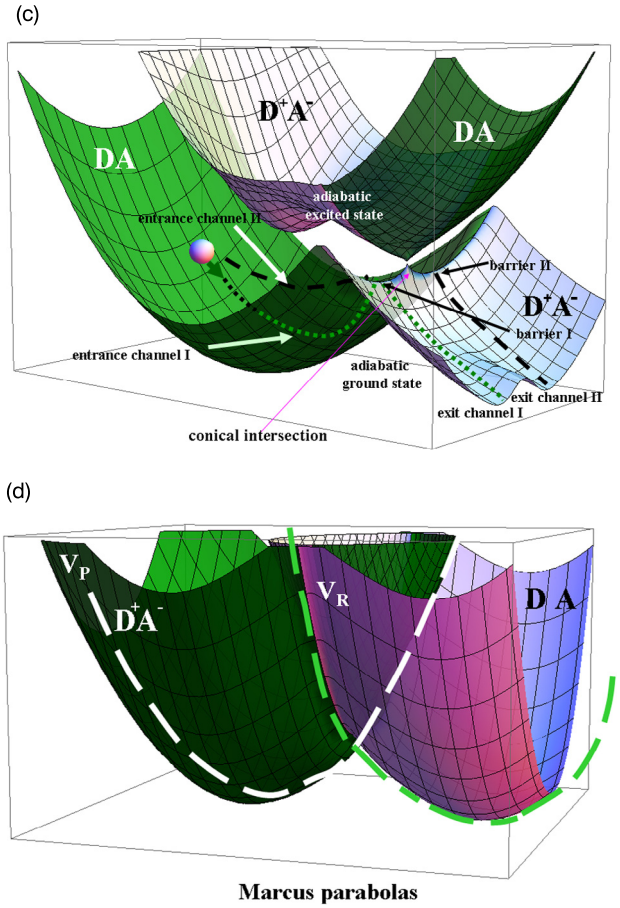


Fig. 6.27. (*continued*)

We therefore usually have *two* barriers. In the example given (H_3) they are identical, but in general they may differ. When the barrier heights are equal because of symmetry, it does not matter which is overcome. When they are different, one of them dominates (usually the lower barrier[77]). The channels shown in the figure are not curved, because we use a coordinate system different from that used in the colinear reaction.

- The Marcus parabolas (Fig. 6.27d) represent a special section (along the collective variable) of the hypersurfaces passing through the conical intersection (parabolas V_R and V_P). Each parabola represents a diabatic state; therefore, a part of each reactant parabola is on the lower hypersurface, while the other part is on the upper hypersurface. We see that the

77 There may be some surprises. Barrier height is not all that matters. Sometimes it may happen that what decides is access to the barrier region, in the sense of its width (this is where the entropy and free energy matter).

parabolas are only an approximation to the hypersurface profile. The reaction is of a thermic character, and as a consequence, the parabolas should not pass through the conical intersection, because it corresponds to high energy; instead it passes through one of the saddle points.

- The "light color" part of the excited state hypersurface runs up to the "dark color" part of the ground-state hypersurface, or *vice versa*. This means that photoexcitation (following the Franck–Condon rule this corresponds to a vertical excitation) means a profound change: the system looks as if it has already reacted (photoreaction).

Quantum mechanical modification

In the Marcus equation (6.72) we assume that in order to make the electron transfer effective, we have to supply at least the energy equal to the barrier height. The formula does not obviously take into account the quantum nature of the transfer. The system may overcome the barrier not only by having its energy higher than the barrier, but also by tunneling,[78] when its energy is lower than the barrier height (cf. p. V1-203). Besides, the reactant and product energies are quantized (vibrational-rotational levels[79]). The reactants may be excited to one of such levels. The reactant vibrational levels will have different abilities to tunnel.

According to Chapter V1-2 only a time-dependent perturbation is able to change the system energy. Such a perturbation may serve the electric field of the electromagnetic wave. When the perturbation is periodic, with the angular frequency ω matching the energy difference of initial state k and one of the states of higher energy (n), then the transition probability between these states is equal to $P_k^n(t) = \frac{2\pi t}{\hbar} |v_{kn}|^2 \, \delta\left(E_n^{(0)} - E_k^{(0)} - \hbar\omega\right)$ (the Fermi Golden Rule, Eq. (V1-2.36) on p. V1-114, is valid for relatively short times t), where $v_{kn} = \langle k|v|n\rangle$, with $v(\mathbf{r})$ representing the perturbation amplitude,[80] $V(\mathbf{r}, t) = v(\mathbf{r})e^{i\omega t}$. The Dirac delta function δ is a quantum mechanical way of saying that the total energy has to be conserved. In phototransfer of the electron, state "k" represents the quantum mechanical state of the reactants and "n" represents a product state, each of diabatic character.[81] In practice the adiabatic approximation is used, in which the reactant and product wave functions are products of the electronic wave functions (which depend on the electronic coordinates $\mathbf{r}$ and, parametrically, on the nuclear configuration $\mathbf{R}$) and the vibrational functions $f(\mathbf{R})$ describing the motion of the nuclei, i.e., $\psi_{k,R}(\mathbf{r}; \mathbf{R}) f_{v_1,R}(\mathbf{R})$ and $\psi_{n,P}(\mathbf{r}; \mathbf{R}) f_{v_2,P}(\mathbf{R})$. The indices v_1 and v_2 in functions f denote the vibrational quantum numbers.

---

78 In a moment we will also consider electron transfer due to optical excitation.

79 For large molecules, we may forget the rotational spectrum, since, because of the large inertia momentum, the rotational states form a quasicontinuum ("no quantization").

80 $\mathbf{r}$ stands for those variables on which the wave functions depend.

81 They will be denoted by the subscripts R and P.

Then the transition probability depends on the integral (Chapter V1-2)

$$|v_{kn}|^2 = \left|\left\langle \psi_{k,R}\left(\mathbf{r};\mathbf{R}\right) f_{v_1,R}(\mathbf{R})|v(\mathbf{r})|\psi_{n,P}\left(\mathbf{r};\mathbf{R}\right) f_{v_2,P}(\mathbf{R})\right\rangle\right|^2 .$$

Let us rewrite it, making the integration over the nuclear and electronic coordinates explicit (where $\mathrm{d}V_{nucl}$ and $\mathrm{d}V_e$ mean that the integrations are over the nuclear and electronic coordinates, respectively):

$$v_{kn} = \int \mathrm{d}V_{nucl}\; f^*_{v_1,R}\left(\mathbf{R}\right)\; f_{v_2,P}\left(\mathbf{R}\right) \int \mathrm{d}V_e\; \psi^*_{k,R}\left(\mathbf{r};\mathbf{R}\right) v\left(\mathbf{r}\right) \psi_{n,P}\left(\mathbf{r};\mathbf{R}\right).$$

Now, let us use the Franck–Condon approximation that the optical perturbation makes the electrons move instantaneously while the nuclei do not keep pace with the electrons and stay in the same positions (we assume therefore equilibrium positions of the nuclei $\mathbf{R}_0$ in the reactants). We have

$$v_{kn} \approx \int \mathrm{d}V_{nucl}\; f^*_{v_1,R}\left(\mathbf{R}\right)\; f_{v_2,P}\left(\mathbf{R}\right) \int \mathrm{d}V_e\; \psi^*_{kR}\left(\mathbf{r};\mathbf{R}_0\right) v\left(\mathbf{r}\right) \psi_{n,P}\left(\mathbf{r};\mathbf{R}_0\right).$$

The last integral represents a constant, and therefore

$$|v_{kn}|^2 = |V_{RP}|^2\, |S_{osc}(v_1, v_2)|^2 ,$$

where

$$V_{RP} = \int \mathrm{d}V_e\; \psi^*_{k,R}\left(\mathbf{r};\mathbf{R}_0\right) v\left(\mathbf{r}\right) \psi_{n,P}\left(\mathbf{r};\mathbf{R}_0\right),$$

$$S_{osc}\left(v_1, v_2\right) = \int \mathrm{d}V_{nucl}\; f^*_{v_1,R}\left(\mathbf{R}\right)\; f_{v_2,P}\left(\mathbf{R}\right). \tag{6.75}$$

The $|S_{osc}(v_1, v_2)|^2$ is called the *Franck–Condon factor*.

FRANCK–CONDON FACTOR

A Franck–Condon factor is the square of the absolute value of the overlap integral of the vibrational wave functions, one pertaining to the reactants with $\mathrm{d}V_{nucl}$ vibrational quantum number v_1 and the second pertaining to the products with vibrational quantum number v_2.

The calculation of V_{RP} is not an easy matter; therefore we often prefer an empirical approach, by modeling the integral as[82]

$$V_{RP} = V_0 exp[-\beta(R - R_0)],$$

where R_0 stands for the van der Waals distance of the donor and acceptor, R represents their distance, $\beta > 0$ represents a constant, and V_0 means V_{RP} for the van der Waals distance.[83]

A large Franck–Condon factor means that by exciting the reactants to the vibrational state v_1 there is a particularly high probability of electron transfer (by tunneling) with the products in vibrational state v_2.

Reorganization energy

In the Marcus equation, reorganization energy plays an important role. This energy is the main reason for the electron transfer reaction barrier.

The reorganization pertains to the neighborhood of the transferred electron,[84] i.e., to the solvent molecules, but also to the donors and acceptors themselves.[85] This is why the reorganization energy, in the first approximation, consists of the internal reorganization energy (λ_i) that pertains to the donor and acceptor molecules and the solvent reorganization energy (λ_0), i.e.,

$$\lambda = \lambda_i + \lambda_0.$$

Internal reorganization energy

For the electron to have the chance of jumping from molecule A^- to molecule[86] B, it has to have the neighborhood reorganized in a special way. The changes should make the extra electron's life hard on A^- (together with solvation shells) and seduce it by the alluring shape

[82] Sometimes the dependence is different. For example, in *twisted intramolecular charge transfer* (TICT), after the electron is transferred between the donor and acceptor moieties (a large V_{RP}) the molecule undergoes an internal rotation of the moieties, which causes an important decrease of V_{RP} (K. Rotkiewicz, K.H. Grellmann, Z.R. Grabowski, *Chem. Phys. Letters*, 19(1973)315).

[83] As a matter of fact, such formulae only contain a simple message: V_{RP} decreases *very* fast when the donor–acceptor distance increases.

[84] The neighborhood is adjusted perfectly to the extra electron (to be transferred) in the reactant situation, and very unfavorable for its future position in the products. Thus the neighborhood has to be reorganized to adjusted for the electron transfer products.

[85] It does not matter for an electron what in particular prevents it from jumping.

[86] "–" denotes the site of the extra electron. It does not necessarily mean that A^- represents an anion.

of molecule B and its solvation shells. To do this, work has to be done. First, this is an energy cost for the proper deformation of A^- to the geometry of molecule A, i.e., already without the extra electron (the electron obviously does not like this – this is how it is forced out). Next, molecule B is deformed to the geometry of B^- (this is what makes B attractive to the extra electron – everything is prepared for it in B). These two energy effects correspond to λ_i.

Calculation of λ_i is simple. We have

$$\lambda_i = E(A^-B; \text{geom } AB^-) - E(A^-B; \text{geom } A^-B),$$

where $E(A^-B; \text{geom } AB^-)$ denotes the energy of A^-B calculated for the equilibrium geometry of another species, namely, AB^-, while $E(A^-B; \text{geom } A^-B)$ stands for the energy of A^-B at its optimum geometry.

Usually the geometry changes in AB^- and A^-B attain several percent of the bond lengths or the bond angles. The change is therefore relatively small, and we may represent it by a superposition of the normal mode vectors[87] $\mathbf{L}_k, k = 1, 2, ..., 3N$, described in Chapter V1-7. We may use the normal modes of the molecule A^-B (when we are interested in electron transfer from A^- to B) or of the molecule AB^- (back transfer). What for? Because some normal modes are more effective than others in facilitating electron transfer. The normal mode analysis would show[88] that

> the *most effective normal mode of the reactants deforms them in such a way as to resemble the products*. This vibration reorganizes the neighborhood in the desired direction (for electron transfer to occur), and therefore effectively lowers the reaction barrier.

Solvent reorganization energy

Spectroscopic investigations are unable to distinguish between the internal or solvent reorganization, because nature does not distinguish between the solvent and the rest of the neighborhood. An approximation to the solvent reorganization energy may be calculated by assuming a continuous solvent model. Assuming that the mutual configuration of the donor and acceptor

87 Yet the normal modes are linear combinations of the Cartesian displacements.

88 It usually turns out that there are several such vibrations. They will help electron transfer from A^- to B. The reason is quite obvious, e.g., the empirical formula for V_{RP} says that a vibration that makes the AB distance smaller will increase the transfer probability. This could be visible in what is known as resonance Raman spectroscopy close to a charge transfer optical transition. In such spectroscopy, we have the opportunity to observe particular vibronic transitions. The intensity of the vibrational transitions (usually from $v = 0$ to $v = 1$) of those normal modes which facilitate electron transfer will be highest.

(separated by distance R) allows for enclosing them in nonoverlapping spheres of radii a_1 and a_2, the following formula was derived by Marcus:

$$\lambda_0 = (\Delta e)^2 \left\{ \frac{1}{2a_1} + \frac{1}{2a_2} - \frac{1}{R} \right\} \left\{ \frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0} \right\},$$

where ϵ_∞ and ϵ_0 denote the dielectric screening constants measured at infinite and zero electromagnetic field frequency, respectively, and Δe is equal to the effective electric charge transferred between the donor and acceptor. The dielectric screening constant is related to the polarization of the medium. The value ϵ_0 is larger than ϵ_∞, because, at a constant electric field, the electrons as well as the nuclei (mainly an effect of the reorientation of the molecules) keep pace to adjust to the electric field. At high frequency only the electron keep pace, hence $\epsilon_\infty < \epsilon_0$. The last term in curly brackets takes care of the difference, i.e., of the reorientation of the molecules in space (cf. Chapter 4).

Summary

- A chemical reaction represents a molecular catastrophe, in which the electronic structure, as well as the nuclear framework of the system, changes qualitatively. Most often a chemical reaction corresponds to the breaking of an old and creation of a new bond.
- The simplest chemical reactions correspond to overcoming a single reaction barrier on the way from reactants to products through saddle point along the intrinsic reaction coordinate (IRC). The IRC corresponds to the steepest descent path (in the mass-weighted coordinates) from the saddle point to configurations of reactants and products.
- Such a process may be described as the system passing from the entrance channel (reactants) to the exit channel (products) on the electronic energy map as a function of the nuclear coordinates. For the reaction $A + BC \rightarrow AB + C$ the map shows a characteristic reaction "drain-pipe." Passing along the "drain-pipe" bottom usually requires overcoming a reaction barrier, its height being a fraction of the energy of breaking the "old" chemical bond.
- The reaction barrier reactants $\rightarrow$ products, is as a rule, of different height to the corresponding barrier for the reverse reaction.
- We have shown how to obtain an accurate solution for three-atomic reactions. After introducing the hyperspherical democratic coordinates it is possible to solve the Schrödinger equation (within the Ritz approach). We obtain the rate constant for the state-to-state elementary chemical reaction.

A chemical reaction may be described by the reaction path Hamiltonian in order to focus on the IRC measuring the motion along the "drain-pipe" bottom (reaction path) and the normal mode coordinates orthogonal to the IRC.

- During the reaction, energy may be exchanged between the vibrational normal modes, as well as between the vibrational modes and the motion along the IRC.
- Two atoms or molecules may react in many different ways (reaction channels). Even if at some conditions they do not react (e.g., the noble gases), the reason for this is that their kinetic energy is too low with respect to the corresponding reaction barrier, and the opening of their electronic

closed shells is prohibitively expensive in the energy scale. If the kinetic energy increases, more and more reaction channels open up, because it is possible for higher and higher energy barriers to be overcome.

- A reaction barrier is a consequence of the "quasiavoided crossing" of the corresponding diabatic hypersurfaces; as a result we obtain two adiabatic hypersurfaces ("lower," or electronic ground state, and "upper," or electronic excited state). Each of the adiabatic hypersurfaces consists of two diabatic parts stitched along the border passing through the conical intersection point. On both sides of the conical intersection there are usually two saddle points along the border line leading in general to two different reaction products (Fig. 6.27c).
- The two intersecting diabatic hypersurfaces (at the reactant configuration) represent the electronic ground-state DA and *that electronic excited state that resembles the electronic charge distribution of the products*, usually D^+A^-.
- The barrier appears therefore as the cost of opening the closed shell in such a way as to prepare the reactants for the formation of (a) new bond(s).
- In Marcus electron transfer theory, the barrier also arises as a consequence of the intersection of the two diabatic potential energy curves. The barrier height depends mainly on the (solvent and reactant) reorganization energy.

Main concepts, new terms

acceptor–donor (AD) reaction theory (p. 479)
Aharonov–Bohm effect (p. 458)
Berry phase (p. 458)
bobsleigh effect (p. 451)
collective coordinate (p. 518)
Coriolis coupling (p. 465 and 472)
critical points (p. 444)
cross-section (p. 458)
curvature coupling (p. 465 and 472)
cycloaddition reaction (p. 504)
democratic coordinates (p. 454)
diabatic and adiabatic states (p. 509)
donating mode (p. 472)
early and late reaction barriers (p. 451)
electrophilic attack (p. 495)
entrance and exit channels (p. 451)
exo- and endothermic reactions (p. 467)
femtosecond spectroscopy (p. 444)
Franck–Condon factors (p. 523)
intrinsic reaction coordinate (IRC) (p. 461)
inverse Marcus region (p. 514)
mass-weighted coordinates (p. 460)
mean force potential (p. 518)
MO and AD pictures (p. 481)
natural coordinates (p. 463)
nucleophilic attack (p. 492)
reaction "drain-pipe" (p. 451)
reaction path Hamiltonian (p. 463)
reaction rate (p. 458)
reaction spectator (p. 475)
reaction stages (p. 484)
reactive and nonreactive trajectories (p. 448)
reorganization energy (p. 524)
role of states DA, D^+A^-, D^+A^{-*} (p. 489)
saddle point (p. 444)
skew coordinate system (p. 448)
steepest descent path (SDP) (p. 446)
"trajectory-in-molasses" (p. 461)
vibrational adiabatic approximation (p. 465)
vibrational adiabatic potential (p. 466)
Woodward–Hoffmann rules (p. 501)

From the research front

Chemical reactions represent a very difficult problem for quantum chemistry, because:

- There are a lot of possible reaction channels. Imagine the number of all combinations of atoms in a monomolecular dissociation reaction, also in their various electronic states. We have to select first which reaction to choose, and a good clue may be the lowest possible reaction barrier.
- A huge change in the electronic structure is usually quite demanding for standard quantum mechanical methods.
- Given a chosen single reaction channel we confront the problem of calculating the potential energy hypersurface. Let us recall (Chapters V1-6 and V1-7) the number of quantum mechanical calculations to perform this is of the order of 10^{3N-6}. For number of nuclei as small as $N = 4$ we already have a million computation tasks to perform.
- Despite unprecedented progress in the computational technique, the cutting edge possibilities are limited in *ab initio* calculations to two diatomic molecules.

On the other hand, a chemist always has some additional information on which chemical reactions are expected to occur. Very often the most important changes happen in a limited set of atoms, e.g., in functional groups, their reactivity being quite well understood. Freezing the positions of those atoms which are reaction spectators only allows us to limit the number of degrees of freedom to consider.

Ad futurum

Chemical reactions with the reactants precisely oriented in space will be more and more important in chemical experiments of the future. Here it will be helpful to favor some reactions by supramolecular recognition, docking in reaction cavities, or reactions on prepared surfaces. For theoreticians, such control of orientation will mean the reduction of certain degrees of freedom. This, together with eliminating or simulating the spectator bonds, may reduce the task to manageable size. State-to-state calculations and experiments that will describe an effective chemical reaction that starts from a given quantum mechanical state of the reactants and ends up with another well-defined quantum mechanical state of the products will become more and more important. Even now, we may design with great precision practically any sequence of laser pulses (a superposition of electromagnetic waves, each of a given duration, amplitude, frequency, and phase). For a chemist, this means that we are able to change the shape of the hypersurfaces (ground and excited states) in a controllable way, because every nuclear configuration corresponds to a dipole moment that interacts with the electric field (cf. Chapter 4). The hypersurfaces may shake and undulate in such a way as to make the point representing the system move to the product region. In addition, there are possible excitations and the products may be obtained via excited hypersurfaces. As a result we may have selected bonds broken, and others created in a selective and highly efficient way. This technique demands important developments in the field of chemical reaction theory and experiment, because currently we are far from such a goal.

Note that the most important achievements in the chemical reaction theory pertained to concepts (von Neumann, Wigner, Teller, Woodward, Hoffmann, Fukui, Evans, Polanyi, Shaik) rather than computations. The potential energy hypersurfaces are so complicated that it took the scientists 50 years to

elucidate their main machinery. Chemistry means first of all chemical reactions, and most chemical reactions still represent *terra incognita*. This will change considerably in the years to come. In the longer term this will be the main area of quantum chemistry.

Additional literature

R.D. Levine, R.B. Bernstein, "Molecular Reaction Dynamics and Chemical Reactivity," Oxford University Press, 1987.

An accurate approach to the reactions of small molecules.

H. Eyring, J. Walter, G.E. Kimball, "Quantum chemistry," John Wiley, New York, 1967.

A good old textbook, written by outstanding specialists in the field. To the best of my knowledge, no later textbook has done it in more detail.

R.B. Woodward, R. Hoffmann, "The Conservation of Orbital Symmetry," Acad. Press, New York, 1970.

A summary of the important discoveries made by these authors (Woodward–Hoffmann symmetry rules).

S.S. Shaik, "What Happens to Molecules as They React? Valence Bond Approach to Reactivity," in the *Journal of the American Chemical Society*, 103(1981)3692.

An excellent paper that introduces many important concepts in a simple way.

Questions

1. The IRC represents (N is the number of nuclei):
   a. an arbitrary atom–atom distance, which changes from the value for the reactants' configuration to that of the products.
   b. a steepest descent path from the saddle point to the corresponding minima (using a Cartesian coordinate system in the nuclear configuration space).
   c. a steepest descent path from the saddle point to the corresponding minima (within a Cartesian coordinate system of the nuclei scaled by the square roots of the atomic masses).
   d. a curve in the $(3N - 6)$-dimensional space that connects two minima corresponding to two stable structures separated by a barrier.
2. In the vibrational adiabatic approximation within the reaction path Hamiltonian theory:
   a. the potential energy for the motion of the nuclei depends on the frequencies and excitations of the normal modes.
   b. the vibrational contribution to the potential energy for the nuclei depends on the value of the coordinate s along the IRC.
   c. excitations of some particular vibrational modes may lower the reaction barrier.
   d. no normal mode can change its vibrational quantum number at any s.

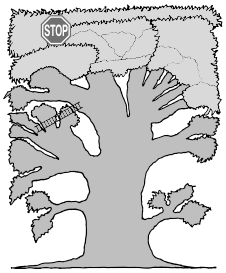
3. The donation mode:
   a. is that one which offers the largest value of the zero-point energy in the entrance channel.
   b. when excited lowers the reaction barrier.
   c. has a large value of the Coriolis coupling with at least one mode in the exit channel.
   d. means a mode that has a large value of the curvature coupling in the entrance channel.
4. A spontaneous endothermic reaction proceeds (at $T > 0$), because:
   a. the density of the vibrational energy levels is larger in the exit channel than in the entrance channel.
   b. the energy of the bottom of the entrance channel is higher than that of the exit channel.
   c. what decides about the direction is not the energy, but the free energy.
   d. the exit channel is much wider than the entrance channel.
5. In the acceptor–donor theory:
   a. the reactants correspond to the DA structure.
   b. in the D^+A^- structure the acceptor's bond becomes weaker, due to the electron occupation of its antibonding molecular orbital.
   c. the D^+A^- structure represents a low-energy excited state of the products.
   d. the reaction barrier comes from an intersection of the diabatic potential energy hypersurfaces for the DA and D^+A^- structures.
6. In the acceptor–donor theory at the intermediate stage of the reaction (I):
   a. an electron of the donor molecule jumps on the acceptor and occupies its antibonding orbital.
   b. the structure $D^{2+}A^{2-}$ becomes the dominating one.
   c. the dominating structures are D^+A^- and D^+A^{-*}.
   d. the dominating structures are DA and D^+A^-.
7. At the conical intersection for H_3:
   a. the excited-state adiabatic PES increases linearly with the distance from the conical intersection point.
   b. the ground-state adiabatic PES increases linearly with the distance from the conical intersection point.
   c. the excited-state adiabatic PES is composed of two diabatic surfaces corresponding to different electronic distributions.
   d. the ground- and excited-state adiabatic PESs coincide for any equilateral triangle configuration of H_3.
8. The conical intersection:
   a. pertains to relative position (in energy scale) of two electronic states as functions of the configuration of the nuclei.
   b. moving on the ground-state PES from its one diabatic part to the other corresponds to a thermally induced chemical reaction.
   c. thanks to the conical intersection, a UV excitation from the ground state may end up in a nonradiative transition back to the ground state.
   d. an electronic excitation (according to the Franck–Condon rule) of the reactants leads to a product-looking compound having the geometry of the reactants.
9. In the electron transfer theory of Marcus:
   a. the reorganization energy can be calculated as the energy needed to change the geometry of the reactants (from the equilibrium one) to the optimum nuclear configuration of the products.
   b. the larger the absolute value of the energy difference between the products and the reactants, the faster the reaction.

c. the activation energy is equal to the reorganization energy.
d. when the reactants and the products are the same, the barrier equals $\frac{1}{4}$ of the reorganization energy.

10. In the electron transfer theory of Marcus:
    a. one assumes the same force constants for the reactants and for the products.
    b. the reorganization energy can be calculated as the energy needed to change the geometry of the products (from the equilibrium one) to the optimum nuclear configuration of the reactants.
    c. one assumes the harmonic approximation for either of the diabatic states.
    d. the reason why there is an energy barrier for the electron transfer in the reaction $Fe^{2+} + Fe^{3+} \longrightarrow Fe^{3+} + Fe^{2+}$ is the reorganization energy of the solvent.

Answers

1c,d, 2a,b,c,d, 3b,d, 4a,c,d, 5a,b,d, 6a,d, 7a,c,d, 8a,b,c,d, 9a,d, 10a,b,c,d



CHAPTER 7

Information Processing – The Mission of Chemistry☆

Concern for man and his fate must always form the chief interest of all technical endeavors. Never forget this in the midst of your diagrams and equations.
Albert Einstein

Where are we?

We are on the top of the TREE crown.

An example

Information is of key importance, not only for human civilization, but also for functioning of any biological system. It is especially visible in biology that the hardware and the software of the information processing is based on chemistry. Thus, molecules, so interesting by themselves for chemists, theoreticians, and experimentalists, also so difficult to be correctly described by theory, participate in their "second life," of completely different character – they *process information*. In this special life they cooperate with thousands of other molecules in a precise spatiotemporal algorithm that has its own goal. Do you recall the first words of the introduction to the present book? Those about who we are? The present book made a loop. Imagine molecules are devoid of this "second function"; this would mean a disaster for our world.[1] *Despite its striking importance, this "second life" is virtually absent in nowadays chemistry,* which does not even have an appropriate language for describing information flows.

What is it all about?

Chemistry has played, and continues to play, a prominent role in human civilization. If you doubt it, just touch *any* surface around you – most probably it represents a product of the chemical industry.[2] Pharma-

☆ This chapter is based in part on a lecture given by the author in the Warsaw Academic Laser Center, December 7, 1999, as well as on *Reports of the Advanced Study Institute* of Warsaw Technical University, 2(2013).

1 We would be unable to notice it.

2 Just a quick test around myself (random choice of surfaces): laptop (polymers), marble table (holes filled with a polymer), pencil (wood, but coated by a polymer), box of paper tissue (dyes and polymer coat), etc.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464248-6.00015-6
Copyright © 2020 Elsevier B.V. All rights reserved.

ceutical chemistry may be seen as a real benefactor, for it makes our lives longer and more comfortable. Is the mission of chemistry, therefore, the production of better dyes, polymers, semiconductors, drugs? *No, its true mission is much more exciting.*

Multilevel supramolecular structures (statics) (⊣□) **p. 537**

- Complex systems
- Self-organizing complex systems
- Cooperative interactions
- Combinatorial chemistry – molecular libraries

Chemical feedback – a steering element (dynamics) (⊣□) **p. 543**

- A link to mathematics – attractors
- Bifurcations and chaos
- Brusselator with diffusion
- Brusselator with diffusion – dissipative structures
- Hypercycles
- From self-organization and complexity to information

Information and informed matter (⊣□) **p. 556**

- Abstract theory of information
- Teaching molecules
- Dynamic information processing of chemical waves
- Molecules as computer processors
- The mission of chemistry

Why is this important?

In this book we have dealt with many problems of quantum chemistry. If this book were only about quantum chemistry, I would not have written it. My goal was to focus on perspectives and images, rather than on pixel-like separate problems. Before being quantum chemists, we are scientists, happy eye witnesses of miracles going on around us. We are also human beings, and have the right to ask ourselves where we are heading. *Why* is the Schrödinger equation to be solved? *Why* do we want to understand the chemical foundations of the world? Just for curiosity? Well, should curiosity legitimate *any* investigation?[3] What will the future role of chemistry be?

Chemistry is on the threshold of a big leap forward. Students of today will participate in this revolution. The limits will be set by our imagination, maybe by our responsibility as well. The way that future progress in chemistry and biochemistry will be chosen will determine the fate of human civilization. This *is* important...

[3] Do not answer "yes" too easily, for it gives other people the right to perform any experiments on you and me.

What is needed?

- Elements of chemical kinetics,
- elements of differential equations,
- your natural curiosity.

Classical works

The classic papers pertain to three, at first sight unrelated, topics: molecular recognition, differential equations, and information processing. These topics evolved virtually separately within chemistry, mathematics, and telecommunication, and only now[4] begin to converge. ★ It seems that the first experiment with an oscillatory chemical reaction was reported by Robert Boyle in the 17th century (oxidation of phosphorus). Then several new reports on chemical oscillations were published (including books). *All these results did not attract any significant interest in the scientific community, because they contradicted the widely known, all important, and successful equilibrium thermodynamics*. ★ In 1903 Jules Henri Poincaré published in *Journal de Mathématiques Pures et Appliquées*, 7(1881)251, the article "*Mémoire sur les courbes définies par une équation différentielle*," in which he showed that a wide class of two coupled nonlinear differential equations leads to oscillating solutions that tend to a particular behavior *independently of the initial conditions* (called the *limit cycle*). ★ Emil Hermann Fischer was the first to stress the importance of molecular recognition. In "*Einfluss der Configuration auf die Wirkung der Enzyme*," published in *Berichte*, 27(1894)2985, Fischer used the self-explanatory words "key-lock" for the perfect fit of an enzyme and its ligand. ★ Independently, there was a continuing parallel progress in oscillatory solutions in mathematics. In 1910 Alfred J. Lotka in "*Contributions to the theory of chemical reactions*," published in the *Journal of Physical Chemistry*, 14(1910)271, proposed some differential equations that corresponded to the kinetics of an autocatalytic chemical reaction, and then with Vito Volterra gave a differential equation that describes a prey–predator *feedback* (oscillation) known as the Lotka–Volterra model. Chemistry of that time turned out to be nonprepared for such an idea. ★ In another domain Harry Nyquist published the article "Certain Factors affecting Telegraph Speed" in *The Bell Systems Technical Journal*, 3(1924)324; four years later, in the same journal, 7(1928)535, Ralph V.L. Hartley published the paper "Transmission of Information," in which for the first time the quantitative notion of information together with the effectiveness of information transmission was considered. ★ Twenty years later, the same topic was developed by Claude E. Shannon in "*A Mathematical Theory of Communication*," also published in *The Bell Systems Technical Journal*, 27(1948)379, 623, in which he related the notion of information and that of entropy. ★ The Soviet general Boris Belousov finally agreed to publish his only unclassified paper "*Periodichesky deystvouyoushchaya rieakcya i yeyo miekhanism*" in an obscure Soviet medical journal, *Sbornik Riefieratow Radiacjonnoj Miediciny*, *Medgiz, Moskwa*, 1(1959)145, reporting spectacular color oscillations in his test tube: yellow Ce^{4+} and then colorless Ce^{3+}, and again yellow, etc. (nowadays called the Belousov–Zhabotinsky reaction). Information about this oscillatory reaction diffused to Western science in the 1960 and made a real breakthrough. ★ Belgian scientists Ilya Prigogine and Gregoire Nicolis in the paper "*On Symmetry-Breaking Instabilities in Dissipative Systems*," published in the *Journal of Chemical Physics*, 46(1967)3542, introduced the notion of the dissipative structures. ★ Charles John Pedersen reopened (after the pioneering work of Emil Fischer) the field of supramolecular chemistry, publishing an article "*Cyclic Polyethers and their Complexes with*

[4] The aim of the present chapter is to highlight these connections.

Metal Salts," which appeared in the *Journal of the American Chemical Society*, 89(1967)7017, and dealt with molecular recognition (cf. Chapter 5). This, together with later works of Jean-Marie Lehn and Donald J. Cram, introduced the new paradigm of chemistry known as supramolecular chemistry. ★ Manfred Eigen and Peter Schuster, in three articles "*The Hypercycle. A Principle of Natural Self-Organization*," in *Naturwissenschaften*, 11(1977), 1(1978), and 7(1978), introduced to chemistry the idea of the hypercycles and of the natural selection of molecules. ★ Mathematician Leonard Adleman published "*Molecular Computation of Solutions to Combinatorial Problems*," in *Science*, 266(1994)1021, in which he described *his own* chemical experiments that shed new light on the role molecules can play in processing information. ★ Ivan Huc and Jean-Marie Lehn in the paper "*Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly*," published in *Proceedings of the National Academy of Sciences (USA)*, 94(1997)2106, stressed the importance in chemistry of the idea of molecular library as an easy-to-shift equilibrium *mixture* of molecular complexes, which was in contradiction with usual chemical practice of purifying substances.

* * *

What are the most important problems in chemistry? Usually we have no time to compose such a list, not to speak of presenting it to our students. The choice made reflects the author's personal point of view. The author tried to keep in mind that he is writing for mainly young (undergraduate and graduate) students, who are seeking not only detailed research reports, but also new *guidelines* in chemistry and some general *trends* in it, and who want to establish strong and general *links* among mathematics, physics, chemistry, and biology. An effort was made to expose the ideas, not only to students' minds but also to their hearts.

It is good to recall from time to time that all of us, physicists, chemists, and biologists, share the same electrons and nuclei as the objects of our investigation. It sounds trivial, but sometimes there is an impression that these disciplines investigate three different worlds. In the triad physics–chemistry–biology, chemistry plays a bridging role. By the middle of the 20th century, chemistry had closed the period of the exploration of its basic building blocks: elements and chemical bonds, their typical lengths, and typical values of angles between chemical bonds. Future discoveries in this field are not expected to change our ideas fundamentally. Now we are in a period of using this knowledge for the construction of what we only could dream of. In this chapter I will refer now and then to mathematicians and mathematics, who deal with ideal worlds. For some strange reason, at the foundation of (almost[5]) everything there are logic and mathematics. Physics, while describing the real rather than the ideal world, more than other natural sciences is symbiotic with mathematics.

[5] Yes, almost; e.g., generosity is not included here.

7.1 Multilevel supramolecular structures (statics)

7.1.1 Complex systems

Even a relatively simple system (e.g., an atom) often exhibits strange properties. Understanding simple objects seemed to represent a key for description of complex systems (e.g., molecules). Complexity *can* be explained using the first principles.[6] However, the complexity itself may add some important features. In a complex system some phenomena may occur, which would be extremely difficult to foresee from knowledge of their component parts. Most importantly, sometimes the behavior of a complex system is universal, i.e., independent of the properties of the parts of which it is composed and related to the very fact that the system consists of many small parts interacting in a simple way.

The behavior of a large number of argon atoms represents a difficult task for theoretical description, but is still quite predictable. When the number of atoms increases, they pack together in compact clusters similar to those we would have with the densest packing of tennis balls (the maximum number of contacts). We may have to do here with phenomena that are complicated (similar to chemical reactions) and connected to the different stability of the clusters (e.g., "magic numbers" related to particularly robust closed shells[7]). Yet, the interaction of the argon atoms, however difficult for quantum mechanical description, comes from the quite primitive two-body, three-body, etc., interactions (Chapter 5).

7.1.2 Self-organizing complex systems

Chemistry offers a plethora of intermolecular interactions.

Some intermolecular interactions are specific, i.e., a substrate A interacts with a particular molecule B_i from a set B_1, B_2, ... B_N (N is large) much more strongly than with others. The reasons for this are their shape, the electric field[8] fitness, a favorable hydrophobic interaction, etc., resulting either in the "key-lock" or "hand-glove" types of interaction (cf. Chapter 5). A molecule may provide a set of potential contacts localized in space (called synthon), which may fit to another synthon of another molecule.

This idea is used in supramolecular chemistry.[9] Suppose a particular reaction does not proceed with sufficient yield. Usually the reason is that to run just this reaction the molecules have

[6] In the 1920s, after presenting his equation, Paul Dirac said that now chemistry is explained. Yet, from the equation to foreseeing the function of the ribosome in our body is a long, long way.

[7] Similar closed shells are observed in nuclear matter, where the "tennis balls" correspond to nucleons.

[8] Both molecules carry their charge distributions, their interaction at a certain geometry may considerably lower the Coulombic energy.

[9] C.J. Pedersen, *J. Am. Chem. Soc*., 89(1967)2495, 7017; B. Dietrich, J.-M. Lehn, J.-P. Sauvage, *Tetrahedron Lett.*, (1969)2885, 2889; D.J. Cram, J.M. Cram, *Science (Washington)*, 183(1974)803.

to find themselves in a very specific position in space (a huge entropy barrier to overcome), but before this happens they undergo some unwanted reactions. We may however "instruct" the reactants by substituting them with such synthons that the latter lock the reactants in the right position in space. The reaction we want to happen becomes inevitable. The driving force for all this is the particularly high interaction energy of the reactants. Very often however, the interaction energy has to be high, but not too high, in order to enable the reaction products to separate. This reversibility is one of the critically important features for "intelligent" molecules, which could adapt to external conditions in a flexible way.

If the system under consideration is relatively simple, even if the matching of corresponding synthons is completed, we would still have a relatively primitive spatial structure. However, we may imagine far more interesting situations, when:

- The molecules are *chosen* in such a way as to ensure *some* intermolecular interaction is particularly attractive. A specific matching is called molecular recognition.
- The molecular complexes formed this way may recognize themselves again by using synthons previously existing or created *in situ*. In this way a multilevel structure can be formed, each level characterized by its own stability (cf. p. 424).
- The multilevel molecular structure may depend very strongly on its environment. When this changes, the structure may decompose, and eventually another structure may emerge.

Therefore,

> a *hierarchical* multilevel structure is formed, where the levels may exhibit different stability with regard to external perturbations.

An example is shown in Fig. 7.1.

There is nothing accidental in this system. The helices are composed of such amino acids that ensure the external surface of the helices is hydrophobic, and therefore enter the hydrophobic lipid bilayer of the cell walls. The peptide links serve to *recognize* and dock some particular signaling molecules. The seven-helix systems serve in biology as a universal sensor, with variations to make them specific for some particular molecular recognition and the processes that occur afterwards. After docking with a ligand or by undergoing photochemical isomerization of the retinal, some conformational changes take place, which after involving several intermediates, finally results in a signal arriving at a nerve cell. We see what this structure is able to do in dynamics, not statics.

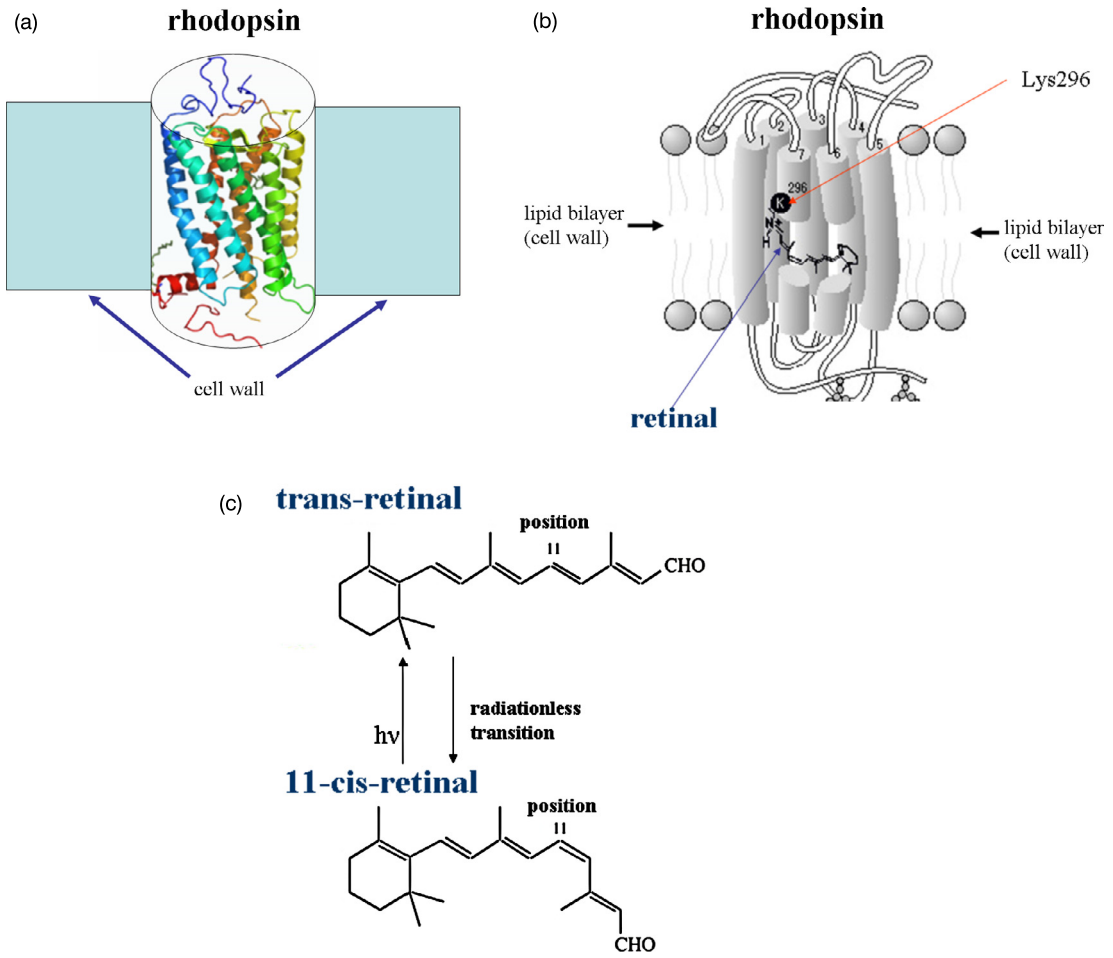


Fig. 7.1. A "universal" biological sensor based on rhodopsin (protein), a schematic view. (a) The sensor consists of seven α-helices (shown here as ribbons) connected in a sequential way by some oligopeptide links. The molecule is anchored in the cell wall (lipid bilayer), due to the hydrophobic effect: the rhodopsin's lipophilic amino acid residues are distributed on the rhodopsin surface. (b) The α-helices (this time shown for simplicity as cylinders) form a cavity. Some of the cylinders have been cut out to display a cis-retinal molecule bound (in one version of the sensor) by the Schiff bond (an ionic bond) to amino acid 296 (lysine denoted as K, in helix 7). (c) The cis-retinal (a chain of alternating single and double bonds) is able to absorb a photon and change its conformation to trans (at position 11). This triggers the cascade of processes responsible for our vision. The protruding protein loops exhibit specific interactions with some drugs. Such a system is at the basis of interaction with about 70% of drugs.

7.1.3 Cooperative interactions

Some events may cooperate. Suppose we have an extended object, which may undergo a set of events, A, B, C, ..., each taking place separately and locally with a small probability. However, it may happen that for a less extended object the events cooperate, i.e., event A makes it easier for event B to occur, and when A and then B happens this makes it easier for event C to happen, etc.

Self-organization is possible without cooperativity, but cooperativity may greatly increase its effectiveness. The hemoglobin molecule may serve as an example of cooperativity in intermolecular interactions, where its interaction with the first oxygen molecule makes its interaction with the second easier despite a considerable separation of the two binding events in space.

An example is shown in Fig. 7.2.

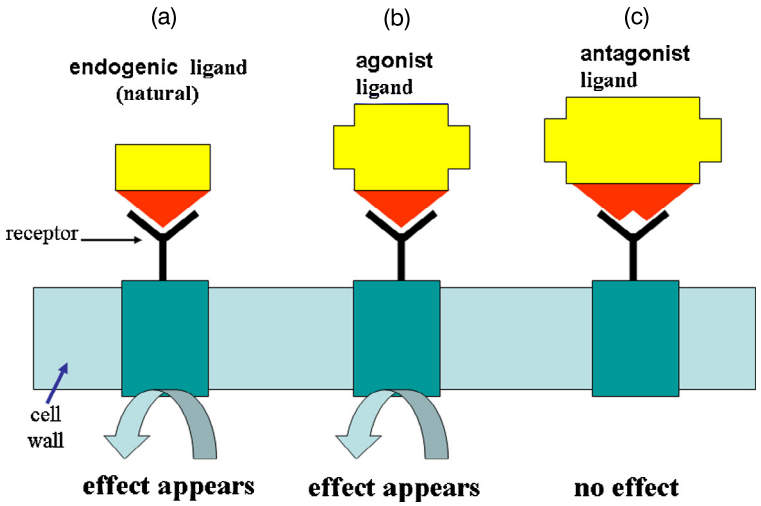


Fig. 7.2. Some examples of cooperativity (appearance of a phenomenon makes another phenomenon easier to occur). (a) Molecular recognition of synthons (upper side, endogenic ligand–receptor) causes an effect, e.g., release of a signaling molecule. (b) A similar perfect molecular recognition, the effect appears, in spite of the fact that the ligand is not endogenic (known as agonist – its shape differs from the natural, i.e., endogenic one). (c) An example of an antagonist: the recognition is so imperfect that the effect does not appear, although the ligand is blocking the receptor.

Fig. 7.2a shows a perfect molecular recognition of an endogenic (i.e., functioning in living organism, native) ligand by the receptor, followed by the release of a signaling molecule. (b) An agonist type of ligand differs from the native one, but performs the same function. (c) An imperfect recognition: the ligand (antagonist) binds, but fails to release the signaling molecule.

7.1.4 Combinatorial chemistry – molecular libraries

Jean-Marie Lehn (born 1939), French chemist, emeritus professor at the University of Strasbourg. Lehn, together with Pedersen and Cram (all three received the Nobel Prize for Chemistry in 1987), changed the paradigm of chemistry by stressing the importance of molecular recognition, which developed into the field he termed supramolecular chemistry. By proposing dynamic molecular libraries, Lehn broke with another long-standing idea, that of pure substances as the only desirable products of chemical reaction, by stressing the potential of diversity and instructed mixtures.

Chemistry is often regarded as dealing with pure, uniquely defined substances, which is obviously very demanding. Such role of chemistry was crucial in analysis, determinig the rules of chemistry and then in checking this knowledge in synthesis of new substances. Nature does not operate that way, it usually works on mixtures of substances. Nowadays there are cases, when a chemist is interested also in a mixture of all possible isomers instead of a single isomer or in a mixture of components that can form different complexes.

A complex system in a labile equilibrium may adjust itself to an external stimulus by changing its molecular composition. This is known as the dynamic combinatorial library,[10] but in fact it corresponds rather to a quasidynamics for we are interested mainly in shifting equilibrium due to an external stimulus. Liquid water may be regarded as a molecular combinatorial library of various clusters, all of them being in an easy-to-shift equilibrium. This is why water is able to hydrate a nearly infinite variety of molecular shapes, shifting the equilibrium towards the clusters that are just needed to wrap the solute by a water coat.

The immune system in our body is able to fight against a lot of enemies and win, irrespective of their shape and molecular properties (charge distribution). How is this possible? Would the organism be prepared for everything? Well, yes and no.

Let us imagine a system of molecules (building blocks) having some synthons and able to create some van der Waals complexes (Fig. 7.3). Since the van der Waals forces are quite weak, the complexes are in dynamic equilibrium. All complexes are present in the solution, possibly none of the complexes dominates.

[10] I. Huc., J.-M. Lehn, *Proc. Natl. Acad. Sci. (USA)*, 94(1997)2106.

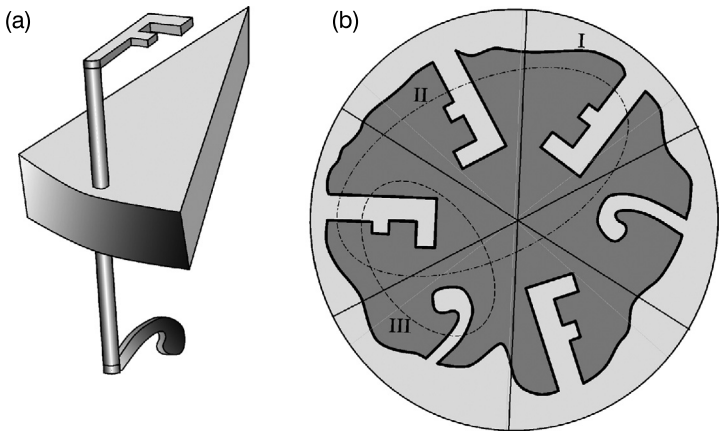


Fig. 7.3. A model of the immune system. (a) The figure shows schematically some monomers in a solvent. They have the shape of a piece of pie with two synthons: protruding up and protruding down, differing in shape. The monomers form some side-by-side aggregates containing from two to six monomers, each aggregate resulting in some pattern of synthons. We have then a library of all possible associates in thermodynamic equilibrium. Say, there are plenty of monomers, a smaller number of dimers, even fewer trimers, etc., up to a tiny concentration of hexamers. (b) The attacking factor I (the irregular body shown) is best recognized and bound by one of the hexamers. If the concentration of I is sufficiently high, the *equilibrium among the aggregates shifts towards the hexamer mentioned above, which therefore binds all the molecules of I*, making them harmless. If the attacking factor was II and III, binding could be accomplished with some trimers or dimers (as well as some higher aggregates). The defence is highly specific and at the same time highly flexible (adjustable).

Now, let us introduce some "enemy molecules." The building blocks use part of their synthons for binding the enemies (that have complementary synthons), and at the same time bind among themselves in order to strengthen the interaction. Some of the complexes are especially effective in this binding. Now, the Le Chatelier rule comes into game and the equilibrium shifts to produce as many of the most effective binders as possible. On top of this, the most effective binder may undergo a chemical reaction that replaces the weak van der Waals forces by strong chemical forces (the reaction rate is enhanced by the supramolecular interaction). The enemy was tightly secured, the invasion is over.[11]

[11] A simple model of immunological defense, similar to that described above, was proposed by F. Cardullo, M. Crego Calama, B.H.M. Snelling-Ruël, J.-L. Weidmann, A. Bielejewska, R. Fokkens, N.M.M. Nibbering, P. Timmerman, D.N. Reinhoudt, *J. Chem. Soc. Chem. Commun*., 367(2000).

7.2 Chemical feedback – a steering element (dynamics)

Steering is important for stability of the system and for its desirable reaction upon the external conditions imposed. The idea of feedback is at the heart of any steering, since it uses to control the output of a device to correct its input data. In a sense the feedback means also a self-using of information about itself.

7.2.1 A link to mathematics – attractors

Systems often exhibit dynamic, or time-dependent, behavior (chemical reactions, nonequilibrium thermodynamics).

Dynamic systems have been analyzed first in mathematics. When applying an iterative method of finding a solution to an equation one first decides which operation is supposed to bring us closer to the solution as well as what represents a reasonable zero-order guess (a starting point being a number, a function, a sequence of functions). Then one forces an evolution ("dynamics") of the approximate solutions by applying the operation first to the starting point, then to the result obtained by the operation on the starting point, and then again and again until the convergence is achieved.

As an example may serve the equation $\sin(x^2 + 1) - x = 0$. There is an iterative way to solve this equation numerically: $x_{n+1} = \sin(x_n^2 + 1)$, where n stands for the iteration number. The iterative scheme means choosing any x_0, and then applying many times a sequence of four keys on the calculator keyboard (square, +, 1, sin). The *result* (0.0174577) *is independent of the starting point chosen*. The number 0.0174577 represents an attractor or a *fixed point* for the operation. As a chemical analog of the fixed point may serve the thermodynamic equilibrium of a system (e.g., solving a substance in water), the same attained from any starting point (e.g., various versions of starting solution).

Mitchell Feigenbaum (born 1944), American physicist, employee of the Los Alamos National Laboratory, then professor at Cornell University and at Rockefeller University. Feigenbaum discovered attractors after making some observations just playing with a pocket calculator.

There are other types of attractors. Already in 1881 Jules Henri Poincaré has shown that a wide class of two coupled nonlinear differential equations leads to the solutions that tend to a particular oscillatory behavior *independently of the initial conditions* (*limit cycle*).

Jules Henri Poincaré (1854–1912), French mathematician and physicist, professor at the Sorbonne, made important contributions to the theory of differential equations, topology, celestial mechanics, and probability theory and the theory of functions.

Let us take an example of a set of two coupled differential equations. We have

$$\dot{r} = \mu r - r^2,$$
$$\dot{\theta} = \omega + br^2$$

(a dot over a symbol means a derivative with respect to time t), where r and θ are polar coordinates on a plane (Fig. 7.4) and $\omega, \mu, b > 0$ are constants. Just by looking at the equations we recognize what will happen. When r is small, one may neglect the r^2 term with respect to μr. The resulting equation $\dot{r} = \mu r > 0$ means that r starts to *increase* (while bringing a constant angular velocity $\dot{\theta} = \omega$). However, when r gets large enough, the $-r^2$ term starts to enter the game, and this means an increasing tendency to *diminish* r. As a result a compromise is achieved and one gets a stable trajectory.

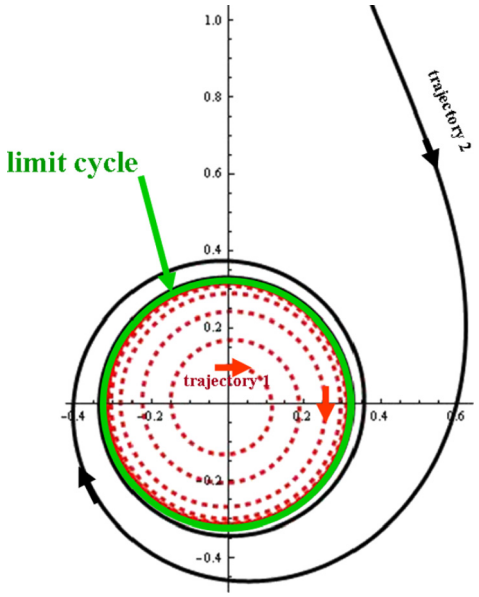


Fig. 7.4. An example of a limit cycle. Two special solutions (trajectories) of a set of two differential equations. Trajectories 1 (dashed line) and 2 (solid line) correspond to a start from a point close to the origin and from a point very distant from the origin, respectively. As $t \to \infty$ the two trajectories merge to a single circle (limit cycle).

Whatever the starting r and θ are, the trajectory tends to the circle of radius $r = \sqrt{\mu}$, and is rotating at a constant velocity $\dot{\theta} = \omega + b\mu$. This circle represents a limit cycle. Thus, *a limit cycle may be viewed as a feedback, a prototype of any steering, also a chemical one*.

Steering of chemical concentrations is at the heart of how the biological systems control the concentrations of thousands of substances.

This type of behavior has been anticipated by a work in the domain of mathematics, already seeking some connection to chemistry (not yet prepared for that). In 1910 Alfred Lotka proposed some differential equations that corresponded to the kinetics of an autocatalytic chemical reaction,[12] and then with Vito Volterra gave a differential equation that describes a general feedback mechanism (oscillations) known as the Lotka–Volterra model.

7.2.2 Bifurcations[13] and chaos

Nonlinear dynamics turned out to be extremely sensitive to coupling with some external parameters (representing the "neighborhood").

Let us take what is called the *logistic equation*

$$x = Kx(1 - x),$$

where $K > 0$ is a constant. This is a quadratic equation and there is no problem in solving it by the traditional method. However, here we will focus *on an iterative scheme*

$$x_{n+1} = Kx_n(1 - x_n),$$

which is obviously related to the iterative solution of the logistic equation. The biologist Robert May gave a numerical exercise to his Australian graduate students. They had to calculate how a rabbit population evolves when we let it grow according to the rule $x_{n+1} = Kx_n(1 - x_n)$, where the natural number n denotes the current year, while x_n stands for the (relative) population of, say, rabbits in a given and limited-area field, $0 \leq x_n \leq 1$. The number of the rabbits in year $(n + 1)$ is proportional to their popu-

[12] A.J. Lotka, *J. Phys. Chem.*, 14(1910)271.

[13] A bifurcation (corresponding to a parameter p) denotes in mathematics a doubling of an object when the parameter exceeds a value p_0. For example, when the object corresponds to the number of solutions of equation $x^2 + px + 1 = 0$, then the bifurcation point $p_0 = 2$. Another example of bifurcation is branching of roads, valleys, etc.

lation in the preceding year (x_n), because they reproduce very fast, but the rabbits eat grass and the field has a finite size. The larger x_n, the lower the amount of grass to eat, which makes the rabbits a bit weaker and less able to reproduce (this effect corresponds to $1 - x_n$).

The logistic equation contains a feedback mechanism.

The constant K measures the population–grass coupling strength (low-quality grass means a small K). What interests us is the fixed point of this operation, i.e., the final population the rabbits develop after many years at a given coupling constant K. For example, for $K = 1$ the evolution leads to a steady self-reproducing population x_∞, and x_∞ depends on K (the larger K the larger x_∞). The graduate students took various values of K. *Nobody* imagined this quadratic equation could hide a mystery. This mathematical phenomenon was carefully investigated and the results were really amazing (Fig. 7.5).

If K were small ($0 \leq K < 1$, extremely poor grass), the rabbit population would simply vanish (the first part of Fig. 7.5). If K increased (the second part of the plot, $1 \leq K < 3$), the herd would survive, and would even flourish if the grass quality would improve (larger K). When K exceeded 3 this flourishing would give, however, an unexpected twist: instead of reaching a fixed point, the system would oscillate between two sizes of the population (every second year the population was the same, but two consecutive years have different populations). This resembles the limit cycle described above – the system just repeats the same cycle all the time.

Further increase in K introduces further qualitative changes. First, for $3 \leq K < 3.44948$ the oscillations have period two (*bifurcation*), then at $3.44948 \leq K < 3.5441$ the oscillations have period four (next bifurcation, the four-member limit cycle), then for $3.5441 \leq K < 3.5644$ the period is eight (next bifurcation).[14]

[14] Mitchell Feigenbaum was interested to see at which value K (n) the next bifurcation into 2^n branches occurs. It turned out that there is a certain regularity, namely, $\lim_{n\to\infty} \frac{K_{n+1}-K_n}{K_{n+2}-K_{n+1}} = 4.669201609... \equiv \delta$. To the astonishment of scientists, the value of δ turned out to be "universal," i.e., characteristic for many *very different* mathematical problems, and therefore reached a status similar to that of the numbers π and e. The numbers π and e satisfy the exact relation $e^{i\pi} = -1$, but so far no similar relation was found for the Feigenbaum constant. There is an *approximate* relation (used by physicists in phase transition theory) which is satisfied: $\pi + \tan^{-1} e^{\pi} = 4.669201932 \approx \delta$.

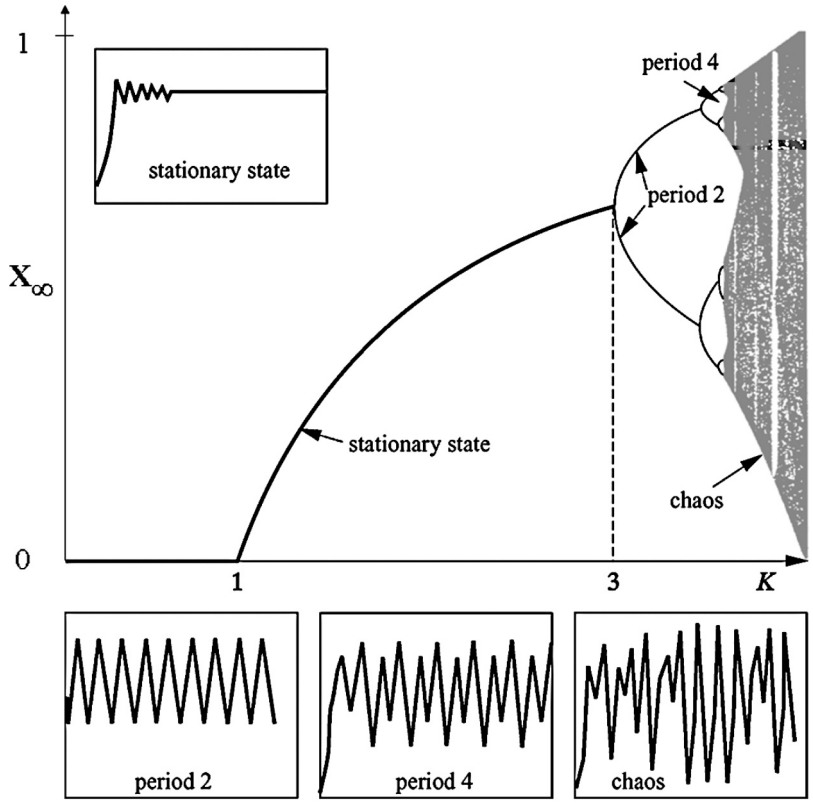


Fig. 7.5. The diagram of the fixed points and the limit cycles for the logistic equation as a function of the coupling constant K. From J. Gleick, "*Chaos*," Viking, New York, 1988, reproduced with permission of the author.

Then, the next surprise: exceeding $K = 3.56994$ we obtain populations that do not exhibit any regularity (no limit cycle, or just *chaos*). A further surprise is that this is not the end of the surprises. Some sections of K began to exhibit *odd-period* behavior, separated by some sections of chaotic behavior.

7.2.3 Brusselator without diffusion

Could we construct chemical feedback? What for? Those who have ever seen feedback working know the answer[15] – this is the very basis of control. Such control of chemical concentrations is at the heart of how biological systems operate.

[15] For example, an oven heats until the temperature exceeds an upper bound, then it switches off. When the temperature reaches a lower bound, the oven switches *itself* on (therefore, we have temperature *oscillations*).

Ilya Prigogine (1917–2003) Belgian physicist, professor at the Université Libre de Bruxelles. In 1977 he received the Nobel Prize "*for his contributions to non-equilibrium thermodynamics, particularly the theory of dissipative structures*." I wish to commemorate this great Belgian scientist, and *en passent*, just to indicate on the example of this stamp some regrettable problems menacing our culture.

The first idea is to prepare such a system in which an increase in the concentration of species X triggers the process of its decreasing. The decreasing occurs by replacing X by a very special substance Y, each molecule of which, when disintegrating, produces several X molecules. Thus we would have a scheme (X denotes a large concentration of X, x denotes a small concentration of X; similarly for the species Y): (X,y)→(x,Y)→(X,y), or *oscillations of the concentration of X and Y in time*.[16]

Imagine we carry out a complex chemical reaction in flow conditions,[17] i.e., the reactants A and B are pumped with a constant speed into a long narrow tube reactor, there is intensive stirring in the reactor, and then the products flow out to the sink (Fig. 7.6). After a while a steady state is established.[18]

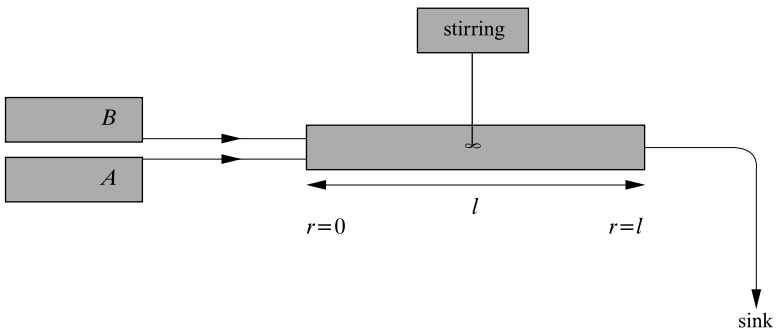


Fig. 7.6. A flow reactor (a narrow tube – in order to make a one-dimensional description possible) with stirring (no space oscillations in the concentrations). The concentrations of A and B are kept constant at all times (the corresponding fluxes are constant).

After A and B are supplied, the substances[19] X and Y appear, which play the role of catalysts, i.e., they participate in the reaction, but in total their amounts do not change. To model such a

16 Similar to the temperature oscillations in the feedback of the oven.

17 Such reaction conditions are typical for industry.

18 To be distinguished from the thermodynamic equilibrium state, where the system is isolated (no energy or matter flows).

19 Due to the chemical reactions running.

situation let us assume the following chain of chemical reactions[20]:

$$\begin{aligned} \mathrm{A} &\rightarrow \mathrm{X} \\ \mathrm{B}+\mathrm{X} &\rightarrow \mathrm{Y}+\mathrm{D} \\ 2\mathrm{X}+\mathrm{Y} &\rightarrow 3\mathrm{X} \\ \mathrm{X} &\rightarrow \mathrm{E} \\ \text{in total} &: \\ \mathrm{A}+\mathrm{B}+4\mathrm{X}+\mathrm{Y} &\rightarrow \mathrm{D}+\mathrm{E}+4\mathrm{X}+\mathrm{Y}. \end{aligned}$$

This chain of reactions satisfies our feedback postulates. In step 1 the concentration of X increases, in step 2 Y is produced at the expense of X, in step 3 substance Y enhances the production of X (at the expense of itself, this is an *autocatalytic step*), then again X transforms to Y (step 2), etc.

If we shut down the fluxes in and out, after a while a thermodynamic equilibrium is attained with all the concentrations of the six substances (A, B, D, E, X, Y; their concentrations will be denoted as A, B, D, E, X, Y, respectively) being constant in space (along the reactor) and time. On the other hand, when we fix the in and out fluxes to be constant (but nonzero) for a long time, we force the system to be in a steady state and as far from thermodynamic equilibrium as we wish. In order to simplify the kinetic equations, let us assume irreversibility of all the reactions considered (as shown in the reaction equations above) and put all the velocity constants equal to 1. This gives the kinetic equations for what is called the *Brusselator model* (of the reactor)

$$\begin{aligned} \frac{dX}{dt} &= A-(B+1)X+X^2Y, \\ \frac{dY}{dt} &= BX-X^2Y. \end{aligned} \tag{7.1}$$

These two equations, plus the initial concentrations of X and Y, totally determine the concentrations of all the species as functions of time (due to the stirring there will be no dependence on position in the reaction tube).

Steady state

A steady state (at constant fluxes of A and B) means $\frac{dX}{dt}=\frac{dY}{dt}=0$ and therefore we easily obtain the corresponding steady-state concentrations X_s, Y_s by solving Eq. (7.1), i.e.,

$$0=A-(B+1)X_s+X_s^2Y_s,$$

[20] See, e.g., A. Babloyantz, "*Molecules, Dynamics and Life*," Wiley, New York, 1987.

$$0 = BX_s - X_s^2 Y_s.$$

Please check that these equations are satisfied by

$$X_s = A,$$
$$Y_s = \frac{B}{A}.$$

Evolution of fluctuations from the steady state

Any system undergoes some spontaneous concentration fluctuations, or we may perturb the system by injecting a small amount of X and/or Y. What will happen to the stationary state found a while before if such a fluctuation happens?

Let us see. We have fluctuations x and y from the steady state, so

$$X(t) = X_s + x(t), \tag{7.2}$$
$$Y(t) = Y_s + y(t).$$

What will happen next?

After inserting (7.2) in (7.1) we obtain the equations describing how the fluctuations evolve in time, i.e.,

$$\frac{dx}{dt} = -(B+1)x + Y_s(2X_s x + x^2) + y(X_s^2 + 2xX_s + x^2), \tag{7.3}$$
$$\frac{dy}{dt} = Bx - Y_s(2X_s x + x^2) - y(X_s^2 + 2xX_s + x^2).$$

Since a mathematical theory for arbitrarily large fluctuations does not exist, we will limit ourselves to small x and y. Then all the quadratic terms of these fluctuations can be neglected (*linearization* of (7.3)). We obtain

$$\frac{dx}{dt} = -(B+1)x + Y_s(2X_s x) + yX_s^2, \tag{7.4}$$
$$\frac{dy}{dt} = Bx - Y_s(2X_s x) - yX_s^2.$$

Let us assume fluctuations of the form[21]

[21] Such a form allows for very versatile behavior, i.e., exponential growth ($\omega > 0$), decay ($\omega < 0$), or staying constant ($\omega = 0$), as well as periodic behavior ($\mathrm{Re}\omega = 0, \mathrm{Im}\omega \neq 0$), quasiperiodic growth, ($\mathrm{Re}\omega > 0, \mathrm{Im}\omega \neq 0$) or decay ($\mathrm{Re}\omega < 0, \mathrm{Im}\omega \neq 0$).

$$x = x_0 \exp(\omega t), \tag{7.5}$$
$$y = y_0 \exp(\omega t)$$

and represent particular solutions of (7.4) provided proper values of ω, x_0, and y_0 are chosen. After inserting (7.5) in Eqs. (7.4) we obtain the following set of equations for the unknowns ω, x_0, and y_0:

$$\omega x_0 = (B-1)x_0 + A^2 y_0, \tag{7.6}$$
$$\omega y_0 = -Bx_0 - A^2 y_0.$$

This represents a set of homogeneous linear equations with respect to x_0 and y_0, and this means we have to ensure that the determinant, composed of the coefficients multiplying the unknowns x_0 and y_0, vanishes (*characteristic equation*, cf. secular equation, in another, quite unrelated domain, p. V1-271)

$$\begin{vmatrix} \omega - B + 1 & -A^2 \\ B & \omega + A^2 \end{vmatrix} = 0.$$

This equation is satisfied by some special values[22] of ω, i.e.,

$$\omega_{1,2} = \frac{T \pm \sqrt{T^2 - 4\Delta}}{2}, \tag{7.7}$$

where

$$T = -\left(A^2 - B + 1\right), \tag{7.8}$$
$$\Delta = A^2. \tag{7.9}$$

Fluctuation stability analysis

Now it is time to pick the fruits of our hard work.

How the fluctuations depend on time is characterized by the roots $\omega_1(t)$ and $\omega_2(t)$ of Eq. (7.7), because x_0 and y_0 are nothing but some constant amplitudes of the changes. We have the following possibilities (Fig. 7.7, Table 7.1):

- Both roots are real, which happens only if $T^2 - 4\Delta \geq 0$. Since $\Delta > 0$, the two roots are of the same sign (sign of T). If $T > 0$, then both roots are positive, which means that the fluctuations $x = x_0 \exp(\omega t)$, $y = y_0 \exp(\omega t)$ *increase over time and the system will never return to the steady state ("unstable node")*. Thus the steady state represents a *repeller* of the concentrations X and Y.

[22] They represent an analog of the normal mode frequencies from Chapter V1-7.

Table 7.1. Fluctuation stability analysis, i.e., what happens if the concentrations undergo a fluctuation from the steady-state values. The analysis is based on the values of ω_1 and ω_2 from Eq. (7.7); they may have real (subscript r) as well as imaginary (subscript i) parts; hence we write $\omega_{r,1}, \omega_{i,1}, \omega_{r,2}, \omega_{i,2}$.

T	Δ	$T^2 - 4\Delta$	$\omega_{r,1}$	$\omega_{i,1}$	$\omega_{r,2}$	$\omega_{i,2}$	Stability
+	+	+	+	0	+	0	unstable node
−	+	+	−	0	−	0	stable node
−	+	0	−	0	−	0	stable stellar node
+	+	0	+	0	+	0	unstable stellar node
−	+	−	−	$i\omega$	−	$-i\omega$	stable focus
+	+	−	+	$i\omega$	+	$-i\omega$	unstable focus
0	+	−	0	$i\omega$	0	$-i\omega$	center of marginal stability

- If as in the previous case $T^2 - 4\Delta \geq 0$, but this time $T < 0$, then both roots are negative, and this means that the fluctuations from the steady state will vanish ("*stable node*"). It looks as if we had in the steady state an *attractor* of the concentrations X and Y.
- Now let us take $T^2 - 4\Delta = 0$, which means that the two roots are equal ("degeneracy"). This case is similar to the two previous ones. If the two roots are positive then the point is called the *stable stellar node* (attractor), if they are negative it is called the *unstable stellar node* (repeller).
- If $T^2 - 4\Delta < 0$, we have an interesting situation: both roots are complex conjugate $\omega_1 = \omega_r + i\omega_i$, $\omega_2 = \omega_r - i\omega_i$, or $\exp \omega_{1,2} t = \exp \omega_r t \exp(\pm i\omega_i t) = \exp \omega_r t (\cos \omega_i t \pm i \sin \omega_i t)$. Note that $\omega_r = \frac{T}{2}$. We have therefore three special cases:
  - $T > 0$. Because of $\exp \omega_r t$ we have, a monotonic increase in the *fluctuations*, and at the same time because of $\cos \omega_i t \pm i \sin \omega_i t$ the two concentrations oscillate. Such a point is called the *unstable focus* (and represents a repeller).
  - $T < 0$. In a similar way we obtain the *stable focus*, which means some damped vanishing concentration oscillations (attractor).
  - $T = 0$. In this case $\exp \omega_{1,2} t = \exp(\pm i\omega_i t)$, i.e., we have the *undamped oscillations* of *X and Y about the stationary point* X_s, Y_s, which is called, in this case, the *center of marginal stability*. This point is for us the most interesting: it explains the chemical oscillations like in Belousov–Zhabotinsky reaction.

Qualitative change

Can we qualitatively change the behavior of the reaction? Yes. It is sufficient just to change the concentrations A or B (i.e., to rotate the reactor taps). For example, let us gradually change B. Then, from (7.8) it follows that the key parameter T begins to change, which leads to an *abrupt qualitative change* in the behavior. Such changes may be of great importance (also in the sense of information processing), and as the control switch may serve to regulate the concentrations of some substances in the reaction mixture.

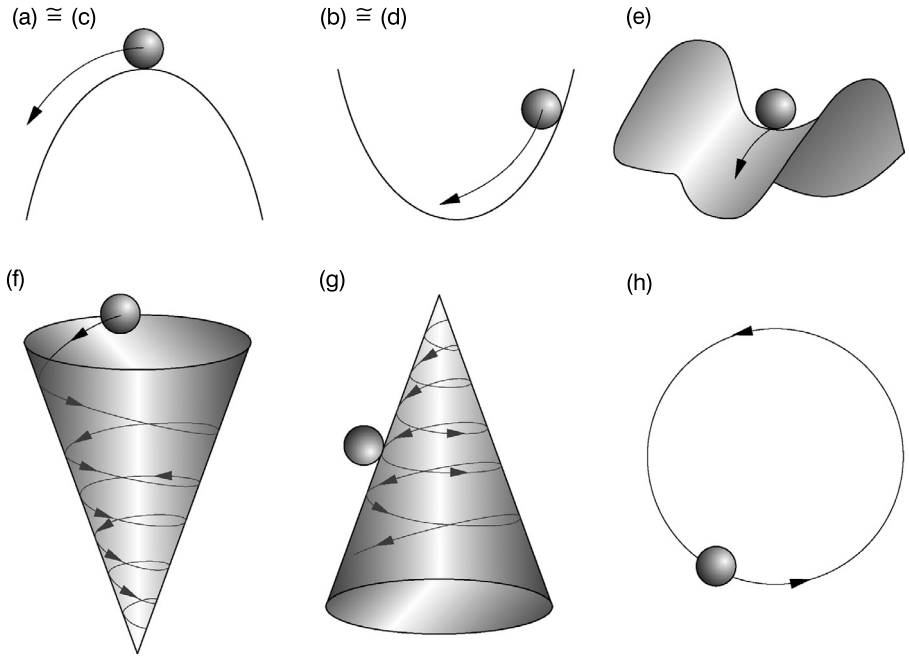


Fig. 7.7. Evolution types of fluctuations from the reaction steady state. The classification is based on the numbers ω_1 and ω_2 of Eq. (7.7). The individual figures correspond to the rows of Table 7.1. The behavior of the system (in the space of chemical concentrations) resembles sliding of a point or rolling a ball over certain surfaces in a gravitational field directed downward: (a) unstable node resembles sliding from the top of a mountain, (b) stable node resembles moving inside a bowl-like shape, (c) the unstable stellar node is similar to case (a), with a slightly different mathematical reason behind it, (d) similarly for the stable stellar node (resembles case (b)), (e) saddle – the corresponding motion is similar to a ball rolling over a cavalry saddle, (f) stable focus – the motion resembles rolling a ball over the interior surface of a cone pointing downward (attractor), (g) unstable focus – a similar rolling but on the external surface of a cone that points up (repeller), (h) center of marginal stability corresponds to a circular motion (oscillation) – a case most important for oscillatory (in time) chemical reactions.

7.2.4 Brusselator with diffusion – dissipative structures

If the stirrer were removed from the reactor, (7.1) would have to be modified by adding diffusion terms

$$\frac{dX}{dt} = A - (B+1)X + X^2Y + D_X\frac{\partial^2 X}{\partial r^2}, \tag{7.10}$$

$$\frac{dY}{dt} = BX - X^2Y + D_Y\frac{\partial^2 Y}{\partial r^2}. \tag{7.11}$$

A stability analysis similar to that carried out a moment before results *not only in oscillations in time, but also in space, i.e., in the reaction tube there are waves of the concentrations* of X and Y *moving in space* (*dissipative structures*). Now, look at the photo of a zebra (Fig. 7.8) and at the bifurcation diagram in the logistic equation (Fig. 7.5).

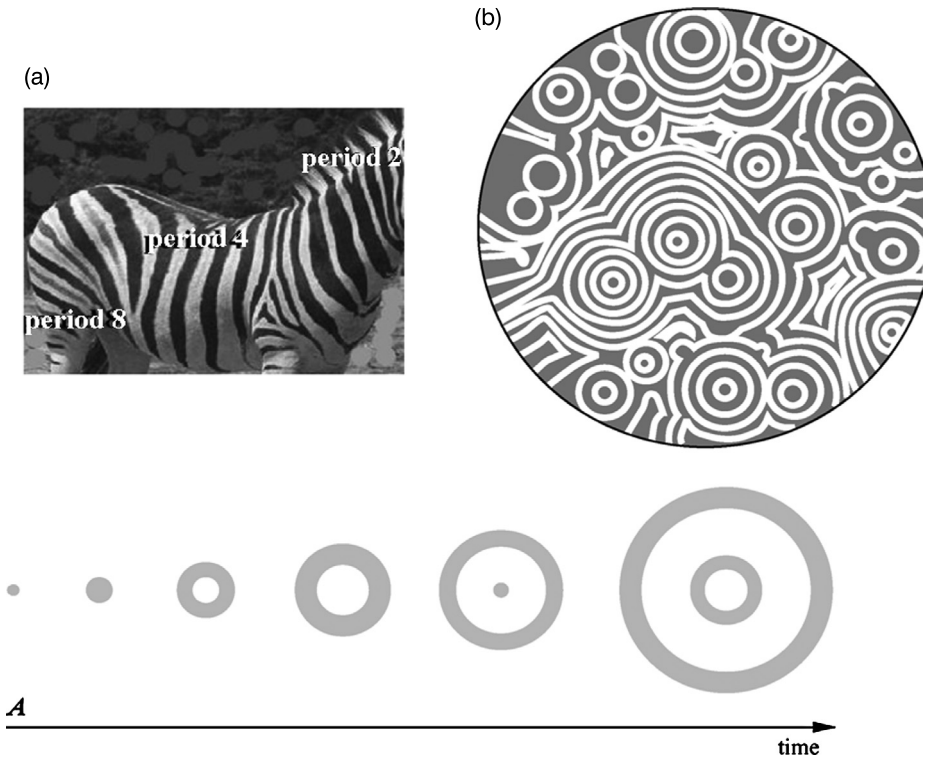


Fig. 7.8. Concentration waves. (a) Such an animal "should not exist." Indeed, how did the *molecules* know that they have to make a beautiful pattern? I looked many times at zebras, but only recently I was struck by the observation that what I see on the zebra's skin is described by the logistic equation. The skin on the zebra's neck exhibits quasiperiodic oscillations of the black and white color (period 2), in the middle of the zebra's body we have a *period doubling* (period 4), the zebra's back has period 8. (b) The waves of the chemical information (concentration oscillations in space and time) in the Belousov–Zhabotinsky reaction from several sources in space. A "freezing" (for any reason) of the chemical waves leads to a striking similarity with the zebra's skin. (c) Similar waves of an epidemic in a rather immobile society. The epidemic broke out in center A. Those who have contact with the sick person get sick, but after some time they regain their health, and *for some time* become immune. When the immune period is over these people get sick again, because there are a lot of microbes around. This is how the epidemic waves propagate.

7.2.5 Hypercycles

Let us imagine a system with a chain of consecutive chemical reactions. There are a lot of such reaction chains around, and it is difficult to single out an elementary reaction without such a chain being involved. They end up with a final product and everything stops. What would happen, however, if at a given point of the reaction chain, a substance X were created, the same as one of the reactants at a previous stage of the reaction chain? The X would take control over its own fate, by the Le Chatelier rule. In such a way, feedback would have been established, and instead of the chain we would have a catalytic cycle. A system with feedback may adapt to changing external conditions, reaching a steady or oscillatory state. Moreover, in our system a number of such independent cycles may be present. However, when two of them share a common reactant X, both cycles would begin to cooperate, usually exhibiting a very complicated stability/instability pattern or an oscillatory character. We may think of coupling many such cycles in a *hypercycle*, etc.[23]

Cooperating hypercycles based on multilevel supramolecular structures could behave in an extremely complex way when subject to variable fluxes of energy and matter. No wonder, then, that a single photon produced by the prey hidden in the dark and absorbed by the retinal in the lynx's eye may trigger an enormous variety of hunting behaviors. Or, maybe from another domain: a single glimpse of a girl may change the fate of many people,[24] and sometimes the fate of the world. This is because the retinal in the eye hit by the photon of a certain energy changes its conformation from cis to trans. This triggers a cascade of further processes, which end up as a nerve impulse traveling to the brain, and it is over.

7.2.6 From self-organization and complexity to information

Using multilevel supramolecular architectures one may tailor new materials exhibiting desired properties, e.g., adapting themselves to changes of the neighborhood ("smart materials"). The shape and the stability of such architectures may depend on more or less subtle interplay of external stimuli. Such materials by the chemical synthesis of their building blocks are taught to have a function to perform, i.e., an action like ligand binding and/or releasing, or transporting a ligand, an electron, or a photon. A molecule may perform several functions. Sometimes these functions may be *coupled*, getting a functional cooperativity.

A vast majority of chemistry nowadays deals with mastering the structure and/or studying how this structure behaves when allowing its dynamics in a closed system attaining equilibrium

[23] M. Eigen, P. Schuster, *Naturwissenschaften*, 11(1977), 1(1978), 7(1978).
[24] Well, think of a husband, children, grandchildren, etc.

(a "beaker-like" approach). In the near future, chemistry will face the nonlinear, far-from-equilibrium dynamics of systems composed of multilevel supramolecular architectures. To my knowledge no such endeavors have been undertaken yet. These will be very complex and very sensitive systems. It is probably a matter of several years to see such systems in action. It seems much more difficult for contemporary chemistry to control/foresee theoretically what kind of behavior to expect.

Biology teaches us that unbelievable effects are possible: molecules may spontaneously form some large aggregates with a very complex dynamics and the whole system is searching for energy-rich substances to keep itself running. The molecular functions of very many molecules may be coupled in a complex space-temporal relationship at several time and space scales involving enormous transport problems over distances of the size of our body. All this spans the time scale from femtoseconds to years, and on the spatial scale from angstroms until meters.

Future chemists will deal with molecular scenarios involving interplay of sophisticated, multilevel structures transforming in a nonlinear dynamic processes into an object that has a purpose, plays a certain complex role. The achievements of today, such as molecular switches, molecular wires, etc., are important, but they represent just simple elements of the space-temporal molecular interplay of tomorrow.

7.3 Information and informed matter

A body acts as a medium in which massive information processing is going on. What living organisms do is a permanent exchanging information not only at the level of an individual, but also at the molecular, cellular, and tissue levels. *The corresponding hardware and software are chemical systems.* It looks as if the chemical identity were much less important than the function the molecules perform. For example, the protein with the generic name cytochrome C is involved in electron transfer in all organisms, from yeast to humans. Each species has however a specific sequence of amino acids in its cytochrome C that differs from the cytochromes of all other species, the differences ranging from a single amino acid to 50% of amino acids. Yet the function is preserved in all these molecules.

Thus, the most advanced chemistry we are dealing with is used for information processing. Chemistry is still in a stage in which one does not consider quantitatively the exchange of information. Information became however an object of quantitative research in telecommunication. As soon as in 1924 Harry Nyquist studied the efficiency of information channels using a given number of electric potential entries in telegraphs. A few years later Ralph Hartley published an article on measuring information. Twenty years later Claude E. Shannon introduced the notion of information entropy.

7.3.1 Abstract theory of information

As a natural measure of the amount of information contained in a binary sequence (a message, e.g., 00100010...) of length N one may propose just the number N, i.e., the message length. To reconstruct the message one has to ask N questions: *is the next position equal to "*1*"?*

The number of all possible messages of length N is equal to $M = 2^N$. Hence the amount of information in a message (I, measured in bits) is

$$I \equiv N = \log_2 M.$$

Assumption of equal probabilities

If one assumes that the probability of picking out a particular message to be sent is the same for all the messages, $p = \frac{1}{M}$, one obtains the amount of information in a particular message as defined by Hartley,

$$I = -\log_2 p.$$

The receiver does not know M and therefore judges the amount of information (its importance in bits) by his own estimation of p *before* the particular information comes. Hence, I may be viewed as a measure of receiver's surprise that this particular information came true (a message about a marginally probable event contains a lot of information). *This means that the amount of information in a message does not represent a feature of this message, but in addition tells us about our knowledge about the message.* If one receives the same message twice, the amount of the first information coming is $I_1 = -\log_2 p_1$ (where p_1 is the probability of the event described by the message as judged by the receiver), while the second (identical) information carries $I_2 = -\log_2 1 = 0$, because there is no surprise. The situation becomes more ambiguous when there are several receivers, each of them having his own estimation of p. Thus, the amount of information received by each of them may be different.[25]

[25] On top of that it is not clear how to define p, i.e., a chance for some event to happen. The classical Laplace definition says that p can be calculated as frequencies if *one assumes not having any reason to think* that some events are more probable than other events. This means subjectivity. Such a procedure is useless in case no repeating is possible, like for estimating a chance to win a battle, etc. An alternative definition of p by Thomas Bayes is subjective as well ("a degree of *someone's conviction*"). Both definitions are used in practice.

The probabilities may differ – the Shannon entropy of information

Claude Elwood Shannon (1916–2001), American mathematician, professor at the Massachusetts Institute of Technology. His professional life was associated with the Bell Laboratories. His idea, now so obvious, that information may be transmitted as a sequence of "0" and "1" was shocking in 1948. It was said that Shannon used to understand problems "*in zero time*."

Suppose we have an alphabet of letters $a_1, a_2, ...a_m$. For a particular language these are the letters of the language ($m = 26$ for the English language), for the genetic code $m = 4$ (adenine, thymine, guanine, cytosine), for proteins $m = 20$ (the number of the native amino acids), etc. A letter a_i appears N_i times in a large object (animal, plant, the biosphere, a given language, etc.). Using the Laplace definition one may calculate for each letter its probability to appear in the object $p_i = \frac{N_i}{\sum_i N_i}$ for $i = 1, 2, ...m$. Let a long message (e.g., a letter, a portion of DNA, a protein) contain N such letters. The number of different messages of length N one is able to construct from a given set of letters can be computed if we know how many times every letter appears in the message. One may estimate these numbers from the probabilities p_i if one assumes that the message is not only long, but also typical for the language. Then the letter a_i will appear most probably Np_i times. Since the permutations of the same letters do not lead to different messages, the number of different messages of length N is

$$M = \frac{N!}{(Np_1)!(Np_2)!...(Np_m)!}. \tag{7.12}$$

By analogy with the case of messages of the same probability one defines the expected amount of information as the *entropy of information*, or the *Shannon information entropy*[26]

[26] We use the Stirling formula, valid for large N, i.e., $\ln N! \approx N \ln N - N$. Indeed,

$$\log_2 \frac{N!}{(Np_1)!(Np_2)!...(Np_m)!} = \log_2 N! - \sum_i \log_2 (Np_i)! = \frac{\ln N!}{\ln 2} - \sum_i \frac{\ln (Np_i)!}{\ln 2} \approx$$
$$\frac{N \ln N - N}{\ln 2} - \sum_i \frac{Np_i \ln (Np_i) - Np_i}{\ln 2} = \frac{N \ln N}{\ln 2} - \sum_i \frac{Np_i \ln (Np_i)}{\ln 2} =$$
$$\frac{1}{\ln 2}\left[N \ln N - \sum_i Np_i \ln N - \sum_i Np_i \ln p_i \right] = -N \sum_i p_i \log_2 p_i .$$

$$\langle I\rangle = \log_2 M = \log_2 \frac{N!}{(Np_1)!(Np_2)!...(Np_m)!} \approx N\sum_i^m \left[-p_i \log_2 p_i\right].$$

The expected amount of information per letter of the message is therefore

$$\langle I\rangle_N \equiv \frac{\langle I\rangle}{N} \approx \sum_{i=1}^{m} \left[-p_i \log_2 p_i\right]. \tag{7.13}$$

Properties of the information entropy

- Here $\langle I\rangle_N$ may be treated as a function of a discrete probability distribution,[27] $\langle I\rangle_N = \langle I\rangle_N (p_1, p_2, ...p_m)$, or as a functional of a continuous distribution $p(x)$, x being the random variable or variables.
- The information entropy attains a *minimum* (equal to 0) for any of the following situations: $\langle I\rangle_N (1, 0, ...0) = \langle I\rangle_N (0, 1, ...0) = \langle I\rangle_N (0, 0, ...1) = 0$.
- The information entropy attains a *maximum* for uniform probability distribution $\langle I\rangle_N (\frac{1}{m}, \frac{1}{m}, ...\frac{1}{m})$ and in such a case $\langle I\rangle_N = \sum_{i=1}^{m} \left[-\frac{1}{m}\log_2 \frac{1}{m}\right] = \log_2 m$.

Thus, the information entropy, $N\langle I\rangle_N$ or $\langle I\rangle_N$, represents a measure of our lack of knowledge about which particular message one may expect given the probability distribution of random variables. If the probability distribution is of the kind $(1, 0, ...0)$, the message is fully determined, i.e., $a_1a_1a_1...a_1$ (N times), and the lack of our knowledge is equal to 0. The worst case, i.e., the largest lack of knowledge of what one may expect as a message, corresponds to a uniform distribution (maximum of information entropy).

7.3.2 Teaching molecules

Molecular recognition represents one of possible areas of application of information theory. Suppose we consider the formation of a molecular complex composed of two molecules, A and E. Molecule E is considered as a "teacher" and molecule A as a "student." The teacher molecule does not change its structure, while the student molecule will be able to adapt its structure to satisfy the teacher. If all possible stable interaction energy minima on the potential energy surface (PES) are of the same depth, the lack of knowledge (information entropy) which particular configuration one observes reaches maximum. If the depth of one energy well is very much larger than the depth of other energy valleys, the calculated Boltzmann factor normalized to probability is very close to 1 and the corresponding Shannon entropy reaches minimum. In

[27] It is a symmetric function of variables $p_i > 0$, i.e., $\langle I\rangle_N (p_1, p_2, ...p_m) = \langle I\rangle_N (p_2, p_1, ...p_m)$, similarly for all other permutations of variables.

all other cases, calculating the Boltzmann probability distribution one may calculate the information entropy $\langle I_{AE} \rangle$ for A recognizing E. Then molecule A may be modified chemically (yielding molecule B) and the new molecule ("a better prepared student") may fit molecule E better. Also in this case one can calculate the information entropy $\langle I_{BE} \rangle$. If $\langle I_{BE} \rangle < \langle I_{AE} \rangle$, the change from A to B means acquiring an amount of knowledge by the first molecule (that depends on temperature through the Boltzmann probability) and we can calculate the information associated to this *molecular lesson* as $\Delta I = \langle I_{AE} \rangle - \langle I_{BE} \rangle$, also as a function of temperature. Such a lesson is meaningless if one considers the educated molecule alone, without any "teacher" to check the level of education. Information is related to the teaching-and-learning process through molecular interaction.

Fig. 7.9 shows an example of such a teaching[28] in a very simple model of the teacher molecule E and the taught molecules: A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D (in three lessons: A $\rightarrow$ B, B $\rightarrow$ C, C $\rightarrow$ D). The model is simple, because we intend to bring (as far as we know for the first time) the idea of molecular teaching, rather than being interested in particular molecules, not speaking about currently prohibitively time consuming study of several PESs for real molecules of necessarily considerable size (both in quantum mechanical *ab initio* or semiempirical-type calculations). The following simplifying assumptions have been made in our model:

- The "model molecules" are planar objects (see Fig. 7.9) and approach each other within a common plane, with possible reversing of the plane of one of the molecules.
- For the sake of simplicity in counting the contact positions each molecule on the contact side is built in a certain spatial resolution (the square structure unit shown at the bottom of Fig. 7.9).
- The interaction energy is zero unless the molecules form a contact along a line length equal or larger than the structure unit side. The interaction energy is assumed to be equal to $\frac{1}{3}n$ (in kcal/mol), where the natural number n is the number of contact units.

All possible contacts have been taken into account and the normalized Boltzmann probability distributions, as well as the Shannon information entropies, have been calculated for each pair of the molecules (i.e., AE, BE, CE, DE) as functions of temperature T within the range from 100 K to 400 K. The results (Fig. 7.9) show that student molecule A (with the simplest structure) does not recognize molecule E at any temperature T under study. The ignorance (Shannon's entropy) of A with respect to E is of the order of 5 bits. Transforming molecule A into B improves much the situation, but only at low temperatures, teaching the molecule by 5 bits at $T = 100$ K, at room temperature by about 2 bits, and by only 1 bit at $T = 400$ K. The second lesson (teaching B to C) makes the recognition almost perfect below 200 K and very good below 300 K. Finally,

[28] L. Piela, *Rep. Advan. Study Instit.*, Warsaw Technical University, 2(2012).

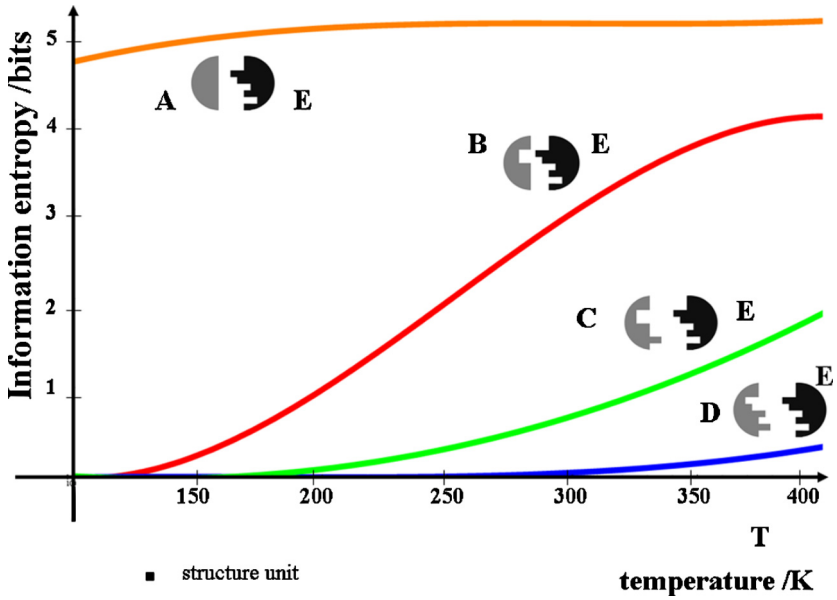


Fig. 7.9. A model of the temperature-dependent molecular teaching. The left-hand side object ("student molecule") recognizes the shape of the right-hand side molecule ("teacher molecule" E) in three steps ("lessons"): A → B → C → D. Each of the complexes AE, BE, CE, DE is characterized by its state of ignorance (Shannon information entropy), while each lesson (A → B, B → C, C → D) is characterized by the knowledge acquired (in bits).

the last step (C → D) gives perfect matching below 300 K and still very good even at 400 K.

7.3.3 Dynamic information processing of chemical waves

Mathematical model

The Belousov–Zhabotinsky (BZ) reactive solution can be prepared in a special stable state that exhibits a remarkable sensitivity ("excitable state"). When a local small perturbation is applied in such a state, the system comes back to the initial state very quickly, but when the local perturbation exceeds a certain threshold, the system undergoes very large changes before a long way of coming back to the stable state. Another remarkable feature of such a system is that it becomes refractable to any consecutive excitation (at this locality) during a certain time after the first excitation (the refraction time). A local perturbation may propagate further in space, while the originally perturbed place relaxes to the stationary state. *This means profile preserving traveling waves are possible.*

Boris Pavlovich Belousov (1893–1970) looked for an inorganic analog of the biochemical Krebs cycle. The investigations began in 1950 in a Soviet secret military institute. Belousov studied mixtures of potassium bromate with citric acid, and a small admixture of a catalyst: a salt of cerium ions. He expected a monotonic transformation of the yellow Ce^{4+} ions into the colorless Ce^{3+}. Instead, he found oscillations of the color of the solvent (colorless–yellow–colorless, etc., also called by Russians "vodka–cognac–vodka..."). He wrote a paper and sent it to a Soviet journal, but the paper was rejected with a referee's remark that what the author had described was simply impossible. His involvement in classified research caused him to limit himself to bringing (by intermediacy of somebody) a piece of paper with reactants and his phone number written on it. He refused to meet anybody. Finally, Schnoll persuaded him to publish his results. Neither Schnoll nor his PhD student Zhabotinsky ever met Belousov, though all they lived in Moscow.

Belousov's first chemistry experience was at the age of 12, while engaged in making bombs in the Marxist underground. Stalin thought of everything. When, formally underqualified, Belousov had problems as head of the lab, a hand-written message by Stalin, written with a blue pencil on an ordinary piece of paper, "*Has to be paid as a head of laboratory as long as he has this position*," worked miracles.
After S.E. Schnoll, "*Gheroy i zladieyi rossiyskoy nauki*," Kron-Press, Moscow, 1997.

Such a system can be described by a mathematical model known as the FitzHugh–Nagumo (FHN) scheme, originally applied to the propagation of electric excitation pulses in nerve tissues.[29] In the chemical context these (mathematical) waves may correspond to time- and space-dependent concentration of some particular substance.

The BZ system may be viewed as a medium able to process information, e.g., an excess of a chemical species at a point of the reactor may be treated as "1" (or "true") and its deficit as "0" (or "false"). Alternatively, the arrival or the absence of a wave in a predefined part of the reactor may be interpreted in a similar way. This was the basis of chemical reactors using the BZ reaction, which can behave as logical gates (AND, OR, NOT).[30]

Following Sielewiesiuk and Górecki[31] we consider here the FHN model of the BZ reaction in the case of a square planar reactor containing a thin layer of the BZ reactive solution (the

29 R. FitzHugh, *Biophysics J.*, 1(1961)445; J. Nagumo, S. Arimoto, S. Yoshizawa, *Proc. IRE*, 50(1962)2061.

30 A. Toth, K. Showalter, *J. Chem. Phys.*, 103(1995)2058; O. Steinbock, P. Kettunen, K. Showalter, *J. Phys. Chem.*, 100(1996)18970.

31 J. Sielewiesiuk, J. Górecki, *Acta Phys. Pol. B*, 32, 1589(2001); J. Sielewiesiuk, J. Górecki, *GAKUTO Intern. Series Math. Sci. Appl.*, 17(2001).

position in the reactor is given by (x, y), the solution's depth is assumed to be negligible). Two quantities, $u\,(x, y)$ and $v\,(x, y)$, denote the concentration amplitudes of two chemical substances, one called the *activator* (which is present in the solution) and the other the *inhibitor* (which is uniformly immobilized on the bottom of the reactor). The time evolution of their values is modeled by the following set of FHN equations[32]:

$$\tau \frac{\partial u}{\partial t} = -\gamma\,[ku(u-\alpha)(u-1)+v] + \tau D_u \nabla^2 u, \tag{7.14}$$

$$\frac{\partial v}{\partial t} = \gamma u, \tag{7.15}$$

where t stands for time and τ, k, α, D_u are constants, the last one being the diffusion constant for the activator (there is no diffusion of the inhibitor). The system of Eqs. (7.14) and (7.15) was found to correspond to an excitable state of the BZ solution for the following values of the parameters[33]: $k = 3, \tau = 0.03, \alpha = 0.02, D_u = 0.00045$. The quantity γ is a parameter that defines the architecture of the reactor; $\gamma = 1$ is set everywhere in the reactor, except the predefined regions called "passive" ones, where $\gamma = 0$. In the passive regions no production of the activator, only its diffusion, takes place according to Eq. (7.14). The passive regions are inhibitor-free, since $v = const$ follows from Eq. (7.15), and the constant is set to zero.

Reactor's geometry

After the excitation, a traveling wave appears that propagates freely in the regions with $\gamma = 1$, while the passive regions ($\gamma = 0$) represent for the wave a kind of barrier to overcome. The wave penetrates the passive region. For a passive stripe of a certain width the penetration depth depends on the stripe/incident wave impact angle,[34] the most efficient being a perpendicular impact (with the parallel wave front and the stripe):

- if the stripe is too wide, the wave disappears,
- if the stripe is sufficiently narrow the wave passes through it.

We will consider a particular architecture of the chemical reactor called later on the double-cross reactor, in which the passive regions form a double cross (stripes), as shown in Fig. 7.10a.

---

32 R. FitzHugh, *Biophysics J.*, 1(1961)445; J. Nagumo, S. Arimoto, S. Yoshizawa, *Proc. IRE*, 50(1962)2061.

33 Given the values of parameters for the excitable solution one can see that for small u (Eq. (7.14)) the linear and quadratic terms in u have the positive coefficient at the right-hand side, i.e., those terms lead to increases of u. Only at very large u enters the negative $-\gamma k u^3$ term into play, dumping finally the value of u. As one can see v acts as an inhibitor, since it enters the right-hand side of Eq. (7.14) with the negative coefficient for $\frac{\partial u}{\partial t}$. As seen from Eq. (7.15), the production of the inhibitor v is enhanced by the presence of u, therefore diminishing u through Eq. (7.14). Thus, a stimulation of u up to very large amplitude before any significant damping by the $-\gamma k u^3$ term begins means that Eqs. (7.14) and (7.15) may indeed describe an excitable system.

34 I.N. Motoike, K. Yoshikawa, *Phys. Rev. E*, 59(1999)5354.

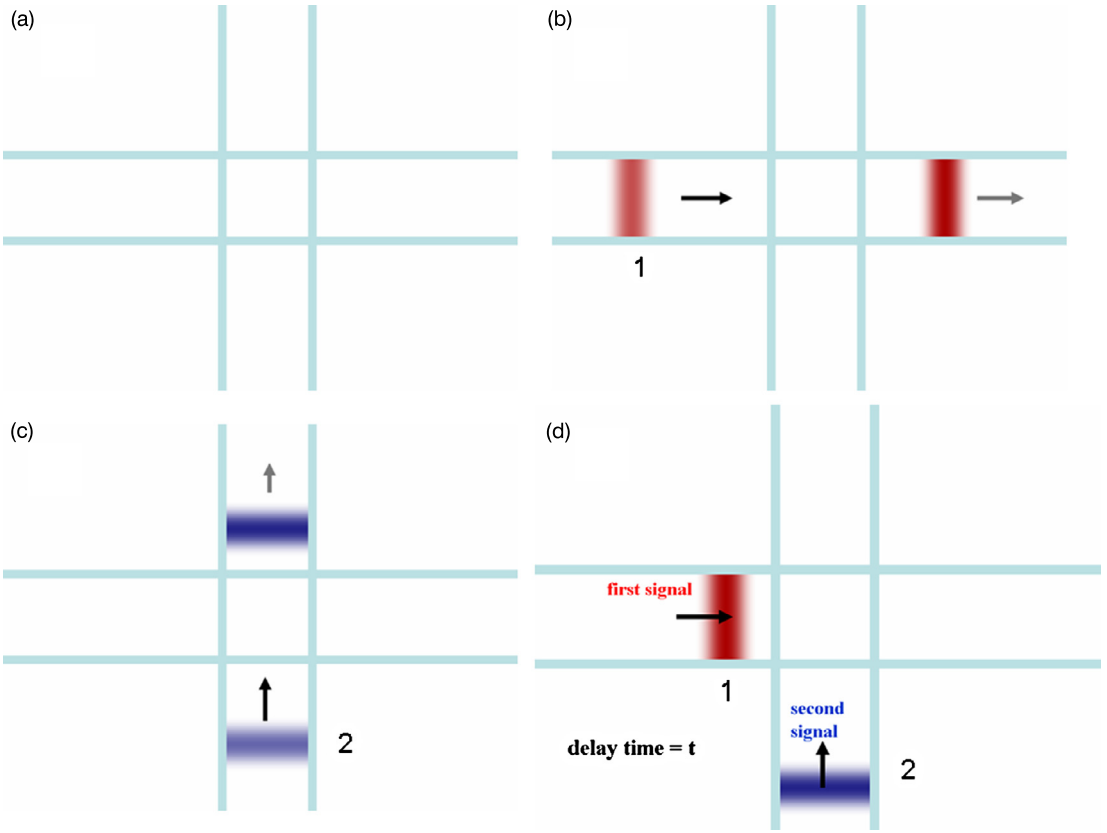


Fig. 7.10. Geometry of the double-cross reactor in the FHN mathematical model. (a) The gray stripes forming the double cross are passive regions, i.e., they correspond to the absence of the inhibitor ($v = 0, \gamma = 0$). This means that no autocatalytic reaction is taking place within the stripes, only diffusion is possible. (b,c) Two chemical waves of u are shown on each figure at two time values (earlier – lighter color, later – darker color), one heading East (b), the other one North (c). It was shown in calculations (Sielewiesiuk and Górecki) that both waves cross the perpendicular stripes, but are able to propagate along a single corridor without any leaking sideways. (d) Two such waves before the collision in the double-cross reactor. Wave 1 (heading East) comes to the reactor's center first, wave 2 (heading North) has a delay time t with respect to wave 1.

The narrow lines forming the double cross are the inhibitor-free stripes of a certain width (passive stripes). The role of the double cross is to provide a partitioning of the reactor into cells, within which the chemical wave can travel freely, while the stripes play the role of barriers for such a motion. In the model described their width is taken sufficiently small to allow the wave to cross perpendicular stripes, but sufficiently large to prohibit sideways leaking of the wave when the wave glides along the corridor between two parallel stripes (the wave front at right angle with the stripes).

Test waves

Fig. 7.10b,c shows a wave front 1 (2) moving with a constant velocity in the Eastwards (Northwards) corridor. The wave is displayed at two instants of time: the lighter color for the earlier time snapshot, the darker color for the later snapshot. It is seen that the waves move straight along their corridors. The reason why the waves are numbered is to distinguish them *after* two such waves collide. Fig. 7.10d shows preparation to such a collision: we see the two incoming wave fronts arriving at the center (the second one with delay t).

Wave collisions

After the collision, the outgoing waves (darker colors) are formed. Fig. 7.11a–g shows how sensitive the output's dependence on the delay time between the two waves is (1 and 2; lighter color means going in, darker color means going out). The output looks complicated, but in fact it can be understood by applying a simple analogy, related to the FHN equations. Imagine the white areas in Fig. 7.11a–g to be covered by grass, the double-cross gray stripes being bare ground (no grass). The variable u may be thought to be a fire activator (like sparks). Since the stripes are sufficiently narrow, the sparks may diffuse through them (especially at the right-angle impact), despite the fact that there is no grass within the stripes that could support their production over there. If the fire front is perpendicular to the stripe (like when going along the corridors), it is unable to set fire behind the stripe (unless it receives some help from another fire front). After the fire front passes, the grass behind begins to grow and then, after a certain refraction time, becomes able to catch fire again. Indeed:

- When the delay ($t > 0$) is smaller than $t = 0.26$ time units (Fig. 7.11a), there are two outgoing chemical waves: one heading Southwest, the other one heading in the opposite direction. Both waves have been labeled 1/2 in order to stress that in this particular case one cannot tell which of the incoming waves turn out to be the outcoming ones. Our analogy gives an explanation: both fire fronts when meeting about the lower-left corner of the central square help one another (across the stripes) to set fire in the corner, first in the outer one, then in the inner one (thus creating the two $\frac{1}{2}$ fire fronts).
- When $0.28 < t < 3.45$ (Fig. 7.11b), wave 1 goes through, while wave 2 disappears. Again, it stands to reason, because fire front 1 comes first and leaves ashes in the center, thus causing wave 2 to die.
- For $3.48 < t < 3.79$ (Fig. 7.11c) wave 1 goes through, while wave 2 deflects, with a rather strange deflection angle of $\frac{3\pi}{4}$. In our analogy this corresponds to the situation when the fire front 1 coming first to the central square just leaves the square and meets at the central

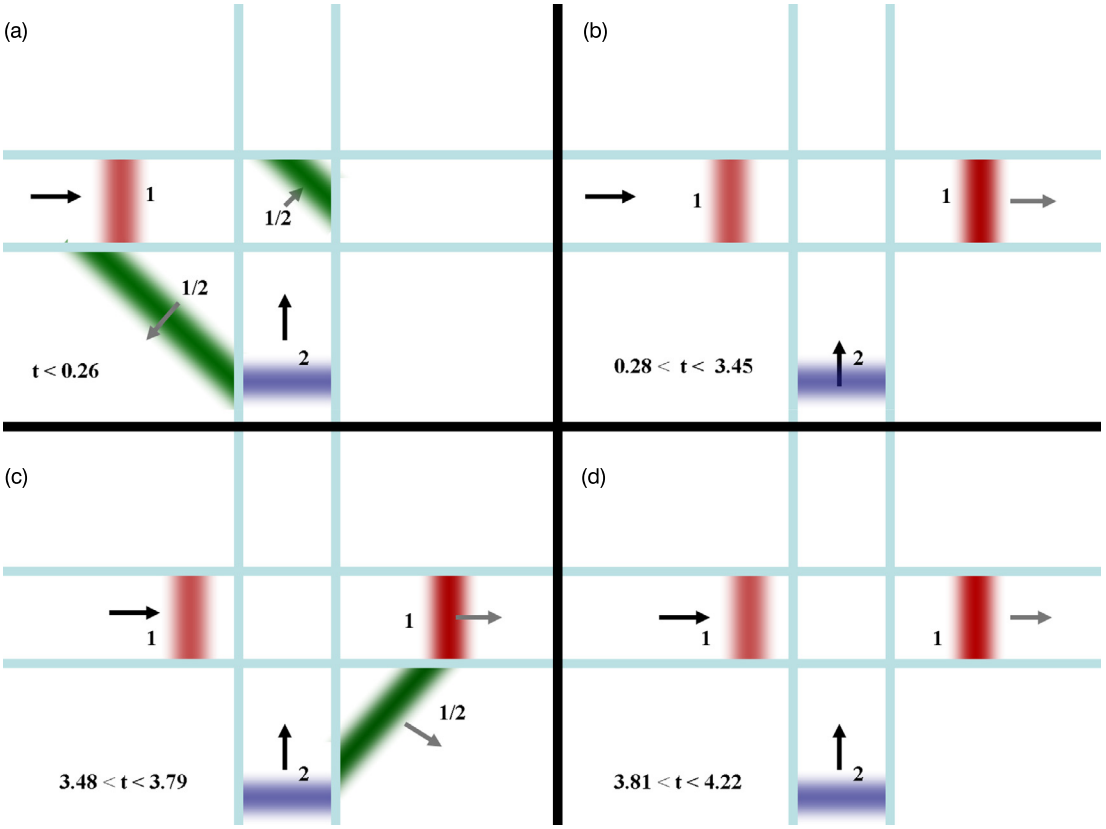


Fig. 7.11. Several solutions of the FHN equations simulating an excitable BZ reaction mixture in the double-cross reactor. The two incoming wave fronts (1 and 2) in two entrance channels (from the East and South directions, lighter color) differ in time of arrival to the central square. The first wave (1) is heading to the center from the West direction, the second one (2) goes North with a delay of $0 < t < 5.5$ time units with respect to wave 1. The outcoming waves (darker color) depend on the delay time t. If an outcoming wave can be identified with the corresponding incoming one, they are denoted by the same label (number). For $t < 0.26$ there are two outgoing chemical waves: one heading Southwest, the other one heading in the opposite direction. Both waves have been labeled by 1/2 in order to stress that in this case one cannot tell which of the incoming waves turns out to be the outcoming one.

lower-right corner the incoming fire front 2. Both are able to set fire, the front (denoted therefore 1/2) heading Southeast.

- For $3.81 < t < 4.22$ (Fig. 7.11d) one has a similar behavior as in the case of $0.28 < t < 3.45$ (Fig. 7.11b). Wave 1 goes through, while wave 2 disappears. This also looks reasonable, since fire front 1 is already in the exit corridor, while fire front 2 does not find the grass in the center, because the refraction time is longer that the delay time.

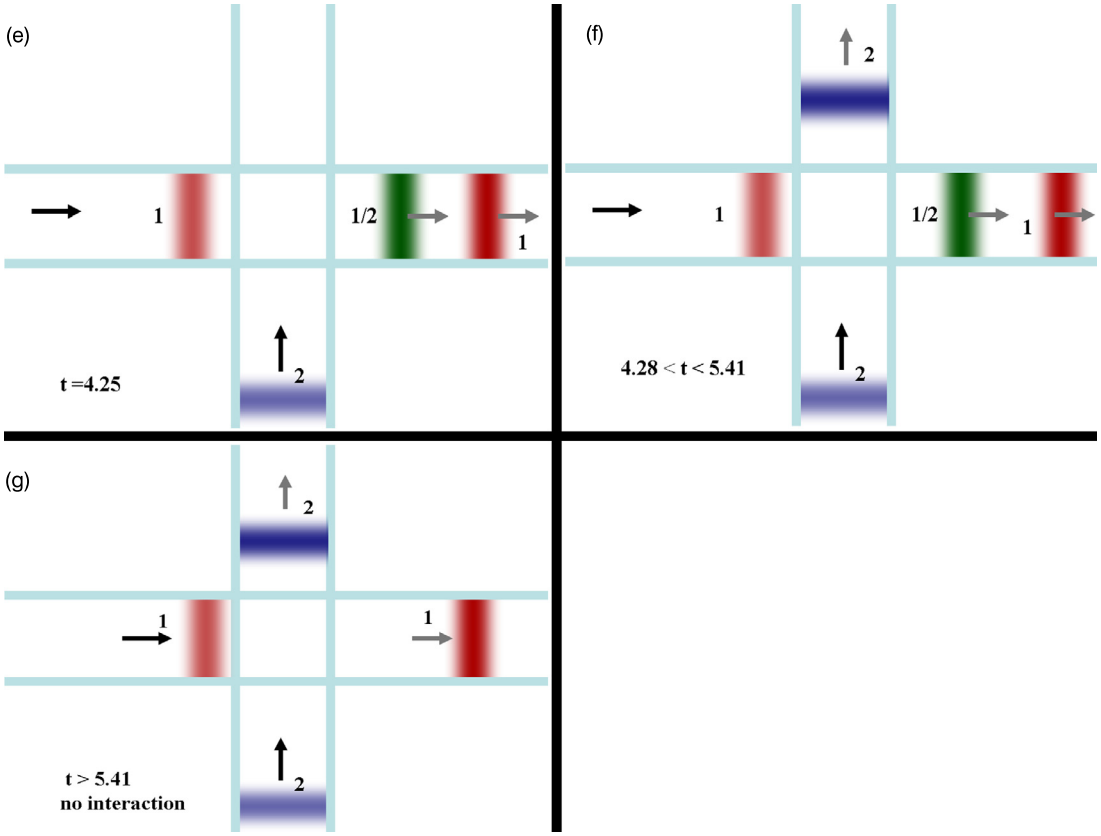


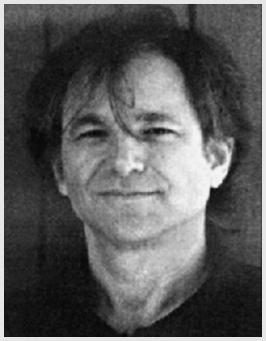
Fig. 7.11. (*continued*)

- For $t = 4.25$ (Fig. 7.11e) wave 1 continues its motion (Eastwards), while wave 2 ... turns right following the first one, thus a double-front wave going Eastwards is formed. Comparing this to the previous case, now the refraction time is smaller than the delay time, the grass is already grown, but there are also some sparks left from fire front 1. This causes a propensity for fire front 2 to turn right.
- For $4.28 < t < 5.41$ (Fig. 7.11f) there is a qualitative change once more: wave 2 splits and therefore in addition to the double wave described above we get a second wave traveling North. In our analogy, since the refraction time is smaller than the delay time the two fire fronts move almost independent in their corridors, except that when fire front 2 passes the center it receives some additional sparks from wave front 1, which already passed the center. This sets fire in the new-grown grass in the horizontal corridor.
- Finally, for $t > 5.41$ (Fig. 7.11g) the delay is so large that the two waves do not interact and pass through unchanged.

The extraordinary versatility of this behavior witnesses about a large potential for such dynamic information processing. By using one parameter (time delay) one is able to ignite the preselected cells in the reactor. Such a chemical wave coming to a cell may trigger a cascade of other processes.

7.3.4 Molecules as computer processors

Computers have changed human civilization. Their speed doubles every year or so, but the expectations are even greater. A possible solution is parallel processing, or making lots of computations at the same time; another is miniaturization. As will be seen in a moment, both these possibilities could be offered by molecular computers, in which the elementary devices would be the individual molecules chemists work with all the time. This stage of technology is not yet achieved. The highly elaborated silicon lithographic technology makes it possible to create electronic devices of a size of the order of 1000 Å. Chemists would be able to go down to the hundreds or even tens of Å. Besides, the new technology would be based on self-organization (supramolecular chemistry) and self-assembling. In 1 cm^3 we could store the information of a billion CD-ROMs. People thought a computer had to have the form of a metallic box. However...



Leonard M. Adleman (born 1945), American mathematician, professor of computer science and of molecular biology at the University of California, Los Angeles. As a young boy he dreamed of becoming a chemist, then a medical doctor. These dreams led him to the discovery described here.

In 1994 mathematician Leonard M. Adleman[35] began his experiments in one of the American genetics labs, while learning biology in the evenings. Once, reading in bed Watson's textbook "*The Molecular Biology of the Gene*," he recognized that the features of the polymerase molecule interacting with the DNA strand described in the textbook perfectly match the features of what is called the Turing machine, or an *abstract representation of a computing device*, elaborated just before the Second World War by Alan Turing.

35 L. Adleman, *Science*, 266(1994)1021.

Alan Mathison Turing (1912–1954), British mathematician, defined a device ("Turing machine") which consists of a read/write head that scans a one-dimensional tape divided into squares, each of which contains a "0" or "1." The behavior of the machine is completely characterized by its current state, the content of the square it is just reading, and a table of instructions.

Turing is known also for decoding during the Second World War further versions of the German Enigma code. This was a continuation of the 1933 breaking the Enigma code by three young Polish mathematicians, Marian Rejewski, Jerzy Różycki, and Henryk Zygalski. About a month before the outbreak of the Second World War (on July 25, 1939, in Pyry) they delivered to British and French Intelligence the deciphering technology, including the constructed computer-like deciphering "bombe." The British further massive development was especially successful. Alan Turing – a mathematical genius – was the key person in this effort.

Therefore, the polymerase and the DNA (and certainly some other molecules) could be used as a computer. If we think about it *now*, the computer in our head is more similar to water than to a box with hard disks, etc. The achievement of Adleman was that he was able to translate a known and important mathematical problem into the language of laboratory recipes, and then using a chemical procedure he was able to solve the mathematical problem.

Fig. 7.12a shows the original problem of Adleman: a graph with 14 airplane flights involving seven cities.

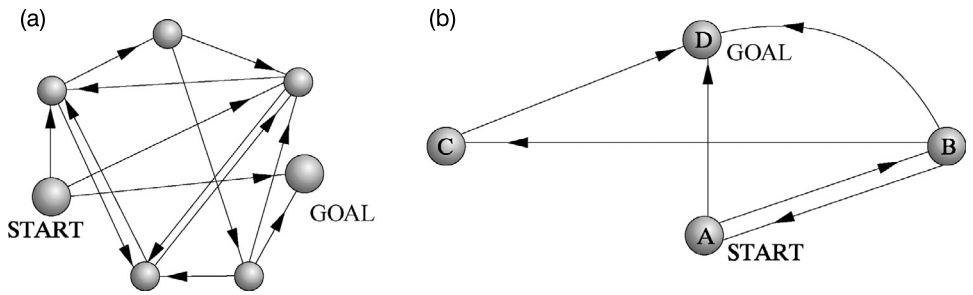
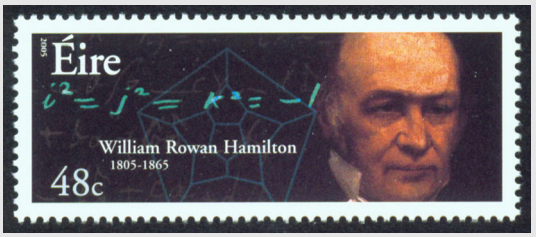


Fig. 7.12. A graph of airplane flights. Is the graph of the Hamilton type? This was a mathematical question for the molecular computer. (a) The graph from Adleman's experiment. (b) A simplified graph described in this book.

The task is called the *traveling salesman* problem, notorious in mathematics as extremely difficult.[36] The salesman begins his journey from the city START and wants to go to the city GOAL, visiting every other city precisely once. This is feasible only for some flight patterns. Those graphs for which it is feasible are called the *Hamilton graphs*. When the number of cities is small, such a problem may be quite effectively solved by the computer in our head. For seven cities it takes on average 56 s, as stated by Adleman; for larger numbers we need a desk computer, but *for a hundred* cities *all the computers of the world would be unable to provide the answer*. However, a molecular computer would have the answer within a second.

William Rowan Hamilton (1805–1865) was an Astronomer Royal in Ireland. At the age of 17 he found an error in the famous "Celestial Mechanics" by Laplace. This drew the attention of scientists and was the beginning of Hamilton's scientific career. In the present book his name occurs many times (e.g. Hamiltonian).



How does a molecular computer work?

Let us recall two important examples of complementary synthons: guanine and cytosine (GC) and adenine and thymine (AT, cf. Chapter 5, p. 424).

Let us repeat Adleman's algorithm for a much simpler graph (Fig. 7.12b). What Adleman did was the following.

1. He assigned for every city some particular section of DNA (sequence), an oligomer composed of eight nucleotides:

City A	A	C	T	T	G	C	A	G
City B	T	C	G	G	A	C	T	G
City C	G	G	C	T	A	T	G	T
City D	C	C	G	A	G	C	A	A

2. Then to each existing flight X → Y, another eight-base DNA sequence was assigned, composed of the second half of the sequence of X and the first half of the sequence of Y:

[36] The problem is what is called NP-hard (NP from *nonpolynomial*), because the difficulties increase faster than any polynomial with the size of the problem.

Flight A→B	G	C	A	G	T	C	G	G
Flight A→D	G	C	A	G	C	C	G	A
Flight B→C	A	C	T	G	G	G	C	T
Flight B→D	A	C	T	G	C	C	G	A
Flight B→A	A	C	T	G	A	C	T	T
Flight C→D	A	T	G	T	C	C	G	A

3. Then, the synthesis of the DNA sequences[37] of the flights and the DNA sequences complementary to the cities was ordered:

co-City A	T	G	A	A	C	G	T	C
co-City B	A	G	C	C	T	G	A	C
co-City C	C	C	G	A	T	A	C	A
co-City D	G	G	C	T	C	G	T	T

4. To run the molecular computer all these substances are to be mixed together, dissolved in water, and supplemented with a bit of salt and an enzyme called ligase.[38]

How to read the solution

What happened in the test tube? First of all matching and pairing of the corresponding synthons took place. For example, the DNA strand that codes the AB flight (i.e., GCAGTCGG) matched in the solution with the complementary synthon of city B (i.e., the co-city AGCCTGAC) and, because of the molecular recognition mechanism, made a strong intermolecular complex (cf. Chapter 5, p. 424):

flights	G	C	A	G	T	C	G	G				
					⋮	⋮	⋮	⋮				
co-cities					A	G	C	C	T	G	A	C

where the upper part is flights, and the lower part is co-cities. Note that the flights are the only feasible ones, because only feasible flights' DNA sequences were synthesized. The role of a co-city's DNA is to provide the information that there is the possibility to land and take-off in this particular city.

In the example just given, the complex will also find the synthon that corresponds to flight B → C, i.e., ACTGGGCT, and we obtain a more extended strand

[37] Nowadays it is a matter of commercial activity.

[38] To be as effective as Nature, we want to have conditions similar to those in living cells.

```
GC   A   G   T   C   G   G  |A   C   T   G   G   G   C   T
             ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮   ⋮
             A   G   C   C   T   G   A   C
```

In this way from the upper part[39] of the intermolecular complexes we can read a particular itinerary. The ligase was needed, because this enzyme binds the loose ends of the DNA strands (thus removing the perpendicular separators shown above). Therefore,

> every possible itinerary is represented in solution by a DNA oligomer. If the graph were Hamiltonian, then there would be in the solution the DNA molecule that encodes the right itinerary. In our example this molecule corresponds to itinerary A → B → C → D and is composed of 24 nucleotides: GCAGTCGGACTGGGCTATGTCCGA.

Eliminating wrong trajectories

Practically, independent of how large N is, after a second the solution to the traveling salesman problem is ready. The only problem now is to be able to read the solution. This will currently take much more than a second, but in principle only depends linearly on the number of cities.

To get the solution we use three techniques: polymerase chain reaction (PCR), electrophoresis, and separation through affinity. The machinery behind all this is supramolecular chemistry, with the recognition of synthons and cosynthons (called in biochemistry hybridization[40]).

The itineraries coded by hybridization are mostly wrong. *One* of the reasons is that they do not start from the START (A) and do not end up at GOAL (D). Using the PCR technique[41] it is possible to increase the concentration of only those itineraries which start from START and end at GOAL to such an extent that concentrations of all the other molecules may be treated as marginal.

Still there are a lot of wrong itineraries. First of all there are a lot of itineraries that are too long or too short. This problem may be fixed by electrophoresis,[42] which allows the separation of

---

39 From the lower part as well.

40 A bit misleadingly, if we think of the hybridization of atomic orbitals (p. V1-562).

41 The PCR technique is able to copy a chosen DNA sequence and to grow its population even from a single molecule to a high concentration by using the repeated action of the enzyme DNA polymerase.
The reaction was invented by Kary B. Mullis (born 1944), an American technical chemist in an industrial company. In 1983 Mullis was driving to his favorite California surfing area, when the idea of a DNA copying molecular machine struck him suddenly. He stopped the car and made a note of the reaction. In 1993 Kary Mullis received the Nobel Prize in chemistry "*for his invention of the polymerase chain reaction (PCR) method*."

42 Electrophoresis is able to physically separate DNA sequences according to their length. It is based on the electrolysis in a gel medium. Since DNA is an anion, it will travel through the gel to the anode. The shorter the

DNA strands of a given length, in our case the 24-city itineraries. In this way we have itineraries starting from START and ending at GOAL and having 24 cities. They can be copied again by PCR to increase their concentration for further operations.

Now we have to eliminate wrong itineraries from these 24-long sequences: those which repeat some transit cities and leave others unvisited. This is done by the affinity separation method.[43] First, the cosynthon for the first transit city (in our case: C) on the list of transit cities (in our case: C and D) is prepared and attached to the surface of iron balls. The iron balls are then added to the solution and after allowing a second to bind to those itineraries that contain the city, they are picked out using a magnet. The balls are then placed in another test tube, the attached "itineraries" are released from the surface of the iron balls, and the empty iron balls are separated. Then the oligomers are multiplied by PCR. Thus, we have in a test tube the "itineraries" that begin and end correctly, have the correct number of 24 nucleotides and certainly go through the first transit city (C) on our list of transit cities.

The process is repeated for the second, third, etc., transit cities.

> If, in the last test tube, any molecular "itinerary" swims, it has to be the Hamiltonian-like and is identified by the described procedure. The answer to the salesman problem is, therefore, positive. Otherwise the answer is negative.

Thus, a mathematical problem was solved using a kind of molecular biocomputer. From the information processing point of view, this was possible because parallel processing was under way – a lot of DNA oligomers interacted with themselves at the same time. The number of such molecular processors is of the order of 10^{23}. This number is so huge, that such a biocomputer is able to check (virtually) all possibilities and to find the solution.

7.3.5 The mission of chemistry

There is an impression that chemistry's role in biology is only just a possibility, a kind of substitute, a pretext, no more than a material carrier of some mission of the whole organism. Contemporary textbooks of biochemistry do not say much about chemistry, they talk about

molecule, the longer distance it will reach. The DNA molecules of a given length can then be picked out by cutting the particular piece of gel and then they can be multiplied by PCR.

[43] Affinity separation makes it possible to separate particular sequences from a mixture of DNA sequences. This is achieved by providing its cosynthon attached to iron spheres. The particular sequence we are looking for binds to the surface of the iron ball, which may afterwards be separated from the solution using a magnet.

molecular functions to perform, in a sense about metachemistry. A particular molecule seems not to be so important. What counts is its function. A good example are enzymes. One *type* of enzyme may perform the same or similar functions in many different organisms (from fungi to man). The function is the same, but the composition of the enzyme changes from species to species: two species may differ by as much as 70% in amino acid sequence. However, those amino acids that are crucial for enzyme function are preserved in all species.

Our world is full of information, is operating because of information, it exists because of information processing: from impregnation till writing poems, from cellular phones till the human individual or collective decisions. Information seems to be the central idea, which in our world is based on chemistry, but chemistry seems ...to ignore it. It is time to introduce information to future chemistry. A great adventure! This unprecedented chemical task would be collecting, transporting, changing, dispatching, and transferring of information.

Chemistry, as we develop it, is far from such a masterpiece. What we are doing currently might be compared to chemical research by a Martian with a beautifully edited "Auguries of Innocence" by William Blake. The little green guy would perform a chemical analysis of the paper (he would probably even make a whole branch of science of that), examine the chemical composition of the printing dye; with other Martian professors he would make some crazy hypotheses on the possible source of the leather cover, list the 26 different black signs as well as their perpendicular and horizontal clusters, analyze their frequencies, etc. The Martian science would flourish! He would, however, be very far from the *information* the book contains, including the boring matrix of black marks.

To see a world in a grain of sand
And heaven in a wild flower
Hold infinity in the palm of your hand
And eternity in an hour

Most importantly, he could not even imagine his heart[44] beating any faster after reading this passage, because of thousands of associations *he could never have had*... We are close to what the Martian professor would do. We have wonderful matter in our hands from which we could make chemical poems, but so far we have been able to do only very little.

Molecules could play much more demanding roles than those we have foreseen for them: *they can process information*.

[44] If any...

Summary

Chemistry has attained such a stage that soon a new quality can be achieved:

- chemistry entered the second half of the 20th century with detailed knowledge of the main building blocks of molecular structures: atoms, chemical bonds, bond angles, and intermolecular interactions;
- the accumulated knowledge now serves to build more and more complex molecular architectures;
- in these architectures we may use *chemical bonds* (with energy of the order of 50–150 kcal/mol) to build molecules, as well as *intermolecular interactions* (with energy of about 1–20 kcal/mol) to construct supramolecular structures from them;
- in supramolecular chemistry we operate with synthons, i.e., some special systems of functional groups that fit together perfectly, giving rise to molecular recognition;
- the interaction leads to a molecular complex that facilitates further evolution of the system, either by a chemical reaction going on selectively at such a configuration of the molecules or by further self-organization due to next-step molecular recognition of the newly formed synthons;
- this may result in the formation of complex systems of multilevel architecture, each level characterized by its own stability;
- the self-organization may take place with significant interaction nonadditivity effects ("nonlinearity" in mathematical terms) that may lead to cooperation in the formation of the multilevel structure;
- the self-organized structures may interact with other such structures (chemical reactions or association);
- in particular they may create autocatalytic cycles, which represent chemical feedback;
- such cycles may couple in a higher-order cycle forming hypercycles;
- a dynamic system with hypercycles, when perturbed by an external stimulus, reacts in a complex and nonlinear way;
- one of the possibilities in nonequilibrium conditions are the limit cycles, which lead to dissipative structures, which may exhibit periodicity (in space and time) as well as chaotic behavior;
- some dynamic systems may represent molecular libraries with the proportions of species strongly depending on external conditions (cf. the immune system);
- molecules may act (e.g., transfer photons, electrons, protons, or ions, induce a conformational change, etc.) thus performing a function;
- several functions may cooperate exhibiting a space/time organization of the individual functions;
- some molecules may serve for effective information processing;
- information processing seems to represent the ultimate goal of future chemistry;
- molecules can be "taught" by adjusting its shape to the requirements of a "teacher molecule" and the temperature-dependent amount of knowledge acquired can be measured in bits;
- chemical waves can carry information very effectively, changing their behavior that depends strongly on the spatial and time restrictions imposed;
- molecules may serve in the future as massively parallel computer processors in a way that is fundamentally distinct from performance of contemporary computers.

Main concepts, new terms

activator (p. 563)
attractors (p. 543)
autocatalysis (p. 547)
Belousov–Zhabotinsky reaction (p. 563)
bifurcation (p. 546)
Brusselator (p. 549)
center of marginal stability (p. 552)
chaos (p. 546)
chemical waves (p. 563)
combinatorial chemistry (p. 541)
complex systems (p. 537)
cooperativity (p. 540)
dissipative structures (p. 553)
DNA computing (p. 568)
DNA hybridization (p. 572)
excitable state (p. 563)
feedback (p. 547)
FitzHugh–Nagumo model (p. 563)
fixed point (p. 543)
fluctuation (p. 550)
focus (stable and unstable, p. 552)
Hamilton graph (p. 569)
hypercycles (p. 555)
information (p. 558)
limit cycle (p. 543)
logistic equation (p. 546)
molecular evolution (p. 555)
molecular lesson (p. 559)
molecular libraries (p. 541)
nodes (stable and unstable, p. 552)
nonadditivity (p. 540)
NP-hard problem (p. 570)
PCR (p. 572)
polymerase (p. 572)
reaction center (p. 552)
repellers (p. 552)
saddle point of reaction (p. 552)
self-organization (p. 538)
separation by affinity (p. 572)
Shannon information entropy (p. 558)
stable focus (p. 552)
stable node (p. 552)
stable stellar node (p. 552)
stellar nodes (stable and unstable, p. 552)
student molecule (p. 559)
teaching molecule (p. 559)
traveling salesman problem (p. 570)
Turing machine (p. 568)
unstable focus (p. 552)
unstable node (p. 552)
unstable stellar node (p. 552)

From the research front

To say that organic chemists are able to synthesize almost any molecule one may think of is certainly an exaggeration, but the statement seems sometimes to be very close to reality. Chemists were able to synthesize the five-olympic-ring molecule, the three interlocked Borromean rings, the football made of carbon atoms, the "cuban" (a hydrocarbon cube), the "basketan" (in the form of an apple basket), a molecule in the form of the Möbius band, etc. Now we may ask why the enormous synthetic effort was undertaken and what these molecules were synthesized for. Well, the answer seems to be that contemporary chemists are fascinated by the art of making complex and yet perfect and beautiful molecular objects. The main goal apparently was to demonstrate the mastership of modern chemistry. However, high symmetry does not necessarily mean particular usefulness. No doubt the future synthetic targets will be identified by the careful planning of molecular *functions*, rather than molecular beauty.

Ad futurum

We may expect that more and more often, chemical research will focus on molecular function, and (later) on the space/time cooperation of the functions. Research projects will be formulated in a way that will highlight the role of the molecular function, and will consist of several (interrelated) steps:

- first, the technical goal will be defined;
- the molecular functions will be identified which will make this goal achievable;
- theoreticians will design and test with computers ("*in silico*") the molecules which will exhibit the above functions;
- synthetic chemists will synthesize the molecules designed;
- physicochemists will check whether the molecules can perform the expected molecular functions;
- finally, the material will be checked against the technical goal.

We will be able to produce "intelligent" materials which will respond to external conditions in a previously designed, complex, yet (we hope) predictable way. The materials that will be created this way will not resemble the materials of today, which mostly carry out one, often primitive, function. The drugs of today are usually quite simple molecules, which enter the extremely complex system of our body. The drugs of tomorrow will involve much larger molecules (like proteins). Will we be clever enough to avoid unpredictable interactions with our body? What in principle do we want to achieve?

What will the motivation of our work be? Will we take into account the psychological needs of the human being, equilibrium of their minds, his/her longing for the Truth, the Beauty and the Good?

What will the future of the human family be, which was able in the past to create such wonderful music, the Chartres cathedral, breathtaking paintings, moving poetry, and abstract mathematics, and proudly landed on other celestial bodies[45]? In the past nothing could stop their curiosity and ingeniousness; they were able to resist the harshest conditions on their planet. Humans have reached nowadays the technical level that probably will ensure avoiding the next glaciation,[46] maybe allow a *small* asteroid to be pushed off-target by nuclear weapons, if it were aimed dangerously at the Earth, or erase in a nuclear war most of its own population together with the wonders of our civilization.

What is the goal of these beings and what will be the final limit of their existence? What they are aiming at? Do we want to know the smell of fresh bread, to be charmed by the Chartres cathedral with all it is for, to use our knowledge to cherish the friendship of the human family, or will it be sufficient to pack newborns into a personal container and make computers inject substances that will stimulate their neural system and make them as happy as if they were in the Seventh Heaven[47]?

Which of the goals, as chemists, do we want to participate in?

Additional literature

M. Eigen, P. Schuster, "The Hypercycle. A Principle of Natural Organization," Springer Verlag, Berlin, 1979.

45 Basically thanks to the last twelve millenia global warming (glacier melting).

46 Well, it is expected within the next 500 years.

47 This shows that we cannot rely on voting alone, it may lead us to abyss.

An excellent, comprehensible book, written by the leading specialists in the domain of molecular evolution.

I. Prigogine, "From Being to Becoming. Time and Complexity in Physical Sciences," Freeman, San Francisco, 1980.

A book written by the most prominent specialist in the field.

A. Babloyantz, "Molecules, Dynamics and Life," Wiley, New York, 1987.

The author describes the scientific achievements of Prigogine and his group, which she participated in. An excellent, competent book, the most comprehensible among the first three recommended books.

J.-M. Lehn, "Supramolecular chemistry: Concept and Perspectives," VCH, 1995.

A vision of supramolecular chemistry given by one of its founders.

K. Szaciłowski, "Infochemistry: Information Processing at the Nanoscale," Wiley, 2012.

A rare link to laboratory experiments in this domain.

Questions

1. An oscillatory solution of differential equations:
   a. represents an attractor.
   b. has been discovered in the 20th century.
   c. when met in chemistry means concentration oscillations.
   d. is a limit cycle.
2. A dissipative structure:
   a. may appear in thermodynamic equilibrium.
   b. represents a space- and/or time-dependent distribution of concentrations of chemical substances.
   c. depends on the matter and energy fluxes in the system.
   d. may appear when the system is sufficiently far from a thermodynamic equilibrium.
3. A molecular library:
   a. cannot exist in thermodynamic equilibrium.
   b. in the case of a mixture of A and B substances, contains molecular complexes A_nB_m with various m, n.
   c. has the ability to shift the equilibrium under the influence of some other substances.
   d. means a complete set of books on molecular physics.
4. A molecular self-organization:
   a. means a spontaneous formation of molecular complexes resulting in the supramolecular structures exhibiting a short-range and/or a long-range order.
   b. is possible only in nonequilibrium.

c. is a result of the molecular recognition through spatial and electrical matching.
d. is impossible without a chemist's planned action.

5. In the iterative solution $x_{n+1} = K x_n(1 - x_n)$ of the logistic equation:
   a. one obtains a fixed point at any K.
   b. any attempt of increasing K leads to a bifurcation.
   c. there is a range of K that corresponds to chaotic behavior of the solution x.
   d. at a sufficiently small K the population vanishes.
6. In the Brusselator without diffusion (x and y stand for fluctuations of the substances X and Y):
   a. a stable node corresponds to an exponential vanishing of x and y.
   b. a stable focus means vanishing oscillations of x and y.
   c. at least one of the reactions should have an autocatalytic character.
   d. X is a catalyst but Y is not.
7. In an isolated system:
   a. the entropy does not change.
   b. after a sufficiently long time the gradient of the temperature must attain zero.
   c. the concentration gradients are zero.
   d. one cannot observe dissipative structures.
8. Information entropy:
   a. is equal to $-\sum_i p_i \log_2 p_i$.
   b. represents a measure of our ignorance about a coming message.
   c. is equal to the mean number of questions necessary to define the probability distribution.
   d. attains the minimum for all p_i being equal.
9. An event has only four possible outputs with *a priori* probabilities $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$. Reliable information arrives, telling us that in fact the probabilities are different; $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, $p_4 = \frac{1}{8}$. This information had I_1 bits, and I_1 is equal to:
   a. 1 bit.
   b. 0.5 bit.
   c. 2 bits.
   d. 0.25 bit.
10. The situation corresponds to question 9, but a second piece of reliable information arrives, telling us that the situation changed once more and now $p_1 = \frac{1}{2}$, $p_2 = 0$, $p_3 = 0$, $p_4 = \frac{1}{2}$. The second piece of information had I_2 bits. We pay for information in proportion to its quantity. Therefore, for the second piece of information we have to pay:
    a. the same as for the first piece of information.
    b. twice as much as for the first piece of information.
    c. half of the price for the first piece of information.
    d. three time more than for the first piece of information.

Answers

1a,c,d, 2b,c,d, 3b,c, 4a,c, 5c,d, 6a,b,c, 7b, 8a,b, 9d, 10d

APPENDIX A

Dirac Notation for Integrals

The integrals over the spatial *and* spin coordinates (ϕ are the spin orbitals, φ the orbitals) in the *Dirac notation* will be denoted with angle brackets $\langle\rangle$ ($\hat{h}$ denotes a one-electron operator and r_{12} denotes the distance between electrons 1 and 2).

For one-electron integrals we write

$$\left\langle i|\hat{h}|j\right\rangle \equiv \sum_{\sigma_1}\int dx_1 dy_1 dz_1 \phi_i^*(1)\hat{h}\phi_j(1), \tag{A.1}$$

and for two-electron integrals we write

$$\langle ij|kl\rangle \equiv \sum_{\sigma_1}\sum_{\sigma_2}\int dx_1 dy_1 dz_1 \int dx_2 dy_2 dz_2 {}_1\phi_i^*(1)\phi_j^*(2)\frac{1}{r_{12}}\phi_k(1)\phi_l(2). \tag{A.2}$$

The integrals over the spatial coordinates (only) will be denoted by parentheses (). For one-electron integrals we write

$$\left(i|\hat{h}|j\right) \equiv \int dx_1 dy_1 dz_1 \varphi_i^*(1)\hat{h}\,(1)\,\varphi_j(1), \tag{A.3}$$

and for two-electron integrals we write

$$(ij|kl) \equiv \int dx_1 dy_1 dz_1 \int dx_2 dy_2 dz_2 \varphi_i^*(1)\varphi_j^*(2)\frac{1}{r_{12}}\varphi_k(1)\varphi_l(2). \tag{A.4}$$

This is called Dirac notation (for the integrals).

APPENDIX B

Hartree–Fock (or Molecular Orbitals) Method

The method belongs to a class of variational methods (p. V1-265), i.e., the mean value $\varepsilon = \frac{\langle \psi | \hat{H} \psi \rangle}{\langle \psi | \psi \rangle}$ of the Hamiltonian $\hat{H}$ is to be minimized, producing the Hartree–Fock energy $\varepsilon_{\min}$. The variational wave function ψ *is restricted (approximation*[1]*) to a single Slater determinant,*

$$\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}, \tag{B.1}$$

where the *molecular spin orbital* $\phi_i(j) \equiv \phi_i(\mathbf{r}_j, \sigma_j)$ represents a function of electron j (i.e., of its spatial x_j, y_j, z_j and spin σ_j coordinates $(x_j, y_j, z_j, \sigma_j) \equiv (\mathbf{r}_j, \sigma_j)$). The spin orbitals $\{\phi_i\}$ are assumed to form an orthonormal (i.e., normalized and mutually orthogonal) set. From this follows that:

- two (or more) electrons cannot be described by the same ϕ_i (identical columns make the determinant equal to zero);
- two electrons with the same spin coordinate cannot touch (identical columns make the determinant equal to zero) – the Pauli exclusion principle satisfied ("exchange hole" or "Fermi hole" present);
- two electrons with different spin coordinate can touch (Coulombic hole absent).

Molecular orbitals

The spin orbitals ϕ_i are constructed from the *molecular orbitals* and the α and β spin functions, i.e.,

$$\phi_i(\mathbf{r}, \sigma) = \varphi_i(\mathbf{r})\alpha(\sigma)$$

or

$$\phi_i(\mathbf{r}, \sigma) = \varphi_i(\mathbf{r})\beta(\sigma),$$

[1] The Slater determinant is an antisymmetric function, but an antisymmetric function does not necessarily need to take the shape of a Slater determinant.

where $\varphi_i(\mathbf{r})$ denotes the i-th *orbital* and the spin basis functions are defined as

$$\alpha(\sigma) = \begin{cases} 1 & \text{for } \sigma = \frac{1}{2}, \\ 0 & \text{for } \sigma = -\frac{1}{2} \end{cases} \quad \text{and} \quad \beta(\sigma) = \begin{cases} 0 & \text{for } \sigma = \frac{1}{2}, \\ 1 & \text{for } \sigma = -\frac{1}{2}. \end{cases}$$

Restricted Hartree–Fock (RHF) Method

When the number of electrons is even, the spin orbitals are usually formed out of orbitals by multiplication of each orbital by the spin functions α or β, i.e.,

$$\phi_{2i-1}(\mathbf{r}, \sigma) = \varphi_i(\mathbf{r})\alpha(\sigma), \tag{B.2}$$

$$\phi_{2i}(\mathbf{r}, \sigma) = \varphi_i(\mathbf{r})\beta(\sigma), \tag{B.3}$$

$$i = 1, 2, \ldots \frac{N}{2}, \tag{B.4}$$

where – as can be clearly seen – there are twice as few occupied orbitals φ as occupied spin orbitals ϕ (*occupation means that a given spin orbital appears in the Slater determinant*). Thus we introduce an artificial *restriction* for spin orbitals. This is why the method is called the restricted Hartree–Fock (RHF). Nothing forces us to do so (not introducing the restriction means the unrestricted Hartree–Fock [UHF] method), the criterion is simplicity. There are as many spin orbitals as electrons, and there can be a maximum of two electrons per RHF orbital.

The variational method represents a way to obtain the best orbitals φ_i, i.e., those producing ψ that ensures the lowest mean value of energy $\varepsilon = \frac{\langle \psi | \hat{H} \psi \rangle}{\langle \psi | \psi \rangle}$. In the case of the RHF method for such an optimal $\psi \equiv \psi_{RHF}$ one obtains the Hartree–Fock energy $E_{RHF} = \frac{\langle \psi_{RHF} | \hat{H} \psi_{RHF} \rangle}{\langle \psi_{RHF} | \psi_{RHF} \rangle} \equiv \varepsilon_{\min}$. One might find such optimal φ's by a trial-and-error method when minimizing ε, but there is a much more interesting direct way: variational calculus gives us an equation satisfied by the optimal φ's. This is the Fock equation,

$$\hat{\mathcal{F}}(1)\varphi_i(1) = \varepsilon_i \varphi_i(1), \tag{B.5}$$

$i = 1, 2, \ldots$ numbering its multiple solutions and ε_i being the orbital energy (in ascending order by convention).

For the double occupancy the Fock operator $\hat{\mathcal{F}}$ is defined as

$$\hat{\mathcal{F}}(1) = \hat{h}(1) + 2\hat{\mathcal{J}}(1) - \hat{\mathcal{K}}(1), \tag{B.6}$$

where the first term

$$\hat{h}(1) = -\frac{1}{2}\Delta_1 - \sum_{a=1}^{M} \frac{Z_a}{r_{a1}} \tag{B.7}$$

is the one-electron operator (in atomic units) of the kinetic energy of the electron plus the operator of the nucleus–electron attraction (there are M nuclei), the next two terms, i.e., Coulombic[2] $\hat{\mathcal{J}}$ and exchange $\hat{\mathcal{K}}$ operators, are connected with the potential energy of the interaction of electron 1 with all electrons in the system, and they are defined via the action on any function χ of the spatial position of electron 1, i.e.,

$$\hat{\mathcal{J}}(1)\chi(1) \equiv \sum_{i}^{\mathrm{MO}} \int \mathrm{d}V_2 \frac{\varphi_i^*(2)\varphi_i(2)}{r_{12}} \chi(1), \tag{B.8}$$

$$\hat{\mathcal{K}}(1)\chi(1) \equiv \sum_{i}^{\mathrm{MO}} \int \mathrm{d}V_2 \frac{\varphi_i^*(2)\chi(2)}{r_{12}} \varphi_i(1), \tag{B.9}$$

where the MO limit means summation over the orbitals $i = 1, ..., N/2$, the integration is now exclusively over the spatial coordinates[3] of electron 2.

Although Eq. (B.5) has many solutions, to build the Slater determinant (B.1) one needs only $\frac{N}{2}$ orbitals φ_i with the lowest ε_i. They are known as the occupied orbitals; the most important of them is the Highest Occupied Molecular Orbital (HOMO). The next orbitals (with the ascending order of ε_i) are known as the virtual orbitals, the Lowest Unoccupied Molecular Orbital is known as LUMO. The important message for a chemist is that HOMO and LUMO decide about the main features of a molecule's chemistry.

At the end there are two problems left.

The first one pertains to Eq. (B.5), which looks as "a Schrödinger equation for electron 1," however with unknown ...all mathematical objects involved: $\hat{\mathcal{F}}(1)$, $\varphi_i(1)$, and ε_i, because of unknown φ's... Fortunately usually this difficulty is easy to overcome, by using an iterative method (with some starting φ's).

The second problem is to provide an expression for the Hartree–Fock total energy. The corresponding formula reads as

$$E_{RHF} = 2\sum_{i}^{\mathrm{MO}} h_{ii} + \sum_{i,j}^{\mathrm{MO}} [2\mathcal{J}_{ij} - \mathcal{K}_{ij}] + V_{nn}, \tag{B.10}$$

where the integrals (in the Dirac notation, see p. 581) $h_{ii} = (i|\hat{h}|i)$, $\mathcal{J}_{ij} \equiv (ij|ij)$, $\mathcal{K}_{ij} \equiv (ij|ji)$, and V_{nn} is the internuclear repulsion

[2] Here factor 2 results from the *double* occupancy of the orbitals.

[3] The summation over the spin coordinates has already been performed when deriving the equation for the mean value of the Hamiltonian.

$$V_{nn} = \sum_{a<b} \frac{Z_a Z_b}{R_{ab}}. \tag{B.11}$$

To calculate E_{RHF} and ψ_{RHF} one needs to know the occupied orbitals only.

APPENDIX C

Second Quantization

When we work with a basis set composed of the Slater determinants, we are usually confronted with a large number of the matrix elements involving one- and two-electron operators. The Slater–Condon rules (Appendix V1-N) are doing the job to express these matrix elements by the one-electron and two-electron integrals. However, we may introduce an even easier tool, called second quantization, which is equivalent to the Slater–Condon rules.

Vacuum state

In the second quantization formalism we introduce for the system under study a reference state, which is a Slater determinant (usually the Hartree–Fock wave function) composed of N orthonormal spin orbitals, where N is the number of electrons. This function will be denoted in short by Φ_0 or as a particular case of $\Phi^N\,(n_1, n_2, ..., n_\infty)$, where the last notation means that we have to do with a normalized N-electron Slater determinant, and in the parentheses we give the occupancy list ($n_i = 0, 1$) for the infinite number of the orthonormal spin orbitals considered in the basis set and listed one by one in the parentheses. This simply means that some spin orbitals are present in the determinant (they have $n_i = 1$), while some other are absent[1] ($n_i = 0$). Hence, $\sum_i n_i = N$. The reference state is often called the *vacuum state*. The subscript 0 in Φ_0 means that we consider this single-determinant as an approximation to the *ground state wave function*. Besides the reference state some normalized Slater determinants of the excited states will be considered, with other occupancies, *including those corresponding to the number of electrons that differs from N*.

[1] For example, the symbol Φ^2 (001000100000...) means a normalized Slater determinant of dimension 2, containing the spin orbitals 3 and 7. The symbol Φ^2 (001000...) makes no sense, because the number of "ones" has to be equal to 2, etc.

Creation and annihilation of electrons

Let us make a quite strange move and consider operators that change the number of electrons in the system. To this end let us define the creation operator[2] $\hat{k}^\dagger$ of the electron going to occupy the spin orbital k and the annihilation operator $\hat{k}$ of an electron disappearing from the spin orbital k:

CREATION AND ANNIHILATION OPERATORS

$$\hat{k}^\dagger \Phi^N (...n_k...) = \theta_k (1 - n_k) \Phi^{N+1} (...1_k, ...),$$
$$\hat{k} \Phi^N (...n_k...) = \theta_k n_k \Phi^{N-1} (...0_k, ...),$$

where

$$\theta_k = (-1)^{\Sigma_{j<k} n_j}.$$

The symbol 1_k means that the spin orbital k is present in the Slater determinant, while 0_k means that this spin orbital is empty, i.e., is not present in the Slater determinant. The factors $(1 - n_k)$ and n_k ensure an important property of these operators, namely, that

any attempt of *creation* of the electron on an already *occupied spin orbital* gives zero; similarly, any attempt of *annihilation of an empty spin orbital* also gives zero.

It can be easily shown[3] that (as the symbol suggests) $\hat{k}^\dagger$ is simply the adjoint operator with respect to $\hat{k}$.

The above operators have the following properties that make them equivalent to the Slater–Condon rules:

[2] The domain of the operators represents the space spanned by the Slater determinants built of spin orbitals. Richard Feynman in one of his books says jokingly that he could not understand the very sense of the operators. If we annihilate or create an electron, then what about the system's electroneutrality? Happily enough these operators will always act in creator–annihilator pairs.

[3] *Proof.* Let us take two Slater determinants $\Phi_a = \Phi^{N+1} (...1_k...)$ and $\Phi_b = \Phi^N (...0_k...)$. In both of them the occupancies of all other spin orbitals are identical. Let us write the normalization condition for Φ_b in the following way: $1 = \left\langle \Phi_b | \theta_k \hat{k} \Phi_a \right\rangle = \theta_k \left\langle \Phi_b | \hat{k} \Phi_a \right\rangle = \theta_k \left\langle \hat{k}^\# \Phi_b | \Phi_a \right\rangle$, where as $\hat{k}^\#$ has been denoted the operator adjoint to $\hat{k}$, θ_k appeared in order to compensate ($\theta_k^2 = 1$) the θ_k produced by the annihilator. On the other hand, from the normalization condition of Φ_a we see that $1 = \langle \Phi_a | \Phi_a \rangle = \theta_k \left\langle \hat{k}^\dagger \Phi_b | \Phi_a \right\rangle$. Hence,

$$\theta_k \left\langle \hat{k}^\# \Phi_b | \Phi_a \right\rangle = \theta_k \left\langle \hat{k}^\dagger \Phi_b | \Phi_a \right\rangle \text{ or } \hat{k}^\# = \hat{k}^\dagger,$$

quod erat demonstrandum.

ANTICOMMUTATION RULES

$$[\hat{k}, \hat{l}]_+ = 0,$$
$$[\hat{k}^\dagger, \hat{l}^\dagger]_+ = 0,$$
$$[\hat{k}^\dagger, \hat{l}]_+ = \delta_{kl},$$

where the symbol $[\hat{A}, \hat{B}]_+ = \hat{A}\hat{B} + \hat{B}\hat{A}$ is called the *anticommutator*.[4] It is simpler than the Slater–Condon rules, isn't it? Let us check the rule $[\hat{k}^\dagger, \hat{l}]_+ = \delta_{kl}$. We have to check how it works for all possible occupancies of the spin orbitals k and l, $(n_k, n_l) : (0,0), (0,1), (1,0)$, and $(1,1)$.

Case: $(n_k, n_l) = (0, 0)$

We have

$$[\hat{k}^\dagger, \hat{l}]_+ \Phi^N(\ldots 0_k \ldots 0_l \ldots) = [\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger]\Phi^N(\ldots 0_k \ldots 0_l \ldots) =$$
$$\hat{k}^\dagger \hat{l}\Phi^N(\ldots 0_k \ldots 0_l \ldots) + \hat{l}\hat{k}^\dagger \Phi^N(\ldots 0_k \ldots 0_l \ldots) =$$
$$0 + \hat{l}\theta_k \Phi^{N+1}(\ldots 1_k \ldots 0_l \ldots) = \theta_k \hat{l}\Phi^{N+1}(\ldots 1_k \ldots 0_l \ldots) = \theta_k \delta_{kl}\theta_k \Phi^N(\ldots 0_k \ldots) = \delta_{kl}\Phi^N(\ldots 0_k \ldots).$$

So far so good.

Case: $(n_k, n_l) = (0, 1)$

We have

$$[\hat{k}^\dagger, \hat{l}]_+ \Phi^N(\ldots 0_k \ldots 1_l \ldots) = [\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger]\Phi^N(\ldots 0_k \ldots 1_l \ldots) =$$
$$\hat{k}^\dagger \hat{l}\Phi^N(\ldots 0_k \ldots 1_l \ldots) + \hat{l}\hat{k}^\dagger \Phi^N(\ldots 0_k \ldots 1_l \ldots) =$$
$$\theta_k \theta_l \Phi^N(\ldots 1_k \ldots 0_l \ldots) - \theta_k \theta_l \Phi^N(\ldots 1_k \ldots 0_l \ldots) = \delta_{kl}\Phi^N(\ldots 0_k \ldots 1_l \ldots).$$

That is what we expected.[5]

---

[4] The above formulae are valid under the (common) assumption that the spin orbitals are orthonormal. If this assumption is not true, then only the last anticommutator changes to the form $[\hat{k}^\dagger, \hat{l}]_+ = S_{kl}$, where S_{kl} stands for the overlap integral of spin orbitals k and l.

[5] What decided is the change of sign (due to θ_k) when the order of the operators has changed.

Case: $(n_k, n_l) = (1, 0)$

We have

$$[\hat{k}^\dagger, \hat{l}]_+ \Phi^N(...1_k...0_l...) = [\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger]\Phi^N(...1_k...0_l...) =$$
$$\hat{k}^\dagger \hat{l}\Phi^N(...1_k...0_l...) + \hat{l}\hat{k}^\dagger \Phi^N(...1_k...0_l...) =$$
$$(0+0)\,\Phi^N(...1_k...0_l...) = \delta_{kl}\Phi^N(...1_k...0_l...).$$

That is OK.

Case: $(n_k, n_l) = (1, 1)$

We have

$$[\hat{k}^\dagger, \hat{l}]_+ \Phi^N(...1_k...1_l...) = [\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger]\Phi^N(...1_k...1_l...) =$$
$$\hat{k}^\dagger \hat{l}\Phi^N(...1_k...1_l...) + \hat{l}\hat{k}^\dagger \Phi^N(...1_k...1_l...) =$$
$$\hat{k}^\dagger \hat{l}\Phi^N(...1_k...1_l...) + 0 = \theta_k^2 \delta_{kl}\Phi^N(...1_k...1_l...) = \delta_{kl}\Phi^N(...1_k...).$$

OK, the thing is over.

Operators in the second quantization

The creation and annihilation operators may be used to represent the one- and two-electron operators. The resulting matrix elements with Slater determinants[6] *correspond exactly to the Slater–Condon rules* (see Appendix V1-N, p. V1-707).

One-electron operators

The operator $\hat{F} = \sum_i \hat{h}(i)$ is a sum of the one-electron operators[7] $\hat{h}(i)$ acting on functions of the coordinates of electron i.

Slater–Condon rule I says (see Appendix V1-N) that for the Slater determinant ψ built of the spin orbitals ϕ_i the matrix element $\left\langle \psi | \hat{F} \psi \right\rangle = \sum_i h_{ii}$, where $h_{ij} = \left\langle \phi_i | \hat{h} \phi_j \right\rangle$.

[6] The original operator and its representation in the language of the second quantization are not identical though. The second ones can act only on the Slater determinants or their combinations, while the first ones have a larger domain. Since we are going to work with the creation and annihilation operators in those methods only that use Slater determinants (CI, MC SCF, etc.), the difference is irrelevant.

[7] Most often this will be the kinetic energy operator, the nuclear attraction operator, the interaction with the external field, or the multipole moment.

In the second quantization

$$\hat{F} = \sum_{ij}^{\infty} h_{ij} \hat{i}^{\dagger} \hat{j}.$$

Interestingly, the summation extends to infinity, and, therefore, the operator is independent of the number of electrons in the system.

Let us check whether the formula is correct. Let us insert $\hat{F} = \sum_{ij} h_{ij} \hat{i}^{\dagger} \hat{j}$ into $\left\langle \psi | \hat{F} \psi \right\rangle$. We have

$$\left\langle \psi | \hat{F} \psi \right\rangle = \left\langle \psi | \sum_{ij} h_{ij} \hat{i}^{\dagger} \hat{j} \psi \right\rangle = \sum_{ij} h_{ij} \left\langle \psi | \hat{i}^{\dagger} \hat{j} \psi \right\rangle = \sum_{ij} h_{ij} \delta_{ij} = \sum_{i} h_{ii}.$$

This is a correct result.

What about Slater–Condon rule II (the Slater determinants ψ_1 and ψ_2 differ by a single spin orbital: the spin orbital i in ψ_1 is replaced by the spin orbital i' in ψ_2)? We have

$$\left\langle \psi_1 | \hat{F} \psi_2 \right\rangle = \sum_{ij} h_{ij} \left\langle \psi_1 | \hat{i}^{\dagger} \hat{j} \psi_2 \right\rangle.$$

The Slater determinants that differ by one spin orbital produce the overlap integral equal to zero,[8] and therefore $\left\langle \psi_1 | \hat{F} \psi_2 \right\rangle = h_{ii'}$. Thus, the operator in the form $\hat{F} = \sum_{ij} h_{ij} \hat{i}^{\dagger} \hat{j}$ ensures equivalence with all the Slater–Condon rules.

Two-electron operators

Similarly, we may use the creation and annihilation operators to represent the two-electron operators $\hat{G} = \frac{1}{2} \sum'_{ij} \hat{g}(i, j)$. In most cases $\hat{g}(i, j) = \frac{1}{r_{ij}}$ and $\hat{G}$ has the following form:

$$\hat{G} = \frac{1}{2} \sum_{ij}{}' \frac{1}{r_{ij}} = \frac{1}{2} \sum_{ijkl}^{\infty} \langle ij|kl \rangle \, \hat{j}^{\dagger} \hat{i}^{\dagger} \hat{k} \hat{l}.$$

[8] It is evident that if in this situation the Slater determinants ψ_1 and ψ_2 differed by more than a single spin orbital, then we would get zero (Slater–Condon rules III and IV).

Here also the summation extends to infinity and the operator is independent of the number of electrons in the system.

The proof of Slater–Condon rule I relies on the following chain of equalities:

$$\left\langle \psi | \hat{G} \psi \right\rangle = \frac{1}{2} \sum_{ijkl} \langle ij|kl\rangle \left\langle \psi | \hat{\jmath}^\dagger \hat{\imath}^\dagger \hat{k} \hat{l} \psi \right\rangle = \frac{1}{2} \sum_{ijkl} \langle ij|kl\rangle \left\langle \hat{\imath}\hat{\jmath}\psi | \hat{k}\hat{l}\psi \right\rangle =$$

$$\frac{1}{2} \sum_{ijkl} \langle ij|kl\rangle \left(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}\right) = \frac{1}{2} \sum_{ij} \left(\langle ij|ij\rangle - \langle ij|ji\rangle\right),$$

because the overlap integral $\left\langle \hat{\imath}\hat{\jmath}\psi | \hat{k}\hat{l}\psi \right\rangle$ of the two Slater determinants $\hat{\imath}\hat{\jmath}\psi$ and $\hat{k}\hat{l}\psi$ is nonzero in two cases only: either if $i = k$, $j = l$ or if $i = l$, $j = k$ (then the sign has to change). This is what we get from the Slater–Condon rules.

For Slater–Condon rule II we have (instead of the spin orbital i in ψ_1 we have the spin orbital i' in ψ_2)

$$\left\langle \psi_1 | \hat{G} \psi_2 \right\rangle = \frac{1}{2} \sum_{Ijkl} \langle Ij|kl\rangle \left\langle \psi_1 | \hat{\jmath}^\dagger \hat{I}^\dagger \hat{k} \hat{l} \psi_2 \right\rangle = \frac{1}{2} \sum_{Ijkl} \langle Ij|kl\rangle \left\langle \hat{I}\hat{\jmath}\psi_1 | \hat{k}\hat{l}\psi_2 \right\rangle, \tag{C.1}$$

where the summation index I has been introduced in order not to mix up with spin orbital i. In the overlap integral $\left\langle \hat{I}\hat{\jmath}\psi_1 | \hat{k}\hat{l}\psi_2 \right\rangle$ the sets of the spin orbitals in the Slater determinant $\hat{I}\hat{\jmath}\psi_1$ and in the Slater determinant $\hat{k}\hat{l}\psi_2$ have to be *identical*; otherwise the integral will be equal to zero. However, already in ψ_1 and ψ_2 we have a difference of one spin orbital. Thus, first of all we have to get rid of just these spin orbitals (i and i')! For the integral to survive[9] we have to have at least one of the following conditions to be satisfied:

- $I = i$ and $k = i'$ (and then $j = l$),
- $j = i$ and $k = i'$ (and then $I = l$),
- $I = i$ and $l = i'$ (and then $j = k$),
- $j = i$ and $l = i'$ (and then $I = k$).

This means that when taking into account the above cases in Eq. (C.1) we obtain

$$\left\langle \psi_1 | \hat{G} \psi_2 \right\rangle = \frac{1}{2} \sum_{j} \left\langle ij|i'j\right\rangle \left\langle \hat{\imath}\hat{\jmath}\psi_1 | \hat{\imath}'\hat{\jmath}\psi_2 \right\rangle + \frac{1}{2} \sum_{l} \left\langle li|i'l\right\rangle \left\langle \hat{l}\hat{\imath}\psi_1 | \hat{\imath}'\hat{l}\psi_2 \right\rangle$$

$$+ \frac{1}{2} \sum_{j} \left\langle ij|ji'\right\rangle \left\langle \hat{\imath}\hat{\jmath}\psi_1 | \hat{\jmath}\hat{\imath}'\psi_2 \right\rangle + \frac{1}{2} \sum_{k} \left\langle ki|ki'\right\rangle \left\langle \hat{k}\hat{\imath}\psi_1 | \hat{k}\hat{\imath}'\psi_2 \right\rangle = \frac{1}{2} \sum_{j} \left\langle ij|i'j\right\rangle - \frac{1}{2} \sum_{l} \left\langle li|i'l\right\rangle,$$

[9] This is necessary, not a sufficient condition.

$$
\begin{aligned}
-\frac{1}{2}\sum_j \langle ij|ji'\rangle + \frac{1}{2}\sum_k \langle ki|ki'\rangle &= \frac{1}{2}\sum_j \langle ij|i'j\rangle - \frac{1}{2}\sum_j \langle ji|i'j\rangle, \\
-\frac{1}{2}\sum_j \langle ij|ji'\rangle + \frac{1}{2}\sum_j \langle ji|ji'\rangle &= \frac{1}{2}\sum_j \langle ij|i'j\rangle - \frac{1}{2}\sum_j \langle ij|ji'\rangle, \\
-\frac{1}{2}\sum_j \langle ij|ji'\rangle + \frac{1}{2}\sum_j \langle ij|i'j\rangle &= \sum_j \langle ij|i'j\rangle - \sum_j \langle ij|ji'\rangle,
\end{aligned}
$$

where in the two sums the coordinates of electrons 1 and 2 have been exchanged, and it has been noted that the overlap integrals $\langle \hat{i}\hat{j}\psi_1|\hat{i}'\hat{j}\psi_2\rangle = \left\langle \hat{k}\hat{i}\psi_1|\hat{k}\hat{i}'\psi_2 \right\rangle = 1$, because the Slater determinants $\hat{i}\psi_1$ and $\hat{i}'\psi_2$ are identical. Also, from the anticommutation rules $\left\langle \hat{l}\hat{i}\psi_1|\hat{i}'\hat{l}\psi_2 \right\rangle = \langle \hat{i}\hat{j}\psi_1|\hat{j}\hat{i}'\psi_2\rangle = -1$. Thus, Slater–Condon rule II has been correctly reproduced: $\left\langle \psi_1|\hat{G}\psi_2 \right\rangle = \sum_j \left[\langle ij|i'j\rangle - \langle ij|ji'\rangle\right]$.

We may conclude that the definition of the creation and annihilation operators and the simple anticommutation relations are equivalent to the Slater–Condon rules. This opens up for us the space spanned by the Slater determinants, i.e., all the integrals involving Slater determinants can be easily transformed into the one- and two-electron integrals involving spin orbitals.

APPENDIX D

Population Analysis

On p. 194 the electronic density ρ is defined. If the wave function is a Slater determinant (p. V1-465) and assuming double occupancy of orbitals φ_i, we have (see (3.9))

$$\rho(\mathbf{r}) = 2\left[|\varphi_1(\mathbf{r})|^2 + |\varphi_2(\mathbf{r})|^2 + \ldots \left|\varphi_{\frac{N}{2}}(\mathbf{r})\right|^2\right]. \tag{D.1}$$

The density distribution ρ may be viewed as a cloud carrying a charge $-Ne$ and Eq. (D.1) says that the cloud is composed of the interpenetrating individual clouds of the molecular orbitals, each carrying two electrons. On the other hand, in the LCAO approximation any molecular orbital is represented by a sum of atomic orbitals. If we insert the LCAO expansion into ρ, then ρ becomes a sum of the contributions, each being a product of two atomic orbitals. There is a temptation to go even further and to divide ρ somehow *into contributions of particular atoms*, calculate the charge corresponding to such contribution, and locate the (point) charge right on the nucleus.[1] We might say therefore, what is the "electron population" residing on the particular atoms (hence the name: population analysis)?

Mulliken population analysis

Such tricks are of course possible, and one of them is the so-called Mulliken population analysis. From (D.1), after using the LCAO expansion $\varphi_i = \sum_r c_{ri}\chi_r$, we have ($S_{rs}$ stand for the overlap integrals between the atomic orbitals r and s, c are the corresponding LCAO coefficients)

$$N = \int \rho(\mathbf{r})\,\mathrm{d}V = 2\sum_{i=1}^{N/2}\int |\varphi_i(\mathbf{r})|^2\,\mathrm{d}V = \sum_i\sum_{rs} 2c_{ri}^* c_{si} S_{rs} = \sum_{rs} P_{rs}S_{rs} = \mathrm{Tr}(\mathbf{PS}), \tag{D.2}$$

where $\mathbf{P}$ is the so-called charge and bond-order matrix

$$P_{sr} = \sum_i 2c_{ri}^* c_{si}. \tag{D.3}$$

[1] This number need not be integer.

The summation over r and s may be carried out with highlighting from which atom A the particular atomic orbital comes (we assume that the AOs are centered on the nuclei). We get an equivalent formula (A and B denote atoms), i.e.,

$$N = \sum_A \sum_{r \in A} \sum_B \sum_{s \in B} P_{rs} S_{rs}.$$

Afterwards we may choose the following partitionings.

Atomic partitioning. We have

$$N = \sum_A q_A,$$

$$q_A = \sum_{r \in A} \left(\sum_B \sum_{s \in B} P_{rs} S_{rs} \right),$$

where q are the so-called Mulliken charges. They are often computed in practical applications and serve as information about how much of the electronic density ρ is concentrated on atom A. Such a quantity is of interest, because it may be directly linked to the reactivity of atom A, often identified with its ability to be attacked by a nucleophilic or an electrophilic agent.[2] Also, if somebody measures the dipole moments, then he/she would like to know why this moment is particularly large in a molecule. Performing the Mulliken analysis we are able to identify those atoms that are responsible for that. This might be of value when interpreting experimental data.

Atomic and bond partitioning. The summation may be performed also in a slightly different way. Then we have

$$N = \sum_A \sum_{r,s \in A} P_{rs} S_{rs} + \sum_{A<B} 2 \sum_{r \in A} \sum_{s \in B} P_{rs} S_{rs} = \sum_A \bar{q}_A + \sum_{A<B} \bar{q}_{AB}.$$

The first term represents the contributions $\bar{q}_A$ of the atoms, the second term pertains to the atomic pairs $\bar{q}_{AB}$.

The last populations are large and positive for those pairs of atoms for which chemists assign chemical bonds.

[2] We have to remember that besides electrons this atom has a nucleus. This has to be taken into account when calculating the atomic net charge.

The bond population $\bar{q}_{AB}$ may be treated as a measure whether in the A–B atomic interaction the bonding or antibonding character prevails.[3] If for two atoms $\bar{q}_{AB} < 0$, then we may say that they are not bound by any chemical bond; if $\bar{q}_{AB}$ is large, then we may treat it as an indication that these two atoms are bound by a chemical bond or bonds.

Example 1 (Hydrogen molecule). Let us take the simplest example. First, let us consider the electronic ground state in the simplest molecular orbital approximation, i.e., the two electrons are described by the normalized orbital in the form (a, b denote the $1s$ atomic orbitals centered on the corresponding nuclei; note that what we take is the famous bonding orbital)

$$\varphi_1 = N_1 (a + b) ,$$

where $N_1 = (2 + 2S)^{-\frac{1}{2}}$ and $S \equiv (a|b)$. Then $P_{sr} = \sum_i 2c^*_{ri} c_{si} = 2c^*_{r1} c_{s1} = (1 + S)^{-1}$, independently of the indices r and s. Of course, $\mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}$, and therefore $\mathbf{PS} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Thus, $\mathrm{Tr}\,(\mathbf{PS}) = 2 =$ the number of electrons $= P_{11}S_{11} + P_{22}S_{22} + 2P_{12}S_{12} = q_A + q_B + q_{AB}$, with $q_A = q_B = (1 + S)^{-1}$, and $q_{AB} = \frac{2S}{1+S} > 0$. Thus, we immediately see that the H–H bond *has the electronic population greater than zero*, i.e., the atom–atom interaction is *bonding*.

Let us now consider H_2 with the two electrons occupying the normalized orbital of different character,[4] $\varphi_2 = N_2 (a - b)$, with $N_2 = (2 - 2S)^{-\frac{1}{2}}$. Then $P_{sr} = \sum_i 2c^*_{ri} c_{si} = 2c^*_{r2} c_{s2} = (1 - S)^{-1}$ for $(r, s) = (1, 1)$ and $(r, s) = (2, 2)$ while $P_{rs} = -(1 - S)^{-1}$ for $(r, s) = (1, 2)$ and $(r, s) = (2, 1)$.

Now, let us calculate $\mathbf{PS} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and $\mathrm{Tr}(\mathbf{PS}) = 2 =$ the number of electrons $= P_{11}S_{11} + P_{22}S_{22} + 2P_{12}S_{12} = q_A + q_B + q_{AB}$, but now $q_A = q_B = (1 - S)^{-1}$, and $q_{AB} = -\frac{2S}{1-S} < 0$. Thus, q_{AB} tells us that this time the atoms are interacting in the antibonding way.

A similar analysis for polyatomic molecules gives more subtle and more interesting results.

---

[3] P_{rs} is a sum (over the occupied orbitals) of products of the LCAO coefficients of two atoms in each of the occupied molecular orbitals. Equal signs of these coefficients (with $S_{rs} > 0$) mean a *bonding* interaction (recall Chapter V1-8 and Appendix V1-S on p. V1-735) and such a contribution increases P_{rs}. Opposite signs of the coefficients (with $S_{rs} > 0$) correspond to the antibonding interactions and in such a case the corresponding contribution decreases P_{rs}. If $S_{rs} < 0$, then the words "bonding" and "antibonding" above have to be exchanged, but the effect remains the same. This means that the product $P_{rs}S_{rs}$ in all cases correctly controls the bonding ($P_{rs}S_{rs} > 0$) or antibonding ($P_{rs}S_{rs} < 0$) effects.

[4] We do not want to suggest anything, but the orbital is notorious for its antibonding character.

Other population analyses

Partitioning of the electron cloud of N electrons according to the Mulliken population analysis represents only one of possible choices. For a positively definite matrix[5] $\mathbf{S}$ (and the overlap matrix is always positively definite) we may introduce the powers of the matrix[6] $\mathbf{S}^x$, where x is an arbitrary real number (in a way shown in Appendix V1-K on p. V1-697), and we have $\mathbf{S}^{1-x}\mathbf{S}^x = \mathbf{S}$. Then, we may write[7]

$$N = \mathrm{Tr}(\mathbf{PS}) = \mathrm{Tr}(\mathbf{S}^x\mathbf{PS}^{1-x}). \tag{D.4}$$

Now, we may take any x and for this value construct the corresponding partition of N electronic charges into atoms. If $x = 0$ or 1, then one has the Mulliken population analysis; if $x = \frac{1}{2}$, then we have the so-called *Löwdin population analysis*.

Multipole representation

Imagine a charge distribution $\rho(\mathbf{r})$. Let us choose a Cartesian coordinate system. We may compute the Cartesian moments of the distribution $\int \rho(\mathbf{r})\,\mathrm{d}V$, i.e., the total charge, $\int x\rho(\mathbf{r})\,\mathrm{d}V$, $\int y\rho(\mathbf{r})\,\mathrm{d}V$, $\int z\rho(\mathbf{r})\,\mathrm{d}V$, i.e., the components of the dipole moment, $\int x^2\rho(\mathbf{r})\,\mathrm{d}V$, $\int y^2\rho(\mathbf{r})\,\mathrm{d}V$, $\int z^2\rho(\mathbf{r})\,\mathrm{d}V$, $\int xy\rho(\mathbf{r})\,\mathrm{d}V$, $\int xz\rho(\mathbf{r})\,\mathrm{d}V$, $\int yz\rho(\mathbf{r})\,\mathrm{d}V$, i.e., the components of the quadrupole moment, etc. The moments mean a complete description of $\rho(\mathbf{r})$ as concerns its interaction with another (distant) charge distribution. The higher the powers of x, y, z (i.e., the higher the moment), the more important are distant parts (from the origin) of $\rho(\mathbf{r})$. If $\rho(\mathbf{r})$ extends to infinity (and for atoms and molecules it does), higher-order moments tend to infinity. This means trouble when the consecutive interactions of the multipole moments are computed (multipole expansion, Appendix G) and indeed, the multipole expansion "explodes," i.e., diverges.[8] This would not happen if the interacting charge distributions did not overlap.

There is also another problem: where to locate the origin of the coordinate system with respect to which the moments are calculated? The answer is: *anywhere*. Wherever such origin is located it is OK from the point of view of mathematics. However, such choices may differ enormously from the practical point of view. For example, let us imagine a spherically symmetric charge distribution. If the origin is located in its center (as "most people would do"), then one has a quite simple description of $\rho(\mathbf{r})$ by using the moments, namely, the only nonzero moment is

[5] That is, all the eigenvalues positive.

[6] They are symmetric matrices as well.

[7] We easily check that $\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB})$. Indeed, $\mathrm{Tr}(\mathbf{ABC}) = \sum_{i,k,l} A_{ik}B_{kl}C_{li}$, while $\mathrm{Tr}(\mathbf{CAB}) = \sum_{i,k,l} C_{ik}A_{kl}B_{li}$. Changing summation indices $k \to i,\ l \to k,\ i \to l$ in the last formula, we obtain $\mathrm{Tr}(\mathbf{ABC})$.

[8] Although the first terms (i.e., before the "explosion") may give accurate results.

the charge, i.e., $\int \rho(\mathbf{r})\,\mathrm{d}V$. If, however, the origin were located off-center, then all the moments would be nonzero. All they are needed to calculate accurately the interaction (with something) of such simple object as a sphere. As we can see, it is definitely better to locate the origin in the center of $\rho(\mathbf{r})$.

Well, what if the charge distribution $\rho(\mathbf{r})$ were divided into segments and each segment were represented by a set of multipoles? It would be all right, although, in view of the above example, it would be better to locate the corresponding origins in the centers of the segments. It is clear that in particular it would be good if the segments were very small, e.g., if the cloud is cut into tiny cubes and one considers every cube's content as a separate cloud.[9] But, well, what are the multipoles for? Indeed, it would be sufficient to take the charges of the cubes only, because they approximate the original charge distribution. In this situation higher multipoles would be certainly irrelevant! Thus, we have two extreme cases:

- a single origin and an infinite number of multipoles, or
- an infinite number of centers and the monopoles (charges) only.

It is seen that when the origins are located on atoms we have an intermediary situation and it might be sufficient to have a few multipoles per atom.[10] This is what the concept of the so-called cumulative multipole moments (CAMM[11]) is all about. Besides the isotropic atomic charges $q_a = M_a^{(000)}$ computed in an arbitrary population analysis we have in addition higher multipoles $M_a^{(klm)}$ (atomic dipoles, quadrupoles, octupoles, etc.) representing the anisotropy of the atomic charge distribution (i.e., they describe the deviations of the atomic charge distributions from spherical ones)

$$M_a^{(klm)} = Z_a x_a^k y_a^l z_a^m - \sum_{r\in a}\sum_{s} D_{sr}\left(r|x^k y^l z^m|s\right)$$
$$-\sum_{k'\le k}\sum_{l'\le l}\sum_{m'\le m,(k'l'm')\neq(klm)} \begin{pmatrix} k \\ k' \end{pmatrix}\begin{pmatrix} l \\ l' \end{pmatrix}\begin{pmatrix} m \\ m' \end{pmatrix} x_a^{k-k'} y_a^{l-l'} z_a^{m-m'} \cdot M_a^{k'l'm'},$$

where $M_a^{(klm)}$ is the multipole moment of the "klm" order with respect to the Cartesian coordinates x, y, z located on atom a ($M_a^{(000)}$ standing for atomic charge, e.g., from the Mulliken population analysis), Z_a denotes the nuclear charge of atom a, $\left(r|x^k y^l z^m|s\right)$ stands for the

[9] The clouds might eventually overlap.

[10] If the clouds overlapped, the description of each center by an infinite number of multipoles would lead to a redundancy ("overcompleteness"). I do not know about any troubles of that kind, but in my opinion troubles would come if the numbers of the origins were large. This is in full analogy with the overcompleteness of the LCAO expansion. These two examples differ by a secondary feature: in the LCAO instead of moments we have the s,p,d,... orbitals, i.e., some moments multiplied by exponential functions.

[11] W.A. Sokalski, R. Poirier, *Chem. Phys. Lett.*, 98(1983)86; W.A. Sokalski, A. Sawaryn, *J. Chem. Phys.*, 87(1987)526.

one-electron integral of the corresponding multipole moment, and $D_{sr}\chi_r^*\chi_s$ represents the electronic density contribution related to AOs χ_s and χ_r and calculated by any method (LCAO MO SCF, CI, MP2, DFT, etc.). We may also use the multipole moments expressed by the spherical harmonic functions as proposed by Stone.[12]

[12] A.J. Stone, *Chem. Phys. Lett.*, 83(1981)233; A.J. Stone, M. Alderton, *Mol. Phys.*, 56(1985)1047.

APPENDIX E

Pauli Deformation

Two molecules, when isolated (say, at infinite distance), are independent and the wave function of the total system might be taken as a *product* of the wave functions for the individual molecules. When the same two molecules are at a finite distance, any product-like function represents only an approximation, sometimes a very poor[1] one, because according to a postulate of quantum mechanics the wave function has to be antisymmetric with respect to the exchange of electronic labels, while the product does not fulfill that condition. More exactly, the approximate wave function has to belong to the irreducible representation of the symmetry group of the Hamiltonian (see Appendix V1-C, p. V1-605) to which the ground-state wave function belongs. This means first of all that the Pauli exclusion principle is to be satisfied.

PAULI DEFORMATION

The product-like wave function has to be antisymmetrized. This makes some changes in the electronic charge distribution (electronic density), which will be called the Pauli deformation.

The Pauli deformation may be viewed as a mechanical distortion of both interacting molecules due their mutual pushing. The reason why two gum balls deform when pushed against each other is the same: the electrons of one ball cannot occupy the same space with the electrons (with the same spin coordinates) of the second ball. The most dramatic deformation takes place close to the contact area of these balls.

The norm of the difference of $\varphi^{(0)}$ and $\psi^{(0)}$ represents a very stringent measure of difference between two functions: any deviation gives a contribution to the measure. We would like to know how the electronic density changes, where the electrons flows from, and where they go to. The electron density ρ (a function of the position in space) is defined as a sum of densities ρ_i of the particular electrons, i.e.,

$$\rho\,(x, y, z) = \sum_{i=1}^{N} \rho_i\,(x, y, z)\,,$$

[1] For example, when the intermolecular distance is short, the molecules push one another and deform (maybe strongly), and the product-like function is certainly inadequate.

$$\rho_i(x_i, y_i, z_i) = \sum_{\sigma_i=-\frac{1}{2}}^{+\frac{1}{2}} \int \frac{d\tau}{d\tau_i} |\psi|^2, \tag{E.1}$$

where $d\tau = d\tau_1 d\tau_2 \ldots d\tau_N$, and therefore the integration goes over the coordinates (space and spin) of all the electrons except electron i. In addition, there is also a summation over the spin coordinate of electron i, because we are not interested in its value. As is seen the integral of $\rho(x, y, z)$ over x, y, z is equal to N; therefore, $\rho(x, y, z)$ represents an electron cloud carrying N electrons, the same as defined on p. 194. We make the two molecules approach without changing their charge distribution (the system is described by the electron density corresponding to the wave function $\psi = \varphi^{(0)}$), and then we allow the Pauli exclusion principle to operate in order to ensure the proper symmetry of the wave function (the system is therefore described by a new wave function $\psi = \psi^{(0)}$) by applying a suitable projection operator. What happens to the electronic density? Will it change or not?

Let us see what will happen when we make such an approach of two hydrogen atoms and then of two helium atoms.

H_2 case

In the case of two hydrogen atoms[2] $\varphi^{(0)} = 1s_a(1)\alpha(1)1s_b(2)\beta(2) \equiv a(1)\alpha(1)b(2)\beta(2)$, where we have used the abbreviations $1s_a(1) \equiv a$ and $1s_b(1) \equiv b$. After inserting $\psi = \varphi^{(0)}$ into (E.1), integration over space, and summation over spin coordinates we obtain $\rho^{(0)} = \rho_1(x, y, z) + \rho_2(x, y, z)$, where

$$\rho_1(x, y, z) = \sum_{\sigma_1=-\frac{1}{2}}^{+\frac{1}{2}} \int \frac{d\tau}{d\tau_1} |a(1)\alpha(1)b(2)\beta(2)|^2 = \sum_{\sigma_1=-\frac{1}{2}}^{+\frac{1}{2}} \int d\tau_2 |a(1)\alpha(1)b(2)\beta(2)|^2 = a^2.$$

Similarly,

$$\rho_2(x, y, z) = \sum_{\sigma_2=-\frac{1}{2}}^{+\frac{1}{2}} \int \frac{d\tau}{d\tau_2} |a(1)\alpha(1)b(2)\beta(2)|^2 =$$

$$\sum_{\sigma_2=-\frac{1}{2}}^{+\frac{1}{2}} \int d\tau_1 |a(1)\alpha(1)b(2)\beta(2)|^2 = b^2.$$

[2] We assign arbitrarily the spin function α to electron 1 and the spin function β to electron 2. We might have done this in the opposite way, but it does not change anything.

Thus, finally $\rho^{(0)} = a^2 + b^2$. This density is normalized to 2 – as it has to be, because the electron cloud $\rho(x, y, z)$ carries two electrons. Now, let us do the same for the wave function $\psi^{(0)} = N\hat{\mathcal{A}}\varphi^{(0)}$, where $\hat{\mathcal{A}}$ stands for the idempotent projection operator (5.27), the normalization constant $N = \frac{2}{\sqrt{1+S^2}}$ with $S = (a|b)$, and all quantities are as described in Chapter 5 about the symmetry-adapted perturbation theory. We obtain

$$\rho(x, y, z) = \rho_1(x, y, z) + \rho_2(x, y, z),$$

$$\rho_1(x, y, z) = \sum_{\sigma_1 = \pm\frac{1}{2}} \int d\tau_2 \left|\psi^{(0)}\right|^2 =$$

$$N^2 \frac{1}{8} \int dV_2 \left[a(1)\, b(2) + a(2)\, b(1)\right]^2 \sum_{\sigma_1} \sum_{\sigma_2} \frac{1}{2} \left[\alpha(1)\, \beta(2) - \alpha(2)\, \beta(1)\right]^2 =$$

$$N^2 \frac{1}{8} \int dV_2 \left[a(1)\, b(2) + a(2)\, b(1)\right]^2 = N^2 \frac{1}{8} \left(a^2 + b^2 + 2abS\right) =$$

$$\frac{1}{2\left(1 + S^2\right)} \left(a^2 + b^2 + 2abS\right),$$

$$\rho_2(x, y, z) = \rho_1(x, y, z).$$

As is seen, the density $\rho_1(x, y, z)$ is normalized to 1 – this is what we would get after integration over dV_1. A similar calculation for ρ_2 would give the same result, because $\left|\psi^{(0)}\right|^2$ is symmetric with respect to the exchange of electrons[3] 1 and 2. Therefore, the *change* of the electron density due to the proper symmetry projection (including the Pauli exclusion principle)

$$\rho - \rho^{(0)} = \frac{a^2 + b^2 + 2abS}{1 + S^2} - (a^2 + b^2) = \frac{2S}{1 + S^2} ab - \frac{S^2}{1 + S^2} a^2 - \frac{S^2}{1 + S^2} b^2. \tag{E.2}$$

Thus, it turns out that as a result of the Pauli exclusion principle (i.e., of the antisymmetrization of the wave function) an electron density $a^2S^2/(1 + S^2)$ flows from atom a, and a similar thing happens to atom b, where the electronic density decreases by $b^2S^2/(1 + S^2)$. Both these electronic clouds go to the bond region – we find them as an electron cloud $2abS/(1 + S^2)$, the integral of $\rho - \rho^{(0)}$ is equal to zero (Fig. E.1a).

Thus,

> in the hydrogen molecule the Pauli exclusion principle causes a sticking of the two atoms together (the two electrons increase their probability to be in the region between the two nuclei).

[3] This was not the case for $\varphi^{(0)}$.

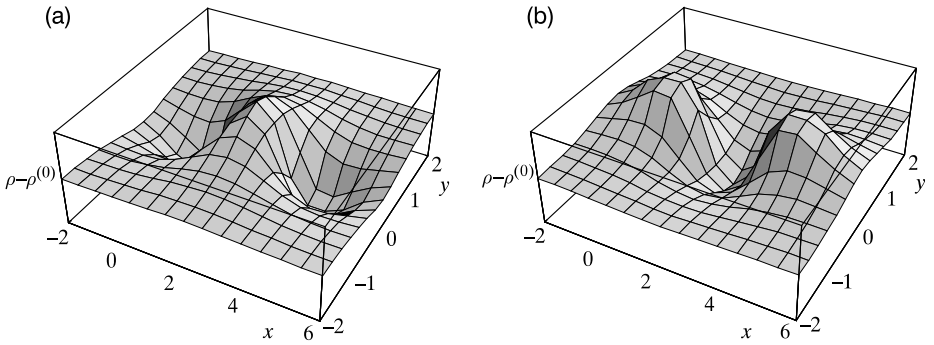


Fig. E.1. Comparison of the Pauli deformation for two hydrogen atoms and for two helium atoms. (a) Two hydrogen atoms. Visualization of $\rho - \rho^{(0)}$ calculated in the plane containing the nuclei ("the net result is zero"). One of the protons is located at the origin, the other has the coordinates $(0, R, 0)$, with $R = 1.4$ a.u. (close to equilibrium). For this distance the overlap integral (see Appendix V1-S, p. V1-735) $S = (1 + R + \frac{R^2}{3})exp(-R)$ amounts to 0.752943. As we can see the electronic density has flown from the nuclei to the bond.(b) Two helium atoms. The only difference with respect to (a) is that two electrons have been added. The visualization of $\rho - \rho^{(0)}$ reveals a *completely different pattern.* This time the electronic density has been removed from the bond region and increased in the region of the nuclei.

This is what the Pauli exclusion principle dictates. Besides that we have, of course, all the physical interactions (electron repulsion, attraction with the nuclei) and the kinetic energy, but none of these effects has yet been taken into account.[4] Fig. E.1a shows the deformation that results only from forcing the proper symmetry in the wave function.

He_2 case

Let us see what happens if we make similar calculations for two helium atoms. In order to compare the future result with the H_2 case, let us keep everything the same (the internuclear distance R, the atomic orbitals, the overlap integral S, etc.), except that the number of electrons changes from 2 to 4. This time the calculation will be a little bit more tedious, because

[4] Indeed, all these effects have been ignored, because we neither calculated the energy, nor used the Hamiltonian. However, the very fact that we write $\varphi^{(0)} = a(1)\alpha(1)b(2)\beta(2)$, where a and b stand for the properly centered $1s$ orbitals, means that the electron–nucleus interaction has been *implicitly* taken into account (this is why the $1s$ orbital appears). Similarly, when we project the product-like function and obtain $\psi^{(0)}$ proportional to $[a(1)b(2) + a(2)b(1)]\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big]$, then besides the abovementioned electron–nucleus interactions (manifested by the $1s$ orbitals) we obtain an interesting effect: when one electron is on nucleus a, then the second electron runs to nucleus b. It looks as if they repelled each other! This is, however, at the level of the *mathematical formula of the function* ("function design"), as if the function has already been designed quite well for the *future*, taking into account the physical interactions.

four-electron wave functions are more complicated than two-electron functions. For example, the function $\varphi^{(0)}$ can be approximated as the product of the two Slater determinants – one for atom a, the other for atom b – i.e.,

$$\varphi^{(0)} = N \begin{vmatrix} a\alpha(1) & a\alpha(2) \\ a\beta(1) & a\beta(2) \end{vmatrix} \begin{vmatrix} b\alpha(3) & b\alpha(4) \\ b\beta(3) & b\beta(4) \end{vmatrix} = \tag{E.3}$$

$$N' a(1)a(2)b(3)b(4) \left[\frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right]\right]\left[\frac{1}{\sqrt{2}}\left[\alpha(3)\beta(4) - \alpha(4)\beta(3)\right]\right],$$

where the normalization constant $N' = 1$ (easy to verify: just square the function and integrate – all that by heart). Then we obtain directly from the definition[5]

$$\rho^{(0)} = \rho_1 + \rho_2 + \rho_3 + \rho_4 = 2a^2 + 2b^2,$$

which after integration gives four electrons, as it should. The function $\varphi^{(0)}$ is "illegal," because it does not fulfill the Pauli exclusion principle; e.g., the exchange of electrons 1 and 3 does not lead to a change of sign of the wave function.

Now let us focus on $\psi^{(0)}$. Please note that $\varphi^{(0)}$ of Eq. (E.3) may be written alternatively as

$$\varphi^{(0)} = N \begin{vmatrix} a\alpha\,(1) & a\alpha\,(2) & 0 & 0 \\ a\beta\,(1) & a\beta\,(2) & 0 & 0 \\ 0 & 0 & b\alpha\,(3) & b\alpha\,(4) \\ 0 & 0 & b\beta\,(3) & b\beta\,(4) \end{vmatrix},$$

where N is a normalization constant.

Antisymmetrization of $\varphi^{(0)}$, in which electrons 1 and 2 occupy orbital a and electrons 3 and 4 occupy orbital b, is equivalent to completing the Slater determinant[6] in a way allowing for the exchange of electrons between the subsystems, i.e.,

$$\psi^{(0)} = N\frac{1}{2}(1+I)\,\hat{A}\varphi^{(0)} = N\frac{1}{2}(1+I) \begin{vmatrix} a\alpha\,(1) & a\alpha\,(2) & a\alpha\,(3) & a\alpha\,(4) \\ a\beta\,(1) & a\beta\,(2) & a\beta\,(3) & a\beta\,(4) \\ b\alpha\,(1) & b\alpha\,(2) & b\alpha\,(3) & b\alpha\,(4) \\ b\beta\,(1) & b\beta\,(2) & b\beta\,(3) & b\beta\,(4) \end{vmatrix} =$$

$$N \begin{vmatrix} a\alpha\,(1) & a\alpha\,(2) & a\alpha\,(3) & a\alpha\,(4) \\ a\beta\,(1) & a\beta\,(2) & a\beta\,(3) & a\beta\,(4) \\ b\alpha\,(1) & b\alpha\,(2) & b\alpha\,(3) & b\alpha\,(4) \\ b\beta\,(1) & b\beta\,(2) & b\beta\,(3) & b\beta\,(4) \end{vmatrix},$$

[5] This also may be calculated in memory (note that the spin functions in the square brackets are normalized).

[6] The Slater determinant containing linearly independent spin orbitals *guarantees* the antisymmetry.

where according to (5.27) $\hat{A}$ stands for the idempotent antisymmetrization operator, and $\frac{1}{2}(1+I)$ represents an idempotent symmetrization operator acting on the nuclear coordinates. The last equality follows from the fact that this particular Slater determinant is already symmetric with respect to the exchange of nuclei,[7] which is equivalent to $a \leftrightarrow b$.

Any determinant is invariant with respect to addition of any linear combination of its rows (columns) to a given row (column). For reasons that will become clear soon, let us make a series of such operations. First, let us add the third row to the first one, then multiply the third row by 2 (any multiplication is harmless for the determinant, because at the end it will be normalized), and subtract the first row from the third one. Then let us perform a similar series of operations applied to rows 2 and 4 (instead of 1 and 3), and at the end let us multiply rows 1 and 2 by $\frac{1}{\sqrt{2(1+S)}}$ and rows 3 and 4 by $\frac{1}{\sqrt{2(1-S)}}$. The result of these operations is the Slater determinant with the doubly occupied bonding molecular orbital $\sigma = \frac{1}{\sqrt{2(1+S)}}(a+b)$ and the doubly occupied antibonding molecular orbital $\sigma^* = \frac{1}{\sqrt{2(1-S)}}(a-b)$, and

$$\psi^{(0)} = \frac{1}{\sqrt{4!}} \begin{vmatrix} \sigma\alpha\,(1) & \sigma\alpha\,(2) & \sigma\alpha\,(3) & \sigma\alpha\,(4) \\ \sigma\beta\,(1) & \sigma\beta\,(2) & \sigma\beta\,(3) & \sigma\beta\,(4) \\ \sigma^*\alpha\,(1) & \sigma^*\alpha\,(2) & \sigma^*\alpha\,(3) & \sigma^*\alpha\,(4) \\ \sigma^*\beta\,(1) & \sigma^*\beta\,(2) & \sigma^*\beta\,(3) & \sigma^*\beta\,(4) \end{vmatrix}.$$

All the spin orbitals involved are orthonormal (in contrast to the original determinant) and the corresponding electronic density is easy to write down – it is the sum of squares of the molecular orbitals multiplied by their occupancies (cf. p. 595), i.e.,

$$\rho(x,y,z) = 2\sigma^2 + 2\left(\sigma^*\right)^2.$$

Now let us compute the Pauli deformation

$$\begin{aligned} \rho - \rho^{(0)} &= \frac{a^2+b^2+2ab}{1+S} + \frac{a^2+b^2-2ab}{1-S} - 2(a^2+b^2) \\ &= -\frac{4S}{1-S^2}ab + \frac{2S^2}{1-S^2}a^2 + \frac{2S^2}{1-S^2}b^2. \end{aligned} \tag{E.4}$$

Integration of the difference gives zero, as it should. Note that the formula is similar to what we have obtained for the hydrogen molecule, but this time the electron flow is completely different (Fig. E.1b).

[7] This corresponds to the exchange of rows in the determinant: the first with the third, and the second with the fourth. A single exchange changes the sign of the determinant; therefore, the two exchanges leave the determinant invariant.

In the case of He_2 the Pauli exclusion principle makes the electron density decrease in the region between the nuclei and increase close to the nuclei. In the case of the hydrogen molecule the two atoms stick together, while the two helium atoms deform as if they were gum spheres squeezed together (Pauli deformation).

The *only* thing that has been changed with respect to the hydrogen molecule is the increase of the number of electrons from two to four (we have not changed the orbital exponents equal to 1 as well as the internuclear distance equal to 4 a.u.). This change results in a qualitative difference in the Pauli deformation.

Two large molecules

For two helium atoms the Pauli deformation means a decrease of the electron density in the region between the nuclei and the corresponding increase of the density on the nuclei. This looks dangerous! What if instead of the two helium atoms we had two closed shell large molecules A and B that touch each other by their terminal parts? *Would the Pauli deformation be local, or would it extend over the whole system?* Maybe the distant parts of the molecules would deform as much as the contact regions?

The answer may be deduced from Eq. (E.4). The formula suggests that the electronic density change pertains to the whole system. When the formula was derived we concentrated on two helium atoms. However, *nothing* would change in the derivation if we had in mind a doubly occupied molecular orbital a that extends over the whole polymer A and a similar orbital b that extends over B, Fig. E.2. In such a case Eq. (E.4) would be identical. The formula says: the three deformation contributions cancel if we integrate them over the total space.[8] The first deformation means a density deficiency (minus sign), the two others mean density gains (plus sign). The first of these contributions is *for sure located close to the contact region* of A and B. The two others (of the same magnitude) have the spatial form such as a^2 and b^2 (i.e., extend over the whole polymer chains A and B!), but are scaled by the factor $2S^2/(1-S^2)$. Since the contributions cancel out in space (when integrated), *the density gain extends over the polymeric molecules and, therefore, is very small locally; the larger the interacting molecules, the smaller the local change*. The situation is therefore similar to an inflatable balloon pressed by your finger. We have a large deformation at the contact region corresponding to $-\frac{4S}{1-S^2}ab$, but in fact the *whole* balloon deforms. Just because of this deformation that has to extend over the whole balloon, the local deformation on the other side of the toy is extremely small. Therefore, by common sense we obtained a quantum mechanical explanation.[9]

[8] But of course at a given point they do not cancel in general.

[9] Good for both of them.

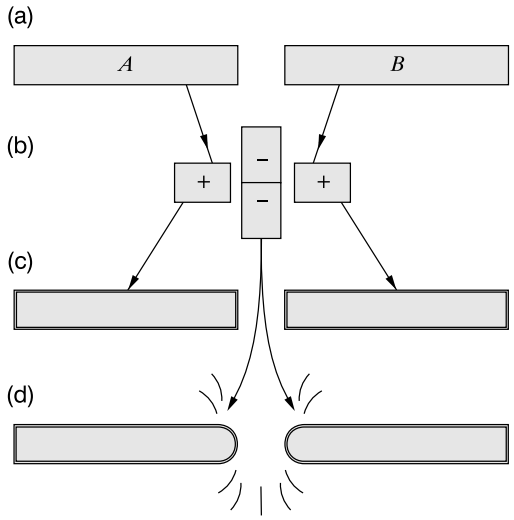


Fig. E.2. The locality of the Pauli deformation (scheme). (a) Two polymeric chains A and B (with the electronic densities in the form of the elongated rectangles corresponding to the isolated molecules A and B) approach one another. (b) The Pauli deformation consists of the two density gains (the rectangles with $+$) and a single electron loss (the rectangles with $-$). Let us assume that the surfaces of the rectangles are equal to the corresponding integrals of the charge distributions $-4S/(1-S^2)ab$ in the contact region, $2S^2/(1-S^2)a^2$ on the molecule A and $2S^2/(1-S^2)b^2$ on the polymer B – this is why the electron density loss has the rectangle twice as large as any of the electron density gains. (c) A partial Pauli deformation: the density gain $2S^2/(1-S^2)a^2$ for molecule A has been added to the initial density distribution, and similarly for molecule B (the rectangles became larger, but *locally the corresponding increase is small*). (d) In order to represent the total Pauli deformation from the result obtained in point c we subtracted the density distribution $4S/(1-S^2)ab$, located in the contact region. As a result *the Pauli deformation, when viewed locally, is large only in the contact region*.

This means that the Pauli deformation has a *local character*: it takes place almost exclusively in the region of the contact of both molecules.

APPENDIX F

Hydrogen Atom in Electric Field – Variational Approach

Polarization of an atom or molecule can be calculated by using the finite field method described on p. 284. Let us apply this method to the hydrogen atom. Its polarizability was already calculated by using a simple version of the perturbation theory (p. 282). This time we will use the variational method.

The Hamiltonian for the isolated hydrogen atom (within the Born–Oppenheimer approximation) reads as

$$\hat{H}^{(0)} = -\frac{1}{2}\Delta_e - \frac{1}{r},$$

where the first term is the electronic kinetic energy operator, and the second is its Coulomb interaction energy with the nucleus (proton–electron distance is denoted by r). The atom is in the uniform electric field $\boldsymbol{\mathcal{E}} = (0, 0, \mathcal{E})$ with $\mathcal{E} > 0$ and, similarly as in perturbation theory (p. 282), the total Hamiltonian has the form

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

with $V = z\mathcal{E}$, where z denotes the coordinate of the electron and the proton is in the origin (the derivation of the formula is given on p. 282, the exchange of z to x does not matter).

The variational wave function ψ is proposed in the form

$$\psi = \chi_1 + c\chi_2, \tag{F.1}$$

where $\chi_1 = \frac{1}{\sqrt{\pi}}\exp(-r)$ is the $1s$ orbital of the hydrogen atom (ground state) and χ_2 is the normalized[1] p-type orbital

$$\chi_2 = Nz\exp(-\zeta r).$$

There are two variational parameters, c and ζ. Let us assume for a while that we have fixed the value of ζ, so the only variational parameter is c. The wave function ψ is a linear combination of

[1] N can be easily calculated from the normalization condition $1 = N^2\int[z\exp(-\zeta r)]^2 dV = N^2\int_0^\infty dr r^4\exp(-2\zeta r)\int_0^\pi d\theta\sin\theta\cos^2\theta\int_0^{2\pi}d\phi = N^2 4!\,(2\zeta)^{-5}\frac{2}{3}2\pi = N^2\frac{\pi}{\zeta^5}$. This gives $N = \sqrt{\frac{\zeta^5}{\pi}}$.

two expansion functions ("two-state model"), χ_1 and χ_2. Therefore, the optimal energy follows from the Ritz method according to case III of Appendix V1-D on p. V1-655, i.e.,

$$E = E_{ar} \pm \sqrt{\Delta^2 + h^2}, \tag{F.2}$$

where the arithmetic mean energy $E_{ar} \equiv \frac{H_{11}+H_{22}}{2}$, while $\Delta \equiv \frac{H_{11}-H_{22}}{2}$ and $h \equiv H_{12} = H_{21}$ with

$$H_{ij} \equiv \left\langle \chi_i | \hat{H} \chi_j \right\rangle = \left\langle \chi_i | \hat{H}^{(0)} \chi_j \right\rangle + \langle \chi_i | V \chi_j \rangle .$$

Let us compute all the ingredients of the energy given by (F.2).

First, let us note that $H_{11} \equiv \left\langle \chi_1 | \hat{H}^{(0)} \chi_1 \right\rangle = -\frac{1}{2}$ a.u., since χ_1 is the ground state of the isolated hydrogen atom (p. V1-232), and $\langle \chi_1 | V \chi_1 \rangle = 0$, because the integrand is antisymmetric with respect to $z \to -z$.

Now, let us compute $H_{22} = H_{22}^{(0)} + V_{22}$. Note that $V_{22} = 0$ for the same reason as V_{11}. We have

$$H_{22}^{(0)} = -\frac{1}{2} \langle \chi_2 | \Delta_e \chi_2 \rangle - \left\langle \chi_2 | \frac{1}{r} \chi_2 \right\rangle .$$

The second integral is $\left\langle \chi_2 | \frac{1}{r} \chi_2 \right\rangle = N^2 \int_0^\infty dr r^3 \exp(-2\zeta r) \int_0^\pi d\theta \sin\theta \cos^2\theta \int_0^{2\pi} d\phi = \frac{\zeta^5}{\pi} \cdot 3!\,(2\zeta)^{-4} \cdot \frac{2}{3} \cdot 2\pi = \frac{1}{2}\zeta$, where the dots separate the values of the corresponding integrals[2]. In Appendix V1-S the reader will find the main ingredients needed to compute the first integral of $H_{22}^{(0)}$:

$$\langle \chi_2 | \Delta_e \chi_2 \rangle = N^2 \left\langle r\cos\theta \exp(-\zeta r) | \begin{bmatrix} \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \sin\theta \frac{\partial}{\partial \theta} + \\ \frac{1}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \phi^2} \end{bmatrix} r\cos\theta \exp(-\zeta r) \right\rangle =$$

$$N^2 \begin{bmatrix} \left\langle r\cos\theta \exp(-\zeta r) | \cos\theta \frac{1}{r^2} \frac{\partial}{\partial r} [r^2 \exp(-\zeta r) - \zeta r^3 \exp(-\zeta r)] \right\rangle + \\ \left\langle r\cos\theta \exp(-\zeta r) | \frac{(-2\cos\theta)}{r^2} r \exp(-\zeta r) \right\rangle + 0 \end{bmatrix} =$$

$$N^2 \begin{bmatrix} \left\langle r\cos\theta \exp(-\zeta r) | \cos\theta [\frac{2}{r} - \zeta - 3\zeta + \zeta^2 r] \exp(-\zeta r) \right\rangle + \\ \left\langle r\cos\theta \exp(-\zeta r) | \frac{(-2\cos\theta)}{r} \exp(-\zeta r) \right\rangle \end{bmatrix} =$$

$$\frac{\zeta^5}{\pi} (\frac{2}{3} \cdot 2\pi) \left[2 \cdot 2 \cdot (2\zeta)^{-3} - 4\zeta \cdot 3! \cdot (2\zeta)^{-4} + \zeta^2 \cdot 4! \cdot (2\zeta)^{-5} - 2 \cdot 2! \cdot (2\zeta)^{-3} \right] = -\zeta^2 .$$

[2] Note that in the spherical coordinates the volume element $dV = r^2 \sin\theta dr d\theta d\phi$. In derivations of this appendix (and not only) we often use the equality $\int_0^\infty dx x^n \exp(-\alpha r) = n! \alpha^{-(n+1)}$.

Thus, we obtain $H_{22} = \frac{1}{2}\zeta^2 - \frac{1}{2}\zeta$. This formula looks good, since for $\chi_2 = 2p_z$, i.e., for $\zeta = \frac{1}{2}$, we get correctly (see p. V1-232) $H_{22} = E_{2p} = -\frac{1}{8}$ a.u., the energy of orbital $2p$.

Let us turn to the nondiagonal matrix element of the Hamiltonian, $H_{12} = H_{12}^{(0)} + V_{12}$. Note that $H_{12}^{(0)} = 0$, because χ_1 is an eigenfunction of $\hat{H}^{(0)}$ and $\langle \chi_1 | \chi_2 \rangle = 0$. Thus,

$$h = N\mathcal{E}\left\langle r\cos\theta \exp(-\zeta r) | r\cos\theta \frac{1}{\sqrt{\pi}} \exp(-r) \right\rangle =$$
$$N\mathcal{E}\frac{1}{\sqrt{\pi}} \int_0^\infty dr r^4 \exp[-(\zeta+1)r] \int_0^\pi d\theta \sin\theta \cos^2\theta \int_0^{2\pi} d\phi =$$
$$\mathcal{E}\frac{\sqrt{\zeta^5}}{\pi} \cdot 4!(\zeta+1)^{-5} \cdot \frac{2}{3} \cdot 2\pi = 32\frac{\sqrt{\zeta^5}}{(\zeta+1)^5}\mathcal{E}.$$

Now we can write Eq. (F.2) as a function of ζ, i.e.,

$$E = \frac{1}{4}(\zeta^2 - \zeta - 1) - \sqrt{\frac{1}{16}(\zeta^2 - \zeta + 1)^2 + \zeta^5 \left(\frac{2}{\zeta+1}\right)^{10} \mathcal{E}^2}. \tag{F.3}$$

We would like to expand this expression in power series of $\mathcal{E}$ in order to highlight the coefficient at $\mathcal{E}^2$, because this coefficient is related to the polarizability. The expansion gives (in a.u.)

$$E \approx \frac{1}{4}(\zeta^2 - \zeta - 1) - \frac{1}{4}(\zeta^2 - \zeta + 1) - \frac{1}{2}\alpha_{zz}\mathcal{E}^2 + \ldots = -\frac{1}{2} - \frac{1}{2}\alpha_{zz}\mathcal{E}^2 + \ldots,$$

where according to Eq. (4.24) the polarizability (in a.u.) reads as

$$\alpha_{zz} = 4 \cdot \frac{\zeta^5}{|\zeta^2 - \zeta + 1|} \left(\frac{2}{\zeta+1}\right)^{10}. \tag{F.4}$$

Several numerical values of α_{zz} computed by using (F.3) and (F.4) are given on p. 284. They are compared with the exact result $\alpha_{zz} = 4.5$ a.u.

APPENDIX G

Multipole Expansion

What is the multipole expansion for?

In the perturbational theory of intermolecular interactions (Chapter 5) the perturbation operator (V) plays an important role. The operator contains all the Coulombic charge–charge interactions, where one of the point charges belongs to subsystem A, the other to B. Therefore, according to the assumption behind the perturbational approach (large intermolecular distance) there is a *guarantee* that both charges are distant in space. For example, for two interacting hydrogen atoms (electron 1 at nucleus a, electron 2 at nucleus b, a.u. are used)

$$V = -\frac{1}{r_{a2}} + \frac{1}{r_{12}} - \frac{1}{r_{b1}} + \frac{1}{R}, \tag{G.1}$$

where R stands for the internuclear distance. A short inspection convinces us that the mean value of the operator $-\frac{1}{r_{a2}} + \frac{1}{r_{12}}$, with the wave function[1] $\psi_{A,n_1}(1)\psi_{B,n_2}(2)$, would give something close to zero, because both distances in the denominators are almost equal to each other (Fig. G.1a). The same can be said about the two other terms of V. This is why the situation is similar (see Chapter 5) to weighing the captain's hat, criticized so harshly by us in the supermolecular approach to the supermolecular forces (see Fig. 5.4).

What could we do to not lose accuracy? This is precisely the goal of the multipole expansion for each of the operators $\frac{1}{r_{ij}}$.

Coordinate system

What is the multipole expansion really? We will explain this in a moment. Let us begin quietly by introduction of two Cartesian coordinate systems: one on molecule A, the other on molecule B (Fig. G.1).

[1] $\psi_{A,n_1}(1)$ means an excited state (n_1 is the corresponding quantum number) of atom A, $\psi_{B,n_2}(2)$ similarly for atom B. Note that electron 1 is always close to nucleus a, and electron 2 is close to nucleus b, while A and B are far away.

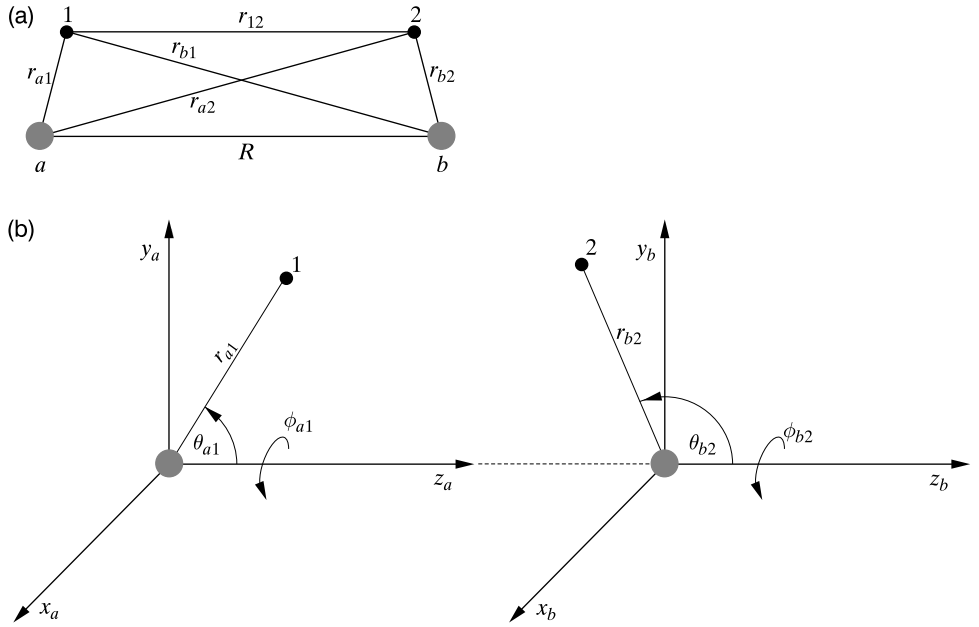


Fig. G.1. The coordinate system used in the multipole expansion. (a) The relevant distances. The large dots denote the origins of the two Cartesian coordinate systems, labeled by a and b, respectively. One assumes that particle 1 resides always close to a, particle 2 resides always close to b. The figure gives the notation related to the distances considered. (b) Two Cartesian coordinate systems (and their polar counterparts): one associated with the center a, the other with the center b (the x and y axes are parallel in both systems, the z axes are colinear). For any point in three-dimensional space $x_a = x_b$, $y_a = y_b$, $z_a = z_b + R$. Note that the two coordinate systems are not on the same footing: the z axis of a points towards b, while the z axis of system b does not point to a. Sometimes in the literature one introduces an alternative coordinate system with the "equal footing" by changing $z_b \to -z_b$ (then the two coordinate systems point to each other), but this leads to different "handedness" ("right-" or "left-handed") of the systems and subsequently to complications for chiral molecules. Let us stick to the "nonequivalent choice."

This can be done in several ways. Let us begin by choosing the origins of the coordinate systems. Where to choose them? Is it irrelevant? It turns out that the choice is important. Let us stop the problem here and come back to it later on. Just for clarity, let me communicate the conclusion: the origins should be chosen in the neighborhood of the centers of mass (charges) of the interacting molecules. Let us introduce the axes by taking the z axes (z_a and z_b) colinear pointing in the same direction, and axes x_a and x_b as well as y_a and y_b pairwise parallel.

Multipole series and the multipole operators of a particle

With such a coordinate system, the Coulomb interaction of particles 1 and 2 (with charges q_1 and q_2) can be expanded using the following approximation,[2] Fig. G.1b:

$$\frac{q_1 q_2}{r_{12}} \cong \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-s}^{m=+s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_a^{(k,m)}(1)^* \hat{M}_b^{(l,m)}(2), \tag{G.2}$$

where the coefficient

$$A_{kl|m|} = (-1)^{l+m} \frac{(k+l)!}{(k+|m|)!(l+|m|)!}, \tag{G.3}$$

whereas

MULTIPOLE MOMENT OPERATORS

$\hat{M}_a^{(k,m)}(1)$ and $\hat{M}_b^{(l,m)}(2)$ represent the m-th components of the 2^k-pole and 2^l-pole of particle 1 in the coordinate system on a and of particle 2 in the coordinate system on b, respectively, i.e.,

$$\hat{M}_a^{(k,m)}(1) = q_1 r_{a1}^k P_k^{|m|}(\cos\theta_{a1}) \exp(\mathrm{i}m\phi_{a1}), \tag{G.4}$$

$$\hat{M}_b^{(l,m)}(2) = q_2 r_{b2}^l P_l^{|m|}(\cos\theta_{b2}) \exp(\mathrm{i}m\phi_{b2}), \tag{G.5}$$

where r, θ, ϕ stand for the spherical coordinates of a particle (in the coordinate system a or b, Fig. G.1), the associated Legendre polynomials $P_k^{|m|}$ with $|m| \le k$ are defined as (cf. p. V1-229)

$$P_k^{|m|}(x) = \frac{1}{2^k k!}(1-x^2)^{|m|/2} \frac{\mathrm{d}^{k+|m|}}{\mathrm{d}x^{k+|m|}}(x^2-1)^k, \tag{G.6}$$

n_k and n_l in principle have to be equal to ∞, but in practice one takes finite integer values, and s is the lower of the summation indices k, l. That is it!

Maybe an additional remark concerning the nomenclature would be in place. Any multipole may be called a 2^k-pole (however strange this name looks), because this "multi" means the number 2^k. If we know how to make powers of two, and in addition got some contact with the world of ancient Greeks and Romans, then we will know how to compose the names of the successive multipoles: $2^0 = 1$, hence *monopole*; $2^1 = 2$, hence *dipole*; $2^2 = 4$, hence,

[2] It represents an approximation because it is not valid for $R < |\mathbf{r}_1 - \mathbf{r}_2|$, and this may happen in real systems (the electron clouds extend to infinity), also because n_k, n_l are finite instead of ∞.

quadrupole; etc. The names, however, are of no importance. Important are the formulae for the multipoles.

Multipole moment operators for many particles

A while before, a definition of the multipole moments of a single point-like charged particle was introduced. However, the multipole moments will be calculated in the future practically always for a molecule.

THE TOTAL MULTIPOLE MOMENT OPERATOR

The total multipole moment operator for molecule A represents a sum of the same operators for the individual particles (of course, all them have to be computed in the same coordinate system), i.e., $\hat{M}_a^{(k,m)}(A) = \sum_{i \in A} \hat{M}_a^{(k,m)}(i)$.

The first thing we have to stress about multipole moments is that in principle they depend on the choice of the coordinate system (Fig. G.2).

This will be seen in a while when inspecting the formulae for the multipole moments.

Examples

Let us take a few examples for particle 1 in the coordinate system a (for the sake of simplicity we skip the indices). The case with $k = 0$ is obviously the simplest one, and we should always begin by the simplest things. If $k = 0$, then (because of $P_k^{|m|}$) $m = 0$, and the monopole has therefore a single component $M^{(00)}$, i.e.,

$$\hat{M}^{(0,0)} = qr^0 P_0^0(\cos\theta)\exp(i0\phi) = q. \tag{G.7}$$

Hence,

MONOPOLE

the monopole for a particle means simply its charge.

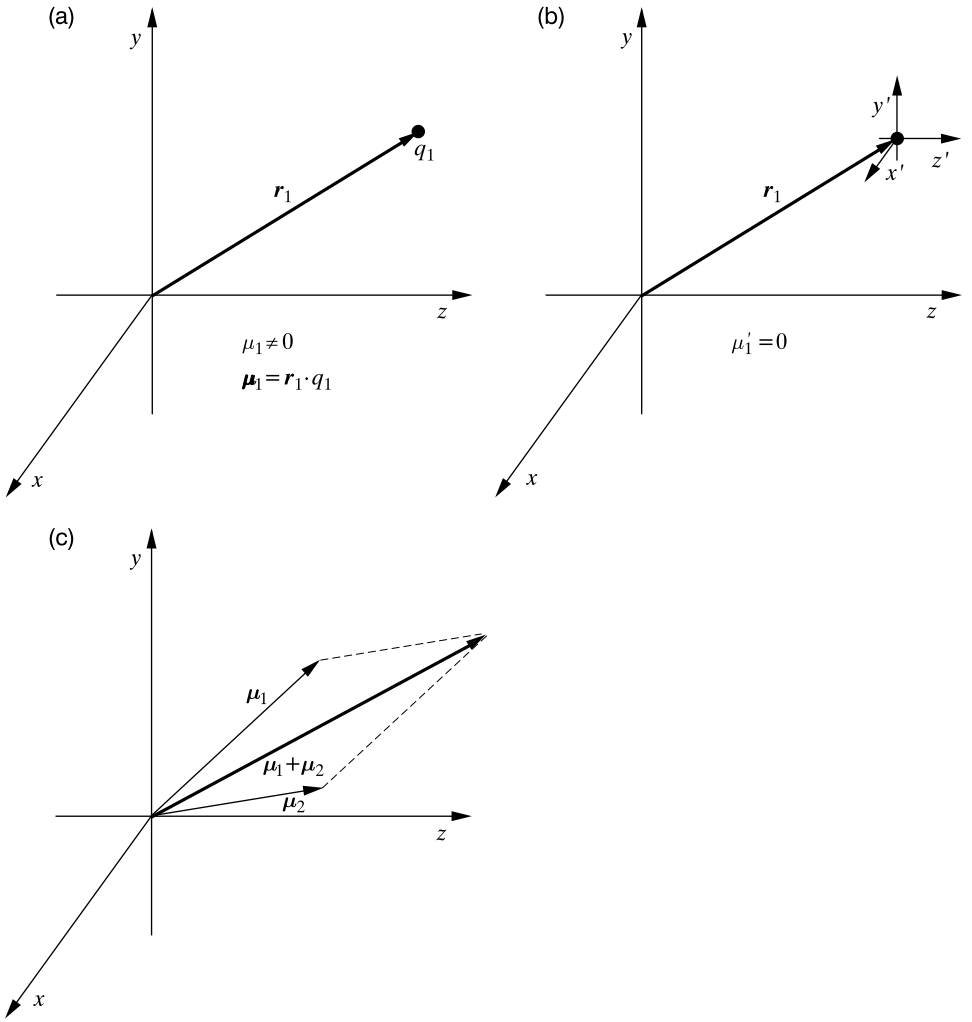


Fig. G.2. The multipole moments (or simply multipoles) in general depend on the choice of the coordinate system. (a) The dipole moment of a point-like particle with charge q_1 is equal to $\boldsymbol{\mu}_1$. (b) The dipole moment of the same particle in the coordinate system with the origin on the particle. In such a case we obtain $\boldsymbol{\mu}_1' = 0$. (c) The dipole moment of two particles represents a sum of the dipole moments of the individual particles (in a common coordinate system).

Let us go to $k = 1$, i.e., to the dipole moment. Since $m = -1, 0, +1$, the dipole moment has three components. First, let us consider $\hat{M}^{(1,0)}$. We have

$$\hat{M}^{(1,0)} = qr^1 P_1^0(\cos\theta)\exp(\mathrm{i}0\phi) = qr\cos\theta = qz. \tag{G.8}$$

DIPOLE MOMENT OPERATOR

Thus, the z component of the dipole moment operator of a single particle is equal to qz. The other components are

$$\begin{aligned} M^{(1,1)} &= qr^1 P_1^1(\cos\theta)\exp(\mathrm{i}\phi) = qr\sin\theta(\cos\phi + \mathrm{i}\sin\phi) \\ &= q(x + \mathrm{i}y), \\ M^{(1,-1)} &= qr^1 P_1^1(\cos\theta)\exp(-\mathrm{i}\phi) = qr\sin\theta(\cos\phi - \mathrm{i}\sin\phi) \\ &= q(x - \mathrm{i}y). \end{aligned}$$

After a careful (but a little boring) derivation we arrive at the following table of multipoles (up to the octupole). Just to make the table simpler, every multipole moment of the particle has been divided by q.

k	m			
	0	**±1**	**±2**	**±3**
0 charge	1	—	—	—
1 dipole	z	$x + \mathrm{i}y$ $x - \mathrm{i}y$	—	—
2 quadrupole	$\frac{1}{2}(3z^2 - r^2)$	$3z(x + \mathrm{i}y)$ $3z(x - \mathrm{i}y)$	$3(x + \mathrm{i}y)^2$ $3(x - \mathrm{i}y)^2$	— —
3 octupole	$\frac{1}{2}(5z^3 - 3zr^2)$	$\frac{3}{2}(x + \mathrm{i}y)(5z^2 - r^2)$ $\frac{3}{2}(x - \mathrm{i}y)(5z^2 - r^2)$	$15z(x + \mathrm{i}y)^2$ $15z(x - \mathrm{i}y)^2$	$15(x + \mathrm{i}y)^3$ $15(x - \mathrm{i}y)^3$

Thus, the operator of the 2^k-pole moment of a charged particle represents simply a k-th-degree polynomial of x, y, z.

The multipoles depend on the coordinate system chosen

Evidently any multipole moment value (except monopole) depends on my caprice (Fig. G.2), because I am free to choose any coordinate system I want and, e.g., the z coordinate of the particle in such a system will depend on me! It turns out that if we calculate the multipole moments, then

the lowest nonvanishing multipole moment does not depend on the coordinate system translation; the other moments in general do depend on it.

This is not peculiar for the moments defined by Eqs. (G.4) or (G.5), but represents a property of every term of the form $x^n y^l z^m$. Indeed, $k = n + l + m$ tells us that we have to do with a 2^k-pole. Let us shift the origin of the coordinate system by the vector $\mathbf{L}$. Then the $x^n y^l z^m$ moment computed in the new coordinate system, i.e., $x'^n y'^l z'^m$, is equal to

$$\left(x'\right)^n \left(y'\right)^l \left(z'\right)^m = (x + L_x)^n (y + L_y)^l (z + L_z)^m = x^n y^l z^m + \text{ a linear combination of } \textit{lower multipole moments}. \tag{G.9}$$

If, for some reason, all the lower moments were equal to zero, then this would mean the invariance of the moment with respect to the choice of the coordinate system.

Let us take, e.g., the system $ZnCl^+$. In the first approximation the system may be approximated by *two point-like charges* Zn^{2+} and Cl^-. Let us locate these charges on the z axis in such a way that Zn^{2+} has the coordinate $z = 0$ and Cl^- has the coordinate $z = 5$. Now, we would like to compute the z component of the dipole moment,[3] i.e., $M^{(1,0)} = \mu_z = q_1 z_1 + q_2 z_2 = (+2)0 + (-1)5 = -5$. What if we had chosen another coordinate system? Let us check what we would see if the origin of the coordinate system were shifted towards positive z by 10 units. In such a case the ions have the coordinates $z'_1 = -10$ and $z'_2 = -5$, and as the z component of the dipole moment we obtain

$$M^{(1,0)'} = \mu'_z = q_1 z'_1 + q_2 z'_2 = (+2)(-10) + (-1)(-5) = -15. \tag{G.10}$$

Thus, in this case the dipole moment depends on the choice of the coordinate system. However, the monopole of the system is equal to $(+2) + (-1) = +1$ and this number will not change at any shift of the coordinate system. Therefore,

the dipole moment of a molecular ion depends on us through the arbitrary choice of the coordinate system.

[3] Since we have to do with point charges, the computation of the multipole moments reduces just to inserting into the multipole operator the values of the coordinates of the corresponding charges.

Interaction energy of nonpoint-like multipoles

In chemical reasoning about intermolecular interactions the multipole–multipole (mainly dipole–dipole, like for interactions in, e.g., water) interaction plays an important role. The dipolar molecules have nonzero dimensions and therefore they represent something else than just point-like dipoles. Let us clarify that by taking a simple example of two dipolar systems located on the z axis (Fig. G.3). System a consists of the two charges $+1$ at $z = 0$ and -1 at $z = 1$, while system b also has two charges $+1$ with $z = 10$ and -1 with $z = 11$,

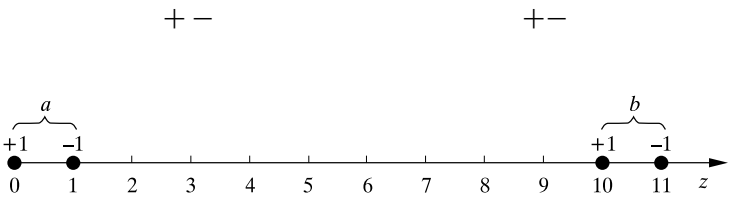


Fig. G.3. Interaction of the *nonpoint-like* dipoles contains also interactions of higher point-like multipoles.

A first idea is that we have to do with the interaction of two dipoles and that is it. Let us check whether everything is OK. The checking is very easy, because what really interacts are the charges, not dipoles whatsoever. Thus, the exact interaction of systems a and b is $(+1)(+1)/10 + (+1)(-1)/11 + (-1)(+1)/9 + (-1)(-1)/10 = 2/10 - 1/11 - 1/9 = -0.0020202$. What would give the dipole–dipole interaction? Such a task immediately poses the question how such an interaction is to be calculated.

> The first advantage of the multipole expansion is that it produces the formulae for the multipole–multipole interactions.

We have the dipole–dipole term in the form $R^{-3}(\mu_{ax}\mu_{bx} + \mu_{ay}\mu_{by} - 2\mu_{az}\mu_{bz}) = -2R^{-3}\mu_{az}\mu_{bz}$, because the x and y components of our dipole moments are equal to zero. Since A and B are neutral, it is absolutely irrelevant which coordinate system is chosen to compute the dipole moment components. Therefore, let us use the global coordinate system, in which the positions of the charges have been specified. Thus, $\mu_{az} = (+1) \cdot 0 + (-1) \cdot 1 = -1$ and $\mu_{bz} = (+1) \cdot 10 + (-1) \cdot 11 = -1$.

What is R?

Now, we are confronting a serious problem (which we *always* encounter in the multipole expansion); what is R? We are forced to choose the two local coordinate systems in A and B.

We *arbitrarily* decide here to locate these origins in the middle of each dipolar system, and therefore $R = 10$. It looks like a reasonable choice, and as will be shown later on, it really is. We are all set to compute the dipole–dipole interaction: $-2 \cdot 10^{-3}(-1)(-1) = -0.0020000$. Close! The computed exact interaction energy is -0.0020202. Where is the rest? Is there any error in our dipole–dipole interaction formula? We simply forgot that our dipolar systems not only represent the dipole moments, but also have nonzero octupole moments (the quadrupoles are equal to zero) and nonzero higher odd-order multipoles, and we did not take them into account. If somebody computed all the interactions of such multipoles, then we would recover the correct interaction energy with any desired accuracy. How come, however, that such a simple dipolar system has also a nonzero octupole moment? The answer is simple: this is because the dipole is not point-like. The conclusion from this story is that the reader has to pay attention whether we have to do with point-like or nonpoint-like multipole moments.

Just to have a small training, the table below reports which multipole moments are zero and which are nonzero for a few simple chemical systems. Everything shown follows from the symmetry of their nuclear framework in the electronic ground state.

	Li^+	HCl	H_2	CH_4	HCl^+
monopole					
$k = 0$	q	0	0	0	q
dipole					
$k = 1$	0	μ	0	0	μ
quadrupole					
$k = 2$	0	Q	Q	0	Q
octupole					
$k = 3$	0	Oct	0	Oct	Oct

Properties of the multipole expansion

In practice of the multipole expansion at least three simple questions arise:

a) How to truncate the expansion, i.e., how to choose the values of n_k and n_l?
b) Since the multipole moments depend in general on the coordinate system chosen, what sort of miracle make the multipole expansion *of the energy* independent of the coordinate system?
c) When does the multipole expansion make sense, i.e., when does it converge?

Truncating the multipole expansion and its coordinate system dependence

It turns out that questions a and b are closely related to each other. When n_k and n_l are finite and nonzero,[4] however horrifying it might be, the result of the multipole expansion is in general coordinate-dependent! If, however, n_k and n_l satisfy $n_k + n_l = \text{const}$, then we may shift both coordinate systems (the same translation for both) however we like and the interaction energy computed stays invariant.[5] Such a recipe for n_k and n_l corresponds to taking all the terms with a given power of R^{-1}.

> In other words, if we take all the terms with a given R^{-m} dependence, then the result does not depend on the same translations of both coordinate systems.

This means that in order to maintain the invariance of the energy with respect to equal translations of both coordinate systems we have to compute in the multipole expansion all terms satisfying $n_k + n_l = n_{\max}$. If, e.g., $n_{\max} = 2$, then we have to compute the term proportional to R^{-1} or charge–charge interaction (it will be invariant), proportional to R^{-2} or charge–dipole and dipole–charge terms (their sum is also invariant), and proportional to R^{-3} or charge–quadrupole, quadrupole–charge, and dipole–dipole (their sum is invariant as well).

Imagine a scientist calculating the interaction energy of two molecules. As will be shown later, in his multipole expansion he will have the charges of both interacting *molecules*, their dipole moments, their quadrupole moments, etc. Our scientist is a systematic fellow, therefore I expect he will begin by calculating the multipole moments for each molecule, up to a certain maximum multipole moment (say, the quadrupole; the calculations become more and more involved, which makes his decision easier). Then, he will be ready to compute all the individual multipole–multipole interaction contributions. He will make a table of such interactions (in the rows the multipole moments of A, in the columns the multipole moments of B) and compute all the entries in his table. Then, many of his colleagues would sum up *all* the entries of the table in order not to waste their effort. This will be an error. The scientists might not suspect that due to this procedure their result depends on the choice of the coordinate system, which is always embarrassing. However, our fellow will do something else. He will sum up the entries corresponding to charge–charge, charge–dipole, dipole–charge, charge–quadrupole, quadrupole–charge, and dipole–dipole, and he will throw the other computed entries in the garbage can. Taking this decision the scientist will gain a lot: his interaction energy will not depend on how he translated the a and b coordinate systems.

[4] Zero would introduce large errors in most applications.

[5] L.Z. Stolarczyk, L. Piela, *Int. J. Quantum Chem.*, 15(1979)701.

Now, we will illustrate this by simple formulae and see how it works in practice. We have said before that the complete set of terms with a given dependence on R^{-1} has to be taken. Otherwise, horrible things happen. Let us take such a complete set of terms with $k+l=2$. We will see how nicely they behave upon translation of the coordinate system, and how nasty the behavior of the individual terms is. Let us begin by the charge–dipole term. The term in the multipole expansion corresponds to $k=0$ and $l=2$, i.e.,

$$(-1)^2\frac{2!}{2!R^3}\hat{M}^{(00)}(1)^*\hat{M}^{(20)}(2)=q_1q_2R^{-3}\frac{1}{2}(3z_2^2-r_2^2).$$

The next term ($k=1, l=1$) has three contributions coming from the summation over m:

$$(-1)\frac{2!}{1!1!R^3}\hat{M}^{(10)}(1)^*\hat{M}^{(10)}(2)+(-1)^2\frac{2!}{2!2!R^3}\hat{M}^{(11)}(1)^*\hat{M}^{(11)}(2)$$

$$+(-1)^0\frac{2!}{2!2!R^3}\hat{M}^{(1-1)}(1)^*\hat{M}^{(1-1)}(2)=q_1q_2R^{-3}[(x_1x_2+y_1y_2)-2z_1z_2].$$

For the third term ($k=2, l=0$) we have

$$(-1)^2\frac{2!}{2!R^3}\hat{M}^{(20)}(1)^*\hat{M}^{(00)}(2)=q_1q_2R^{-3}\frac{1}{2}(3z_1^2-r_1^2).$$

Note that each of the computed terms separately depends on the translation along the z axis of the origins of the interacting objects. Indeed, by taking $z+T$ instead of z we obtain the following. For the first term

$$\begin{aligned}&q_1q_2R^{-3}[\frac{1}{2}(3(z_2+T)^2-x_2^2-y_2^2-(z_2+T)^2)\\&=q_1q_2R^{-3}[\frac{1}{2}(3z_2^2-r_2^2)+\frac{1}{2}(6Tz_2+3T^2-2Tz_2-T^2)],\end{aligned}$$

for the second term

$$\begin{aligned}&q_1q_2R^{-3}[(x_1x_2+y_1y_2)-2(z_1+T)(z_2+T)]\\&=q_1q_2R^{-3}[(x_1x_2+y_1y_2)-2z_1z_2]+R^{-3}[-2Tz_1-2Tz_2-2T^2],\end{aligned}$$

and for the third term

$$\begin{aligned}&q_1q_2R^{-3}\frac{1}{2}(3(z_1+T)^2-x_1^2-y_1^2-(z_1+T)^2)\\&=q_1q_2R^{-3}[1/2(3z_1^2-r_1^2)+\frac{1}{2}(6Tz_1+3T^2-2Tz_1-T^2)].\end{aligned}$$

If somebody still had illusions that the coordinate system dependence is negligible, then this is the right time to change opinion. Evidently, each term depends on what we chose as T, and T can be *anything*! If I were really malicious, then I would get monstrous dependence on T.

Now, let us add all the individual terms together to form the complete set for $k + l = 2$. We obtain

$$\begin{aligned} & q_1 q_2 \{R^{-3}[\frac{1}{2}(3z_2 - r_2^2) + (2Tz_2 + T^2)] + R^{-3}[(x_1x_2 + y_1y_2) - 2z_1z_2] \\ & + R^{-3}[-2Tz_1 - 2Tz_2 - 2T^2] + R^{-3}[1/2(3z_1 - r_1^2) + (2Tz_1 + T^2)]\} \\ & = q_1 q_2 R^{-3}\{1/2(3z_2 - r_2^2) + [(x_1x_2 + y_1y_2) - 2z_1z_2] + 1/2(3z_1 - r_1^2)\}. \end{aligned}$$

The dependence on T disappeared as if touched by a magic wand.[6] *The complete set does not depend on T*! This is what we wanted to show.

Convergence of the multipole expansion

I owe the reader an explanation about the convergence of the multipole expansion (point c, Fig. G.4). Well,

> we may demonstrate that the multipole expansion convergence depends on how the molecules are located in space with respect to one another. The convergence criterion reads
>
> $$|\mathbf{r}_2 - \mathbf{r}_1| < |\mathbf{R}|, \tag{G.11}$$
>
> where $\mathbf{r}_1$ denotes the vector pointing at particle 1 from its coordinate system origin, and similarly for vector $\mathbf{r}_2$.

The reader will be easily convinced if he/she draws two spheres that are tangent (this is the most dangerous situation) and then considers possible $\mathbf{r}_1$ and $\mathbf{r}_2$ vectors. Whatever $\mathbf{r}_1$ and $\mathbf{r}_2$ vectors are, our criterion will be fulfilled. The criterion is, however, even more general than allow two non-overlapping spheres. It is easy to find such locations of two particles that are outside the spheres and yet the convergence criterion is fulfilled. For example, let us take two tangent spheres with radii ρ_1 and ρ_2 (their centers are on the x axis) as well as the vectors $\mathbf{r}_1 = (0, \rho_1, 0)$ and $\mathbf{r}_2 = (0, u, 0)$, where $u = \rho_1 + R/10$ and $u > \rho_2$. Then $|\mathbf{r}_2 - \mathbf{r}_1| = R/10 < R$, i.e., the convergence criterion is satisfied, despite the fact that particle 2 is outside of its sphere.

[6] We may prove that equal but arbitrary rotations of both coordinate systems about the z axis lead to a similar invariance of the interaction energy.

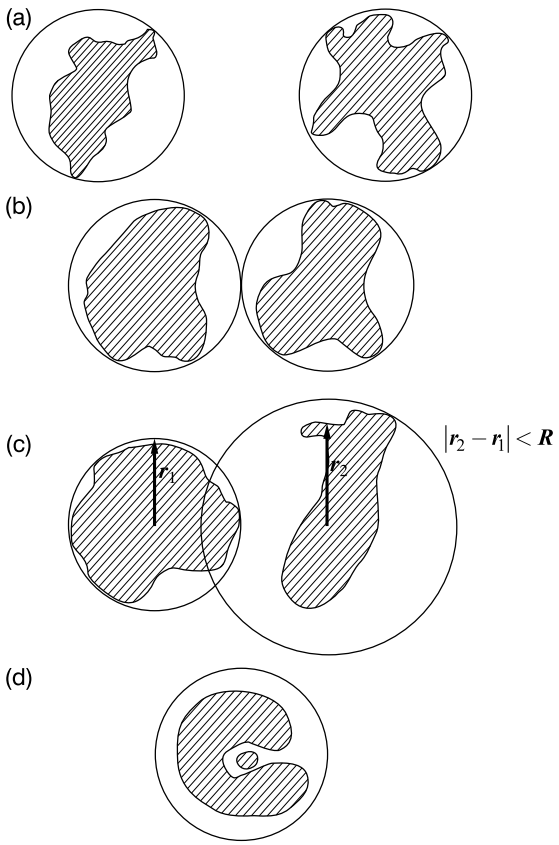


Fig. G.4. Convergence of the multipole expansion for the interacting charge density distributions shown as irregular shapes. In cases (a) and (b) the expansion converges, while in cases (c) and (d) it diverges, despite the fact that for certain positions of particles – those shown by the arrows – the expansion still converges. If the shapes represented molecular electron densities, they could not be enclosed in finite spheres in any of the above cases. This is the reason of the multipole expansion divergence for molecules, still very useful in calculations because of the "asymptotical convergence."

For our purposes it is sufficient to remember that

> when the two particles are in their nonoverlapping spheres, the multipole expansion converges.

Can we make such an assumption? Our goal is application of the multipole expansion in the case of intermolecular interactions. Are we able to enclose both molecules in two nonoverlapping spheres? Sometimes for sure not, e.g., if a small molecule A is to be docked in a cavity

of a large molecule B. This is a very interesting case (Fig. G.4d), but what we have most often in quantum chemistry are two *distant* molecules. Is then everything all right? Apparently the molecules can be enclosed in the spheres, but if we recall that the electronic density extends to infinity (although it decays very fast), then we feel a little scared. *Almost* the whole density distribution could be enclosed inside such spheres, but outside the spheres there is also something. It turns out this very fact causes that

> the multipole expansion for the interaction energy of such diffused charge distributions diverges, i.e., if we go to very high terms we will get infinity.

However strange it could look, in mathematics we are able to extract very useful information also from divergent series if they converge *asymptotically* (see p. V1-284). This is precisely the situation with the multipole expansion applied to the diffuse charge distributions like molecules have. This is why the multipole expansion is useful.[7] It has also the important advantage of being physically appealing, because thanks to it we may interpret the interaction energy in terms of properties of the individual interacting molecules (their charges, dipole, quadrupole, etc., moments). All such terms represent a part of the well established chemical terminology.

[7] If the computation were feasible with a very high accuracy, then the multipole expansion might be of low importance.

APPENDIX H

NMR Shielding and Coupling Constants – Derivation

This section is for those who do not fully believe the author, and want to check whether the final formulae for the shielding and coupling constants in nuclear magnetic resonance (NMR) are indeed valid (Chapter 4).

Shielding constants

Let us begin from Eq. (4.88).

Applying vector identities

We are going to apply some vector identities[1] in the operators $\hat{B}_3$, $\hat{B}_4$, $\hat{B}_5$. The first identity is $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = \mathbf{v}\cdot(\mathbf{w}\times\mathbf{u}) = \mathbf{w}\cdot(\mathbf{u}\times\mathbf{v})$, which simply means three equivalent ways of calculating the volume of a parallelepiped. This identity applied to $\hat{B}_3$ and $\hat{B}_4$ gives

$$\hat{B}_3 = \frac{e}{mc}\sum_A\sum_j \gamma_A \frac{\mathbf{I}_A\cdot\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}, \tag{H.1}$$

$$\hat{B}_4 = \frac{e}{2mc}\sum_j \mathbf{H}\cdot\hat{\mathbf{L}}_{0j}. \tag{H.2}$$

Let us transform the term $\hat{B}_5$ by using the following identity: $(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{w}\times\mathbf{s}) = (\mathbf{u}\cdot\mathbf{w})(\mathbf{v}\cdot\mathbf{s}) - (\mathbf{v}\cdot\mathbf{w})(\mathbf{u}\cdot\mathbf{s})$. We obtain

$$\hat{B}_5 = \frac{e^2}{2mc^2}\sum_A\sum_j \gamma_A \left(\mathbf{H}\times\mathbf{r}_{0j}\right)\cdot\frac{\mathbf{I}_A\times\mathbf{r}_{Aj}}{r_{Aj}^3} =$$

$$\frac{e^2}{2mc^2}\sum_A\sum_j \gamma_A[\left(\mathbf{H}\cdot\mathbf{I}_A\right)(\mathbf{r}_{0j}\cdot\mathbf{r}_{Aj}) - \left(\mathbf{r}_{0j}\cdot\mathbf{I}_A\right)(\mathbf{H}\cdot\mathbf{r}_{Aj})]\cdot\frac{1}{r_{Aj}^3}.$$

[1] The reader may easily check each of them.

Putting things together

We are now all set to put all this baroque furniture into its destination, i.e., into Eq. (4.88) for ΔE,

$$\Delta E = \sum_A \Delta E_A, \tag{H.3}$$

where ΔE_A stands for the contribution of nucleus A, i.e.,

$$\begin{aligned}\Delta E_A = & -\gamma_A \left\langle \psi_0^{(0)} | \left(\mathbf{I}_A \cdot \mathbf{H}\right) \psi_0^{(0)} \right\rangle + \\ & + \frac{e^2}{2mc^2} \gamma_A \left\langle \psi_0^{(0)} | \sum_j \left[(\mathbf{H} \cdot \mathbf{I}_A)(\mathbf{r}_{0j} \cdot \mathbf{r}_{Aj}) - \left(\mathbf{r}_{0j} \cdot \mathbf{I}_A\right)(\mathbf{H} \cdot \mathbf{r}_{Aj}) \right] \cdot \frac{1}{r_{Aj}^3} \psi_0^{(0)} \right\rangle + \\ & + \frac{e^2}{2m^2c^2} \gamma_A \left[\left\langle \psi_0^{(0)} | \left(\sum_j \frac{\mathbf{I}_A \cdot \hat{\mathbf{L}}_{Aj}}{r_{Aj}^3} \right) \hat{R}_0 \left(\sum_j \mathbf{H} \cdot \hat{\mathbf{L}}_{0j} \right) \psi_0^{(0)} \right\rangle + \right. \\ & \left. \left\langle \psi_0^{(0)} | \left(\sum_j \mathbf{H} \cdot \hat{\mathbf{L}}_{0j} \right) \hat{R}_0 \left(\sum_j \frac{\mathbf{I}_A \cdot \hat{\mathbf{L}}_{Aj}}{r_{Aj}^3} \right) \psi_0^{(0)} \right\rangle \right].\end{aligned}$$

Averaging over rotations

The expression for ΔE_A represents a bilinear form with respect to the components of vectors $\mathbf{I}_A$ and $\mathbf{H}$,

$$\Delta E_A = \mathbf{I}_A^T \mathbf{C}_A \mathbf{H},$$

where $\mathbf{C}_A$ stands for a square matrix[2] of dimension 3 and $\mathbf{I}_A$ and $\mathbf{H}$ are vertical three-component vectors.

A contribution to the energy such as ΔE_A cannot depend on our choice of coordinate system axes x, y, z, i.e., on the components of $\mathbf{I}_A$ and $\mathbf{H}$. We will obtain the same energy if we rotate the axes (orthogonal transformation) in such a way as to diagonalize $\mathbf{C}_A$. The resulting diagonalized matrix $\mathbf{C}_{A,diag}$ has three eigenvalues (composing the diagonal) corresponding to the new axes x', y', z'. *The very essence of averaging is that none of these axes are to be privileged in any sense.* This is achieved by constructing the averaged matrix

$$\begin{aligned}&\frac{1}{3}\left[\left(\mathbf{C}_{A,diag}\right)_{x'x'} + \left(\mathbf{C}_{A,diag}\right)_{y'y'} + \left(\mathbf{C}_{A,diag}\right)_{z'z'} \right] \\ &= \left(\bar{\mathbf{C}}_{A,diag}\right)_{x'x'} = \left(\bar{\mathbf{C}}_{A,diag}\right)_{y'y'} = \left(\bar{\mathbf{C}}_{A,diag}\right)_{z'z'} \equiv C_A,\end{aligned}$$

[2] We could write its elements from the equation for ΔE_A, but their general form will turn out to be nonnecessary.

where $\left(\bar{\mathbf{C}}_{A,diag}\right)_{qq'} = \delta_{qq'} C_A$ for $q, q' = x', y', z'$. Note that since the transformation was orthogonal (i.e., the trace of the matrix is preserved), the number C_A may also be obtained from the original matrix $\mathbf{C}_A$ as follows:

$$C_A = \frac{1}{3}\left[\left(\mathbf{C}_{A,diag}\right)_{x'x'} + \left(\mathbf{C}_{A,diag}\right)_{y'y'} + \left(\mathbf{C}_{A,diag}\right)_{z'z'}\right] = \frac{1}{3}\left[\mathbf{C}_{A,xx} + \mathbf{C}_{A,yy} + \mathbf{C}_{A,zz}\right]. \tag{H.4}$$

Then the averaged energy ΔE becomes

$$\Delta\bar{E} = \sum_A \mathbf{I}_A^T \bar{\mathbf{C}}_{A,diag}\mathbf{H} = \sum_A C_A \left(\mathbf{I}_A \cdot \mathbf{H}\right).$$

Thus we obtain the sum of energy contributions over the nuclei, each contribution with its own coefficient averaged over rotations,[3]

$$\Delta\bar{E} = -\sum_A \gamma_A \mathbf{I}_A \cdot \mathbf{H}\left\{1 - \frac{e^2}{2mc^2}\left\langle\psi_0^{(0)}\right|\sum_j \frac{2}{3}(\mathbf{r}_{0j}\cdot\mathbf{r}_{Aj})\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle\right. \tag{H.5}$$
$$\left. -\frac{e^2}{2m^2c^2}\frac{1}{3}\left\langle\psi_0^{(0)}\right|\left[\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\hat{R}_0\left(\sum_j \hat{\mathbf{L}}_{0j}\right) + \left(\sum_j \hat{\mathbf{L}}_{0j}\right)\hat{R}_0\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\right]\left|\psi_0^{(0)}\right\rangle\right\},$$

[3] Indeed, making $C_A = \frac{1}{3}\left[\mathbf{C}_{A,xx} + \mathbf{C}_{A,yy} + \mathbf{C}_{A,zz}\right]$ for the terms of Eq. (H.3) we have the following contributions (term by term):

- $-\gamma_A \frac{1}{3}[1+1+1] = -\gamma_A$,
- $\frac{e^2}{2mc^2}\gamma_A \frac{1}{3}\left[\left\langle\psi_0^{(0)}\right|\sum_j \mathbf{r}_{0j}\cdot\mathbf{r}_{Aj}\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle + \left\langle\psi_0^{(0)}\right|\sum_j \mathbf{r}_{0j}\cdot\mathbf{r}_{Aj}\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle + \left\langle\psi_0^{(0)}\right|\sum_j \mathbf{r}_{0j}\cdot\mathbf{r}_{Aj}\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle\right] = \frac{e^2}{2mc^2}\gamma_A\left\langle\psi_0^{(0)}\right|\sum_j \mathbf{r}_{0j}\cdot\mathbf{r}_{Aj}\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle$,
- $-\frac{e^2}{2mc^2}\gamma_A\left\langle\psi_0^{(0)}\right|\sum_j \frac{1}{3}\left[x_{0j}x_{Aj} + y_{0j}y_{Aj} + z_{0j}z_{Aj}\right]\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle = -\frac{e^2}{2mc^2}\gamma_A\left\langle\psi_0^{(0)}\right|\sum_j \frac{1}{3}\mathbf{r}_{0j}\cdot\mathbf{r}_{Aj}\frac{1}{r_{Aj}^3}\left|\psi_0^{(0)}\right\rangle$,
- $$+\frac{1}{3}\frac{e^2}{2m^2c^2}\gamma_A {\sum_k}' \frac{1}{E_0^{(0)} - E_k^{(0)}} \times$$
$$\times\left[\left\langle\psi_0^{(0)}\right|\left(\sum_j \frac{\hat{L}_{Ajx}}{r_{Aj}^3}\right)\left|\psi_k^{(0)}\right\rangle\left\langle\psi_k^{(0)}\right|\sum_j \hat{L}_{0jx}\left|\psi_0^{(0)}\right\rangle + similarly\ y, z + cc\right] =$$
$$\frac{1}{3}\frac{e^2}{2m^2c^2}\gamma_A {\sum_k}' \frac{1}{E_0^{(0)} - E_k^{(0)}} \times \frac{1}{3}\left[\left\langle\psi_0^{(0)}\right|\left(\sum_j \frac{\hat{\mathbf{L}}_{Aj}}{r_{Aj}^3}\right)\left|\psi_k^{(0)}\right\rangle\left\langle\psi_k^{(0)}\right|\sum_j \hat{\mathbf{L}}_{0j}\left|\psi_0^{(0)}\right\rangle + cc\right],$$

where cc means the "complex conjugate" counterpart. This reproduces Eq. (H.5).

with the matrix elements $\left(\hat{\mathbf{U}}\right)_{kl} = \left\langle \psi_k^{(0)} | \hat{\mathbf{U}} \psi_l^{(0)} \right\rangle$ of the corresponding operators $\hat{\mathbf{U}} = (\hat{U}_x, \hat{U}_y, \hat{U}_z)$.

Finally, after comparing the formula with Eq. (4.81), we obtain the shielding constant for nucleus A (the change of sign in the second part of the formula comes from the change in the denominator) given in Eq. (4.89).

Coupling constants

Averaging over rotations

In each contribution on p. 323 there is a double summation over the nuclear spins, which, after averaging over rotations (similarly as for the shielding constant), gives rise to an energy dependence of the kind $\sum_{A<B} \gamma_A \gamma_B K_{AB} (\hat{\mathbf{I}}_A \cdot \hat{\mathbf{I}}_B)$, which is required in the NMR Hamiltonian. Now, let us take the terms $E_{\mathrm{DSO}}, E_{\mathrm{PSO}}, E_{\mathrm{SD}}, E_{\mathrm{FC}}$ and average them over rotations, producing $\bar{E}_{\mathrm{DSO}}, \bar{E}_{\mathrm{PSO}}, \bar{E}_{\mathrm{SD}}, \bar{E}_{\mathrm{FC}}$. We obtain the following:

- We have

$$\bar{E}_{\mathrm{DSO}} = \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \mathbf{I}_A \cdot \mathbf{I}_B \left\langle \psi_0^{(0)} | \frac{\mathbf{r}_{Aj} \cdot \mathbf{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle$$
$$- \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \frac{1}{3} \mathbf{I}_A \cdot \mathbf{I}_B \left\{ \left\langle \psi_0^{(0)} | \frac{x_{Aj} x_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle + \right.$$
$$\left. \left\langle \psi_0^{(0)} | \frac{y_{Aj} y_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} | \frac{z_{Aj} z_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle \right\},$$

  because the first part of the formula does not need any averaging (it is already in the appropriate form), the second part is averaged according to (H.4). Therefore,

$$\bar{E}_{\mathrm{DSO}} = \frac{e^2}{3mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \mathbf{I}_A \cdot \mathbf{I}_B \left\langle \psi_0^{(0)} | \frac{\mathbf{r}_{Aj} \cdot \mathbf{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle.$$

- We have

$$\bar{E}_{\mathrm{PSO}} = \left\langle \psi_0^{(0)} | \hat{B}_3 \hat{R}_0 \hat{B}_3 \psi_0^{(0)} \right\rangle_{\mathrm{aver}} =$$
$$\left(\frac{i\hbar e}{mc} \right)^2 \sum_{A,B} \sum_{j,l} \gamma_A \gamma_B \left\langle \psi_0^{(0)} | \nabla_j \cdot \frac{\mathbf{I}_A \times \mathbf{r}_{Aj}}{r_{Aj}^3} \hat{R}_0 \nabla_l \cdot \frac{\mathbf{I}_B \times \mathbf{r}_{Bl}}{r_{Bl}^3} \psi_0^{(0)} \right\rangle_{\mathrm{aver}} =$$

$$\left(\frac{i\hbar e}{mc}\right)^2 \sum_{A,B}\sum_{j,l} \gamma_A\gamma_B \left\langle \psi_0^{(0)}|\nabla_j \cdot \frac{\mathbf{r}_{Aj}\times\mathbf{I}_A}{r_{Aj}^3}\hat{R}_0 \nabla_l \cdot \frac{\mathbf{r}_{Bl}\times\mathbf{I}_B}{r_{Bl}}\psi_0^{(0)}\right\rangle_{\text{aver}} =$$
$$-\left(\frac{\hbar e}{mc}\right)^2 \sum_{A,B}\sum_{j,l} \gamma_A\gamma_B \left\langle \psi_0^{(0)}|\mathbf{I}_A \cdot \left(\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}\right)\hat{R}_0 \mathbf{I}_B \cdot \left(\nabla_l \times \frac{\mathbf{r}_{Bl}}{r_{Bl}}\right)\psi_0^{(0)}\right\rangle_{\text{aver}},$$

where the subscript "aver" means the averaging of Eq. (H.4) and the identity $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\times\mathbf{B})\cdot\mathbf{C}$ has been used. We have the following chain of equalities (involving[4] the electronic momenta $\hat{\mathbf{p}}_j$ and angular momenta $\mathbf{L}_{Aj}$ with respect to nucleus A, where j means electron number j):

$$\left(\frac{i\hbar e}{mc}\right)^2 \sum_{A,B}\sum_{j,l} \gamma_A\gamma_B \left\langle \psi_0^{(0)}|\mathbf{I}_A \cdot \frac{1}{i\hbar}\left(\mathbf{r}_{Aj}\times\hat{\mathbf{p}}_j\right)\hat{R}_0 \mathbf{I}_B \cdot \frac{1}{i\hbar}\left(\mathbf{r}_{Bl}\times\hat{\mathbf{p}}_l\right)\psi_0^{(0)}\right\rangle_{\text{aver}} =$$
$$\left(\frac{e}{mc}\right)^2 \sum_{A,B}\sum_{j,l} \gamma_A\gamma_B \left\langle \psi_0^{(0)}|\mathbf{I}_A \cdot \left(\mathbf{r}_{Aj}\times\hat{\mathbf{p}}_j\right)\hat{R}_0 \mathbf{I}_B \cdot \left(\mathbf{r}_{Bl}\times\hat{\mathbf{p}}_l\right)\psi_0^{(0)}\right\rangle_{\text{aver}} =$$
$$\left(\frac{e}{mc}\right)^2 \sum_{A,B}\sum_{j,l} \gamma_A\gamma_B \left\langle \psi_0^{(0)}|\mathbf{I}_A \cdot \hat{\mathbf{L}}_{Aj}\hat{R}_0 \mathbf{I}_B \cdot \hat{\mathbf{L}}_{Bl}\psi_0^{(0)}\right\rangle_{\text{aver}} =$$
$$\left(\frac{e}{mc}\right)^2 \sum_{A,B}\sum_{j,l} \gamma_A\gamma_B \mathbf{I}_A\cdot\mathbf{I}_B \frac{1}{3}\left\{\left\langle \psi_0^{(0)}|\hat{\mathbf{L}}_{Aj,x}\hat{R}_0\hat{\mathbf{L}}_{Bl,x}\psi_0^{(0)}\right\rangle + \left\langle \psi_0^{(0)}|\hat{\mathbf{L}}_{Aj,y}\hat{R}_0\hat{\mathbf{L}}_{Bl,y}\psi_0^{(0)}\right\rangle\right.$$

[4] Let us have a closer look of the operator $\left(\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}\right)$ acting on a function (it is necessary to remember that ∇_j in $\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}$ is not just acting on the components of $\frac{\mathbf{r}_{Aj}}{r_{Aj}^3}$ alone, but in fact on $\frac{\mathbf{r}_{Aj}}{r_{Aj}^3}$ times a wave function) f. We have

$$\left(\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}\right) f = \mathbf{i}\left(\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}\right)_x f + \mathbf{j}\left(\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}\right)_y f + \mathbf{k}\left(\nabla_j \times \frac{\mathbf{r}_{Aj}}{r_{Aj}^3}\right)_z f =$$
$$\mathbf{i}\left(\frac{\partial}{\partial y_j}\frac{z_{Aj}}{r_{Aj}^3} - \frac{\partial}{\partial z_j}\frac{y_{Aj}}{r_{Aj}^3}\right)_x f + \text{similarly with } y \text{ and } z =$$
$$\mathbf{i}\left(-3\frac{y_{Aj}z_{Aj}}{r_{Aj}^4} + \frac{z_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial y_j} + 3\frac{y_{Aj}z_{Aj}}{r_{Aj}^4} - \frac{y_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial z_j}\right)_x f + \text{similarly with } y \text{ and } z =$$
$$\mathbf{i}\left(\frac{z_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial y_j} - \frac{y_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial z_j}\right)_x f + \text{similarly with } y \text{ and } z = \mathbf{i}\left(\frac{z_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial y_j} - \frac{y_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial z_j}\right)_x f$$
$$+\text{similarly with } y \text{ and } z = -\frac{1}{i\hbar}\left(-\mathbf{r}_{Aj}\times\hat{\mathbf{p}}_j\right) f = \frac{1}{i\hbar}\left(\mathbf{r}_{Aj}\times\hat{\mathbf{p}}_j\right) f.$$

$$+\left\langle \psi_0^{(0)}|\hat{\mathbf{L}}_{Aj,z}\hat{R}_0\hat{\mathbf{L}}_{Bl,z}\psi_0^{(0)}\right\rangle\Bigg\}.$$

Thus, finally

$$\bar{E}_{\mathrm{PSO}}=\frac{1}{3}\left(\frac{e}{mc}\right)^2\sum_{A,B}\sum_{j,l}\gamma_A\gamma_B\mathbf{I}_A\cdot\mathbf{I}_B\left\langle\psi_0^{(0)}|\hat{\mathbf{L}}_{Aj}\hat{R}_0\hat{\mathbf{L}}_{Bl}\psi_0^{(0)}\right\rangle.$$

- We have

$$\bar{E}_{\mathrm{SD}}=\left\langle\psi_0^{(0)}|\hat{B}_6\hat{R}_0\hat{B}_6\psi_0^{(0)}\right\rangle_{\mathrm{aver}}=$$
$$\gamma_{el}^2\sum_{j,l=1}^{N}\sum_{A,B}\gamma_A\gamma_B\left\langle\psi_0^{(0)}|\left[\frac{\hat{\mathbf{s}}_j\cdot\mathbf{I}_A}{r_{Aj}^3}-3\frac{\left(\hat{\mathbf{s}}_j\cdot\mathbf{r}_{Aj}\right)\left(\mathbf{I}_A\cdot\mathbf{r}_{Aj}\right)}{r_{Aj}^5}\right]\times\right.$$
$$\left.\hat{R}_0\left[\frac{\hat{\mathbf{s}}_l\cdot\mathbf{I}_B}{r_{Bl}^3}-3\frac{\left(\hat{\mathbf{s}}_l\cdot\mathbf{r}_{Bl}\right)\left(\mathbf{I}_B\cdot\mathbf{r}_{Bl}\right)}{r_{Bl}^5}\right]\psi_0^{(0)}\right\rangle_{\mathrm{aver}}=$$
$$\gamma_{el}^2\sum_{j,l=1}^{N}\sum_{A,B}\gamma_A\gamma_B\mathbf{I}_A\cdot\mathbf{I}_B\frac{1}{3}\Bigg\{\left\langle\psi_0^{(0)}|\left[\frac{\hat{\mathbf{s}}_{j,x}}{r_{Aj}^3}-3\frac{\left(\hat{\mathbf{s}}_j\cdot\mathbf{r}_{Aj}\right)x_{Aj}}{r_{Aj}^5}\right]\times\right.$$
$$\left.\hat{R}_0\left[\frac{\hat{\mathbf{s}}_{l,x}}{r_{Bl}^3}-3\frac{\left(\hat{\mathbf{s}}_l\cdot\mathbf{r}_{Bl}\right)(x_{Bl})}{r_{Bl}^5}\right]\psi_0^{(0)}\right\rangle+$$
$$\left\langle\psi_0^{(0)}|\left[\frac{\hat{\mathbf{s}}_{j,y}}{r_{Aj}^3}-3\frac{\left(\hat{\mathbf{s}}_j\cdot\mathbf{r}_{Aj}\right)y_{Aj}}{r_{Aj}^5}\right]\hat{R}_0\left[\frac{\hat{\mathbf{s}}_{l,y}}{r_{Bl}^3}-3\frac{\left(\hat{\mathbf{s}}_l\cdot\mathbf{r}_{Bl}\right)(y_{Bl})}{r_{Bl}^5}\right]\psi_0^{(0)}\right\rangle+$$
$$\left\langle\psi_0^{(0)}|\left[\frac{\hat{\mathbf{s}}_{j,z}}{r_{Aj}^3}-3\frac{\left(\hat{\mathbf{s}}_j\cdot\mathbf{r}_{Aj}\right)z_{Aj}}{r_{Aj}^5}\right]\hat{R}_0\left[\frac{\hat{\mathbf{s}}_{l,z}}{r_{Bl}^3}-3\frac{\left(\hat{\mathbf{s}}_l\cdot\mathbf{r}_{Bl}\right)(z_{Bl})}{r_{Bl}^5}\right]\psi_0^{(0)}\right\rangle\Bigg\}.$$

Therefore,

$$\bar{E}_{\mathrm{SD}}=\frac{1}{3}\gamma_{el}^2\sum_{j,l=1}^{N}\sum_{A,B}\gamma_A\gamma_B\mathbf{I}_A\cdot\mathbf{I}_B\left\langle\psi_0^{(0)}|\left[\frac{\hat{\mathbf{s}}_j}{r_{Aj}^3}-3\frac{\left(\hat{\mathbf{s}}_j\cdot\mathbf{r}_{Aj}\right)\mathbf{r}_{Aj}}{r_{Aj}^5}\right]\times\right.$$
$$\left.\hat{R}_0\left[\frac{\hat{\mathbf{s}}_l}{r_{Bl}^3}-3\frac{\left(\hat{\mathbf{s}}_l\cdot\mathbf{r}_{Bl}\right)(\mathbf{r}_{Bl})}{r_{Bl}^5}\right]\psi_0^{(0)}\right\rangle.$$

- We have

$$\bar{E}_{\mathrm{FC}} = \left\langle \psi_0^{(0)} | \hat{B}_7 \hat{R}_0 \hat{B}_7 \psi_0^{(0)} \right\rangle =$$

$$\gamma_{el}^2 \sum_{j,l=1} \sum_{A,B} \gamma_A \gamma_B \left\langle \psi_0^{(0)} | \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_j \cdot \mathbf{I}_A \hat{R}_0 \delta\left(\mathbf{r}_{Bl}\right) \hat{\mathbf{s}}_l \cdot \mathbf{I}_B \psi_0^{(0)} \right\rangle_{\mathrm{aver}} =$$

$$\gamma_{el}^2 \sum_{j,l=1} \sum_{A,B} \gamma_A \gamma_B \mathbf{I}_A \cdot \mathbf{I}_B \frac{1}{3} \left\{ \left\langle \psi_0^{(0)} | \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_{j,x} \hat{R}_0 \delta\left(\mathbf{r}_{Bl}\right) \hat{\mathbf{s}}_{l,x} \psi_0^{(0)} \right\rangle + \right.$$

$$\left. \left\langle \psi_0^{(0)} | \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_{j,y} \hat{R}_0 \delta\left(\mathbf{r}_{Bl}\right) \hat{\mathbf{s}}_{l,y} \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} | \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_{j,z} \hat{R}_0 \delta\left(\mathbf{r}_{Bl}\right) \hat{\mathbf{s}}_{l,z} \psi_0^{(0)} \right\rangle \right\}.$$

  Hence,

$$\bar{E}_{\mathrm{FC}} = \frac{1}{3} \left(\frac{8\pi}{3} \right)^2 \gamma_{el}^2 \sum_{j,l=1} \sum_{A,B} \gamma_A \gamma_B \mathbf{I}_A \cdot \mathbf{I}_B \left\langle \psi_0^{(0)} | \delta\left(\mathbf{r}_{Aj}\right) \hat{\mathbf{s}}_j \hat{R}_0 \delta\left(\mathbf{r}_{Bl}\right) \hat{\mathbf{s}}_l \psi_0^{(0)} \right\rangle.$$

The results mean that the coupling constants J are such as reported on p. 324.

APPENDIX I

Acceptor–Donor Structure Contributions in the MO Configuration

In Chapter 6 the Slater determinants have been constructed in three different ways, using:

- molecular orbitals (MO picture),
- acceptor and donor orbitals (AD picture),
- atomic orbitals (VB picture).

Then, a problem appeared how to express one picture by another; in particular this has been of importance for expressing the MO picture in the AD one. More specifically, we are interested in calculating the contribution of an acceptor–donor structure[1] in the Slater determinant written in the MO formalism, where the molecular orbitals are expressed by the donor (n) and acceptor (χ and χ^*) orbitals in the following way:

$$\begin{aligned} \varphi_1 &= a_1 n + b_1 \chi - c_1 \chi^*, \\ \varphi_2 &= a_2 n - b_2 \chi - c_2 \chi^*, \\ \varphi_3 &= -a_3 n + b_3 \chi - c_3 \chi^*. \end{aligned} \tag{I.1}$$

We assume that $\{\varphi_i\}$ form an orthonormal set. For simplicity reasons it is also assumed that in the first approximation the orbitals $\{n, \chi, \chi^*\}$ are also orthonormal. Then we may write that a normalized Slater determinant in the MO picture (denoted by X_i) represents a linear combination of the normalized Slater determinants (Y_j), each of the latter containing exclusively the donor and the acceptor orbitals, i.e.,

$$X_i = \sum_j C_i \left(Y_j\right) Y_j,$$

where, due to the assumed orthogonality, the coefficient $C_i\left(Y_k\right) = \langle Y_k | X_i \rangle$ at the Slater determinant Y_k is the contribution of the acceptor–donor structure Y_k in X_i.

In Chapter 6 three particular cases are highlighted, which will be derived below. We will use the antisymmetrizer $\hat{A} = \frac{1}{N!} \sum_P (-1)^p \hat{P}$ introduced in Chapter 2 ($\hat{P}$ is the permutation operator, and p is its parity).

[1] That is, of a Slater determinant built of the acceptor and of the donor orbitals.

Case $C_0(DA)$

The $C_0(DA)$ coefficient means a contribution of the structure $n^2\chi^2$, i.e., $\Psi(DA) = (4!)^{-\frac{1}{2}} \det\left[n\bar{n}\chi\bar{\chi}\right] = (4!)^{\frac{1}{2}} \hat{A}\left[n\bar{n}\chi\bar{\chi}\right]$ in the ground-state Slater determinant $\Psi_0 = (4!)^{-\frac{1}{2}} \det\left[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\right] = (4!)^{\frac{1}{2}} \hat{A}\left[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\right]$. We have to calculate

$$C_0(DA) = \langle Y_k | X_i \rangle = \langle \Psi(DA) | \Psi_0 \rangle = 4! \left\langle \hat{A}\left[n\bar{n}\chi\bar{\chi}\right] | \hat{A}\left[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\right] \right\rangle =$$
$$4! \left\langle \left[n\bar{n}\chi\bar{\chi}\right] | \hat{A}^2\left[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\right] \right\rangle = 4! \left\langle \left[n\bar{n}\chi\bar{\chi}\right] | \hat{A}\left[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\right] \right\rangle =$$
$$4! \left\langle \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right] | \hat{A}\left[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\right] \right\rangle,$$

where we have used the notion that $\hat{A}$ is Hermitian and idempotent. Next, one has to write down all the 24 permutations $[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)]$ (taking into account their parity) and then perform integration over the coordinates of all the four electrons (together with summation over spin variables). Then we obtain

$$C_0(DA) = \int d\tau_1 d\tau_2 d\tau_3 d\tau_4 \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]^* \sum_P (-1)^p P\left[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\right].$$

The integral to survive *has* to have perfect matching of the spin functions between $\left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]$ and $\hat{P}\left[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\right]$. This makes 20 of those permutations vanish! Only the following four integrals will survive:

$$C_0(DA) = \int d\tau_1 d\tau_2 d\tau_3 d\tau_4 \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]^* \left[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\right] +$$
$$- \int d\tau_1 d\tau_2 d\tau_3 d\tau_4 \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]^* \left[\varphi_1(1)\bar{\varphi}_1(4)\varphi_2(3)\bar{\varphi}_2(2)\right] +$$
$$- \int d\tau_1 d\tau_2 d\tau_3 d\tau_4 \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]^* \left[\varphi_1(3)\bar{\varphi}_1(2)\varphi_2(1)\bar{\varphi}_2(4)\right] +$$
$$\int d\tau_1 d\tau_2 d\tau_3 d\tau_4 \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]^* \left[\varphi_1(3)\bar{\varphi}_1(4)\varphi_2(1)\bar{\varphi}_2(2)\right] =$$
$$\int d\tau_1 n(1)^* \varphi_1(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_1(2) \int d\tau_3 \chi(3)^* \varphi_2(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_2(4) +$$
$$- \int d\tau_1 n(1)^* \varphi_1(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_2(2) \int d\tau_3 \chi(3)^* \varphi_2(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_1(4) +$$
$$- \int d\tau_1 n(1)^* \varphi_2(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_1(2) \int d\tau_3 \chi(3)^* \varphi_1(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_2(4) +$$
$$\int d\tau_1 n(1)^* \varphi_2(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_2(2) \int d\tau_3 \chi(3)^* \varphi_1(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_1(4) =$$

$$(a_1)^2(-b_2)^2 - a_1a_2(-b_2)b_1 - a_2a_1b_1(-b_2) + (a_2)^2(b_1)^2 =$$
$$(a_1)^2(b_2)^2 + a_1a_2b_2b_1 + a_2a_1b_1b_2 + (a_2)^2(b_1)^2 = a_1b_2(a_1b_2 + a_2b_1) + a_2b_1(a_1b_2 + a_2b_1) =$$
$$(a_1b_2 + a_2b_1)^2 = \begin{vmatrix} a_1 & a_2 \\ b_1 & -b_2 \end{vmatrix}^2.$$

Hence,

$$C_0(DA) = \begin{vmatrix} a_1 & a_2 \\ b_1 & -b_2 \end{vmatrix}^2,$$

which agrees with the formula on p. 482, which was our goal.

Case $C_2(DA)$

The $C_2(DA)$ represents the contribution of the structure $\Psi(DA) = (4!)^{\frac{1}{2}}\hat{A}\left[n\bar{n}\chi\bar{\chi}\right]$ in the Slater determinant corresponding to the double excitation $\Psi_{2d} = (4!)^{\frac{1}{2}}\hat{A}[\varphi_1\bar{\varphi}_1\varphi_3\bar{\varphi}_3]$. We are interested in the integral

$$C_2(DA) = \langle\Psi(DA)|\Psi_{2d}\rangle = 4!\left\langle\left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right]|\hat{A}[\varphi_1(1)\bar{\varphi}_1(2)\varphi_3(3)\bar{\varphi}_3(4)]\right\rangle.$$

This case is very similar to the previous one; the only difference is a substitution $\varphi_2 \to \varphi_3$. Everything goes therefore the same way as before, but this time we obtain

$$\begin{aligned} C_2(DA) = &\int d\tau_1 n(1)^*\varphi_1(1)\int d\tau_2\bar{n}(2)^*\bar{\varphi}_1(2)\int d\tau_3\chi(3)^*\varphi_3(3)\int d\tau_4\bar{\chi}(4)^*\bar{\varphi}_3(4) + \\ &- \int d\tau_1 n(1)^*\varphi_1(1)\int d\tau_2\bar{n}(2)^*\bar{\varphi}_3(2)\int d\tau_3\chi(3)^*\varphi_3(3)\int d\tau_4\bar{\chi}(4)^*\bar{\varphi}_1(4) + \\ &- \int d\tau_1 n(1)^*\varphi_3(1)\int d\tau_2\bar{n}(2)^*\bar{\varphi}_1(2)\int d\tau_3\chi(3)^*\varphi_1(3)\int d\tau_4\bar{\chi}(4)^*\bar{\varphi}_3(4) + \\ &\int d\tau_1 n(1)^*\varphi_3(1)\int d\tau_2\bar{n}(2)^*\bar{\varphi}_3(2)\int d\tau_3\chi(3)^*\varphi_1(3)\int d\tau_4\bar{\chi}(4)^*\bar{\varphi}_1(4), \end{aligned}$$

or

$$C_2(DA) = (a_1)^2(b_3)^2 - a_1(-a_3)b_3b_1 - (-a_3)a_1b_1b_3 + (-a_3)^2(b_1)^2 =$$
$$(a_1)^2(b_3)^2 + a_1a_3b_3b_1 + a_3a_1b_1b_3 + (a_3)^2(b_1)^2 =$$
$$(a_1b_3 + a_3b_1)^2 = \begin{vmatrix} a_1 & b_1 \\ -a_3 & b_3 \end{vmatrix}^2.$$

We have

$$C_2(DA) = \begin{vmatrix} a_1 & b_1 \\ -a_3 & b_3 \end{vmatrix}^2,$$

which again agrees with the result used on p. 483.

Case $C_3(DA)$

This time we have to compute the contribution of $\Psi(DA) = (4!)^{\frac{1}{2}} \hat{A}\left[n\bar{n}\chi\bar{\chi}\right]$ in the Slater determinant $\Psi_{3d} = (4!)^{\frac{1}{2}} \hat{A}[\varphi_2\bar{\varphi}_2\varphi_3\bar{\varphi}_3]$. Therefore

$$C_2(DA) = \langle \Psi(DA) | \Psi_{3d} \rangle = 4! \left\langle \left[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\right] | \hat{A}[\varphi_2(1)\bar{\varphi}_2(2)\varphi_3(3)\bar{\varphi}_3(4)] \right\rangle.$$

This is a case similar to the previous one, but we have to exchange $\varphi_1 \to \varphi_2$. We obtain

$$\begin{aligned} C_3(DA) = &\int d\tau_1 n(1)^* \varphi_2(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_2(2) \int d\tau_3 \chi(3)^* \varphi_3(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_3(4) + \\ &- \int d\tau_1 n(1)^* \varphi_2(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_3(2) \int d\tau_3 \chi(3)^* \varphi_3(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_2(4) + \\ &- \int d\tau_1 n(1)^* \varphi_3(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_2(2) \int d\tau_3 \chi(3)^* \varphi_2(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_3(4) + \\ &\int d\tau_1 n(1)^* \varphi_3(1) \int d\tau_2 \bar{n}(2)^* \bar{\varphi}_3(2) \int d\tau_3 \chi(3)^* \varphi_2(3) \int d\tau_4 \bar{\chi}(4)^* \bar{\varphi}_2(4), \end{aligned}$$

or

$$C_3(DA) = (a_2)^2 (b_3)^2 - a_2(-a_3) b_3(-b_2) - (-a_3) a_2(-b_2) b_3 + (-a_3)^2 (-b_2)^2 =$$
$$(a_2)^2 (b_3)^2 - a_2 a_3 b_3 b_2 - a_3 a_2 b_2 b_3 + (a_3)^2 (b_2)^2 = a_2 b_3 [a_2 b_3 - a_3 b_2] - a_3 b_2 [a_2 b_3 - a_3 b_2] =$$
$$(a_2 b_3 - a_3 b_2)^2 = \begin{vmatrix} a_2 & -b_2 \\ -a_3 & b_3 \end{vmatrix}^2.$$

Finally,

$$C_3(DA) = \begin{vmatrix} a_2 & -b_2 \\ -a_3 & b_3 \end{vmatrix}^2,$$

and again agreement with the formula on p. 483 is obtained.

Acronyms and Their Explanation

AD	Acceptor–donor method ✠ A theoretical description of chemical reactions in terms of acceptor molecular orbitals and donor molecular orbitals.
AIM	Atoms-in-molecules ✠ An analysis of the critical points of the molecular electron density distribution that leads to its unique partition into atomic contributions.
AMO	Alternant molecular orbitals ✠ A version of the UHF method, in which the occupied orbitals are modified by admixtures of virtual orbitals
AO	Atomic orbital ✠ A function of an electron's position in space, centered in a point and decaying exponentially (like STO or GTO) at large distances from the center.
BFCS	Body-fixed coordinate system ✠ The coordinate system fixed on the moving molecule.
BO	Born–Oppenheimer approximation ✠ An approximation assuming that the electrons move in the field of the clamped nuclei, while the nuclei move in the potential energy being the electronic energy.
BOAS	Bond-Order Alternating Solution ✠ The electronic density distribution that breaks the translational symmetry of the nuclear framework by doubling the period.
BSSE	Basis set superposition error ✠ An error in calculation of intermolecular interaction energy stemming from using an incomplete basis set of atomic orbitals and calculation of the energies of isolated molecules by using only their own basis sets of atomic orbitals.
B3LYP	Becke–Lee–Young–Parr density functional theory ✠ A semiempirical DFT method of hybrid type, i.e., with the exchange–correlation potential composed of several empirical contributions.
CAS SCF	Complete active space self-consistent field method ✠ The iterative and variational method of solving the Schrödinger equation with the variational wave function in the form of a linear combination of all the Slater determinants (coefficients and spin orbitals are determined variationally) that can be built from a limited set of the spin orbitals (forming the active space).

CC	Coupled cluster method ✠ Nonvariational method of solving the Schrödinger equation with the wave function in the form of an exponential operator to be determined acting on the Hartree–Fock wave function.
CCSD	Coupled cluster singles and doubles ✠ Nonvariational method of solving the Schrödinger equation with the wave function in the form of an exponential operator (with the explicit presence of the single and double excitations, their contribution to be determined in the method) acting on the Hartree–Fock wave function.
CCSD-R12	Coupled cluster singles and doubles with r_{12} ✠ Nonvariational method of solving the Schrödinger equation with the wave function in the form of an exponential operator (with the explicit presence of the single and double excitations, their contribution to be determined in the method) acting on the Hartree–Fock wave function with the correlation factors of the type $(1 + \frac{1}{2}r_{12})$.
CCSD(T)	Coupled cluster singles and doubles with estimated triples ✠ Nonvariational method of solving the Schrödinger equation with the wave function in the form of an exponential operator (with the explicit presence of the single and double excitations, their contribution to be determined in the method, and perturbational approximation of the triple excitation contribution) acting on the Hartree–Fock wave function.
CCSD(T)-R12	Coupled cluster singles and doubles with estimated triples ✠ Nonvariational method of solving the Schrödinger equation with the wave function in the form of an exponential operator (with the explicit presence of the single and double excitations, their contribution to be determined in the method, and perturbational approximation of the triple excitation contribution) acting on the Hartree–Fock wave function with the correlation factors of the type $(1 + \frac{1}{2}r_{12})$.
CI	Configuration interaction ✠ Variational method with the trial wave function in the form of a linear combination of the given set of the Slater determinants.
CIS	Configuration interaction singles ✠ Variational method with the trial wave function in the form of a linear combination of the given set of the singly excited Slater determinants.
CISD	Configuration interaction singles and doubles ✠ Variational method with the trial wave function in the form of a linear combination of the given set of the singly and doubly excited Slater determinants.
CISDT	Configuration Interaction singles, doubles, and triples ✠ Variational method with the trial wave function in the form of a linear combination of the given set of the singly, doubly, and triply excited Slater determinants.

CP	Counterpoise method ✠ A method of elimination of the basis set superposition error (BSSE) in the intermolecular interaction energy by calculating all quantities using the basis set of atomic orbitals of the whole system.
CS	Charge-shift bonding ✠ Two maxima of the electron localization function in a chemical bond, interpreted as a manifestation of resonance of two ionic structures.
CSF	Configuration state function ✠ An expansion function in the configuration interaction method that has the same symmetry and spin state as those of the exact wave function.
CT	Charge transfer ✠ Transition to an excited state with the charge distribution differing substantially from the ground state charge distribution.
DC	Dirac–Coulomb approximation ✠ Approximate and many-electron quasirelativistic theory, in which the one-electron Hamiltonians are the Dirac relativistic Hamiltonians, whereas the electron–electron interaction operators are represented uniquely by the (nonrelativistic) Coulomb interactions.
DF	Density fitting method ✠ Expansion of two orbitals' products into a series of auxiliary set of functions (usually monocentric). Equivalent to the resolution of identity (RI) method.
DF MPn-R12	Møller–Plesset perturbation theory with r_{12} and density fitting ✠ MPn perturbational method (accurate to the n-th order) of solution of the Schrödinger equation with the Hartree–Fock function multiplied by the explicit correlation factor $(1 + \frac{1}{2}r_{12})$ as the zeroth-order approximation, with the density fitting (DF) used for orbital products.
DFT	Density functional theory ✠ A theory in which the total energy of a molecule depends on its electron density distribution.
DODS	Different orbitals for different spins ✠ Another name for the UHF method.
DRC	Distinguished reaction coordinate ✠ A selected distance changing from its value for the reactants to the value for the products of an elementary chemical reaction.
ELF	Electron localization function ✠ A measure, defined in density functional theory, of the tendency to occupy a point of space by an electron pair.
EOM-CC	Equation of motion coupled cluster ✠ A nonvariational method of solving the Schrödinger equation for excited states (related to the equation of motion), with the wave function calculated in the coupled cluster method.
FBZ	First Brillouin zone ✠ The set of vectors of the inverse space in a periodic system, which correspond to all possible distinct Bloch functions.
FCI	Full configuration interaction ✠ A configuration interaction method with all possible excitations from a given finite set of molecular orbitals.

FCIQMC	Full configuration interaction quantum Monte Carlo ✠ A full-configuration-interaction-quality configuration interaction method with stochastically determined initial configuration interaction wave function and its perturbational improvement afterwards.
FEMO	Free electron molecular orbitals ✠ π electrons in a molecule treated as free electrons in a box.
FF	Force field ✠ A simple mathematical expression mimicking the electronic energy as a function of the positions of the nuclei.
FF	Finite field ✠ Solving the Schrödinger equation for a molecule in an external field with the molecule–field interaction term included in the Hamiltonian.
FVAO	Field-variant atomic orbitals ✠ Atomic orbital centers depend on the external electric field intensity.
GEA	Gradient expansion approximation ✠ A class of DFT functionals that take into account a nonlocal character of the exchange–correlation energy through a gradient correction.
GHF	General Hartree–Fock ✠ The Hartree–Fock method with spin orbitals of the most general form.
GIAO	Gauge including atomic orbitals, previously also gauge-invariant atomic orbitals ✠ A method of calculations for a molecule in the magnetic field that ensures the invariance of the results with respect to the choice of the origin of the vector potential describing the magnetic field.
GTO	Gaussian-type orbital ✠ Atomic orbitals with an exponential decaying as $\exp(-\zeta r^2)$, where r stands for the distance from a given point in space ("center") and $\zeta > 0$.
HF	Hartree–Fock method ✠ Variational method with the trial wave function in the form of a single Slater determinant.
HOMO	Highest Occupied Molecular Orbital ✠ The highest (in the energy scale) occupied (be electrons) molecular orbital.
HTS	High-temperature superconductor ✠ The crystalline substances exhibiting superconductivity with an unusually high critical temperature.
IRC	Intrinsic reaction coordinate ✠ The steepest descent curve in the space of the nuclear configurations (with the mass-weighted coordinates) that connects two electronic energy minima through the first-order saddle point (transition state).
KS	Kohn–Sham method ✠ The DFT method in which the electronic density distribution results from a single Slater determinant (Kohn–Sham determinant).
LCAO CO	Linear combination of atomic orbitals, crystal orbitals ✠ Expression of the crystal orbitals as linear combinations of atomic orbitals.

LCAO MO	Linear combination of atomic orbitals, molecular orbitals ✠ Expression of the molecular orbitals as linear combinations of atomic orbitals.
LDA	Local density approximation ✠ A DFT method that estimates the exchange–correlation energy from this energy in the homogeneous electron gas.
LUMO	Lowest Unoccupied Molecular Orbital ✠ The lowest (in the energy scale) unoccupied molecular orbital.
MBPT	Many-body perturbation theory ✠ An iterative perturbation-based method of solving the Schrödinger equation.
MCD	Monte Carlo dynamics ✠ A dynamics with a stochastic choice of configurations of the nuclei and a criterion for accepting or rejecting this choice.
MC SCF	Multiconfigurational self-consistent field ✠ A variational iterative solution of the Schrödinger equation with the trial function in the form of linear combinations of variable Slater determinants.
MD	Molecular dynamics ✠ Solution of the Newton equation of motion for nuclei.
MEP	Molecular electrostatic potential ✠ The electrostatic potential created by a molecule as a function of position in space.
MM	Molecular mechanics ✠ Minimization of the molecular electronic energy, approximated by the force field, as a function of positions of the nuclei.
MO	Molecular orbital ✠ A one-electron function, which is a solution of the Fock equation for a molecule.
MP, MP2, MP4	Møller–Plesset perturbation theory ✠ Perturbational method (up to the second [MP2] or fourth [MP4] order) of solution of the Schrödinger equation with the Hartree–Fock function as the zeroth approximation.
MPn-R12	Møller–Plesset perturbation theory with r_{12} ✠ Perturbational method (accurate to the n-th order) of solution of the Schrödinger equation with the Hartree–Fock function multiplied by the explicit correlation factor $(1 + \frac{1}{2}r_{12})$ as the zeroth-order approximation.
NLDA	Nonlocal density approximation ✠ A DFT method with the exchange–correlation energy correction containing the electron density gradient.
NMR	Nuclear magnetic resonance ✠ A spectroscopic method in which transitions between the energy levels of the nuclear magnetic moments result from their interaction with the local magnetic field and among themselves.
NO	Natural orbital ✠ The molecular orbitals corresponding to the diagonal form of the one-electron density matrix.
PES	Potential energy surface ✠ The electronic energy as a function of configuration of the nuclei.
PW	Plane waves ✠ Function $A_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r})$ used in descriptions of periodic systems with vector $\mathbf{k}$ belonging to the first Brillouin zone.

PW91	Perdew–Wang density functional theory ✠ A semiempirical method of finding the ground-state electronic density distribution within the density functional theory.
QED	Quantum electrodynamics ✠ Quantum theory of charged particles interacting with an electromagnetic field that goes beyond the Dirac theory.
QMC	Quantum Monte Carlo ✠ Stochastic evolution of the time-dependent Schrödinger equation with imaginary time that gives the ground-state energy of the system.
RHF	Restricted Hartree–Fock method ✠ The variational method with a single Slater determinant with doubly occupied molecular orbitals as a trial function.
RI	Resolution of identity method ✠ Expansion of two orbitals' products into a series of auxiliary set of functions (usually monocentric). Equivalent to the density fitting method.
ROHF	Restricted open shell Hartree–Fock method ✠ The variational method with Slater determinant(s) with doubly occupied core molecular orbitals, but different valence molecular orbitals for different spins.
SAPT	Symmetry-adapted perturbation theory ✠ Perturbational method of calculating intermolecular interaction energy with taking into account the Pauli exclusion principle.
SCF	Self-consistent field ✠ Iterative method of solving the Fock equation.
SCF LCAO CO	Self-consistent field linear combination of atomic orbitals, crystal orbitals ✠ Iterative method of solving the Fock equation for crystals (in the LCAO CO approximation).
SCF LCAO MO	Self-consistent field linear combination of atomic orbitals, molecular orbitals ✠ Iterative method of solving the Fock equation for molecule (in the LCAO MO approximation).
SDP	Steepest descent path ✠ Steepest descent trajectory (of lowering the electronic energy as a function of configuration of the nuclei) that connects a first-order saddle point with two adjacent energy minima corresponding to the stable configurations of the reactants and products.
SFCS	Space-fixed coordinate system ✠ The coordinate system of the laboratory in which the molecule is observed and measured.
SE	Single exchange ✠ A contribution to the exchange interaction (valence repulsion of molecules) nonadditivity effect coming from the interaction of the Pauli deformation of the electron cloud due to two interacting molecules with the electric field created by the third molecule.
SHG	Second harmonic generation ✠ Frequency doubling of light in materials with nonlinear electric properties.

SOS	Sum over states ✠ Perturbational corrections with summation over unperturbed states.
STO	Slater-type orbital ✠ Atomic orbitals with the asymptotic exponential decay $\exp(-\zeta r)$, where r means the distance from a certain point in space and $\zeta > 0$.
SUSY	Supersymmetry ✠ A symmetry-like relation between two dissimilar systems that comes from a symmetry of mathematical expressions that describe them.
TE	Triple exchange ✠ A contribution to the exchange interaction (valence repulsion of molecules) nonadditivity effect coming from a single electron exchange between two molecules by mediation of a third one.
THG	Third harmonic generation ✠ Frequency tripling of light in materials with nonlinear electric properties.
UHF	Unrestricted Hartree-Fock method ✠ The variational method with a single Slater determinant as a trial function (without molecular orbital double occupancy restriction).
VB	Valence bond ✠ A variational method with the wave function in the form of a linear combination of the Slater determinants built of atomic spin orbitals.
VSEPR	Valence Shell Electron Pair Repulsion ✠ An algorithm to predict the spatial structure of a molecule by counting the electronic pairs in the valence shell of a central atom with substituents.
ZDO	Zero-differential overlap ✠ Neglecting any product of two atomic orbitals (that describe the same electron) with different centers.

Author Index

Subject Index

Sources of Photographs and Figures

The figures in this book, except those listed below or acknowledged in situ, are manufactured by the author and reproduced thanks to the courtesy of Wydawnictwo Naukowe PWN, Poland from "Idee chemii kwantowej," © 2012 PWN ★ The copies of the public domain postal stamps or medals concerning scientists have been used in biographic boxes (V1: pp. 1, 5, 6, 10, 12, 14, 30, 34, 85×2, 88, 127, 128, 130, 131×2, 147, 309×2, 310, 320, 357, 386, 401, 416, 435, 521; V2: pp. 2, 9, 24, 64, 99, 102, 130, 258, 303, 340×3, 427, 440, 483, 514, 544, 548, 558).

Vol. 1: ★ Wikipedia + websites (Rayleigh 6, Lewis 9, Sommerfeld 11, Born 15, Bose 32, Noether 72, Heisenberg 42, Bell 54, Weyl 90, Minkowski 140, Klein 146, Anderson 149, Breit 171, Condon 358, Jacobi 402, Crick 406, Watson 406, Lennard-Jones 409, Ulam 436, Hartree 463, Fock 463, Slater 467, Hueckel 500, Roothaan 504, Hund 541, Koopmans 546). ★ p. 448 MathWorld – A Wolfram Web Resource. ★ p. 539 Photo of Aleksander Jabłoński, courtesy of the Physics Department, Nicolaus Copernicus University, Toruń. ★ p. 552 Table, Reidel Publish. Co. ★ p. 147 Photo of Charles Galton Darwin – courtesy of Dr. R.C. McGuiness. ★ p. 310 Photo of Christopher Longuet-Higgins – courtesy of Professor J.D. Roberts. ★ p. 327 Courtesy of Professor Bohdan Korybut-Daszkiewicz. ★ p. 385 Courtesy of Professor Andrzej Sobolewski. ★ p. 451 Courtesy of Professors Andrzej Koliński and Krzysztof Bujnicki. ★ p. 451 Courtesy of Professor David Baker. ★ p. 520 Courtesy of Professors Trygve Helgaker, Poul Jørgensen, Jeppe Olsen; Wiley and Sons Ltd. ★ p. 541 Photo of Friedrich Hund – courtesy of Mr. Gerhard Hund.

Vol. 2: ★ Wikipedia + websites (crystals 4, Brillouin 14, Kato 90, Friedmann 101, London 128, Kohn 205, Van Vleck 256, Ramsey 311, Purcell 311, Lee 440, Herschbach 441, Marcus 511, Feigenbaum 543, Belousov 562, Adleman 568, Turing 569, Hamilton 570). ★ pp. 22, 33, 36, 44, 49, 50 Courtesy of Professor Roald Hoffmann, permission from John Wiley and Sons. ★ pp. 45, 46 Courtesy of Professor Jean-Marie André. ★ p. 174 Courtesy of Professor Trygve Helgaker and permission of John Wiley and Sons Ltd. ★ p. 183 Courtesy of Professor Hiroshi Nakatsuji. ★ pp. 198, 201, 204, 205 Courtesy of Professor Richard Bader. ★ Figs. 3.10–3.12 reused from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, G.J. Laming, *J. Chem. Phys.*, 99 (1993) 417. ★ p. 257 Portrait of Hans Hellmann reproduced from a painting by Ms. Tatiana

Livschitz, courtesy of Professor W.H. Eugen Schwartz. ★ p. 287 Courtesy of Professor Sadlej's family. ★ Figures on pp. 474–479, Courtesy of Professor Thom Dunning. ★ Tables 6.1–6.5, courtesy of Professor Sason Shaik. ★ p. 541 Courtesy of Professor Jean-Marie Lehn. ★ p. 547 Courtesy of Mr. John Gleick.

If you are the copyright owner to any of the images we have used without your explicit permission (because we were unable to identify you during our search), please contact Professor Lucjan Piela, Chemistry Department, Warsaw University, 02093 Warsaw, Poland, e-mail: piela@chem.uw.edu.pl, phone (48)-22-7226692. We will be pleased to place the appropriate information on our website at booksite.elsevier.com/978-0-444-59436-5, which supports the present book and represents its integral part.

Tables

Table 1. Units of physical quantities.

Quantity	Unit	Symbol	Value
light velocity		c	299 792.458 km/s
Planck constant		h	$6.6260755 \cdot 10^{-34}$ J·s
mass	electron rest mass	m_0	$9.1093897 \cdot 10^{-31}$ kg
charge	elementary charge = a.u. of charge	e	$1.60217733 \cdot 10^{-19}$ C
action	$\frac{h}{2\pi}$	$\hbar$	$1.05457266 \cdot 10^{-34}$ J·s
length	Bohr radius of hydrogen	a_0	$5.29177249 \cdot 10^{-11}$ m
energy	Hartree = a.u. of energy	E_h	$4.3597482 \cdot 10^{-18}$ J
time	a.u. of time	$\frac{\hbar}{E_h}$	$2.418884 \cdot 10^{-17}$ s
velocity	a.u. of velocity	$\frac{a_0 E_h}{\hbar}$	$2.187691 \cdot 10^{6}$ m/s
momentum	a.u. of momentum	$\frac{\hbar}{a_0}$	$1.992853 \cdot 10^{-24}$ kg · m/s
electric dipole moment	a.u. of electric dipole	ea_0	$8.478358 \cdot 10^{-30}$ C · m (= 2.5415 D)
magnetic dipole	Bohr magneton	$\frac{e\hbar}{2m_0 c}$	$0.92731 \cdot 10^{-20}$ erg/gauss
polarizability		$\frac{e^2 a_0^2}{E_h}$	$1.648778 \cdot 10^{-41}$ C^2m^2/J
electric field		$\frac{E_h}{ea_0}$	$5.142208 \cdot 10^{11}$ V/m
Boltzmann constant		k_B	$1.380658 \cdot 10^{-23}$ J/K
Avogadro constant		N_A	$6.0221367 \cdot 10^{23}$ mol^{-1}

Table 2. Conversion coefficients.

	a.u.	**erg**	**eV**	**kcal/mol**	**1 cm^{-1}**	**1 Hz**	**1 K**
1 a.u.	1	$4.35916 \cdot 10^{-11}$	27.2097	627.709	$2.194746 \cdot 10^{5}$	$6.579695 \cdot 10^{15}$	$3.15780 \cdot 10^{5}$
1 erg	$2.29402 \cdot 10^{10}$	1	$6.24197 \cdot 10^{11}$	$1.43998 \cdot 10^{13}$	$5.03480 \cdot 10^{15}$	$1.50940 \cdot 10^{26}$	$7.2441 \cdot 10^{15}$
1 eV	$3.67516 \cdot 10^{-2}$	$1.60206 \cdot 10^{-12}$	1	23.0693	$8.06604 \cdot 10^{3}$	$2.41814 \cdot 10^{14}$	$1.16054 \cdot 10^{4}$
1 kcal/mol	$1.59310 \cdot 10^{-3}$	$6.9446 \cdot 10^{-14}$	$4.33477 \cdot 10^{-2}$	1	$3.49644 \cdot 10^{2}$	$1.048209 \cdot 10^{13}$	$5.0307 \cdot 10^{2}$
1 cm^{-1}	$4.556336 \cdot 10^{-6}$	$1.98618 \cdot 10^{-16}$	$1.23977 \cdot 10^{-4}$	$2.86005 \cdot 10^{-3}$	1	$2.997930 \cdot 10^{10}$	1.43880
1 Hz	$1.519827 \cdot 10^{-16}$	$6.62517 \cdot 10^{-27}$	$4.13541 \cdot 10^{-15}$	$9.54009 \cdot 10^{-14}$	$3.335635 \cdot 10^{-11}$	1	$4.7993 \cdot 10^{-11}$
1 K	$3.16676 \cdot 10^{-6}$	$1.38044 \cdot 10^{-16}$	$8.6167 \cdot 10^{-5}$	$1.98780 \cdot 10^{-3}$	0.69502	$2.08363 \cdot 10^{10}$	1

IDEAS OF QUANTUM CHEMISTRY

Volume 2: Interactions

Volume 2 in a 2-volume expansion of this accessible guide to key interactions and tools used in the field of quantum chemistry

- Gives deep insight into the physical phenomena behind quantum chemistry
- Features a practical range of quantum chemical problems
- Uses informal language and unique structure to make complex topics accessible and digestible

Ideas of Quantum Chemistry, Volume 2: Interactions highlights the key structures and motions described by quantum chemistry and introduces the models and tools used to assess them. Beginning with a review of the molecular orbital model applied afterwards to periodic systems, the book goes on to explore areas such as the correlation of electronic motions, density functional theory (DFT), molecules subject to electric and magnetic fields as well as the all important intermolecular interactions and chemical reactions. Using an innovative structure to show the logical relationships between different topics, systems, and methods, this book makes of quantum chemistry an organized yet open entity, answers commonly asked questions, and emphasizes knowledge with practical examples throughout.

The 2nd volume in this 2-volume expansion, *Ideas of Quantum Chemistry: Interactions*, has been updated and revised to cover the latest developments in the field and can be used either on its own as a guide to key interactions and tools or in combination with volume 1 to give a complete overview of the field. It is an authoritative introduction for beginners, a useful tool for specialists in advanced areas of quantum chemistry, and a practical overview for all those working in related circles, from computational chemists and material scientists to biotechnologists and drug designers.

Professor Lucjan Piela received his bachelor's degree in 1960 from the historic Konarski College in his home town of Rzeszów, Poland. In 1965, he graduated with a Master of Science from the University of Warsaw and, after obtaining his Ph.D. from the same university 5 years later, went on to become a professor in 1976. In addition to his work in Warsaw, he has carried out research in the Centre Européen de Calcul Atomique et Moléculaire (France), Facultés Universitaires de Namur (Belgium), and Cornell University (USA). In addition to authoring multiple papers published in international journals, professor Piela is an elected member of the Académie Royale des Sciences, Lettres et Beaux-Arts de Belgique, and a member of the European Academy of Sciences.

ISBN 978-0-444-64248-6

9 780444 642486

ELSEVIER elsevier.com/books-and-journals